

УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

ЗБОРНИК РАДОВА

VII СИМПОЗИЈУМ „МАТЕМАТИКА И ПРИМЕНЕ”

4. и 5. новембар 2016.



УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

ЗБОРНИК РАДОВА – VII СИМПОЗИЈУМ „МАТЕМАТИКА И ПРИМЕНЕ”

4. и 5. новембар 2016.

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:

Марек Светлић

Корице:

Славиша Радовић, ГеоГебра Центар Београд

Штампа:

Донат Граф

Тираж:

100 примерака

СИР - Каталогизација у публикацији –

Народна библиотека Србије, Београд

51-7(082)

371.3::51(082)

СИМПОЗИЈУМ Математика и примене (7 ; 2016 ; Београд)

Зборник радова / VII симпозијум Математика и примене,

4. и 5. новембар 2016. ; [организатори] Универзитет

у Београду, Математички факултет [и Српска академија

наука и уметност ; уредник Миљан Кнежевић]. – Београд

: Универзитет, Математички факултет, 2017 (Београд

: Донат Граф). - 104 стр.: илустр. ; 30 см

Радови на срп. и енгл. језику. - Текст лат. и ћир.

- Тираж 100. - Напомене уз радове. - Библиографија уз сваки рад.

ISBN 978-86-7589-122-2

1. Математички факултет (Београд)

а) Математика - Зборници б) Математика - Настава - Зборници

COBISS.SR-ID 250139916

ПРЕДГОВОР

Седми симпозијум „МАТЕМАТИКА И ПРИМЕНЕ”, национални скуп са међународним учешћем, одржан је 4. и 5. новембра 2016. 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, 4. новембра у свечаној сали САНУ. На отварању Симпозијума, учеснике и госте су поздравили Миодраг Матељевић, дописни члан САНУ и председник Програмског одбора Симпозијума, декан Математичког факултета проф. др Зоран Ракић - члан Програмског одбора Симпозијума и др Миљан Кнежевић - члан организационог 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 углавном посвећена актуелним темама које се тичу примена 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.

Седмом симпозијуму „Математика и примене“ је присуствовало око 250 учесника и гостију. Кроз секције Симпозијума, своје резултате представило је око 90 истраживача из реномираних научноистраживачких институција из земље и иностранства. У раду секције Математика и информатика у образовању“ активно је учествовало око 120 професора математике и информатике из основних и средњих школа широм Србије. Са задовољством можемо констатовати да су испуњени главни циљеви скupa: 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 VIII Симпозијуму „Математика и примене“, који ће се одржати 17. и 18. новембра у организацији Математичког факултета Универзитета у Београду и Српске академије наука и уметности.

Програмски одбор VII 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 Зоран Петровић,**
Универзитет у Београду, ванредни професор Математичког факултета,
- **др Борислав Гајић,**
биши научни сарадник, Математички институт САНУ,
- **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
- **доц. др Миљан Кнежевић,**
Универзитет у Београду, доцент Математичког факултета
- **доц. др Марко Обрадовић,**
Универзитет у Београду, доцент Математичког факултета

Организациони одбор VII Симпозијума:

- **доц. др Миљан Кнежевић,**
Универзитет у Београду, доцент Математичког факултета - председник одбора,
- **проф. др Зорица Станимировић,**
Универзитет у Београду, ванредни професор Математичког факултета,
- **доц. др Александра Делић,**
Универзитет у Београду, доцент 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,
- **Душко Вишић,**
Универзитет у Београду, Рачунарска лабораторија Математичког факултета,
- **Катарина Лукић,**
Универзитет у Београду, студент Математичког факултета,

У Београду, октобра 2017.

САДРЖАЈ

1. AHLFORS-SCHWARZ LEMMA, HYPERBOLIC GEOMETRY, THE CARATHÉODORY AND KOBAYASHI METRICS Miodrag Mateljević	1
2. CONORMAL LAGRANGIAN FLOER HOMOLOGY FOR OPEN SUBSETS AND PSS ISOMORPHISM Jelena Katić, Darko Milinković, Jovana Nikolić	42
3. PROBLEM RASPOREĐIVANJA VOZILA PRI OPTIMIZACIJI TROŠKOVA TRANSPORTA POLJOPRIVREDNIH SIROVINA Ana Anokić, Zorica Stanimirović, Đorđe Stakić, Tatjana Davidović	49
4. ТЕОРИЈА БРОЈЕВА – ДЕЉИВОСТ, ПРОСТИ БРОЈЕВИ, КОНГРУЕНЦИЈЕ ПО МОДУЛУ Владимир Балтић	61
5. RAZVOJ GEOMETRIJSKOG MIŠLJENJA KROZ TANGRAM AKTIVNOSTI Nives Baranović, Sanja Lehman	81
6. REALIZACIJA MUPAD PROCEDURE ZA ODREĐIVANJE MOORE–PENROSE-OVOG INVERZA MATRICE Ivana Jovović	93

Ahlfors-Schwarz lemma, Hyperbolic geometry, the Carathéodory and Kobayashi Metrics

Miodrag Mateljević

Faculty of mathematics, University of Belgrade, Studentski Trg 16, Belgrade, Serbia
e-mail: miodrag@matf.bg.ac.rs

Abstract. In this paper we will give a review about some properties of hyperbolic metrics and various versions of Schwarz lemma.

Keywords: Hyperbolic metric; Ultrahyperbolic metric; Quasiregular harmonic mappings; Negatively curved metrics; Bers spaces.

1. Introduction

Throughout this paper, \mathbb{D} will denote the unit disc $\{z : |z| < 1\}$, \mathbb{T} the unit circle $\{z : |z| = 1\}$ and we will use notation $z = x + iy$ and $z = re^{i\theta}$, where $r = |z|$ and $\theta \in \mathbb{R}$ are polar coordinates. For a function h , we use notation $\partial h = \frac{1}{2}(h'_x - ih'_y)$ and $\bar{\partial}h = \frac{1}{2}(h'_x + ih'_y)$; we also use notations $D^c h$ and $\bar{D}^c h$ instead of ∂h and $\bar{\partial}h$ respectively when it seems convenient. By h'_x and h'_y we denote partial derivatives with respect to x and y respectively. We write $D_{zz}^2 h = D(\bar{D}h)$, where $Dh = D^c h$ and $\bar{D}h = \bar{D}^c h$.

Probably the best known equivalent of Euclid's parallel postulate, contingent on his other postulates, is Playfair's axiom, named after the Scottish mathematician John Playfair, which states:

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. Hyperbolic geometry was created in the first half of the nineteenth century in the midst of attempts to understand Euclid's axiomatic basis for geometry. It is one type of non-Euclidean geometry, that is, a geometry that discards one of Euclid's axioms (Euclid's parallel postulate). 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.

Hyperbolic geometry is tightly related to the function theory of one and several complex variables. Using Schwarz's lemma it is proved that

(A) holomorphic functions do not increase the corresponding hyperbolic distances between the corresponding hyperbolic domains.

The Caratheodory and Kobayashi metrics have proved to be important tools in the function theory of several complex variables. In particular, we have:

(B) If G_1 and G_2 are domains in C^n and $f : G_1 \rightarrow G_2$ holomorphic function, then f does not increase the corresponding Caratheodory(Kobayashi) distances.

But they are less familiar in the context of one complex variable. Krantz [31] gathers in one place the basic ideas about these important invariant metrics for domains in the plane and provides some illuminating examples and applications.

In [54], Wong proved:

(a) If G is a hyperbolic manifold in the sense of Kobayashi and the differential Kobayashi metric K_G is of class C^2 , then the holomorphic curvature of K_G is greater than or equal to -4 .

(b) If G is Carathéodory-hyperbolic and the differential Caratheodory metric C_G is of class C^2 , then the holomorphic curvature of C_G is less than or equal to -4 . With this result the author obtain an intrinsic characterization of the unit ball. For (b) see also Burbea [12].

In [27], Earle, Harris, Hubbard and Mitra discuss the Carathéodory and Kobayashi pseudometrics and their infinitesimal forms on complex Banach manifolds. Their discussion includes a very elementary treatment of the Kobayashi pseudometric as an integral of its infinitesimal form. They also prove new distortion theorems for the Carathéodory pseudometric under holomorphic maps from the open unit disk to a complex Banach manifold.

Although this is mainly review paper we treat known results with novelty and outline a few new results. The content of the paper is as follows. In Section 2, we outline how to introduce hyperbolic distances from

the point of complex analysis (more precisely, using Schwarz's lemma). We also shortly consider versions of Ahlfors-Schwarz lemma related to ultrahyperbolic metric and the comparison principle related to curvatures and distances.

In Section 3 we consider hyperbolic geometry, Möbius transformations and Cayley-Klein model in several variables. It is supposedly classical and can be found in the literature that the restriction of the Beltrami-Klein metric on the ball of \mathbb{R}^n to any minimal surface (minimal with respect to the flat metric) has curvature ≤ -1 . Using a heuristic argument we give application to minimal surfaces.

The properties (a) and (b) of holomorphic curvature of Kobayashi and Caratheodory metric are considered in Section 6.

Several years ago, the author communicated at Belgrade seminar (probably around 1980 -1990), some results related to the Carathéodory and Kobayashi pseudometrics and their infinitesimal forms on complex Banach spaces (see also [36]). In Section 7 we outline our approach, which is probably known to the experts in the subject (in particular see Theorems 6 and 7)¹.

2. Schwarz lemma and Hyperbolic geometry

2.1. The Schwarz lemma, Introduction

Throughout this paper by $\mathbb{S}(a, b)$ we denote the set $(a, b) \times \mathbb{R}$, $-\infty \leq a < b \leq \infty$, and in particular $\mathbb{S}_0 = \mathbb{S}(-1, 1)$. Note that $\mathbb{S}(a, b)$ is a strip if $-\infty < a < b < \infty$ and $\mathbb{S}(a, +\infty)$ is a half-plane if a is a real number, and $\mathbb{S}(-\infty, +\infty) = \mathbb{C}$.

If w is complex number by $\Re w$ (or $u = \operatorname{Re} w$) we denote the corresponding real part, and in a similar way if f is complex-valued function by $\Re f$ (or $u = \operatorname{Re} f$) we denote the corresponding real valued function and by $\nabla f(z) = (f'_x, f'_y)$ the gradient of f .

Occasionally by λ_0 and ρ_0 we denote respectively hyperbolic metric on the unit disk and on the strip \mathbb{S}_0 . See [61] and also [62] for discussion in this subsection.

The following result is a corollary of the maximum modulus principle:

Proposition 2.1 (classical Schwarz lemma 1-the unit disk). Suppose that $f : \mathbb{D} \rightarrow \mathbb{D}$ is an analytic map and $f(0) = 0$. The classic Schwarz lemma states : $|f(z)| \leq |z|$ and $|f'(0)| \leq 1$.

It is interesting that this result(which looks simple and elementary at first glance) has far reaching applications and forms. In this paragraph we follow [39, 43]. We will see below, if we do not specify the value of $f(0)$, we get Pick's lemma.

Pick's lemma. Suppose $f : \mathbb{D} \rightarrow \mathbb{D}$ is holomorphic. Then

$$|f'(z)| \leq \frac{1 - |f(z)|^2}{1 - |z|^2},$$

for any $z \in \mathbb{D}$. Equality holds for some $z \in \mathbb{D}$ if and only if f is a conformal self-map of \mathbb{D} (and in that case equality holds everywhere). Pick's lemma leads naturally to the hyperbolic metric on \mathbb{D} (see also Section 9).

2.1.1. the subordination principle

(A) Let $f(z)$ and $g(z)$ be analytic functions in \mathbb{D} , $f(z)$ is said to be subordinate to $g(z)$ in \mathbb{D} written, or $f \prec g$ ($z \in \mathbb{D}$), if there exists a function $w(z)$, analytic in \mathbb{D} with $w(0) = 0$ and $|w(z)| < 1$, ($z \in \mathbb{D}$) such that (i) $f(z) = g(w(z))$, ($z \in \mathbb{D}$).

(B) In particular, if the function $g(z)$ is univalent in \mathbb{D} , the above subordination is equivalent to (ii) $f(0) = g(0)$ and $f(\mathbb{D}) \subset g(\mathbb{D})$.

In the setting (B), then $f(\mathbb{D}_r) \subset g(\mathbb{D}_r)$, for all $0 < r < 1$ and $|f'(0)| \leq |g'(0)|$, where $\mathbb{D}_r = \{z : |z| < r\}$.

Since g is one-to-one, g is in fact a conformal map from \mathbb{D} to $g(\mathbb{D})$. Let $h = g^{-1}$ be the inverse of g . Then $h \circ f$ is holomorphic and maps \mathbb{D} into \mathbb{D} with $(h \circ f)(0) = 0$ since $f(0) = g(0)$. By Schwarz's lemma, we have $|(h \circ f)'(0)| \leq 1$ and $|(h \circ f)(z)| \leq |z|$ for all $z \in \mathbb{D}$. Hence $f(\mathbb{D}_r) \subset g(\mathbb{D}_r)$. By the chain rule, $(h \circ f)'(0) = \frac{f'(0)}{g'(0)}$. So the first inequality gives $|f'(0)| \leq |g'(0)|$.

As an exercise, formulate the condition for equality. Here we give a simple example.

¹Because of limited time author could not settle some details related to this paper.

Example 1. (i) Let $R = \{z : \operatorname{Re}(z) > 0\}$ be the right half plane. Let $f : \mathbb{D} \rightarrow R$ be holomorphic. We claim that $|f'(0)| \leq 2\operatorname{Re}(f(0))$.

(ii) Let $f : \mathbb{D} \rightarrow \mathbb{S}_0$ be holomorphic. Then $|f'(0)| \leq \frac{4}{\pi}$.

(iii) Let $f : \mathbb{D} \rightarrow \mathbb{D} \setminus \{0\}$ be holomorphic. Then $|f'(0)| \leq \frac{2}{e}$.

Proposition 2.2. Suppose (a) ϕ is univalent in \mathbb{D} , f holomorphic in \mathbb{D} and $f(\mathbb{D}) \subset \phi(\mathbb{D})$.

(b) $\phi(z_0) = f(z_0)$, $z_0 \in \mathbb{D}$. Then $|f'(z_0)| \leq |\phi'(z_0)|$.

2.2. The Schwarz lemma

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

$$\varphi_a(\zeta) = \frac{\zeta - a}{1 - \bar{a}\zeta}. \quad (2.1)$$

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

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

In the literature the notation T_a is also used instead of φ_a . Here we define $T_a = -\varphi_a$ and therefore we have $T_a^{-1} = T_a$.

Proposition 2.3 (Schwarz lemma 1). Suppose that $f : \mathbb{D} \rightarrow \mathbb{D}$ is an analytic map and $f(0) = 0$. The classic Schwarz lemma states : $|f(z)| \leq |z|$ and $|f'(0)| \leq 1$.

Proof. 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 $f : \mathbb{D} \rightarrow \mathbb{D}$ 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,

$$|F(\zeta)| = |\varphi_w \circ f \circ \varphi_{-z}(\zeta)| \leq |\zeta|, \quad \zeta \in \mathbb{D}, \quad (2.2)$$

and $|F'(0)| \leq 1$. Hence, since $F'(0) = \varphi'_w(w)f'(z)\varphi'_{-z}(0)$, we find

$$\frac{|f'(z)|}{1 - |f(z)|^2} \leq \frac{1}{1 - |z|^2}, \quad z \in \mathbb{D}, \quad (2.3)$$

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

Hence, equality holds in (2.3) if and only if f is a Möbius transformation of \mathbb{D} onto itself. \square

Let ω be an arbitrary point in \mathbb{D} and $\zeta = \varphi_z(\omega)$, then $\varphi_{-z}(\zeta) = \omega$, and by (2.2), we find $|\varphi_w(f(\omega))| \leq |\varphi_z(\omega)|$.

It is convenient to introduce a pseudo-distance

$$\delta(z, \omega) = |\varphi_z(\omega)| = \left| \frac{z - \omega}{1 - \bar{\omega}z} \right|, \quad (2.4)$$

which is a *conformal invariant*. Thus we get:

Proposition 2.4. The following inequality

$$\delta(f(z), f(\omega)) \leq \delta(z, \omega) \quad (2.5)$$

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

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

$$ds = \lambda(z)|dz| = \frac{2|dz|}{1 - |z|^2}, \quad (2.6)$$

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

$$|\gamma|_{\text{hyp}} = \int_{\gamma} \frac{2|dz|}{1-|z|^2}$$

and $|f \circ \gamma|_{\text{hyp}} = |\gamma|_{\text{hyp}}$, 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 λ ; 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 $h \in T_z$, $A \in \text{Aut}(\mathbb{D})$ and $h_* = A'(z)h$, then $\|h_*\|_{\lambda} = \|h\|_{\lambda}$ for every $z \in \mathbb{D}$ and every $h \in T_z$.

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

$$\lambda(0, r) = \int_0^r \frac{2dt}{1-t^2} = \ln \frac{1+r}{1-r}. \quad (2.7)$$

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

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

$$\lambda(z, \omega) = \ln \frac{1 + \left| \frac{z-\omega}{1-z\bar{\omega}} \right|}{1 - \left| \frac{z-\omega}{1-z\bar{\omega}} \right|}. \quad (2.8)$$

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

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

Exercise 1. Check the formula (2.7).

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

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

2.3. 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 $\text{Aut}D$.

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

$$f = \phi \circ f_0, \quad \phi \in \text{Aut } D. \quad (2.9)$$

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

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

Example 3. Describe $\text{Aut}(\mathbb{H})$. If $A \in \text{Aut}(\mathbb{H})$, 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 ∞ to ∞ . 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 $b_1 = \lambda + sx_0$.

Therefore, every $A \in \text{Aut}(\mathbb{H})$ can be represented in the form

$$f(z) = \frac{az + b}{cz + d}, \quad \text{where } a, b, c, d \in \mathbb{R} \text{ and } D = D(f) = ad - bc = 1. \quad (2.10)$$

If A is represented by (2.10), then

$$Az - \overline{A}\bar{z} = \frac{z - \bar{z}}{|cz + d|^2}. \quad (2.11)$$

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

There is another way to describe $\text{Aut}(\mathbb{H})$ 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 $w = (z, x_2, x_3, x_4)$. \square

If $L \in \text{Aut}(\mathbb{H})$, 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

$$(z_1, \overline{z_1}; z_2, \overline{z_2}) = (Lz_1, L\overline{z_1}; Lz_2, L\overline{z_2}) = (w_1, \overline{w_1}; w_2, \overline{w_2}). \quad (2.12)$$

Set $Tz = \frac{z - z_2}{z - \overline{z_2}}$. 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 z_1 and $\overline{z_1}$ with respect to \mathbb{R} onto points $T(z_1)$ and $T(\overline{z_1})$ 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

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

It is invariant with the group $\text{Aut}(\mathbb{H})$ because of (2.12) and $\delta_{\mathbb{H}}^2(z_1, z_2) = (z_1, \overline{z_1}, z_2, \overline{z_2})$. We will give another proof of this fact in subsection on Schwarz lemma(below).

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 $w \rightarrow z$ we get an infinitesimal invariant $ds = |\mathrm{d}z|/y$ (we drop multiple 2), where $y = \mathrm{Im}z$. For a piecewise continuously differentiable path $\gamma(t) = (x(t), y(t))$, $0 \leq t \leq 1$, in \mathbb{H} , we define $|\gamma|_{\text{hyp}} = \int \mathrm{d}z/y = \int_0^1 \frac{|\gamma'(t)|}{y(t)} dt$. We use this infinitesimal form to obtain Poincaré distance between two points p and q in \mathbb{H} by putting

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

where the infimum is taken over all paths γ 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 $d_{\text{hyp}} = d_{\text{hyp}, \mathbb{H}}$ if it is clear that our considerations is related to \mathbb{H} .

Proposition 2.5. In half-plane model, geodesics are the arcs of circles orthogonal to the real axis. The pseudo-hyperbolic distance and the hyperbolic distance are related by

$$\delta = \tanh(\lambda/2).$$

Proof. To find geodesic which joins p and q we use $A \in \text{Aut}(\mathbb{H})$ 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 I_0 that connects iy_1 and iy_2 . If γ is a path which joins iy_1 and iy_2 , using obvious geometric interpretation, we find $|\gamma|_{\text{hyp}} \geq |I_0|_{\text{hyp}}$ and hence

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

It is interesting to prove this inequality directly (without geometric interpretation). We outline a proof. Suppose that $y_1 \leq y_2$. Since

$|\gamma|_{\text{hyp}} = \int_{\gamma} |\mathrm{d}z|/y = \int_0^1 \frac{|\gamma'(t)|}{y(t)} dt$, we have

$$|\gamma|_{\text{hyp}} \geq \int_0^1 \frac{|y'(t)|}{y(t)} dt \geq \int_0^1 \frac{y'(t)}{y(t)} = \ln y(t)|_0^1 = \ln(y_2/y_1).$$

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 $A(z) = \frac{z - a_1}{z - a_2}$, then $\det(1, -a_1; 1, -a_2) = a_1 - a_2 > 0$ and therefore $A \in \text{Aut}(\mathbb{H})$. Hence it maps K on one half of the imaginary axis. If $A(p) = iy_1$ and $A(q) = iy_2$, the cross ratio ω 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 $y_2 \geq y_1$, we get

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

and since $\delta = \delta_{\mathbb{H}}$ is invariant with respect to $\text{Aut}(\mathbb{H})$, we find $\delta(iy_1, iy_2) = \delta((p, q))$.

Hence $e^\lambda = \frac{1+\delta}{1-\delta}$, i.e.

$$\delta = \tanh(\lambda/2).$$

In a similar way one can prove that this formula is valid if $y_2 < y_1$. \square

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

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

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

$$\delta(z, \omega) = |\varphi_z(\omega)| = \left| \frac{z - \omega}{1 - \bar{\omega}z} \right|, \quad (2.13)$$

which is a *conformal invariant*. It is easy to check that $\delta_{\mathbb{H}}(a, z) := |\phi_a(z)| = \delta_{\mathbb{D}}(T(a), T(z))$.

Moving on to the limit value when $\omega \rightarrow z$ 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 λ is connected with δ through $\delta = \tanh \frac{\lambda}{2}$.

There is another way of calculating that exhibits additivity.

Let γ 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 λ 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 remarks as:

Theorem 1.

$$\lambda_{\mathbb{D}} = \ln \frac{1 + \delta_{\mathbb{D}}}{1 - \delta_{\mathbb{D}}}, \quad \lambda_{\mathbb{H}} = \ln \frac{1 + \delta_{\mathbb{H}}}{1 - \delta_{\mathbb{H}}}. \quad (2.14)$$

2.4. Ahlfors-Schwarz lemma

It was noted by Pick that result can be expressed in invariant form. We refer the following result as Schwarz-Pick lemma.

Theorem 1 (Schwarz-Pick lemma). *Let F be an analytic function from a disk B to another disk U . Then F does not increase the corresponding hyperbolic (pseudo-hyperbolic) distances.*

2.5. Schwarz's Lemma for harmonic mappings

Recall that throughout this paper \mathbb{D} will denote the unit disc $\{z : |z| < 1\}$, \mathbb{T} the unit circle, $\{z : |z| = 1\}$ and we will use notation $z = x + iy$ and $z = re^{i\theta}$, where $r = |z|$ and $\theta \in \mathbb{R}$ are polar coordinates. For a function h , we use notation $\partial h = \frac{1}{2}(h'_x - ih'_y)$ and $\bar{\partial}h = \frac{1}{2}(h'_x + ih'_y)$; we also use notations $D^c h$ and $\bar{D}^c h$ instead of ∂h and $\bar{\partial}h$ respectively when it seems convenient. By h'_x and h'_y we denote partial derivatives with respect to x and y respectively. We write $D_{z\bar{z}}^2 h = D(\bar{D}h)$, where $Dh = D^c h$ and $\bar{D}h = \bar{D}^c h$.

Example 4. Let $\mathbb{S} = \{w : |\operatorname{Re} w| < 1\}$ and $\mathbb{S}_1 = \{w : |\operatorname{Re} w| < \pi/4\}$. The function $\tan : z \mapsto \tan(z)$ maps \mathbb{S}_1 onto \mathbb{D} . Let $B(w) = \frac{\pi}{4}w$ and $f_0 = \tan \circ B$, ie. $f_0(w) = \tan(\frac{\pi}{4}w)$. Then f_0 maps \mathbb{S} onto \mathbb{D} , $u = \frac{\sqrt{2}}{\pi} \arg A_0(iz)$, $v = \frac{\sqrt{2}}{\pi} \arg A_0(z)$, $t = \frac{\sqrt{2}}{\pi} \ln |A_0(z^2)|$.

Let $r < 1$, $A_0(z) = \frac{1+z}{1-z}$ and let $\phi = i \frac{2}{\pi} \ln A_0$. That is $\phi = \phi_0 \circ A_0$, where $\phi_0(\zeta) = i \frac{2}{\pi} \ln(\zeta)$. Let $\hat{\phi}$ be defined by $\hat{\phi}(z) = -\phi(iz)$. Note that $\hat{\phi} = \frac{4}{\pi} \arctan$ is the inverse of f_0 . Then, $\operatorname{Re} \hat{\phi}(z) = \frac{2}{\pi} \arg \frac{1+iz}{1-iz}$ and $|\operatorname{Re} \phi(z)| \leq \frac{4}{\pi} \tan^{-1} |z|$.

Let F be analytic such that $\operatorname{Re} h = \operatorname{Re} F$ on \mathbb{D} with $F(0) = 0$.

By subordination, we show that $|\operatorname{Re} F(z)| \leq \frac{4}{\pi} \arctan(|z|)$.

Example $h(z) = \frac{4}{\pi} \operatorname{Re}(\arctan z) + iay$, $a \in \mathbb{R}$, shows that we can not control growth of h in general at 0. The function h maps \mathbb{D} into \mathbb{S} , but $|(\operatorname{Im} h)_y(0)| = |a|$ can be arbitrarily large.

But if $f : \mathbb{D} \rightarrow \mathbb{D}$ is harmonic and $f(0) = 0$, then the maximal distortion (i) $L_f(0) \leq \frac{4}{\pi}$. (ii) In particular, if f is conformal at 0, then $|f'(0)| \leq \frac{4}{\pi}$. The estimate (i) is sharp.

It seems that if f is conformal at 0, then $|f'(0)| < \frac{4}{\pi}$.

By a normal family argument there is extremal function for the problem: (iii) $D(0) = \sup\{|L'_f(0)|\}$, where supremum is taken over all harmonic maps $f : \mathbb{D} \rightarrow \mathbb{D}$ with $f(0) = 0$. But, extremal functions $f_0(z) = \frac{2}{\pi} \arg \frac{1+z}{1-z}$ maps \mathbb{D} onto $(-1, 1)$.

Lemma 2.1. *Let $\mathbb{S} = \{w : |\operatorname{Re} w| < 1\}$ and let $h : \mathbb{U} \rightarrow \mathbb{D}$ be a harmonic mapping with $h(0) = 0$. Then $|\operatorname{Re} h(z)| \leq \frac{4}{\pi} \tan^{-1} |z|$.*

Proof. Let $M_h(r) = \max\{|h(z)| : z \in T_r\}$. Then there is $z_r = re^{i\alpha}$ such that $R = M_h(r) = |h(z_r)|$. If $h(z_r) = Re^{i\beta}$ and $H(z) = e^{-i\beta}h(e^{i\alpha}z)$, then $M_h(r) = M_H(r) = R$. By Lemma 2.2, $M_H(r) \leq \frac{4}{\pi} \arctan r$ and the proof follows. \square

Lemma 2.2 (Schwarz lemmma for harmonic functions [17]). *Let $\mathbb{S} = \{w : |\operatorname{Re} w| < 1\}$ and let $h : \mathbb{D} \rightarrow \mathbb{S}$ be a harmonic mapping with $h(0) = 0$. Then $|\operatorname{Re} h(z)| \leq \frac{4}{\pi} \tan^{-1} |z|$ and this inequality is sharp for each point $z \in \mathbb{D}$.*

Proof. Let A_1 be defined by $z \mapsto A_0(iz)$. Then A_1 carries the segment $[-i, i]$ onto half circle $T^+ = \{w : |w| = 1, \operatorname{Re} w \geq 0\}$ and $\frac{4}{\pi} \arctan r = \phi_1(r) = \frac{2}{\pi} \arg A_1(r) = -\frac{2}{\pi} \alpha$.

Observe now that the linear fractional mapping $w = \frac{1+z}{1-z}$ carries the circle $K_r : |z| = r < 1$ onto the circle $K_R : |w - w_0| = R$ with center $w_0 = (1+r^2)/(1-r^2)$ and radius $R = 2r/(1-r^2)$. Let $r < 1$, $A_0(z) = \frac{1+z}{1-z}$, $s = A_0(r)$, $R = \frac{s-s^{-1}}{2} = \frac{2r}{1-r^2}$ and α be the maximum of $|\arg w|$ on K_R ; therefore since $w_0^2 - R^2 = 1$, $\tan \alpha = R$ and $\alpha(r) = \arctan R = 2 \arctan r$; recall $\mathbb{S} = \{w : |\operatorname{Re} w| < 1\}$ and let $\phi = i \frac{2}{\pi} \ln A_0$. We prove, if $h : \mathbb{D} \rightarrow \mathbb{S}$ is harmonic, $h(0) = 0$, then $|h(z)| \leq \frac{2}{\pi} \alpha(|z|)$.

The linear fractional transformation A maps the circle $|z| = r$ onto the circle $K(a, R)$, where $a = (1+r^2)/(1-r^2)$ and $R = \frac{2r}{1-r^2}$; and therefore the disk \mathbb{D}_r onto the disk $B(a; R)$ of radius R with the center at a .

Thus,

- (1) $|\arg A_1|$ is bounded by $\alpha(r) = 2 \arctan r$ on \mathbb{D}_r and therefore since $\operatorname{Re} \phi = -\frac{2}{\pi} \arg A_0$,
- (2) $|\operatorname{Re} \phi|$ is bounded by $\alpha(r) = \frac{4}{\pi} \arctan r$ on \mathbb{D}_r .

Thus, (1) says that A_0 maps \mathbb{D}_r in the angle of opening $2\alpha(r) = 4 \arctan r$.

Let F be analytic such that $\operatorname{Re} h = \operatorname{Re} F$ on \mathbb{D} with $F(0) = 0$. By subordination $F(\mathbb{D}_r) \subset \phi_0(B(a; R))$. Hence, by (2) (recall $|\arg z| \leq 2 \arctan r$ on $B(a; R)$),

- (3) $|\operatorname{Re} \phi|$ is bounded by $\alpha(r) = \frac{4}{\pi} \arctan r$ on \mathbb{D}_r .

From (3), it follows

$$|\operatorname{Re} h(z)| = |\operatorname{Re} F(z)| \leq \frac{4}{\pi} \arctan |z|.$$

□

2.5.1. Rado-Kneser-Choquet

Example 5. Let $h(z) = x + i(x^2 - y^2)$. Then $J_h = -2y$, $h(i) = -1$, $h(H) = D = \{(x, y) : y < x^2\}$ and $h(z) = h(\bar{z})$; Since $h(z_1) = h(z_2)$ implies $x_1 = x_2$ and $y_1 = \pm y_2$, h is univalent on the upper half plane. Consider triangle Δ_1 with vertices $A = 0$, $B = 2$ and $C = 1 + i$ and Δ_2 with vertices $A = 0$, $B = 2$ and $D = 1 - i/2$. Quadrilateral $\Delta_3 = \Delta_1 \cup \Delta_2$, $\overline{D} = 1 + i/2$ and let quadrilateral Δ_4 consist of points $A = 0$, \overline{D} , $B = 2$ and C .

Verify that $\mathbf{L}_0 = h(\partial\Delta_3) = h(\partial\Delta_4)$, $h(\Delta_3) \neq \operatorname{int}(h(\partial\Delta_3))$ and that $h(\partial\Delta_3)$ is not a convex set.

Consider a conformal mapping φ of the unit disk onto Δ_3 and $\check{h} = h \circ \varphi$. Check that a \check{h} is harmonic on \mathbb{D} , a homeomorphism of \mathbb{T} onto \mathbf{L}_0 , but \check{h} is not 1-1 onto \mathbb{D} .

Choquet showed For every Jordan domain D which is not convex, there exists a homeomorphism $\phi : \mathbb{T} \rightarrow \partial D$ such that $h = P[\phi]$ is not a homeomorphism in \mathbb{D} .

Theorem 2. Assume that $\Omega \subset \mathbb{R}^2$ is a convex domain with smooth boundary $\partial\Omega$. Given any homeomorphism $\varphi : S^1 \rightarrow \partial\Omega$, there exists a unique harmonic map $h : \mathbb{D} \rightarrow \Omega$ such that $h = \varphi$ on S^1 and h is a diffeomorphism.

Proof. Let $h = (u, v)$. It suffices to show that $\det J(h) \neq 0$. Suppose that $\det J(h)$ is zero at z_0 , that is

$$\begin{vmatrix} u'_x(z_0) & u'_y(z_0) \\ v'_x(z_0) & v'_y(z_0) \end{vmatrix} = 0.$$

Thus vectors $X = (u'_x(z_0), u'_y(z_0))$ and $Y = (v'_x(z_0), v'_y(z_0))$ are linearly dependent and therefore there exists $(\alpha, \beta) \neq (0, 0)$ such that $\alpha X + \beta Y = 0$, that is $U'_x = 0$, $U'_y = 0$ at z_0 , where $U = \alpha u + \beta v$. Let $L = \{z : U(z) = U(z_0)\}$. Since U is a real harmonic function, there is an analytic function F such that $U = \operatorname{Re} F$ in \mathbb{D} and that $F'(z_0) = 0$. By the maximum principle for harmonic functions, no pair of the arcs of L emanating from z_0 can rejoin elsewhere in $\overline{\mathbb{D}}$. Since a neighborhood of z_0 consists of at least four arcs emanating from z_0 , and each of these arcs must extend out to the boundary, which means that L must meet T in at least four distinct points (that is $L \cap \Delta$ contains at least 4 points). On the other hand, h maps L into the line, which meets $\partial\Omega$ in exactly two points because of the assumption that Ω is convex. It follows that h maps at least four points on T onto two points in $\partial\Omega$, contradicting the hypothesis (φ being 1-1). This contradiction proves that the Jacobian cannot vanish in \mathbb{D} , so h is locally univalent. An application of the argument principle completes the proof. □

(A-1) If ϕ conformal mapping of a planar domain D onto \mathbb{D} , we define the ϕ -hyperbolic density on D by $\operatorname{Hyp}_{\mathbb{D}}(\phi z)|\phi'(z)| = \lambda_{\phi, D}(z)$. If another ϕ_1 conformal mapping of the domain D onto \mathbb{D} , set $w = \phi(z)$, $w_1 = \phi_1(z)$, $\omega = \phi_1 \circ \phi^{-1}$ and $w_1 = \omega(w)$. Then $\phi_1 = \omega \circ \phi$, and by the composition rule $\phi'_1(z) = \omega'(w)\phi'(z)$. Since $\omega \in \operatorname{Aut}(\mathbb{D})$, $1 - |w_1|^2 = |\omega'(w)|(1 - |w|^2)$ and hence $1 - |w_1|^2 = |\phi'_1(z)/\phi'(z)|(1 - |w|^2)$.

Therefore $\lambda_{\phi, D} = \lambda_{\phi_1, D}$ and the definition of the hyperbolic density is independent of conformal maps from D onto \mathbb{D} ; and we write $\operatorname{Hyp}_D(z)$ for the hyperbolic density on D at z .

Exercise 3. (I-1) If G and D are simply connected domains different then \mathbb{C} and f conformal mapping of D onto G , then $\operatorname{Hyp}_G(fz)|f'(z)| = \operatorname{Hyp}_D(z)$.

Outline. Let ψ be conformal mappings of the domain D onto \mathbb{D} , $g = f^{-1}$ and $\psi_1 = \psi \circ g$; set also $z_1 = f(z)$. Then $\lambda_D(z) = \lambda_0(\psi z)|\psi' z|$ and $\lambda_G(z) = \lambda_0(\psi_1 z_1)|\psi' z_1|$.

Exercise 4. Suppose that D is a simply connected domain different than \mathbb{C} and ω holomorphic map from D into itself with $\omega(z_0) = z_0$ for some $z_0 \in D$. Then $|\omega'(z_0)| \leq 1$.

If $G_1 = kG$, then $H(z) = kz$ maps G onto G_1 and $H^{-1}(w) = w/k$. Hence $\text{Hyp}_{G_1}(w) = \frac{1}{k} \text{Hyp}_G(w/k)$.

Example 6. 1. If $\Pi = \{w : \text{Re } w > 0\}$, then $B_0(w) = \frac{1-w}{1+w}$ maps Π on \mathbb{D} . Since $1 - |B_0(w)|^2 = \frac{4u}{|1+w|^2}$, where $u = \text{Re } w$, $B'_0(w) = 2(1+w)^{-2}$, and $\text{Hyp}_\Pi(w) = \lambda_0(B_0(w))|B'_0(w)|$, we find

$$\text{Hyp}_\Pi(w) = \frac{1}{\text{Re } w}.$$

2. Since $e = \exp$ maps $S = \{y : |y| < \pi/2\}$ onto Π , we have $\lambda_S(z) = \text{Hyp}_\Pi(e^z)|e^z| = \frac{1}{\cos y}$.

3. Let λ_0 be a hyperbolic density on \mathbb{S}_0 . Then

$$\lambda_0(w) = \text{Hyp}_{\mathbb{S}_0}(w) = \frac{\pi}{2} \frac{1}{\cos(\frac{\pi}{2}u)}. \quad (2.15)$$

4. $\lambda_0(iy_1, iy_2) = \frac{\pi}{2}|y_2 - y_1|$, $y_1, y_2 \in \mathbb{R}$.

5. If $a, b \in \mathbb{R}$, $a < b$, the linear map L defined by $L(w) = \frac{2w-(a+b)}{b-a}$, maps $\mathbb{S}(a, b)$ onto \mathbb{S}_0 and $\rho(w) = \rho_0(Lw)\frac{2}{b-a}$. Hence

$$\rho(w) = \text{Hyp}_{\mathbb{S}(a,b)}(w) = \frac{\pi}{(b-a)} \frac{1}{\cos\left(\frac{\pi}{2}[(2u - (a+b))/(b-a)]\right)}. \quad (2.16)$$

Example 7. Check $\arctan(\frac{2r}{1-r^2}) = 2\arctan(r)$, $0 < r < 1$.

Outline. If $\tan\beta = r$, then $\tan(2\beta) = \frac{2\tan\beta}{1-(\tan\beta)^2}$.

Example 8. For $w_1, w_2 \in \mathbb{S}_0$, $\rho_0(u_1, u_2) \leq \rho_0(w_1, w_2)$.

A plane region D whose complement has at least two points we call a hyperbolic plane domain. On a hyperbolic plane domain there exists a unique maximal ultrahyperbolic metric, and this metric has constant curvature -1 .

Using holomorphic covering $\pi : \mathbb{D} \rightarrow D$, one can define the pseudo-hyperbolic and the hyperbolic metric on D .

2.6. Hyperbolic domains

Now we outline how we can use powerful tools which yield the uniformization theorem to get hyperbolic version of Ahlfors-Schwarz lemma.

Let W and W^* be surfaces and $f : W^* \rightarrow W$ a continuous surjective map such that for every $p \in W$, there exists an open neighborhood V of p , such that is a union of disjoint open sets in W^* and every component of $p^{-1}(V)$ is in one-to-one correspondence with V . When this is so The map f is called the covering map and the pair (W^*, f) is called a covering surface of W . A deck transformation or automorphism of a cover $f : W^* \rightarrow W$ is a homeomorphism $A : W^* \rightarrow W^*$ such that $f \circ A = f$. The set of all deck transformations of A forms a group under composition, the deck transformation group $\text{Aut}(f)$. Deck transformations are also called covering transformations.

In particular, if W and W^* are Riemann surfaces and $f : W^* \rightarrow W$ holomorphic, we call f the holomorphic covering map. If W^* is simply connected, we call the pair (W^*, f) a universal covering.

The uniformization theorem says that every simply connected Riemann surface is conformally equivalent to a disk, the complex plane, or the Riemann sphere. In particular it implies that every Riemann surface admits a Riemannian metric of constant curvature. Every Riemann surface is the quotient of the deck transformation group (a free, proper and holomorphic action of a discrete group on its universal covering) and this universal covering is holomorphically isomorphic (one also says: "conformally equivalent" or "biholomorphic") to the Riemann sphere, the complex plane and the unit disk in the complex plane. If the universal covering of a Riemann surface S is the unit disk we say that S is hyperbolic. Using holomorphic covering $\pi : \mathbb{D} \rightarrow S$, one can define the pseudo-hyperbolic and the hyperbolic metric on S . In particular, if $S = G$ is hyperbolic planar domain we can

use

(I-2) $\text{Hyp}_G(\pi z)|\pi'(z)| = \text{Hyp}_{\mathbb{D}}(z)$ and

(I-3) If G and D are hyperbolic domains and f conformal mapping of D onto G , then $\text{Hyp}_G(fz)|f'(z)| = \text{Hyp}_D(z)$.

Proposition 2.6 (Schwarz lemma, 1-planar hyperbolic domains, Ahlfors-Schwarz; hyperbolic version). (a) If G and D are conformally isomorphic to \mathbb{D} and $f \in \text{Hol}(G, D)$, then

$$\delta_D(fz, fz') \leq \delta_G(z, z'), \quad z, z' \in G.$$

(b) The result holds more generally: if G and D are hyperbolic domains and $f \in \text{Hol}(G, D)$, then

$$\text{Hyp}_D(fz, fz') \leq \text{Hyp}_G(z, z'), \quad z, z' \in G.$$

(c) If $z \in G$, $\mathbf{v} \in T_z \mathbb{C}$ and $\mathbf{v}^* = df_z(\mathbf{v})$, then

$$|\mathbf{v}^*|_{\text{Hyp}} \leq |\mathbf{v}|_{\text{Hyp}}.$$

For a hyperbolic planar domain G the Carathéodory distance $C_G \leq \lambda_G$ with equality if and only if G is a simply connected domain.

2.7. Schwarz lemma at the boundary for holomorphic functions

In this subsection we discuss some known results related to the subject.

2.7.1. Jack's lemma

In connection with Jack's lemma we state:

(T1) Let f be nonconstant analytic function on the unit disk and $f(0) = 0$. Then for given $r \in (0, 1)$ the function $|f|$ attains maximum at some point $z_0 \in T_r$. Then, $z_0 f'(z_0) = kf(z_0)$, for some $k \geq 1$. Using homothety, rotation and translation we consider the following:

(T2) Let $B = B(a_0; a_0)$, $a_0 > 0$, f be analytic function on \overline{B} , $f(B) \subset B$ and $f(0) = 0$. Then $f'(0) = k$, where $k > 0$. Contrary, if $f'(0) = ke^{i\alpha}$, where $0 < \alpha < 2\pi$, then easily we can show that there is a small arc L on the boundary of B centered at the origin such that $f(L)$ is out of \overline{B} .

Q1. If D is domain and f is analytic function on \overline{D} , $f(D) \subset D$ and there is $z_0 \in \partial D$ such that $f(z_0) = z_0$. Whether $f'(z_0) > 0$? It seems that using similar approach as the above we can prove that answer to Q1 is positive if ∂D is smooth at z_0 .

Proposition 2.7. a) Let $f : \mathbb{D} \rightarrow \mathbb{D}$. Assume that (H0) there is a point $b \in \mathbb{T}$ so that f extends continuously to b , $|f(b)| = 1$ (say that $f(b) = c$) and f is \mathbb{R} -differentiable at b .

b) Further assume that there is a function A such that $A : [0, 1] \rightarrow [0, 1]$, $A'(1)$ exists and $M_f(r) \leq A(r)$. Then $|\lambda_f(b)| \geq A'(1)$.

Proposition 2.8. Under the above hypothesis, if there exists $f'(b)$, then

(i) $|f'(b)| \geq A'(1)$.

Proposition 2.9. Let $f(z) = b + c_p(z - a)^p + c_{p+1}(z - a)^{p+1} + \dots$, $c_p > 0$, $p \geq 1$ be a holomorphic function of the disc \mathbb{D} into self and let z_1, z_2, \dots, z_n be zeros of the function $f(z)$ in \mathbb{D} that are different from 0. If we assume (H0), then $|f'(b)| \geq p + \sum_{k=1}^n \frac{1-a_k}{1+a_k}$, where $a_k = |z_k|$.

Proof. Set

$$B_1(z) = z^p \prod_{k=1}^n \frac{z - z_k}{1 - \bar{z}_k z},$$

$B_0(z) = z^p$, $B = B_0 B_1$ and $F = f/B$. Further, set $a_k = |z_k|$, $T_k(r) := \frac{r+|z_k|}{1+r|z_k|}$ and $B_+(r) = \prod_{k=1}^n \frac{r+|z_k|}{1+r|z_k|}$. By Schwarz lemma, $|f| \leq |B|$. Hence $M_f(r) \leq A(r) := r^p B_+(r)$. Without loss of generality we can suppose that $b = 1$. Then $A'(1) = pB_+(1) + B'_+(1)$, Hence, since $T_k(1) = 1$ and $T'_k(1) = \frac{1-a_k}{1+a_k}$, we find $A'(1) = p + B'_+(1)$. For example if $n = 2$, we have $B'_+(1) = \frac{1-a_1}{1+a_1} + \frac{1-a_2}{1+a_2}$. In general, $B'_+(1) = \sum_{k=1}^n \frac{1-a_k}{1+a_k}$. \square

Theorem 3. Let $f : \mathbb{D} \rightarrow \mathbb{D}$ be harmonic. Assume that $f(0) = 0$. Further assume that there is a point $b \in \mathbb{T}$ so that f extends continuously to b , $|f(b)| = 1$ (say that $f(b) = b'$) and f is \mathbb{R} -differentiable at b . Then $|\lambda_f(b)| \geq 2/\pi$.

In [49], R. Osserman offered the following boundary refinement of the classical Schwarz lemma. It is very much in the spirit of the sort of result that we wish to consider here.

Theorem 4. Let $f : \mathbb{D} \rightarrow \mathbb{D}$ be holomorphic. Assume that $f(0) = 0$. a) Further assume that there is a point $b \in \mathbb{T}$ so that f extends continuously to b , $|f(b)| = 1$ (say that $f(b) = c$) and $f'(b)$ exists. Then $|f'(b)| \geq \frac{2}{1+|f'(0)|}$.
c) If f has a zero of order p at 0 , then (ii) $|f'(b)| \geq p$.

Set $k = |f'(0)|$, $g(z) = f(z)/z$ and $F = T_k \circ g$. Then $g = T_{-k} \circ F$, $M_{T_{-k}}(r) \leq T_{-k}(r) = \frac{r+k}{1+rk}$ and therefore $M_f(r) \leq A(r) := r^p \frac{r+k}{1+rk}$.

Hence, since $A'(1) = \frac{2}{1+k}$, we have (i). Let $f : \mathbb{D} \rightarrow \mathbb{D}$ be holomorphic and satisfy $f(0) = 0$. Then $|f(\zeta)| \leq |\zeta| \frac{|\zeta| + |f(0)|}{1+|f(0)||\zeta|}$ for $|\zeta| < 1$. c) $M_f(r) \leq A(r) := r^n$. Hence, since $A'(1) = n$, we have (ii).

Set $r = |z|$ and $k = |f'(0)|$. Then $1 - |f(z)| \geq 1 - r \frac{r+k}{1+rk} = \frac{1-r^2}{1-rk}$ and therefore $\frac{1-|f(z)|}{1-r} \geq \frac{1+r}{1+rk}$. Hence $|f'(b)| \geq \frac{2}{1+|f'(0)|}$. Without loss of generality we reduce the proof to the case $b = 1$ and $f(1) = 1$. By Schwarz lemma $|f(z)| \leq |z|$. Hence $|1 - f(x)| \geq |1 - x|$.

Theorem 5. Let $f : \mathbb{D} \rightarrow \mathbb{D}$ be holomorphic function. Suppose that there is an extension of f at $b \in \mathbb{T}$ such that $|f(b)| = 1$ and there exists $f'(b)$. Then

$$|f'(b)| \geq \frac{2(1 - |f(0)|)}{1 - |f(0)|^2 + |f'(0)|}. \quad (2.17)$$

Now we state a version of Lowner and Velling result.

Proposition 2.10. Let $f : \mathbb{D} \rightarrow \mathbb{D}$ be holomorphic and $f(0) = 0$. Let $S \subset \mathbb{T}$ be a nontrivial arc, and suppose that f extends continuously to S . Further assume that $f(S)$ lies in \mathbb{T} . Let s denote the length of S and σ the length of $f(S)$ (which is also necessarily an arc, since it is a connected subset of the circle). Then $\sigma \geq s \frac{2}{1+|f'(0)|}$.

Proof: By Schwarz reflection, we may take it that f is analytic on the interior of the arc S . Hence it certainly satisfies the hypotheses of the first lemma at each point of the interior of S . The conclusion of that lemma then holds, and integration yields the desired result.

Theorem 6. Let $f : \mathbb{D} \rightarrow \mathbb{D}$ be holomorphic. Further assume that there is a point $b \in \mathbb{T}$ so that f extends continuously to b , $|f(b)| = 1$ (say that $f(b) = c$) and $f'(b)$ exists. Then

$$|f'(b)| \geq \frac{2(1 - |f(0)|)^2}{1 - |f(0)|^2 + |f'(0)|}.$$

Suppose f is an analytic map of \mathbb{D} into itself. If $|b| < 1$, we say b is a fixed point of f if $f(b) = b$. If $|b| = 1$, we say b is a fixed point of f if $\lim_{r \rightarrow 1^-} f(rb) = b$. Julia-Caratheodory Theorem implies If b is a fixed point of f with $|b| = 1$, then $\lim_{r \rightarrow 1^-} f'(rb)$ exists (call it $f'(b)$) and $0 < |f'(b)| \leq \infty$.

Perhaps, we can restate the hypothesis. The following boundary version of the Schwarz lemma was proved in 1938 by Unkelbach in [Un] and then rediscovered and partially improved by Osserman in 2000 [50].

Theorem 7. In addition to hypothesis of Theorem 4, suppose that $f = c_p z^p + o(z^p)$, if z tends 0. Then (iii.1) $|f'(c)| \geq p + \frac{1-|c_p|}{1+|c_p|}$. The equality in (1.7) holds if and only iff f is of the form $f = -z^p \varphi_a$ on \mathbb{D} , for some constant $a \in (-1; 0]$.

Outline: Set $c = c_p$, $k = |c_p|$, $g(z) = f(z)/z^p$ and $F = T_c \circ g$. Then $g = T_{-c} \circ F$, $M_{T_{-c}}(r) \leq T_{-c}(r) = \frac{r+k}{1+rk}$ and therefore $M_f(r) \leq A(r) := r^p \frac{r+k}{1+rk}$.

Hence, since $A'(r) = pr^{p-1}T_{-c}(r) + r^p T'_{-c}(r)$ and $T'_{-c}(r) = (1-k^2)(1+rk)^{-2}$, we have $A'(1) = p + \frac{1-k}{1+k}$, and therefore (iii.1).

The inequality (iii.1) is a particular case of a result due to Dubinin in (see [16]), who strengthened the inequality $|f'(c)| \geq 1$ by involving zeros of the function f .

Suppose (iii.2): Let $f(z) = b + c_p(z-a)^p + c_{p+1}(z-a)^{p+1} + \dots$, $c_p > 0$, $p \geq 1$, be a holomorphic function in the disc \mathbb{D} satisfying $f(a) = b$, $|a| \leq 1$ and

(c) $|f(z) - \alpha| < \alpha$ for $|z| < 1$, where α is a positive real number and $1/2 < \alpha \leq 1$, and $f(z) - b$ has no zeros in \mathbb{D} , except $z = a$. Assume that, for some $c \in \mathbb{T}$, f has an angular limit $f(c)$ at c , $f(c) = 2\alpha$.

There are several papers by B. Örnek and Örnek-Akyel (see for example [48] and [47]) related to the subject. In [47], under the hypothesis (iii.2) the optimal lower estimate for $|f'(c)|$ are obtained, and the following functions are used. Let z_1, z_2, \dots, z_n be zeros of the function $f(z) - b$ in \mathbb{D} that are different from $z = a$. Set

$$B(z) = z^p \prod_{k=1}^n \frac{z - z_k}{1 - \bar{z}_k z},$$

$\phi = \varphi_d$, where $d = f(0)$, $\Upsilon = \phi/B$ and $\kappa = \varphi_e$, where $e = \Upsilon(0)$. Set $B_0(z) = z^p$, $p = \phi/B_0$ and $\Theta(z) = \frac{\ln p(z) - \ln p(0)}{\ln p(z) + \ln p(0)}$. Set $F = F_\alpha = f - \alpha$. Then the hypothesis (c) can be rewritten as (c1): F maps the unit disc into the disc of radius α and $F(c) = \alpha$. The auxiliary function Θ is a holomorphic in the unit disc \mathbb{T} , $|\Theta(z)| < 1$, $\Theta(0) = 0$ and $|\Theta(b)| = 1$ for $b \in \mathbb{T}$. Apply

$$\frac{2}{1 + |\Theta'(0)|} \leq |\Theta'(b)|.$$

Theorem 8 (Burns-Krantz [11]). *Let g be an analytic function of the unit disk \mathbb{D} into self which satisfy (i) $g(z) = z + O(1-z)^4$, when z approaches 1 throughout \mathbb{D} . Then $g = Id$.*

The result is the sharpest possible. Indeed, since $g(z) = z - (z-1)^3/10$ maps the unit disc to itself, this example shows that the exponent 4 in the theorem cannot be replaced by 3. The proof in fact shows that 4 can be replaced by $o((z-1)^3)$.

In [25], a new theory of regular functions over the skew field of Hamilton numbers (quaternions) and in the division algebra of Cayley numbers (octonions) has been recently introduced by Gentili and Struppa (Adv. Math. 216(2007) 279-301). For these functions, among several basic results, the analogue of the classical Schwarz lemma has been already obtained. In this paper, following an interesting approach adopted by Burns and Krantz in the holomorphic setting, they prove some boundary versions of the Schwarz Lemma and Cartan's Uniqueness Theorem for regular functions. We are also able to extend to the case of regular functions most of the related "rigidity" results known for holomorphic functions.

2.8. Denjoy-Wolff Theorem

The Schwarz Lemma is related to the following result.

Theorem (Denjoy-Wolff 1926): Let \mathbb{D} be the open unit disk in \mathbb{C} and let f be a holomorphic function mapping \mathbb{D} into \mathbb{D} which is not an automorphism of \mathbb{D} (i.e. a Möbius transformation). Then there is a unique point z in the closure of \mathbb{D} such that the iterates of f tend to z uniformly on compact subsets of \mathbb{D} . If z lies in \mathbb{D} , it is the unique fixed point of f . The mapping f leaves invariant hyperbolic disks centered on z , if z lies in \mathbb{D} , and disks tangent to the unit circle at z , if z lies on the boundary of \mathbb{D} .

When the fixed point is at $z = 0$, the hyperbolic disks centred at z are just the Euclidean disks with centre 0. Otherwise f can be conjugated by a Möbius transformation so that the fixed point is zero. An elementary proof of the theorem is given below, taken from Shapiro (1993) and Burckel (1981). Two other short proofs can be found in Carleson & Gamelin (1993).

Definition. Suppose f is an analytic map of \mathbb{D} into itself. If $|b| < 1$, we say b is a fixed point of f if $f(b) = b$. If $|b| = 1$, we say b is a fixed point of f if $\lim_{r \rightarrow 1^-} f(rb) = b$.

Julia-Caratheodory Theorem implies that if b is a fixed point of f with $|b| = 1$, then $\lim_{r \rightarrow 1^-} f'(rb)$ exists (call it $f'(b)$) and $0 < f'(b) \leq \infty$.

(Denjoy-Wolff 1926): (a) If f is an analytic map of \mathbb{D} into itself, not the identity map, there is a unique fixed point, a , of f in $\overline{\mathbb{D}}$ such that $|f'(a)| \leq 1$.

(b) If f is not an automorphism of \mathbb{D} (i.e. a Möbius transformation) with fixed point in \mathbb{D} , iterates of f tend to a uniformly on compact subsets of \mathbb{D} .

This distinguished fixed point will be called the Denjoy-Wolff point of f .

The Schwarz-Pick Lemma implies f has at most one fixed point in \mathbb{D} and if f has a fixed point in \mathbb{D} , it must be the Denjoy-Wolff point. See: C. C. Cowen, Iteration and the Solution of Functional Equations for Functions Analytic in the Unit Disk, Trans. Amer. Math. Soc. 265 (1981) 69-95; C. C. Cowen and Ch. Pommerenke, Inequalities for the Angular Derivative of an Analytic Function in the Unit Disk, J. London Math. Soc. 26 (1982) 271-289.

Question 3. Is there a version of this result for quasi-regular mappings? For Danjoy-Wolff theorem see [18], [19]. It seems that using the analytic covering theorem (the uniformization theorem for hyperbolic domains) one can get a version of Theorem Danjoy-Wolff for hyperbolic domains.

Theorem Danjoy-Wolff. Let \mathbb{D} be the open unit disk in \mathbb{C} and let f be a holomorphic function mapping \mathbb{D} into \mathbb{D} which is not an automorphism of \mathbb{D} (i.e. a Möbius transformation). Then there is a unique point z_0 in the closure of \mathbb{D} such that the iterates of f tend to z_0 uniformly on compact subsets of \mathbb{D} . If z_0 lies in \mathbb{D} , it is the unique fixed point of f . The mapping f leaves invariant hyperbolic disks centered on z_0 , if z_0 lies in \mathbb{D} , and disks tangent to the unit circle at z_0 , if z_0 lies on the boundary of \mathbb{D} .

Proof. When the fixed point is at $z_0 = 0$, the hyperbolic disks centred at z_0 are just the Euclidean disks with centre 0. Otherwise f can be conjugated by a Möbius transformation so that the fixed point is zero. An elementary proof of the theorem is given below, taken from Shapiro (1993) and Burckel (1981). Two other short proofs can be found in Carleson & Gamelin (1993)[18].

Case 1(Fixed point in the disk). If f has a fixed point z in \mathbb{D} then, after conjugating by a Möbius transformation, it can be assumed that $z = 0$. Let $M(r)$ be the maximum modulus of f on $|z| = r < 1$. By the Schwarz lemma $|f(z)| \leq \delta(r)|z|$, for $|z| \leq r$, where $\delta(r) = \frac{M(r)}{r}$. Since f is not automorphism of \mathbb{D} , $\delta(r) < 1$. It follows by iteration that $|f^n(z)| \leq \delta(r)^n$ for $|z| \leq r$. These two inequalities imply the result in this case.

Case 2 (No fixed points in the unit disk). When f acts in \mathbb{D} without fixed points, Wolff showed that there is a point z_0 on the boundary such that the iterates of f leave invariant each disk tangent to the boundary at that point. Take a sequence r_k increasing to 1 and set $f_k(z) = r_k f(z)$. By applying Rouché's theorem to $f_k(z) - z$ and $g(z) = z$, f_k has exactly one zero z_k in \mathbb{D} . Passing to a subsequence if necessary, it can be assumed that $z_k \rightarrow z_0$. The point z_0 cannot lie in \mathbb{D} , because, by passing to the limit, z_0 would have to be a fixed point. The result for the case of fixed points implies that the maps f_k leave invariant all Euclidean disks whose hyperbolic center is located at z_k . We leave the interested reader to fill details for proof of the following:

(I1) Explicit computations show that, as k increases, one can choose such disks so that they tend to any given disk tangent to the boundary at z_0 . By continuity, f leaves each such disk B invariant.

To see that f^n converges uniformly on compacta to the constant z_0 , it is enough to show that the same is true for any subsequence f^{n_k} , convergent in the same sense to a function g , say. Such limits exist by Montel's theorem, and if g is non-constant, it can also be assumed that $f^{n_{k+1}-n_k}$ has a limit, h say. Set $m_k = n_{k+1} - n_k$. But then $f^{n_{k+1}} = (f^{n_k})^{m_k}$ and $f^{n_k}(w) \rightarrow g(w)$ and $f^{n_{k+1}}(w) \rightarrow g(w)$. Hence since f^{m_k} holomorphic function does not increase hyperbolic distance on \mathbb{D} , we find $d(f^{m_k}(f^{n_k}(w)), f^{m_k}(g(w))) \leq d(f^{n_k}(w), g(w))$ and therefore $h(g(w)) = g(w)$ for w in \mathbb{D} .

Since h is holomorphic and $g(\mathbb{D})$ open, $h(w) = w$ for all w .

It can also be assumed that f^{m_k-1} is convergent to F say. But then $f^{m_k}(w) = f^{m_k-1}(fw) = f(f^{m_k-1}(w))$ and therefore $f(F(w)) = w = f(F(w))$, contradicting the fact that f is not an automorphism. Hence every subsequence tends to some constant uniformly on compacta in \mathbb{D} .

The invariance of B implies each such constant lies in the closure of each disk B , and hence their intersection, the single point z_0 . By Montel's theorem, it follows that f^n converges uniformly on compacta to the constant z_0 .

For the subject see: What are the most recent versions of The Schwarz Lemma at the Boundary? - ResearchGate. Available from: <https://www.researchgate.net/post/>. \square

2.9. Curvature

Let D be a domain in $z = x + iy$ -plane and a Riemannian metric be given by the fundamental form

$$ds^2 = \sigma |dz|^2 = \sigma(dx^2 + dy^2),$$

which is conformal with euclidian metric. If $M = (D, \sigma |dz|^2)$, then the Gaussian curvature of M is

$$K_M = -\frac{1}{2\sigma} \Delta \ln \sigma.$$

Instead of K_M it is also convenient to use notation K_σ and call σ shortly metric coefficient.

Often in the literature a Riemannian metric is given by $ds = \rho|dz|$, $\rho > 0$, that is by the fundamental form

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

In some situations it is convenient to call ρ shortly metric density.

If $\rho > 0$ is a C^2 function on D and $M = (D, \rho|dz|)$, the Gaussian curvature of M is expressed by the formula

$$K_M = \bar{K}_\rho := -\rho^{-2}\Delta \ln \rho.$$

We also call the above term the Gaussian curvature of a Riemannian metric density ρ on D . Also we write $K(\rho)$ and \bar{K}_ρ instead of K_ρ and $\bar{K}(\rho)$ respectively. It is clear that $\bar{K}_\rho = K(\rho^2)$.

For $a > 0$, $\bar{K}(a\rho) = a^{-2}\bar{K}(\rho)$.

Recall that a pseudohermitian metric on D is a non-negative upper semicontinuous function ρ such the set $\rho^{-1}(0)$ is discrete in D .

If u is an upper semicontinuous function on D and $\omega \in D$, the *lower generalized Laplacian* of u is defined by ([2], see also [24])

$$\Delta_L u(\omega) = 4 \liminf_{r \rightarrow 0} \frac{1}{r^2} \left[\frac{1}{2\pi} \int_0^{2\pi} (u(\omega + re^{it}) - u(\omega)) dt \right].$$

When u is a C^2 function, then the lower generalized Laplacian of u reduces to the usual Laplacian

$$\Delta u = u_{xx} + u_{yy}.$$

The Gaussian curvature of a pseudohermitian metric density on D is defined by the formula

$$K = \bar{K}_\rho = -\rho^{-2}\Delta_L \ln \rho.$$

For all $a > 0$ define the family of functions λ_a

$$\lambda_a(z) = \frac{2}{a(1 - |z|^2)}.$$

Also, it is convenient to write λ instead of λ_1 .

Suppose ρ is a semimetric density on a region G and $f : D \rightarrow G$ is a holomorphic function. The pull-back of ρ by f is $f^*(\rho) = \rho(f(z))|f'(z)|$.

Suppose $a \in D$, $f'(a) \neq 0$, $\rho(f(a)) > 0$ and ρ is of class C^2 at $f(a)$. Then $K_{f^*(\rho)}(a) = K_\rho(f(a))$. The Gaussian curvature of the density λ_a is $\bar{K}(\lambda_a) = -a^2$. This family of Hermitian metrics on \mathbb{D} is of interest because it allows an ordering of all pseudohermitian metrics on \mathbb{D} in the sense of the following ([2]).

Theorem 9. *Let ρ be a pseudohermitian metric density on \mathbb{D} such that*

$$\bar{K}_\rho(z) \leq -a^2$$

for some $a > 0$. Then $\rho \leq \lambda_a$.

This kind of estimate is similar to Ahlfors-Schwarz lemma. Ahlfors lemma can be found in Ahlfors [5].

A metric ρ is said to be ultrahyperbolic in a region Ω if it has the following properties :

- (a) ρ is upper semicontinuous; and
- (b) at every z_0 with $\rho(z_0) > 0$ there exists a supporting metric ρ_0 , defined and class C^2 in a neighborhood V of z_0 , such that $\rho_0 \leq \rho$ and $\bar{K}_{\rho_0} \leq -1$ in V , while $\rho_0(z_0) = \rho(z_0)$.

Theorem 10. (Ahlfors Lemma 1) Suppose ρ is an ultrahyperbolic metric on \mathbb{D} . Then $\rho \leq \lambda$.

The version presented in Gardiner [23] has a slightly modified definition of supporting metric. This modification and formulation is due to Earle. This version has been used (see [23]) to prove that *Teichmüller* distance is less than equal to *Kobayashi's* on *Teichmüller* space.

Ahlfors [5] proved a stronger version of Schwarz's lemma and Ahlfors lemma 1.

Theorem 11. (Ahlfors Lemma 2) Let f be an analytic mapping of \mathbb{D} into a region on which there is given ultrahyperbolic metric ρ . Then $\rho[f(z)]|f'(z)| \leq \lambda$.

The proof consists of observation that $\rho[f(z)]|f'(z)|$ is ultrahyperbolic metric on \mathbb{D} . Observe that the zeros of $f'(z)$ are singularities of this metric.

Note that if f is the identity map on \mathbb{D} we get Theorem 3 (Ahlfors lemma 1) from Theorem 4.

The notation of an ultrahyperbolic metric makes sense, and the theorem remains valid if Ω is replaced by a Riemann surface.

In a plane region Ω whose complement has at least two points, there exists a unique maximal ultrahyperbolic metric, and this metric has constant curvature -1 .

The maximal metric is called the *Poincaré metric* of Ω , and we denote it by λ_Ω . It is maximal in the sense that every ultrahyperbolic metric ρ satisfies $\rho \leq \lambda_\Omega$ throughout Ω .

The hyperbolic density (metric) of a disk $|z| < R$ is given by

$$\lambda_R(z) = \frac{2R}{R^2 - |z|^2}.$$

If ρ is ultrahyperbolic in $|z| < R$, then $\rho \leq \lambda_R$. In particular, if ρ is ultrahyperbolic in the whole plane, then $\rho = 0$. Hence there is no ultrahyperbolic metric in the whole plane.

The same is true of the punctured plane $C^* = \{z : z \neq 0\}$. Indeed, if ρ were ultrahyperbolic metric in the whole plane, then $\rho(e^z)|e^z|$ would be ultrahyperbolic in the whole plane. These are only cases in which ultrahyperbolic metric fails to exist.

Ahlfors [5] used Theorem 4 to prove Bloch and the Picard theorems. Ultrahyperbolic metrics (without the name) were introduced by Ahlfors. They found many applications in the theory of several complex variables.

The comparison principle.

Theorem 2 ([37]). If ρ and σ are two metrics (densities) on the disk \mathbb{D} , σ is complete and $0 > \bar{K}_\sigma \geq \bar{K}_\rho$ on \mathbb{D} , then $\sigma \geq \rho$.

Here, \bar{K} is Gaussian curvature. For the hyperbolic density on the disk we have $\bar{K} = -4$ (or -1 , depends of normalization).

Example 9. If σ is the Poincaré metric with $K_\sigma = -1$ and ρ is any other metric with $K_\rho \leq -1$, then $\rho \leq \sigma$. In particular this holds if $\rho = F^*(\sigma)$ for a holomorphic map $F : \mathbb{D} \rightarrow \mathbb{D}$ (that is, the map F must be conformal with respect to the complex structures induced by the respective metrics).

2.10. An inequality opposite to Ahlfors-Schwarz lemma

Mateljević [37] proved an estimate opposite to Ahlfors-Schwarz lemma.

A metric $H|dz|$ is said to be superhyperbolic in a region Ω if it has the following properties:

- (a) H is continuous (more general, lower semicontinuous) on Ω .
- (b) at every z_0 there exists a supporting metric (from above) H_0 , defined and class C^2 in a neighborhood V of z_0 , such that $H_0 \geq H$ and $K_{H_0} \geq -1$ in V , while $H_0(z_0) = H(z_0)$.

Theorem 12 ([37]). Suppose H is a superhyperbolic metric on \mathbb{D} for which

(c) $H(z)$ tends to $+\infty$ when $|z|$ tends to 1_- .

Then $\lambda \leq H$.

Proof. Let $\rho_r(z) = 2r(1 - |rz|^2)^{-1}$, where $r \in (0, 1)$, and let

$$\Psi_r(z) = \log |H(z)| - \log \rho_r(z).$$

By the hypothesis of theorem Ψ_r has a minimum on \mathbb{D} at a point z_0 . Let H_0 be supporting metric density from above to H at z_0 in a neighborhood V and

$$\tau_r(z) = \log |H_0(z)| - \log \rho_r(z).$$

Function τ_r has a minimum on V at z_0 and so

$$(1) \quad 0 \leq \Delta \tau_r(z_0) = \Delta \log |H_0(z_0)| - \Delta \log \rho_r(z_0).$$

By the hypothesis, we have

$$\bar{K}_{H_0}(z_0) = -H_0(z_0)^{-2}(\Delta \ln H_0)(z_0) \geq -1,$$

that is

$$(\Delta \ln H_0)(z_0) \leq H_0(z_0)^2,$$

and

$$(\Delta \ln \rho_r)(z_0) \leq (\rho_r(z_0))^2.$$

Hence by (1),

$$(2) \quad 0 \leq \Delta \tau_r(z_0) = \Delta \log |H_0(z_0)| - \Delta \log \rho_r(z_0) \leq H_0^2(z_0) - (\rho_r(z_0))^2$$

and therefore $\rho_r(z_0) \leq H_0(z_0)$. Since $H_0(z_0) = H(z_0)$ it follows that Ψ_r has non-negative minimum at z_0 and hence we conclude that $\rho_r \leq H$ for every $z \in \mathbb{D}$. If r tends to 1_- , we find $\rho \leq H$ on \mathbb{D} . \square

By applying a method developed by Yau in [58] (or by generalized maximum principle of Cheng and Yau [15]), it follows that this result holds if we suppose instead of (c) that

(d) H is a complete metric on \mathbb{D} .

Theorem 13. *If ρ and σ are two metrics (density) on \mathbb{D} , σ complete and $0 > \bar{K}_\sigma \geq \bar{K}_\rho$ on \mathbb{D} , then $\sigma \geq \rho$.*

This theorem remains valid if ρ is ultrahyperbolic metric and σ superhyperbolic metric on \mathbb{D} . Also, we can get further generalizations if \mathbb{D} is replaced by a Riemann surface.

Suppose that Ω is a hyperbolic domain and

- (a) $H_0 : \Omega \rightarrow (0, \infty)$ is continuous (more general, lower semicontinuous) on Ω ,
- (b) the generalized Gaussian curvature of H_0 , $K_{H_0} \geq -1$ on Ω .

Then $\lambda_\Omega \leq H$.

Curvature. Let D be a domain in $z = x + iy$ -plane and a Riemannian metric be given by the fundamental form

$$ds^2 = \sigma |dz|^2 = \sigma(dx^2 + dy^2),$$

which is conformal with euclidian metric. If $M = (D, \sigma |dz|^2)$, then the Gaussian curvature of M is

$$K_M = -\frac{1}{2\sigma} \Delta \ln \sigma.$$

Instead of K_M it is also convenient to use notation K_σ and call σ shortly metric coefficient.

Often in the literature a Riemannian metric is given by $ds = \rho|dz|$, $\rho > 0$, that is by the fundamental form

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

In some situations it is convenient to call ρ shortly metric density.

If $\rho > 0$ is a C^2 function on D and $M = (D, \rho|dz|)$, the Gaussian curvature of M is expressed by the formula

$$K_M = \bar{K}_\rho := -\rho^{-2}\Delta \ln \rho.$$

We also call the above term the Gaussian curvature of a Riemannian metric density ρ on D . Also we write $K(\rho)$ and \bar{K}_ρ instead of K_ρ and $\bar{K}(\rho)$ respectively. It is clear that $\bar{K}_\rho = K(\rho^2)$.

Proposition 2.11 ([64]). If $\rho, \rho_0, \tilde{\rho}$ metric densities on $B_0 = B(z_0; r_0)$. Suppose that $\eta := \frac{\tilde{\rho}}{\rho}$ has a local minimum at z_0 , $\bar{K}_{\tilde{\rho}} \geq \bar{K}_{\rho_0}$ at z_0 and $\rho_0 \geq \rho$ and $\rho(z_0) = \rho_0(z_0)$. Then $\rho \leq \tilde{\rho}$ on B_0 .

A metric ρ is said to be ultrahyperbolic in a region Ω if it has the following properties :

- (a) ρ is upper semicontinuous and
 - (b) at every z_0 , with $\rho(z_0) > 0$, there exists a supporting metric ρ_0 , defined and class C^2 in a neighborhood V of z_0 , such that $\rho_0 \leq \rho$ and $\bar{K}_{\rho_0} \leq -1$ in V , while $\rho_0(z_0) = \rho(z_0)$.
- Set $I_u(a, r) := \int_0^{2\pi} (u(a + re^{it}) - u(a)) dt$. If u is a C^2 function in a neighborhood V , then $u(a + re^{it}) - u(a) = Ar \cos t + Br \sin t + Dr^2 \cos^2 t + Er^2 \cos t \sin t + Fr^2 \cos^2 t + o(r^2)$, where $D = u_{xx}(a)/2$ and $F = u_{yy}(a)/2$. Hence $I_u(a, r) = \frac{\pi}{2} r^2 \Delta u(a) + o(r^2)$ and therefore $\frac{2}{\pi} r^{-2} I_u(a, r)$ tends to $\Delta u(a)$, if r tends to 0.

Definition 1. If u is an upper semicontinuous function, the *lower generalized Laplacian* of u is defined by ([2], see also [24])

$$\Delta_L u(\omega) = 4 \liminf_{r \rightarrow 0} \frac{1}{r^2} \left[\frac{1}{2\pi} \int_0^{2\pi} (u(\omega + re^{it}) - u(\omega)) dt \right].$$

When u is a C^2 function, then by (i) we conclude that the lower generalized Laplasian of u reduces to the usual Laplacian

$$\Delta u = u_{xx} + u_{yy}.$$

Definition 2. A region Ω is hyperbolic if $\mathbb{C} \setminus \Omega$ contains at least two points. The hyperbolic metric λ_Ω on Ω is the unique metric on Ω such that $\lambda_D(z) = \lambda_\Omega(z)|f'(z)|$, where $f : D \rightarrow \Omega$ is any holomorphic universal covering projection. The hyperbolic metric has constant curvature -1 .

Theorem 14 (Ahlfors Lemma 1). *Suppose ρ is an ultrahyperbolic metric on \mathbb{D} . Then (1) $\rho \leq \lambda$.*

Definition 3. A conformai metric $\rho(z)|dz|$ on a region Ω is called an SK metric provided $\rho : \Omega \rightarrow [0, +\infty)$ is upper semicontinuous and $\Delta_L \log \rho(a) \geq \rho^2(a)$ at each point $a \in \Omega$ such that $\rho(a) > 0$.

Thus, an SK metric is a conformai metric with generalized curvature at most -1 at each point where it does not vanish.

Here $\lambda_D|dz|$ is the hyperbolic metric on normalized to have curvature -1 . (In some references the curvature is taken to be -4 ; we will translate all such results to the context of curvature -1 without further comment.) Ahlfors did not show that equality in (1) at a single point implied $\rho = \lambda_D$ which would be the analog of the equality statement in Schwarz's lemma. Heins [2]² introduced the class of SK metrics, which includes ultrahyperbolic metrics, and verified that (1) remains valid for SK metrics. In addition, he showed that equality at a single point implied $\rho = \lambda_D$. However, his proof of the equality statement is not as elementary as the proof of Ahlfors' lemma since it relies on an integral representation for a solution of the nonlinear partial differential equation $\Delta u = \exp(2u)$.

In [45] D. Minda also considered the strong form of Ahlfors' lemma and present a relatively elementary proof of the equality statement for Ahlfors' lemma for SK metrics; it relies on the fact that the Laplacian of a real-valued function is nonpositive at any point where the function has a relative maximum. His proof is in the

²See D. Minda [45] for papers Heins [2], Hopf [3], Jorgensen [4].

spirit of Ahlfors' derivation of (1) and is a modification of a method introduced by Hopf [3] for linear partial differential equations. A related proof was given by Jorgensen [4] in the special case of metrics with constant curvature -1 .

Theorem 15. *Let Ω be a hyperbolic region in C and λ_Ω the hyperbolic metric on Ω . If $\rho(z)|dz|$ is an SK metric on Ω , then either $\rho(z) < \lambda_\Omega(z)$ for all $z \in \Omega$ or else $\rho(z) = \lambda_\Omega(z)$ for all $z \in \Omega$.*

Proposition 2.12 ([45]). Suppose G is a region in C , $u : G \rightarrow [-\infty, \infty)$ is upper semicontinuous and there is a positive constant K such that $\Delta u(z) \geq K u(z)$ at any point $z \in G$ with $u(z) > -\infty$. If $\limsup_{z \rightarrow \zeta} u(z) \leq 0$ for all $\zeta \in \partial G$, then either $u(z) < 0$ for all $z \in G$ or else $u(z) = 0$ for all $z \in G$.

Outline. Fix $a \in G$ and take $r > 0$ such that $B = B(a, r)$. There exists $M > 0$ such that $\rho \leq \lambda \leq M$ on B . Now $u = \ln(\rho/\lambda)$ is upper semicontinuous on B , $u(z) \leq 0$, for $z \in B$, and at any point $z \in B$ where $u(z) > -\infty$ (that is, where $\rho(z) > 0$) we have $\Delta u \geq \rho^2 - \lambda^2 \geq 2M(\rho - \lambda)$. Hence $\Delta u \geq 2M^2 u$. Theorem 1 implies that either $\rho(z) < \lambda_G(z)$ for all $z \in B$ or else $\rho(z) = \lambda_G(z)$ for all $z \in B$.

Note that H. L. Royden [The Ahlfors-Schwarz lemma: the case of equality, J. Analyse Math. 46 (1986), 261-270] also established the sharp form of Ahlfors' lemma by a different method.

Theorem 16 ([64]). *If ρ and σ are two metrics (density) on \mathbb{D} , σ complete and $0 > \bar{K}_\sigma \geq \bar{K}_\rho$ on \mathbb{D} (where K_σ and K_ρ are the generalized curvatures), then $\sigma \geq \rho$.*

The method of sub-solutions and super-solutions have been used in study harmonic maps between surfaces cf. [32].

3. Hyperbolic geometry, Möbius transformations and Cayley-Klein model in several variables

The unit sphere in three-dimensional space \mathbb{R}^3 is the set of points (x, y, z) such that $x^2 + y^2 + z^2 = 1$. Let $N = (0, 0, 1)$ be the "north pole", and let M be the rest of the sphere. The plane $z = 0$ contains the center of the sphere; the "equator" is the intersection of the sphere with this plane.

For any point P on M , there is a unique line through N and P , and this line intersects the plane $z = 0$ in exactly one point P' . Define the stereographic projection of P to be this point P' in the plane.

In Cartesian coordinates (x, y, z) on the sphere and (X, Y) on the plane, the projection and its inverse are given by the formulas

$$(X, Y) = \left(\frac{x}{1-z}, \frac{y}{1-z} \right),$$

$$(x, y, z) = \left(\frac{2X}{1+X^2+Y^2}, \frac{2Y}{1+X^2+Y^2}, \frac{-1+X^2+Y^2}{1+X^2+Y^2} \right).$$

If M is a surface in \mathbb{R}^3 space and suppose that f is a difeomorphism of \mathbb{D} onto M we can transfer the Poincaré disk model onto M . For example, L is line on M if $f^{-1}(L)$ is a U-line on \mathbb{D} . We define $d_{hyp, M}(p, q) = d_{hyp, \mathbb{D}}(f^{-1}(p), f^{-1}(q))$, $p, q \in M$. The disk model and M -model are isomorphic under f .

The inverse of the stereographic projection S maps the unit disk onto the hemisphere \mathbb{S}_+^2 and defines a \mathbb{S}_+^2 -hyperbolic model and an orthogonal (orthographic) projection of this model on xy-plane defines the Klein model on \mathbb{D} .

Thus, the two models are related through a projection on or from the hemisphere model.

Shortly, the Klein model is an orthographic projection of the hemisphere model, while the Poincaré disk model is a stereographic projection.

Given two distinct points U and V in the open unit ball of the model in Euclidean space, the unique straight line connecting them intersects the unit sphere at two ideal points A and B , labeled so that the points are, in order along the line, A, U, V, B . Taking the centre of the unit ball of the model as the origin, and assigning position vectors $\mathbf{u}, \mathbf{v}, \mathbf{a}, \mathbf{b}$ respectively to the points U, V, A, B , we have that that $|a - v| > |a - u|$ and $|u - b| > |v - b|$, where $|\cdot|$ denotes the Euclidean norm. Then the distance between U and V in the modelled hyperbolic space is expressed as

$$d(\mathbf{u}, \mathbf{v}) = \frac{1}{2} \log \frac{\|\mathbf{v} - \mathbf{a}\| \|\mathbf{b} - \mathbf{u}\|}{\|\mathbf{u} - \mathbf{a}\| \|\mathbf{b} - \mathbf{v}\|},$$

where the factor of one half is needed to make the curvature -1 .

We will prove below that on the unit ball in \mathbb{R}^n the associated metric tensor is given by the formula: if $\mathbf{v} \in T_{\mathbf{x}}\mathbb{R}^n$, then

$$ds^2(\mathbf{v}) = \text{Kle}(\mathbf{x}, \mathbf{v}) = \frac{\|d\mathbf{v}\|^2}{1 - \|\mathbf{x}\|^2} + \frac{(\sum_{k=1}^n x_k v_k)^2}{(1 - \|\mathbf{x}\|^2)^2}. \quad (3.1)$$

It is supposedly classical and can be found in the literature that the restriction of the Beltrami-Klein metric on the ball of \mathbb{R}^n to any minimal surface (minimal with respect to the flat metric) has curvature ≤ -1 . Unfortunately, the B-K metric is not conformally equivalent to the Euclidean one. Hence, a conformal minimal disk is not isothermal with respect to the B-K metric, and the pull-back is not a hermitian metric on the disk. Probably it is not even quasiconformal.

3.1. The Cayley-Klein model of hyperbolic geometry

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.

By arcosh and arsinh we denote inverses of hyperbolic functions:

$\text{arsinh } x = \ln(x + \sqrt{x^2 + 1})$, $\text{arcosh } x = \ln(x + \sqrt{x^2 - 1})$; $x \geq 1$. By (9.1), we find

$$\cosh d = \frac{1}{2}(e^d + e^{-d}) = \frac{1}{2}\left(\frac{1+\sigma}{1-\sigma} + \frac{1-\sigma}{1+\sigma}\right) = \frac{1+\sigma^2}{1-\sigma^2} = 1 + \frac{2\sigma^2}{1-\sigma^2},$$

where $d = d_{\text{hyp}, \mathbb{H}}$ and $\sigma = \delta_{\mathbb{H}}$. Hence,

$$\cosh d = 1 + 2 \frac{|z_1 - z_2|^2}{|z_1 - \bar{z}_2|^2 - |z_1 - z_2|^2},$$

and since $|z_1 - \bar{z}_2|^2 - |z_1 - z_2|^2 = 4y_1 y_2$, we find

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

Thus, in general, the distance between two points in \mathbb{H} measured in hyperbolic metric along such a hyperbolic geodesic is:

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

Distances in this model are Cayley-Klein metrics. Given two distinct points p and q inside the disk, the unique hyperbolic line connecting them intersects the boundary at two ideal points, a and b , label them so that the points are, in order, a, p, q, b and $|aq| > |ap|$ and $|pb| > |qb|$.

The hyperbolic distance between p and q is then $d(p, q) = \log \frac{|aq| |pb|}{|ap| |qb|}$.

Set $\{p, q\} = \frac{|aq| |pb|}{|ap| |qb|}$. If the ideal points, a and b , label them so that the points are, in order, a, p, q, r, b , then $\{p, q\} \{q, r\} = \{p, r\}$ and therefore $d(p, r) = d(p, q) + d(q, r)$.

The vertical bars indicate Euclidean length of the line segment connecting the points between them in the model (not along the circle arc), \log is the natural logarithm. Both the Poincaré disk model and the Klein disk model are models of the hyperbolic plane. An advantage of the Poincaré disk model is that it is conformal (circles and angles are not distorted); a disadvantage is that lines of the geometry are circular arcs orthogonal to the boundary circle of the disk. This section focuses on the projection of the unit sphere from the north pole onto the plane through the equator. Other formulations are treated in later sections.

The unit sphere in three-dimensional space \mathbb{R}^3 is the set of points (x, y, z) such that $x^2 + y^2 + z^2 = 1$. Let $N = (0, 0, 1)$ be the "north pole", and let M be the rest of the sphere. The plane $z = 0$ contains the center of the sphere; the "equator" is the intersection of the sphere with this plane.

For any point P on M , there is a unique line through N and P , and this line intersects the plane $z = 0$ in exactly one point P' . Define the stereographic projection of P to be this point P' in the plane.

In Cartesian coordinates (x, y, z) on the sphere and (X, Y) on the plane, the projection and its inverse are given by the formulas

$$(X, Y) = \left(\frac{x}{1-z}, \frac{y}{1-z} \right),$$

$$(x, y, z) = \left(\frac{2X}{1+X^2+Y^2}, \frac{2Y}{1+X^2+Y^2}, \frac{-1+X^2+Y^2}{1+X^2+Y^2} \right).$$

If M is a surface in \mathbb{R}^3 space and suppose that f is a difeomorphism of \mathbb{D} onto M we can transfer the Poincaré disk model onto M . For example, L is line on M if $f^{-1}(L)$ is a U-line on \mathbb{D} . We define $d_{hyp,M}(p, q) = d_{hyp,U}(f^{-1}(p), f^{-1}(q))$, $p, q \in M$. The disk model and M -model are isomorphic under f .

The inverse of the stereographic projection S maps the unit disk onto the hemisphere \mathbb{S}_-^2 and defines a \mathbb{S}_-^2 -hyperbolic model and an orthogonal (orthographic) projection of this model on xy -plane defines the Klein model on \mathbb{D} .

Thus, the two models are related through a projection on or from the hemisphere model.

Shortly, the Klein model is an orthographic projection of the hemisphere model, while the Poincaré disk model is a stereographic projection.

Let o be an orthographic projection defined by $o(y_1, y_2, y_3) = (y_1, y_2, 0)$ and denote by S the inverse of the stereographic projection. Then S maps the unit disk onto \mathbb{S}_-^2 and $S_1 = o \circ S$ the unit disk onto itself. S maps circles K orthogonal to \mathbb{T} onto circles $S(K)$ in \mathbb{S}_-^2 orthogonal to T and every $S(K)$ belongs to a plane parallel to e_3 . Let L be a plane parallel to e_3 and let the half-circle K be the intersection of L and \mathbb{S}_-^2 . If $a, b \in S(K)$, c and d are ideal point on K , and $a' = o(a)$ and $b' = o(b)$, then by similarity $|a - d|^2 = 2R|a' - d|$ and $|a - c|^2 = 2R|a' - c|$, where R is the radius of $S(K)$. Hence

(i) $|a', b', c, d| = |a, b, c, d|^2$. Now let $z, w \in U$ be points on the circle K and let points z^*, w^* be the intersection of the unit circle by the circle K . Since $S(z^*) = z^*$, $S(w^*) = w^*$ and the absolute cross ratio is invariant under Möbius,

(ii) $|z, w; z^*, w^*| = |Sz, Sw; z^*, w^*|$.

Let K_l (in honor of Klein) denote the inverse of S_1 . Note that K_l fixes the points on the unit circle \mathbb{T} .

Proposition 3.1. The distance in Klein model is $d_{kle}(z, w) = d_{hyp}(K_l(z), K_l(w))$ and it equals $\frac{1}{2} \ln |z, w; \hat{z}, \hat{w}|$, where \hat{z}, \hat{w} are the intersection of the unit circle by line zw .

When projecting the same lines in both models on one disk both lines go through the same two ideal points (the ideal points remain on the same spot) also the pole of the chord is the centre of the circle that contains the arc.

3.2. The Hyperbolic Metric and Möbius transformations

For Möbius transformations in several dimensions see [6]. By e_1, \dots, e_n we denote the coordinate unit vectors of \mathbb{R}^n . For example, $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, \dots, 0)$ and $e_3 = (0, 0, 1, \dots, 0)$. We denote by x_1, \dots, x_n the coordinates of a point $x \in \mathbb{R}^n$. Thus $x = (x_1, \dots, x_n)$ and $x = x_1e_1 + \dots + x_ne_n$. We denote by $\mathbb{R}_{\infty}^n = \overline{\mathbb{R}}^n = \mathbb{R}^n \cup \{\infty\}$ the one point compactification of \mathbb{R}^n . By $B^n(a; r)$ we denote ball $\{x \in \mathbb{R}^n : |x - a| < r\}$ and by $S^{n-1}(a; r)$ sphere $\{x \in \mathbb{R}^n : |x - a| = r\}$.

Möbius transformation is a mapping which is composition of a finite number of the following:

1. Translation: $f(x) = x + a$;
2. Stretching $f(x) = rx$, $r > 0$;
3. Orthogonal: f is linear and $|f(x)| = |x|$ for all $x \in \mathbb{R}^n$;
4. Inversion in a sphere $S = S(a; r)$: $J(x) = a + r^2 \frac{x-a}{|x-a|^2}$.

Every isometry of \mathbb{R}^n can be uniquely written as the composition $t \circ k$ where t is a translation and k is an isometry fixing the origin.

An $n \times n$ matrix A is called orthogonal if $A^T A = I_n$, or equivalently if $AA^T = I_n$. The geometric meaning of the condition $A^T A = I_n$ is that the columns of A are mutually perpendicular unit vectors (check!). Let $O(n) = O_n(R)$ denote the set of $n \times n$ orthogonal matrices.

The group of similarities consists of all mappings $x \mapsto mx + b$ where $b \in \mathbb{R}^n$ and m is a conformal matrix, i.e. $m = \lambda k$, with $\lambda > 0$ and $k \in O(n)$. Every Möbius can be expressed as a composite of inversions.

The reflection with respect to the unit sphere in \mathbb{R}^n is defined by

$$x \mapsto x^* = Jx = x/|x|^2, \quad J_0 = \infty, \quad J_\infty = 0.$$

The matrix $J'(x)$ has components $J'(x)_{ij} = \frac{1}{|x|^2}(\delta_{ij} - \frac{2x_i x_j}{|x|^2})$. We adapt a special notation for the matrix $Q(x)$ with entries $Q(x)_{ij} = \frac{x_i x_j}{|x|^2}$. This enables us to write $J'(x) = \frac{1}{|x|^2}(I - 2Q(x))$. This is an important formula. From $Q^2 = Q$ we obtain $(I - 2Q)^2 = I$. In higher dimensions, a Möbius transformation is a homeomorphism of $\overline{\mathbb{R}^n}$, the one-point compactification of \mathbb{R}^n , which is a finite composition of inversions in spheres and reflections in hyperplanes. Liouville's theorem in conformal geometry states that in dimension at least three, all conformal transformations are Möbius transformations. Every Möbius transformation can be put in the form

$$f(x) = b + \frac{\alpha A(x - a)}{|x - a|^\epsilon},$$

where $a, b \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$, A is an orthogonal matrix, and ϵ is 0 or 2. The group of Möbius transformations is also called the full Möbius group and denote by $\hat{M}(\mathbb{R}^n)$.

The orientation-preserving Möbius transformations form The sub-group of $\hat{M}(\mathbb{R}^n)$ which we denote by $M(\mathbb{R}^n)$ and which is the connected component of the identity in the Möbius group. For any $\gamma \in M(\mathbb{R}^n)$ we denote by $|\gamma'(x)|$ the positive number such that $\gamma'(x)/|\gamma'(x)| \in O(n)$. In other words, $|\gamma'(x)|$ is the linear change of scale at x which is the same in all directions. In higher dimensions, we define absolute cross ratio $[a, b, c, d] = \frac{|a-c|}{|a-d|} : \frac{|b-c|}{|b-d|}$, which is invariant $[\gamma a, \gamma b, \gamma c, \gamma d] = [a, b, c, d]$, $\gamma \in \hat{M}(\mathbb{R}^n)$. This is clear when γ is a similarity, and for J we obtain $|Jx - Jy|^2 = |J'(x)||J'(y)||x - y|^2$.

For $a \in B^n$ ($a \neq 0$), $R = R(a) = (|a^*|^2 - 1)^{1/2} = \sqrt{1 - |a|^2}/|a|$. Then $S^{n-1}(a^*, R(a))$ is orthogonal to the unit sphere S .

The reflection (inversion) with respect to this sphere is given by

$$\sigma_a x = a^* + R(a)^2(x - a^*)^*. \quad (3.2)$$

Define canonical mapping

$$T_a(x) = (I - 2Q(a))\sigma_a x. \quad (3.3)$$

The explicit expression for $T_a(x)$ is

$$T_a(x) = -a + (1 - |a|^2)(x^* - a)^* = \frac{(1 - |a|^2)(x - a) - |x - a|a}{[x, a]^2}, \quad (3.4)$$

where $[x, a] = |x||x^* - a| = |a||x - a^*|$.

If $\gamma \in M(\mathbb{R}^n)$ maps a in 0, then $\gamma = kv$, where $k \in O(n)$.

Let $x = (x_1, \dots, x_n)$ be the coordinates on \mathbb{R}^n . The Poincaré metric on the unit $\mathbb{B} \subset \mathbb{R}^n$ is given by

$$ds_{\mathbb{B}}^2 = \frac{4|dx|^2}{(1 - |x|^2)^2}.$$

It is conformally equivalent to the Euclidean metric. The 2-dimensional case $n = 2$ is the standard Poincaré metric on the unit disk $\mathbb{D} \subset \mathbb{R}^2 \cong \mathbb{C}$.

The Hyperbolic Metric: Let \mathbb{B}^3 be the unit ball $\{x \in \mathbb{R}^3 : ||x|| < 1\}$ in Euclidean 3-space. Using analogy with the planar unit disk

the hyperbolic density on \mathbb{B}^3 is defined by

$$\lambda(x) = \frac{2}{1 - ||x||^2}.$$

The hyperbolic length of a smooth curve $\gamma : [a, b] \rightarrow \mathbb{B}^3$ is then

$$L(\gamma) = \int_a^b \lambda(\gamma(t)) \|\gamma'(t)\| dt = \int_a^b \frac{2\|\gamma'(t)\|}{1 - \|\gamma(t)\|^2} dt.$$

The hyperbolic metric λ on \mathbb{B}^3 is defined by $\lambda(x_0, x_1) = \inf\{L(\gamma)\}$, where the infimum is taken over all smooth curves γ in \mathbb{B}^3 from x_0 to x_1 .

A curve that attains this infimum is a hyperbolic geodesic from x_0 to x_1 . The arguments used for the hyperbolic metric on the unit disc show that:

Proposition 3.2 (Hyperbolic metric on \mathbb{B}^3). The hyperbolic metric is a metric on the unit ball \mathbb{B}^3 .

Moreover, the hyperbolic geodesic from the origin 0 to any point $x \in \mathbb{B}^3$ is a radial path with hyperbolic length $\log \frac{1 + \|x\|}{1 - \|x\|}$.

Hyperbolic distance between arbitrary point $x, y \in \mathbb{B}^3$ is

$$\lambda(x, y) = \log |y, x, \xi, \eta| = \log \frac{1 + \|T_y(x)\|}{1 - \|T_y(x)\|},$$

where ξ and η are ends of geodesics through x and y , and T_y is defined by (3.4).

We can use the arguments above for any ball in \mathbb{R}_{∞}^3 and obtain a hyperbolic metric on the ball for which the orientation preserving isometries are the Möbius transformations. The most important example is when the ball is the upper half-space: $\mathbb{R}_{+}^3 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 > 0\}$. The boundary of this is the extended complex plane $\mathbb{C}_{\infty} = \mathbb{R}_{\infty}^2$. We can show that any Möbius transformation acting on this boundary extends to an orientation preserving isometry of the upper half-space for the hyperbolic metric with density: $\lambda(x) = \frac{1}{x_3}$. We can also deduce the results for the upper half-space directly from those for the ball \mathbb{B}^3 for inversion in the sphere.

We have seen how to put a hyperbolic metric on the unit ball \mathbb{B}^3 in \mathbb{R}^3 or the upper half-space \mathbb{R}_{+}^3 . We will denote both of these by \mathbb{H}^3 and call them hyperbolic 3-space. The orientation preserving isometries for hyperbolic 3-space have been identified with the group of Möbius transformations acting on the boundary $\partial\mathbb{H}^3$.

3.3. Klein model

We can show that (A) the group $M(\mathbb{H}^n)$ is isomorphic with the group $M(\mathbb{B}^n)$, using a Möbius transformation of \mathbb{H}^n onto \mathbb{B}^n . We choose so that $0, e_n, \infty$ correspond by $y = \sigma x$ to $-e_n, 0, e_n$, where e_n is the last coordinate vector. The restriction of σ on \mathbb{R}^{n-1} is the usual stereographic projection.

The correspondence is given by

$$\begin{aligned} y &= \sigma x = (e_n + 2(x - e_n)^*)^* \\ x &= \sigma^{-1}(y) = e_n + 2(y^* - e_n)^*. \end{aligned}$$

When $x_n = 0$, one verifies that $|y|^2 = 1$, $y^* = y$ and (X) reduces to

$$y_i = \frac{2x_i}{1 + |x|^2}, \quad y_n = \frac{|x|^2 - 1}{1 + |x|^2}, \quad (3.5)$$

and for $|y| = 1$, we find

$$x_i = \frac{y_i}{1 - y_n}, \quad x_n = 0. \quad (3.6)$$

The stereographic projection (3.5) maps ball $B^{n-1} = \{(x_1, x_2, \dots, x_{n-1}, 0) : x_1^2 + x_2^2 + \dots + x_{n-1}^2 < 1\}$ on the lower hemi-sphere S^{n-1} . The composition of the stereographic projection with the mapping $(y_1, y_2, \dots, y_{n-1}, y_n) \mapsto (y_1, y_2, \dots, y_{n-1})$ is

$$y = \frac{2x}{1 + |x|^2}, \quad x \in \mathbb{R}^{n-1}, \quad |x| < 1, \quad (3.7)$$

and it maps \mathbb{B}^{n-1} onto itself. The inverse mapping is L is given by

$$x = Ly = \frac{y}{1 - y_n}, \quad y \in \mathbb{R}^{n-1}, |y| < 1, \quad (3.8)$$

where $-y_n = (1 - |y|^2)^{1/2}$. Note that here $y = (y_1, y_2, \dots, y_{n-1})$.

The equation of an orthogonal circle is of the form $|x - a|^2 = |a|^2 - 1$, $|a| > 1$ or $|x|^2 + 1 = 2xa$ and (3.7) is equivalent to $ay = 1$, the equation of a straight line. This can be used to construct the Klein model of hyperbolic space. In this model the non-euclidean lines are the line segments in B^{n-1} .

Proof of (3.1) (Klein-Finsler norm) (see also 9.2).

Set $\omega_{hyp} = 2(1 - |X|^2)^{-1}|dX|$, $\omega_{hyp}^2 = 4(1 - |X|^2)^{-2}|dX|^2$ and $\omega_{kle} = L^*\omega_{hyp}$. Note that

$$(1) \quad 1 - y_n = \frac{1}{1 + |X|^2},$$

$$(2) \quad 1 - |X|^2 = \frac{2y_n}{1 - y_n},$$

$$(3) \quad y_ndy_n = -\omega, \text{ where } \omega = y_1dy_1 + y_2dy_2 + \dots + y_{n-1}dy_{n-1} = y \cdot dy.$$

Hence

$$dX = \frac{dY}{1 - y_n} + Y \frac{dy_n}{(1 - y_n)^2} = \frac{dY}{1 - y_n} + Y \frac{\omega}{y_n(1 - y_n)^2}$$

and

$(1 - |X|^2)^{-1}|dX| = \frac{1}{2y_n}|\omega_1|$, where $\omega_1 = dY - Y \frac{\omega}{y_n(1 - y_n)}$. Set $Pv = P_Yv = (Y, v)Y/|Y|$ and $Qv = v - Pv$. For $v \in T_Y\mathbb{R}^{n-1}$, we find

$$\omega_1(v) = Pv + Qv - \frac{|Y|^2}{y_n(1 - y_n)}Pv = -\frac{1}{y_n}Pv + Qv,$$

and therefore

$$\omega_{kle}^2(v) = \frac{|v|^2}{y_n^2} + \frac{|(v, Y)|^2}{y_n^4} = \frac{|v|^2}{1 - |Y|^2} + \frac{|(v, Y)|^2}{(1 - |Y|^2)^2}. \quad (3.9)$$

3.4. Conformal minimal immersion

Let $x = (x_1, \dots, x_n)$ be the coordinates on \mathbb{R}^n . The Poincaré metric on the unit $\mathbb{B} \subset \mathbb{R}^n$ is given by

$$ds_{\mathbb{B}}^2 = \frac{4|dx|^2}{(1 - |x|^2)^2}.$$

It is conformally equivalent to the Euclidean metric. The 2-dimensional case $n = 2$ is the standard Poincaré metric on the unit disk $\mathbb{D} \subset \mathbb{R}^2 \cong \mathbb{C}$. Let $S \subset \mathbb{B}$ be a minimal surface (with respect to the Euclidean metric). Let $h = ds_{hyp|S}^2$ be the metric on S obtained by restricting the Poincaré metric from $ds_{\mathbb{B}}^2$ on S (inherited from form $ds_{\mathbb{B}}^2$). Since ds_P^2 is conformally equivalent to the Euclidean metric, it introduces the same conformal structure on S as the Euclidean metric.

Problem 1. Does the Gaussian curvature of (S, h) satisfy $K_h \leq -1$?

If S is euclidean disk then the Gaussian curvature of (S, h) equals -1 .

Under "hyperbolic" we mean the Poincaré metric, then the answer is no, the curvature of minimal submanifolds need not decrease; we get this information via Forstneric [56].

Proposition 3.3. Let $f: \mathbb{D} \rightarrow \mathbb{B} \subset \mathbb{R}^n$ be a conformal immersion, $S = f(\mathbb{D})$ and $K_h \leq -1$. Then $f^*ds_P \leq ds_P$. That is, the pullback of the Poincaré metric on \mathbb{B} to the disk \mathbb{D} is bounded above by the Poincaré metric on \mathbb{D} . By integration we get

$$\operatorname{dist}_{\mathbb{B}}(f(z), f(w)) \leq \operatorname{dist}_{\mathbb{D}}(z, w), \quad z, w \in \mathbb{D}.$$

Proof. The conformal surface (S, h) has a unique Riemann surface structure. In any isothermal local complex coordinate z on S we have $h = \lambda(z)|dz|^2$, so h is a Kähler metric. Furthermore, a conformal parametrization

$f: \mathbb{D} \rightarrow S$ is a holomorphic map (up to a correct choice of orientation). If $K_h \leq -1$ were, we could apply the Ahlfors lemma which tells us that

$$f^*(ds_{\mathbb{B}}^2) = f^*(h) \leq ds_{\mathbb{D}}^2.$$

□

Via Forstneric [56], we get the following information.

It is supposedly classical and can be found in the literature that the restriction of the Beltrami-Klein metric on the ball of \mathbb{R}^n to any minimal surface (minimal with respect to the flat metric) has curvature ≤ -1 . This is what we need. Unfortunately, the B-K metric is not conformally equivalent to the Euclidean one. Hence, a conformal minimal disk is not isothermal with respect to the B-K metric, and the pull-back is not a hermitian metric on the disk. Probably it is not even quasiconformal.

There is a related results related to the estimate of the Gaussian curvature of analytic disks and more generally for complex submanifolds of Hermitian manifolds. For the following result, see [30]:

Theorem 3. *If M' is a complex submanifold of a Hermitian manifold M , then the holomorphic bisectional (sectional) curvature of M' does not exceed that of M .*

It is also interesting fact that The Bergman metric and the Beltrami-Klein metric are tightly related.

The Bergman metric is on the unit ball in \mathbb{C}^n is given by

$$ds^2 = (n+1) \left(\frac{|d\mathbf{z}|^2}{1 - |\mathbf{z}|^2} + \sum_{\mu, \nu=1} \frac{\bar{z}_\mu z_\nu dz_\mu d\bar{z}_\nu}{(1 - |z|^2)^2} \right).$$

More precisely, if $\mathbf{v} \in T_{\mathbf{z}}\mathbb{C}^n$, then

$$ds^2(\mathbf{v}) = \text{Ber}(\mathbf{z}, \mathbf{v}) = (n+1) \left(\frac{\|d\mathbf{v}\|^2}{1 - \|\mathbf{z}\|^2} + \frac{(\sum_{k=1}^n z_k \bar{v}_k)^2}{(1 - \|\mathbf{z}\|^2)^2} \right). \quad (3.10)$$

The restriction of this metric on the unit ball in \mathbb{R}^n is up to the constant the Klein metric. More precisely, if $\mathbf{v} \in T_{\mathbf{x}}\mathbb{R}^n$, then $\text{Ber}(\mathbf{z}, \mathbf{v}) = (n+1)\text{Kle}(\mathbf{x}, \mathbf{v})$.

4. Schwarz lemma in the unit ball

In this section we follow [42]. For further result see [44]. If f is a function on a set X and $x \in X$ sometimes we write fx instead of $f(x)$. We write $z = (z_1, z_2, \dots, z_n) \in \mathbb{C}^n$.

On \mathbb{C}^n we define the standard Hermitian inner product by $\langle z, w \rangle = \sum_{k=1}^n z_k \bar{w}_k$, for $z, w \in \mathbb{C}^n$, and by $|z| = \sqrt{\langle z, z \rangle}$ we denote the norm of vector z . We also use notation (z, w) instead of $\langle z, w \rangle$ on some places. By \mathbb{B}_n we denote the unit ball in \mathbb{C}^n . In particular we use also notation \mathbb{D} for the unit disk in complex plane.

For planar domains G and D we denote by $\text{Hol}(G, D)$ the class of all holomorphic mapping from G into D . For complex Banach manifold X and Y we denote by $\mathcal{O}(X, Y)$ the class of all holomorphic mapping from X into Y .

We need some properties of bi-holomorphic automorphisms of unit ball (see [59] for more details). For a fixed z , $B_z = \{w : (w - z, z) = 0, |w|^2 < 1\}$ and denote by $R(z)$ radius of ball B_z . Denote by $P_a(z)$ the orthogonal projection onto the subspace $[a]$ generated by a and let $Q_a = I - P_a$ be the projection on the orthogonal complement. For $z, a \in \mathbb{B}^n$ we define

$$\tilde{z} = \varphi_a(z) = \frac{a - Pz - s_a Qz}{1 - (z, a)}, \quad (4.1)$$

where $P_a(z) = \frac{\langle z, a \rangle}{\langle a, a \rangle} a$ and $s_a = (1 - |a|^2)^{1/2}$. Set $U^a = [a] \cap \mathbb{B}$, $Q^b = b + [a]^\perp \cap \mathbb{B}_n$,

$$\varphi_a^1(z) = \frac{a - Pz}{1 - (z, a)}, \quad \varphi_a^2(z) = \frac{-s_a Qz}{1 - (z, a)}$$

and $\delta(a, z) = |\varphi_a(z)|$. Then one can check that

(I1) The restriction of φ_a onto U^a is automorphism of U^a and the restriction onto B_z maps it bi-holomorphically mapping onto $B_{\bar{z}}$.

A domain U is called complete circular if whenever $z \in U$ and $|\lambda| \leq 1$ then $\lambda z \in U$. Note in passing that a complete circular domain automatically contains 0.

We need a few results from Rudin [59]. For a we define $s = s_a = \sqrt{1 - |a|^2}$.

Theorem 4 (2.2.2 [59]). *For every $a \in \mathbb{B}$, φ_a has the following properties:*

- (i) $\varphi_a(0) = a$ and $\varphi_a(a) = 0$.
- (ii) $\varphi'_a(0) = -s^2 P - sQ$, $\varphi'_a(a) = -P/s^2 - Q/s$.
- (iii) *The identity*

$$1 - |\varphi_a(z)|^2 = \frac{(1 - |a|^2)(1 - |z|^2)}{|1 - (z, a)|^2}.$$

- (iv) φ_a is an involution: $\varphi_a(\varphi_a(z)) = z$.
- (v) φ_a is a homeomorphism of $\overline{\mathbb{B}}$ onto $\overline{\mathbb{B}}$, and $\varphi_a \in \text{Aut}(\mathbb{B})$.
- (vi) $\text{Aut}(\mathbb{B})$ acts transitively on \mathbb{B} .

We only outline a proof. Since $(1 - (z, a))^{-1} = 1 + \langle z, a \rangle + O(|z|^2)$ and $|a|^2 P z = a \langle z, a \rangle$, $\varphi_a(z) = a - (P + sQ)z + a \langle z, a \rangle + O(|z|^2)$. Hence

$$\varphi_a(z) - \varphi_a(0) = -s^2 P z - sQz + O(|z|^2)$$

and therefore the first formula in (ii) follows; the second one follows from

$$\varphi_a(a + h) = \frac{-Ph - sQh}{s^2 - \langle h, a \rangle}.$$

From (iv), it follows that φ_a is one-to-one of $\overline{\mathbb{B}}$ onto $\overline{\mathbb{B}}$, and that $\varphi_a^{-1} = \varphi_a$. If $a, b \in \mathbb{B}$, $\varphi_b \circ \varphi_a$ is an automorphism of B that takes a to b .

If $f \in \text{Aut}(\mathbb{B})$, $a = f^{-1}(0)$, $J_{\mathbb{R}}f$ denotes real Jacobian, then

$$J_{\mathbb{R}}f(z) = \left(\frac{(1 - |a|^2)}{|1 - (z, a)|^2} \right)^{n+1}. \quad (4.2)$$

Proposition 4.1 (Theorem 8.1.2 [59]). Suppose that (i) G and G' are complete circular domains in \mathbb{C}^n and \mathbb{C}^m , respectively, and

- (ii) G' convex and bounded,
- (iii) $F : G \rightarrow G'$ holomorphic.

Then

- (a) $F'(0)$ maps G into G' and
- (b) $F(rG) \subset rG'$ ($0 < r \leq 1$), if $F(0) = 0$.

The following is an immediate corollary of Proposition 4.1:

Corollary 1. *Suppose that $f \in \mathcal{O}(\mathbb{B}_n, \mathbb{B}_m)$. If $f(0) = 0$, then (A1) $|f'(0)| \leq 1$.*

We give another proof which is more in spirit of this paper.

Proof. For $z^* = z/|z|$ define $D_z = \{\zeta z^* : \zeta \in \mathbb{D}\}$ and $F(\zeta) = f(\zeta z^*)$, $\zeta \in \mathbb{D}$. Let p be projection of \mathbb{B}_m on the slice $D_{f(z)}$. By one dimensional version of Schwarz lemma $|F(\zeta)| \leq |\zeta|$ and in particular for $\zeta = |z|$, $|f(z)| \leq |z|$. Hence (A1) $|f'(0)| \leq 1$. \square

Proposition 4.2 (Theorem 8.1.4 [59]). Suppose that $f : \mathbb{B}_n \rightarrow \mathbb{B}_m$ holomorphic, $a \in \mathbb{B}_n$ and $b = f(a)$. Then $|\varphi_b(f(z))| \leq |\varphi_a(z)|$, $z \in \mathbb{B}_n$, or equivalently,

$$\frac{|1 - (fz, fa)|^2}{(1 - |fa|^2)(1 - |fz|^2)} \leq \frac{|1 - (z, a)|^2}{(1 - |a|^2)(1 - |z|^2)}. \quad (4.3)$$

Set

$$\sigma_n(z, a) := \frac{|1 - (z, a)|^2}{(1 - |a|^2)(1 - |z|^2)}.$$

For $z, w \in \mathbb{C}^n$, $|1 - \langle z, w \rangle|^2 = 1 + |\langle z, w \rangle|^2 - (|z|^2 + |w|^2) + |z - w|^2$ and therefore
(A1) $|1 - \langle z, w \rangle|^2 \leq (s_z s_w)^2 + |z - w|^2$ and $|1 - \langle z, w \rangle|^2 = (s_z s_w)^2 + |z - w|^2$, $z, w \in \mathbb{C}$. That is

$$(B1) \quad \sigma_n(z, w) \leq 1 + \frac{|z - w|^2}{(s_z s_w)^2}, \quad \sigma_1(z, w) = 1 + \frac{|z - w|^2}{(s_z s_w)^2}, \quad z, w \in \mathbb{C}.$$

Theorem 5. Suppose that $f \in \mathcal{O}(\mathbb{B}_n, \mathbb{B}_m)$, $a \in \mathbb{B}_n$ and $b = f(a)$.

(i) Then $s_a^2 |f'(a)| \leq s_b$, i.e. $(1 - |a|^2) |f'(a)| \leq \sqrt{1 - |f(a)|^2}$.

(ii) If $m = 1$, then $s_a^2 |f'(a)| \leq s_b^2$, and

(iii) If $m > 1$, the inequality (a) $\sigma_m(fz, fw) \leq \sigma_n(z, w)$, $z, w \in \mathbb{B}_n$, does not hold in general, but if $f \in \mathcal{O}(\mathbb{B}_n, \mathbb{B}_1)$, then $\sigma_1(fz, fw) \leq \sigma_n(z, w)$, that is the following inequality holds:

$$\sigma_1(fz, fa) = \frac{|fz - fa|^2}{(1 - |fa|^2)(1 - |fz|^2)} \leq \frac{|z - a|^2}{(1 - |a|^2)(1 - |z|^2)}, \quad z \in \mathbb{B}_n. \quad (4.4)$$

Proof. (i) Suppose first that $f(0) = 0$ and take $z \in \mathbb{B}_n$. Hence (A1) $|f'(0)| \leq 1$.

For $u \in T_a C^m$, by Theorem 4(ii), $v = \varphi'_a(a)u = -Pu/s^2 - Qu/s$ and, by Pitagora's theorem,

$$|u| = \sqrt{|Pu|^2 + |Qu|^2}, |v|^2 = |Pu|^2/s^4 + |Qu|^2/s^2$$

and therefore we find

$$(B1) \quad \frac{|u|}{s} \leq |\varphi'_a(a)u| \leq \frac{|u|}{s^2}.$$

If $f(a) = b$, set $h = \varphi_b \circ f \circ \varphi_a$. By the chain rule $h'(0) = \varphi'_b(b) \circ f'(a) \circ \varphi'_a(0)$.

Set $u \in T_a C^n$, $v = f'(a)u \in T_a C^m$, $u' = \varphi'_a(a)u$ and $v' = \varphi'_b(b)v$. By (A1), $|v'| \leq |u'|$. Since, by (B1),

$$\frac{|v|}{s_b} \leq |v'| \quad \text{and} \quad |u'| \leq \frac{|u|}{s_a^2},$$

hence $s_a^2 |f'(a)| \leq s_b$, i.e. $(1 - |a|^2) |f'(a)| \leq \sqrt{1 - |f(a)|^2}$ and therefore (i) is proved.

(ii) If $m = 1$, then $s_b^2 |v'| = |v|$ and (ii) follows.

(iii) By (B1) and (4.3),

$$\sigma_1(fz, fa) = 1 + \frac{|fz - fa|^2}{(s_{fz} s_{fa})^2} \leq \sigma_n(z, a) = \frac{|1 - (z, a)|^2}{(s_z s_a)^2} \leq 1 + \frac{|z - a|^2}{(s_z s_a)^2}$$

and therefore we get (4.4). If z tends a , (ii) also follows from (4.4). If (a1) holds, then (b1) $s_a^2 |f'(a)| \leq s_b^2$. For function $f_0 = \varphi_b \circ \varphi_a$ we have $(1 - |a|^2) |f'_0(a)| = (1 - |b|^2)$, which yields a contradiction with (b1). \square

5. Contraction properties of holomorphic functions with respect to Kobayashi distances

The author also published a paper [36] about holomorphic fixed point theorem on Riemann surfaces.

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

We define the distance function on G by integrating the pseudometric k_G : for $z, z_1 \in G$

$$\text{Kob}_G(z, z_1) = \inf_{\gamma} \int_0^1 k_G(\gamma(t), \dot{\gamma}(t)) dt, \quad (5.1)$$

where the infimum is over all piecewise paths $\gamma: [0, 1] \rightarrow G$ with $\gamma(0) = z$ and $\gamma(1) = z_1$.

One can prove

Theorem 6. Suppose that G and G_1 are bounded connected open subset of complex Banach space and $f: G \rightarrow G_1$ is holomorphic. Then

$$\text{Kob}_{G_1}(fz, fz_1) \leq \text{Kob}_G(z, z_1), \quad (5.2)$$

for all $z, z_1 \in G$.

Let $A = \{1 < |x| < 4\}$, $A^* = \{2 < |x| < 3\}$, $l(t) = 2 + \frac{1}{3}(t - 1)$ and $f(x) = -l(|x|)x$. l maps the interval $(1, 4)$ onto the interval $(2, 3)$ and therefore f maps A onto $A^* \subset A$, but f has no fixed point (there is no point $x \in A$ such that $f(x) = x$). Hence this example shows that there is no metric d on G such that f is a contraction wrt d . The situation is completely different for analytic functions.

Theorem 7. Suppose that G is bounded connected open subset of complex Banach space and $G_* \subset G$, $s_0 = \text{dist}(G_*, G^c)$, $d_0 = \text{diam}(G)$ and $q_0 = \frac{d_0}{d_0 + s_0}$. Then

- (i) $\text{Kob}_G \leq q_0 \text{Kob}_{G_*}$ on G_* .
- (ii) In addition if $f: G \rightarrow G_*$ is holomorphic, then

$$\text{Kob}_{G_*}(fz, fz_1) \leq q_0 \text{Kob}_{G_*}(z, z_1), \quad (5.3)$$

for $z, z_1 \in G_*$.

$$\text{Kob}_G(fz, fz_1) \leq q_0 \text{Kob}_G(z, z_1), \quad (5.4)$$

for $z, z_1 \in G$.

Proof. Suppose that $p \in G_*$, $v \in T_p G_*$ and $\phi: \mathbb{D} \rightarrow G$ is a holomorphic function such that $\phi(0) = p$ and $d\phi(h) = v$. Set $R_s = \frac{d_0+s}{d_0}$ and $q_s = \frac{d_0}{d_0+s}$. For $h \in \mathbb{D}$ define $\phi_s(h) = p + R_s(\phi(h) - p)$. Then $\phi_s(h) - \phi(h) = (R_s - 1)(\phi(h) - p)$ and therefore $|\phi_s(h) - \phi(h)| \leq s$. For $s < s_0$, ϕ_s maps \mathbb{D} into G and $d\phi_s(h) = R_s v$. Hence $k_G(p, v) \leq q_s k_{G_*}(p, v)$ and if s approaches s_0 we first get (i) $k_G(p, v) \leq q_0 k_{G_*}(p, v)$ and by a standard procedure $\text{Kob}_G \leq q_0 \text{Kob}_{G_*}$. Now, by (5.2), we have (ii) $\text{Kob}_{G_*}(fz, fz_1) \leq \text{Kob}_G(z, z_1)$. Combining (i) and (ii) we get (5.3) and (5.4). \square

If $d_0 = \text{diam}(G)$ is not finite, elementary example $H_a = \{z : \text{Im } z > a\}$, with $f(z) = z + ia$, which maps H onto H_a , shows that the theorem does not hold.

Theorem 8. Let $D \subset C^n$ domain for which Kobayashi (Carthédory) pseudo-distance is distance and $f: D \rightarrow D$ holomorphic mapping such that $f(D)$ is a compact subset of D . Then f is contraction with respect to Kobayashi (Carthédory) metric on D . In particular f has fixed points in D .

It is a corollary of Theorem 7. A version of Theorems 6-7 was proved in 1968 by Clifford Earle and Richard Hamilton [20] (see subsections 5.2 for further comments).

5.1. Addition to the proof of Theorem 5(iii) and Theorems 6-7

The Schwarz-Pick lemma states that every holomorphic function from the unit disk \mathbb{D} to itself, or from the upper half-plane \mathbb{H} to itself, will not increase the Poincaré distance between points.

It is convenient to introduce a pseudo-distance

$$\delta(z, \omega) = |\varphi_z(\omega)| = \left| \frac{z - \omega}{1 - \bar{\omega}z} \right|, \quad z, \omega \in \mathbb{D}, \quad (5.5)$$

which is a *conformal invariant*.

Shwarz-Pick lemma: If f holomorphic function from the unit disk \mathbb{D} to itself, then

$$\delta(f(z), f(\omega)) \leq \delta(z, \omega), \quad z, \omega \in \mathbb{D}, \quad (5.6)$$

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

For $z, w \in \mathbb{C}$, set $a = (1 - |z|^2)(1 - |w|^2)$, $b = |z - w|^2$, $A = (1 - |fz|^2)(1 - |fw|^2)$, and $B = |fz - fw|^2$. By this notation,

- (A2) $|1 - \langle z, w \rangle|^2 = 1 + |\langle z, w \rangle|^2 - (|z|^2 + |w|^2) + |z - w|^2 = a + b$,
(B2) $|1 - \langle fz, fw \rangle|^2 = A + B$.

If $f \in \mathcal{O}(\mathbb{B}_1, \mathbb{B}_1)$, using (A2) and (B2) Shwarz-Pick lemma can be rewritten in the form $\frac{B}{b} \leq \frac{A+B}{a+b}$ and therefore $Ba \leq Ab$. That is

$$(I) \quad |fz - fw| \sqrt{(1 - |z|^2)} \sqrt{(1 - |w|^2)} \leq \sqrt{(1 - |fz|^2)} \sqrt{(1 - |fw|^2)} |z - w|.^3$$

Question 1 (D. Jocić). If $f \in \mathcal{O}(\mathbb{B}_n, \mathbb{B}_m)$ whether (I) holds?

For $z, w \in \mathbb{C}^n$ we have

$$|z - w|^2 = |z|^2 + |w|^2 - 2\operatorname{Re} \langle z, w \rangle$$

and

$$|1 - \langle z, w \rangle|^2 = 1 - 2\operatorname{Re} \langle z, w \rangle + |\langle z, w \rangle|^2.$$

Hence

- $|1 - \langle z, w \rangle|^2 = 1 + |\langle z, w \rangle|^2 - (|z|^2 + |w|^2) + |z - w|^2$ and
 $|1 - \langle fz, fw \rangle|^2 = 1 + |\langle fz, fw \rangle|^2 - (|fz|^2 + |fw|^2) + |fz - fw|^2$, thus by Cauchy-Shwarz inequality
 $|\langle z, w \rangle|^2 \leq |z||w|$ and therefore
(C2) $|1 - \langle z, w \rangle|^2 \leq a_n + b_n$, where $a_n = (1 - |z|^2)(1 - |w|^2)$, and $b_n = |z - w|^2$.
Set $A_m = (1 - |fz|^2)(1 - |fw|^2)$ and $B_m = |fz - fw|^2$. By (C2) and (4.3),

$$\sigma_1(fz, fw) = \frac{A_1 + B_1}{A_1} \leq \sigma_n(z, w) = \frac{|1 - \langle z, w \rangle|^2}{a_n} \leq \frac{a_n + b_n}{a_n},$$

so we get (4.4).

We show that (I) does not hold in general. Contrary suppose that (I) holds and that $f \in \mathcal{O}(\mathbb{B}_n, \mathbb{B}_m)$, $a \in \mathbb{B}_n$ and $b = f(a)$.

Recall if $m > 1$ we proved,

$$(II) \quad (1 - |a|^2)|f'(a)| \leq \sqrt{1 - |f(a)|^2}.$$

Note that for function $f_0 = \varphi_b \circ \varphi_a$ we have equality in (II).

If (I) holds and z tends a then we have,

$$(III) \quad (1 - |a|^2)|f'(a)| \leq (1 - |b|^2).$$

An application of (II) and (III) to f_0 shows that $s_b \leq s_b^2$ and consequently $s_b \geq 1$. Since $s_b < 1$ for $b \neq 0$, we have a contradiction.

5.2. Further comments related to Theorems 6-7

We have worked on the subject from time to time between 1980 -1990 and in that time we proved Theorems 6-7⁽⁴⁾. 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 [20] 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.

6. Curvature of Kobayashi and Carathéodory metric

6.1. On the Hessian of the Carathéodory metric

Here we collect some materials from Burbea's paper [12].

In [12], the generalized lower Hessian of an upper semi-continuous function near a point z in \mathbb{C}^n is introduced (for $n = 1$ see Heins[29]). With this Burbea introduces a "sectional curvature" and he proves that the sectional

³D. Jocić turns my attantion on this form and after communication with him we have added the proof of (4.4))

⁴we found a my hand written manuscript 1990 and did not pay much attention to it at that time

curvature of the Carathéodory-Reiffen metric is always ≤ -4 . This generalizes a result of Saito [51] in the one dimensional case. The sectional curvatures of the ball and polydisk are always -4 . A few other properties of the Hessian of the above metric are shown.

Now we give more details.

For ζ in D we write $H_\zeta(D, \mathbb{D}) = \{f \in \text{Hol}(D, \mathbb{D}) : f(\zeta) = 0\}$. For each ζ in D , $C_D(\zeta, \cdot)$ is the function defined on the complex tangent space of D at ζ by $C_D(\zeta, \mathbf{v}) = \sup\{|<\partial f(\zeta), \mathbf{v}>| : f \in H_\zeta(D, \mathbb{D})\}$.

Exercise 5. For example, in the polydisk $C_{\mathbb{D}^n}(0, z) = \max_k \ln \frac{1+|z_k|}{1-|z_k|}$ and in the ball $C_{\mathbb{B}^n}(0, z) = \ln \frac{1+|z|}{1-|z|}$.

Let (z, w) and (z', w') be points in \mathbb{D}^2 and define $A(z, w) = (T_{z'}(z), T_{w'}(w))$. Then $A \in \text{Aut}(\mathbb{D}^2)$ and $A(z', w') = (0, 0)$ and if $|T_{z'}(z)| \geq |T_{w'}(w)|$, then

$$C_{\mathbb{D}^2}((z, w), (z', w')) = \ln \frac{1 + |T_{z'}(z)|}{1 - |T_{z'}(z)|}.$$

Since $\text{Hol}(D, \mathbb{D})$ is a normal family, the supremum in the definition of $C_D(\zeta, \mathbf{v})$ is attained by some $F \in H_\zeta(D, \mathbb{D})$. Here $F(z) = F(z; \zeta, \mathbf{v})$. By a normal family argument $C_D(\zeta, \mathbf{v})$ is continuous in (ζ, \mathbf{v}) .

The Hessian: Let f be upper semi-continuous near $z \in \mathbb{C}^n$ and let $\mathbf{u} \in \mathbb{C}^n \setminus \{0\}$. The generalized lower Hessian (or "Laplacian") of f at z along the direction \mathbf{u} is defined by

$$\Delta_{\mathbf{u}} f(z) = 4 \liminf_{r \rightarrow 0} \frac{1}{r^2} \left[\frac{1}{2\pi} \int_0^{2\pi} (f(z + re^{it}\mathbf{u}) - f(z)) dt \right].$$

Note that, if f is a C^2 function near z , then $\Delta_{\mathbf{u}} f(z)$ reduces to four times the usual Hessian of f at z along \mathbf{u} , that is

$$\Delta_{\mathbf{u}} f(z) = 4 \sum D_{z_i \bar{z}_j}^2 f(z) u_i \bar{u}_j = 4 H_z(f, \mathbf{u}).$$

If u is the restriction of f on the complex line $z = l(z^0 + \zeta \mathbf{u})$, that is $u(\zeta) = f \circ l(\zeta)$, then using the chain rule we have

$$D_{\zeta \bar{\zeta}}^2 u = H_z(f, \mathbf{u}) = \sum D_{z_i \bar{z}_j}^2 f(z) u_i \bar{u}_j.$$

Hence, since $\Delta u = 4 D_{\zeta \bar{\zeta}}^2 u$, $\Delta u = \Delta_{\mathbf{u}} f(z) = 4 H_z(f, \mathbf{u})$.

Especially, if f is a C^1 function near the point z , and $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{C}^n$, then $<\partial f, \mathbf{v}> = \sum_{j=1}^n D_j^c f v_j$. Let $v \in \mathbb{C}^n \setminus \{0\}$ and consider $F(z) = F(z, \zeta, \mathbf{v})$ as before. Define

$$\lambda(z; \mathbf{u}) = \frac{|<\partial F(z), \mathbf{u}>|}{1 - |F(z)|^2}.$$

Therefore, $\ln \lambda(z; \mathbf{u}) = \ln |<\partial F(z), \mathbf{u}>| - \ln(1 - |F(z)|^2)$. The first term on the right is pluriharmonic and hence its Hessian along any direction (independently of \mathbf{u}) is zero. Consequently, $\Delta_{\mathbf{u}} \ln \lambda(z; \mathbf{u}) = 4 \lambda(z; \mathbf{u})^2$, for each direction $\mathbf{u} \in \mathbb{C}^n$. Especially, $\Delta_{\mathbf{u}} \ln \lambda(\zeta; \mathbf{v}) = 4 \lambda(\zeta; \mathbf{v})^2$. Note that $\lambda(\zeta; \mathbf{v}) = C_D(\zeta; \mathbf{v})$.

Theorem 9. Let $\zeta \in D$ and $\mathbf{v} \in \mathbb{C}^n \setminus \{0\}$ be fixed. Then $\Delta_{\mathbf{u}} \ln C_D(\zeta; \mathbf{v}) \geq 4 \lambda(\zeta; \mathbf{u})^2$, for each direction $\mathbf{u} \in \mathbb{C}^n$, and thus again $\log C_D(\zeta; \mathbf{v})$ is plurisubharmonic.

Let $\mathbf{v} \in \mathbb{C}^n \setminus \{0\}$ and assume that the metric density $\mu(z; \mathbf{v})$ is a positive upper semi-continuous function at z . The "curvature" of $\mu(z; \mathbf{v})$ at z in the direction \mathbf{v} is given by

$$K(\mu; z, \mathbf{v}) = -\frac{1}{\mu(z; \mathbf{v})^2} \Delta_{\mathbf{v}} \ln \mu(z; \mathbf{v}).$$

If $\rho = \mu_{\mathbf{v}}$ is the restriction of μ on the complex line $z = l(z^0 + \omega \mathbf{v})$, then $\rho = \mu_{\mathbf{v}}$ is function of one complex variable ω . If we consider ρ as a metric density then $K_{\rho} = K(\mu; z, \mathbf{v})$.

The "curvature" of $\lambda(z; \mathbf{v})$ is -4 at $z = \zeta$.

Theorem 10. The curvature of $C_D(\zeta; \mathbf{v})$ is always ≤ -4 .

Proof. By Theorem 9, $\triangle_{\mathbf{v}} \ln C_D(\zeta; \mathbf{v}) \geq 4\lambda(\zeta; \mathbf{v})^2$ and, since $\lambda(\zeta; \mathbf{v}) = C_D(\zeta; \mathbf{v})$, the assertion follows. \square

We also note that the Carathéodory metric has the "distance-decreasing" property that is, if $f : D \rightarrow D^*$ is a holomorphic mapping, then $C_D(f(z); f^*(v)) \leq C_D(z; v)$. The Carathéodory metric may be defined on arbitrary complex manifolds, although it may be zero in some directions v . Clearly the Carathéodory metric for the unit disk \mathbb{D} is given by (B1) $C_{\mathbb{D}}(z; \mathbf{v}) = \frac{|\mathbf{v}|}{1-|z|^2}$.

Proposition 6.1. $C_D(z; \mathbf{v})||t; ||/d(z)$, where $d(z)$ is the distance of $z \in D$ from the boundary of D .

Proposition 6.2. Let $v \in \mathbb{C}^n \setminus \{0\}$ be fixed. Then $\log C_D(z; \mathbf{v})$ is plurisubharmonic in $z \in D$.

6.2. Kobayashi and Carathéodory metric

For complex Banach manifold X and Y we denote by $\mathcal{O}(X, Y)$ the class of all holomorphic mapping from X into Y . A complex Finsler metric F on a complex (Banach) manifold M is an upper semicontinuous function $F : T^{1,0}M \rightarrow \mathbb{R}^+$ satisfying

- (i) $F(p; \mathbf{v}) > 0$ for all $p \in M$ and $\mathbf{v} \in T_p^{1,0}M$ with $v \neq 0$;
- (ii) $F(p; \lambda \mathbf{v}) = |\lambda|F(p; \mathbf{v})$ for all $p \in M$, $\mathbf{v} \in T_p^{1,0}M$ and $\lambda \in \mathbb{C}$.

Wong proved the following interesting result, see also [12]:

Theorem 11 ([54]). (A) If G is a hyperbolic manifold in the sense of Kobayashi and the differential Kobayashi metric K_G is of class C^2 , then the holomorphic curvature of K_G is greater than or equal to -4 .

(B) If G is Carathéodory-hyperbolic and the differential Caratheodory metric C_G is of class C^2 , then the holomorphic curvature of C_G is less than or equal to -4 .

Here we shortly outline Wong approach [54].

Lemma 1. Let M, N be complex manifolds, and N complete hyperbolic in the sense of Kobayashi, suppose that we fix two points x_1 and x_2 in M and N respectively. Then $S = \{f \in \mathcal{O}(M, N), f(x_1) = x_2\}$ is compact in $\mathcal{O}(M, N)$ with respect to the compact open topology.

Definition 4. (a) Suppose that F is a C^2 hermitian Finsler metric on a complex one dimensional manifold. It is obvious that in this case F is just a C^2 hermitian metric in the usual sense of differential geometry. Then the holomorphic curvature of F is given by the following formula:

$$K(F) = -\frac{D_{z\bar{z}}^2 \ln F}{F}. \quad (6.1)$$

(b) Let G be a complex manifold as before and $M_p(\mathbf{v})$ any complex one dimensional submanifold through the point p and whose tangent space at p is spanned by $\{\mathbf{v}, J\mathbf{v}\}$. In the following, $G(\mathbf{v})_p$ is the set of all $M_p(\mathbf{v})$. The holomorphic curvature $k_F(p, \mathbf{v})$ of a C^2 hermitian Finsler metric F at $(p, \mathbf{v}) \in T(G)$ is defined to be the following number:

$$k_F(p, \mathbf{v}) = \sup_{G(\mathbf{v})_p} \{\text{the holomorphic curvature of the restriction of } F \text{ to } M_p(\mathbf{v})\}. \quad (6.2)$$

If X and Y are complex Banach manifolds by $\mathcal{O}(X, Y)$ (the notation $\text{Hol}(X, Y)$ is also used in the literature) we denote the family of holomorphic mappings of X into Y . Let G be a complex manifold and $T(G)$ the tangent bundle; we define the differential Caratheodory metric as follows: $C_G : T(G) \rightarrow \mathbb{R}^+ \cup \{0\}$,

$$C_G(z, v) = \sup\{|df_z(v)| : f \in \mathcal{O}(G, \mathbb{D}), f(z) = 0\}. \quad (6.3)$$

One can obtain a mapping f belonging to $\mathcal{O}(G, \mathbb{D})$ satisfying the following conditions:

- (1) $f(p) = 0$ and
- (2) $C_G(p, v) = |df_p(v)|$.

We observe that $df_p \neq 0$. For any $M_p(v)$ one can choose a neighborhood U of the origin 0 in \mathbb{D} such that $f : M_p(v) \rightarrow U$ is a biholomorphism (for sufficiently small choice of $M_p(v)$).

Proof. (A) Let us fix a tangent (p, \mathbf{v}) at p . It is clear from Definition 4(b) that it suffices to prove the holomorphic curvature of the restriction of C_G to any $M_p(\mathbf{v})$ of $G(\mathbf{v})_p$ is less than -4 .

With respect to the local coordinates $\{z, \bar{z}\}$ of $M_p(\mathbf{v})$, $f^*(B_{\mathbb{D}})$ (the pullback of $B_{\mathbb{D}}$ by f restricted to $M_p(\mathbf{v})$) and the restriction of C_G to $M_p(\mathbf{v})$ can be written as follows:

The restriction of $f^*(B_{\mathbb{D}})$ to $M_p(\mathbf{v})$ is $hdz d\bar{z}$; The restriction of C_G to $M_p(\mathbf{v})$ is $gdz d\bar{z}$, where h and $g = g_C$ are smooth functions on $M_p(\mathbf{v})$.

It is important to point out here that the Caratheodory metrics enjoy the distance decreasing property under holomorphic mappings. Therefore we have the following inequality: $f^*(B_{\mathbb{D}}) \leq C_G$ (i.e. $h \leq g$), where $B_{\mathbb{D}}$ is Finsler form of Poincaré metric \mathbb{D} .

We let $u = h/g$. From (2) we have $h(p) = g(p)$ (f realizes C_G at the point p). Together with the above inequality ($h \leq g$ on $M_p(\mathbf{v})$) and the definition of u , one can obtain the following two conditions of u :

- (a) $u(p) = 1$ (i.e. $\ln u(P) = 0$),
- (b) $u \leq 1$ on $M_p(\mathbf{v})$ (i.e. $\ln u \leq 0$).

This means $\ln u$ attains a maximum at p . Therefore we have $D_{z\bar{z}}^c \ln u(p) \leq 0$. From the fact that $\ln u = \ln h - \ln g$, we have the following inequality:

$$D_{z\bar{z}}^c \ln h(p) \leq D_{z\bar{z}}^c \ln g(p).$$

However, since $h(P) = g(P)$ and $1/h \geq 1/g$, we easily get

$$-\frac{1}{h} D_{z\bar{z}}^c \ln h(p) \geq -\frac{1}{g} D_{z\bar{z}}^c \ln g(p),$$

that is $-4 = K_h(p) \geq K_g(p)$.

The left-hand side is just the holomorphic curvature of the Poincaré metric on \mathbb{D} , which is equal to -4 . The right-hand side is the holomorphic curvature of the restriction of C_G to $M_p(\mathbf{v})$. This completes the proof.

(B) From the definition of the differential Kobayashi metric there exists a sequence of holomorphic functions $\{f_i\}$ in $\mathcal{O}(\mathbb{D}, G)$, such that $f_i(0) = p$, $K(P, v) = \lim |df_i(w_i)|$, where $df_i(w_i) = v$, w_i is a tangent at the origin of \mathbb{D} , and $|w_i|$ is taken with respect to the Poincaré metric in the unit disc \mathbb{D} .

G is assumed to be hyperbolic in the sense of Kobayashi, so that it is a tight manifold in the sense of Wu (see [55]). Therefore there exist neighborhoods U_1, U_2 of 0 and p , respectively, such that $f_i(U_1) \subset U_2$ for all i . Furthermore, U_2 can be chosen to be complete hyperbolic (for example, biholomorphic to the unit ball).

Applying the lemma in part (A) again, we obtain a holomorphic mapping $f : U_1 \rightarrow U_2$ satisfying the following conditions: $f : U_1 \rightarrow U_2$ satisfying the following conditions:

- (i) $f(0) = p$, $(df)_0(w_0) = v$ and $|w_0| = K_G(p, v)$.

Since f does not increase the corresponding distances we have

- (ii) $\text{Hyp}_{\mathbb{D}}(z, w) \geq K_G(f(z), (df)_0(w))$, for all $(z, w) \in T(D)$, $z \in U_1$.

Let $gdz d\bar{z}$ ($g = g_K$) be the pullback of the restriction of K_G to $M_p(\mathbf{v})$ by f and let $hdz d\bar{z}$ be the Poincaré metric of the unit disc \mathbb{D} . We let $u = h/g$. Clearly, by (i), $\log u$ attains a minimum at 0 in \mathbb{D} . Hence we have the following inequalities: $D_{z\bar{z}}^2 \ln u(0) \geq 0$ and $K_h(0) \leq K_g(0)$. One observes that the left-hand side of the above second inequality is the holomorphic curvature of the Poincaré metric, which is identically equal to -4 ($K_h(0) = -4$). The right-hand side is equal to the holomorphic curvature of the restriction of K_G to $M_p(\mathbf{v})$ at the point p . Our proof is therefore completed. \square

Exercise 6. (a) Check that in unit ball $K_{\mathbb{B}}(0, z) = C_{\mathbb{B}}(0, z) = \ln \frac{1+|z|}{1-|z|}$.

(b) Fill the details for the proofs: $D_{z\bar{z}}^c \ln h(p) \leq D_{z\bar{z}}^c \ln g_C(p)$ in the case (A) and $-4 = K_h(0) \leq K_g(0)$, where $g = g_K$.

The following question is fundamental in hyperbolic complex analysis. If G is complete hyperbolic, does K_G satisfy the maximum modulus principle in $T(G)$.

A Schwarz-Pick system is a functor, denoted by $X \mapsto d_X$, that assigns to each complex Banach manifold X a pseudometric d_X so that the following conditions hold: (a) The pseudometric assigned to \mathbb{D} is the Poincaré metric (b) If X and Y are complex Banach manifolds then (2.2) $d_Y(f(x_1), f(x_2)) \leq d_X(x_1, x_2)$ if $x_1 \in X$, $x_2 \in X$ and $\forall f \in \mathcal{O}(X, Y)$.

Because of conditions (a) and (b) the sets $\mathcal{O}(\mathbb{D}, X)$ and $\mathcal{O}(X, \mathbb{D})$ provide upper and lower bounds for d_X . These upper and lower bounds lead to the definitions of the Kobayashi and Caratheodory pseudometrics, which we shall study in the remainder of this paper.

In this paper $d_{\mathbb{D}}$ will always be the Poincaré metric (2.1) on the unit disk \mathbb{D} .

Definition 5. A Schwarz-Pick pseudometric on the complex Banach manifold X is a pseudometric d such that (3.1) $d(f(z), f(w)) \leq d_{\mathbb{D}}(z, w)$ for all z and w in \mathbb{D} and f in $\mathcal{O}(\mathbb{D}, X)$. If $X \mapsto d_X$ is a Schwarz-Pick system, then d_X is obviously a Schwarz-Pick pseudometric on X for every complex Banach manifold X .

The Carathéodory length of a piecewise C^1 curve $\gamma : [a, b] \rightarrow X$ in X is $\tilde{L}_X(\gamma) = \int_a^b c_X(\gamma(t), \gamma'(t))dt$ and the distance $\tilde{C}_X(x, y)$ is the infimum of the lengths of all piecewise C^1 curves joining x to y . Observe that the integrand in (4.6) is piecewise continuous. The functor assigning \tilde{C}_X to each complex Banach manifold X is a Schwarz-Pick system. In particular, if x and y are points in X , then $d(fx, fy) \leq \tilde{C}_X(x, y)$ for all f in $\mathcal{O}(X, \mathbb{D})$. Definition (4.3) therefore implies that $C_X(x, y) \leq \tilde{C}_X(x, y)$ for all x and y in X .

Complex geodesics. Since $X \mapsto \tilde{C}_X$ is a Schwarz-Pick system, \tilde{C}_X is a Schwarz-Pick pseudometric on X for every complex Banach manifold X . Therefore $\tilde{C}_X \leq K_X$ for every X . Combining that inequality with (4.7) we obtain $C_X(f(z), f(w)) \leq \tilde{C}_X(f(z), f(w)) \leq K_X(f(z), f(w)) \leq d_{\mathbb{D}}(z, w)$, whenever X is a complex Banach manifold, $f \in \mathcal{O}(\mathbb{D}, X)$, and z and w are points of \mathbb{D} .

Following Vesentini [Ves81], we call f in $\mathcal{O}(\mathbb{D}, X)$ a complex geodesic (more precisely a complex C_X -geodesic) if there is a pair of distinct points z and w in \mathbb{D} , with

$$C_X(f(z), f(w)) = d_{\mathbb{D}}(z, w), \quad (6.4)$$

so that none of the inequalities in (5.1) is strict.

Definition 6. A holomorphic map $\varphi : \mathbb{D} \rightarrow X$ in a complex manifold X is a complex geodesic if it is an isometry between the Poincaré distance d_{hyp} and the Kobayashi distance k_X .

It is a well-known result of Lempert [34] that on convex domains the Kobayashi and Carathéodory distances (resp metrics) coincide.

In his famous 1981 paper [34], Lempert proved that given a point in a strongly convex domain the complex geodesics (i.e., the extremal disks) for the Kobayashi metric passing through that point provide a very useful fibration of the domain.

In communication with Forstneric and the author the following question has been mentioned:

Question. Whether, in the ball or a bounded convex domains of R^n , there exist minimal geodesics, i.e. conformal minimal (=harmonic) disks which are extremal at every point. This holds for holomorphic disks in any bounded convex domain in C^n by a famous theorem of Lempert (1981).

6.3. Calculation of the curvature

Let G be a bounded domain and $K(z; w)$ be the Bergman Kernel on G . Write $\phi(z) = \log K(z, z)$. The Bergman metric

$$ds^2 = \sum_{\mu, \nu=1}^n D_{z_\mu \bar{z}_\nu}^2 \phi dz_\mu d\bar{z}_\nu \quad (6.5)$$

is the Kahler metric with Kahler form $i\partial\bar{\partial}\phi$. We use notation b_D or $Berg_D$ for the Bergman distance on D . Note that the distance in the Bergman metric from the origin in the unit ball $\mathbb{B} \subset \mathbb{C}^n$ is

$$b_B(0, z) = \sqrt{n+1} \ln \frac{1+|z|}{1-|z|} = \sqrt{n+1} b_B(0, z). \quad (6.6)$$

In the polydisk $b_{\mathbb{D}}(0, z) = \frac{1}{\sqrt{2}} \sqrt{\sum_{k=1}^n \ln \frac{1+|z_k|}{1-|z_k|}}$ and $C_{\mathbb{D}}(0, z) = \max_k \ln \frac{1+|z_k|}{1-|z_k|}$.

Let (z, w) and (z', w') be points in \mathbb{D}^2 and define $A(z, w) = (T_{z'}(z), T_{w'}(w))$. Then $A \in Aut(\mathbb{D}^2)$ and $A(z', w') = (0, 0)$ and if $|T_{z'}(z)| \geq |T_{w'}(w)|$, then $C_{\mathbb{D}}^2((z, w), (z', w')) = \ln \frac{1+|T_{z'}(z)|}{1-|T_{z'}(z)|}$.

For $u > 0$

$$\Delta \ln u = \frac{u\Delta u - |\nabla u|^2}{u^2}.$$

Let $g, f \in \mathcal{O}(\mathbb{D}, \mathbb{C}^n)$ and let f map \mathbb{D} into the unit ball in \mathbb{C}^n . Thus,

$$|g|\Delta|g| = 2|g'(z)|^2 - \frac{|\langle g, g' \rangle|^2}{|g|^2},$$

$$|\nabla|g||^2 = \frac{|\langle g, g' \rangle|^2}{|g|^2} := A,$$

$$\Delta|f^2| = 4|f'|^2,$$

$$|\nabla|f|^2|^2 = 4|\langle f', f \rangle|^2.$$

Set $u = |g|$, $g = f'$, $\lambda_1 := \ln u$, $v := 1 - |f|^2$, $\lambda_2 := \ln v$, $I_1 = \Delta\lambda_1$ and $I_2 = -\Delta\lambda_2$. Then

$$I_1 = u^{-2}2(|g'|^2 - A) \quad \text{and} \quad I_2 = -\Delta\lambda_2 = 4v^{-2}(v|f'|^2 + |\langle f', f \rangle|^2).$$

Set

$$\rho = |f'|(1 - |f|^2)^{-1}.$$

Hence $I = \Delta \ln \rho = I_1 + I_2$. Since

$$I_2 = 4v^{-2}(|f'|^2 - |f|^2|f'|^2 + |\langle f', f \rangle|^2),$$

it seems that for $n = 1$, $I = \Delta \ln \rho \geq 4\rho^2$. But, for $n > 1$, we have some difficulties. Note that $I_1 \geq 0$. Perhaps, we can try to apply Schwarz lemma

$$(1 - |z|^2)^{-1}|f'(z)| \leq v = 1 - |f|^2$$

to estimate

$$R = I_1 + 4v^{-2}(-|f|^2|f'|^2 + |\langle f', f \rangle|^2).$$

Wikipedia says that the metric

$$h = \frac{4|dx|^2}{(1 - |x|^2)^2} \tag{6.7}$$

on the ball $\mathbb{B} = \{|x|^2 < 1\} \subset \mathbb{R}^n$ (for any $n \in \mathbb{N}$) is the Poincare model of the hyperbolic space; presumably it has constant (Gaussian?) curvature -1 . However, it seems that this metric might not be the most suitable for our purposes.

There is another model of a hyperbolic n -space, the so called *Beltrami-Klein model*, which is represented by the ball $\mathbb{B} \subset \mathbb{R}^n$ by the metric

$$g = 4 \frac{|dx|^2 + (x \cdot dx)^2}{(1 - |x|^2)^2}; \quad x \cdot dx = \sum_{j=1}^n x_j dx_j. \tag{6.8}$$

In the complex case, replacing \mathbb{R}^n by \mathbb{C}^n , the metric h is obviously not Kähler. The natural standard Kähler metric on the ball is the *Bergman metric*. Up to a normalizing constant, the Bergman kernel for the ball $\mathbb{B}_n \subset \mathbb{C}^n$ on the diagonal $z = w$ equals

$$K_{\mathbb{B}}(z) = \frac{1}{(1 - |z|^2)^{n+1}}. \tag{6.9}$$

The Bergman metric on $\mathbb{B} = \mathbb{B}_n$ is defined by the Kähler $(1, 1)$ -form

$$k_{\mathbb{B}} = -i(n+1)\partial\bar{\partial}\log(1 - |z|^2) = \frac{i(n+1)}{(1 - |z|^2)^2} \sum_{j,k=1}^n \bar{z}_j z_k dz_j \wedge d\bar{z}_k. \tag{6.10}$$

That is, up to a constant we have

$$ds_{\mathbb{B}}^2 = \frac{n+1}{(1 - |z|^2)^2} \sum_{j,k=1}^n \bar{z}_j z_k dz_j \otimes d\bar{z}_k. \tag{6.11}$$

This is similar to the expression for g (6.8), while h (6.7) amounts to the diagonal terms.

Now, if $F = (F_1, \dots, F_n) : \mathbb{D} \rightarrow \mathbb{B}^n$ is a holomorphic map the the pull-back F^* commutes with the ∂ and $\bar{\partial}$ operators. Hence, the induced metric on \mathbb{D} is given by the Kähler form

$$F^*(k_B) = -i(n+1)\partial\bar{\partial}\log(1-|F|^2).$$

Up to a constant, the metric is therefore given by

$$F^*(ds_{\mathbb{B}}^2) = \frac{n+1}{2}\Delta\log(1-|F(\zeta)|^2)|d\zeta|^2.$$

It might now be possible to complete the calculation of the curvature and use the comparison principle.

Concerning the conformal minimal disks, there might be another problem with using the Beltrami-Klein metric (6.8). This metric is not conformally equivalent to the Poincaré metric (6.7). Hence, disks which are conformal harmonic in the standard flat (Euclidean) metric (and hence in the Poincaré metric) are no longer conformal in the Beltrami-Klein metric. Maybe this is not essential since in calculations we get various expressions of standard inner products which may anyhow simplify if the the disk is conformal in the Euclidean metric.

7. Further results

7.1. Boundary Schwarz lemma and harmonic functions

For the next results see [64, 65]. Define the harmonic density

$$\text{Har}(w) = \frac{1}{\sqrt{2}} \frac{1}{\hat{R}(w)} \quad (7.1)$$

on \mathbb{U} , where $\hat{R}(w) = \sqrt{1-|w|^2}$ and denote by d_{har} the corresponding distance.

Theorem 17. *If f is a harmonic mapping from the unit disk \mathbb{D} into self, then*

$$d_{\text{har}}(fz, fz') \leq d_{\text{hyp}}(z, z'), \quad z, z' \in \mathbb{D}.$$

Define $A(z) = (1+z)(1-z)^{-1}$, $A_g = (1+g)(1-g)^{-1}$, $B_g = R_g = A_g - A$ and $h = \text{Re}(A_g - A)$. Then $A' = 2(1-z)^{-2}$, and if g is a holomorphic function, we have $A'_g = 2(1-g)^{-2}g'$, and $R'_g = 2(1-g)^{-2}g' - 2(1-z)^{-2}$.

Ornek-Akyel [46, 47] use the following form of maximum principle:

Proposition 7.1. If u is harmonic on the unit disk and for every $w \in \mathbb{T}$, $\liminf_{z \rightarrow w} u(z) \geq 0$, then $u \geq 0$.

Theorem 18. Let $B : \mathbb{D} \rightarrow \mathbb{D}$ be a finite Blaschke product which equal w_0 on a finite set $A \subset \mathbb{T}$ and let f be a holomorphic function in the unit disc and $|f(z) - 1| < 1$, for $|z| < 1$. Suppose the following condition is satisfied
(ii.2) For all $a \in A$, $f(z) = 1 + B(z) + o(z-a)^2$, $z \in \mathbb{T}$, $z \rightarrow a$.

Then $u(z) = \text{Re}A_F - \text{Re}A_B$ is continuous on $\mathbb{D} \cup A$ and satisfies the condition

(ii.1) $\liminf_{z \rightarrow w} u(z) \geq 0$ for every $w \in \mathbb{T}$, and it is non-negative on \mathbb{D} .

For every $a \in A$, $u(z) = o(1)$ if z tends a .

Theorem 19. Let $B : \mathbb{U} \rightarrow \mathbb{U}$ be a finite Blaschke product which equal w_0 on a finite set $A \subset \mathbb{T}$ and let f be a holomorphic function in the unit disc and $|f(z) - 1| < 1$, for $|z| < 1$. Suppose in addition to (ii.2) the following conditions is satisfied

(ii.3) there is a $a_0 \in A$, such that $f(z) = 1 + B(z) + o(z-a_0)^3$, $z \in \mathbb{T}$, $z \rightarrow a_0$. Then $u(z) = o(z-1)$, if z tends 1, and $f = 1 + B$.

In joint work with M. Knežević and the author, it is proved:

Proposition 7.2 (The unit disk euclidean-qch version, [32]). Let f be a k -quasiconformal euclidean harmonic mapping from the unit disc \mathbb{D} into itself. Then for all $z \in \mathbb{D}$ we have

$$|f_z(z)| \leq \frac{1}{1-k} \frac{1-|f(z)|^2}{1-|z|^2}.$$

Using $L_f(z) \leq (1+k)|f_z(z)|$, we get

$$(A) L_f(z) \leq K \frac{1-|f(z)|^2}{1-|z|^2} \text{ and therefore}$$

$$(A1) \lambda(fz', fz) \leq K \lambda(z', z), z', z \in \mathbb{D}.$$

In proof we use the metric density

$$\sigma_f(z) = (1-k)^2 \lambda(f(z)) |f_z(z)|^2, z \in \mathbb{D},$$

and check that the curvature $\mathbf{K}(\sigma)(z) \leq -1$.

In communication with Pavlović appears the following question:

Question 2. Whether (A1) holds if f is k-qr? We announce a positive answer to this question in [42]:

Theorem 12. (i): Let f be a k -quasiregular euclidean harmonic mapping from the unit disc \mathbb{D} into itself.

Then for any two points z_1 and z_2 in \mathbb{D} we have

$$\lambda(f(z_1), f(z_2)) \leq \frac{1+k}{1-k} \lambda(z_1, z_2).$$

8. Further comments

Because of limited space we mention only a few papers related to Schwarz lemma holomorphic maps. We first recall the definition of normality. Let X, Y be two complex manifolds; a family F of holomorphic maps from X to Y is normal if every sequence in F admits either a convergent subsequence or a compactly divergent subsequence. A complex manifold X is taut if $\text{Hol}(\mathbb{D}, X)$ is a normal family. Let X be a taut complex manifold. Then $\text{Hol}(Y, X)$ is a normal family for every complex manifold Y . A connected complex manifold X is (Kobayashi) hyperbolic if k_X is a true distance. Every complete hyperbolic manifold is taut.

1. For a review about Schwarz lemma holomorphic maps between Kahler manifolds see Jianbo Chen [14], Abstract. In Section 1, we introduce some background knowledge of complex geometry. In Section 2, classical Schwarz lemma and its interpretation is discussed. In Section 3, we study the Ahlfors-Schwarz's lemma and its generalization to holomorphic maps between the unit disk and Kahler manifolds with holomorphic sectional curvature bounded from above by a negative constant. In Section 4, we focus on the case when equality holds at a certain point is discussed for holomorphic maps between the unit disk and classical bounded symmetric domains of type I, II and III. In Section 5, two higher-dimensional generalizations of the Ahlfors-Schwarz lemma for holomorphic maps from a compact Kahler manifold to another Kahler manifold, both of which satisfy respective conditions on curvature, are studied. In Section 6, we investigate two applications of various versions of Ahlfors-Schwarz lemma.

2. James J. Faran [22], Abstract. The problem of local equivalence of Hermitian Finsler metrics under holomorphic changes of coordinates is solved. On such a Finsler metric we find some differential conditions which imply that the Finsler metric is the Kobayashi metric of the underlying manifold (these conditions are satisfied if the metric is the Kobayashi metric on a bounded, strictly convex domain in \mathbb{C}^n with smooth boundary).

3. Gunnar Pór Magnússon [35], Abstract. Lars Ahlfors proved a differential geometric version of the classical Schwarz lemma in 1938. His version of the lemma gives an interesting connection between the existence of non-constant entire functions with values in a given domain and metrics with negative curvature on such domain. We recall the classical Schwarz lemma and review the notions necessary to understand Ahlfors' lemma, before proving both the new form of the lemma and giving some applications.

4. In the survey [21] (see Abstract), C. Frosini and F. Vlacci give geometric interpretations of some standard results on boundary behaviour of holomorphic self-maps in the unit disc of \mathbb{C} and generalize them for holomorphic self-maps of some particular domains of \mathbb{C}^n .

5. Marco Abate [3], Abstract. These are the notes of a short course I gave in the school "Aspects métriques et dynamiques en analyse complète", Lille, May 2015. The aim of this notes is to describe how to use a geometric structure (namely, the Kobayashi distance) to explore and encode analytic properties of holomorphic functions and maps defined on complex manifolds. We shall first describe the main properties of the Kobayashi distance, and then we shall present applications to holomorphic dynamics in taut manifolds, strongly pseudo convex domains and convex domains, and to operator theory in Bergman spaces (Carleson measures and Toeplitz operators).

Theorem 20 (Theorem 2.40 Budzynska; Abate-Raissy, [3]). *Let $D \subset \subset \mathbb{C}^n$ be a bounded strictly convex domain, and take $f \in \text{Hol}(D, D)$ without fixed points. Then the sequence of iterates $\{f^k\}$ converges to a constant map defined by $z_0 \in \partial D$.*

As often happens with objects introduced via a general definition, the Kobayashi pseudodistance can seldom be explicitly computed. "Besides the cases listed in Proposition 1.17 [3], as far as we know there are formulas only for some complex ellipsoids [39], bounded symmetric domains [38], the symmetrized bidisk [11] and a few other scattered examples." Recall that in particular (see [30], p.47):

$$\text{Kob}(z, w) = \max\{\text{Kob}(z_k, w_k) : k = 1, \dots, n\}.$$

On the other hand, it is possible and important to estimate the Kobayashi distance; see Subsection 1.5 [3].

6. For Iteration theory of holomorphic maps on taut manifolds we refer the interested reader to M. Abate monograph[4], in which a Wolff-Denjoy theorem for hyperbolic Riemann surfaces is proved:

Theorem 21 (Theorem 1.3.12, [4]). *Let X be a hyperbolic Riemann surface, and let $f \in \text{Hol}(X, X)$. Then either:*

- (i) f has an attractive fixed point in X , or
- (ii) f is a periodic automorphism, or
- (iii) f is a pseudoperiodic automorphism, or
- (iv) the sequence $\{f^k\}$ is compactly divergent. Furthermore, the case (iii) can occur only if X is either simply connected (and f has a fixed point) or doubly connected (and f has no fixed points).

If X is compact, Theorem 1.3.12 drastically simplifies, becoming: Corollary 1.3.13: Let X be a compact hyperbolic Riemann surface. Then every function $f \in \text{Hol}(X, X)$ is either constant or a periodic automorphism.

7. Suzuki, Masaaki, citeSuzu-Abstract: In this paper we study the intrinsic metrics for the circular domains in \mathbb{C}^n . We calculate the Kobayashi (pseudo-) metric at its center for pseudoconvex complete circular domain D using the result of Sadullaev. From this we have that such D is hyperbolic iff D is bounded. If a convex complete circular domain is complete hyperbolic, then the Caratheodory and Kobayashi metrics coincide at the center. Using this and the results of Hua we explicitly compute the intrinsic metrics of the classical domains. Furthermore we define the extremal function and extremal disc for intrinsic metrics and compute them in some special cases.

8. In [1], the author (Simonic) introduces Ahlfors' generalization of the Schwarz lemma. With this powerful geometric tool of complex functions in one variable, he is able to prove some theorems concerning the size of images under holomorphic mappings, including the celebrated Picard's theorems. The article concludes with a brief insight into the theory of Kobayashi hyperbolic complex manifolds.

9. Filippo Bracci, John Erik Fornaess, Erlend Fornaess Wold, [10]-Abstract: We prove that for a strongly pseudoconvex domain $D \subset \mathbb{C}^n$, the infinitesimal Carathéodory metric $C(z, v)$ and the infinitesimal Kobayashi metric $K(z, v)$ coincide if z is sufficiently close to bD and if v is sufficiently close to being tangential to bD . Also, we show that every two close points of D sufficiently close to the boundary and whose difference is almost tangential to bD can be joined by a (unique up to reparameterization) complex geodesic of D which is also a holomorphic retract of D . The same continues to hold if D is a worm domain, as long as the points are sufficiently close to a strongly pseudoconvex boundary point. We also show that a strongly pseudoconvex boundary point of a worm domain can be globally exposed, this has consequences for the behavior of the squeezing function.

10. For Pluriharmonic Functions in Balls see Rudin [60]-Abstract: It is proved that a function is pluriharmonic in the open unit ball of \mathbb{C}^n if and only if it is harmonic with respect to both the ordinary Laplacian and the invariant Laplace-Beltrami operator.

Theorem 22 ([64]). *Suppose that $G = G_n = (a, b) \times \mathbb{R}^{n-1}$, $-\infty < a < b \leq \infty$, $f : B_n \rightarrow G_n$. If f_1 is pluriharmonic on B_n and f is K -qc at $a \in B_n$, then $k_G(fa)|f'(a)| \leq Kk_B(a)$, where k is quasi-hyperbolic density.*

9. Appendix- Schwarz lemma 2

The model of the hyperbolic plane is the half-plane model. The underlying space of this model is the upper half-plane model \mathbb{H} in the complex plane \mathbb{C} , defined to be $\mathbb{H} = \{z \in \mathbb{C} : \text{Im}(z) > 0\}$. In coordinates (x, y) the line element is defined as $ds^2 = \frac{1}{y^2}(dx^2 + dy^2)$. The geodesics of this space are semicircles centered on the x-axis and vertical half-lines. The geometrical properties of the figures on the half-plane are studied by considering quantities invariant under an action of the general Möbius group, which consists of compositions of

Möbius transformations and reflections [4]. The curvilinear triangle formed by circular arcs of three intersecting semicircles is one of the principal figures of the upper half-plane model \mathbb{H} . The hyperbolic laws of sines-cosines for that triangle are proved by using properties of the Möbius group and the upper half-plane \mathbb{H} .

In [57] Yamaleev suggests another way of construction of proofs of the sines-cosines theorems of the Poincaré model. The curvilinear triangle formed by circular arcs is the figure of the Euclidean plane; consequently, on the Euclidean plane we have to find relationships antecedent to the sines-cosines hyperbolic laws. Therefore, first of all, we establish these relationships by making use of axioms of the Euclidean plane, only. Secondly, we prove that these relationships can be formulated as the hyperbolic sine-cosine theorems. For that purpose we refer to the general complex calculus and within its framework establish a relationship between exponential function and the cross-ratio. In this way the hyperbolic trigonometry emerges on Euclidean plane in a natural way.

For the benefit of the reader we add some details concerning some parts of subsection 2.2. 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

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

Let γ 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 λ as the oriented distance which changes the sign of the permutation z_1 and z_2 . Additivity of the distance on geodesics follow from $(z_2, z_1, w_1, w_2) = (z_2, z_3, w_1, w_2)(z_3, z_1, w_1, w_2)$.

Let $K = K(z_1, z_2)$ circle orthogonal on T throughout points z_1 i z_2 and denote by a and b the intersection points K and T . Usually we denote the intersection points such that z_1 is between a and z_2 .

Recall $[z_1, z_2; a, b] = \frac{z_1-a}{z_1-b} : \frac{z_2-a}{z_2-b}$. For example if $a = -1, b = 1$, then according convention about the notation we write $z_1 = 0$ i $z_2 = r, 0 < r < 1$.

Since $[0, r; -1, 1] = \frac{1-r}{1+r}$, it is $0 < [0, r; -1, 1] < 1$. Therefore it is convenient to define $\{z_1, z_2\} = [z_2, z_1; a, b]_2 = \frac{z_2-a}{z_2-b} : \frac{z_1-a}{z_1-b}$.

Check that according our convention on notation $\{z_1, z_2\} > 1$ and $\{z_1, z_2\} \cdot \{z_2, z_3\} = \{z_1, z_3\}$. Define $d_{hyp}(z_1, z_2) = \ln\{z_1, z_2\}$. Check $\{0, r\} = \frac{1+r}{1-r}$.

Define $T_{z_1}(z) = \frac{z-z_1}{1-\bar{z}_1z}, \varphi_{z_1} = -T_{z_1}$ and

$\delta(z_1, z_2) = |T_{z_1}(z_2)| = |\frac{z-z_1}{1-\bar{z}_1z}|$. Schwarz's lema yields motivation to introduce the hyperbolic distance: If $f \in \text{Hol}(\mathbb{D}, \mathbb{D})$, then $\delta(fz_1, fz_2) \leq \delta(z_1, z_2)$.

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

Hence

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

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

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

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

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

If γ 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 γ in \mathbb{D} joining the points z_1 and z_2 .

We summarize

$$\lambda_{\mathbb{D}} = \ln \frac{1 + \delta_{\mathbb{D}}}{1 - \delta_{\mathbb{D}}}, \quad \lambda_{\mathbb{H}} = \ln \frac{1 + \delta_{\mathbb{H}}}{1 - \delta_{\mathbb{H}}} \quad (9.1)$$

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

Proposition 9.1. If $f \in \text{Hol}(\mathbb{D}, \mathbb{D})$, then $|\mathbf{v}^*|_\rho \leq |\mathbf{v}|_\rho$.

If γ 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 γ in \mathbb{D} joining the points z_1 and z_2 .

Let G be a simply connected domain different from \mathbb{C} and let $\phi : G \rightarrow \mathbb{D}$ be a conformal isomorphism. Define the pseudo hyperbolic distance on G by $\varphi_a^G(z) = \varphi_b(\phi(z))$, where $b = \phi(a)$, and $\delta_G(a, z) = |\varphi_a^G(z)| = \delta_{\mathbb{D}}(\phi(a), \phi(z))$. Verify that the pseudo hyperbolic distance on G is independent of conformal mapping ϕ . In particular, using conformal isomorphism $A(w) = A_{w_0}(w) = \frac{w-w_0}{w-\bar{w}_0}$ of \mathbb{H} onto \mathbb{D} , we find $\varphi_{H,w_0}(w) = A(w)$ and therefore $\delta_H(w, w_0) = |A(w)|$.

Proposition 9.2. The definition of δ_G is independent of conformal isomorphism.

Proof. Let $\phi, \phi_1 : G \rightarrow \mathbb{D}$ be two conformal isomorphism and let z_1, z_2 be two points in G and $w_k = \phi(z_k)$ and $w'_k = \phi_1(z_k)$, $k = 1, 2$. If $A = \phi_1^{-1} \circ \phi$, then $A \in \text{Aut}(\mathbb{D})$ and $A(w_k) = w'_k$, $k = 1, 2$. Hence $\delta_{\mathbb{D}}(w_1, w_2) = \delta_{\mathbb{D}}(w'_1, w'_2)$. \square

For a domain G in \mathbb{C} and $z, z' \in G$ we define $\delta_G(z, z') = \sup \delta_{\mathbb{D}}(\phi(z'), \phi(z))$, where the supremum is taken over all $\phi \in \text{Hol}(G, \mathbb{D})$.

Proposition 9.3. (a) If G and D are conformally isomorphic to \mathbb{D} and $f \in \text{Hol}(G, D)$, then

$$\delta_D(fz, fz') \leq \delta_G(z, z'), \quad z, z' \in G.$$

(b) The result holds more generally if G and D are hyperbolic domains.

Proof. Let $\phi : G \rightarrow \mathbb{D}$ and $\phi_1 : D \rightarrow \mathbb{D}$ be two conformal isomorphisms and set $F = \phi_1 \circ f \circ \phi^{-1}$. Next let z_1, z_2 be two points in G and $w_k = f(z_k)$, $\zeta_k = \phi(z_k)$ and $\zeta'_k = \phi_1(w_k)$, $k = 1, 2$. Then $\delta_{\mathbb{D}}(\zeta'_1, \zeta'_2) \leq \delta_{\mathbb{D}}(\zeta_1, \zeta_2)$ and the result follows.

A proof of (b) can be based on the fact if $\phi \in \text{Hol}(D, \mathbb{D})$, then $\phi \circ f \in \text{Hol}(G, \mathbb{D})$. \square

Set $d^0(z, w) = |\frac{z-w}{1-\bar{w}z}|$. In a response to a nice post, Principle of subordination (see [61]), says: "In most textbooks, the hyperbolic metric g is given first and then it is deduced that the isometry group is exactly the group of all conformal self maps of the disk i.e. disk-preserving holomorphic and antiholomorphic functions. Suppose we want to reverse this process and want to find a (hyperbolic) metric which is preserved by the conformal self maps. We observe from the equality case of the Pick's lemma that for $w = w(z)$ to be a conformal self map, $|\frac{dw}{dz}| = \frac{1-|w|^2}{1-|z|^2}$, i.e. $\frac{|dw|}{1-|w|^2} = \frac{|dz|}{1-|z|^2}$. Thus we find that the hyperbolic metric is given by (up to a positive factor) $g = \frac{|dz|^2}{(1-|z|^2)^2} = \frac{dx^2+dy^2}{(1-x^2-y^2)^2}$. Interestingly, the other form of Pick's lemma is given by $d^0(fz, fw) \leq d^0(z, w)$ with equality hold iff f is a conformal self map. It suggests (without integrating) that the hyperbolic distance is actually given by (up a scaling) $d^0(z, w) = |\frac{z-w}{1-\bar{w}z}|$. However, it is different from the usual definition $d(z, w) = \tanh^{-1}(|\frac{z-w}{1-\bar{w}z}|)$, I can't think if any way to "see" (except integrating) that the correct definition is the later one instead of the former one (actually I haven't checked if the former one really satisfies the triangle inequality)."

Note that the former one really satisfies the triangle inequality and we call it the pseudo-hyperbolic distance. The pseudo-hyperbolic distance is not additive along hyperbolic geodesics.

9.1. Klein model

Recall for given two distinct points U and V in the open unit ball of the model in Euclidean space, the unique straight line connecting them intersects the unit sphere at two ideal points A and B , labeled so that the points are, in order along the line, A, U, V, B . Taking the centre of the unit ball of the model as the origin, and assigning position vectors u, v, a, b respectively to the points U, V, A, B , we have that that $|a - v| > |a - u|$ and $|u - b| > |v - b|$, where $|\cdot|$ denotes the Euclidean norm. Then the distance between U and V in the modelled hyperbolic space is expressed as

$$d(\mathbf{u}, \mathbf{v}) = \frac{1}{2} \log \frac{\|\mathbf{v} - \mathbf{a}\| \|\mathbf{b} - \mathbf{u}\|}{\|\mathbf{u} - \mathbf{a}\| \|\mathbf{b} - \mathbf{v}\|},$$

where the factor of one half is needed to make the curvature -1 .

The associated metric tensor is given by

$$ds^2 = g(\mathbf{x}, d\mathbf{x}) = \frac{\|d\mathbf{x}\|^2}{1 - \|\mathbf{x}\|^2} + \frac{(\mathbf{x} \cdot d\mathbf{x})^2}{(1 - \|\mathbf{x}\|^2)^2}.$$

More precisely, if $\mathbf{v} \in T_{\mathbf{x}}\mathbb{R}^n$, then

$$ds^2(\mathbf{v}) = \text{Kle}(\mathbf{x}, \mathbf{v}) = \frac{\|\mathbf{v}\|^2}{1 - \|\mathbf{x}\|^2} + \frac{(\sum_{k=1}^n x_k v_k)^2}{(1 - \|\mathbf{x}\|^2)^2}. \quad (9.2)$$

Acknowledgement. We have discussed the subject at Belgrade Analysis seminar (in 2016) and in particular in connection with minimal surfaces with F. Forstnerič and get useful information about the subject via Forstnerič [56]. We are indebted to the members of the seminar and to professor F. Forstnerič for useful discussions.

References

- [1] Aleksander Simonic, *The Ahlfors lemma and Picard's theorems*, arXiv:1506.07019v1 [math.CV] 22 Jun 2015
- [2] Abate M., Patrizio G. , *Holomorphic curvature of Finsler metrics and complex geodesics*, Preprint, Max-Planck-Institut fur Matematik, Bonn (1992)
- [3] Marco Abate, The Kobayashi distance in holomorphic dynamics and operator theory, <http://www.dm.unipi.it/abate/articoli/artric/files/AbateLille.pdf>
- [4] Abate,M.: Iteration theory of holomorphic maps on taut manifolds. Mediterranean Press, Cosenza(1989) [See also <http://www.dm.unipi.it/abate/libri/libriric/libriric.html>
- [5] Ahlfors, L., *Conformal invariants*, McGraw-Hill Book Company, 1973.
- [6] Ahlfors,L.V.: Möbius transformations in several dimensions. -Lecture Notes, University of Minnesota, 1981
- [7] A. F. Beardon and D. Minda, *The hyperbolic Metric and Geometric Function Theory*, The Proceedings of IWQCMA05, Quasiconformal Mappings and their applications, Editors: S. Ponnusamy, T. Sugawa and M. Vuorinen, 2007, Narosa Publishing House, New Delhi, India, p. 9-56
- [8] F. Bracci, G. Patrizio, Monge-Ampere foliations with singularities at the boundary of strongly convex domains. *Math. Ann.* 332 (2005), 499-522.
- [9] F. Bracci, G. Patrizio, S. Trapani, The pluricomplex Poisson kernel for strongly convex domains. *Trans. Amer. Math. Soc.* 361 (2009), 979-1005.
- [10] Filippo Bracci, John Erik Fornaess, Erlend Fornaess Wold, *Comparison of invariant metrics and distances on strongly pseudoconvex domains and worm domains*, arXiv:1710.04192v2 [math.CV] 15 Oct 2017.
- [11] D.M. Burns and S.G. Krantz, Rigidity of holomorphic mappings and a new Schwarz Lemma at the boundary, *J. Amer. Math. Soc.* 7 (1994), 661-676.
- [12] J. Burbea, *On the Hessian of the Caratheodory metric*, *Rocky Mountain Journal of Mathematics*, Volume 8, Number 3, Summer 1978
- [13] 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>
- [14] Jianbo Chen, *Generalizations of Schwarz Lemma*, The final year report of Math 2999 under the guidance of Professor Ngaiming Mok, <https://www.stat.berkeley.edu/~jianbo/schwarz.pdf>
- [15] Cheng, S.Y. and Yau, S.T., *Differential equations on Riemannian manifolds and their geometric applications*, *Comm.Pure and Appl. Math.* **28** (1975) 333-354
- [16] V. N. Dubinin, The Schwarz inequality on the boundary for functions regular in the disc, *J. Math. Sci.* 122 (2004), 3623-3629.
- [17] P. DUREN, *Harmonic mappings in the plane*, Cambridge Univ. Press, 2004.
- [18] L. Carleson, T. D. W. Gamelin, Complex dynamics, Universitext: Tracts in Mathematics, Springer-Verlag, ISBN 0-387-97942-5,(1993)
- [19] C. C. Cowen, Fixed Points of Functions Analytic in the Unit Disk, Conference on Complex Analysis, University of Illinois, May 22, 2010, <https://www.math.iupui.edu/~ccowen/Talks/FixPts1005.pdf>
- [20] C. Earle and R. Hamilton, A fixed point theorem for holomorphic mappings, *Proc. Symp. Pure Math.*, Vol. XVI, 1968, 61-65.
- [21] C. Frosini, F. Vlacci, *A survey on geometric properties of holomorphic self-maps in some domains of \mathbb{C}^n* , <http://web.math.unifi.it/users/frosini/SurveyFrosini-Vlacci.pdf>
- [22] James J. Faran, *Hermitian Finsler metrics and the Kobayashi metric*. *J. Differential Geom.* 31 (1990), no. 3, 601–625. doi:10.4310/jdg/1214444630.

- [23] Gardiner,F.P., *Teichmüller Theory and Quadratic Differentials*, New York: Awiley-Interscience Publication, 1987.
- [24] Gentili G. and Visintin B., *Finsler Complex Geodesics and Holomorphic Curvature*, Rendiconti Accad. XL Memorie di Matematica bf 111 (1993) 153-170
- [25] Graziano Gentili and Fabio Vlacci, *Rigidity for regular functions over Hamilton and Cayley numbers and a boundary Schwarz' Lemma*, Indag. Mathem., N.S., 19 (4), 535-545 December, 2008
- [26] Robert E. Greene, Kang-Tae Kim, Steven G Krantz, Applications of Bergman Geometry, The Geometry of Complex Domains - Page 90 <https://books.google.rs/books/isbn=0817646221>
- [27] Clifford J. Earle, Lawrence A. Harris, John H. Hubbard, Sudeb Mitra, Schwarz's lemma and the Kobayashi and Carathéodory pseudometrics on complex Banach manifolds, Kleinian Groups and Hyperbolic 3-Manifolds: Proceedings of the Warwick Workshop September 2001, Lond. Math. Soc. Lec. Notes **299**, 363-384, edited by Y. Komori, V. Markovic, C. Series
- [28] Lawrence A. Harris, Fixed Point Theorems for Infinite Dimensional Holomorphic Functions, J. Korean Math. Soc. 41(2004), No. 1, pp. 175-192, http://www.ms.uky.edu/larry/paper.dir/fixed_pt.pdf
- [29] M. Heins, *On a class of conformal metrics*, Nagoya Math. J. Volume 21 (1962), 1-60.
- [30] S. Kobayashi, Hyperbolic manifolds and holomorphic mappings, Marcel-Dekker, New York, 1970.
- [31] Steven G. Krantz, *The Carathéodory and Kobayashi Metrics and Applications in Complex Analysis*, arXiv:math/0608772v1 [math.CV] 31 Aug 200
- [32] M. Knežević, M. Mateljević, On the quasi-isometries of harmonic quasi-conformal mappings J. Math. Anal. Appl, 2007, 334(1), 404-413.
- [33] M. Knežević, Harmojska i kvazikonformna preslikavanja.kvazi - izometrije i krivina,2014, http://elibrary.maf.bg.ac.rs/bitstream/handle/123456789/4280/phdKNEZEVIC_MILJAN.pdf
- [34] L. Lempert, *La métrique de Kobayashi et la représentation des domaines sur la boule*, Bull. Soc. Math. France 109 (1981),427-474
- [35] Gunnar Pór Magnússon, *A geometric version of the Schwarz lemma*, A talk given at KAUS 2010,<https://notendur.hi.is/gthm1/files/kaus2010-schwarz.pdf>
- [36] Mateljević, M., *Holomorphic fixed point theorem on Riemann surfaces*, Math. Balkanica **12** (1-2) (1998), 1-4.
- [37] M. Mateljević, *Ahlfors-Schwarz lemma and curvature*, Kragujevac Journal of Mathematics (Zbornik radova PMF), Vol. **25**, 2003, 155-164.
- [38] M. Mateljević: Kompleksne Funkcije 1 & 2, Društvo matematičara Srbije, 2006.
- [39] M. MATELJEVIĆ, *Topics in Conformal, Quasiconformal and Harmonic maps*, Zavod za udžbenike, Beograd 2012.
- [40] M. MATELJEVIĆ, *Hyperbolic geometry and Schwarz lemma*, VI Simpozijum Matematika i primene, p.1-17,Beograd, November 2016
- [41] M. Mateljević, (a) *Schwarz lemma and Kobayashi Metrics for holomorphic and pluriharmonic functions*, arXiv:1704.06720v1[math.CV] 21 Apr 2017
(b) *Schwarz lemma, Kobayashi Metrics and FPT*, preprint November 2016.
- [42] M. Mateljević, *Schwarz Lemma and Kobayashi Metrics for Holomorphic Functions*, Filomat Volume 31, Number 11, 2017, 3253-3262
- [43] M. Mateljević, Hyperbolic geometry and Schwarz lemma, VI Simpozijum Matematika i primene, p.1-17,Beograd, November 2016
- [44] M. Mateljević, (a) *Schwarz lemma and Kobayashi Metrics for holomorphic and pluriharmonic functions*, arXiv:1704.06720v1 [math.CV] 21 Apr 2017
(b) *Schwarz lemma, Kobayashi Metrics and FPT*, preprint November 2016.
- [45] D. Minda, *The strong form of Ahlfors' lemma*. Rocky Mountain J. Math. Volume 17, Number 3 (1987), 457-462.
- [46] B. Örnek and T. Akyel, A Representation with majorant of the Schwarz lemma at the boundary, manuscript July 2015.
- [47] T. Akyel and B. Örnek, Some Remarks on Schwarz lemma at the boundary, manuscript August 2015.
- [48] B. Örnek, Estimates for holomorphic functions concerned with Jack's lemma (manuscript 15 156 Publ).
- [49] Robert Osserman, From Schwarz to Pick to Ahlfors and Beyond, Notice AMS **46**, 8, 868-873.
- [50] R. Osserman, *A sharp Schwarz inequality on the boundary*, Proc. Amer. Math. Soc. 128 (2000), 3513-3517.
- [51] N. Saito, *On a metric induced by analytic capacity. I.*, Kodai Math. Sem. Rep. 25 (1973), 215-218
- [52] B. V. Sabat, Vvedenie v kompleksnyi analiz,I, II 1976, Introduction to Complex Analysis, translated by American mathematical society, 1992.
- [53] M. VUORINEN: *Conformal Geometry and Quasiregular Mappings*, Lecture Notes in Math. 1319, Springer-Verlag, Berlin-New York, 1988.
- [54] B. Wong, *On the Holomorphic Curvature of Some Intrinsic Metrics*, Proceedings of the American Mathematical Society Vol. 65, No. 1 (Jul., 1977), pp. 57-61
- [55] H. Wu, *Normal families of holomorphic mappings*, Acta Math. 119 (1967), 193-233. MR 31# 468.
- [56] Communication with Forstnerič.
- [57] Robert M. Yamaleev, *Hyperbolic Cosines and Sines Theorems for the Triangle Formed by Arcs of Intersecting Semicircles on Euclidean Plane*, Journal of Mathematics Volume 2013, Article ID 920528, 10

pages,<http://dx.doi.org/10.1155/2013/920528>

- [58] Yau, S., T., *Harmonic functions on complete Riemannian manifolds*, Comm.Pure and Appl. Math. **28** (1975) 201-228
- [59] Rudin, W., *Function Theory in the Unit Ball of C^n* , Springer-Verlag, Berlin Heidelberg New York, 1980.
- [60] W. Rudin, *Pluriharmonic Functions in Balls*, Proceedings of the American Mathematical Society Vol. 62, No. 1 (Jan., 1977), pp. 44-46
- [61] <https://cuhkmath.wordpress.com/2012/03/04/principle-of-subordination/>, Principle of subordination Posted on 04/03/2012 by leonardwong
- [62] SUBORDINATION PRINCIPLE IN UNIVALENT FUNCTIONS THEORY. Available from: https://www.researchgate.net/publication/242278646_SUBORDINATION_PRINCIPLE_IN_UNIVALENT_FUNCTIONS THEORY [accessed Oct 3, 2016]. bibitemUn H. Unkelbach, *Über die Randverzerrung bei konformer Abbildung*, Math. Z., 43 (1938), 739-742.
- [63] Suzuki, Masaaki. *The intrinsic metrics on the circular domains in C^n* . Pacific J. Math. 112 (1984), no. 1, 249–256. <http://projecteuclid.org/euclid.pjm/1102710113>.
- [64] M. Mateljević, Communications at Analysis Seminar, University of Belgrade.
- [65] https://www.researchgate.net/post/What_are_the_most_recent_versions_of_The_Schwarz_Lemma [accessed Jul 31, 2017].
How to solve an extremal problems related to harmonic functions?. Available from: https://www.researchgate.net/post/How_to_solve_a_extremal_problems_related_to_harmonic_functions [accessed Aug 3, 2017].
- [66] P. Melentijević, Invariant gradient in refinements of Schwarz lemma and Harnack inequalities, manuscript.
- [67] M. Marković, On harmonic functions and the hyperbolic metric, Indag. Math., 26(1):19-23, 2015.

Conormal Lagrangian Floer homology for open subsets and PSS isomorphism

Jelena Katić

Matematički fakultet, Studentski trg 16, 11000 Belgrade, Serbia
e-mail: jelenak@matf.bg.ac.rs

Darko Milinković

Matematički fakultet, Studentski trg 16, 11000 Belgrade, Serbia
e-mail: milinko@matf.bg.ac.rs

Jovana Nikolić

Matematički fakultet, Studentski trg 16, 11000 Belgrade, Serbia
e-mail: jovanadj@matf.bg.ac.rs

Abstract. We sketch the construction of PSS isomorphism between Morse homology of an open subset U of a closed manifold M and Floer homology of U defined by Kasturirangan and Oh in [16]. We used this isomorphism in [18] to construct symplectic invariants for the open set U .

Keywords: Morse homology; Lagrangian submanifolds; Floer homology; PSS isomorphism.

1. Morse homology

Let M be a closed manifold and $f : M \rightarrow \mathbb{R}$. Denote by $\text{Crit}(f)$ the set of critical points of f . Morse chain complex is a \mathbb{Z}_2 -vector space (or a \mathbb{Z} -module) generated by $\text{Crit}(f)$.

For two critical point $p, q \in \text{Crit}(f)$, consider an ordinary differential equation:

$$\dot{\gamma} = -\nabla f \quad (1)$$

with the boundary condition

$$\gamma(-\infty) = p, \quad \gamma(+\infty) = q.$$

This equation does not have to have a solution, or can have infinitely many of them. However, for a suitable choice of p, q (depending on Morse index) and the generic Riemannian metric defining the gradient ∇ , there are only finitely many solutions. In that case, let $n(p, q)$ denote its number.

Now define a boundary operator as:

$$\partial_M : p \mapsto \sum_q n(p, q)q,$$

for those q for which this number is finite.

We will denote by $HM(f)$ the corresponding homology group, known as Morse homology.

It turns out that Morse homology does not depend neither on the choice of a Morse function nor on the choice of Riemannian metric (that defines the gradient ∇ in (1)). Moreover, it is isomorphic to the singular homology of M :

$$HM(f) \cong H^{\text{sing}}(M).$$

See [2, 22] for the details on Morse theory.

2. Lagrangian Floer homology

Lagrangian Floer homology is a generalization of Morse homology. Namely, let T^*M be a cotangent bundle over M , and O_M its zero-section. Note that the differential df of f is an one-form and can be considered as a map

$$df : O_M \rightarrow T^*M.$$

The set $\text{Crit}(f)$ is the intersection

$$O_M \cap df(O_M).$$

Now, both O_M and $df(O_M)$ are *Lagrangian submanifolds* of T^*M , meaning that the canonical symplectic form $\omega_0 := \sum dq_j \wedge dp_j$ vanishes on them, and that their dimension is half of the dimension of T^*M .

If (P, ω) is any symplectic manifold of dimension $2n$, then $L \subset P$ is *Lagrangian submanifold* if

$$\dim L = n, \quad \omega|_{TL} = 0.$$

For two Lagrangian submanifolds $L_0, L_1 \subset P$, we consider

$$L_0 \cap L_1$$

as the set of generators of *Floer chain complex*. *Floer boundary operator* is defined via pseudo-holomorphic strips with Lagrangian boundary conditions. More precisely, consider the partial differential equation

$$u : \mathbb{R} \times [0, 1] \rightarrow P, \quad \bar{\partial}u = 0, \quad u(s, j) \in L_j. \quad (2)$$

The set of u 's satisfying (2) is a union of finite-dimensional manifolds, whose dimensions depend on $u(\pm\infty, t) \in L_0 \cap L_1$.

Let $x \in L_0 \cap L_1$. Floer boundary operator ∂_F is defined as

$$\partial_F : x \mapsto \sum_y n(x, y)y$$

where $n(x, y)$ is defined as the cardinality of the set of u solving (2) and the condition

$$u(-\infty, t) = x, \quad u(+\infty, t) = y,$$

for those y for which this number is finite.

The homology groups of this complex are called *Floer homology groups* and denoted by $HF(L_0, L_1)$. It was introduced and developed by Floer in the series of papers [5–12].

Lagrangian Floer homology is not always well defined. We also refer the reader to [13] for the details on Lagrangian intersection Floer theory.

One case when Lagrangian Floer homology is well defined is the following. Let M be a closed manifold, $L_0 = O_M \subset T^*M$, $L_1 = \phi_H^1(O_M)$, where ϕ_H^1 is a time one map of the Hamiltonian flow $\{\phi_H^t\}$ induced by a smooth Hamiltonian $H : T^*M \rightarrow \mathbb{R}$. Floer homology $HF(O_M, \phi_H^1(O_M))$ is then well defined and independent of the choice of H . Moreover, it is isomorphic to the singular homology of M :

$$HF(O_M, \phi_H^1(O_M)) \cong HM^{\text{sing}}(M).$$

The above isomorphism was proved by Floer. He actually proved that the Morse and Floer chain groups and boundary operators coincide:

$$CF(O_M, \phi_H^1(O_M)) = CM(f), \quad \partial_F = \partial_M$$

for the special choice of f and H . Then, since both homology do not depend on parameters involved (i.e. f, H), it is true always.

3. PSS isomorphism

As we already mentioned, there exists a canonical isomorphism for Morse homology for different Morse functions. We will denote it by

$$S_{\alpha\beta} HM(f_\alpha) \rightarrow HM(f_\beta),$$

for two Morse functions $f_\alpha, f_\beta : M \rightarrow \mathbb{R}$. The isomorphism $S_{\alpha\beta}$ is defined in a similar way as the boundary operator ∂_M , via the number of certain gradient trajectories (see [22]). In Floer case, we denote the corresponding canonical isomorphism by

$$T_{\alpha\beta} : HF(O_M, \phi_{H_\alpha}^1(O_M)) \rightarrow HF(O_M, \phi_{H_\beta}^1(O_M)).$$

The map $T_{\alpha\beta}$ is defined by the number of pseudo-holomorphic discs with Lagrangian boundary condition (similarly to ∂_F). Since the isomorphism between Morse and Floer homology is realized even on chain level, it is not clear whether the diagram

$$\begin{array}{ccc} HM(f_\alpha) & \xrightarrow{S_{\alpha\beta}} & HM(f_\beta) \\ PSS_\alpha \downarrow & & \downarrow PSS_\beta \\ HF(O_M, \phi_{H_\alpha}^1(O_M)) & \xrightarrow{T_{\alpha\beta}} & HF(O_M, \phi_{H_\beta}^1(O_M)) \end{array} \quad (3)$$

commutes.

The PSS isomorphism

$$PSS : HM(F) \rightarrow HF(O_M, \phi_H^1(O_M))$$

is defined by a number of combined objects (γ, u) , where γ is a negative gradient trajectory and u is a holomorphic strip with Lagrangian boundary conditions (see Figure 1). It is originally defined by Piunikhin, Salamon and Schwarz [20] for the case of Hamiltonian periodic case (which is a special case of Lagrangian intersection). PSS isomorphism for the case of cotangent bundle (as given in (3)) is constructed in [17] and it is generalized by Albers [1] for some wider class of symplectic manifolds (P, ω) and Lagrangian submanifolds L . PSS isomorphism is functorial, meaning that the diagram (3) commutes.

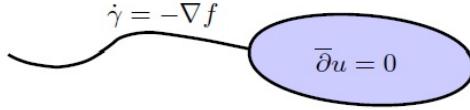


Figure 1. Mixed-type object that defines PSS

4. Floer homology for submanifolds

Let $N \subset M$ be a closed submanifold of a closed manifold M . Its conormal bundle is defined as:

$$\nu^*N := \{\alpha \in T^*M|_N \mid \alpha_{TN} = 0\}.$$

It is a Lagrangian submanifold of symplectic manifold (T^*M, ω_0) . In his thesis [21], Poźniak proved that Floer homology $HF(\nu^*N, \phi_H^1(O_M))$ is isomorphic to $H_{\text{sing}}(N)$. He also used Morse homology of N as a model for a singular one and the special choice of a Morse function f and a Hamiltonian H . This isomorphism is also established on the chain level and it is not functorial with respect to canonical isomorphisms in Morse and Floer homology.

PSS type isomorphism

$$HM(f, N) \xrightarrow{\cong} HF(\nu^*N, \phi_H^1(O_M))$$

is constructed in [3]. This isomorphism is now functorial, i.e. the diagram similar to (3) commutes.

5. Floer homology for open subsets

Let $U \subset M$ be an open set, such that ∂U is a smooth $(n - 1)$ -dimensional submanifold of M . In [16] Kasturirangan and Oh constructed Floer homology for U in the following way. Define

$$\nu_-^*(\partial U) := \{\alpha \in \nu^*(\partial U) \mid \alpha(\mathbf{n}) \leq 0, \text{ for } \mathbf{n} \text{ outward normal to } \partial U\}$$

and

$$\nu_-^*\overline{U} := O_U \cup \nu_-^*(\partial U).$$

We call $\nu_-^*\overline{U}$ a *conormal to \overline{U}* . The problem is that $\nu_-^*\overline{U}$ is not a smooth manifold, but there are approximations Υ_ε of $\nu_-^*\overline{U}$ (see Figure 2), such that Υ_ε is a smooth Lagrangian submanifold of (T^*M, ω_0) and that the Lagrangian Floer homology $HF(\phi_H^1(O_U), \Upsilon_\varepsilon)$ is well defined.

There exists a certain partial order on the set of approximations $\{\Upsilon_s\}$, and homomorphisms

$$F_{ab} : HF(\phi_H^1(O_U), \Upsilon_a) \rightarrow HF(\phi_H^1(O_U), \Upsilon_b)$$

such that

$$F_{bc} \circ F_{ab} = F_{ac}, \quad \text{for } \Upsilon_a \leq \Upsilon_b \leq \Upsilon_c.$$

Now define Floer homology of an open set as a direct limit

$$HF(U) := \varinjlim_s HF(\phi_H^1(O_U), \Upsilon_s).$$

Kasturirangan and Oh proved that

$$HF(U) \cong HM(f, U) \cong H^{\text{sing}}(M).$$

Again they used special choices of f and H , which do not guarantee any functoriality of type (3).

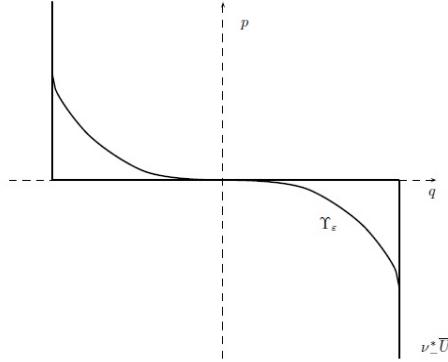


Figure 2. Singular Lagrangian $\nu_-^*\overline{U}$ and approximation Υ

6. PSS for open subset

In [18] we used PSS isomorphisms for an open subset to construct certain numerical invariants, called spectral invariants, for the case of open subset. We now give the sketch of the construction of PSS for this case (see [18] for technical details).

In order to avoid some analytical troubles, we choose a Morse function $f : M \rightarrow \mathbb{R}$ such that

- $\text{Crit}(f) \cap \partial U = \emptyset$,
 - ∇f points outward ∂U .
- (4)

The first step is the construction of PSS for approximations

$$\text{PSS}^\Upsilon : HM(f, U) \rightarrow HF(\phi_H^1(O_U), \Upsilon^s).$$

Next, we prove that PSS^Υ commutes with F_{ab} that defines direct limit. This enables us to define

$$\text{PSS} : HM(f, U) \rightarrow HF(U), \quad \text{PSS}([p]) := [\text{PSS}(p)].$$

To prove the functoriality of PSS one needs to prove that all PSS^Υ are functorial (see [18]).

7. Poincaré duality

There are some analytical difficulties arising in the attempt of defining the inverse of PSS. Namely, it is usually (in [1, 17, 20]) defined via a number of (u, γ) (recall that PSS is defined via a number of (γ, u)). In our situation, the choice (4) of f that enabled us to construct of PSS, disables us to construct the other way round isomorphism. Therefore we need to use the Poincaré duality isomorphisms.

In Morse case, if we take $-f$ instead of f , we obtain Morse Poincaré dual:

$$\text{PD}_M : HM(f, U) \xrightarrow{\cong} HM(-f, U)$$

which, in singular homology, corresponds to

$$H^{\text{sing}}(U) \cong HM(f, U) \cong HM(-f, U) \cong H^{\text{sing}}(U, \partial U).$$

In Floer case, instead of $\nu_-^*(\partial U)$ we need to consider

$$\nu_+^*(\partial U) := \{\alpha \in \nu^*(\partial U) \mid \alpha(\mathbf{n}) \geq 0\}.$$

Let

$$\zeta : T^*M \rightarrow T^*M, \quad \zeta : x = (q, p) \mapsto \bar{x} := (q, -p)$$

be the symplectic anti-involution map. Note that

$$\zeta(\nu_-^*\partial U) = \nu_+^*\partial U$$

and that, if Υ is an Lagrangian approximation of $\nu_-^*\overline{U}$, then $\overline{\Upsilon} := \zeta(\Upsilon)$ is the Lagrangian approximation of $\nu_+^*\overline{U}$. Setting

$$\overline{H}(x, t) := H(\zeta(x), t)$$

gives us the identification

$$\phi_H^1(O_U) \cap \Upsilon \cong \phi_{\overline{H}}^1(O_U) \cap \overline{\Upsilon}.$$

Similarly:

$$\overline{J} := \zeta^* J$$

leads to the bijection of the spaces of holomorphic discs defining the boundary operation. Therefore, we have

$$\zeta_* : HF(\phi_H^1(O_U), \Upsilon : J) \xrightarrow{\cong} HF(\phi_{\overline{H}}^1(O_U), \overline{\Upsilon} : \overline{J})$$

and we define

$$\text{PD}_F := \zeta_*.$$

8. PSS is isomorphism

We define Floer homology for open set via positive approximations of $\nu_+^* \overline{U}$ in the way analogous to the construction for negative ones. We denote this homology by $HF^+(U)$.

Now, for f satisfying (4), we can define the inverse isomorphism for the approximations

$$HM(-f, U) \xrightarrow{\cong} HF(\phi_H^1(O_U), \overline{\Upsilon}),$$

using a number of (u, γ) . We denote this isomorphism by Ψ^Υ . The diagram:

$$\begin{array}{ccc} HM(f, U) & \xrightarrow{\text{PSS}^\Upsilon} & HF(\phi_H^1(O_U), \Upsilon) \\ \text{PD}_M \downarrow \cong & & \text{PD}_F \downarrow \cong \\ HM(-f, U) & \xleftarrow{\Psi} & HF(\phi_H^1(O_U), \overline{\Upsilon}) \end{array}$$

commutes. Therefore, we conclude that PSS^Υ is an isomorphism.

We next prove that PD_M , PD_F and Ψ^Υ all commute with homomorphisms defining a direct limits, and obtain the commutativity of

$$\begin{array}{ccc} HM(f, U) & \xrightarrow{\text{PSS}} & HF(U) \\ \text{PD}_M \downarrow \cong & & \text{PD}_F \downarrow \cong \\ HM(-f, U) & \xleftarrow{\Psi} & HF^+(U). \end{array}$$

This proves that PSS is indeed an isomorphism.

Acknowledgements. This work is partially supported by Ministry of Education and Science of Republic of Serbia Project #ON174034.

References

- [1] P. Albers, *A Lagrangian Piunikhin-Salamon-Schwarz morphism and two comparison homomorphisms in Floer homology*, Int. Math. Res. Not. IMRN , no. 4, 56pp, 2008.
- [2] A. Banyaga, D. Hurtubise, *Lectures on Morse Homology*, Kluwer Academic Publishers, 2004.
- [3] J. Đuretić, *Piunikhin-Salamon-Schwarz isomorphisms and spectral invariants for conormal bundle*, preprint, arXiv:1411.0852, 2014.
- [4] J. Đuretić, J. Katić, D. Milinković, *Comparison of spectral invariants in Lagrangian and Hamiltonian Floer theory*, Filomat, **30**, no. 5, 1161–1174, 2016.
- [5] A. Floer, *A relative Morse index for the symplectic action*, Comm. Pure Appl. Math. 41 (1988), 393–407.
- [6] A. Floer, *Morse theory for Lagrangian intersections*, J. Differential Geometry 28 (1988), 513–547.
- [7] A. Floer, *The unregularized gradient flow of the symplectic action*, Comm. Pure Appl. Math. 41 (1988), 775–813.
- [8] A. Floer, *Witten’s complex and infinite dimensional Morse theory*, J. Differential Geometry 30 (1989), 207–221.
- [9] A. Floer, *Symplectic fixed points and holomorphic spheres*, Comm. Math. Phys. 120 (1989), 575–611.
- [10] A. Floer, H. Hofer, *Coherent orientations for periodic orbit problems in symplectic geometry*, Mathematische Zeitschrift 212 (1993), 13–38.
- [11] A. Floer, H. Hofer, *Symplectic homology I: Open sets in C^n* , Mathematische Zeitschrift 215 (1994), 37–88.
- [12] A. Floer, H. Hofer and K. Wysocki, *Applications of symplectic homology I*, Mathematische Zeitschrift 217 (1994), 577–606.
- [13] K. Fukaya, Y.-G. Oh, H. Ohta, K. Ono, *Lagrangian intersection Floer theory*, Kyoto University preprint.
- [14] J. Katić, D. Milinković, J. Nikolić *PSS homomorphisms and Spectral Invariants in Lagrangian Floer homology of open subset*, preprint
- [15] R. Kasturirangan, Y.-G. Oh, *Quantization of Eilenberg–Steenrod axioms via Fary functors*, RIMS preprint, 2000.
- [16] R. Kasturirangan, Y.-G. Oh, *Floer homology for open subsets and a relative version of Arnold’s conjecture*, Math. Z. 236, 151–189, 2001.
- [17] J. Katić, D. Milinković, *Piunikhin-Salamon-Schwarz isomorphisms for Lagrangian intersections*, Differential Geom. Appl. **22**, no. 2, 215–227, 2005.

- [18] J. Katić, D. Milinković, J. Nikolić *PSS homomorphisms and Spectral Invariants in Lagrangian Floer homology of open subset*, preprint.
- [19] Y.-G. Oh, *Symplectic topology as the geometry of action functional I – relative Floer theory on the cotangent bundle*, J. Differential Geom. 45, 499–577, 1997.
- [20] S. Piunikhin, D. Salamon, M. Schwarz, *Symplectic Floer–Donaldson theory and quantum cohomology*, in: Contact and symplectic geometry, Publ. Newton Instit. 8, Cambridge Univ. Press, Cambridge, pp. 171–200, 1996.
- [21] M. Poźniak, *Floer homology, Novikov rings and clean intersections*, Ph.D. thesis, University of Warwick, 1994.
- [22] M. Schwarz, *Morse Homology*, Birkhäuser, 1993.

Problem raspoređivanja vozila pri optimizaciji troškova transporta poljoprivrednih sirovina

Ana Anokić

*Faculty of Agriculture, University of Belgrade, Nemanjina 6, 11 080 Zemun, Belgrade, Serbia
e-mail: anokicana@agrif.bg.ac.rs*

Zorica Stanimirović

*Faculty of Mathematics, University of Belgrade, Studentski trg 16/IV, 11 000 Belgrade, Serbia
e-mail: zoricast@matf.bg.ac.rs*

Dorđe Stakić

*Faculty of Mathematics, University of Belgrade, Studentski trg 16/IV, 11 000 Belgrade, Serbia
e-mail: djordjes@matf.bg.ac.rs*

Tatjana Davidović

*Mathematical Institute of the Serbian Academy of Science and Arts, Kneza Mihaila 36, 11000 Belgrade, Serbia
e-mail: tanjad@mi.sanu.ac.rs*

Apstrakt. U radu je razmatrana varijanta problema raspoređivanja vozila (engl. *Vehicle Scheduling Problem - VSP*) koja se odnosi na optimizaciju transporta u jednoj fabričkoj šećeri u Srbiji. Predstavljen je precizan matematički model problema VSP koji je oblika mešovitog celobrojnog programiranja sa kvadratnim ograničenjima (engl. *Mixed integer quadratically constrained program - MIQCP*). MIQCP model je korišćen u okviru egzaktnog rešavača Lingo 17 u cilju dobijanja rešenja na skupu realnih i generisanih test primera. Kako su optimalna ili dopustiva rešenja dobijena samo za test primere malih dimenzija, razvijena je Opšta metoda promenljivih okolina (engl. *General Variable Neighborhood Search - GVNS*), kao metaheuristički pristup rešavanju razmatranog problema VSP. Izložene su karakteristike predložene GVNS metode i opisani njeni elementi, prilagođeni karakteristikama posmatrane varijante problema VSP. Analiza eksperimentalnih rezultata pokazuje da predložena GVNS metoda brzo dostiže sva poznata optimalna rešenja na test primerima malih dimenzija i daje rešenja na test primerima srednjih i većih dimenzija za koje optimalno rešenje nije poznato.

Ključne reči: problem raspoređivanje vozila; optimizacija transportnih troškova; MIQCP; metaheuristike, opšta metoda promenljivih okolina.

1. Uvod

Potreba za optimizacijom transporta u brojnim transportnim sistemima prirodno vodi ka velikom interesovanju istraživača za različite varijante problema raspoređivanja vozila (engl. *Vehicle Scheduling Problem - VSP*) sa teorijskog i praktičnog aspekta. Problemi VSP se u opštem slučaju svode na određivanje niza tura koje vozilo realizuje tokom radnog dana sa ciljem smanjenja troškova transporta. Dantzig i Fulkerson su u radu [7] 1954. godine predložili model linearнog programiranja za problem određivanja minimalnog broja tankera u cilju realizacije unapred poznatog plana transporta goriva. Njihov rad podstakao je rad drugih istraživača širom sveta i samim tim i nastanak brojnih varijanti problema raspoređivanja vozila i razvoj metoda za njihovo rešavanje.

Značajan broj varijanti problema VSP odnosi se na optimizaciju prevoza ljudi u urbanim sredinama. Baita i sar. su u studiji [4] za rešavanje problema optimizacije javnog prevoza predložili tri pristupa: *tradicionalni*, zasnovan na skaliranju različitih kriterijuma, *leksikografski*, koji koristi principe logičkog programiranja i *metod Pareto optimizacije* implementiran pomoću genetskih algoritama. Haghani i sar. su u radu [9] analizirali tri modela za raspoređivanja vozila pri autobuskom prevozu putnika. Wang i Shen su u studiji [21] razmatrali problem raspoređivanja električnih autobusa i za njegovo rešavanje primenili algoritam zasnovan na metodi optimizacije kolonijom mrava. Pregled problema raspoređivanja vozila može se pronaći u [2].

Problemi raspoređivanja vozila se često posmatraju zajedno sa problemima rutiranja vozila (engl. *Vehicle Routing Problem - VRP*). Preciznije, za dati skup vozila neophodno je definisati raspored po kojem vozila prave rute, a istovremeno se vrši optimalan izbor lokacija u okviru svake rute. Takvi problemi nazivaju se integrisani

problemi rutiranja i raspoređivanja vozila. Moons i sar. su u radu [17] ukazali na značaj ovakvog pristupa, nagnavši da odvojeno posmatranje aspekata raspoređivanja i rutiranja vodi ka suboptimalnim rešenjima problema. Agustina i sar. su u studiji [1] razmatrali integrisan problem raspoređivanja i rutiranja vozila pri transportu hrane, u cilju minimizacije ukupnih troškova, koji uključuju transportne troškove i troškove penala u slučaju ranije ili kasnije dostave od planirane. Metodu simuliranog kaljenja primenili su Xiao i sar. u [22] za rešavanje zelenog problema rutiranja i raspoređivanja vozila. Matematička formulacija problema predloženog u [22] uključuje tri funkcije cilja koje se, redom, odnose na: minimizaciju ukupne CO_2 emisije, ukupno pređeno rastojanje i ukupno vreme putovanja. Androutsopoulos i Zografos su u radu [3] integralski problem rutiranja i raspoređivanja vozila pri rešavanju realnog problema određivanja ruta opsluživanja i rasporeda putovanja tako da ukupni troškovi transporta i potrošnje goriva budu minimalni, pri čemu su uzeti u obzir i promenljivi uslovi u saobraćaju tokom vremena. Aspekti rutiranja i raspoređivanja integrirani su u algoritmu zasnovanom na optimizaciji mrvljinim kolonijama predloženom u [3] za rešavanje posmatranog problema.

Organizacija transporta poljoprivrednih dobara ima svoje specifičnosti koje potiču uglavnom od niske cene sirovina na tržištu. S obzirom da troškovi transporta čine glavne troškove proizvodnje, ušteda koja se može ostvariti u ovoj fazi je od velikog značaja. Kako bi motivisali individualne proizvođače da nastave saradnju, kompanije koje se bave otkupom i preradom poljoprivrednih sirovina, organizuju i finansiraju transport od svakog proizvođača do fabrike. Preciznije, kompanije obično iznajmljuju vozila i angažuju radnike koji vrše poslove transporta, utovara i istovara tokom proizvodne sezone. Iz tog razloga, efikasna organizacija transporta uz ostvarenu uštedu utrošenog vremena i novca je od velikog značaja u preradi poljoprivrednih proizvoda. U literaturi postoji nekoliko radova koji se odnose na razne varijante problema transporta poljoprivrednih sirovina. Higgins je u [11] razmatrao problem raspoređivanja vozila pri transportu šećerne trske u australijskoj fabričkoj šećera. Autor je predložio model mešovitog celobrojnog linearnog programiranja (engl. *Mixed Integer Linear Program - MILP*) koji opisuje razmatrani problem i za rešavanje realnih test primera razvio dve metaheurističke metode, tabu pretraživanje (engl. *Tabu Search - TS*) i metodu promenljivih okolina (engl. *Variable Neighborhood Search - VNS*). Eksperimentalni rezultati u [11] pokazali su da je su obe dizajnirane metaheuristike uspele da pronađu rešenja sa prosečnim smanjenjem redova čekanja za oko 90% u poređenju sa dodatašnjim metodama primenjivanim u praksi. Milan i sar. su u studiji [14] razmatrali problem transporta šećerne trske na Kubi, sa ciljem smanjenja transportnih troškova integrisanjem drumskog i železničkog saobraćaja. Za razmatrani problem predložen je MILP model, koji je potom testiran korišćenjem komercijalnog rešavača HyperLINDO u cilju pronalaženja optimalnih rešenja na realnim test primerima. Kako HyperLINDO nije uspeo da reši ceo model za 200h izvršavanja, autori su najpre rešili potproblem polaznog problema, a zatim iskoristili dobijeno rešenje za konstrukciju dopustivog rešenja polaznog modela. Thuankaewski i sar. su u radu [18] razmatrali problem transprtata šećerne trske na Tajvanu koji za cilj ima maksimizaciju očekivanih prinosa pod istim uslovima za sve proizvođače.

Predmet ovog rada je problem raspoređivanja vozila pri transportu šećerne repe u jednoj kompaniji u Srbiji. Varijanta razmatranog problema VSP se razlikuje od problema iz literature [11, 14, 18] zbog razlika između šećerne trske i šećerne repe u njihovoj održivosti na otvorenom, tipu korišćenih vozila pri transportu i fabričkih resursa. Razmatrani problem je formulisan u vidu mešovitog celobrojnog programa sa kvadratnim ograničenjima (engl. *Mixed Integer Quadratically Constrained Program - MIQCP*). Razvijeni MIQCP model je korišćen u okviru egzaktnog rešavača Lingo 17 u cilju dobijanja rešenja na skupu realnih i generisanih test primera. Dobijena su optimalna rešenja samo na test primerima manjih dimenzija. Iz tog razloga heuristički pristup predstavlja logičan izbor za rešavanje instanci većih dimenzija.

Metoda promenljivih okolina (VNS) je metaheuristička metoda koja je do sada uspešno primenjena na brojne probleme optimizacije, uključujući i probleme rutiranja i raspoređivanja vozila [10]. Macedo i sar. su u [13] primenili adaptivni metod promenljivih okolina (engl. *Skewed Variable Neighborhood Search - SVNS*) na lokacijski problem rutiranja i raspoređivanja vozila (engl. *Location Routing Scheduling Problem*). De Armas i Melian-Batista su u studiji [8] predložili VNS za rešavanje dinamičkog problema rutiranja vozila sa vremenskim okvirima (engl. *Dynamic Vehicle Routing Problem with Time Windows*). U radu Cheikh i sar. [6], razvijena je VNS metoda za rešavanje problema rutiranja vozila sa višestrukim putovanjima (engl. *Vehicle Routing Problem with Multiple Trips*).

Brojni primeri uspešne primene VNS metode na probleme rutiranja i raspoređivanja vozila bile su motivacija za dizajniranje i implementaciju opšte metode promenljivih okolina (engl. *General Variable Neighborhood Search - GVNS*) za rešavanje razmatranog VSP. Predložena GVNS metoda koristi adekvatne strukture okolina u fazi razmrdavanja, a umesto klasične lokalne pretrage koristi metodu promenljivog spusta (engl. *Variable Neighborhood Descent - VND*). Eksperimentalni rezultati na realniminstancama manjih dimenzija pokazuju da predložena

GVNS metoda brzo dostiže sva poznata optimalna rešenja dobijena egzaktnim rešavačem Lingo 17, dok je na instancama većih dimenzija, GVNS metoda dostigla ili popravila gornje granice optimalnih rešenja dobijenih Lingo rešavačem za veoma kratko vreme izvršavanja. GVNS se pokazao efikasnim i pri rešavanju generisanih instanci većih dimenzija, koje prate strukturu realnih instanci.

Rad je koncipiran na sledeći način. U sekciji 2 detaljno je izložen razmatrani problem raspoređivanja vozila pri transportu šećerne repe, a odgovarajući MIQCP model prikazan je u sekciji 3. U sekciji 4 opisana je dizajnirana GVNS metoda za rešavanje razmatranog problema. Eksperimentalni rezultati su izloženi u sekciji 5 i data je njihova analiza. Konačno, u sekciji 6 izvedeni su zaključci i ukazano je na moguće pravce budućih istraživanja.

2. Opis problema

Sezona sakupljanja šećerne repe u Srbiji, tzv. kampanja počinje u septembru i traje uglavnom do decembra ili januara sledeće godine. Okvirni plan transporta šećerne repe definiše se u kasno proleće, kada proizvođači mogu da procene količine šećerne repe koje će spremiti za otkup. Međutim, precizan plan transporta se formira na dnevnom nivou tokom kampanje, tako što se za svaki datum u toku posmatranog perioda unapred priprema lista lokacija koja će biti posećena toga dana. Pritom, veoma važan uslov je da dnevne potrebe fabrike, u smislu količine pristigle robe, moraju biti zadovoljene. Kada se pokrenu fabrička postrojenja, proces proizvodnje ne bi smeо da se zaustavi, s obzirom da ponovo pokretanje mašina stvara značajne dodatne troškove. Takođe, najavljeni vremenski uslovi kao što su velike kiše, mogu primorati proizvođače da pripreme robu nekoliko dana pre dogovorenog datuma. Šećerna repa, izvađena iz zemlje, ne bi trebalo da stoji previše dugo na otvorenom kako ne bi izgubila kvalitet. U slučaju da se nekoj lokaciji repa nalazi duže od dozvoljenog broja dana, posmatrana lokacija se smatra hitnom i mora se isprazniti tokom radnog dana.

Prepostavka je da su vozila koja se koriste za transport homogena, odnosno imaju isti kapacitet i prosečnu brzinu. Sva vozila su locirana u krugu fabrike, koja je početna i krajnja tačka svake ture. Dnevni raspored vozila definisan je nizom tura i odgovarajućim vremenima polazaka iz kruga fabrike. U svakoj turi, vozilo posećuje samo jednu lokaciju i vraća se u fabriku. Količine pripremljene šećerne repe na svakoj lokaciji značajno su veće od kapaciteta vozila, pa je stoga potrebno svaku lokaciju posetiti nekoliko puta kako bi se ispraznila. Kada se vozilo vrati u krug fabrike, uzimaju se uzorci i vrši se procena kvaliteta transportovane šećerne repe. Vozila ne prave rute kako bi se izbeglo mešanje repe sa različitim lokacijama. Čak i u slučaju kada vozilo nije potpuno natovareno (što se može desiti pri poslednjoj poseti lokaciji) ono se vraća u fabriku, opslužujući samo jednu lokaciju. Posle završenog istovara, vozilo može započeti novu turu. Zbog jednostavnosti, prepostavlja se da postoji dovoljno radne snage i prostora na svakoj lokaciji i u krugu fabrike za istovremeni utovar i istovar neograničenog broja vozila.

Raspored vozila je definisan nizom tura i odgovarajućim vremenima polazaka iz kruga fabrike. Cilj problema je pronaći optimalan raspored vozila koji zadovoljava postavljena ograničenja i pritom minimizira maksimalno radno vreme na skupu posmatranih vozila, odnosno trenutak kada sva vozila završe svoje ture za posmatrani radni dan. U nastavku je dat pregled karakteristika razmatranog problema raspoređivanja vozila pri transportu šećerne repe:

- Vozila su homogena i koriste se više puta tokom radnog dana;
- Krug fabrike predstavlja početnu i krajnju tačku svake ture;
- Svako vozilo u svakoj turi posećuje samo jednu lokaciju;
- Dnevne potrebe fabrike, u smislu količine prevezene robe, moraju biti zadovoljene;
- Postoji gornja granica u broju dana koliko roba može da stoji na otvorenom;
- Svaka lokacija na kojoj roba stoji na otvorenom duže nego što je dozvoljeno mora biti opslužena;
- Cilj je minimizovati vremenski trenutak kada sva vozila završe svoje poslednje ture.

3. MICP model razmatranog problema

Opisani problem je formulisan u obliku mešovitog celobrojnog programa sa kvadratnim ograničenjima (engl. *Mixed integer quadratically constrained problem - MIQCP*). U narednim podsekcijama opisani su datalji MIQCP formulacije razmatranog problema.

3.1. Notacija

Za predstavljanje MIQCP modela uvedena je sledeća notacija:

- J : Skup lokacija;
- I : Skup vozila;
- K : Skup tura;
- n : Ukupan broj lokacija;
- m : Ukupan broj vozila;
- k_{max} : Maksimalan broj tura koje vozilo može da napravi u toku dana;
- c_j : Količina sakupljene sirovine na lokaciji $j \in J$;
- d_j : Rastojanje od farike do lokacije $j \in J$;
- CV : Kapacitet vozila;
- C : Dnevne potrebe fabrike;
- v : Prosečna brzina vozila;
- u : Prosečno vreme zadržavanja u fabriki po završetku ture, potrebno za istovar i analiziranje uzoraka;
- w : Prosečno vreme trajanja utovara;
- t_j : Broj dana koliko roba стоји na otvorenom na lokaciji $j \in J$;
- t_0 : Maksimalan broj dana koliko roba može da стоји na otvorenom bez gubitka kvaliteta;
- T_j : Binarna veličina dodeljena lokaciji $j \in J$, definisana sa:
 $T_j = 1$ ako je $t_j > t_0$ i $T_j = 0$ u suprotnom.
- ε : Mala pozitivna konstanta, manja od CV i c_j/CV , $j \in J$.
- t_{start} : Početak radnog vremena;
- t_{end} : Kraj radnog vremena;

3.2. Koncept virtualnih tura

Različita vozila ne moraju realizovati isti broj tura tokom radnog dana. Broj tura jednog vozila direktno zavisi od udaljenosti lokacija koje to vozilo poseće od fabrike. Udaljenost direktno određuje vreme trajanja putovanja u oba smera, a samim tim i broj realizovanih tura tokom ograničenog radnog vremena. U cilju jednostavnije formulacije matematičkog modela, uveden je koncept *virtualne ture*, koji ne narušava ograničenja problema. Preciznije, izjednačen je broj tura svih vozila tako što je dopušteno da vozila u svom nizu tura tokom radnog dana imaju jednu ili više virtualnih tura. Pretpostavka je da, u toku virtualne ture, vozilo ostaje u krugu fabrike, tj. ne poseće ni jednu lokaciju, pa je trajanje virtualne ture jednak nuli. Imajući u vidu da je cilj problema minimizacija maksimalnog od vremena završetaka radnog dana svih vozila, virtualne ture ne utiču na vrednost funkcije cilja, te se zanemaruju pri računanju vremena završetka radnog dana vozila. Zbog jednostavnosti, virtualne ture se nalaze na kraju niza tura koji odgovara vozilu, odmah iza poslednje ture u kojoj vozilo opslužuje neku od lokacija. Dodavanje virtualnih tura, ukoliko je potrebno, omogućava dopunu broja tura svakog vozila do maksimalnih k_{max} bez uticaja na ograničenja problema i na vrednost funkcije cilja.

3.3. Promenljive

MIQCP formulacija koristi sledeće promenljive:

- Binarna promenljiva x_{ik}^j ima vrednost 1 ako vozilo $i \in I$ poseće lokaciju $j \in J$ u turi $k \in K$, a 0 u suprotnom. Ako je tura $k \in K$ vozila $i \in I$ virtualna, tada je $\sum_{j \in J} x_{ik}^j = 0$;
- Realna promenljiva t_{ik} definiše vreme polaska vozila $i \in I$ u turi $k \in K$ iz kruga fabrike;
- Binarna promenljiva y_j , $j \in J$ je uvedena u cilju praćenja ukupne količine prevezene repe do fabrike sa lokacije $j \in J$. Vrednost promenljive y_j jednaka je 0 ako je lokacija $j \in J$ ispraznjena, a 1 u suprotnom. Ako je $y_j = 1$, količina šećerne repe koja je transportovana sa lokacije $j \in J$ u fabriku jednaka je c_j . U suprotnom, količina robe dovezene u krug fabrike sa lokacije $j \in J$ računa se kao proizvod kapaciteta vozila CV i ukupnog broja poseta lokaciji $j \in J$ od strane svih vozila;
- Realna promenljiva T se koristi za formulisanje funkcije cilja. Označava vremenski trenutak kada sva vozila završe svoju poslednju turu.

3.4. MIQCP formulacija

MIQCP formulacija razmatranog problema raspoređivanja vozila je:

$$\min \quad T \quad (1)$$

pri uslovima

$$\sum_{j \in J} x_{ik}^j \leq 1 \quad \forall i \in I, \quad \forall k \in K, \quad (2)$$

$$CV \sum_{(i,k) \in I \times K} x_{ik}^j - CV + \varepsilon \leq c_j \quad \forall j \in J, \quad (3)$$

$$T_j + y_j \leq 1 \quad \forall j \in J, \quad (4)$$

$$\sum_{j \in J} (1 - y_j) \cdot c_j + CV \cdot \sum_{(i,j,k) \in I \times J \times K} y_j x_{ik}^j \geq C, \quad (5)$$

$$\sum_{(i,k) \in I \times K} x_{ik}^j \geq (c_j/CV) \cdot (1 - y_j) \quad \forall j \in J, \quad (6)$$

$$\sum_{(i,k) \in I \times K} y_j x_{ik}^j + \varepsilon \leq c_j/CV \quad \forall j \in J, \quad (7)$$

$$t_{i,k+1} \geq t_{ik} + \frac{2}{v} \sum_{j \in J} d_j x_{ik}^j + (u+w) \sum_{j \in J} x_{ik}^j \quad \forall i \in I, \quad \forall k \in K \setminus \{k_{max}\}, \quad (8)$$

$$t_{i,k_{max}} + \frac{2}{v} \sum_{j \in J} d_j x_{ik_{max}}^j + (u+w) \sum_{j \in J} x_{ik_{max}}^j \leq T \quad \forall i \in I, \quad (9)$$

$$x_{ik}^j \in \{0, 1\} \quad \forall i \in I, \quad \forall j \in J, \quad \forall k \in K, \quad (10)$$

$$y_j \in \{0, 1\} \quad \forall j \in J, \quad (11)$$

$$t_{ik} \in [t_{start}, t_{end}] \quad \forall i \in I, \quad \forall k \in K. \quad (12)$$

Funkcija cilja (1) zajedno sa ograničenjem (9) minimizuje vremenski trenutak u kojem se završava dnevni transport za sva vozila, odnosno trenutak u kojem poslednje pristiglo vozilo završi istovar robe u svojoj poslednjoj turi.

Ograničenje (2) obezbeđuje da svako vozilo u svakoj turi može da opsluži najviše jednu lokaciju. Dakle, svako vozilo u svakoj turi ima mogućnost da poseti neku lokaciju ili napravi virtualnu turu. Ukupna količina robe koja se preveze sa lokacije $j \in J$ u fabriku ne može biti veća od pripremljene količine c_j na datoj lokaciji, što je obezbeđeno ograničenjem (3). Umanjilac CV koji se nalazi na levoj strani ograničenja (3) odražava činjenicu da poslednje vozilo koje posećuje neku lokaciju ne mora biti u potpunosti natovareno. Kako bi količina koja se transportuje sa lokacije $j \in J$ umanjena za veličinu kapacita vozila CV bila strogo manja od prikupljene količine c_j na lokaciji $j \in J$, na levoj strani ograničenja (3) je dodata mala pozitivna konstanta ε . To je neophodno u slučaju kada je c_j jednak proizvodu prirodnog broja i kapaciteta vozila CV tj. celobrojnom umnošku veličine CV . Bez stroge nejednakosti, bilo bi moguće da vozila nepotrebno naparave turu više do lokacije $j \in J$. Na primer, ako je $c_j = 120t$ i $CV = 30t$, lokaciju j je potrebno posetiti najviše 4 puta, međutim, bez konstante ε u ograničenju (3), postoji mogućnost da se lokacija j poseti 5 puta, kako je $5 \cdot 30 - 30 \leq 120$.

Ograničenje (4) obezbeđuje uslov da se sve količine sa lokacije $j \in J$ moraju prevesti tokom radnog dana, ukoliko na ovoj lokaciji roba stoji na otvorenom duže od t_0 dana. Imajući u vidu definiciju binarnih veličina T_j (videti podsekciju 3.1), u slučaju kada je $t_j \leq t_0$, važi $T_j = 0$, a desna strana ograničenja (4) je jednaka 1. Stoga je, u ovom slučaju, ograničenje (4) zadovoljeno nezavisno od transportovane količine sa lokacije j . Sa druge

strane, ako je $t_j > t_0$, tada je $T_j = 1$ i lokacija j mora biti ispražnjena tokom radnog dana, što znači da y_j mora imati vrednost 0.

Ograničenje (5) obezbeđuje zadovoljenje dnevnih potreba fabrike. Prva suma na levoj strani relacije (5) odnosi se na količine koje su dovezene sa lokacija koje su ispražnjene, a druga predstavlja količinu robe, dovezene sa neispražnjenih lokacija.

Vrednosti promenljivih y_j su definisane ograničenjima (6) i (7). Ako je lokacija $j \in J$ ispražnjena, y_j uzima vrednost 0 i relacija (6) obezbeđuje da je lokacija j posećena ne manje od c_j/CV puta. U slučaju kada lokacija $j \in J$ nije ispražnjena, tada je $y_j = 1$ i ograničenje (6) je zadovoljeno. Ograničenje (7) obezbeđuje da je svaka neispražnjena lokacija $j \in J$ posećena strogo manje od c_j/CV puta. U slučaju da se radi o neispražnjenoj lokaciji, ograničenje (7) nije aktivno.

Ograničenje (8) odražava činjenicu da svako vozilo $i \in I$ ne može početi novu turu dok se prethodna ne završi. Vreme polaska svake ture ne sme biti manje od vremena polaska prethodne ture, uvećanom za vreme trajanja vožnje u oba smera i zbirom dužine trajanja utovara i istovara. Ako je tura $k \in K \setminus \{k_{max}\}$ vozila $i \in I$ virtualna, obe sume na desnoj strani ograničenja (8) su jednake 0. Na kraju, ograničenja (10)–(12) definišu tip promenljivih koje su korišćene u predloženom modelu.

4. Opšta metoda promenljivih okolina za predloženi VSP

Metoda promenljivih okolina (VNS) je poznata metaheuristika koju su uveli Mladenović i Hansen 1997. godine [15]. Osnovna ideja VNS metode je pretraga prostora rešenja problema korišćenjem različitih struktura okolina. U opštem slučaju, algoritam VNS metode se sastoji iz tri osnovna koraka: faze razmrdavanja (engl. *Shaking Phase*), lokalne pretrage (engl. *Local Search*) i pomeraja (engl. *Move or Not*). U fazi razmrdavanja, bira se rešenje iz okoline tekućeg rešenja, u cilju diversifikacije rešenja, sto pomaže algoritmu da ne završi u lokalnom optimumu. Zadatak lokalne pretrage je pokušaj poboljšanja rešenja iz faze razmrdavanja korišćenjem jedne ili više strukture okolina. Ukoliko lokalna pretraga dostigne rešenje bolje od tekućeg, vrši se zamena tekućeg rešenja novodobijenim, što predstavlja fazu pomeraja. Ova tri koraka se smenjuju do zadovoljenja kriterijuma zaustavljanja. U literaturi postoje brojne varijante VNS metode koje su uspešno primenjene na različite probleme optimizacije: Osnovna VNS metoda (engl. *Basic VNS - BVNS*), Redukovana VNS metoda (engl. *Reduced VNS - RVNS*), Metoda promenljivog spusta (engl. *Variable Neighborhood Descent - VND*), Opšta VNS metoda (engl. *General VNS - GVNS*), Adaptivna VNS metoda (engl. *Skewed VNS - SVNS*), Primalno-dualna VNS metoda (engl. *Primal-dual VNS - PD-VNS*) i druge. Pregled varijanti VNS metode i njihovih primena može se pronaći u [10].

Za rešavanje problema razmatranog u ovom radu predložena je Opšta metoda promenljivih okolina (GVNS). GVNS metoda koristi VND umesto klasične lokalne pretrage, što znači da GVNS pokušava da poboljša rešenje pretražujući okoline na deterministički način [5, 16, 19, 20]. Glavni koraci predložene GVNS za VSP predstavljeni su u Algoritmu 1. Nakon generisanja inicijalnog rešenja, smenjuju se faza razmrdavanja, VND i faza pomeraja sve dok se ne zadovolji kriterijum zaustavljanja. Rešenje dobijeno u fazi razmrdavanja se poboljšava VND pretragom. Ukoliko je poboljšanje bolje od tekućeg rešenja, vrši se zamena tekućeg rešenja novim i pretraga počinje od prve okoline. U suprotnom pretraga nastavlja od naredne okoline u nizu korišćenih struktura okolina. Ovi koraci se smenjuju sve dok se ne zadovolji kriterijum zaustavljanja, koji u predloženoj GVNS implementaciji predstavlja dostignutu gornju granicu vremena izvršavanja (t_{max}). U nastavku ovog odeljka opisani su elementi dizajnirane GVNS metode za opisani VSP.

4.1. Reprezentacija rešenja. Računanje vrednosti funkcije cilja

Rešenje S predloženog problema VSP predstavljeno je u matricom celih nenegativnih brojeva, koja se sastoji iz m vrsta dužine k_{max} . Vrsta i matrice S odgovara vozilu $i \in \{1, \dots, m\}$ i sadrži cele nenegativne brojeve $s_{ik} \in \{0, 1, \dots, n\}$, gde s_{ik} označava indeks lokacije koju vozilo i posećuje u turi $k \in \{1, \dots, k_{max}\}$. Preciznije, ako vozilo i poseće lokaciju $j \in \{1, \dots, n\}$ u turi k važi $s_{ik} = j$. U slučaju kada je tura k vozila i virtualna, $s_{ik} = 0$. Zbog jednostavnosti, virtualne ture svakog vozila nalaze se uvek na kraju vrste koja odgovara posmatranom vozilu u matrici S .

Svako generisano rešenje potrebno je evaluirati. Za dopustivo rešenje, vrednosti s_{ik} , $i \in I$, $k \in K$ određuju raspored tura svih vozila. Na osnovu vrednosti s_{ik} , računaju se vrednosti t_{ik} koje predstavljaju vremena polazaka vozila iz kruga fabrike za svako vozilo $i \in I$ i svaku turu $k \in K$, na sledeći način. Za svaku lokaciju izračuna se vreme opsluživanja te lokacije koje je jednakom sumi vremena trajanja putovanja od fabrike do lokacije i nazad i vremena trajanja utovara i istovara. Kako sva vozila započinju svoje prve ture u isto vreme, za svako vozilo $i \in I$, t_{i1} ima vrednost jednaku vremenu početka radnog dana (t_{start}). Vreme polaska t_{id} vozila $i \in I$ u turi

Algoritam 1 GVNS za VSP

```

1: Procedure GVNS(ProblemData, $h_{max}$ ,  $t_{max}$ )
2: Generate an initial solution  $S$ ;
3: repeat
4:    $h \leftarrow 1$ ;
5:   while  $h \leq h_{max}$  do
6:      $S' \leftarrow \text{Shaking}(S, h)$ ;
7:      $S'' \leftarrow \text{Variable Neighborhood Descent}(S')$ ;
8:     if  $S''$  is better than  $S$  then
9:        $S \leftarrow S''$ ;
10:       $h \leftarrow 1$ ;
11:    else  $h \leftarrow h + 1$ ;
12:   end while
13: until SessionTime  $\geq t_{max}$ 

```

$d \in \{2, \dots, k_{max}\}$ jednako je vremenu polaska istog vozila u njegovoј prethodnoј turi, uvećanom za vreme potrebno da se opsluži lokacija posećena u prethodnoј turi. Na taj način, obezbeđuje se da vozilo može početi novu turu tek po završetku prethodne. Izračunata vremena polazaka t_{ik} smeštaju se u matricu T_d . Vreme završetka radnog dana za svako vozilo računa se kao vreme povratka u krug fabrike u poslednjoј nevirtualnoј turi, uvećano za neophodno vreme trajanja istovara. Konačno, prema (1) i (9), vrednost funkcije cilja T predstavlja maksimum skupa vremena završetaka radnog dana svih vozila.

4.2. Generisanje inicijalnog rešenja

Inicijalno rešenje u predloženoј GVNS metodi generisano je pohlepnom procedurom, koja počinje sortiranjem lokacija na osnovu dva kriterijuma. Najpre se svaka lokacija označava kao hitna ili ne, imajući u vidu broj dana koliko roba stoji na otvorenom na toj lokaciji. Preciznije, ako se roba na lokaciji $j \in J$ nalazi na otvorenom t_j dana i $t_j > t_0$, lokacija j se smatra hitnom. Za hitnu lokaciju j , nije od značaja za koliko dana t_j prelazi gornju granicu od t_0 dana. Hitne lokacije imaju prioritet i moraju se opslužiti tokom radnog dana. U slučaju kada se količine repe prikupljene na hitnim lokacijama ne mogu prevesti raspoloživim skupom vozila u okviru unapred određenog broja tura i pre kraja radnog vremena, smatra se da problem nema dopustivo rešenje. Drugi kriterijum za izbor lokacija koje će biti opslužene je njihova udaljenost od fabrike, u smislu da lokacija $j \in J$ sa manjom udaljenošću od fabrike d_j ima prioritet. Konačno, sortirana lista lokacija ima sledeću strukturu: na vrhu liste nalaze se hitne lokacije sortitane u neopadajući niz prema udaljenosti od fabrike, a za njima dolaze lokacije koje nisu hitne, takođe sortirane u neopadajući niz prema udaljenosti od fabrike.

Nakon opisanog sortiranja lokacija, matrica S inicijalnog rešenja se formira na sledeći način. Uzima se jedna po jedna lokacija iz sortirane liste i njenim indeksom popunjava se prva kolona matrice rešenja. Kada su prve ture za sva vozila formirane, odnosno kada je popunjena prva kolona matrice rešenja, formiraju se druge ture tj. popunjava se druga kolona matrice, zatim treće itd. Pritom se elementi matrice popunjavaju indeksom jedne lokacije sve dok se posmatrana lokacija ne isprazni, nakon čega se uzima sledeća lokacija sa liste prioriteta. Poslednja nevirtualna tura svakog vozila može početi ukoliko se posećena lokacija može opslužiti u okviru radnog vremena. Ako nije moguće dodeliti lokaciju vozilu tako da ona bude opslužena u okviru radnog vremena, posmatrana tura i sve naredne ture tog vozila postaju virtualne, a proces generisanja rešenja nastavlja se od sledećeg vozila. Kada se potrebe fabrike zadovolje prebroje se ture svakog vozila. Ako je indeks poslednje nevirtualne ture k_i vozila $i \in I$ manji od k_{max} , sve preostale ture $k_i + 1, k_i + 2, \dots, k_{max}$ postaju virtualne.

Kada je matrica S generisana, računaju se vrednosti matrice T_d . Vremena polazaka svih vozila u prvoj turi jednaka su početku radnog vremena t_{start} , tj. $t_{i1} = t_{start}$, za svako vozilo $i \in I$. Vremena polazaka u preostalim turama, t_{id} , $d = 2, \dots, k_{max}$ računaju se uvećavanjem vremena polaska istog vozila u prethodnoј turi vremenom koje je neophodno da bi se opslužila posećena lokacija. Na kraju, prebrojavanjem poseta svakoj lokaciji, računaju se vrednosti promenljivih y_j . Ako broj poseta lokaciji j nije manji od c_j/CV , promenljiva y_j ima vrednost 0, u suprotnom je $y_j = 1$.

Lako se može primetiti da je ovako generisano inicijalno rešenje dopustivo. Svaki element s_{ik} matrice S uzima jednu vrednost skupa celih nenegativnih brojeva koja odgovara indeksu lokacije $j \in J$ posećene od strane vozila i u turi k . Preciznije, $s_{ik} = j$ označava da je $x_{ik}^j = 1$, sa izuzetkom $s_{ik} = 0$ u slučaju virtualne ture. Dakle,

ograničenje (2), koje obezbeđuje uslov da svako vozilo u svakoj turi može posetiti najviše jednu lokaciju ili napraviti praznu turu, je ispunjeno. Kolone matrice inicijalnog rešenja se popunjavaju indeksima lokacija sa liste prioriteta dok se sve hitne lokacije ne isprazne i dok se ne zadovolje potrebe fabrike. Iz tog razloga su ograničenja (4) i (5) takođe zadovoljena. Kako se prilikom popunjavanja matrice rešenja, uzima jedna po jedna lokacija sa liste prioriteta, imajući u vidu količine robe na svakoj od njih, odnosno indeks lokacije se unosi u matricu rešenja sve dok se ta lokacija ne isprazni, ograničenje (3) će biti ispunjeno. Takođe, prilikom popunjavanja matrice rešenja, vozilo započinje nevirtualnu turu jedino u slučaju kada ona može da se završi u okviru radnog vremena, što znači da će sva vozila završiti svoje poslednje ture do t_{end} . Vrednosti promenljivih y_j su definisane tako da zadovoljavaju ograničenja (6) i (7). Vrednosti promenljivih t_{ik} , x_{ik}^j i y_j , dobijene opisanim postupkom očigledno zadovoljavaju ograničenja (8)-(12).

4.3. Strukture okolina

Predložena GVNS metoda koristi četiri strukture okolina, označenih sa: N^I , N^{II} , N^{III} i N^{IV} . Okolina N^I rešenja S definisana je na sledeći način: N^I -sused S' od S se dobija izborom jednog vozila i jedne njegove nevirtualne ture ka lokaciji koja nije hitna, koja se zatim menja nekom drugom neispräžnenom lokacijom. Okolina N^{II} se sastoji iz svih rešenja koja se dobijaju zamenom prve virtualne ture u nizu tura jednog vozila, turom ka nekoj od neispräžnenih lokacija. Okolinu N^{III} rešenja S čine sva rešenja koja se dobijaju kada se jedna nevirtualna tura ka nehitnoj lokaciji vozila sa najdužim vremenom završetka radnog dana u rešenju S , zameni virtualnom turom, nakon čega se nova virtualna tura pomera na kraj niza tura posmatranog vozila. Konačno, okolina N^{IV} je definisana na sledeći način: N^{IV} -sused S' od S se dobija tako što vozilo sa najdužim vremenom završetka radnog dana i vozilo sa najkraćim vremenom završetka radnog dana razmenjuju po jednu svoju turu. Treba primetiti da potez kojim je definisana okolina N^{III} može narušiti dopustivost rešenja, jer se može dogoditi da posle poteza u okolini N^{III} potrebe fabrike ne budu zadovoljene. Iz tog razloga, potez u ovoj okolini se prihvata samo ako novo rešenje zadovoljava i ovaj uslov. U suprotnom potez u okolini N^{III} se ponavlja.

4.4. Faza razmrdavanja

Faza razmrdavanja predložene GVNS metode koristi dve strukture okolina: N^I i N^{II} , pri čemu se prvo istražuje okolina N^I za sve vrednosti reda $r = 1, \dots, r_{max}$, a potom okolina N^{II} samo za $r = 1$. Razmrdavanje u okolini N^I reda r se izvodi ponavljanjem poteza kojim je ta okolina definisana r puta, pri čemu se potezi biraju na slučajan način.

4.5. Faza VND

VND faza koristi strukture okolina N^{III} i N^{IV} na sledeći način. Prvo se deterministički istražuje okolina N^{III} , tako što pokušava da zameni jednu po jednu turu vozila sa najvećim vremenom završetka virtualnom turom. Nakon svakog takvog poteza, proverava se uslov zadovoljenosti potreba fabrike i pretraga u okolini N^{III} posmatranog rešenja se završava najboljim pronađenim susedom u ovoj okolini. Ako je dobijeno poboljšanje bolje od tekućeg VND rešenja, vrši se odgovarajuća zamena i ponavlja se pretraga u okolini N^{III} , u suprotnom se prelazi na pretragu u okolini N^{IV} . Okolina N^{IV} se pretražuje na sledeći način. Pronalaze se dva vozila, m_1 sa najdužim vremenom završetka i m_2 sa najkraćim vremenom završetka radnog dana, i vrši se zamena ture k_1 vozila m_1 ka najudaljenijoj lokaciji u nizu svih njegovih tura sa turom k_2 vozila m_2 ka najbližoj lokaciji od svih lokacija koje vozilo m_2 posećuje. Ukoliko se dobije poboljšanje na nivou dva posmatrana vozila, odnosno ako je maksimum vrednosti njihovih novih vremena završetaka manji od odgovarajuće vrednosti pre izvršene zamene, novo rešenje postaje tekuće VND rešenje i postupak se ponavlja još jednom u okolini N^{IV} . Primetimo da poboljšanje maksimalnog vremena završetaka posmatranog para vozila ne mora obavezno da predstavlja i poboljšanje na nivou celog rešenja. Opisano pretraživanje okoline N^{IV} može dati poboljšanje rešenja u slučaju kada postoje dva vozila sa istom vrednosti vremena završetka koja je pritom i maksimalna u skupu svih vozila. Nakon pretrage okoline N^{IV} , VND algoritam evaluira dobijeno rešenje. Ako je postignuto poboljšanje u odnosu na tekuće VND rešenje, tekuće rešenje se zamenuje novim i VND pretraga nastavlja istražujući okolinu N^{III} . U suprotnom završava se jedna VND iteracija. Opisane VND iteracije se smenjuju sve dok se dobija poboljšanje tekućeg VND rešenja. Ako su nekoj iteraciji ne ostvari poboljšanje rešenja, VND pretraga se završava, vraćajući najbolje pronađeno rešenje.

5. Eksperimentalni rezultati

Svi eksperimentalni rezultati predstavljeni u ovom odeljku dobijeni su na računaru sa Intel Core i5 - 3320M procesorom, na 2,60 GHz i sa 2 GB RAM memorije pod Windows 7 operativnim sistemom. Eksperimenti su izvršeni nad dva skupa instanci:

- Realne instance dobijene na osnovu podatka iz posmatrane fabrike šećera u Srbiji. Ovaj skup instanci uključuje 43 test primera sa do 15 lokacija, 40 vozila i maksimalno 20 tura tokom radnog dana;
- Generisane instance koje uključuju do 150 lokacija, 100 vozila i maksimalno 100 tura tokom radnog dana. Ovaj skup obuhvata 12 test primera, koji su generisani prateći strukturu realnih instanci.

Za dobijanje optimalnih rešenja problema VSP korišćen je komercijalni rešavač Lingo 17, koji je dizajniran za rešavanje velikog broja različitih tipova matematičkih modela [12]. MIQCP formulacija razmatranog VSP problema je korišćena u okviru Lingo 17 rešavača u cilju dobijanja optimalnih rešenja ili bar gornjih granica. Pritom je postavljena gornja granica vremena rada Lingo rešavača od 5h. Parametar koji se odnosi na kriterijum zaustavljanja u predloženoj GVNS metodi uzima sledeće vrednosti: $t_{max} = 1\text{s}$ za male realne instance, $t_{max} = 10\text{s}$ za realne primere srednjih dimenzija i $t_{max} = 100\text{s}$ za generisane instance. Ove vrednosti paramatara su izabrane tako da se postigne kompromis između kvaliteta dobijenih rešenja i vremena izvršavanja algoritma. Na svakom razmatranom test primeru, GVNS metoda je izvršavana po 30 puta.

Eksperimentalni rezultati na skupu realnih instanci malih dimenzija za koje je dobijeno optimalno rešenje korišćenjem Lingo rešavača predstavljeni su u Tabeli 1, koja je organizovana na sledeći način. Prva kolona $T_{n,m,k_{max}}$ sadrži naziv instance. U drugoj i trećoj koloni, označenim sa $opt.sol.$ i $t(s)$, predstavljeni su vrednost funkcije cilja koja odgovara optimalnom rešenju dobijenim korišćenjem Lingo rešavača i odgovarajuće vreme rada. Naredne tri kolone se odnose na rezultate testiranja dizajnirane GVNS metode i sadrže: vrednost funkcije cilja najboljeg pronađenog dopustivog rešenja (*best sol.*), pri čemu je korišćena oznaka *opt* u slučaju kada se GVNS rešenje poklapa sa *opt.sol.*, zatim prosečno vreme koje je bilo neophodno GVNS metodi da dostigne svoje najbolje rešenje u 30 uzastopnih izvršavanja i na kraju, prosečno procentualno odstupanje (*avg. gap (%)*) najboljeg GVNS rešenja od optimalnog u 30 uzastopnih izvršavanja.

Na osnovu rezultata predstavljenih u Tabeli 1 može se zaključiti da GVNS za vrlo kratko vreme (u proseku 0,001s) dostiže sva optimalna rešenja koje je pronašao Lingo rešavač za 6,431s u proseku. Može se zaključiti da je GVNS više od 6000 puta brži od Lingo rešavača na skupu realnih test primera malih dimenzija. Osim toga, dizajnirana GVNS metoda je izuzetno stabilna na skupu realnih instanci malih dimenzija, s obzirom da prosečno procentualno odstupanje najboljih GVNS rešenja od optimalnih ima veoma malu vrednost (0,092 %).

Tabela 2 sadrži eksperimentalne rezultate na skupu realnih instanci za koje Lingo rešavač nije uspeo da pronađe optimalno rešenje za 5h, kao i na skupu generisanih instanci većih dimenzija. Prva kolona $T_{n,m,k_{max}}$ sadrži naziv instance. Naredne dve kolone sadrže rezultate testiranja MIQCP formulacije Lingo rešavačem: vrednost funkcije cilja najboljeg pronađenog dopustivog rešenja (*UB*) i odgovarajuću donju granicu vrednosti funkcije cilja (*LB*). Ako u okviru datog vremena izvršavanja nije pronađeno optimalno ili dopustivo rešenje, oznaka / je upisana u odgovarajuće polje. Poslednje tri kolone odnose se na rezultate testiranja predložene GVNS metode i predstavljene su na isti način kao u Tabeli 1.

Levi deo Tabele 2 sadrži rezultate testiranja na skupu realnih instanci srednjih dimenzija, za koje je Lingo rešavač u najboljem slučaju dao dopustivo rešenje za 5h izvršavanja. U slučaju 9 od 12 realnih instanci srednjih dimenzija Lingo rešavac je pronašao dopustiva rešenja koja je GVNS metoda potvrdila ili poboljšala, osim pri testiranju instance $T_{8,6,5}$, a kod koje je GVNS pronašla za nijansu lošije rešenje od Lingo rešavača. Za preostale 3 instance na kojima Lingo rešavač nije uspeo da pronađe čak ni dopustivo rešenje, GVNS je dala rešenja za relativno kratko vreme. Prosečno GVNS vreme izvršavanja na skupu realnih instanci srednjih dimenzija iznosi 0,950s što je više od 18 900 puta brže u poređenju sa vremenom izvršavanja Lingo rešavača (5h). Pored toga, niska vrednost prosečnog procentualnog odstupanja dobijenih GVNS rešenja od najboljih pronađenih u 30 uzastopnih izvršavanja (0,055 %) pokazuje stabilnost predložene GVNS metode na skupu realnih instanci srednjih dimenzija.

Rezultati testiranja generisanih instanci većih dimenzija predstavljeni su u desnom delu Tabele 2. Lingo rešavač nije uspeo da pronađe dopustivo rešenje ni na jednoj generisanoj instanci, dok je GVNS metoda pronašla svoja najbolja rešenja za 24,495s u proseku, sa malim prosečnim procentualnim odstupanjem od najboljih rešenja, dobijenih u 30 uzastopnih puštanja (0,514 %).

Tabela 1. Eksperimentalni rezultati na realniminstancama manjih dimenzija rešenih do optimalnosti

Naziv instance	Lingo 17		GVNS		
	$T_{n,m,k_{max}}$	opt. sol.	t(s)	best sol.	t(s)
$T_{3,2,4}$	16,874	0,40	<i>opt</i>	0,001	0,000
$T_{3,3,2}$	9,780	0,13	<i>opt</i>	0,001	0,000
$T_{3,3,3}$	13,727	0,37	<i>opt</i>	0,001	0,000
$T_{3,3,4}$	16,303	4,33	<i>opt</i>	0,001	0,000
$T_{3,4,2}$	10,580	0,30	<i>opt</i>	0,002	0,000
$T_{3,4,3}$	13,727	0,27	<i>opt</i>	0,001	0,000
$T_{3,4,4}$	16,303	52,90	<i>opt</i>	0,001	0,000
$T_{3,5,2}$	12,294	0,30	<i>opt</i>	0,002	0,000
$T_{4,2,3}$	12,870	0,22	<i>opt</i>	0,001	0,000
$T_{4,2,4}$	13,446	0,26	<i>opt</i>	0,001	0,000
$T_{4,3,2}$	10,580	0,13	<i>opt</i>	0,001	0,000
$T_{4,3,3}$	12,299	2,94	<i>opt</i>	0,001	0,000
$T_{4,4,2}$	11,151	0,17	<i>opt</i>	0,001	0,000
$T_{4,4,3}$	14,299	6,35	<i>opt</i>	0,002	0,000
$T_{4,4,4}$	17,446	71,06	<i>opt</i>	0,001	0,000
$T_{4,5,2}$	10,866	0,39	<i>opt</i>	0,002	2,717
$T_{5,2,3}$	10,584	0,46	<i>opt</i>	0,000	0,000
$T_{5,2,4}$	13,156	1,40	<i>opt</i>	0,002	0,000
$T_{5,3,2}$	10,580	0,13	<i>opt</i>	0,002	0,000
$T_{5,3,3}$	13,156	6,04	<i>opt</i>	0,002	0,000
$T_{5,3,4}$	16,589	9,42	<i>opt</i>	0,001	0,000
$T_{5,4,3}$	13,156	5,23	<i>opt</i>	0,001	0,145
$T_{5,5,2}$	11,437	3,59	<i>opt</i>	0,001	0,000
$T_{6,2,3}$	11,727	0,72	<i>opt</i>	0,001	0,000
$T_{6,2,4}$	13,160	9,98	<i>opt</i>	0,001	0,000
$T_{6,3,2}$	10,294	0,60	<i>opt</i>	0,001	0,000
$T_{6,3,3}$	12,870	6,74	<i>opt</i>	0,001	0,000
$T_{6,4,2}$	10,009	0,32	<i>opt</i>	0,001	0,000
$T_{6,5,2}$	10,866	3,43	<i>opt</i>	0,001	0,000
$T_{7,3,3}$	12,299	4,86	<i>opt</i>	0,004	0,000
$T_{7,5,2}$	10,580	6,69	<i>opt</i>	0,004	0,000
Prosek	12,800	6,431	<i>opt</i>	0,001	0,092

Tabela 2. Eksperimentalni rezultati na realniminstancama srednjih dimenzija i generisaniminstancama za koje Lingo 17 nije pronašao optimalno rešenje

Naziv instance	Lingo 17			GVNS			Naziv instance	Lingo 17			GVNS		
	$T_{n,m,k_{max}}$	UB	LB	best sol.	t(s)	avg. gap (%)		$T_{n,m,k_{max}}$	UB	LB	best sol.	t(s)	avg. gap (%)
$T_{4,5,7}$	15,287	14,315	15,287	0,993	0,000		$T_{20,40,20}^r$	/	/	42,020	25,120	0,136	
$T_{5,5,5}$	18,879	18,210	18,879	0,001	0,000		$T_{30,30,25}^r$	/	/	47,017	12,345	0,454	
$T_{5,5,10}$	17,839	16,540	17,781	0,003	0,000		$T_{30,60,15}^r$	/	/	29,831	24,745	1,527	
$T_{5,10,10}$	17,129	15,974	17,014	0,627	0,000		$T_{40,60,55}^r$	/	/	103,966	19,595	0,354	
$T_{5,20,20}$	/	/	24,591	2,321	0,205		$T_{45,65,70}^r$	/	/	110,774	1,551	1,001	
$T_{6,6,6}$	16,626	16,169	16,569	3,355	0,076		$T_{50,70,70}^r$	/	/	117,526	4,260	1,467	
$T_{6,7,6}$	11,511	10,765	11,479	0,065	0,000		$T_{70,60,60}^r$	/	/	172,653	23,346	0,149	
$T_{7,5,6}$	16,426	15,785	16,140	0,773	0,110		$T_{80,65,65}^r$	/	/	182,556	45,854	0,272	
$T_{8,6,5}$	14,246	13,911	14,250	0,097	0,017		$T_{90,80,80}^r$	/	/	159,383	6,956	0,255	
$T_{8,40,10}$	/	/	27,920	0,001	0,000		$T_{100,85,85}^r$	/	/	185,36	40,487	0,140	
$T_{10,10,10}$	16,924	14,998	16,243	0,870	0,000		$T_{120,90,90}^r$	/	/	216,256	31,136	0,010	
$T_{15,20,15}$	/	/	26,014	2,298	0,248		$T_{150,100,100}^r$	/	/	240,303	40,148	0,025	
Prosek	/	/	18,514	0,950	0,055		Prosek	/	/	142,625	24,495	0,514	

6. Zaključak

Predmet ovog rada je varijanta problema raspoređivanja vozila koja se odnosi na optimizaciju transporta šećerne repe u Srbiji. Problem je formulisan u vidu Mešovitog celobrojnog programa sa kvadratnim ograničenjima

(MIQCP), koji uključuje sva specifična ograničenja problema i cilj optimizacije. MIQCP model je korišćen u okviru Lingo 17 rešavača, pri čemu su dobijena optimalna rešenja samo za realne test primere malih dimenzija. U cilju rešavanja instanci problema većih dimenzija, dizajnirana je Opšta metoda promenljivih okolina (GVNS) koja je prilagođena razmatranoj varijanti problema raspoređivanja vozila. Predložena GVNS metoda sastoji se iz faze razmrdavanja, metode promenljivog spusta (VND), umesto klasične lokalne pretrage i faze pomeraja. Eksperimentalni rezultati pokazuju da GVNS za vrlo kratko vreme izvršavanja dostiže sva optimalna rešenja dobijena Lingo rešavačem i dostiže ili poboljšava gornje granice na realniminstancama srednjih dimenzija koje nisu rešene do optimalnosti (sa izuzetkom jedne instance). Takođe, dizajnirana GVNS metoda obezbeđuje rešenja na generisaniminstancama većih dimenzija za relativno kratko vreme izvršavanja, imajući u vidu dimenzije problema.

Jedan od pravaca budućih istraživanja je kombinovanje dizajnirane GVNS metode sa drugim metodama optimizacije kako bi se poboljšao kvalitet dobijenih rešenja ili dodatno smanjilo vreme izvršavanja, pri rešavanju razmatranog VSP. Takođe, predloženi VSP problem se može proširiti dodavanjem novih ograničenja, na primer, uključivanjem vozila koja nisu homogena ili optimizacijom transporta do dve ili više fabrika za preradu.

Zahvalnica. Istraživanja na ovom radu su delimično podržana od strane Ministarstva obrazovanja, nauke i tehnološkog razvoja, Republike Srbije, kroz projekte pod brojem OI174010 i OI174033.

Bibliografija

- [1] **D. Agustina, C. Lee, R. Piplani.** Vehicle scheduling and routing at a cross docking center for food supply chains. *International Journal of Production Economics*, 2014, 152, pp. 29-41.
- [2] **A. Allahverdi.** The third comprehensive survey on scheduling problems with setup times/costs. *European Journal of Operational Research*, 2015, 246 (2), pp. 345-378.
- [3] **K. Androutsopoulos, K. Zografos.** An integrated modelling approach for the bicriterion vehicle routing and scheduling problem with environmental considerations. *Transportation Research Part C: Emerging Technologies*, 2017, 82, pp. 180-209.
- [4] **F. Baita, R. Pesenti, W. Ukovich, D. Favaretto.** A comparison of different solution approaches to the vehicle scheduling problem in a practical case. *Computers & Operations Research*, 2000, 27 (13), pp. 1249-1269.
- [5] **J. Brimberg, N. Mladenović, R. Todosijević, D. Urošević.** General variable neighborhood search for the uncapacitated single allocation p-hub center problem. *Optimization Letters*, 2015, 11 (2), pp. 377-388.
- [6] **M. Cheikh, M. Ratli, O. Mkaouar, B. Jarboui.** A variable neighborhood search algorithm for the vehicle routing problem with multiple trips. *Electronic Notes in Discrete Mathematics*, 2015, 47, pp. 277-284.
- [7] **G.B. Dantzig, D. R. Fulkerson.** Minimizing the number of tankers to meet a fixed schedule. *Naval Research Logistics Quarterly*, Wiley Subscription Services, Inc., A Wiley Company, 1954, 1, pp. 217 - 222.
- [8] **J. de Armas, B. Melián-Batista.** Variable neighborhood search for a dynamic rich vehicle routing problem with time windows. *Computers and Industrial Engineering*, 2015, 85, pp. 120-131.
- [9] **A. Haghani, M. Banihashemi, K.H. Chiang.** A comparative analysis of bus transit vehicle scheduling models. *Transportation Research Part B: Methodological*, 2003, 37 (4), pp. 301 - 322.
- [10] **P. Hansen, N. Mladenović, J.A.M. Pérez.** Variable neighbourhood search: methods and applications. *Annals of Operations Research*, 2010, 175 (1), pp. 367-407.
- [11] **A. Higgins.** Scheduling of road vehicles in sugarcane transport: A case study at an Australian sugar mill. *European Journal of Operational Research*, 2006, 170 (3), pp. 987 - 1000.
- [12] **Lindo Systems Inc.**, <http://www.lindo.com>, 2016.
- [13] **R. Macedo, C. Alves, S. Hanafi, B. Jarboui, N. Mladenović, B. Ramos, J.M.V. Carvalho.** Skewed general variable neighborhood search for the location routing scheduling problem. *Computers & Operations Research*, 2015, 61, pp. 143-152.
- [14] **E.L. Milan, S.M. Fernandez, L.M.P. Aragones.** Sugar cane transportation in Cuba, a case study. *European Journal of Operational Research*, 2016, 174 (1), pp. 374 - 386.
- [15] **N. Mladenović, P. Hansen.** Variable neighborhood search. *Computers & Operations Research*, 1997, 24 (11), pp. 1097 - 1100.
- [16] **N. Mladenović, D. Urošević, A. Ilić.** A general variable neighborhood search for the one-commodity pickup-and-delivery travelling salesman problem. *European Journal of Operational Research*, 2012, 220(1), 270-285.
- [17] **S. Moons, K. Ramaekers, A.A.Y. Caris.** Integrating production scheduling and vehicle routing decisions at the operational decision level: A review and discussion. *Computers and Industrial Engineering*, 2017, 104, pp. 224 - 245.
- [18] **S. Thuankaewsing, S. Khamjan, K. Piewthongngam, S. Pathumnakul.** Harvest scheduling algorithm to equalize supplier benefits: A case study from the Thai sugar cane industry. *Computers and Electronics in Agriculture*, 2015, 110, pp. 42 - 55.

- [19] **R. Todosijević, A. Mjirda, M. Mladenović, S. Hanafi, B. Gendron.** A general variable neighborhood search variants for the travelling salesman problem with draft limits. *Optimization Letters*, 2017, 11 (6), pp. 1047 - 1056.
- [20] **R. Todosijević, D. Urošević, N. Mladenović, S. Hanafi.** A general variable neighborhood search for solving the uncapacitated r-allocation p-hub median problem. *Optimization Letters*, 2017, 11 (6), pp. 1109 – 1121.
- [21] **H. Wang, J. Shen.** Heuristic approaches for solving transit vehicle scheduling problem with route and fueling time constraints. *Applied Mathematics and Computation*, 2007, 190 (2), pp. 1237 - 1249.
- [22] **Y. Xiao, A. Konak.** A simulating annealing algorithm to solve the green vehicle routing & scheduling problem with hierarchical objectives and weighted tardiness. *Applied Soft Computing*, 2015, 34, pp. 372 - 388.

Теорија бројева – дељивост, прости бројеви, конгруенције по модулу

Владимир Балтић

Висока школа за електротехнику и рачунарство, Београд
e-mail: baltic@matf.bg.ac.rs

Апстракт. У овом предавању даћемо краћи преглед већег дела Теорије бројева, која се јавља у редовној и додатној настави у основној и средњој школи. Увешћемо основне појмове из теорије бројева, као што су дељивост, највећи заједнички делитељ, прости и сложени бројеви, а затим ћемо прећи на конгруенције и њихове особине. Формулисаћемо и најважније теореме Теорије бројева, које су повезане са појмовима дељивости и когруенција. Све ово ћемо илустровати примерима.

Кључне речи: теорија бројева; конгруенције; прости и сложени бројеви.

1. Увод

Један од највећих математичара свих времена, Карл Фридрих Гаус је рекао: "Математика је краљица наука, а теорија бројева је краљица математике!" Још од античких времена људе су занимала нека својства бројева. На пример, у Вавилону, хиљаду година пре Питагоре, математичари су знали како систематски да нађу Питагорине бројеве, тј. целе бројеве који чине странице правоуглог троугла. На радовима који су остали иза Питагоре, Никомаха, Еуклида, Архимеда, Диофанта, Хипасије (она је била прва жена математичар) развијала се и средњовековна и ренесансна теорија бројева преко књига Фиbonачија, Луке Паћолија и Ферма. Потпуно независно, у другим деловима света теоријом бројева бавили су се кинески математичар Сун-Це, индијски Брамагупта, као и неколицина средњовековних арапских математичара. Осамнаест век доноси процват ове области кроз радове математичких величина Ојлера, Гауса, као и Лагранжа, Лежандра, Воринга, Фуријеа. У деветнаестом веку значајан траг у овој области оставили су Дирихле, Кронекер (који је рекао "Бог је створио целе бројеве, а све остало је дело човека!"), Кумер, Марков, Пеано (који је увео аксиоматику), док су прошли век обележили Виноградов, Пол Ердош, наш Александар Ивић и свакако Ендрју Вајлс (доказао Велику Фермаову теорему апаратом који излази ван оквира елементарне теорије бројева). Поред ових великана математике, теорија бројева је област која је заинтересовала и велики број математичара аматера (свако најпознатији од њих је француски правник Пјер Ферма), јер њени проблеми заузимају распон од оних разумљивих и лајцима до оних који су били нерешени неколико векова (или су и даље без решења).

У првом делу даћемо дефиницију дељивости, затим ћемо навести и неке од основних особина ове релације, навешћемо разне критеријеме дељивости са неким од природних бројева. На крају ћемо увести појам остатка при дељењу (који ће имати главну улогу у трећем делу), дати Алгоритам делења и навести појмове највећег заједничког делиоца (НЗД) и најмањег заједничког садржаоца (НЗС) са њиховим најважнијим особинама.

У другом делу се бавимо простим и сложеним бројевима. Наводимо метод Ератостеновог сита, поступке за растављање на просте чиниоце и тражење НЗД и НЗС и дајемо неке од најважнијих особина простих бројева (нпр. Основни став аритметике, Еуклидову теорему).

У трећем делу уводимо конгруенције по модулу, а природан разлог за то је што бројеви који дају исти остатак при дељењу бројем m имају мноштво заједничких особина. Одговарајућу нотацију увео је још Гаус у књизи "Disquisitiones arithmeticæ", која је била објављена 1801. године, када је Гаусу било свега 24 године. Навешћемо и доказати велики број особина конгруенција.

У четвртом делу ћемо навести (без доказа) велики број важних теорема Теорије бројева, које су у вези са појмовима дељивости и конгруенција: Ојлерову, малу Фермаову, Лагранжову, Вилсонову, Лукасову и Кинеску теорему о остатцима.

На основу њих ћемо објаснити када има решења конгруенција $ax \equiv b \pmod{m}$, као и како се она могу одредити. Осврнућемо се и на неке полиномне конгруенције.

Овим смо се дотакли великог дела Теорије бројева. Поменимо и теме која нисмо овде обухватили, а значајне су, попут квадратних остатака, Питагориних тројки и Диофантових једначина.

2. Дељивост бројева

Дефиниција 1. Шео број a је дељив целим бројем $b \neq 0$, ако постоји цео број q такав да је $a = bq$. Ако број b дели број a , то се означава са $b | a$ (Руси користе и ознаку $a : b$ за a је дељив бројем b). Тада још кажемо да је b делилац броја a . Ако b не дели a , то се пише $b \nmid a$.

Пример 1. Важи $3 | 15$, мј. $15 : 3$, али $6 \nmid 15$, јер је $15 = 3 \cdot 5$ и $15 = 6 \cdot 2 + 3$. \triangle

Већ на основу претходне дефиниције можемо да докажемо неколико особина дељивости целих бројева:

Теорема 1. Нека су n, m, a, b и d произвољни цели бројеви, тада важи

- а) $n | n$. (особина рефлексије)
- б) $d | n$ и $n | m$ повлачи $d | m$. (особина транзитивности)
- в) $d | n$ и $d | m$ повлачи $d | (an + bm)$. (особина линеарности)
- г) $d | n$ повлачи $ad | an$. (особина мултипликативности)
- д) $ad | an$ и $a \neq 0$ повлачи $d | n$. (особина скраћивања)
- ђ) $1 | n$. (1 дели сваки број)
- е) $n | 0$. (сваки број дели 0)
- ж) $d | n$ повлачи $-d | n$ и $d | -n$ и $d | -n$.
- з) $d | n$ и $n \neq 0$ повлачи $|d| \leq |n|$.
- и) $d | n$ и $n | d$ повлачи $|d| = |n|$, мј. $d = n$ или $d = -n$.
- ј) $d | n$ повлачи $\frac{n}{d} | n$.

За сваки природан број > 2 можемо да направимо критеријум дељивости са њим. За неке бројеве постоји и више различитих критеријума. То ћемо илустровати кроз наредне примере.

Пример 2. Сада ћемо дати неке од критеријума дељивости природних бројева са неким од бројева.

- 2: Број n је дељив са 2 (или другачије речено, број n је паран) $\Leftrightarrow n$ се завршава парном цифром, односно неком од цифара 0, 2, 4, 6 или 8.
- 3: Број n је дељив са 3 \Leftrightarrow збир цифара броја n је дељив са 3.
- 4: Број n је дељив са 4 $\Leftrightarrow n$ се завршава двоцифреним бројем који је дељив са 4.
- 5: Број n је дељив са 5 $\Leftrightarrow n$ се завршава неком од цифара 0 или 5.
- 6: Број n је дељив са 6 $\Leftrightarrow n$ је дељив и са 2 и са 3.
- 7: Број $n = \overline{a_k a_{k-1} \dots a_3 a_2 a_1}$ је дељив са 7 \Leftrightarrow и број $m = \overline{a_3 a_2 a_1} - \overline{a_6 a_5 a_4} + \overline{a_9 a_8 a_7} - \dots$ је дељив са 7.
- 7: Број n је дељив са 7 \Leftrightarrow и број m , који се добија тако што избршишемо цифру јединицу броја n и од тог броја одузмемо двоструку цифру јединица броја n , је дељив са 7.
- 8: Број n је дељив са 8 $\Leftrightarrow n$ се завршава троцифреним бројем који је дељив са 8.
- 9: Број n је дељив са 9 \Leftrightarrow збир цифара броја n је дељив са 9.
- 10: Број n је дељив са 10 $\Leftrightarrow n$ се завршава цифром 0.
- 11: Број $n = \overline{a_k a_{k-1} \dots a_3 a_2 a_1}$ је дељив са 11 \Leftrightarrow и број $m = a_1 - a_2 + a_3 - a_4 + \dots$ је дељив са 11.
- 11: Број $n = \overline{a_k a_{k-1} \dots a_3 a_2 a_1}$ је дељив са 11 \Leftrightarrow и број $m = \overline{a_2 a_1} + \overline{a_4 a_3} + \overline{a_6 a_5} + \overline{a_8 a_7} + \dots$ је дељив са 11.
- 12: Број n је дељив са 12 $\Leftrightarrow n$ је дељив и са 3 и са 4.

- 13: Broj $n = \overline{a_k a_{k-1} \dots a_3 a_2 a_1}$ je deљив са 13 \Leftrightarrow и број $m = \overline{a_3 a_2 a_1} - \overline{a_6 a_5 a_4} + \overline{a_9 a_8 a_7} - \dots$ је дељив са 13.
- 13: Broj n је дељив са 13 \Leftrightarrow и број m , који се добија тако што избришемо цифру јединица броја n и том броју додамо четворооструку цифру јединица броја n , је дељив са 13.
- 14: Broj n је дељив са 14 $\Leftrightarrow n$ је дељив и са 2 и са 7.
- 15: Broj n је дељив са 15 $\Leftrightarrow n$ је дељив и са 3 и са 5.
- 16: Broj n је дељив са 16 $\Leftrightarrow n$ се завршава четвороцифреним бројем који је дељив са 16.
- 25: Broj n је дељив са 25 $\Leftrightarrow n$ се завршава двоцифреним бројем који је дељив са 25.
- 27: Broj $n = \overline{a_k a_{k-1} \dots a_3 a_2 a_1}$ је дељив са 27 \Leftrightarrow и број $m = \overline{a_3 a_2 a_1} + \overline{a_6 a_5 a_4} + \overline{a_9 a_8 a_7} + \dots$ је дељив са 27.
- 37: Broj $n = \overline{a_k a_{k-1} \dots a_3 a_2 a_1}$ је дељив са 37 \Leftrightarrow и број $m = \overline{a_3 a_2 a_1} + \overline{a_6 a_5 a_4} + \overline{a_9 a_8 a_7} + \dots$ је дељив са 37.
- 100: Broj n је дељив са 100 $\Leftrightarrow n$ се завршава са две цифре 0 (mj. са 00). \triangle

Напомена 1. Може се показати и јаче тврђење:

n и збир цифара броја n дају исте остатке при дељењу са 3 (исто важи и за 9, као и за све претходно наведене критеријуме, сем оних где се брише последња цифра!).

Неке од претходних критеријума ћемо илустровати кроз следеће примере, а сусретаћемо их и касније у задацима.

Пример 3. Broj $n = 123\ 456\ 789 = 9 \cdot 3\ 607 \cdot 3\ 803$ је дељив са 9, а то потврђујемо и на основу критеријума који посматра његов збир цифара m :

$$m = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 = 45 = 9 \cdot 5. \checkmark$$

Broj $n = 1\ 234\ 567 = 9 \cdot 137\ 174 + 1$ није дељив са 9, а ни

$$m = 1 + 2 + 3 + 4 + 5 + 6 + 7 = 28 = 9 \cdot 3 + 1. \checkmark \triangle$$

Видимо и у овом примеру и да m и n дају исте остатке при дељењу са 9 (прво 0, па онда 1).

Пример 4. Broj $n = 4\ 976\ 761\ 801 = 11 \cdot 452\ 432\ 891$ је дељив са 11, као и

$$m = 1 - 0 + 8 - 1 + 6 - 7 + 6 - 7 + 9 - 4 = 11 = 11 \cdot 1. \checkmark$$

Broj $n = 25945 = 11 \cdot 2358 + 7$ није дељив са 11, као и

$$m = 5 - 4 + 9 - 5 + 2 = 7. \checkmark$$

До истих резултата долазимо и са другим критеријумом:

$$m = 01 + 18 + 76 + 76 + 49 = 220 = 11 \cdot 20. \checkmark \quad m = 45 + 59 + 2 = 106 = 11 \cdot 9 + 7. \checkmark \triangle$$

Пример 5. Broj $n = 1\ 123\ 477\ 641 = 13 \cdot 86\ 421\ 357$, је дељив са 13, као и

$$m = 641 - 477 + 123 - 1 = 286 = 13 \cdot 22. \checkmark$$

Broj $n = 865\ 366 = 13 \cdot 66\ 566 + 8$, није дељив са 13, као и

$$m = 366 - 865 = -499 = 13 \cdot (-39) + 8. \checkmark$$

До истих резултата долазимо и са другим критеријумом, само ћемо уместо $n = 1\ 123\ 477\ 641$ узети мањи број $n = 1118 = 13 \cdot 86$:

$$n = 1118 \Rightarrow m_1 = 111 + 4 \cdot 8 = 143$$

$$n = 143 \Rightarrow m_2 = 14 + 4 \cdot 3 = 26 = 13 \cdot 2. \checkmark$$

Број $n = 865\ 366 = 13 \cdot 66\ 566 + 8$.

$$n = 865366 \Rightarrow m_1 = 86536 + 4 \cdot 6 = 86560$$

$$n = 86560 \Rightarrow m_2 = 8656 + 4 \cdot 0 = 8656$$

$$n = 8656 \Rightarrow m_3 = 865 + 4 \cdot 6 = 889$$

$$n = 889 \Rightarrow m_4 = 88 + 4 \cdot 9 = 124$$

$$n = 124 \Rightarrow m_5 = 12 + 4 \cdot 4 = 28 = 13 \cdot 2 + 2.$$



У последњем примеру видимо да број n и сви бројеви m_1, m_2, m_3, m_4, m_5 дају све различите остатке при дељењу са 13 (редом то су: 8, 6, 11, 5, 7, 2).

Следећа теорема говори о томе да ако посматрамо два цела броја a и b и извршимо дељење a/b ($a \neq 0$), онда постоји само један цео број q (делимични количник) и само један ненегативан цео број r мањи од $|b|$ (остатак) тако да је

$$a = bq + r, \quad 0 \leq r < |b|. \quad (1)$$

Убрзо ћемо видети колико је јединственост овог записа важна. Једноставности ради, теорему ћемо доказати само за случај када је $b > 0$. Доказ за $b < 0$ је потпуно аналоган.

Теорема 2. (Алгоритам дељења).

Ако је $a \in \mathbb{Z}$ и $b \in \mathbb{N}$, онда a може на јединствен начин да се представи као

$$a = bq + r, \quad (q, r \in \mathbb{Z}, 0 \leq r < b).$$

Доказ. Да би доказ ове теореме био потпун, прво морамо да докажемо да бројеви q и r уопште постоје, а затим су јединствени:

Егзистенција: Посматрајмо аритметичку прогресију

$$\dots, a - 3b, a - 2b, a - b, a, a + b, a + 2b, a + 3b, \dots$$

која се у оба смера протеже у бесконачност. Онда мора да постоји најмањи ненегативни елемент прогресије, рецимо r . Тада је $r \geq 0$, а мора бити $r < b$, јер је у противном $r - b$ ненегативни члан прогресије мањи од r , што противречи избору r . Како r припада прогресији мора бити $r = a - bq$ за неки цео број q . Тиме смо утврдили егзистенцију бројева q и r .

Јединственост: Докажимо да су бројеви r и q једнозначно одређени.

Наиме ако је $a = bq_1 + r_1$ ($0 \leq r_1 < b$), $a = bq_2 + r_2$ ($0 \leq r_2 < b$), онда следи

$$a|q_1 - q_2| = |r_1 - r_2|.$$

Ако би било $q_1 \neq q_2$, онда је $a|q_1 - q_2| \geq a$, док је с друге стране $|r_1 - r_2| < a$, па мора бити $q_1 = q_2$ те онда следи и $r_1 = r_2$, чиме смо доказали јединственост q и r . \square

Пример 6. Поделимо 35 са 11 и -51 са -7.

$$35 = 11 \cdot 3 + 2, \quad -51 = (-7) \cdot 8 + 5.$$



Дефиниција 2. Џео број d за који важи $d | a$ и $d | b$ назива се заједнички делилац бројева a и b . Џео број s за који важи $a | s$ и $b | s$ назива се заједнички садржалац бројева a и b .

Како делилац бројева не може бити већи од апсолутне вредности тог броја, видимо да међу заједничким делиоцима бројева a и b постоји највећи. Слично међу садржаоцима постоји најмањи (само ћемо ту посматрати најмањи позитиван садржалац, тј. садржалац који је природан број).

Дефиниција 3. Највећи број у скупу делилаца бројева a и b се назива највећи заједнички делилац та два броја и обележава се са $\text{НЗД}(a, b)$ или (a, b) . Најмањи (позитиван) број у скупу садржалаца бројева a и b се назива најмањи заједнички садржалац та два броја и обележава се са $\text{НЗС}(a, b)$ или $[a, b]$.

Пример 7. Заједнички делиоци бројева 12 и 18 су бројеви: 1, 2, 3 и 6. Највећи од њих је 6, па важи $\text{НЗД}(12, 18) = 6$.

Заједнички садржоци бројева 12 и 18 су бројеви: 36, 72, 108, ... тј. бројеви облика $36k$, $k \in \mathbb{N}$. Најмањи од њих је 36, па важи $\text{НЗС}(12, 18) = 36$. \triangle

Ако су два броја узајамно проста, тј. ако немају заједничких фактора, онда им је највећи заједнички делилац 1, тј. $(a, b) = 1$. Због тога се чињеница да су a и b узајамно прости симболички пише баш тако, $(a, b) = 1$.

На основу Алгоритма дељења (Теорема 2) базиран је поступак за одређивање НЗД два природна броја. То је такозвани **Еуклидов алгоритам**. Он се састоји у следећем.

Нека су дати природни бројеви a и b , при чему без умањења општости можемо узети да је $a > b$ (у супротном ако је $b > a$ имали бисмо један корак више у Еуклидовом алгоритму). Тада ако поделимо a са b , према Алгоритму дељења (Теорема 2), постоје једнозначно одредјени бројеви q_1 и r_1 такви да је $a = b \cdot q_1 + r_1$, при чему је $0 \leq r_1 < b$. Даље, можемо поновити поступак са b и r_1 , $b > r_1$, тј. поделићемо b са r_1 и добити $b = r_1 \cdot q_2 + r_2$, при чему је $0 \leq r_2 < r_1$. Овај поступак ћемо настављати даље, све док не добијемо да је неки остатак једнак нули, $r_{k+1} = 0$. То можемо представити на следећи начин (овде су сви бројеви q_i и r_i , $1 \leq i \leq k+1$, единствено одређени на основу Алгоритма дељења):

$$\begin{aligned} a &= b \cdot q_1 + r_1, & (0 < r_1 < b) \\ b &= r_1 \cdot q_2 + r_2, & (0 < r_2 < r_1) \\ r_1 &= r_2 \cdot q_3 + r_3, & (0 < r_3 < r_2) \\ &\vdots \\ r_{k-3} &= r_{k-2} \cdot q_{k-1} + r_{k-1}, & (0 < r_{k-1} < r_{k-2}) \\ r_{k-2} &= r_{k-1} \cdot q_k + r_k, & (0 < r_k < r_{k-1}) \\ r_{k-1} &= r_k \cdot q_{k+1}, & (r_{k+1} = 0). \end{aligned}$$

Низ r_1, r_2, \dots, r_k је опадајући низ природних бројева мањих од b , што повлачи да се горњи поступак мора завршити у коначном броју корака. Следеће тврђење нам даје везу Еуклидовог алгоритма и највећег заједничког делиоца.

Теорема 3. Када применимо Еуклидов алгоритам на природне бројеве a и b , последњи позитиван остатак једнак је највећем заједничком делиоцу бројева a и b , тј. $\text{НЗД}(a, b) = r_k$.

Пример 8. Одредимо $\text{НЗД}(768, 180)$.

Решење. У Еуклидовом алгоритму имамо следећи низ делења:

$$\begin{aligned} 768 &= 180 \cdot 4 + 48, & (0 < 48 < 180) \\ 180 &= 48 \cdot 3 + 36, & (0 < 36 < 48) \\ 48 &= 36 \cdot 1 + 12, & (0 < 12 < 36) \\ 36 &= 12 \cdot 3, & (r = 0), \end{aligned}$$

па је $\text{НЗД}(768, 180) = 12$. \triangle

Последица Еуклидовог алгоритма је и следеће тврђење.

Теорема 4. Постоје цели бројеви x_0 и y_0 тако да је

$$(a, b) = ax_0 + by_0,$$

ако барем један од бројева a и b није нула.

Доказ претходног тврђења изостављамо (добили би га када би ишли од првог до претпоследњег корака у Еуклидовом алгоритму), али ћемо га илустровати на конкретном примеру бројева из претходног примера.

Пример 9. Представити НЗД($768, 180$) = 12 у облику $768x_0 + 180y_0$.

Решење. Из првог корака Еуклидовог алгоритма имамо да је $48 = 768 - 180 \cdot 4$. Формално ову једнакост би могли да запишемо као $48 = 768 \cdot 1 + 180 \cdot (-4)$ да би имали облик као у тврђењу претходне теореме. То нећемо чинити. Претходну једнакост уврстимо у једнакост из другог корака Еуклидовог алгоритма:

$$180 = 48 \cdot 3 + 36 = (768 - 180 \cdot 4) \cdot 3 + 36,$$

одакле након сређивања добијамо да је $36 = 180 \cdot 13 - 768 \cdot 3$.

Сада ћемо претходно добијено уврстити у једнакост из другог корака Еуклидовог алгоритма:

$$48 = 36 \cdot 1 + 12 \quad \Rightarrow \quad 768 - 180 \cdot 4 = (180 \cdot 13 - 768 \cdot 3) \cdot 1 + 12,$$

одакле након сређивања добијамо да је $12 = 768 \cdot 4 + 180 \cdot (-17)$, тј. $x_0 = 4$ и $y_0 = -17$. \triangle

Помоћу овог тврђења можемо показати и Питагорин резултат да број $\sqrt{2}$ није рационалан. То ћемо урадити у другом доказу.

Теорема 5. (Питагорина теорема).

Број $\sqrt{2}$ је ирационалан.

Доказ 1. Претпоставимо да је број $\sqrt{2}$ рационалан. Тада се он може представити у облику $\sqrt{2} = \frac{a}{b}$, где су a и b узајамно прости природни бројеви, $(a, b) = 1$. Квадрирањем преходне једнакости добијамо да је $a^2 = 2b^2$. Како је на десној страни паран број, следи да је и a^2 паран, што је могуће само уколико је и a паран, тј. $a = 2c$. Када то заменимо у претходну једнакост добијамо да је $(2c)^2 = 2b^2$, одакле је $b^2 = 2c^2$, па је b^2 паран, што повлачи и да је b паран, тј. $b = 2d$. Али онда смо добили да су и a и b парни, што је у супротности са тиме да су они узајамно прости. Како смо добили контрадикцију, полазна претпоставка да је $\sqrt{2}$ рационалан није тачна чиме је доказана теорема. \square

Доказ 2. Претпоставимо да је $\sqrt{2}$ рационалан тј. $\sqrt{2} = \frac{a}{b}$, где је $a, b \in \mathbb{N}$, $(a, b) = 1$. Према Теореми 4, постоје цели бројеви r и s тако да важи једнакост $ar + bs = 1$. Како из $\sqrt{2} = \frac{a}{b}$ имамо да $a = b\sqrt{2}$, а кад помножимо са $\sqrt{2}$ имамо $2b = a\sqrt{2}$, то повлачи да је

$$\sqrt{2} = \sqrt{2} \cdot 1 = \sqrt{2}(ar + bs) = (a\sqrt{2})r + (b\sqrt{2})s = 2br + as.$$

На основу претходне једнакости имао да је $\sqrt{2} = 2br + as \in \mathbb{Z}$, што је немогуће. Тиме је показано да полазна претпоставка није тачна. \square

Напомена 2. Штавише, важи да је (a, b) најмања позитивна вредност функције $ax + by$ кад x и y пролазе скупом \mathbb{Z} .

Пример 10. На основу претходне напомене и претходног примера имамо да је НЗД($180, 48$) = 12 најмања позитивна вредност функције $768x + 180y$ кад x и y пролазе скупом \mathbb{Z} .

Због тога имамо и да се ниједан од бројева из скупа $\{1, 2, \dots, 11\}$ не може представити у облику $768x + 180y$. \triangle

Теорема 6. а) $(ma, mb) = m(a, b)$, за $m \in \mathbb{N}$.

б) Ако $d | a, d | b$ и $d > 0$, онда је $\left(\frac{a}{d}, \frac{b}{d}\right) = \frac{1}{d}(a, b)$.

в) $\left(\frac{a}{(a, b)}, \frac{b}{(a, b)}\right) = 1$.

г) Ако је $(a, m) = (b, m) = 1$, онда је $(ab, m) = 1$.

д) Ако $c | ab$ и $(b, c) = 1$, онда $c | a$.

Везу између највећег заједничког делиоца и најмањег заједничког садржаоца бројева a и b даје нам следећа теорема:

- Теорема 7.** а) Ако $a_1 | b, a_2 | b, \dots, a_n | b$, онда $[a_1, \dots, a_n] | b$.
 б) $[ma, mb] = m[a, b]$.
 в) $(a, b) \cdot [a, b] = |a \cdot b|$.

Сада ћемо дефинисати НЗД и НЗС за више од 2 цела броја.

Дефиниција 4. Нека су a_1, a_2, \dots, a_n ненегативни цели бројеви, при чему је бар један различит од 0. Највећи природан број d који дели све те бројеве је њихов *највећи заједнички делилац* и означавамо га са $\text{НЗД}(a_1, a_2, \dots, a_n)$. Најмањи природан број који је садржалац сваког од бројева a_1, a_2, \dots, a_n је њихов *најмањи заједнички садржалац* и означавамо га са $\text{НЗС}(a_1, a_2, \dots, a_n)$.

Теорема 8. Нека су $a_1, a_2, \dots, a_n \in \mathbb{N}$ и $n \geq 3$.
 Тада важи да је $\text{НЗД}(a_1, a_2, \dots, a_n) = \text{НЗД}(\text{НЗД}(a_1, a_2, \dots, a_{n-1}), a_n)$.

На основу претходне теореме следи да НЗД више природних бројева можемо одредити вишеструком применом Еуклидовог алгоритма. У наредном поглављу ћемо се бавити факторизацијом природног броја на просте чиниоце и тада ћемо дати још неке методе за одређивање највећег заједничког делиоца и најмањег заједничког садржаоца.

3. Прости бројеви и основни став аритметике

Дефиниција 5. (Прости и сложени бројеви) Природни бројеви $p \in \mathbb{N}$ који имају тачно два делиоца називају се *прости бројеви*. Природни бројеви ($\neq 1$) који нису прости су *сложени бројеви*.

Наравно да два делиоца су 1 и p . Напоменимо да број 1 није прост број.

Низ простих бројева почиње овако:

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, ...

Теорема 9. Најмањи прост делилац сложеног броја n је мањи или једнак од \sqrt{n} .

Доказ. Прво ћемо показати да n има делилац који је већи од 1, али мањи или једнак од \sqrt{n} .
 Како је n сложен, тада се он може представити у облику производа $n = d_1 \cdot d_2$, где су $d_1 > 1$, $d_2 > 1$.

Ако би било $d_1 > \sqrt{n}$ и $d_2 > \sqrt{n}$, онда би важило

$$n = d_1 \cdot d_2 > \sqrt{n} \cdot \sqrt{n} = n,$$

што је контрадикција.

Даље, без умањења општости можемо узети да је $d_1 \leq \sqrt{n}$. Тада ће сваки прост делилац броја d_1 бити мањи или једнак од \sqrt{n} , па добијамо да најмањи прост делилац q броја d_1 задовољава неједнакост $q \leq \sqrt{n}$. Али како је сваки делилац од d_1 и делилац од n , то за најмањи прост делилац p броја n важи $p \leq q$, те за p важи неједнакост $p \leq \sqrt{n}$. Тиме је теорема доказана. \square

Следећи алгоритам, тзв. Ератостеново сито, је базиран на претходном тврђењу. Он је добио име по старогрчком математичару Ератостену из Кирене, који га је применио још у 3. веку пре нове ере.

Да одредимо све просте бројеве мање од n , можемо искористити метод познат под називом **Ератостеново сито**:

- прво испишемо све природне бројеве од 2 до $n - 1$;
- затим уоквиримо број 2 и редом из списка елиминишемо све парне бројеве изузев броја 2 ;
- затим уоквиримо број 3 и елиминишемо све бројеве дељиве са 3 , изузев самог броја 3 ;
- затим слично за $5, 7$ итд. понављамо поступак све до последњег простог броја мањег од \sqrt{n} .

Приметимо да после елиминације парних бројева није више потребно тражити бројеве дељиве са $4, 6$ итд. јер су ти сви бројеви парни, па су већ избачени. Слично, после елиминације бројева дељивих са 3 , више није потребно тражити бројеве дељиве са $6, 9$ итд. То је разлог, што нам као заокружени бројеви (тј. једини преостали бројеви) остају само прости бројеви. Притом, да би одредили присте бројеве мање од n користимо присте бројеве мање од \sqrt{n} (то је последица претходне теореме).

Пример 11. Ератостеновим ситом одредити све присте бројеве < 100 .

Решење. Да бисмо помоћу Ератостеновог сита одредили све присте бројеве < 100 , треба редом да елиминишемо све бројеве дељиве са $2, 3, 5$ и 7 (то су присти бројеви мањи од $\sqrt{100} = 10$). На почетку на табли имамо исписане следеће бројеве:

2	3	4	5	6	7	8	9	10	
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	

Затим уоквиримо 2 и прецртамо остале парне бројеве:

$$\begin{array}{ccccccccc} \boxed{2} & 3 & \times & 5 & \times & 7 & \times & 9 & \times \\ 11 & \times & 13 & \times & 15 & \times & 17 & \times & 19 & \times \\ 21 & \times & 23 & \times & 25 & \times & 27 & \times & 29 & \times \\ 31 & \times & 33 & \times & 35 & \times & 37 & \times & 39 & \times \\ 41 & \times & 43 & \times & 45 & \times & 47 & \times & 49 & \times \\ 51 & \times & 53 & \times & 55 & \times & 57 & \times & 59 & \times \\ 61 & \times & 63 & \times & 65 & \times & 67 & \times & 69 & \times \\ 71 & \times & 73 & \times & 75 & \times & 77 & \times & 79 & \times \\ 81 & \times & 83 & \times & 85 & \times & 87 & \times & 89 & \times \\ 91 & \times & 93 & \times & 95 & \times & 97 & \times & 99 & \end{array}$$

Сада је најмањи број који није ни уоквирен, ни прецртан 3 , па њега уоквиримо, а све остале

бројеве делимиве са 3 прецртамо:

$$\begin{array}{ccccccccc}
 & \boxed{2} & \boxed{3} & \times & 5 & \times & 7 & \times & \times \\
 11 & \times & 13 & \times & \times & \times & 17 & \times & 19 \\
 & \times & \times & 23 & \times & 25 & \times & \times & 29 \\
 31 & \times & \times & \times & 35 & \times & 37 & \times & \times \\
 41 & \times & 43 & \times & \times & \times & 47 & \times & 49 \\
 & \times & \times & 53 & \times & 55 & \times & \times & 59 \\
 61 & \times & \times & \times & 65 & \times & 67 & \times & \times \\
 71 & \times & 73 & \times & \times & \times & 77 & \times & 79 \\
 & \times & \times & 83 & \times & 85 & \times & \times & 89 \\
 91 & \times & \times & \times & 95 & \times & 97 & \times &
 \end{array}$$

Сада је најмањи број који није ни уоквирен, ни прецртан 5, па њега уоквиримо, а све остале бројеве делимиве са 5 прецртамо:

$$\begin{array}{ccccccccc}
 & \boxed{2} & \boxed{3} & \times & \boxed{5} & \times & 7 & \times & \times \\
 11 & \times & 13 & \times & \times & \times & 17 & \times & 19 \\
 & \times & \times & 23 & \times & \times & \times & \times & 29 \\
 31 & \times & \times & \times & \times & \times & 37 & \times & \times \\
 41 & \times & 43 & \times & \times & \times & 47 & \times & 49 \\
 & \times & \times & 53 & \times & \times & \times & \times & 59 \\
 61 & \times & \times & \times & \times & \times & 67 & \times & \times \\
 71 & \times & 73 & \times & \times & \times & 77 & \times & 79 \\
 & \times & \times & 83 & \times & \times & \times & \times & 89 \\
 91 & \times & \times & \times & \times & \times & 97 & \times &
 \end{array}$$

Сада је најмањи број који није ни уоквирен, ни прецртан 7, па њега уоквиримо, а све остале бројеве делимиве са 7 прецртамо:

$$\begin{array}{ccccccccc}
 & \boxed{2} & \boxed{3} & \times & \boxed{5} & \times & \boxed{7} & \times & \times \\
 11 & \times & 13 & \times & \times & \times & 17 & \times & 19 \\
 & \times & \times & 23 & \times & \times & \times & \times & 29 \\
 31 & \times & \times & \times & \times & \times & 37 & \times & \times \\
 41 & \times & 43 & \times & \times & \times & 47 & \times & \times \\
 & \times & \times & 53 & \times & \times & \times & \times & 59 \\
 61 & \times & \times & \times & \times & \times & 67 & \times & \times \\
 71 & \times & 73 & \times & \times & \times & \times & \times & 79 \\
 & \times & \times & 83 & \times & \times & \times & \times & 89 \\
 & \times & \times & \times & \times & \times & 97 & \times &
 \end{array}$$

Више нема неуоквирених простих бројева мањих од $\sqrt{100} = 10$, па стајемо са алгоритмом.

На табли су остали непрецртани само прости бројеви мањи од 100:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.$$

Укупно има 25 простих бројева мањих од 100. △

Приметимо да је 2 једини паран прост број. Такође и 3 је једини делив са 3.

Сви остали прости бројеви су облика $6k \pm 1$ што се често користи у задацима.

Означимо са $\pi(x)$ број простих бројева који нису већи од x .

На основу претходног примера имамо да је $\pi(100) = 25$.

Због неправилног распореда простих бројева, не можемо очекивати једноставну формулу за $\pi(x)$. Ипак, један од најимпресивнијих резултата напредне Теорије бројева даје асимптотско понапање за $\pi(x)$ и њега ћемо само формулисати у наредној теореми.

Теорема 10. (о броју простих бројева).

Важи једнакост

$$\lim_{x \rightarrow \infty} \pi(x) \frac{\ln x}{x} = 1,$$

где $\ln x$ означава природни логаритам (тј. логаритам са основом $e \approx 2,71\dots$).

Теорема 11. Сваки природан број $n > 1$ је или прост или је производ простих бројева.

Теорема 12. Ако прост број $p \nmid a$ онда је $(a, p) = 1$.

Теорема 13. Ако $p \mid ab$, онда $p \mid a$ или $p \mid b$. Општије, ако $p \mid a_1a_2\dots a_n$, тада p дели барем један од бројева a_1, a_2, \dots, a_n .

Теорема 14. (Основни став аритметике).

Сваки природан број $n > 1$ може се представити као производ простих бројева на јединствен начин (са тачношћу до њиховог поретка).

Ова теорема је кључни став тзв. мултипликативне теорије бројева. У факторизацији броја n неки прост број се може појавити као фактор више пута, рецимо $24 = 2 \cdot 2 \cdot 2 \cdot 3$. Ако су p_1, \dots, p_k сви различити прости фактори броја n , онда се n може представити као

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \dots \cdot p_k^{\alpha_k} = \prod_{j=1}^k p_j^{\alpha_j},$$

где су $\alpha_1, \alpha_2, \dots, \alpha_k$ једнозначно одређени природни бројеви. Ово се назива **канонска репрезентација** или (**каноничко разлагanje**) природног броја n . Она се врло често користи у теорији бројева.

Пример 12. (Републичко за VI разред, 2000.)

$$\text{Решити једначину } \frac{2}{73} = \frac{1}{60} + \frac{1}{219} + \frac{1}{292} + \frac{1}{x}.$$

Решење. Када извршимо факторизацију свих именилаца добијамо да је 73 прост, а остали су $60 = 2^2 \cdot 3 \cdot 5$, $219 = 3 \cdot 73$ и $292 = 2^2 \cdot 73$. Зато када све бројеве пребадимо на једну страну и сведемо их на заједнички именилац добијамо:

$$\frac{1}{x} = \frac{2}{73} - \frac{1}{60} - \frac{1}{219} - \frac{1}{292} = \frac{2 \cdot 2^2 \cdot 3 \cdot 5 - 73 - 2^2 \cdot 5 - 3 \cdot 5}{2^2 \cdot 3 \cdot 5 \cdot 73} = \frac{12}{2^2 \cdot 3 \cdot 5 \cdot 73} = \frac{1}{5 \cdot 73} = \frac{1}{365},$$

одакле је $x = 365$

△

У наредном примеру ћемо искористити канонску репрезентацију да одредимо НЗД и НЗС два природна броја.

Пример 13. Нека су дата 2 природна броја:

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \dots \cdot p_k^{\alpha_k} = \prod_{j=1}^k p_j^{\alpha_j}, \quad m = p_1^{\beta_1} \cdot p_2^{\beta_2} \cdot \dots \cdot p_k^{\beta_k} = \prod_{j=1}^k p_j^{\beta_j},$$

где су $\alpha_1, \alpha_2, \dots, \alpha_k, \beta_1, \beta_2, \dots, \beta_k$ ненегативни цели бројеви (нпр. ако у n нема фактора p_j , онда је $\alpha_j = 0$; слично ако у m нема фактора p_j , онда је $\beta_j = 0$).

Највећи заједнички делилац и најмањи заједнички садржалац ова два броја дат је са:

$$\text{НЗД}(n, m) = p_1^{\min(\alpha_1, \beta_1)} \cdot p_2^{\min(\alpha_2, \beta_2)} \cdot \dots \cdot p_k^{\min(\alpha_k, \beta_k)} = \prod_{j=1}^k p_j^{\min(\alpha_j, \beta_j)},$$

$$H3C(n, m) = p_1^{\max(\alpha_1, \beta_1)} \cdot p_2^{\max(\alpha_2, \beta_2)} \cdot \dots \cdot p_k^{\max(\alpha_k, \beta_k)} = \prod_{j=1}^k p_j^{\max(\alpha_j, \beta_j)}.$$

Број n дели m , $n \mid m$, ако и само ако важи $0 \leq \alpha_i \leq \beta_i$, за $i = 1, 2, \dots, k$. \triangle

Резултате претходног примера можемо искористити за ефективно израчунавање НЗД и НЗС два броја. Ту не морамо да одредимо канонске факторизације оба броја, па да применимо горње формуле са \min и \max . Ако одређујемо НЗД онда у сваком кораку издвајамо просте чиниоце који деле оба броја (ту рецимо користимо својство из Теореме 6. б), а ако одређујемо НЗС онда у сваком кораку издвајамо просте факторе који деле бар један број. То ћемо илустровати на следећем конкретном примеру.

Пример 14. Одредити $H3D(180, 168)$ и $H3C(180, 168)$. Проверити једнакост из Теореме 7. в). Да ли 168 дели 180?

Решење. 1. На основу следећих растављања на просте чиниоце:

$$\begin{array}{r|l} 180 & 2 \\ 90 & 2 \\ 45 & 3 \\ 15 & 3 \\ 5 & 5 \\ 1 & \end{array} \quad \begin{array}{r|l} 168 & 2 \\ 84 & 2 \\ 42 & 2 \\ 21 & 3 \\ 7 & 7 \\ 1 & \end{array}$$

дебијамо да је $n = 180 = 2^2 \cdot 3^2 \cdot 5$ и $m = 168 = 2^3 \cdot 3 \cdot 7$.

Представимо оба броја да имају заједничке просте факторе:

$$n = 180 = 2^2 \cdot 3^2 \cdot 5^1 \cdot 7^0 \quad \text{и} \quad m = 168 = 2^3 \cdot 3^1 \cdot 5^0 \cdot 7^1.$$

Даље, према резултатима Примера 13 имамо да је

$$\begin{aligned} H3D(n, m) &= 2^{\min(2, 3)} \cdot 3^{\min(2, 1)} \cdot 5^{\min(1, 0)} \cdot 7^{\min(0, 1)} = 2^2 \cdot 3^1 \cdot 5^0 \cdot 7^0 = 12, \\ H3C(n, m) &= 2^{\max(2, 3)} \cdot 3^{\max(2, 1)} \cdot 5^{\max(1, 0)} \cdot 7^{\max(0, 1)} = 2^3 \cdot 3^2 \cdot 5^1 \cdot 7^1 = 2520. \end{aligned}$$

Како је $12 \cdot 2520 = 30240 = 180 \cdot 168$, видимо да важи једнакост из Теореме 7. в).

Како степен двојке у m је већи од степена двојке у n , $3 > 2$, добијамо да $m \nmid n$ (могли смо да посматрамо и степене 7). \triangle

Решење. 2. Сада ћемо без факторисања бројева 180 и 168 да добијемо просте чиниоце од $H3D(180, 168)$ и $H3C(180, 168)$.

На левој страни ћемо оба броја да делимо са најмањим простим бројем са којим су оба дељива (и поступак се завршава када добијемо два узајамно прости броја), док ћемо на десној страни оба броја да делимо са најмањим простим бројем са којим је бар један од њих дељив – уколико је само један од тих бројева дељив са p , онда тај који је дељив поделимо са p , а други оставимо какав је био (и поступак се завршава када добијемо две 1):

$$\begin{array}{r|l} 180, 168 & 2 \\ 90, 84 & 2 \\ 45, 42 & 2 \\ 45, 21 & 3 \\ 15, 7 & 3 \\ 5, 7 & 5 \\ 1, 7 & 7 \\ 1, 1 & \end{array}$$

дебијамо да је $H3D(180, 168) = 2^2 \cdot 3^1 = 12$ и $H3C(180, 168) = 2^3 \cdot 3^2 \cdot 5 \cdot 7 = 2520$. \triangle

Пример 15. Број делилаца броја n се означава са $\tau(n)$. Ако број n има канонску репрезентацију онда је сваки делилац броја n облика

$$n = p_1^{\beta_1} \cdot p_2^{\beta_2} \cdot \dots \cdot p_k^{\beta_k},$$

где су бројеви $\beta_i (i = 1, 2, \dots, k)$ такви да важи $0 \leq \beta_i \leq \alpha_i (i = 1, 2, \dots, k)$. Због јединствености факторизације, која се односи и на d , сваки делилац броја n је на јединствен начин одређен избором експонената $\beta_i (i = 1, 2, \dots, k)$. Како β_i можемо да изаберемо на $\alpha_i + 1$ начина, видимо да је укупан број избора, односно укупан број делилаца броја n једнак

$$\tau(n) = (\alpha_1 + 1) \cdot (\alpha_2 + 1) \cdot \dots \cdot (\alpha_k + 1).$$

У теорији бројева је веома битна и Ојлерова фу-функција, $\varphi(n)$ која представља број природних бројева мањих од n и узајамно простих са n :

$$\varphi(n) = n \cdot \left(1 - \frac{1}{p_1}\right) \cdot \left(1 - \frac{1}{p_2}\right) \cdot \dots \cdot \left(1 - \frac{1}{p_k}\right).$$

Функције $\tau(n)$ и $\varphi(n)$ су мултипликативне функције, тј. важи: ако су бројеви m и n узајамно прости, $(m, n) = 1$, тада је

$$\tau(m \cdot n) = \tau(m) \cdot \tau(n) \quad \varphi(m \cdot n) = \varphi(m) \cdot \varphi(n).$$

Ова особина се може искористи за рачунање Ојлерове функције јер је за прост број p и $k \in \mathbb{N}_0$ испуњено

$$\varphi(p^k) = p^k - p^{k-1}.$$

Тако број $24 = 2^3 \cdot 3$ има $\tau(24) = (3+1) \cdot (1+1) = 8$ делилаца и то су: 1, 2, 4, 8, 3, 6, 12 и 24. Природни бројеви < 24 и узајамно прости са 24 су: 1, 5, 7, 11, 13, 17, 19 и 23 и њих има $\varphi(24) = 24 \cdot \left(1 - \frac{1}{2}\right) \cdot \left(1 - \frac{1}{3}\right) = 8$. Сада још да проверимо мултипликативност ових функција. $\tau(8) = (3+1) = 4$ (делиоци су 1, 2, 4, 8), $\tau(3) = (1+1) = 2$ (делиоци су 1, 3) и $\tau(24) = \tau(8) \cdot \tau(3) = 4 \cdot 2 = 8$. $\varphi(8) = \varphi(2^3) = 2^3 - 2^2 = 8 - 4 = 4$ (узајамно прости су 1, 3, 5, 7), $\varphi(3) = \varphi(3^1) = 3^1 - 3^0 = 3 - 1 = 2$ (узајамно прости су 1, 2), $\varphi(24) = \varphi(8) \cdot \varphi(3) = 4 \cdot 2 = 8$. \triangle

Одговор на питање колико има простих бројева знали су још старогрчки математичари. Еуклид у IX књизи својих "Елемената" даје следећи доказ чињенице да је скуп простих бројева бесконачан:

Теорема 15. (Еуклидова теорема).

Не постоји највећи прост број.

Доказ. Претпоставимо да постоји само коначно много простих бројева: p_1, p_2, \dots, p_n и нека је p_n највећи међу њима. Ако формирамо број

$$P = p_1 p_2 \dots p_n + 1,$$

видимо да је $P > p_n$ и да за P постоје две могућности:

1º P је прост број — што је у супротности са претпоставком да је p_n највећи прост број;
2º P је сложен број — P при делењу са сваким од p_1, p_2, \dots, p_n даје остатак 1, тј. P није делив ни са једним од њих, односно P има неке друге просте факторе, што је у супротности са претпоставком да су p_1, p_2, \dots, p_n сви прости бројеви.

Добијањем контрадикције смо показали да простих бројева нема коначно много, тј. да је скуп простих бројева бесконачан. \square

У наредном примеру искористићемо доказ Еуклидове теореме, као и још неке процене да би ограничили величину n -тог простог броја.

Пример 16. Одредимо од чега је мањи n -ти прост број.

Решење. На основу Еуклидовог доказа Теореме 15 имамо да важи $p_{n+1} \leq p_1 p_2 \dots p_n + 1$ за $n \geq 1$. Ако смањимо све индексе за 1 добијамо неједнакост:

$$p_n \leq p_1 p_2 \dots p_{n-1} + 1 \quad \text{за } n \geq 2.$$

Са малом модификацијом Еуклидовог резоновања, претходна неједнакост може бити побољшана на:

$$p_n \leq p_1 p_2 \dots p_{n-1} - 1 \quad \text{за } n \geq 3.$$

Рецимо за $n = 6$ имамо да је

$$p_6 = 13 \leq 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 - 1 = 2309,$$

па видимо да је ово веома груба процена. Мало боље ограничење даје **Бонсова неједнакост** (енг. Bonse's inequality):

$$(p_n)^2 < p_1 p_2 \dots p_{n-1} \quad \text{за } n \geq 5.$$

Из ње имамо да је $(p_6)^2 < 2310$, тј. $p_6 \leq 48$.

Још бољу процену добијамо из неједнакости:

$$p_{2n} < p_2 p_3 \dots p_n \quad \text{за } n \geq 3.$$

Одатле је $p_6 < p_2 \cdot p_3 - 2 = 3 \cdot 5 - 2 = 13$. △

Проблем код свих процена из претходног примера је што за њихову примену морамо да знамо вредности свих мањих простих бројева. Ограниччење које не зависи од претходних простих бројева је дато у наредној теореми.

Теорема 16. Нека p_n означава n -ти прост број. Тада важи $p_n \leq 2^{2^{n-1}}$.

Следећа последица ове теореме је занимљива.

Последица 1. За $n \geq 1$ постоји бар $n + 1$ простих бројева мањих од 2^{2^n} .

Навешћемо још две познате теореме везане за просте бројеве, које се често користе на међународним такмичењима, а њихови докази су далеко од елементарних. Након тога ћемо навести и неколико отворених проблема Теорије бројева (за та тврђења још увек није познато да ли су тачна или не).

Теорема 17. (Берtranов постулат – показао га је Чебишев).

За сваки природан број $n > 1$ постоји прост број p , такав да важи $n < p < 2n$.

Резултат и овог тврђења можемо искористи за боље процене величине n -тог простог броја.

Последица 2. Важи $p_n < 2^n$ за $n \geq 2$.

Последица 3. Важи $p_{n+1} < 2p_n$ за $n \geq 2$.

Теорема 18. (Дирихлеова теорема).

Ако су a и b узајамно прости бројеви, $(a, b) = 1$, тада постоји бесконечно много простих бројева облика $an + b$, где је $n = 1, 2, \dots$

Неки од најпознатијих отворених проблема теорије бројева (везаних за просте бројеве) су:

- **Голдбахова хипотеза.** Сваки паран број већи од 2 може се представити у облику збира 2 прости броја, а сваки непаран број већи од 7 може се представити као збир 3 непарна прости броја (оват други део је 1937. године показао руски математичар Виноградов аналитичким методама, за довољно велике непарне бројеве $n > n_0$; Виноградов није дао процену за n_0 , а то је први учинио Бородкин 1956. године – $n_0 < 3^{3.15}$, док је 1989. граница за n_0 смањена на $10^{43.000}$).
- Не зна се да ли између свака 2 узастопна квадрата n^2 и $(n+1)^2$ постоји прост број.
- Не зна се да ли постоји бесконачно много простих бројева облика $4k^2 + 1$.
- Кронекер је дао претпоставку да се сваки позитиван паран број може представити као разлика 2 прости броја на бесконачно много начина (специјалан случај овог тврђења је и хипотеза да постоји бесконачно много парова простих бројева близанаца, тј. парова простих бројева који се разликују за 2).

4. Конгруенције

Бројеви који дају исти остатак при дељењу бројем m имају заједничких особина. Због тога се и уводе следећа дефиниција и ознака:

Дефиниција 6. За целе бројеве a и b који при дељењу са $m \neq 0$ дају исте остатке (тј. ако цео број m дели $a - b$; ова два исказа су еквивалентна због Теореме 2) каже се да су *конгруентни по модулу m* . Симболички се то пише

$$a \equiv b \pmod{m}.$$

Ако m не дели $a - b$, каже се да a *није конгруентно* b по модулу m и пише се

$$a \not\equiv b \pmod{m}.$$

Ову нотацију увео је Гаус у књизи "Disquisitiones arithmeticæ", која је била објављена 1801. године, када је Гаусу било свега 24 године.

Како је $a - b$ дељиво са m ако и само ако је дељиво са $-m$, можемо без ограничења општости у конгруенцијама претпоставити да су модули m природни бројеви. Стога ћемо од сад па надаље претпостављати да је m природан број.

Пример 17. На основу Примера 6, из једнакости $35 = 11 \cdot 3 + 2$ следи да је $35 \equiv 2 \pmod{11}$, а из $-51 = (-7) \cdot 8 + 5 = 7 \cdot (-8) + 5$ следи да је $-51 \equiv 5 \pmod{7}$. \triangle

Конгруенције имају и многе заједничке особине са једнакостима. Неке од тих особина следе непосредно из дефиниције конгруенције и садржи их следећа теорема.

Теорема 19. Нека су a, b, c, d, x и y произвољни цели бројеви. Тада важи:

- а) $a \equiv a \pmod{m}$. (особина рефлексивности)
- б) $a \equiv b \pmod{m}$, $b \equiv a \pmod{m}$ и $a - b \equiv 0 \pmod{m}$ су еквивалентна тврђења. (особина симетричности)
- в) Ако је $a \equiv b \pmod{m}$ и $b \equiv c \pmod{m}$ онда је и $a \equiv c \pmod{m}$. (особина транзитивности)
- г) Ако је $a \equiv b \pmod{m}$ и $c \equiv d \pmod{m}$ онда је и $ax + cy \equiv bx + dy \pmod{m}$. (особина линеарности)
- д) Ако је $a \equiv b \pmod{m}$, онда постоји цео број q такав да је $a = mq + b$.
- ђ) Ако је $a \equiv b \pmod{m}$ и $c \equiv d \pmod{m}$ онда је и $ac \equiv bd \pmod{m}$. (особина мултипликативности)
- е) Ако је $a \equiv b \pmod{m}$ онда је и $a^k \equiv b^k \pmod{m}$.
- ж) Ако је $a \equiv b \pmod{m}$ и $f(x)$ полином са целим коефицијентима, онда је $f(a) \equiv f(b) \pmod{m}$.
- з) Ако је $a \equiv b \pmod{m}$ и $d | m$, онда је $a \equiv b \pmod{d}$.

За реалне бројеве важи особина скраћивања: ако је $ax = ay$ и $a \neq 0$, онда је $x = y$. Код скраћивања конгруенција потребно је више опреза, као што показује следећа теорема.

Теорема 20. а) $ax \equiv ay \pmod{m}$ ако и само ако је $x \equiv y \pmod{\frac{m}{(a,m)}}$ ($a \neq 0$).

б) Ако је $ax \equiv ay \pmod{m}$ и $(a,m) = 1$, онда је $x \equiv y \pmod{m}$.

в) $a \equiv b \pmod{m_i}$ за $i = 1, 2, \dots, k$ ако је $a \equiv b \pmod{H3C(m_1, \dots, m_k)}$.

Пример 18. Важи $2 \cdot 5 \equiv 2 \cdot 14 \pmod{6}$, али не можемо скратити са 2, јер је $5 \not\equiv 14 \pmod{6}$.

Како је $(2, 6) = 2$ и $\frac{6}{(2, 6)} = \frac{6}{2} = 3$, према Теореми 20 а) имамо да је $5 \equiv 14 \pmod{3}$. \triangle

На основу особина а), б) и в) Теореме 19 следи да је конгруенција по модулу m релација еквиваленције. Класе еквиваленције су бројеви који при дељењу са m дају исти остатак. У раду са бројевима по модулу m у суштини вршимо уобичајене операције аритметике $+$ и \cdot , али занемарујемо умношке броја m . За свако $a \geq 1$ нека је

$$a = qm + r \quad (0 \leq r < m)$$

по алгоритму дељења (Теорема 2). Тада је $a \equiv r \pmod{m}$, и види се да је сваки цео број конгруентан по модулу m неком од бројева $0, 1, \dots, m-1$. Такође је јасно да међу последњим бројевима никоја два нису конгруентна по модулу m . Каже се да тих m бројева образује потпуни систем остатака по модулу m . У општем случају бројеви x_1, x_2, \dots, x_m образују потпуни систем остатака по модулу m ако $x_i \not\equiv x_j \pmod{m}$ за $i \neq j$. Тада је опет сваки број a конгруентан једном (и само једном) од бројева x_i по модулу m . Јасно је да потпуних система остатака има бесконачно много, јер се рецимо додањем истог броја сваком елементу из потпуног система остатака добијају бројеви који опет чине потпуни систем остатака.

Теорема 21. а) Ако је $a \equiv b \pmod{m}$, онда је $(a, m) = (b, m)$.

б) Ако је $a \equiv b \pmod{m}$, $a \equiv b \pmod{n}$ и $(m, n) = 1$ онда је $a \equiv b \pmod{mn}$.

Каже се да бројеви r_1, r_2, \dots, r_t образују редукованни (или сведенни) систем остатака по модулу m ако је $(r_i, m) = 1$ за $i = 1, 2, \dots, t$ и $r_i \not\equiv r_j$ за $i \neq j$, и ако сваки x за које је $(x, m) = 1$ задовољава $x \equiv r_i \pmod{m}$ за неко i . Јасно је да се редуковани систем остатака по модулу m може добити ако се из потпуног система остатака по модулу m одстране сви они бројеви x_i за које је $(x_i, m) > 1$, тј. задрже они бројеви x_i за које је $(x_i, m) = 1$. Једнозначно одређени број t представља Ојлерову функцију, тј. $t = \varphi(m)$. Јасно је да је $\varphi(1) = 1$. Ако је p прост број и $k \geq 1$, онда је

$$\varphi(p^k) = p^k - p^{k-1},$$

јер су бројеви $n \leq p$ за које је $(n, p^k) > 1$ бројеви $p, 2p, 3p, \dots, p^k$, а њих је укупно p^{k-1} .

Пример 19. Ако посматрамо остатке по модулу 8 (тј. $m = 8$), тада бројеви $0, 1, 2, 3, 4, 5, 6, 7$ чине потпуни, а $1, 3, 5, 7$ редукованни систем остатака по модулу 8. Сваки цео број је конгруентан по модулу 8 неком од бројева из потпуног система остатака по модулу 8, а сваки број узайамно прост са 8 је конгруентан по модулу 8 неком од бројева из редукованог система остатака по модулу 8, тј. неком од бројева 1, 3, 5 или 7. $t = 4 = \varphi(8) = 8 \cdot (1 - \frac{1}{2})$. \triangle

5. Основне теореме теорије бројева

У овом поглављу ћемо само навести (без доказа) основне теореме теорије бројева.

Теорема 22. (Ојлерова теорема).

Ако је $(a, m) = 1$, онда је

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

У посебном случају Ојлерова теорема, када је $m = p$ прост број (тада је $\varphi(p) = p - 1$), је

$$a^{p-1} \equiv 1 \pmod{p} \quad \text{за} \quad (a, p) = 1,$$

а множењем са a добија се следећи резултат, који је у литератури познат као *мала Фермаова теорема* (велика Фермаова теорема је тврђење да једначина $x^n + y^n = z^n$ нема решења у скупу природних бројева ако је $n \geq 3$ и она је била вековима изазов за математичаре, а решио је 1997. године Ендрју Вајлс).

Теорема 23. (мала Фермаова теорема).

Ако је p прост број и a цео број који није делјив са p , онда је

$$a^{p-1} \equiv 1 \pmod{p}.$$

Ова теорема се често задаје и у облику $a^p \equiv a \pmod{p}$, где је p прост број, а a произвољан број.

Мала Фермаова теорема налази практичну примену у рачунарству: за потребе кодирања потребно је одредити велике просте бројеве (једини 100% ефикасан начин за установљавање сложености неког броја n је Ератостеново сито, али тај поступак изискује дosta времена) и тада ако нам неки број задовољава малу Фермаову теорему за неколико (што више то боље — већа је вероватноћа да је тај број заиста прост) различитих вредности a онда га можемо сматрати као прост за генерирање неког кода.

Пример 20. Помоћу Ојлерове теореме можемо да решимо конгруенцију

$$ax \equiv b \pmod{m} \quad ((a, m) = 1), \tag{2}$$

где је x непозната, а величине a , b и m су дате.

Решење. Ова једначина не може да има више од једног решења по модулу m . Наиме, ако x_1 и x_2 задовољавају (2), онда следи

$$ax_1 \equiv ax_2 \pmod{m},$$

а по Теореми 20 б) добија се да је $x_1 \equiv x_2 \pmod{m}$. С друге стране

$$x_1 = a^{\varphi(m)-1}b$$

је очигледно решење једначине (2), јер је

$$ax_1 = a^{\varphi(m)}b \equiv 1 \cdot b \pmod{m}$$

по Ојлеровој теореми. Другим речима једначина (2) има јединствено решење по модулу m . \triangle

Теорема 24. Линеарна конгруенција $ax \equiv b \pmod{m}$ има решења ако и само ако $d \mid b$, где је $d = \text{НЗД}(a, m)$. Ако $d \mid b$ тада конгруенција има d међусобно неконгруентних решења по модулу m :

$$x_0, x_0 + \frac{m}{d}, x_0 + 2\frac{m}{d}, \dots, x_0 + (d-1)\frac{m}{d},$$

где је x_0 неко решење дате конгруенције.

На основу претходне теореме имамо да ако $d \nmid b$ тада линеарна конгруенција $ax \equiv b \pmod{m}$ има решења, што се често користи у задацима. Такође ово тврђење за конгруенције, може се свести на линеарну једначину, што је дато у наредној последици.

Последица 4. Линеарна једначина $ax + my = b$ (ово је једначина по x и y) има решења у скупу целих бројева ако и само ако $d \mid b$, где је $d = \text{НЗД}(a, m)$.

Пример 21. (Међуопштинско за V разред, 1987.)

Збир два разломка са једноцифреним имениоцима је $\frac{11}{18}$. Одреди о којим разломцима је реч. Колико решења има?

Решење. Прво ћемо разматрати ако су разломци већ скраћени, тј. да за њихове имениоце и бројиоце важи $(p, q) = 1$ и $(r, s) = 1$. Онда се задатак своди на једначину $\frac{p}{q} + \frac{r}{s} = \frac{11}{18}$, где су $(p, q) = 1$ и $(r, s) = 1$.

За имениоце ова два разломка, q и s , мора да важи да је $\text{НЗС}(q, s) = 18$. Из услова задатка су q и s једноцифрени бројеви, а они су и различити од 1. Зашто? Уколико би неки од њих био једнак 1, онда би њихов НЗС био једнак оном другом броју, који је једноцифрен, а не 18.

Како се у задатку ништа не каже за бројиоце, они могу бити произвољни цели бројеви (0 не могу бити из истог разлога, као што 1 није могло да буде у имениоцу).

На основу претходог, имамо следеће случајеве:

$$1^\circ \quad \frac{p}{2} + \frac{r}{9} = \frac{11}{18}, \quad 2^\circ \quad \frac{p}{3} + \frac{r}{6} = \frac{11}{18}, \quad 3^\circ \quad \frac{p}{6} + \frac{r}{9} = \frac{11}{18}.$$

1° када једанкост $\frac{p}{2} + \frac{r}{9} = \frac{11}{18}$ помножимо са 18, добијамо линеарну једначину $9p + 2r = 11$. Код ње $d = \text{НЗД}(9, 2) = 1 \mid 11$, па према претходној последици она има решења. Ако њу посматрамо по модулу 2, добијамо (еквивалентну) конгруенцију $9p \equiv 11 \pmod{2}$, тј. добијамо $p \equiv 1 \pmod{2}$ и њено решење чине сви непарни цели бројеви p , тј. $p = 2k + 1$, $k \in \mathbb{Z}$. Када то уврстимо у $9p + 2r = 11$ добијамо да је $r = 1 - 9k$.

2° добијамо линеарну једначину $6p + 3r = 11$. Код ње $d = \text{НЗД}(6, 3) = 3 \nmid 11$, па према претходној последици она има решења (до тога смо могли доћи јер је лева страна дељива са 3, а десна није).

3° добијамо линеарну једначину $3p + 2r = 11$. Код ње $d = \text{НЗД}(3, 2) = 1 \mid 11$, па и она има решења. Опет посматрамо по модулу 2, па добијамо $3p \equiv 11 \pmod{2}$, тј. добијамо $p \equiv 1 \pmod{2}$ и њено решење чине сви непарни цели бројеви p , тј. $p = 2k + 1$, $k \in \mathbb{Z}$. Када то уврстимо у $3p + 2r = 11$ добијамо да је $r = 4 - 3k$.

Коначно добијамо да задатак има бесконачно много решења облика:

$$\frac{2k+1}{2} + \frac{1-9k}{9} \quad \text{и} \quad \frac{2k+1}{6} + \frac{4-3k}{9}, \quad \text{где је } k \in \mathbb{Z}.$$

Уколико би се у задатку тражило да су имениоци природни бројеви, онда би требало одредити вредности за k тако да су истовремено $2k + 1 \in \mathbb{N}$ и $1 - 9k \in \mathbb{N}$ (у првом случају), као и да су истовремено $2k + 1 \in \mathbb{N}$ и $4 - 3k \in \mathbb{N}$ (у трећем случају). То нам даје само решења $\frac{1}{2} + \frac{1}{9}$ и $\frac{1}{6} + \frac{4}{9}$ (где су сви разломци скраћени), али како од првог можемо добити још 3 решења код којих је именилац једноцифрен, али разломци нису скраћени, коначно добијамо да би у овој варијанти задатак имао 5 решења: $\frac{1}{2} + \frac{1}{9}$, $\frac{2}{4} + \frac{1}{9}$, $\frac{3}{6} + \frac{1}{9}$, $\frac{4}{8} + \frac{1}{9}$ и $\frac{1}{6} + \frac{4}{9}$. \triangle

По аналогији са проблемима решавања алгебарских једначина поставља се природно и проблем решавања конгруенција. У општем случају се решава конгруенција

$$f(x) \equiv 0 \pmod{m},$$

где је f полином n -тог степена (због тога је потребно да је $a_n \not\equiv 0 \pmod{m}$ јер се у противном добија конгруенција са полиномом степена мањим од n) $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ чији су коефицијенти a_j цели бројеви.

Ако је скуп $\{x_1, x_2, \dots, x_m\}$ потпуни систем остатака по модулу m , број решења конгруенције $f(x) \equiv 0 \pmod{m}$ се дефинише као број оних x_i за које је $f(x_i) \equiv 0 \pmod{m}$. По Теореми 19 е) јасно је да је број решења $f(x) \equiv 0 \pmod{m}$ независан од избора потпуног система остатака по модулу m , као и да тај број решења никада не прелази m .

Пример 22. Одредимо разне полиномне конгруенције колико имају решења.

- $x^3 - x \equiv 0 \pmod{3}$ има три, тј. максималан број решења (то су 0, 1, 2).
- $x^2 + 1 \equiv 0 \pmod{7}$ нема решења.
- $x^2 + 1 \equiv 0 \pmod{5}$ има два решења (то су 2, 3).
- $x^2 - 1 \equiv 0 \pmod{8}$ има четири решења (то су 1, 3, 5, 7).

△

Теорема 25. (Лагранжова теорема).

Нека је p прост број и $p \nmid a_n$ и нека је $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ дати полином са целим коефицијентима. Тада конгруенција

$$f(x) \equiv 0 \pmod{p}$$

има највише n решења по модулу p .

Тврђење не мора бити тачно ако p није прост број, као што рецимо показује пример конгруенције $x^2 - 1 \equiv 0 \pmod{8}$, која има четири решења: 1, 3, 5 и 7.

Последица 5. Нека је p прост број и нека је $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$ дати полином са целим коефицијентима. Ако конгруенција

$$f(x) \equiv 0 \pmod{p}$$

има више од n решења по модулу p , онда је сваки коефицијент полинома $f(x)$ делив са p .

Имајући у виду Лагранжову теорему и њену последицу, сада можемо лако да докажемо један од класичних ставова елементарне теорије бројева, који је у литератури познат као Вилсонова теорема. Следећа теорема је занимљива, јер даје формулу која важи за све просте бројеве и ни за један други број (то није случај са малом Фермаовом теоремом – видети задатак ??). На жалост, коришћење те формуле за испитивање сложености бројева, односно одређивање простих бројева, није нимало практично. Ову теорему је први открио Лајбниц 1682. године (видети задатак ??), али данас носи Вилсоново име. $(p-1)!$ представља производ првих $p-1$ природних бројева, тј. $(p-1)! = 1 \cdot 2 \cdot \dots \cdot (p-1)$.

Теорема 26. (Вилсонова теорема).

Конгруенција

$$(p-1)! \equiv -1 \pmod{p}$$

важи ако и само ако је p прост број.

Теорема 27. (Лукасова теорема).

Нека за $n, a \in \mathbb{N}$ важи

$$a^{n-1} \equiv 1 \pmod{n} \quad \text{и} \quad a^{\frac{n-1}{p}} \not\equiv 1 \pmod{n}$$

за сваки прост делилац p броја $n - 1$. Тада је n прост број.

Систем од две или више конгруенција не мора да има решења, премда свака појединачна конгруенција има решења. Рецимо не постоји x које истовремено задовољава $x \equiv 1 \pmod{2}$ и $x \equiv 0 \pmod{4}$, мада свака од појединачних конгруенција има решења. Разлог томе је што модули конгруенција 2 и 4 нису узајамно прости. Следећа теорема, позната у литератури као *Кинеска теорема о остацима* даје услове под којима више линеарних конгруенција има заједничко решење ако су модули конгруенција узајамно прости у паровима. Мада је прво опште решење за проблеме овог типа дао Ојлер, ова теорема се назива кинеском јер је кинески математичар Сун-Цу у I веку решио задатак који се своди на налажење целих бројева x који при дељењу са 3, 5 и 7 дају редом остатке 2, 3 и 2. Тада проблем је еквивалентан следећем систему конгруенција:

$$\left. \begin{array}{l} x \equiv 2 \pmod{3} \\ x \equiv 3 \pmod{5} \\ x \equiv 2 \pmod{7} \end{array} \right\} \Rightarrow x \equiv 23 \pmod{105}.$$

105 је једнако $105 = 3 \cdot 5 \cdot 7$.

Теорема 28. (Кинеска теорема о остацима).

Нека су m_1, m_2, \dots, m_r природни бројеви који задовољавају $(m_i, m_j) = 1$ за $i \neq j$, и нека су b_1, b_2, \dots, b_r произвољни цели бројеви. Тада систем конгруенција

$$\left. \begin{array}{l} x \equiv b_1 \pmod{m_1} \\ x \equiv b_2 \pmod{m_2} \\ \vdots \\ x \equiv b_r \pmod{m_r} \end{array} \right\}$$

има тачно једно решење x_0 по модулу $m = m_1 \cdot m_2 \cdot \dots \cdot m_r$.

Важи и следеће уопштење претходне теореме.

Теорема 29. Нека су m_1, m_2, \dots, m_r природни бројеви који су узајамно прости у паровима $(m_i, m_j) = 1$ за $i \neq j$ и нека су a_1, a_2, \dots, a_r и b_1, b_2, \dots, b_r произвољни цели бројеви. Тада систем конгруенција

$$\left. \begin{array}{l} a_1x \equiv b_1 \pmod{m_1} \\ a_2x \equiv b_2 \pmod{m_2} \\ \vdots \\ a_rx \equiv b_r \pmod{m_r} \end{array} \right\}$$

има тачно једно решење x_0 по модулу $m = m_1 \cdot m_2 \cdot \dots \cdot m_r$.

Пример 23. Сада ћемо на примеру Сун-Цеовог задатка да илуструјемо како се користи Кинеска теорема о остацима.

Решење. Из система

$$\left. \begin{array}{l} x \equiv 2 \pmod{3} \\ x \equiv 3 \pmod{5} \\ x \equiv 2 \pmod{7} \end{array} \right\}$$

налазимо да је $m_1 = 3$, $m_2 = 5$, $m_3 = 7$, $m = m_1 \cdot m_2 \cdot m_3 = 105$ и $\frac{m}{m_1} = 35$, $\frac{m}{m_2} = 21$, $\frac{m}{m_3} = 15$ и $b_1 = 2$, $b_2 = 3$, $b_3 = 2$. Сада треба да решимо три конгруенције (сваку понаособ) по непознатим c_1 , c_2 и c_3 :

$$\begin{array}{ll} 35c_1 \equiv 1 \pmod{3} & -c_1 \equiv 1 \pmod{3} \\ 21c_2 \equiv 1 \pmod{5} & c_2 \equiv 1 \pmod{5} \\ 15c_3 \equiv 1 \pmod{7} & c_3 \equiv 1 \pmod{7} \end{array} \Rightarrow$$

Одмах добијамо решења $c_1 = 2$, $c_2 = 1$ и $c_3 = 1$, што нам даје основно решење

$$x_0 = 35 \cdot 2 \cdot 2 + 21 \cdot 1 \cdot 3 + 15 \cdot 1 \cdot 2 = 233 \equiv 23 \pmod{105},$$

а одатле налазимо и коначно решење

$$x = 105k + 23, \quad k \in \mathbb{Z}.$$

△

Помоћу Кинеске теореме о остацима може се показати да се решавање полиномне конгруенције $f(x) \equiv 0 \pmod{m}$ може свести на решавање једноставнијих конгруенција $f(x) \equiv 0 \pmod{p^k}$, а затим се може ићи и даље и показати да је довољно ограничити се на конгруенције типа $f(x) \equiv 0 \pmod{p}$.

Теорема 30. *Систем линеарних конгруенција*

$$\begin{aligned} ax + by &\equiv r \pmod{m} \\ cx + dy &\equiv s \pmod{m} \end{aligned}$$

има јединствено решење када је $H3D(ad - bc, m) = 1$.

Као што смо видели конгруенције играју веома битну улогу у Теорији бројева. Неке од најзначајнијих теорема у елементарној теорији бројева се показују помоћу апарате који користи конгруенције. Даље у изучавању опште теорије бројева следе квадратни остаци које је увео Лежандр ради решавања квадратних конгруенције, а основни став је дао Гаус и по њему се зове Гаусов закон квадратног реципроцитета.

Библиографија

- [1] James A. Anderson, "Discrete Mathematics with Combinatorics", Pearson Prentice Hall, 2004.
- [2] Alan Baker, "A concise introduction to the theory of numbers", Cambridge University Press, 1984
- [3] Владимиран Балтић, Душан Ђукић, Ђорђе Кртић, Иван Матић, *Препремни задаци за тајмичења из математике ученика средњих школа*, ДМС, Београд, 2008.
- [4] David M. Burton, "Elementary Number Theory", McGraw-Hill 2002.
- [5] Александар Адольфович Бухштаб, "Теорија чисела", Москва, 1960.
- [6] Иван Матвеевич Виноградов "Основы теории чисел", Москва, 1965.
- [7] Richard K. Guy, "Unsolved Problems in Number Theory", Springer-Verlag, 1994.
- [8] Leonard Eugene Dickson, "History of the Theory of Numbers", New York, 1952
- [9] Александар Ивић "Увод у аналитичку теорију бројева", Београд 1995.
- [10] Бранислав Кисачанин "Мала математика", Нови Сад 1995.
- [11] Марија Станић, Небојша Икодиновић, "Теорија бројева: збирка задатака", Завод за уџбенике, Београд, 2004.
- [12] Марија Станић (редактор), "Тангента 600 – Збирка задатака из часописа Тангента", ДМС, Београд, 2014.
- [13] Ратко Тошић, Вања Вукославчевић, "Елементи теорије бројева", Алеф, Нови Сад, 1995.
- [14] Дмитрий Константинович Фаддеев, Илья Самуилович Соминский, "Сборник задач по высшей алгебре", Москва, 1977.
- [15] *On-line encyclopedia of integer sequences*,
<http://oeis.org>
- [16] Weisstein, E.W, *The World of Mathematics*,
<http://mathworld.wolfram.com/topics/NumberTheory.html>
<http://mathworld.wolfram.com/NumberTheory.html>

Razvoj geometrijskog mišljenja kroz tangram aktivnosti

Nives Baranović

Filozofski fakultet u Splitu
e-mail: nives@ffst.hr

Sanja Lehman

Četvrta gimnazija Marko Marulić Split
e-mail: sanja.lehman@gmail.com

Apstrakt. Na prvi pogled, drevna kineska slagalica tangram i ne izgleda kao ozbiljan matematički alat, ali taj skup od 7 jednostavnih geometrijskih likova pruža mnoštvo mogućnosti za (ozbiljnu) primjenu u nastavi matematike.

Brojna znanstvena istraživanja o učenju i poučavanju geometrije zadnjih desetljeća otkrivaju razloge zašto mnogi učenici imaju teškoće pri učenju geometrijskih sadržaja te daju smjernice koje bi mogle pomoći u savladavanju tih teškoća (vidjeti [1], [2], [4], [5], [7] - [13]). Neka od njih podupiru tangram aktivnosti u nastavi matematike u korist razvoja geometrijskog mišljenja i procesa zaključivanja. Jer, osim što učenje pomoću tangrama potiče aktivno sudjelovanje učenika, njihovu motivaciju i interes, kreativnost i maštu, kroz tangram aktivnosti učenici obogaćuju matematički rječnik, razvijaju strategiju rješavanja problema, vizualno-prostorne vještine; samostalno istražuju, uočavaju pravilnosti i postavljaju tvrdnje, uče s većim razumijevanjem itd. (vidjeti [5], [8], [9], [10], [13]).

U radu se opisuju tangram aktivnosti koje potiču razvoj geometrijskog mišljenja, vizualno-prostornih i strateških vještina koje podupiru učenje geometrije s većim razumijevanjem. Aktivnosti su usmjerene na otkrivanje teorema o tangramu i njegovih lema kroz usmjereno opažanje, vizualno-analitičkom metodom.

Nastavnici, koji bi opisane ili neke druge tangram aktivnosti željeli provesti u učionici, najprije sami trebaju samostalno istraživati te iskustveno doživjeti pravu matematičku snagu i nastavni potencijal ove slagalice, a tek onda tome poučavati učenike (vidjeti [9], [10]).

Ključne reči: geometrijsko mišljenje; tangram aktivnosti; usmjereno opažanje; van Hiele-ova teorija, vizualno-analitička metoda.

1. Uvod

„Kada pažljivo njegujete geometrijsko mišljenje učenika, oni će biti uspješniji u savladavanju i Euklidove matematike.“

Pierre M. van Hiele

Matematika, a posebno geometrija kao njezin bitan dio, omogućava prikazivanje, opisivanje i svakodnevno snalaženje u svijetu u kojem živimo. Zato je jedan od glavnih ciljeva matematičkog obrazovanja podizanje razine geometrijskog mišljenja učenika, koje je neophodno za primjenu geometrije (vidjeti [7]).

Međutim, geometrija se često zanemaruje u školskom kurikulumu i marginalizira kroz nastavu matematike, što je prema nekim autorima najčešće posljedica nedostatka primjerenih nastavnih sredstava, ali i nedovoljne stručnosti nastavnika (vidjeti [7], [12]). Mnogi učenici imaju teškoća pri učenju geometrije, pri vizualizaciji određenih geometrijskih sadržaja, kao i pri zaključivanjima koja se na njima temelje, što je često rezultat nepri-mjerenog poučavanja, a ono je pak posljedica nedovoljnog poznавanja geometrijskih koncepata i razina mišljenja potrebnih za njihov razvoj (vidjeti [12]).

Zahvaljujući tehnološkom razvoju i primjeni računala u gotovo svim sferama ljudskog djelovanja na prije-lazu s 20. na 21. stoljeće, potreba za geometrijskim znanjima stalno je u porastu. Sve veća važnost pridaje se vizualizaciji koja je neophodna u primjeni raznih računalnih programa, a čiji razvoj se u velikoj mjeri temelji na geometriji. Budući je geometrija neophodna za razvoj geometrijskog mišljenja, vizualno-prostornih vještina, de-duktivnog zaključivanja, strateškog djelovanja, vještina rješavanja problema, tj. vještina potrebnih za snalaženje i funkcioniiranje u 21. stoljeću, ona se opet postupno vraća u kurikulum matematičkog obrazovanja na svim razinama (vidjeti [2]). Različita obrazovna istraživanja potvrđuju da učenici spomenute vještine mogu razvijati kroz

ciljano odabране i primjereno osmišljene tangram aktivnosti. Tako Tchoshanov opisuje tangram aktivnosti kroz koje učenici poboljšavaju konceptualno razumijevanje, usavršavaju proceduralne te izgrađuju strateške vještine, razvijaju sposobnost logičkog promišljanja i kritičkog osvrta, posebno naglašavajući da kroz aktivno sudjelovanje i samostalno istraživanje učenici uspješno povezuju različite sadržaje u funkcionalnu cjelinu, a koje inače kroz nastavu matematike uče izolirano (vidjeti [10, str. 16-22]).

Cilj ovog rada je prikazati neke tangram aktivnosti koje učenicima omogućavaju da kao aktivni sudionici samostalno istražuju, otkrivaju, izvode zaključke, svoje rezultate opisuju i uspoređuju i na taj način geometrijska znanja usvajaju s većim razumijevanjem i bez straha. Uz to se demonstrira metoda usmjerjenog opažanja te vizualno-analitička metoda, koja uvelike doprinosi razvoju vizualno-prostornih i strateških vještina te geometrijskog mišljenja.

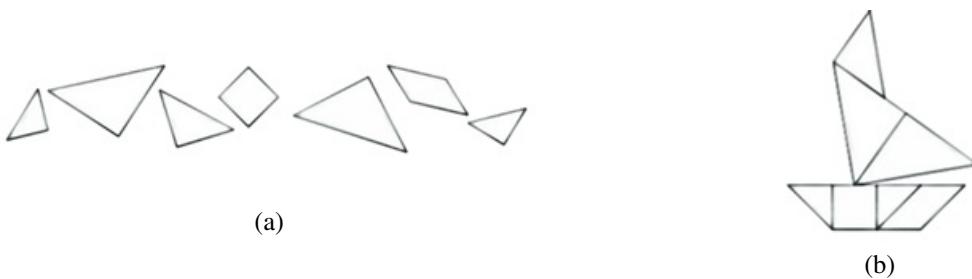
Tangram aktivnosti, koje će se opisati u ovom radu, temelje se na teoremu o tangramu i njegovom dokazu (kojeg bi analitički mogli izvesti i razumjeti tek napredniji srednjoškolci), a obrađuju se na način koji je primijeren svim učenicima 8. razreda osnovne škole pa na dalje. Time se želi ukazati na to kako se apstraktna geometrijska znanja, primjenom vizualizacije u istraživanju algebarsko - geometrijskog problema, mogu približiti učenicima na njima razumljiv i prihvatljiv način.

2. Osnovno o tangramu

Tangram je prihvaćen kao drevna kineska slagalica iako se o porijeklu te slagalice zapravo jako malo zna. Ni porijeklo samog naziva nije u potpunosti razjašnjeno, premda su neki nazivi vremenom bili učestaliji pa je tako naziv *tangram* danas postao opće prihvaćen. Zanimljiv je jedan stari kineski naziv, *Ch'i ch'iao t'u*, koji se na engleski prevodi sa „Seven ingenious plan”, što u našem prijevodu znači „Sedam pločica mudrosti” (vidjeti [4]).

Pojavom prvih tiskanih izdanja početkom 19. stoljeća, tangram postaje popularna rekreacijska slagalica ne samo u Kini, već i u Americi, Europi i Aziji, a njezina popularnost stalno raste sve do danas. Svojom posebnošću, osim u svrhu zabave, tangram sve više privlači stručnjake raznih profila: dizajneri izrađuju namještaj i posude u obliku tangrama, umjetnici stvaraju nakit nalik tangram likovima, menadžeri i filatelisti tangram likovima promoviraju razne ideje ..., a matematičari ga koriste u svrhu učenja raznih matematičkih sadržaja i razvijanja matematičkog mišljenja, posebno geometrije i geometrijskog mišljenja.

Prema jednoj od poznatih legendi, tangram slagalica je nastala kada se sluga, noseći kralju staklo kvadratnog oblika za prozor njegove palače, spotakao i staklo razbio. Ali na iznenadenje svih, staklo nije bilo uništeno već razlomljeno na 7 točno određenih geometrijskih oblika (slika 1(a)). U nastojanju da se pred kraljem opravda, prepričavajući svoje naporno putovanje, sluga je oblikovao razne oblike radi što veće vjerodostojnosti svoje priče. Tako je oblikom sa slike 1(b) prikazao brod kojim je prešao rijeku. Kralj je, oduševljen mogućnostima slaganja raznovrsnih oblika, dao izraditi identične drvene geometrijske likove i tako je nastala tangram slagalica, koju je prihvatio cijelo njegovo kraljevstvo (vidjeti [6]).



Slika 1. Dijelovi tangram slagalice i model broda sastavljen od tih dijelova

Kao što se može vidjeti sa slike 1, tangram slagalica sastoji se od pet jednakokračnih pravokutnih trokuta (među kojima su dva para sukladnih) te od jednog kvadrata i jednog paralelograma. Svaki od tih dijelova naziva se **tangram dio** ili najčešće samo **tan**. Naime, kako bi istaknuli pripadnost određenog lika tangram slagalici, za najmanji trokut se kaže mali trokut tangrama ili mali tan trokut, za kvadrat se kaže kvadrat tangrama ili tan kvadrat itd. (Slika 1(a)).

Kada se od svih sedam tanova oblikuje novi lik onda se on naziva **tangram lik**. Na slici 1(b) lik oblikovan od svih sedam tanova prikazuje brod pa se naziva tangram brod. Slično, na slici 2 prikazan je kvadrat koji je

oblikovan od svih sedam tanova te se za njega kraće kaže tangram kvadrat. Ako se pri slaganju od svih sedam tanova oblikuje konveksni lik, onda se za njega kraće kaže **konveksni tangram lik**.



(a)



(b)

Slika 2. Tangram kvadrat

Pomoću svih sedam tanova može se oblikovati na tisuće različitih oblika, među kojima se opet mogu proučavati posebne skupine kao što su konveksni tangram likovi, tangram zvjezdice, tangram uvale i sl. Ovaj rad bavi se posebno konveksnim tangram likovima.

3. Tangram kao nastavno sredstvo

Kao što je u uvodu spomenuto, tangram slagalica može biti korisno **didaktičko sredstvo** u nastavi matematike, a posebno u nastavi geometrije. Međutim, preduvjet za učinkovito korištenje tangrama u nastavi matematike je iskustvo nastavnika. Tek kad sami nastavnici steknu određena iskustva u tangram aktivnostima, onda mogu ciljano osmišljavati te iste aktivnosti za učenike.

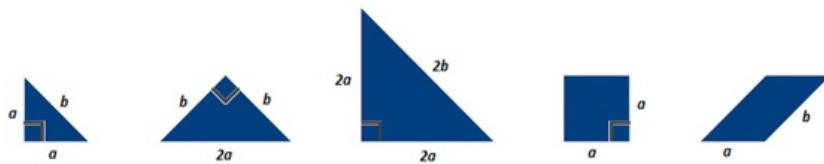
Prema Siew & Abdullah, nastavnici su, sudjelujući u tangram aktivnostima, osvijestili da oni sami imaju nedovoljno razvijeno geometrijsko mišljenje i vizualno-prostorne sposobnosti. Nakon provedenih istraživačkih aktivnosti, crtanja, otkrivanja pravilnosti i postavljanja tvrdnji, nastavnici su istaknuli kako pozitivnije gledaju na geometriju, više je cijene i osjećaju veće samopouzdanje u rješavanju geometrijskih problema te da bi takvom načinu rada trebalo ostaviti više prostora u nastavi matematike (vidjeti [8, str. 256]).

Za efikasno uvođenje tangram u nastavu matematike postoje razne preporuke (vidjeti [4], [11]) i svi se slažu u jednom: prilikom prvog susreta s tangramom, korisno je dopustiti učenicima da najprije rade zadatke **otvorenog tipa**, ovisno o njihovom uzrastu i interesima. Koristeći svih sedam tanova, učenici mogu oblikovati: slova engleskog alfabeta, znamenke, interpunkcijske znakove, različite vrste simbola, oblike koji prikazuju osobe, životinje, biljke, predmete itd. Na primjer, tangram slovima može se zapisati riječ tangram (slika 3).



Slika 3. Riječ tangram zapisana tangram slovima

Odmah nakon toga, važno je da učenici samostalno istraže i opišu osnovne karakteristike tanova, služeći se pritom odgovarajućim definicijama, aksiomima i teoremima jer se te karakteristike dalje koriste u mnogim tangram aktivnostima. Na primjer, ako se radi s modelima tangrama, preklapanjem malih (velikih) tan trokuta do podudarnosti stranica i kutova jednostavno se provjerava da se radi o sukladnim trokutima; uzastopnom primjenom aksioma: *Ako su dva objekta jednakata trećem objektu, onda su i oni međusobno jednakci* jednostavno se izvodi zaključak da su svi tan trokuti jednakokračni; a primjenom poučka o ispruženom kutu: *Ako dva sukladna kuta zajedno čine ispruženi kut, onda su to pravi kutovi* jednostavno se izvodi zaključak da su svi tan trokuti pravokutni. Do istih zaključaka može se doći i proučavanjem tangram kvadrata sa slike 2(a), a posebno ako se smjesti u kvadratnu točkastu mrežu kao na primjer na slici 2(b). Nепосредно nakon takvog istraživanja, korisno je na jednom mjestu istaknuti odnose između stranica i kutova svih tanova (slika 4) jer su ti odnosi važni u dalnjim istraživanjima. Više o osnovnim karakteristikama tanova može se pročitati u [4, str. 10]).



Slika 4. Odnosi stranica svih tanova

Na sličan način mogu se istražiti tanovi s obzirom na njihove površine. Na primjer, jedna tangram aktivnost može biti pronalaženje svih mogućih načina oblikovanja velikog tan trokuta pomoću ostalih tanova (slika 5). Zatim, ako se uzme da je P mjeru površine malog tan trokuta, onda se na temelju te aktivnosti može izvesti zaključak da je mjeru površine srednjeg tan trokuta, tan kvadrata i tan paralelograma jednaka $2P$, a mjeru površine velikog tan trokuta je $4P$. Na kraju ove aktivnosti prikidan je trenutak da se komentira kako svi tangram likovi imaju jednaku površinu ($16P$), bez obzira na oblik. Na primjer, sva tangram slova sa slike 3 su jednake površine.



Slika 5. Odnosi površina tanova

Prethodno opisane tangram aktivnosti osiguravaju prirođan put za uvođenje koncepta mjerne jedinice i mjerjenja površine, kao i pojma sukladnosti likova. Na sličan način se mogu razvijati i koncepti drugih matematičkih pojmovima: omjera, razlomka, postotka, klasične vjerojatnosti, sličnosti itd.

Osim za razvoj različitih koncepata, tangram aktivnosti mogu biti usmjerene na razvoj različitih vještina kroz crtanje, istraživanje, otkrivanje i postavljanja matematičkih zakonitosti, a time i na razvoj matematičkog (geometrijskog) rječnika i procesa mišljenja. Više o korištenju tangrama kao matematičkog alata može se pročitati u [4] i [11].

U ovom radu će se opisati odgovarajuće tangram aktivnosti primjerene učenicima za samostalno istraživanje, otkrivanje i postavljanje matematičkih zakonitosti, a vezanih za konveksne tangram likove.

4. Konveksni tangram likovi

Poseban podskup geometrijskih likova su konveksni likovi. Za lik u ravnini kažemo da je **konveksan lik** ako taj lik za svake svoje dvije točke sadrži i dužinu koja te dvije točke spaja. Na slici 6(a) i 6(d) prikazani su konveksni likovi, a na slici 6(b) i 6(c) nekonveksni.



Slika 6. Konveksni i nekonveksni likovi

Na sličan način, među tisućama različitih tangram likova mogu se promatrati samo konveksni. Dva kineska matematičara, Fu Traing Wang i Chuan-Chih Hsiung, 1942. godine dokazali su da postoji **samo 13 konveksnih tangram likova** (vidjeti [14]).

Njihov dokaz temelji se na četiri leme koje opisuju karakteristike slaganja 16 malih tan trokuta u konveksan oblik, a glavni dio dokaza temelji se na činjenici da se svaki konveksan lik može upisati u pravokutnik. Iz toga izvode kvadratnu diofantsku jednadžbu sa šest nepoznanica i četiri nejednakosti. Postavljena jednadžba ima 48 različitih rješenja, koja određuju 20 različitih konveksnih likova, među kojima je samo 13 tangram likova (Slika 7).



Slika 7. Konveksni tangram likovi

Svijetlo obojani likovi sa slike 7, konveksni su likovi koji se ne mogu popločati tanovima pa ne spadaju u konveksne tangram likove. Tamno obojani likovi sa slike 7, čine preostalih 13 konveksnih tangram likova, među kojima se nalazi: jedan trokut, šest četverokuta, dva peterokuta i četiri šesterokuta.

U dalnjem radu, metodom usmjerjenog opažanja otkrivaju se iskazi lema i samog teorema, a vizualno-analitičkom metodom svi konveksni likovi, prikazani na slici 7.

4.1. Otkrivanje iskaza lema primjenom metode usmjerjenog opažanja

Istraživanje u učionici se može provesti kroz osam aktivnosti, u obliku individualnog ili grupnog rada. Za provođenje svih aktivnosti potrebno je osigurati konture svih 13 konveksnih tangram likova te za svakog sudionika model tangram slagalice, kvadratnu točkastu mrežu, jednu grafitnu olovku i dvije olovke različitih boja. Pritom, konture konveksnih tangram likova trebaju odgovarati modelu tangram slagalice, koju učenici koriste. Iz praktičnih razloga, ovdje se koristi kontura samo jednog konveksnog tangram lika - pravokutnog trapeza.

Aktivnost 1: Oblikovati lik zadan konturom (pravokutni trapez, slika 8(a)) koristeći sve dijelove tangrama - tanove. Napomena: Površinu omeđenu konturom treba cijelu prekriti tanovima, bez praznina i bez preklapanja¹. Jedan od načina popločavanja konveksnog tangram trapeza prikazan je na slici 8(b).

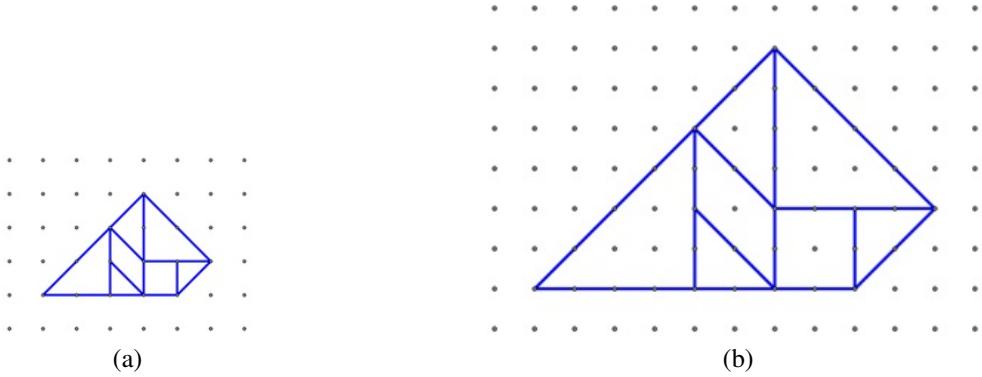


Slika 8. Kontura tangram trapeza i rješenje

Aktivnost 2: Rješenje dobiveno u 1. aktivnosti treba skicirati u točkastoj kvadratnoj mreži, grafitnom olovkom. Napomena: Dimenzije tangram likova, pri skiciranju, potrebno je proporcionalno umanjiti, a vrhove svih tanova smjestiti u točke mreže (slika 9(a) ili 9(b)).

Ove dvije aktivnosti potiču razvoj vizualno-prostornih vještina te uspostavljanje odnosa među pojedinim dijelovima, a te vještine osiguravaju razvoj mišljenja na 1. i 2. razini prema van Hiele-ovojo teoriji (vidjeti [9], [12]). Učenici koji nisu u potpunosti savladali 1. razinu mišljenja imat će teškoće pri crtanjtu srednjeg tan trokuta jer se nalazi u nestandardnom položaju. Učenici koji se nalaze na 2. razini mišljenja i nisu u potpunosti savladali 3. razinu, imat će teškoće pri crtanjtu lika sa svim dijelovima u kvadratnu mrežu na opisani način jer neće biti u mogućnosti uspostaviti odgovarajuće odnose među stranicama likova koji se dodiruju. Bez ikakvih teškoća lik će nacrtati učenici koji se nalaze na 3. razini mišljenja i više.

¹Ovaj postupak prekrivanja površine naziva se popločavanje.



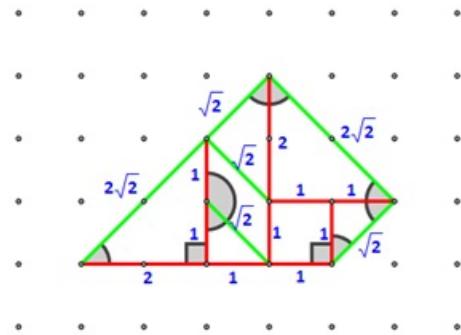
Slika 9. Tangram trapez u točkastoj kvadratnoj mreži

Aktivnost 3: Pod pretpostavkom da je tan kvadrat **jedinični kvadrat**, odrediti **duljine stranica** promatranog tangram trapeza i duljine stranica svakog od sastavnih tanova. Opisati na koji način se dodiruju susjedni tanovi, s obzirom na stranice duž kojih se dodiruju. Ishod 3. aktivnosti prikazan je na slici 10, na primjeru pravokutnog tangram trapeza.

U dalnjem radu, sve stranice čije su duljine iskazane racionalnim brojevima nazivaju se **racionalne stranice**, a stranice čije su duljine iskazane iracionalnim brojevima nazivaju se **iracionalne stranice**.

Aktivnost 4: U liku prikazanom u točkastoj kvadratnoj mreži, jednom bojom istaknuti sve racionalne stranice, a drugom sve iracionalne stranice. Zatim odgovoriti na sljedeće: (a) Kakve su stranice prikazanog tangram lika: racionalne / iracionalne? (b) Proučite i opišite **veze među stranicama** ovisno o bojama (vrstama stranica).

Aktivnost 5: Odrediti **mjere unutarnjih kutova** prikazanog konveksnog tangram lika i mjere unutarnjih kutova svih sastavnih tanova. Zatim proučiti i opisati **veze među stranicama i kutovima** ovisno o bojama (vrstama stranica).



Slika 10. Duljine stranica i kutovi u tangram trapezu

Na temelju 4. aktivnosti može se iskazati **Lema 1:** U konveksnom tangram liku, sastavni tanovi su posloženi tako da se racionalna stranica nalazi uz racionalnu, a iracionalna uz iracionalnu (vidjeti [14, str. 596]).

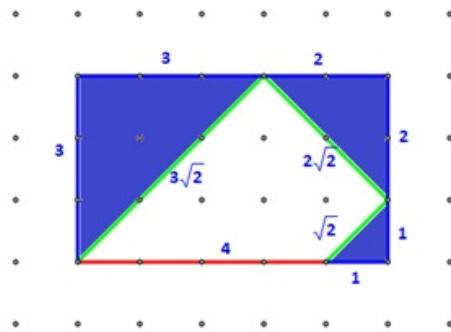
Zaključci koji se mogu izvesti na temelju 3., 4. i 5. aktivnosti zapravo su tvrdnje **Leme 2:** (a) Stranice konveksnog tangram lika su ili racionalne (slika 10, crvena boja) ili iracionalne (slika 10, zelena boja); (b) Racionalne stranice konveksnog lika sastavljene su od racionalnih stranica sastavnih tanova, a iracionalne od iracionalnih stranica sastavnih tanova. (c) Stranice iste vrste su međusobno paralelne ili međusobno okomite. (d) Mjere svih unutarnjih kutova su: 45° , 90° ili 135° ; a stranice različitih vrsta zatvaraju kut od 45° ili kut od 135° (vidjeti [14, str. 597]).

Nadalje, na temelju 5. aktivnosti može se izvesti zaključak: Kako je zbroj svih unutarnjih kutova nekog konveksnog n -terokuta $K = (n - 2) \cdot 180^\circ$, a najveći kut konveksnog tangram lika je 135° , to znači da zbroj svih unutarnjih kutova ne može premašiti $n \cdot 135^\circ$, tj. vrijedi da je: $K \leq n \cdot 135^\circ$. Iz dobivene nejednakosti $(n - 2) \cdot 180^\circ \leq n \cdot 135^\circ$ slijedi da je $n \leq 8$. Drugim riječima, pomoću svih sedam tanova može se oblikovati najviše konveksni osmerokut, što je zapravo tvrdnja **Leme 3** (vidjeti [14, str. 597]).

Aktivnost 6: Konveksni tangram lik, prikazan u točkastoj kvadratnoj mreži, **upisati u pravokutnik** tako da racionalne stranice pripadaju stranicama pravokutnika. Zatim odgovoriti na sljedeće: (a) Kakvi su nastali dijelovi pravokutnika, koji prikazani tangram lik nadopunjaju do pravokutnika? (b) Kolika je **mjera površine** konveksnog tangram lika, uz pretpostavku da je tan kvadrat jedinični kvadrat?

Upisivanje promatranog pravokutnog trapeza u odgovarajući pravokutnik prikazano je na slici 11. Pozornim opažanjem uočava se da su unutarnji kutovi pravokutnog trapeza 45° , 90° i 135° pa su šiljati kutovi nastalih trokuta 45° , a treći kut je pravi jer se radi o kutu pravokutnika. Dakle, konveksni tangram lik do pravokutnika nadopunjaju jednakokračni pravokutni trokuti. Do tog zaključka može se doći i korištenjem svojstava kvadratne mreže.

Kako racionalne stranice, upisanog konveksnog tangram lika, pripadaju stranicama pravokutnika, iracionalne stranice tog lika su hipotenuze promatralih trokuta. Stoga, kada je iracionalna stranica konveksnog tangram lika duljine $a\sqrt{2}$, to će ujedno biti i duljina hipotenuze promatranog trokuta, a pripadne katete će biti duljine a , itd.



Slika 11. Pravokutni tangram trapez upisan u pravokutnik

Na temelju ove aktivnosti može se iskazati **Lema 4:** Konveksni tangram lik se može upisati u pravokutnik tako da racionalne stranice pripadaju stranicama pravokutnika. Analogno razmatranje se može provesti i sa iracionalnim stranicama (vidjeti [14, str. 597]).

Mjera površine promatranog konveksnog tangram lika iznosi $P = 8$ i može se odrediti na različite načine: (1) prebrojavanjem jediničnih kvadrata unutar lika smještenog u mrežu; (2) očitavanjem duljina osnovica i visine trapeza na temelju mreže te izračunavanjem preko formule; (3) dijeljenjem trapeza na trokute ili neke druge likove, računanjem mjera njihovih površina te na kraju zbrajanjem svih; (4) određivanjem mjera površina pravokutnika i promatralih trokuta te oduzimanjem površina trokuta od površine pravokutnika; (5) ili se jednostavno poslužimo prethodnim zaključkom da je površina konveksnog tangram trapeza mjere $16P$, gdje je P mjera površine malog tan trokuta. Površina svakog tangram lika (bilo da je konveksan ili nekonveksan) je mjere 8, uz pretpostavku da je tan kvadrat jedinični kvadrat.

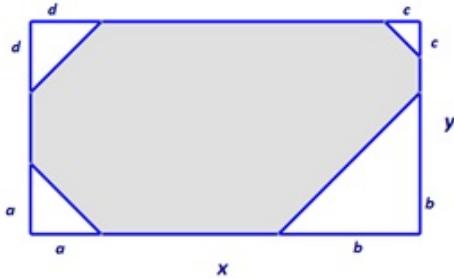
Učenici u samostalnom istraživanju odabiru različite pristupe rješavanja problema, koje je uvijek dobro međusobno usporediti i komentirati, kako je Descartes rekao: *Bolje je riješiti jedan zadatak na više različitim načinima nego više zadataka na sličan način.*

Nakon što su prethodno opisane aktivnosti u potpunosti završene, sve je spremno za postavljanje glavnog odnosa u obliku diofantske jednadžbe, na temelju koje se izvodi dokaz postavljenog teorema.

4.2. Glavna jednadžba i iskaz teorema o tangramu

Na temelju Leme 2 i Leme 4, može se postavi problem: u pravokutnik treba upisati konveksni osmerokut, kao najveći potencijalni tangram lik, na prethodno opisani način. Neka su stranice pravokutnika duljine x i y , a katete pravokutnih jednakokračnih trokuta a, b, c i d (slika 12).

Aktivnost 7: Uspostaviti vezu između svih veličina istaknutih na slici 12.



Slika 12. Konveksni lik upisan u pravokutnik

Veza se može uspostaviti primjenom metode površine. S jedne strane, mjera površine pravokutnika računa se po formuli $P = xy$. S druge strane, ista mjera se može odrediti zbrajanjem mjera pojedinačnih površina: četiriju pravokutnih trokuta duljina kateta a, b, c i d redom te površine upisanog konveksnog tangram lika, koja iznosi 8. Stoga je: $P = \frac{a^2}{2} + \frac{b^2}{2} + \frac{c^2}{2} + \frac{d^2}{2} + 8$. Izjednačavanjem i sređivanjem, dobiva se:

$$a^2 + b^2 + c^2 + d^2 = 2xy - 16, \quad (1)$$

pri čemu su $x, y \in \mathbb{N}$ i $a, b, c, d \in \mathbb{N}_0$.

Nadalje, za duljine stranica pravokutnika x i y te duljine kateta a, b, c i d promatranih jednakokračnih trokuta

$$\begin{aligned} a + b &\leq x \\ c + d &\leq x \\ a + d &\leq y \\ b + c &\leq y. \end{aligned} \quad (2)$$

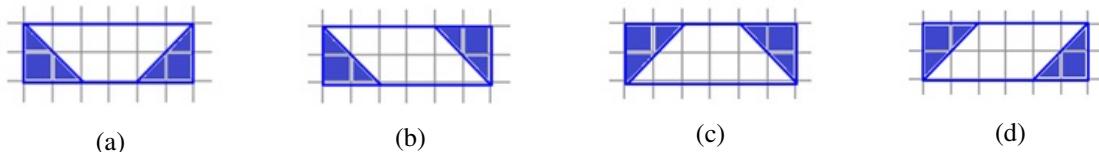
vrijedi

Analitičkim rješavanjem postavljene kvadratne diofantske jednadžbe (1) uz uvjete (2) dobiva se 48 rješenja, uz pretpostavku da je $x \geq y$. Kada se isključe sukladni likovi preostaje 20 različitih konveksnih likova. Među tih 20 likova, 7 je onih koji se ne mogu popločati svim tanovima pa njihovim isključivanjem preostaje samo 13 likova i to su svi mogući konveksni tangram likovi pa **Teorem o tangramu** glasi: **Pomoću tangram slagalice može se oblikovati točno 13 konveksnih mnogokuta.**

Analitički dokaz dan je u [14] ali samo u kratkim crtama vjerojatno zbog ograničenog prostora pisanja, a kompletan izvod nije trivijalan i zahtjeva određena matematička znanja i vještine operiranja s nejednakostima. U tom obliku imalo bi ga smisla raditi sa starijim i vještijim srednjoškolcima. Kako je cilj ove sadržaje približiti i učenicima osnovne škole, u ovom radu se predstavlja primjena vizualno - analitičke metode, kojom se također mogu odrediti sva rješenja jednadžbe (1) uz uvjete (2).

4.3. Određivanje konveksnih tangram likova primjenom vizualno - analitičke metode

Vizualno-analitička metoda sastoji se u sljedećem: od pravokutnika površine $x \cdot y$, koji se crta u kvadratnoj mreži, režu se njegovi vrhovi u obliku jednakokračnih pravokutnih trokuta duljina kateta a, b, c i d redom, tako da površina preostalog lika bude 8. Nakon toga se rješenje jednadžbe (1) očitava sa slike i zapisuje u obliku uređene šestorke: (x, y, a, b, c, d) , a preostali lik je konveksni lik. Na primjer, za $x = 6$ i $y = 2$, površina pravokutnika je veličine 12 pa je potrebno odrezati površinu veličine 4, kako bi preostali lik bio površine 8. To se može učiniti na 4 različita načina, kako je prikazano na slici 13.



Slika 13. Rezanje pravokutnika do konveksnog lika

Sa svake slike se očitava pripadno rješenje. Tako sa slike 13(a) čitamo da je pripadno rješenje jednadžbe: $(6, 2, 2, 2, 0, 0)$, sa slike 13(b) očitava se rješenje $(6, 2, 2, 0, 2, 0)$, sa slike 13(c) očitava se rješenje $(6, 2, 0, 0, 2, 2)$ i konačno sa slike 13(d) očitava se rješenje $(6, 2, 0, 2, 0, 2)$. Prvo i treće rješenje određuju sukladne jednakokračne trapeze, a drugo i četvrti rješenje sukladne paralelograme pa od četiri rješenja izdvajamo samo dva, koja određuju različite likove. Oba su tangram likovi jer se mogu popločati sa svih sedam tanova (slika 14).



Slika 14. Jednakokračni tangram trapez i tangram paralelogram

Aktivnost 8: Crtanjem pravokutnika i *rezanjem* jednakokračnih pravokutnih trokuta, odrediti sve različite konveksne likove, površine veličine 8.

Radi efikasnosti, istraživanja se može podijeliti u dva slučaja: (a) $x = y$, (b) $x > y$. Promatranjem mjera površina odrezanih trokuta, uz kraću analizu, dolazi se do zaključka da za stranice pravokutnika u promatranim slučajevima vrijedi: (a) $3 \leq x \leq 5$, (b) $9 \geq x > y \geq 1$.

Nakon što je aktivnost 8 u potpunosti završena, imate svih **48 različitih rješenja** ($x \geq y$) diofantske jednadžbe (1) uz uvjete (2), koja određuju 48 konveksnih likova, ali među kojima ima i sukladnih (slika 13). Nakon izdvajanja podudarnih likova, preostaje samo **20 različitih konveksnih likova**. Međutim, među njima je 7 onih koji se ne mogu popločati sa svih sedam tanova (slika 15).



Slika 15. Konveksni lik koji nije tangram lik

Preostalih **13** likova su traženi **konveksni tangram likovi**. Sva rješenja su dana u tablici 1, a uz malo vježbe, pripadna skica rezanja se lako rekonstruira na temelju danih rješenja.

Aktivnost 8 zahtjeva dosta vremena i primjenu različitih vještina: vizualno - prostorne vještine, vještinu računanja, kombiniranja, prebrojavanja itd. Stoga je dobro da učenici aktivnost 8 provode samostalno za domaću zadaću i time dodatno razvijaju potrebne vještine.

Kad su konačno određeni svi konveksni tangram likovi, može se uočiti da se među njima nalazi jedan jednakokračni pravokutni trokut, šest četverokuta (kvadrat, pravokutnik, paralelogram, dva pravokutna i jedan jednakokračni trapez), dva peterokuta i četiri šesterokuta (tablica 2).

Tabela 1. Sva rješenja diofantske jednadžbe (1) za $x \geq y$

Lik	x	y	a	b	c	d	Napomena
1	8	1	0	0	0	0	Nije tangram lik
2	9	1	1	1	0	0	Nije tangram lik
3	9	1	0	0	1	1	$3 \cong 2$ ²
4	9	1	1	0	1	0	Nije tangram lik
5	9	1	0	1	0	1	$5 \cong 4$
6	4	2	0	0	0	0	Tangram lik
7	4	3	2	2	0	0	Tangram lik
8	4	3	0	0	2	2	$8 \cong 9$
9	4	3	2	0	2	0	Tangram lik
10	4	3	0	2	0	2	$10 \cong 9$
11	5	2	1	1	1	1	Tangram lik
12	5	2	2	0	0	0	Tangram lik
13	5	2	0	2	0	0	$13 \cong 12$
14	5	2	0	0	2	0	$14 \cong 12$
15	5	2	0	0	0	2	$15 \cong 12$
16	5	3	0	1	2	3	Tangram lik
17	5	3	2	3	0	1	$17 \cong 16$
18	5	3	1	0	3	2	$18 \cong 16$
19	5	3	3	2	1	0	$19 \cong 16$
20	5	3	1	3	0	2	Tangram lik
21	5	3	0	2	1	3	$21 \cong 20$
22	5	3	3	1	2	0	$22 \cong 20$
23	5	3	0	1	2	3	$23 \cong 20$
24	6	2	2	2	0	0	Tangram lik
25	6	2	0	0	2	2	$25 \cong 24$
26	6	2	2	0	2	0	Tangram lik
27	6	2	0	2	0	2	$27 \cong 26$
28	6	4	4	0	4	0	Nije tangram lik
29	6	4	0	4	0	4	$29 \cong 28$
30	9	8	8	0	8	0	Nije tangram lik
31	9	8	0	8	0	8	$31 \cong 30$
32	3	3	1	1	0	0	Tangram lik
33	3	3	0	1	1	0	$33 \cong 32$
34	3	3	0	0	1	1	$34 \cong 32$
35	3	3	1	0	0	1	$35 \cong 32$
36	3	3	1	0	1	0	Tangram lik
37	3	3	0	1	0	1	$37 \cong 36$
38	4	4	4	0	0	0	Tangram lik
39	4	4	0	4	0	0	$39 \cong 38$
40	4	4	0	0	4	0	$40 \cong 38$
41	4	4	0	0	0	4	$41 \cong 38$
42	4	4	2	2	2	2	Tangram lik
43	5	5	5	0	3	0	Nije tangram lik
44	5	5	0	5	0	3	$44 \cong 43$
45	5	5	3	0	5	0	$45 \cong 43$
46	5	5	0	3	0	5	$46 \cong 43$
47	5	5	4	1	4	1	Nije tangram lik
48	5	5	1	4	1	4	$48 \cong 47$
							48 različitih rješenja
							13 tangrama

²Skraćeni zapis za iskaz da je lik pod brojem 3 sukladan liku pod brojem 2.

Tabela 2. Svi konveksni tangram likovi

Trokut	Četverokuti	Peterokuti	Šesterokuti
			
			
			
			
			
			

5. Daljnja istraživanja tangram likova

Kada se odrede svi konveksni tangram likovi, istraživanje o njima se može nastaviti. Mogu se istražiti svi mogući načini popločavanja jednog te istog lika. Na primjer, tangram trokut se može popločati na dva bitno različita načina, dok se ostali načini popločavanja razlikuju do na određenu izometriju (rotaciju, simetriju). Tangram kvadrat se može popločati samo na jedan način, a pravokutni tangram trapez na više od dva načina, itd.

Nadalje, s obzirom da svi konveksni tangram likovi imaju jednake površine, mogu se istražiti odnosi njihovih opsega i na temelju toga izvesti odgovarajuće tvrdnje. Na primjer, postoje tangram likovi jednakog i opsega i površine, a da nisu sukladni.

Osim konveksnih tangram likova, mogu se proučavati još četiri posebne vrste tangram likova: tangram zvezde, čvrsti tangrami, tangram uvale i tangram prozori, ali i odnosi među posebnim vrstama. Više o tome može se pročitati u [3].

Zatim, mogu se proučavati tangram paradoksi - prividno različiti tangram likovi od kojih je jedan podskup drugog, iako su im površine jednake (vidjeti [4, str. 21]). Itd.

6. Zaključak

Različita obrazovna istraživanja potvrđuju da učenje matematike, posebno geometrije, kroz tangram aktivnosti osigurava poticajno okruženje unutar kojeg učenici rado sudjeluju, a ono što uče na takav način bolje razumiju

i više cijene (vidjeti [9]). Jer, kada učenici samostalno provode aktivnosti, onda o tome slobodnije raspravljuju, postaju odgovorniji za svoj rad, stvaraju osjećaj izgradnje vlastitog znanja itd. čime se povećava njihovo samopouzdanje, a strah od matematike smanjuje. Aktivnim sudjelovanjem učenici stvaraju pozitivan stav i veći interes da se time i dalje bave, za razliku od situacija u kojima puko reproduciraju bez razumijevanja ono što slušaju od svojih nastavnika, te brzo odustaju od učenja, a naučeno zaboravljaju. Opisane tangram aktivnosti u ovom radu provjereno pružaju mogućnost otkrivanja, istraživanja, raspravljanja, postavljanja zakonitosti . . . , a sve to doprinosi postupnom razvoju mišljenja, posebno geometrijskog i to barem do razine neformalne dedukcije prema van Hiele-u. Kako bi nastavnici provodili opisane tangram aktivnosti, najprije sami trebaju steći iskustvo učenja geometrije (matematike) na ovaj način i poznavati van Hiele-ove razine mišljenja, a tek onda tome mogu učinkovito poučavati i svoje učenike. Misao bih zaključila riječima van Hiele-a: *Ako netko od učenika očekuje nešto više od sposobnosti ropskog ponavljanja onoga što nastavnik tvrdi ili se nada da će učenik zauzeti kritički stav prema tim tvrđenjima i htjeti raditi na njihovom poboljšanju, on mora poučavati na nov, neuobičajen način.* (vidjeti [12, str. 113]).

Bibliografija

- [1] **Baranović, N. (2015).** O razvoju geometrijskog mišljenja u nastavi matematike prema van Hieleovoj teoriji, Simpozijum Matematika i primene, Matematički fakultet, Univerzitet u Beogradu, 2015, Vol. V(1), 100-109.
- [2] **Fisher, J.** Geometric thinking concept map, Ministry of Education, Wellington, New Zealand C, dostupno na <https://arbs.nzcer.org.nz/conceptual-maps#top> (Kolovoz 2016).
- [3] **Graber, R.B., Pollard, S. & Read, R. C. (2016).** Star Tangrams. Recreational Mathematics Magazine, Number 5, pp. 47-60.
- [4] **Kavajin, A. (2016).** Tangram i njegova primjena. Diplomski rad. Filozofski fakultet u Splitu.
- [5] **Khairiree, K. (2015).** Creative Thinking in Mathematics with Tangrams and The Geometer's Sketchpad. Proceedings of the 20th Asian Technology Conference in Mathematics. Leshan, China.
- [6] Legenda o tangram slagalici, dostupno na: <http://www.tangram-channel.com/legend-of-the-tangram/> (Svibanj 2016).
- [7] **Olkun, S., Sinoplu, N. B. & Deryakulu, D. (2005).** Geometric Explorations with Dynamic Geometry Applications based on van Hiele Levels. International Journal for Mathematics Teaching and Learning. Dostupno na: <http://www.ex.ac.uk/cimt/ijmtl/ijmenu.htm> (Lipanj 2010).
- [8] **Siew, N. M. & Abdullah, S. (2012).** Learning Geometry in a Large - Enrollment Class: Do Tangrams Help in Developing Students' Geometric Thinking?. British Journal of Education, Society & Behavioural Science. Vol. 2(3): 239-259.
- [9] **Siew, N. M., Chong, C. L. & Abdullah, M. R. (2013).** Facilitating student's geometric thinking through van Hiele's phase-based learning using tangram. Journal of Social Sciences. 9 (3): 101-111.
- [10] **Tchoshanov, M. (2011).** Building Students' Mathematical Proficiency: Connecting Mathematical Ideas Using the Tangram. Learning and Teaching Mathematics, 10, 16-23.
- [11] The super source Tangrams. Grades 5-6. ETA/Cuisenaire. Vernon Hills, IL, 2007.
- [12] **Van Hiele, P. M. (1986).** Structure and Insight: A Theory of Mathematics Education. London: Academic Press, Inc.
- [13] **Van Hiele, P.M. (1999).** Developing geometric thinking through activities that begin with play. Teaching children mathematic, 5(6), 310-316.
- [14] **Wang F. T. & Hsiung, C. C. (1942).** A Theorem on the Tangram. The American Mathematical Monthly, Vol. 49, No. 9, pp. 596-599.

Realizacija Mupad procedure za određivanje Moore–Penrose-ovog inverza matrice

Ivana Jovović

Elektrotehnički fakultet, Univerzitet u Beogradu, Bulevar kralja Aleksandra 73, 11000 Beograd, Srbija
e-mail: ivana@efn.rs

Apstrakt. U ovom radu data je jedna Mupad procedura za nalaženje Moore-Penrose-ovog inverza proizvoljne matrice nad poljem realnih brojeva. Svaka matrica nad poljem realnih brojeva ima jedinstven Moore-Penrose-ov inverz. Ideja ovog rada je da se prvo odredi normalna forma date matrice, kao i matrice elementarnih transformacija vrsta i kolona u postupku određivanja normalne forme. Zatim, se određuje opšti $\{1\}$ -inverz polazne matrice u blokovskoj reprezentaciji. Koristeći opšti oblik $\{1\}$ -inverza računaju se $\{2\}$, $\{3\}$ i $\{4\}$ -inverz. Na kraju dobijamo i Moore-Penrose-ov inverz polazne matrice. Vredi napomenuti da u programskom alatu Mupad postoji gotova naredba za računanje Moore-Penrose-ovog inverza, ali ne i za računanje $\{1\}$, $\{2\}$, $\{3\}$ i $\{4\}$ -inverza.

Ključne reči: Mupad; Moore-Penrose-ov inverz matrice; opšti $\{1\}$ -inverz matrice; normalna forma matrice.

1. Uvod

Neka je $A \in \mathbb{R}^{n \times n}$ kvadratna matrica reda n nad poljem realnih brojeva. Za matricu A kažemo da je regularna ako ima inverz, tj. ako postoji matrica $A^{-1} \in \mathbb{R}^{n \times n}$ takva da je $A \cdot A^{-1} = A^{-1} \cdot A = I_n$, gde je $I_n \in \mathbb{R}^{n \times n}$ jedinična matrica reda n . Singularne matrice ili one matrice koje nisu kvadratne nemaju inverz. U raznim granama primenjene matematike javila se potreba za uopštenjem pojma inverza matrice. Pod uopštenim inverzom matrice $A \in \mathbb{R}^{m \times n}$ smatra se matrica $X \in \mathbb{R}^{n \times m}$ za koju važi da:

- se određuje i za matrice koje nisu regularne;
- ima neke osobine koje ima i inverz;
- se svodi na inverz u slučaju kada je matrica regularna.

U literaturi se najčešće sreće definicija uopštenog inverza matrice A kao skupa matrica X za koje važi $A \cdot X \cdot A = A$, pogledati knjige [1], [2], [3], [4], [6], [8].

Jedna od najpoznatijih primena teorije matrica je na sisteme linearnih algebarskih jednačina. Neka je dat sistem u matričnom obliku $A \cdot x = b$, gde je $A \in \mathbb{R}^{m \times n}$ matrica sistema, $x \in \mathbb{R}^{n \times 1}$ kolona-matrica nepoznatih i $b \in \mathbb{R}^{m \times 1}$ kolona-matrica slobodnih članova. Ako je $m = n$ i matrica A regularna, sistem ima jedinstveno rešenje $x = A^{-1} \cdot b$. U ostalim slučajevima sistem ili nema rešenje ili ih ima beskonačno, u zavisnosti od toga da li se kolona-matrica b može predstaviti kao linearna kombinacija kolona matrice A ili ne. Ako je kolona-matrica b jednak linearnoj kombinaciji kolona matrice A , onda postoji kolona-matrica $h \in \mathbb{R}^{n \times 1}$ takva da je $b = A \cdot h$ i prema Koneker-Kapeljevoj teoremi sistem $A \cdot x = b$ ima rešenje. Ako je matrica X uopšteni inverz matrice A definisan sa $A \cdot X \cdot A = A$ imamo da je $x = X \cdot b$ rešenje datog sistema. Zaista, $A \cdot x = A \cdot (X \cdot b) = (A \cdot X) \cdot b = (A \cdot X) \cdot (A \cdot h) = (A \cdot X \cdot A) \cdot h = A \cdot h = b$. Uopšteno, može se pokazati da sistem $A \cdot x = b$ ima rešenje ako i samo ako važi da je $A \cdot X \cdot b = b$, gde je X proizvoljan uopšteni inverz matrice A definisan sa $A \cdot X \cdot A = A$. U ovom slučaju rešenje datog sistema je $x = X \cdot b + (I_n - X \cdot A) \cdot y$, gde je $y \in \mathbb{R}^{n \times 1}$ proizvoljna kolona-matrica.

Ako želimo da nađemo srednje kvadratno rešenje sistema linearnih algebarskih jednačina koji nije saglasan ili rešenje koje ima najmanju normu, onda je potrebno dodati nove relacije za matricu i njen uopšteni inverz i na taj način suziti klasu uopštenih inverza. U nastavku rada mi ćemo razmatrati neke osnovne klase uopštenih inverza koje dobijamo dodajući najčešće korišćene restrikcije na uopšteni inverz. Vođeni tom logikom doći ćemo do Moore-Penrose-ovog inverza matrice koji zauzima centralno mesto u ovom radu i koji je jedinstven. Prateći izlaganje iz [7] ili [5] formiraćemo jednu proceduru u programskom alatu Mupad za nalaženje Moore-Penrose-ovog inverza matrice. Data procedura računa i neke druge tipove uopštenog inverza za datu matricu.

Rad je podeljen u dve celine. U prvom delu rada biće dat pregled uopštenih inverza, njihove definicije i osnovne osobine. U drugom delu biće prezentovana odgovarajuća Mupad procedura, sa kratkim opisom i prime-rom rada.

2. Uopšteni inverzi

Pre nego što krenemo sa detaljnim razmatranjem uopštenih inverza matrica podsetimo se pojmove teorije matrica koji će nam biti potrebni za bolje razumevanje izložene problematike. Pre svega, izlistajmo tipove elementarnih matrica i opišimo njihovu ulogu u određivanju normalne forme proizvoljne matrice.

2.1. Elementarne matrice i normalna forma matrice

Elementarna matrica I tipa je kvadratna matrica reda n koja na dijagonali ima sve elemente izuzev na poziciji (i, i) , $1 \leq i \leq n$, jednake 1, na poziciji (i, i) je element u različit od 0, a elementi van dijagonale su jednaki 0; odnosno to je matrica koja se od jedinične matrice I_n razlikuje samo u jednom elementu na dijagonali i koja je oblika

$$E(i; u) = \begin{bmatrix} 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & u & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

Primetimo da je matrica $E(i; u)$ regularna i da je njen inverz $E(i; u)^{-1} = E(i; u^{-1})$ takođe elementarna matrica I tipa. Množenje sleva matrice A matricom $E(i; u)$ kao rezultat daje matricu A kod koje je i -ta vrsta pomnožena elementom u . Množenje zdesna matrice A matricom $E(i; u)$ kao rezultat daje matricu A kod koje je i -ta kolona pomnožena elementom u . Jedinična matrica I_n je elementarna matrica I tipa. **Elementarna transformacija vrsta I tipa** matrice A je množenje jedne vrste matrice A elementom koji je različit od 0. **Elementarna transformacija kolona I tipa** matrice A je množenje jedne kolone matrice A elementom koji je različit od 0. Izvršavanje elementarne transformacije vrsta (kolona) I tipa matrice A je u stvari množenje matrice A sleva (zdesna) elementarnom matricom I tipa.

Zatim, **elementarna matrica II tipa** je matrica dobijena od jedinične matrice I_n zamenom mesta i -toj i j -toj vrsti, $1 \leq i, j \leq n$, odnosno to je matrica oblika

$$E(i, j) = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix}.$$

Datu matricu $E(i, j)$ možemo dobiti i zamenom mesta i -toj i j -toj koloni u jediničnoj matrici I_n . Matrica $E(i, j)$ je regularna i njen inverz je ona sama. Množenje sleva matrice A matricom $E(i, j)$ kao rezultat daje matricu A kod koje su i -ta i j -ta vrsta zamenile mesta. Množenje zdesna matrice A matricom $E(i, j)$ kao rezultat daje matricu A kod koje su i -ta i j -ta kolona zamenile mesta. Otuda zaključujemo da je $E(i, j)^2 = I_n$, t.j. da je matrica $E(i, j)$ sama sebi inverz. **Elementarna transformacija vrsta II tipa** matrice A je zamena mesta dvema vrstama date matrice. **Elementarna transformacija kolona II tipa** matrice A je zamena mesta dvema kolonama matrice A . Izvršavanje elementarne transformacije vrsta (kolona) II tipa matrice A je množenje matrice A sleva (zdesna) elementarnom matricom II tipa.

Elementarna matrica III tipa je kvadratna matrica reda n koja na dijagonali ima sve elemente jednake 1, na poziciji (i, j) , $i \neq j$, $1 \leq i, j \leq n$, proizvoljan element a , i sve ostale elemente jednake 0; odnosno to je matrica koja se od jedinične matrice I_n razlikuje samo u jednom elementu van dijagonale i oblika je

$$E(i, j; a) = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & \dots & a & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix}.$$

Primetimo da je matrica E regularna i da je njen inverz $E(i, j; a)^{-1} = E(i, j; -a)$ takođe elementarna matrica III tipa. Množenje sleva matrice A matricom $E(i, j; a)$ kao rezultat daje matricu A kod koje je i -ta vrsta zamenjena zbirom i -te vrste i j -te vrste pomnožene sa a . Množenje zdesna matrice A matricom $E(i, j; a)$ kao rezultat daje matricu A kod koje je j -ta kolona zamenjena zbirom j -te kolone i i -te kolone pomnožene sa a . **Elementarna transformacija vrsta III tipa** matrice A je zamena jedne vrste matrice A sa zbirom te vrste i neke druge vrste pomnožene brojem. **Elementarna transformacija III tipa** na kolonama matrice A je zamena jedne kolone matrice A sa zbirom te kolone i neke druge kolone pomnožene brojem. Izvršavanje elementarne transformacije vrsta (kolona) III tipa matrice A je u stvari množenje matrice A sleva (zdesna) elementarnom matricom III tipa.

Elementarne matrice I, II ili III tipa dobijamo primenom elementarnih transformacija vrsta (kolona) I, II ili III tipa na jediničnu matricu I_n . **Elementarna matrica** je elementarna matrica I, II ili III tipa. **Elementarna transformacija vrsta** je elementarna transformacija vrsta I, II ili III tipa. **Elementarna transformacija kolona** je elementarna transformacija kolona I, II ili III tipa.

Matrice A_1 i A_2 su **vrsta-ekvivalentne** ako se matrica A_2 može dobiti od matrice A_1 primenom konačnog broja elementarnih transformacija vrsta. Matrice A_1 i A_2 su **kolona-ekvivalentne** ako se matrica A_2 može dobiti od matrice A_1 primenom konačnog broja elementarnih transformacija kolona. Matrice A_1 i A_2 su **ekvivalentne**, u oznaci $A_1 \cong A_2$, ako se matrica A_2 može dobiti od matrice A_1 primenom konačnog broja elementarnih transformacija vrsta i kolona. Odnosno, matrice A_1 i A_2 su vrsta-ekvivalentne ako postoji konačan niz elementarnih matrica P_1, P_2, \dots, P_r takvih da važi $A_2 = P_r \cdot \dots \cdot P_2 \cdot P_1 \cdot A_1$. Matrice A_1 i A_2 su kolona-ekvivalentne ako postoji konačan niz elementarnih matrica Q_1, Q_2, \dots, Q_s takvih da važi $A_2 = A_1 \cdot Q_1 \cdot Q_2 \cdot \dots \cdot Q_s$. I na kraju, matrice A_1 i A_2 su ekvivalentne ako postoje elementarne matrice P_1, P_2, \dots, P_r i Q_1, Q_2, \dots, Q_s takve da važi $A_2 = P_r \cdot \dots \cdot P_2 \cdot P_1 \cdot A_1 \cdot Q_1 \cdot Q_2 \cdot \dots \cdot Q_s$. U daljem tekstu ćemo proizvod matrica P_1, P_2, \dots, P_r i Q_1, Q_2, \dots, Q_s označavati redom sa P i Q . Prema tome, matrice A_1 i A_2 su ekvivalentne ako postoje matrice P i Q , koje su jednake proizvodu elementarnih matrica, takve da je $A_2 = P \cdot A_1 \cdot Q$. Vrsta-ekvivalencija, kolona-ekvivalencija i ekvivalencija su relacije ekvivalencije na skupu svih matrica istog tipa. Zaista, refleksivnost sledi iz činjenice da je jedinična matrica elementarna matrica. Simetričnost sledi iz činjenice da su elementarne matrice regularne. Pokažimo tranzitivnost. Neka za matrice A_1, A_2 i A_3 važi $A_1 \cong A_2$ i $A_2 \cong A_3$. Tada postoje elementarne matrice $P_1, P_2, \dots, P_r, P'_1, P'_2, \dots, P'_{r'}$ i $Q_1, Q_2, \dots, Q_s, Q'_1, Q'_2, \dots, Q'_{s'}$ takve da važi $A_2 = P_r \cdot \dots \cdot P_2 \cdot P_1 \cdot A_1 \cdot Q_1 \cdot Q_2 \cdot \dots \cdot Q_s$ i $A_3 = P'_{r'} \cdot \dots \cdot P'_2 \cdot P'_1 \cdot A_2 \cdot Q'_1 \cdot Q'_2 \cdot \dots \cdot Q'_{s'}$. Pa imamo $A_3 = P'_{r'} \cdot \dots \cdot P'_2 \cdot P'_1 \cdot P_r \cdot \dots \cdot P_2 \cdot P_1 \cdot A_1 \cdot Q_1 \cdot Q_2 \cdot \dots \cdot Q_s \cdot Q'_1 \cdot Q'_2 \cdot \dots \cdot Q'_{s'}$, odnosno $A_1 \cong A_3$.

Neka je $A \in \mathbb{R}^{m \times n}$ proizvoljna matrica tipa $m \times n$ nad poljem realnih brojeva. Primenjujući elementarne trasformacije vrsta i kolona može se pokazati da je matrica A ekvivalentna matrici:

$$E_r = \begin{bmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} = \left[\begin{array}{c|c} I_r & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{O} \end{array} \right],$$

gde je I_r jedinična matrica reda r i r rang matrice A . Matrica E_r se naziva **normalnom formom** matrice A . Za matrice P i Q koje su proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrica A i E_r važi da je $E_r = P \cdot A \cdot Q$.

2.2. {1}-inverz ili uopšteni inverz

U ovom poglavlju ćemo se detaljno baviti uopštenim inverzima, njihovom definicijom, konstrukcijom i najvažnijim osobinama.

Opšti {1}-inverz ili uopšteni inverz matrice $A \in \mathbb{R}^{m \times n}$ je bilo koja matrica $X \in \mathbb{R}^{n \times m}$ koja zadovoljava jednačinu $A \cdot X \cdot A = A$. Svaka matrica oblika $X = Q \cdot \begin{bmatrix} I_r & X_1 \\ \hline X_2 & X_3 \end{bmatrix} \cdot P$, gde je I_r jedinična matrica reda r , r rang matrice A , X_1, X_2 i X_3 proizvoljne matrice redom tipa $r \times (m-r), (n-r) \times r$ i $(n-r) \times (m-r)$, a P i Q proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrice A i njene normalne

forme, jeste $\{1\}$ -inverz matrice A . Zaista, kako su matrice P i Q regularne važi da je $A = P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1}$, pa prema tome imamo:

$$\begin{aligned} A \cdot X \cdot A &= P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1} \cdot Q \cdot \begin{bmatrix} I_r & X_1 \\ X_2 & X_3 \end{bmatrix} \cdot P \cdot P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1} \\ &= P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} I_r & X_1 \\ X_2 & X_3 \end{bmatrix} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1} \\ &= P^{-1} \cdot \begin{bmatrix} I_r & X_1 \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1} = P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1} = A. \end{aligned}$$

Može se pokazati i obrat tvrđenja, tj. da se svaki $\{1\}$ -inverz može predstaviti u obliku proizvoda matrica Q , $\begin{bmatrix} I_r & X_1 \\ X_2 & X_3 \end{bmatrix}$ i P . U opštem $\{1\}$ -inverzu matrice A broj slobodnih parametara je $N_1 = n \cdot m - r^2$. Za konkretno

izabrane matrice X_1, X_2 i X_3 matricu $X = Q \cdot \begin{bmatrix} I_r & X_1 \\ X_2 & X_3 \end{bmatrix} \cdot P$ nazivamo **partikularnim $\{1\}$ -inverzom** matrice

A . Proizvoljan $\{1\}$ -inverz matrice A označavamo $A^{(1)}$. Primetimo da ukoliko je matrica A regularna njen jedini $\{1\}$ -inverz jeste njen inverz. Množenjem jednakosti $A \cdot X \cdot A = A$ sleve i zdesne strane sa A^{-1} dobijamo da je $X = A^{-1}$.

Naredne tri teoreme se bave konstrukcijom parcijalnih $\{1\}$ -inverza.

Teorema 1. Neka je $A \in \mathbb{R}^{m \times n}$ matrica tipa $m \times n$ nad poljem realnih brojeva ranga r . Neka je $P \in \mathbb{R}^{m \times m}$ matrica za koju važi $P \cdot A = \begin{bmatrix} I_r & C \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$. Tada je jedan $\{1\}$ -inverz matrice A dat sa $X = \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot P$.

Dokaz. Za $\{1\}$ -inverz X matrice A važi $A \cdot X \cdot A = A$. Matrica P je regularna matrica kao proizvod elementarnih matrica reda m . Prema tome, za matricu A važi $A = P^{-1} \cdot \begin{bmatrix} I_r & C \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$. Imamo:

$$A \cdot X \cdot A = P^{-1} \cdot \begin{bmatrix} I_r & C \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot P \cdot P^{-1} \cdot \begin{bmatrix} I_r & C \\ \mathbb{O} & \mathbb{O} \end{bmatrix} = P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} I_r & C \\ \mathbb{O} & \mathbb{O} \end{bmatrix} = P^{-1} \cdot \begin{bmatrix} I_r & C \\ \mathbb{O} & \mathbb{O} \end{bmatrix} = A. \quad \square$$

Teorema 2. Neka je $A \in \mathbb{R}^{m \times n}$ matrica tipa $m \times n$ nad poljem realnih brojeva ranga r . Neka je $Q \in \mathbb{R}^{n \times n}$ matrica za koju važi $A \cdot Q = \begin{bmatrix} I_r & \mathbb{O} \\ C & \mathbb{O} \end{bmatrix}$. Tada je jedan $\{1\}$ -inverz matrice A dat sa $X = Q \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$.

Dokaz. Dokaz se izvodi analogno kao i dokaz Teoreme 1. \square

Teorema 3. Neka je $A = \begin{bmatrix} A_0 & A_1 \\ A_2 & A_3 \end{bmatrix} \in \mathbb{R}^{m \times n}$ blok matrica za koju važi $\text{rang}(A) = \text{rang}(A_0) = r$, gde je $A_0 \in \mathbb{R}^{r \times r}$. Tada je jedan parcijalni $\{1\}$ -inverz matrice A dat sa $X = \begin{bmatrix} A_0^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$.

Dokaz. Vrste matrice A razmatramo kao vektore u vektorskom prostoru \mathbb{R}^n nad poljem \mathbb{R} . Kako je rang matrice A_0 jednak r , prvih r vrsta matrice A su linearne nezavisne. Rang matrice A je takođe jednak r , pa se poslednjih $m - r$ vrsta može predstaviti kao linearna kombinacija prvih r . Zaista, postoji matrica $T \in \mathbb{R}^{(m-r) \times r}$ takva da važi $[A_2 | A_3] = T \cdot [A_0 | A_1] = [T \cdot A_0 | T \cdot A_1]$. Rang matrice A_0 je jednak njenom redu, pa je matrica A_0 regularna. Kako je matrica A_0 regularna, iz $A_2 = T \cdot A_0$ zaključujemo da je $T = A_2 \cdot A_0^{-1}$. Zamenom u matričnu jednakost $A_3 = T \cdot A_1$ dobijamo $A_3 = A_2 \cdot A_0^{-1} \cdot A_1$. Matrica X je $\{1\}$ -inverz matrice A ako važi $A \cdot X \cdot A = A$. Prema tome, imamo da je:

$$\begin{aligned} A \cdot X \cdot A &= \begin{bmatrix} A_0 & A_1 \\ A_2 & A_3 \end{bmatrix} \cdot \begin{bmatrix} A_0^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} A_0 & A_1 \\ A_2 & A_3 \end{bmatrix} = \begin{bmatrix} I_r & \mathbb{O} \\ A_2 \cdot A_0^{-1} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} A_0 & A_1 \\ A_2 & A_3 \end{bmatrix} \\ &= \begin{bmatrix} A_0 & A_1 \\ A_2 \cdot A_0^{-1} \cdot A_0 & A_2 \cdot A_0^{-1} \cdot A_1 \end{bmatrix} = \begin{bmatrix} A_0 & A_1 \\ A_2 & A_3 \end{bmatrix} = A. \end{aligned} \quad \square$$

2.3. $\{1, 2\}$ -inverz ili refleksivni uopšteni inverz

Opšti $\{2\}$ -inverz matrice $A \in \mathbb{R}^{m \times n}$ je bilo koja matrica $X \in \mathbb{R}^{n \times m}$ koja zadovoljava jednačinu $X \cdot A \cdot X = X$. Svaka matrica oblika $X = Q \cdot \begin{bmatrix} X_0 & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot P$, gde su X_0, X_1 i X_2 matrice redom tipa $r \times r, r \times (m-r)$ i $(n-r) \times r$, za koje važi da je $X_0^2 = X_0$ (matrica X_0 je idempotentna), $X_0 \cdot X_1 = X_1$ i $X_2 \cdot X_0 = X_2$, a P i Q proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrice A i njene normalne forme, jeste $\{2\}$ -inverz matrice A . Zaista, kako su matrice P i Q regularne važi da je $A = P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1}$, pa prema tome imamo:

$$\begin{aligned} X \cdot A \cdot X &= Q \cdot \begin{bmatrix} X_0 & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot P \cdot P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1} \cdot Q \cdot \begin{bmatrix} X_0 & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot P \\ &= Q \cdot \begin{bmatrix} X_0 & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} X_0 & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot P = Q \cdot \begin{bmatrix} X_0 & \mathbb{O} \\ X_2 & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} X_0 & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot P \\ &= Q \cdot \begin{bmatrix} X_0^2 & X_0 \cdot X_1 \\ X_2 \cdot X_0 & X_2 \cdot X_1 \end{bmatrix} \cdot P = Q \cdot \begin{bmatrix} X_0 & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot P = X. \end{aligned}$$

Može se pokazati i obrat tvrđenja, tj. da se svaki $\{2\}$ -inverz može predstaviti u obliku proizvoda matrica $Q, \begin{bmatrix} X_0 & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix}$ i P , pri čemu važi $X_0^2 = X_0, X_0 \cdot X_1 = X_1$ i $X_2 \cdot X_0 = X_2$. U opštem $\{2\}$ -inverzu matrice A broj slobodnih parametara je $N_2 = n \cdot m - (n-r) \cdot (m-r) = r \cdot (n+m-r)$. Za konkretno izabrane matrice X_0, X_1 i X_2 matricu $Q \cdot \begin{bmatrix} X_0 & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot P$ nazivamo **partikularnim $\{2\}$ -inverzom** matrice A . Proizvoljan $\{2\}$ -inverz matrice A označavamo $A^{(2)}$. Primetimo da je matrica X $\{1\}$ -inverz matrice A ako i samo ako je A $\{2\}$ -inverz matrice X .

Opšti $\{1, 2\}$ -inverz ili refleksivni uopšteni inverz matrice $A \in \mathbb{R}^{m \times n}$ je bilo koja matrica $X \in \mathbb{R}^{n \times m}$ koja zadovoljava jednačine $A \cdot X \cdot A = A$ i $X \cdot A \cdot X = X$. Svaka matrica oblika $X = Q \cdot \begin{bmatrix} I_r & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot P$, gde su X_1 i X_2 proizvoljne matrice redom tipa $r \times (m-r)$ i $(n-r) \times r$, a P i Q proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrice A i njene normalne forme, jeste $\{1, 2\}$ -inverz matrice A . U opštem $\{1, 2\}$ -inverzu matrice A broj slobodnih parametara je $N_{1,2} = n \cdot m - r^2 - (n-r) \cdot (m-r) = r(n+m-2r)$. Za konkretno izabrane matrice X_1 i X_2 matricu $Q \cdot \begin{bmatrix} I_r & X_1 \\ X_2 & X_2 \cdot X_1 \end{bmatrix} \cdot P$ nazivamo **partikularnim $\{1, 2\}$ -inverzom** matrice A . Proizvoljan $\{1, 2\}$ -inverz matrice A označavamo $A^{(1,2)}$.

2.4. S i T matrice

Neka je $A \in \mathbb{R}^{m \times n}$ proizvoljna matrica tipa $m \times n$ nad poljem realnih brojeva i neka su P i Q proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrice A i njene normalne forme, $P \cdot A \cdot Q = E_r = \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$. Neka su S i T matrice

$$S = P \cdot P^T = \begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix} \quad \text{i} \quad T = Q^T \cdot Q = \begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix}, \quad (1)$$

gde su S_1, S_2, S_3 i S_4 matrice tipa $r \times r, r \times (m-r), (m-r) \times r$ i $(m-r) \times (m-r)$, a T_1, T_2, T_3 i T_4 matrice tipa $r \times r, r \times (n-r), (n-r) \times r$ i $(n-r) \times (n-r)$.

Tvrđenje 1. Neka je $A \in \mathbb{R}^{m \times n}$ proizvoljna matrica tipa $m \times n$ nad poljem realnih brojeva i neka su S i T matrice definisane u (1). Tada važi:

- matrice S i T su simetrične, i važi $S_1^T = S_1, S_2^T = S_3, S_4^T = S_4$ i $T_1^T = T_1, T_2^T = T_3$ i $T_4^T = T_4$;

- matrice S i T su pozitivno definitne, tj. za svaku nenula kolona-matricu $x \in \mathbb{R}^{m \times 1}$ važi da je $x^T \cdot S \cdot x > 0$ i za svaku nenula kolona-matricu $y \in \mathbb{R}^{n \times 1}$ važi da je $y^T \cdot T \cdot y > 0$;
- matrice S i T su regularne i važi da je

$$S^{-1} = \begin{bmatrix} (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & -S_1^{-1} \cdot S_2 \cdot (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \\ -S_4^{-1} \cdot S_3 \cdot (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \end{bmatrix} \quad i$$

$$T^{-1} = \begin{bmatrix} (T_1 - T_2 \cdot T_4^{-1} \cdot T_3)^{-1} & -T_1^{-1} \cdot T_2 \cdot (T_4 - T_3 \cdot T_1^{-1} \cdot T_2)^{-1} \\ -T_4^{-1} \cdot T_3 \cdot (T_1 - T_2 \cdot T_4^{-1} \cdot T_3)^{-1} & (T_4 - T_3 \cdot T_1^{-1} \cdot T_2)^{-1} \end{bmatrix}.$$

2.5. {1, 3}-inverz i {1, 4}-inverz

Opšti {3}-inverz matrice $A \in \mathbb{R}^{m \times n}$ je bilo koja matrica $X \in \mathbb{R}^{n \times m}$ koja zadovoljava jednačinu $(A \cdot X)^T = A \cdot X$. Svaka matrica oblika $X = Q \cdot \begin{bmatrix} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \\ X_2 & X_3 \end{bmatrix} \cdot P$, gde su P i Q proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrice A i njene normalne forme, S_1, S_2, S_3 i S_4 blokovi matrice $S = P \cdot P^T = \begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix}$ redom tipa $r \times r, r \times (m-r), (m-r) \times r$ i $(m-r) \times (m-r)$, a X_0, X_2 i X_3 proizvoljne matrice redom tipa $r \times r, (n-r) \times r$ i $(n-r) \times (m-r)$, za koje važi $(S_1 - S_2 \cdot S_4^{-1} \cdot S_3) \cdot X_0^T = X_0 \cdot (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)$, jeste {3}-inverz matrice A . Zaista, kako su matrice P i Q regularne važi da je $A = P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1}$, pa prema tome imamo:

$$(A \cdot X)^T = \left(P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot Q^{-1} \cdot Q \cdot \begin{bmatrix} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \\ X_2 & X_3 \end{bmatrix} \cdot P \right)^T =$$

$$\left(P^{-1} \cdot \begin{bmatrix} I_r & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \\ X_2 & X_3 \end{bmatrix} \cdot P \right)^T = \left(P^{-1} \cdot \begin{bmatrix} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot P \right)^T =$$

$$P^T \cdot \begin{bmatrix} X_0^T & \mathbb{O} \\ (-X_0 \cdot S_2 \cdot S_4^{-1})^T & \mathbb{O} \end{bmatrix} \cdot (P^{-1})^T = P^T \cdot \begin{bmatrix} X_0^T & \mathbb{O} \\ -S_4^{-1} \cdot S_3 \cdot X_0^T & \mathbb{O} \end{bmatrix} \cdot (P^T)^{-1} =$$

$$(P^{-1} \cdot P) \cdot P^T \cdot \begin{bmatrix} X_0^T & \mathbb{O} \\ -S_4^{-1} \cdot S_3 \cdot X_0^T & \mathbb{O} \end{bmatrix} \cdot (P^T)^{-1} \cdot (P^{-1} \cdot P) = P^{-1} \cdot (P \cdot P^T) \cdot \begin{bmatrix} X_0^T & \mathbb{O} \\ -S_4^{-1} \cdot S_3 \cdot X_0^T & \mathbb{O} \end{bmatrix} \cdot (P \cdot P^T)^{-1} \cdot P =$$

$$P^{-1} \cdot \begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix} \cdot \begin{bmatrix} X_0^T & \mathbb{O} \\ -S_4^{-1} \cdot S_3 \cdot X_0^T & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & -S_1^{-1} \cdot S_2 \cdot (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \\ -S_4^{-1} \cdot S_3 \cdot (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \end{bmatrix} \cdot P =$$

$$P^{-1} \cdot \begin{bmatrix} S_1 \cdot X_0^T - S_2 \cdot S_4^{-1} \cdot S_3 \cdot X_0^T & \mathbb{O} \\ S_3 \cdot X_0^T - S_4 \cdot S_4^{-1} \cdot S_3 \cdot X_0^T & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & -S_1^{-1} \cdot S_2 \cdot (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \\ -S_4^{-1} \cdot S_3 \cdot (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \end{bmatrix} \cdot P =$$

$$P^{-1} \cdot \begin{bmatrix} (S_1 - S_2 \cdot S_4^{-1} \cdot S_3) \cdot X_0^T & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & -S_1^{-1} \cdot S_2 \cdot (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \\ -S_4^{-1} \cdot S_3 \cdot (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \end{bmatrix} \cdot P =$$

$$P^{-1} \cdot \begin{bmatrix} X_0 \cdot (S_1 - S_2 \cdot S_4^{-1} \cdot S_3) & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot \begin{bmatrix} (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & -S_1^{-1} \cdot S_2 \cdot (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \\ -S_4^{-1} \cdot S_3 \cdot (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)^{-1} & (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \end{bmatrix} \cdot P =$$

$$P^{-1} \cdot \begin{bmatrix} X_0 & -X_0 \cdot (S_1 - S_2 \cdot S_4^{-1} \cdot S_3) \cdot S_1^{-1} \cdot S_2 \cdot (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot P =$$

$$P^{-1} \cdot \begin{bmatrix} X_0 & -X_0 \cdot (S_2 - S_2 \cdot S_4^{-1} \cdot S_3 \cdot S_1^{-1} \cdot S_2) \cdot (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot P =$$

$$P^{-1} \cdot \begin{bmatrix} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \cdot (S_4 - S_3 \cdot S_1^{-1} \cdot S_2) \cdot (S_4 - S_3 \cdot S_1^{-1} \cdot S_2)^{-1} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \cdot P =$$

$$P^{-1} \cdot \left[\begin{array}{c|c} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \\ \hline \mathbb{O} & \mathbb{O} \end{array} \right] \cdot P = P^{-1} \cdot \left[\begin{array}{c|c} I_r & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{O} \end{array} \right] \cdot \left[\begin{array}{c|c} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \\ \hline X_2 & X_3 \end{array} \right] \cdot P =$$

$$P^{-1} \cdot \left[\begin{array}{c|c} I_r & \mathbb{O} \\ \hline \mathbb{O} & \mathbb{O} \end{array} \right] \cdot Q^{-1} \cdot Q \cdot \left[\begin{array}{c|c} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \\ \hline X_2 & X_3 \end{array} \right] \cdot P = A \cdot X.$$

Može se pokazati i obrat tvrđenja, tj. da se svaki $\{3\}$ -inverz može predstaviti u obliku proizvoda matrica Q , $\left[\begin{array}{c|c} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \\ \hline X_2 & X_3 \end{array} \right]$ i P , pri čemu važi $(S_1 - S_2 \cdot S_4^{-1} \cdot S_3) \cdot X_0^T = X_0 \cdot (S_1 - S_2 \cdot S_4^{-1} \cdot S_3)$. U opštem $\{3\}$ -inverzu matrice A broj slobodnih parametara je $N_3 = n \cdot m - r \cdot (m - r)$. Za konkretno izabrane matrice X_0 , X_2 i X_3 matricu $Q \cdot \left[\begin{array}{c|c} X_0 & -X_0 \cdot S_2 \cdot S_4^{-1} \\ \hline X_2 & X_3 \end{array} \right] \cdot P$ nazivamo **partikularnim $\{3\}$ -inverzom** matrice A . Proizvoljan $\{3\}$ -inverz matrice A označavamo $A^{(3)}$.

Opšti $\{1, 3\}$ -inverz matrice $A \in \mathbb{R}^{m \times n}$ je bilo koja matrica $X \in \mathbb{R}^{n \times m}$ koja zadovoljava jednačine $A \cdot X \cdot A = A$ i $(A \cdot X)^T = A \cdot X$. Svaka matrica oblika $X = Q \cdot \left[\begin{array}{c|c} I_r & -S_2 \cdot S_4^{-1} \\ \hline X_2 & X_3 \end{array} \right] \cdot P$, gde su P i Q proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrice A i njene normalne forme, S_2 i S_4 blokovi matrice $S = P \cdot P^T = \left[\begin{array}{c|c} S_1 & S_2 \\ \hline S_3 & S_4 \end{array} \right]$ redom tipa $r \times (m - r)$ i $(m - r) \times (m - r)$, a X_2 i X_3 proizvoljne matrice redom tipa $(n - r) \times r$ i $(n - r) \times (m - r)$, jeste $\{1, 3\}$ -inverz matrice A . U opštem $\{1, 3\}$ -inverzu matrice A broj slobodnih parametara je $N_{1,3} = (n - r) \cdot r + (n - r) \cdot (m - r) = (n - r) \cdot m$. Za konkretno izabrane matrice X_2 i X_3 matricu $Q \cdot \left[\begin{array}{c|c} I_r & -S_2 \cdot S_4^{-1} \\ \hline X_2 & X_3 \end{array} \right] \cdot P$ nazivamo **partikularnim $\{1, 3\}$ -inverzom** matrice A . Proizvoljan $\{1, 3\}$ -inverz matrice A označavamo $A^{(1,3)}$.

Opšti $\{4\}$ -inverz matrice $A \in \mathbb{R}^{m \times n}$ je bilo koja matrica $X \in \mathbb{R}^{n \times m}$ koja zadovoljava jednačinu $(X \cdot A)^T = X \cdot A$. Svaka matrica oblika $X = Q \cdot \left[\begin{array}{c|c} X_0 & X_1 \\ \hline -T_4^{-1} \cdot T_3 \cdot X_0 & X_3 \end{array} \right] \cdot P$, gde su P i Q proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrice A i njene normalne forme, T_1, T_2, T_3 i T_4 blokovi matrice $T = Q^T \cdot Q = \left[\begin{array}{c|c} T_1 & T_2 \\ \hline T_3 & T_4 \end{array} \right]$ redom tipa $r \times r, r \times (n - r), (n - r) \times r$ i $(n - r) \times (n - r)$, a X_0, X_1 i X_3 proizvoljne matrice redom tipa $r \times r, r \times (m - r)$ i $(n - r) \times (m - r)$, za koje važi $X_0^T \cdot (T_1 - T_2 \cdot T_4^{-1} \cdot T_3) = (T_1 - T_2 \cdot T_4^{-1} \cdot T_3) \cdot X_0$, jeste $\{4\}$ -inverz matrice A . U opštem $\{4\}$ -inverzu matrice A broj slobodnih parametara je $N_4 = n \cdot m - (n - r) \cdot r$. Za konkretno izabrane matrice X_0, X_1 i X_3 matricu $Q \cdot \left[\begin{array}{c|c} X_0 & X_1 \\ \hline -T_4^{-1} \cdot T_3 \cdot X_0 & X_3 \end{array} \right] \cdot P$ nazivamo **partikularnim $\{4\}$ -inverzom** matrice A . Proizvoljan $\{4\}$ -inverz matrice A označavamo $A^{(4)}$.

Opšti $\{1, 4\}$ -inverz matrice $A \in \mathbb{R}^{m \times n}$ je bilo koja matrica $X \in \mathbb{R}^{n \times m}$ koja zadovoljava jednačine $A \cdot X \cdot A = A$ i $(X \cdot A)^T = X \cdot A$. Svaka matrica oblika $X = Q \cdot \left[\begin{array}{c|c} I_r & X_1 \\ \hline -T_4^{-1} \cdot T_3 & X_3 \end{array} \right] \cdot P$, gde su P i Q proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrice A i njene normalne forme, T_3 i T_4 blokovi matrice $T = Q^T \cdot Q = \left[\begin{array}{c|c} T_1 & T_2 \\ \hline T_3 & T_4 \end{array} \right]$ redom tipa $(n - r) \times r$ i $(n - r) \times (n - r)$, a X_1 i X_3 proizvoljne matrice redom tipa $r \times (m - r)$ i $(n - r) \times (m - r)$, jeste $\{1, 4\}$ -inverz matrice A . U opštem $\{1, 4\}$ -inverzu matrice A broj slobodnih parametara je $N_{1,4} = r \cdot (m - r) + (n - r) \cdot (m - r) = n \cdot (m - r)$. Za konkretno izabrane matrice X_1 i X_3 matricu $Q \cdot \left[\begin{array}{c|c} I_r & X_1 \\ \hline -T_4^{-1} \cdot T_3 & X_3 \end{array} \right] \cdot P$ nazivamo **partikularnim $\{1, 4\}$ -inverzom** matrice A . Proizvoljan $\{1, 4\}$ -inverz matrice A označavamo $A^{(1,4)}$.

2.6. Moore–Penrose-ov inverz

Neka je $A \in \mathbb{R}^{m \times n}$ matrica tipa $m \times n$ nad poljem realnih brojeva. Matricu $X \in \mathbb{R}^{n \times m}$ tipa $n \times m$ nad poljem realnih brojeva koja je rešenje sistema **Penrose-ovih jednačina**:

$$A \cdot X \cdot A = A, \quad X \cdot A \cdot X = X, \quad (A \cdot X)^T = A \cdot X, \quad (X \cdot A)^T = X \cdot A, \quad (2)$$

nazivamo **Moor-Penrose-ov inverz**.

Teorema 4. Neka je $A \in \mathbb{R}^{m \times n}$ matrica tipa $m \times n$ nad poljem realnih brojeva. Tada je Moore–Penrose-ov inverz matrice A jedinstven.

Dokaz. Prepostavimo da su $X \in \mathbb{R}^{n \times m}$ i $Y \in \mathbb{R}^{n \times m}$ dva Moore–Penrose-ov inverz matrice A , tj. matrice za koje važe prethodne četiri matrične jednačine (2). Imamo da je:

$$\begin{aligned} A \cdot X &= (A \cdot X)^T = X^T \cdot A^T = X^T \cdot (A \cdot Y \cdot A)^T = X^T \cdot (A^T \cdot Y^T \cdot A^T) = (X^T \cdot A^T) \cdot (Y^T \cdot A^T) \\ &= (A \cdot X)^T \cdot (A \cdot Y)^T = (A \cdot X) \cdot (A \cdot Y) = (A \cdot X \cdot A) \cdot Y = A \cdot Y \quad \text{i} \\ X \cdot A &= (X \cdot A)^T = A^T \cdot X^T = (A \cdot Y \cdot A)^T \cdot X^T = (A^T \cdot Y^T \cdot A^T) \cdot X^T = (A^T \cdot Y^T) \cdot (A^T \cdot X^T) \\ &= (Y \cdot A)^T \cdot (X \cdot A)^T = (Y \cdot A) \cdot (X \cdot A) = Y \cdot (A \cdot X \cdot A) = Y \cdot A. \end{aligned}$$

Dalje sledi da je:

$$X = X \cdot A \cdot X = (X \cdot A) \cdot X = (Y \cdot A) \cdot X = Y \cdot (A \cdot X) = Y \cdot (A \cdot Y) = Y \cdot A \cdot Y = Y,$$

odakle zaključujemo da je Moore–Penrose-ov inverz jedinstven. \square

Neka je $A \in \mathbb{R}^{m \times n}$ matrica tipa $m \times n$ nad poljem realnih brojeva. Matrica oblika

$$X = Q \cdot \left[\begin{array}{c|c} I_r & -S_2 \cdot S_4^{-1} \\ \hline -T_4^{-1} \cdot T_3 & T_4^{-1} \cdot T_3 \cdot S_2 \cdot S_4^{-1} \end{array} \right] \cdot P,$$

gde su P i Q proizvodi elementarnih matrica reda m , odnosno n , koje uspostavljaju ekvivalenciju između matrice A i njene normalne forme, S_2 i S_4 blokovi matrice $S = P \cdot P^T = \left[\begin{array}{c|c} S_1 & S_2 \\ \hline S_3 & S_4 \end{array} \right]$ redom tipa $r \times (m-r)$ i $(m-r) \times (m-r)$, a T_3 i T_4 blokovi matrice $T = Q^T \cdot Q = \left[\begin{array}{c|c} T_1 & T_2 \\ \hline T_3 & T_4 \end{array} \right]$ redom tipa $(n-r) \times r$ i $(n-r) \times (n-r)$, jeste **Moor-Penrose-ov inverz** matrice A . Na osnovu Teoreme 4 zaključujemo da Moore–Penrose-ov inverz $X = Q \cdot \left[\begin{array}{c|c} I_r & -S_2 \cdot S_4^{-1} \\ \hline -T_4^{-1} \cdot T_3 & T_4^{-1} \cdot T_3 \cdot S_2 \cdot S_4^{-1} \end{array} \right] \cdot P$ matrice A ne zavisi od izbora matrica P i Q . Navedeni postupak za računanje Moore–Penrose-ovog inverza se može naći u [7] i [5].

3. Realizacija Mupad procedure za nalaženje Moore–Penrose-ovog inverza

Prva procedura koju smo realizovali *NormalnaForma(A)* je procedura za nalaženje normalne forme $n(A)$ matrice A kao i odgovarajućih matrica P i Q koje su proizvodi elementarnih matrica koje uspostavljaju ekvivalenciju između matrica A i $n(A)$. U realizaciji date procedure prilagođena je već postojeća Mupad funkcija *linalg::hermiteForm(A,All)* iz paketa *linalg*. Procedura ima jedan ulazni parametar matricu A , a u toku realizacije ispisuju se samsa matrica, broj vrsta i kolona matrice A , kao i njen rang $r(A)$, normalna forma $n(A)$ i matrice elementarnih transformacija vrsta i kolona u postupku određivanja normalne forme P i Q . U nastavku čitalac može naći odgovarajuću proceduru.

```
NormalnaForma:=proc(A)
begin
    print(NoNL, "Matrice A jednaka je"), print(A);
    print(Unquoted, "Broj vrsta matrice A jednak je ".linalg::matdim(A)[1]." .");
```

```

print(Unquoted, "Broj kolona matrice A jednak je ".linalg::matdim(A)[2] . ".");
r(A):=linalg::rank(A):
print (Unquoted, "Rang matrice A jednak je ".r(A) . ".");
h(A):=linalg::hermiteForm(A,All)[1]:
P:=linalg::hermiteForm(A,All)[2]:
n(A):=linalg::transpose(linalg::hermiteForm(linalg::transpose(h(A)),All)[1]):
Q:=linalg::transpose(linalg::hermiteForm(linalg::transpose(h(A)),All)[2]):
for i from 1 to r(A) do
  if n(A)[i,i]=0 then
    nr:=linalg::row(n(A), i):
    n(A):=linalg::delRow(n(A), i):
    n(A):=linalg::stackMatrix(n(A),nr):
    pi:=linalg::row(P, i):
    P:=linalg::delRow(P, i):
    P:=linalg::stackMatrix(P,pi):
    nc:=linalg::col(n(A), i):
    n(A):=linalg::delCol(n(A), i):
    n(A):=linalg::concatMatrix(n(A),nc):
    qi:=linalg::col(Q, i):
    Q:=linalg::delCol(Q,i):
    Q:=linalg::concatMatrix(Q,qi):
  end_if:
end_for:
for i from 1 to r(A) do
  Q:=linalg::multCol(Q, i, 1/n(A)[i,i]):
  n(A):=linalg::multCol(n(A), i, 1/n(A)[i,i]):
end_for:
print(NoNL, "Normalna forma matrice A je matrica"), print(n(A));
print(NoNL, "Matrica elementarnih transformacija vrsta u postupku odredjivanja
normalne forme matrice A je matrica P ="), print(P);
print(NoNL, "Matrica elementarnih transformacija kolona u postupku odredjivanja
normalne forme matrice A je matrica Q ="), print(Q);
end_proc:
```

Sledeća procedura *Inverz1(A)* je procedura za određivanje $\{1\}$ -inverza ulazne matrice A . U okviru ove procedure formira se blok-matrica $e(A)$ i $\{1\}$ -inverz matrice A kao proizvod matrica $Q \cdot e(A) \cdot P$. Naredbe procedure *Inverz1(A)* se nalaze u nastavku.

```

Inverz1:=proc(A)
begin
  NormalnaForma(A):
  i(A):=linalg::submatrix(n(A), 1..r(A), 1..r(A)):
  e(A):=matrix(linalg::matdim(A)[2],linalg::matdim(A)[1],x):
  e(A):=linalg::substitute(e(A), i(A), 1, 1):
  X:=Q*e(A)*P:
  print(Unquoted, "Opsti {1}-inverz matrice A je matrica"), print(X);
  print(NoNL, "koja se dobija kao proizvod matrica Q, E i P, gde je E = "), print(e(A));
end_proc:
```

U proceduri *Inverz2(A)* formira se matrica $f(A)$ tako što se blok X_3 matrica $e(A)$ zamenjuje sa $X_2 \cdot X_1$ i konstruiše se $\{1, 2\}$ -inverz matrice A kao proizvod matrica $Q \cdot f(A) \cdot P$. Procedura sledi.

```

Inverz2:=proc(A)
begin
  Inverz1(A):
  if (linalg::matdim(A)[1]=r(A) or linalg::matdim(A)[2]=r(A))
  then
    print(Unquoted, "Opsti {1,2}-inverz matrice A jednak je opstem {1}-inverzu matrice A.");
  else
    X1:=linalg::submatrix(e(A), 1..r(A), r(A)+1..linalg::matdim(A)[1]):
    X2:=linalg::submatrix(e(A), r(A)+1..linalg::matdim(A)[2], 1..r(A)):
    X2*X1:
    f(A):=linalg::substitute(e(A), X2*X1, r(A)+1, r(A)+1):
    Y:=Q*f(A)*P:
    print(Unquoted, "Opsti {1,2}-inverz matrice A je matrica"), print(Y);
    print(NoNL, "koja se dobija kao proizvod matrica Q, E i P, gde je E = "), print(f(A));
  end_if:
end_proc:
```

Zatim prilažemo dve procedure za nalaženje matrica S i T . Kao rezultat ovih procedura ispisuju se matrice S i T i odgovarajući blokovi $S_1, S_2, S_3, S_4, T_1, T_2, T_3$ i T_4 .

```

Smatrica:=proc(A)
begin
  S:=P*linalg::transpose(P):
  print(Nonl, "S = P*P^T ="), print(S);
  S1:=linalg::submatrix(S, 1..r(A), 1..r(A)):
  S2:=linalg::submatrix(S, 1..r(A), r(A)+1..linalg::matdim(S)[2]):
  S3:=linalg::submatrix(S, r(A)+1..linalg::matdim(S)[1], 1..r(A)):
  S4:=linalg::submatrix(S, r(A)+1..linalg::matdim(S)[1], r(A)+1..linalg::matdim(S)[2]):
  print(Unquoted, "Odgovarajuci blokovi matrice S su:");
  print(Nonl, "S_1 ="), print(S1);
  print(Nonl, "S_2 ="), print(S2);
  print(Nonl, "S_3 ="), print(S3);
  print(Nonl, "S_4 ="), print(S4);
end_proc;

Tmatrica:=proc(A)
begin
  T:=linalg::transpose(Q)*Q:
  print(Nonl, "T = Q^T*Q ="), print(T);
  T1:=linalg::submatrix(T, 1..r(A), 1..r(A)):
  T2:=linalg::submatrix(T, 1..r(A), r(A)+1..linalg::matdim(T)[2]):
  T3:=linalg::submatrix(T, r(A)+1..linalg::matdim(T)[1], 1..r(A)):
  T4:=linalg::submatrix(T, r(A)+1..linalg::matdim(T)[1], r(A)+1..linalg::matdim(T)[2]):
  print(Unquoted, "Odgovarajuci blokovi matrice T su:");
  print(Nonl, "T_1 ="), print(T1);
  print(Nonl, "T_2 ="), print(T2);
  print(Nonl, "T_3 ="), print(T3);
  print(Nonl, "T_4 ="), print(T4);
end_proc;

```

U proceduri $Inverz3(A)$ formira se matrica $g(A)$ tako što se blok X_1 matrica $e(A)$ zamenjuje sa $-S_2 \cdot S_4^{-1}$ i konstruiše se $\{1, 3\}$ -inverz matrice A kao proizvod matrica $Q \cdot g(A) \cdot P$. Prilažemo proceduru.

```

Inverz3:=proc(A)
begin
  Inverz1(A);
  if linalg::matdim(A)[1]=r(A)
  then
    print(Unquoted, "Opsti {1,3}-inverz matrice A jednak je opstem {1}-inverzu matrice A.");
  else
    Smatrica(A):
    -S2*S4^(-1):
    g(A):=linalg::substitute(e(A), -S2*S4^(-1), 1, r(A)+1):
    Z:=Q*g(A)*P:
    print(Unquoted, "Opsti {1,3}-inverz matrice A je matrica"), print(Z);
    print(NoNL, "koja se dobija kao proizvod matrica Q, E i P, gde je E = "), print(g(A));
  end_if;
end_proc;

```

U proceduri $Inverz4(A)$ formira se matrica $h(A)$ tako što se blok X_2 matrica $e(A)$ zamenjuje sa $-T_4^{-1} \cdot T_3$ i konstruiše se $\{1, 4\}$ -inverz matrice A kao proizvod matrica $Q \cdot h(A) \cdot P$. Procedura je priložena u produžetku.

```

Inverz4:=proc(A)
begin
  Inverz1(A);
  if linalg::matdim(A)[2]=r(A)
  then
    print(Unquoted, "Opsti {1,4}-inverz matrice A jednak je opstem {1}-inverzu matrice A.");
  else
    Tmatrica(A):
    -T4^(-1)*T3:
    h(A):=linalg::substitute(e(A), -T4^(-1)*T3, r(A)+1, 1):
    W:=Q*h(A)*P:
    print(Unquoted, "Opsti {1,4}-inverz matrice A je matrica"), print(W);
    print(NoNL, "koja se dobija kao proizvod matrica Q, E i P, gde je E = "), print(h(A));
  end_if;
end_proc;

```

Na kraju dajemo i proceduru *MoorePenroseInverz(A)* koja se sastoji iz prethodno navedenih procedura. Na početku procedure je izvršeno grananje, u zavisnosti da li je ulazna matrica regularna, ili joj je broj vrsta, odnosno kolona, jednak rangu ili ništa od navedenog. Na završetku date procedure Moore-Penrose-ov inverz ulazne matrice A je dobijen kao proizvod matrica $Q \cdot P_{\text{om}} \cdot P$, gde je matrica P_{om} dobijena u dva koraka, prvo zamjenom bloka X_1 u $\{1,4\}$ -inverzu $h(A)$ sa $-S_2 \cdot S_4^{-1}$, a zatim zamjenom bloka X_3 u tako dobijenoj matrici sa $T_4^{-1} \cdot T_3 \cdot S_2 \cdot S_4^{-1}$. Matrice P i Q su matrice elementarnih transformacija vrsta i kolona dobijene u proceduri *NormalnaForma(A)*, a matrice S_2 , S_4 , T_3 i T_4 su blokovi-matrice dobijeni u procedurama *Smatrica(A)* i *Tmatrica(A)*.

```

MoorePenroseInverz:=proc (A)
begin
  r(A):=linalg::rank(A):
  if (linalg::matdim(A)[1]=r(A) and linalg::matdim(A)[2]=r(A))
  then
    print(NoNL, "Matrica A je regularna i Moore_Penrose-ov inverz matrice A \n
    jednak je njenom inverzu"), print(A^(-1));
  elif (linalg::matdim(A)[1]=r(A) and linalg::matdim(A)[2]<>r(A))
  then
    Inverz2(A):
    print(Unquoted, "Opsti {1,3}-inverz matrice A jednak je opstem {1,2}-inverzu matrice A.");
    Tmatrica(A):
    -T4^(-1)*T3:
    h(A):=linalg::substitute(e(A), -T4^(-1)*T3, r(A)+1, 1):
    W:=Q*h(A)*P:
    print(Unquoted, "Opsti {1,4}-inverz matrice A je matrica"), print(W);
    print(NoNL, "koja se dobija kao proizvod matrica Q, E i P, gde je E = "), print(h(A));
    print(Unquoted, "Moore--Penrose-ov inverz matrice A jednak je {1,4}-inverzu matrice A,"),
    print(W);
  elif (linalg::matdim(A)[1]<>r(A) and linalg::matdim(A)[2]=r(A))
  then
    Inverz2(A):
    Smatrica(A):
    -S2*S4^(-1):
    g(A):=linalg::substitute(e(A), -S2*S4^(-1), 1, r(A)+1):
    Z:=Q*g(A)*P:
    print(Unquoted, "Opsti {1,3}-inverz matrice A je matrica"), print(Z);
    print(NoNL, "koja se dobija kao proizvod matrica Q, E i P, gde je E = "),
    print(g(A));
    print(Unquoted, "Opsti {1,4}-inverz matrice A jednak je opstem {1,2}-inverzu matrice A.");
    print(Unquoted, "Moore--Penrose-ov inverz matrice A jednak je {1,3}-inverzu matrice A,"),
    print(Z);
  else
    Inverz2(A):
    Smatrica(A):
    -S2*S4^(-1):
    g(A):=linalg::substitute(e(A), -S2*S4^(-1), 1, r(A)+1):
    Z:=Q*g(A)*P:
    print(Unquoted, "Opsti {1,3}-inverz matrice A je matrica"), print(Z);
    print(NoNL, "koja se dobija kao proizvod matrica Q, E i P, gde je E = "), print(g(A));
    Tmatrica(A):
    -T4^(-1)*T3:
    h(A):=linalg::substitute(e(A), -T4^(-1)*T3, r(A)+1, 1):
    W:=Q*h(A)*P:
    print(Unquoted, "Opsti {1,4}-inverz matrice A je matrica"), print(W);
    print(NoNL, "koja se dobija kao proizvod matrica Q, E i P, gde je E = "), print(h(A));
    GH:=linalg::substitute(g(A), -T4^(-1)*T3, r(A)+1, 1):
    Pom:=linalg::substitute(GH, T4^(-1)*T3*S2*S4^(-1), r(A)+1, r(A)+1):
    MP:=Q*Pom*P:
    print(Unquoted, "Moore--Penrose-ov inverz matrice A je matrica"), print(MP);
    print(NoNL, "koja se dobija kao proizvod matrica Q, E i P, gde je E = "), print(Pom);
  end_if:
end_proc:
```

Navodimo i jedan primer rada procedure *MoorePenroseInverz(A)*.

```
A:=matrix([[1,1],[3,1],[1,2]]):
```

MoorePenroseInverz (A);
 Matrice A jednaka je $\begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 1 & 2 \end{pmatrix}$.

Broj vrsta matrice A jednak je 3.

Broj kolona matrice A jednak je 2.

Rang matrice A jednak je 2.

Normalna forma matrice A je matrica $\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$.

Matrica elementarnih transformacija vrsta u postupku odredjivanja normalne forme matrice A je matrica $P = \begin{pmatrix} 2 & 0 & -1 \\ -1 & 0 & 1 \\ 5 & -1 & -2 \end{pmatrix}$.

Matrica elementarnih transformacija kolona u postupku odredjivanja normalne forme matrice A je matrica $Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Opsti {1}-inverz matrice A je matrica $\begin{pmatrix} 5x(1,3) + 2 & -x(1,3) & -2x(1,3) - 1 \\ 5x(2,3) - 1 & -x(2,3) & 1 - 2x(2,3) \end{pmatrix}$,

koja se dobija kao proizvod matrica Q, E i P, gde je $E = \begin{pmatrix} 1 & 0 & x(1,3) \\ 0 & 1 & x(2,3) \end{pmatrix}$.

Opsti {1,2}-inverz matrice A jednak je opstem {1}-inverzu matrice A.

$$S = P * P^T = \begin{pmatrix} 5 & -3 & 12 \\ -3 & 2 & -7 \\ 12 & -7 & 30 \end{pmatrix}$$

Odgovarajuci blokovi matrice S su:

$$S_{-1} = \begin{pmatrix} 5 & -3 \\ -3 & 2 \end{pmatrix}$$

$$S_{-2} = \begin{pmatrix} 12 \\ -7 \end{pmatrix}$$

$$S_{-3} = (12 \quad -7)$$

$$S_{-4} = (30)$$

Opsti {1,3}-inverz matrice A je matrica $\begin{pmatrix} 0 & \frac{2}{5} & -\frac{1}{5} \\ \frac{1}{6} & -\frac{7}{30} & \frac{8}{15} \end{pmatrix}$.

koja se dobija kao proizvod matrica Q, E i P, gde je $E = \begin{pmatrix} 1 & 0 & -\frac{2}{5} \\ 0 & 1 & \frac{7}{30} \end{pmatrix}$.

Opsti {1,4}-inverz matrice A jednak je opstem {1,2}-inverzu matrice A.

Moore--Penrose-ov inverz matrice A jednak je {1,3}-inverzu matrice A, $\begin{pmatrix} 0 & \frac{2}{5} & -\frac{1}{5} \\ \frac{1}{6} & -\frac{7}{30} & \frac{8}{15} \end{pmatrix}$.

4. Zaključak

Iako u paketu *linalg* programskog alata Mupad postoji već gotova funkcija *linalg::pseudoInverse(A)* za računanje Moore–Penrose-ovog inverza matrice A, značaj realizovane procedure *MoorePenroseInverz(A)* jeste u ispisivanju svih međukoraka; normalne forme $n(A)$ matrice A, matrica P i Q koje su proizvodi elementarnih matrica koje uspostavljaju ekvivalenciju između matrica A i $n(A)$, zatim {1}, {1,2}, {1,3} i {1,4}-inverza.

Zahvalnica. I. Jovović je podržana od strane Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije, na projektu Analiza i algebra sa primenama, iz programa Osnovna istraživanja, broj projekta 174032.

Bibliografija

- [1] **R.B. Bapat.** Linear Algebra and Linear Models. Springer–Verlag London, 2012.
- [2] **A. Ben-Israel, T.N.E. Greville.** Generalized Inverses: Theory and Applications. Springer–Verlag New York, 2003.
- [3] **A. Björhammar.** A Generalized Matrix Algebra. Transaction Royal Institute of Technology, Stockholm, 1958, 124.
- [4] **S.L. Campbell, C.D. Meyer.** Generalized Inverses of Linear Transformations. SIAM, 2009, 56.
- [5] **B. Malešević.** Grupna funkcionalna jednačina (Magistarska teza). Univerzitet u Beogradu, 1998.
- [6] **N. Ravishanker, D.K. Dey.** A First Course in Linear Model Theory. Chapman & Hall / CRC Press, 2001.
- [7] **C.A. Rhode.** Contributions to the Theory, Computation and Application of Generalized Inverses (PhD dissertation). North Carolina State University, 1964.
- [8] **N.S. Urquhart.** Computation of Generalized Inverse Matrices which Satisfy Specified Conditions. SIAM Review, 1968, 10, 2, pp. 216–218.



УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

СЕДМИ СИМПОЗИЈУМ „МАТЕМАТИКА И ПРИМЕНЕ“
НАЦИОНАЛНИ СКУП СА МЕЂУНАРОДНИМ УЧЕШЋЕМ

ПРОГРАМ

1. ДАН, ПЕТАК 4. НОВЕМБАР 2016.

10:00 – 11:45, Свечана сала САНУ

10:00 – 11:45
Отварање скупа:
Зоран Ракић , декан Математичког факултета Универзитета у Београду
Миодраг Матељевић , Универзитет у Београду, Математички факултет, дописни члан САНУ - Председник програмског одбора Симпозијума
„Interior estimates for Poisson type inequality and qc hyperbolic harmonic mappings”
Барбара Дриновец Дриновшек , Faculty of Mathematics and Physics, University of Ljubljana
„Holomorphic discs in complex manifolds”

I СЕКЦИЈА: МАТЕМАТИКА И ПРИМЕНЕ ДАНАС

12:10 – 12:40
Марко Радовановић , Универзитет у Београду, Математички факултет „Gröbner bases for flag manifolds”
12:40 – 13:10
Драгољуб Кечкић , Универзитет у Београду, Математички факултет Златко Лазовић , Универзитет у Београду, Математички факултет „Compact and "compact" operators on the standard Hilbert module over a W^* -algebra”
13:10 – 13:40
Миљан Кнежевић , Универзитет у Београду, Математички факултет „Резултати Семинара за Комплексну анализу Математичког факултета у периоду од 2007. до 2016. године“

14:10 – 14:40
Милош Арсеновић , Универзитет у Београду, Математички факултет „Atomic decompositions in weighted Bergman spaces of analytic functions”
14:40 – 15:10
Јелена Катић , Универзитет у Београду, Математички факултет Јована Николић , Универзитет у Београду, Математички факултет Дарко Милинковић , Универзитет у Београду, Математички факултет „Конормална Лагранжева Флорова хомологија за отворене скупове и PSS изоморфизам“
15:10 – 15:40
Маријан Марковић , Универзитет Црне Горе, Природно-математички факултет „The Khavinson conjecture”
15:40 – 16:10
Ђорђе Баракић , Математички институт САНУ, Универзитет у Београду Душан Јокановић , Факултет за производњу и менаџмент, Требиње Марина Милићевић , Факултет за производњу и менаџмент, Требиње „Варијације Штајнеровог поризма“

12:10 – 12:40	
Татјана Давидовић, Математички институт САНУ	
„Ефикасно моделирање реалних оптимизационих проблема“	
12:40 – 13:10	
Ана Анокић, Универзитет у Београду, Пољопривредни факултет	
Зорица Станимировић, Универзитет у Београду, Математички факултет	
Ђорђе Стакић, Универзитет у Београду, Математички факултет	
Татјана Давидовић, Математички институт САНУ	
„Проблем распоређивања возила при оптимизацији трошкова транспорта пољопривредних сировина“	
13:10 – 13:40	
Бојана Милошевић, Универзитет у Београду, Математички факултет	
Марко Обрадовић, Универзитет у Београду, Математички факултет	
„Тестови симетрије засновани на новим карактеризацијама“	

14:10 – 14:40	
Анђелка Ковачевић, Универзитет у Београду, Математички факултет	
Лука Ч. Поповић, Астрономска опсерваторија Београд	
Alla I. Shapovalova, Special Astrophysical Observatory of the Russian AS, Russia	
Драгана Илић, Универзитет у Београду, Математички факултет	
„Periodicity: the case study of light curves of a quasar E1821+643“	
14:40 – 15:10	
Бранко Малешевић, Универзитет у Београду, Електротехнички факултет	
„Развој једног доказивача неких класа аналитичких неједнакости“	
15:10 – 15:40	
Ivana Milović, University of Vienna, Department of Statistics	
Hannes Leeb, University of Vienna, Department of Statistics	
„Conditional means of low-dimensional projections from high-dimensional data. Explicit error bounds.“	
15:40 – 16:10	
Татјана Станковић, Електротехничка школа „Никола Тесла”, Панчево	
Јасна Бошковић, Електротехничка школа „Никола Тесла”, Панчево	
„Massword Path-креативно и безбедно“	

2. ДАН, СУБОТА 5. НОВЕМБАР 2016.

I СЕКЦИЈА: МАТЕМАТИКА И ПРИМЕНЕ ДАНАС

10:00 – 10:30 Давид Калај , Универзитет Црне Горе, Природно-математички факултет „A result for HQC mappings from the unit disk onto a Jordan domain”
10:30 – 11:00 Божидар Јовановић , Математички институт САНУ „Геометрија и динамика“
11:00 – 11:30 Зоран Петровић , Универзитет у Београду, Математички факултет Маја Рославцев , Универзитет у Београду, Математички факултет „О коначно генерисаним идеалима у полиномијалним прстенима над ненетериним валуационим доменима“
11:30 – 12:00 Горан Ђанковић , Универзитет у Београду, Математички факултет Никола Лелас , Универзитет у Београду, Математички факултет „Коваријанса између уопштених фон Манголтових и дивизорских функција у $F_q[t]$ “

12:30 – 13:00 Татјана Дошеновић , Универзитет у Новом Саду, Технолошки факултет Зоран Каделбург , Универзитет у Београду, Математички факултет Стојан Раденовић , Универзитет у Београду, Машички факултет „A note on a paper of Jachymski and Klima”
13:00 – 13:30 Драгана Недић , Универзитет у Источном Сарајеву, Саобраћајни факултет Добој Раде Лазовић , Универзитет у Београду, Факултет организационих наука Миленко Пикула , Универзитет у Источном Сарајеву, Филозофски факултет Пале „Решавање инверзног граничног задатка типа Штурм-Лиувила са хомогеним кашњењем помоћу Фредholmовог нелинеарног оператора“
13:30 – 14:00 Биљана Војводић , Министарство науке и технологије Републике Српске Миленко Пикула , Универзитет у Источном Сарајеву, Филозофски факултет Владимир Владичић , Универзитет у Источном Сарајеву, Филозофски факултет „Инверзни проблем оператора типа Штурм-Лиувила са два константна кашњења“

10:00 – 10:30	Сандра Хоцић, Универзитет у Београду, Математички факултет <i>,„Апроксимација 1D и 2D једначине субдифузије“</i>
10:30 – 11:00	Александар Ђенић, Универзитет у Београду, Математички факултет <i>,„Решавање проблема одређивања положаја аутобуских терминалата применом паралелизоване методе променљивих околина“</i>
11:00 – 11:30	Јелена Станојевић, Универзитет у Београду, Економски факултет Катарина Кукић, Универзитет у Београду, Саобраћајни факултет <i>,„Од Фиbonачија до хаоса. Преглед економских модела у којима се појављује хаос“</i>
11:30 – 12:00	Сеад Решић, ПМФ Тузла Јусуф Мемић, Основна школа Сапна Елвир Чајић, Европски универзитет Калос, Тузла <i>,„Примена Banachovog teorema o fiksnoj tački u kompresiji slike“</i>

II СЕКЦИЈА: МАТЕМАТИКА И ИНФОРМАТИКА У ОБРАЗОВАЊУ

10:00 – 11:00
Миодраг Матељевић , Универзитет у Београду, Математички факултет „Комплексни бројеви, еуклидска и нееуклидска геометрија”
11:00 – 11:20
Зоран Петровић , Универзитет у Београду, Математички факултет „О дељивости у прстену полинома са више неодређених”
11:20– 11:40
Оливера Михић , Универзитет у Београду, Факултет организационих наука „Фибоначијев низ и златни пресек - свуда око нас”
11:40 – 12:00
Синиша Јешић , Универзитет у Београду, Електротехнички факултет „Вишекритеријумски алтернативни тестови у математици”

12:30 – 13:00
Владимир Балтић , Државни Универзитет у Новом Пазару; Математичка гимназија „Дељивост бројева у основној и средњој школи”
13:00 – 13:30
Ивана Јововић , Универзитет у Београду, Електротехнички факултет „Реализација Mirad процедуре за одређивање Moore-Penrose-овог инверза матрице”
13:30 – 13:50
Анђелка Симић , Гимназија „Бранислав Петронијевић“ Уб „Спој математике, уметности и информатике”
13:50 – 14:10
Наталија Будински , Основна и средња школа са домом ученика „Петро Кузмјак”, Руски Крстур Драгица Милинковић , Универзитет у Источном Сарајеву, Педагошки факултет Бијељина „Увођење појма функције кроз моделовање проблема у реалном контексту”

14:40 – 15:00

Снежана Тошовић, Основна школа „Танаско Рајић“ Чачак

„Математика у графитима као мотивација за осмаке“

15:00 – 15:20

Светлана Албијанић, ЈП Службени гласник, Београд

„Приказ монографије Педесет година Математичке гимназије“

15:30 – 16:30

Нивес Барановић, Филозофски факултет у Сплиту, Свеучилиште у Сплиту

„Развој геометријског мишљења кроз танграм активности (РАДИОНИЦА)“

III СЕКЦИЈА: НАУЧНОИСТРАЖИВАЧКИ И СТРУЧНИ РАД СТУДЕНТА

10:00 – 10:20
Божидар Радивојевић , Microsoft Development Center Serbia <i>„Радити у MDCS-у: искуства бившег студента!”</i>
10:20 – 10:40
Немања Јурић , Astria/Fincore, Београд Мирослав Марић , Универзитет у Београду, Математички факултет <i>„eШкола веба”</i>
10:40 – 11:00
Младен Видић , Саобраћајни факултет Добој, Универзитет у Источном Сарајеву <i>„Вишекритеријумско вредновање система за управљање базама података методом АХП”</i>
11:00 – 11:20
Стефан Спалевић , Универзитет у Београду, Математички факултет Зорица Станимировић , Универзитет у Београду, Математички факултет <i>„Интерполяција функција више променљивих”</i>
11:20 – 11:40
Дино Спасовски , Универзитет у Београду, Математички факултет <i>„Use of Statistical Methods for Predicting and Preventing Terroristic Attacks”</i>
11:40 – 12:00
Владимир Ђошовић , Универзитет у Београду, Математички факултет, Катедра за астрономију Бранислав Вукотић , Астрономска опсерваторија Београд Милан Ђирковић , Астрономска опсерваторија Београд Немања Мартиновић , Астрономска опсерваторија Београд <i>„Ђелијски модел галактичке настањиве зоне”</i>

12:30 – 12:50	Бранко Малешевић , Универзитет у Београду, Електротехнички факултет Бојан Бањац , Универзитет у Београду, Електротехнички факултет; Универзитет у Новом Саду, Факултет техничких наука Миша Јовановић , Универзитет у Београду, Електротехнички факултет; NORDEUS Николина Марковић , Гимназија у Лазаревцу
	<i>„Тродимензионална визуелизација модификованог модела Хајгенсовог планетаријума”</i>
12:50 – 13:10	Rowaida Elmarghani , University of Belgrade, Faculty of Mathematics <i>„Quasi-regularity of harmonic maps based on Blaschke products”</i>
13:10 – 13:30	Shadia Shalandi , University of Belgrade, Faculty of Mathematics <i>„Bi-Lipschicity of Quasiconformal Harmonic Mappings in n-dimensional space with respect to k-metric”</i>
13:30 – 13:50	Вук Пајовић , Универзитет у Београду, Електротехнички факултет Милош Вујадиновић , Универзитет у Београду, Електротехнички факултет Александар Левић , Универзитет у Београду, Електротехнички факултет
	<i>„Модерни методи решавања Проблема трговачког путника”</i>
13:50 – 14:10	Ивана Милић Житник , Астрономска опсерваторија Београд <i>„Утицај сile Јарковског и резонанци у средњем кретању на кретање астероида у Главном астероидном појасу”</i>