Möbiusove transformacije in periodični verižni ulomki

Seminar

Nejc Zajc
Fakulteta za matematiko in fiziko
Oddelek za matematiko

18. marec 2020

1 Uvod

Z verižnimi ulomki oblike

$$b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \ddots}},$$

se večinoma prvič srečamo pri teoriji števil, kjer so njihovi členi naravna števila. A ko to ne velja več in so členi poljubna kompleksna števila, hitro opazimo potrebo po novih pristopih. V članku si bomo ogledali pristop z Möbiusovimi transformacijami. To so funkcije oblike

$$g(z) = \frac{az+b}{cz+d}, \tag{1}$$

kjer so a, b, c in d kompleksna števila, za katera velja $ad - bc \neq 0$.

Njihov pomen pri obravnavi verižnih ulomkov opazimo, če si definiramo $s_1(z) = \frac{az+1}{z} = a + \frac{1}{z}$ in $s_2(z) = \frac{bz+1}{z} = b + \frac{1}{z}$; tedaj je namreč

$$s(z) = s_1(s_2(z)) = a + \cfrac{1}{b + \cfrac{1}{z}}$$

končen verižni ulomek in hkrati Möbiusova transformacija. Transformacije bomo natančno definirali v poglavju 3 in jih uporabili v dokazu glavnega izreka tega članka, za začetek pa si natančneje oglejmo verižne ulomke.

2 Verižni ulomki

Osredotočimo se na enostavne verižne ulomke. Enostaven verižni ulomek ima vse $a_i = 1$. Vpeljimo tudi krajši zapis, ki je v končni obliki enak

$$[b_0, b_1, \ldots, b_n] = b_0 + \cfrac{1}{b_1 + \cfrac{1}{\cdots + \cfrac{1}{b_n}}},$$

kjer je b_0 celo število, $b_1, b_2, \ldots$ pa naravna števila; enostaven verižni ulomek je nato enak

$$[b_0, b_1, b_2, \ldots] = b_0 + \cfrac{1}{b_1 + \cfrac{1}{b_2 + \ddots}} = \lim_{n\to\infty} [b_0, b_1, \ldots, b_n]. \tag{2}$$

Limita v (2) vedno obstaja, saj približki $[b_0, b_1, \ldots, b_n]$ verižnega ulomka strogo naraščajo za sode n, so v tem primeru navzgor omejeni z $[b_0, b_1]$ in posledično konvergirajo. Za lihe n ti približki strogo padajo, so omejeni z $[b_0]$ in zato prav tako konvergirajo. Ker je razlika med zaporednima približkoma obratno sorazmerna z n^2, sta limiti enaki, posledično pa limita (2) obstaja.

Za verižni ulomek $[b_0, b_1, b_2, \ldots]$ rečemo, da je *periodičen* s periodo k, če je $b_n = b_{n+k}$ za vsa naravna števila n. To zapišemo $\overline{[b_0, \ldots, b_{k-1}]}$. Za ulomek rečemo, da je *sčasoma periodičen*, če je $b_n = b_{n+k}$ za vse dovolj velike n. Opazimo, da v primeru, ko je $[b_0, b_1, b_2, \ldots]$ periodičen s periodo k, velja $b_0 = b_k \geq 1$. Člen b_0 je torej naravno število, vrednost periodičnega verižnega ulomka pa tako vedno večja od 1.

2.1 Kvadratna iracionalna števila

Definicija 1. *Realno število x je* ***kvadratno iracionalno število*** *(kvadratni iracional), če je iracionalno število in ničla kvadratnega polinoma P s celoštevilskimi koeficienti.*

Naj bo x kvadratni iracional. Tedaj je ničla celoštevilskega kvadratnega polinoma in je zato oblike $x = \frac{a+b\sqrt{c}}{d}$, kjer so a, b, c in d cela števila, izmed katerih b, c in d ne smejo biti enaki nič, $c > 0$ pa ni popolni kvadrat. Ko vstavimo x v kvadratni celoštevilski polinom P, katerega ničla je, vidimo da je polinom do množenja s skalarjem enolično določen. Druga ničla polinoma P je *algebraična konjugirana vrednost* x, ki jo označimo z $x^* = \frac{a-b\sqrt{c}}{d}$.

Za zapis realnih števil z enostavnimi verižnimi ulomki velja, da lahko vsako racionalno število zapišemo kot končen verižni ulomek, vsakemu iracionalnemu številu pa pripada enolično določen verižni ulomek oblike (2). O verižnih ulomkih kvadratnih iracionalov lahko povemo še več, za iracionalno število $x = [b_0, b_1, b_2, \ldots]$ namreč veljata naslednji lastnosti.

Trditev 1. *Verižni ulomek* $[b_0, b_1, b_2, \ldots]$ *je sčasoma periodičen natanko tedaj, ko je* x *kvadratni iracional.*

Trditev 2. *Verižni ulomek* $[b_0, b_1, b_2, \ldots]$ *je periodičen natanko tedaj, ko je* x *kvadratni iracional, katerega algebraična konjugirana vrednost* x^* *leži na intervalu* $(-1, 0)$.

Prvo ekvivalenco sta dokazala Euler, ki je pokazal, da sčasoma periodičen ulomek predstavlja kvadratni iracional, in Lagrange, ki je dokazal obrat. Primera dokazov lahko najdemo v [5] kot dokaza trditev 176 in 177. Drugo trditev je dokazal Galois.

Osrednji namen tega članka je pokazati, kako lahko s pomočjo Möbiusovih transformacij na verižnih ulomkih dokažemo naslednji Galois-ev izrek.

Izrek (Galois-ev izrek). *Za* $x = \overline{[b_0, \ldots, b_{k-1}]}$ *velja* $\overline{[b_{k-1}, \ldots, b_0]} = -\frac{1}{x^*}$.

Za konec poglavja si oglejmo zgled, ki pokaže veljavnost Galois-evega izreka na ulomku s periodo dolžine 2.

Zgled 1. *Naj* $a,\ b \in \mathbb{N}$ *in* $\alpha = \overline{[a, b]}$. *S substitucijo dobimo* $\alpha = a + 1/(b + 1/\alpha)$. *Torej je* α *negibna točka* $s(z) = a + 1/(b + 1/z)$. *Za negibni točki funkcije* s *velja, da sta rešitvi enačbe*

$$bz^2 - abz - a = 0. \tag{3}$$

To sta torej α *in* α^*. *Ker iz Vietovih formul sledi* $\alpha\alpha^* = -\frac{a}{b} < 0$, *velja* $\alpha > 0 > \alpha^*$.
Definirajmo še $\beta = \overline{[b, a]}$. *Enak premislek nas pripelje do ugotovitve, da sta* β *in* β^* *rešitvi* $az^2 - abz - b = 0$ *ter da velja* $\beta > 0 > \beta^*$. *Če na zadnji enačbi uporabimo transformacijo* $w = -1/z$, *dobimo enačbo* (3). *Torej za rešitve enačbe* (3) *velja* $\{\alpha, \alpha^*\} = \{-1/\beta, -1/\beta^*\}$ *in zato* $\beta = -1/\alpha^*$, *kar bi nam povedal tudi Galois-ev izrek.*

3 Kompleksna ravnina

Pred dokazom izreka, si bomo v tem poglavju ogledali definicijo in lastnosti Möbiusovih transformacij in kompleksne ravnine.

3.1 Möbiusove transformacije

Ko kompleksni ravnini $\mathbb{C}$ dodamo novo točko ∞, s tem tvorimo *razširjeno kompleksno ravnino* oz. *Riemmanovo sfero*, ki jo označimo $\mathbb{C}_\infty = \mathbb{C} \cup \{\infty\}$.

Definicija 2. *Funkcija g z domeno $\mathbb{C}_\infty$ je **Möbiusova transformacija**, če jo lahko zapišemo v obliki* (1)*, kjer so a, b, c in d kompleksna števila za katera velja $ad - bc \neq 0$.*
Če je $c \neq 0$, potem v (1) *velja $g(\infty) = \frac{a}{c}$ in $g(-\frac{d}{c}) = \infty$, sicer je $g(\infty) = \infty$.*

Vsaka Möbiusova transformacija g je bijekcija $\mathbb{C}_\infty$, saj ima inverz funkcije oblike (1) enak $g^{-1}(z) = \frac{-dz+b}{cz-a}$. Vidimo, da je g^{-1} Möbiusova transformacija, kratek račun pa nam utemelji, da to velja tudi za kompozitum dveh transformacij. Množica vseh Möbiusovih transformacij je torej grupa. Pri njihovem komponiranju si lahko pomagamo z množenjem matrik, ki kot člene vsebujejo koeficiente funkcije. Da dobimo na ustreznih mestih enake koeficiente, nam utemeljita naslednji enakosti

$$\frac{a_1\dfrac{a_2z+b_2}{c_2z+d_2}+b_1}{c_1\dfrac{a_2z+b_2}{c_2z+d_2}+d_1} = \frac{(a_1a_2+b_1c_2)z+(a_1b_2+b_1d_2)}{(c_1a_2+d_1c_2)z+(c_1b_2+d_1d_2)},$$

$$\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} \cdot \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix} = \begin{bmatrix} a_1a_2+b_1c_2 & a_1b_2+b_1d_2 \\ c_1a_2+d_1c_2 & c_1b_2+d_1d_2 \end{bmatrix}.$$

Ob predstavitvi Möbiusove transformacije z matriko pripada preslikavi g matrika M s koeficienti iz g. Za M velja, da ima neničelno determinanto, torej pripada *splošni linearni grupi* $\mathrm{GL}(2,\mathbb{C})$. Isti preslikavi g pa pripadajo tudi vse matrike $N = \lambda M$, za neničelna kompleksna števila λ. To pomeni, da je grupa Möbiusovih transformacij izomorfna kvocientu obrnljivih matrik, po neničelnih večkratnikih identitete $\mathrm{GL}(2,\mathbb{C})/((\mathbb{C} \setminus \{0\})I)$. Zadnji kvocient se imenuje *projektivna linearna grupa*, označimo pa ga z $\mathrm{PGL}(2,\mathbb{C})$. Za Möbiusovo transformacijo g velja, da je natanko določena, ko poznamo njene slike treh različnih kompleksnih točk. Za poljubni trojici različnih točk lahko namreč postopoma konstruiramo transformacijo, ki slika elemente prve trojice zaporedoma v elemente druge. Ker velja tudi, da iz lastnosti, da g slika trojico $(0, 1, \infty)$ zaporedoma v $(0, 1, \infty)$, sledi, da je ta g enaka identiteti, lahko nato pokažemo tudi, da je naš izbor transformacije enoličen. Natančnejši dokaz te lastnosti, si lahko preberemo v članku [6].

Označimo *razširjeno realno os* kot $\mathbb{R}_\infty = \mathbb{R} \cup \infty$. V primeru, ko so vsi koeficienti Möbiusove transformacije g realna števila, g ohranja $\mathbb{R}_\infty$. Tedaj velja

$$\mathrm{Im}[g(z)] = \frac{(ad - bc)\mathrm{Im}[\mathrm{z}]}{|cz + d|^2},$$

kar pomeni, da g ohranja zgornjo kompleksno polravnino $\mathbb{H} = \{x + iy\,; x, y \in \mathbb{R}, y > 0\}$ natanko tedaj, ko velja $ad - bc > 0$. Primer za to je funkcija $h(z) = -\frac{1}{z}$. Nasprotno pa se v primeru, ko $ad - bc = -1$, kompleksni polravnini ravno zamenjata, kot je to pri $k(z) = \frac{1}{z}$.

Pomembna je tudi sama geometrija delovanja Möbiusovih transformacij. Ohranjanje premic namreč ni značilno le za R_∞. V Ci se vse premice podaljšajo do ∞ in iz njih nastanejo krožnice. Möbiusove transformacije tako na $\mathbb{C}_\infty$ slikajo krožnice v krožnice.

Omenimo še, kako $\mathbb{C}_\infty$ opremimo z metriko. Stereografska projekcija je znan homeomorfizem med $\mathbb{C}$ in enotsko sfero $\mathbb{S}$ brez ene točke v $\mathbb{R}^3$. To projekcijo lahko razširimo do homeomorfizma med $\mathbb{C}_\infty$ in celotno sfero $\mathbb{S}$ ter nato prenesemo Evklidsko metriko iz $\mathbb{S}$ v metriko χ na $\mathbb{C}_\infty$. Za metrični prostor $(\mathbb{C}_\infty, \chi)$ je nato vsaka Möbiusova transformacija g homeomorfizem prostora $\mathbb{C}_\infty$ samega vase.

V primeru zgornje polravnine $\mathbb{H}$ pa ob vpeljavi norme $||z|| = |z|/y$, kjer je $|z|$ absolutna vrednost kompleksnega števila z in $y = \mathrm{Im}[z]$, dobimo Poincaréjev model polravnine, ki je eden izmed standardnih modelov hiperbolične ravnine. Tu so Möbiusove transformacije, ki ohranjajo $\mathbb{H}$ ($ad - bc > 0$), ravno vse izometrije $\mathbb{H}$. Meja prostora ustreza $\mathbb{R}_\infty$.

3.2 Modularna grupa

Definicija 3. ***Modularna grupa*** Γ *je grupa vseh Möbiusovih transformacij oblike* (1) *s celoštevilskimi koeficienti a, b, c in d, za katere velja $ad - bc = 1$.*

Kot smo že omenili, elementi Γ na $\mathbb{H}$ delujejo kot izometrije hiperbolične metrike, njihovo delovanje na $\mathbb{R}_\infty$ pa je tesno povezano s teorijo verižnih ulomkov. V grupi namreč med drugim leži tudi funkcija $s(z) = a + 1/(b + 1/z)$.

Posebno zanimive so *loksodromične izometrije* $\mathbb{H}$. To so Möbiusove transformacije, ki ohranjajo $\mathbb{H}$ in imajo dve različni negibni točki. Primer takšne funkcije je $z \mapsto 2z$, katere negibni točki sta 0 in ∞. Ob njihovi obravnavi pridemo do pomembne ugotovitve glede kvadratnih iracionalov.

Trditev 3. *Realno število x je kvadratno iracionalno število natanko tedaj ko je negibna točka nekega loksodromičnega elementa g modularne grupe Γ. Tedaj je algebraična konjugirana vrednost x^* druga negibna točka g.*

Tudi te trditve ne bomo dokazovali, uporabili pa jo bomo pri dokazovanju izreka. V ta namen si oglejmo še eno pomembno lastnost. Če je g loksodromična funckija z negibnima točkama u in v, potem je ena izmed njih, recimo u, *privlačna negibna točka*, druga (v tem primeru v) pa je *odbojna negibna točka*. To pomeni da ob večkratni aplikaciji funkcije g na elementu $z \neq v$ v limiti $n \to \infty$ velja $g^n(z) = g(g(\cdots(g(z)))) \to u$. V že omenjenem primeru $z \mapsto 2z$ je ∞ privlačna negibna točka, 0 pa je odbojna. Praviloma velja, da je negibna točka w poljubne funkcije f privlačna oziroma odbojna, če velja zaporedoma $|f'(w)| < 1$ oziroma $|f'(w)| > 1$.

4 Dokaz izreka

S pridobljenim znanjem bomo v tem poglavju dokazali Galois-ev izrek. Dokaz temelji na naslednji lemi, ki posploši pomen algebraične konjugirane vrednosti števila, saj b_i v lemi niso nujno cela števila. Za lažji zapis bomo v lemi namesto komponiranja uporabljali množenje, kot na primer $s_1 s_2(z) = s_1(s_2(z))$.

Lema 1. *Za funkcije s_i, $i \in \{1, \ldots, k\}$ oblike $s : z \mapsto b + 1/z$, kjer je $b \geq 1$, ima končni kompozitum $S = s_1 \cdots s_k$ privlačno negibno točko $\zeta \in (1, \infty)$ in odbojno negibno točko $\tilde{\zeta} \in (-1, 0)$.*

Dokaz. Posebej obravnavamo primera $k = 1$ in $k = 2$.
Pri $k = 1$ sta negibni točki $S(z) = b_1 + 1/z$ ravno rešitvi kvadratne enačbe $z^2 - b_1 z - 1 = 0$. To sta števili

$$z_{1,2} = \frac{b_1 \pm \sqrt{b_1^2 + 4}}{2},$$

kjer v prvem primeru velja $z_1 > b_1 \geq 1$ in v drugem $z_2 \in (-1, 0)$. Pozitivna rešitev je privlačna negibna točka, ker velja, da S slika v interval $[b_1, b_1 + 1]$ in $b_1 \geq 1$.
Za $k = 2$ je obravnavana funkcija

$$S(z) = b_1 + \cfrac{1}{b_2 + \cfrac{1}{z}}.$$

Tudi tu podobno izračunamo rešitvi pripadajoče kvadratne enačbe $S(z) = z$ in vidimo, da je prva večja od 1, druga pa leži na $(-1, 0)$. Funkcija S v tem primeru slika na interval $[b_1 + 1/(b_2 + 1), b_1 + 1/b_2]$. Vidimo torej, da so vse funkcijske vrednosti $S(z) \geq b_1 + 1/(b_2 + 1) = a_1 > 1$.
Naj bo zdaj $S = s_1 \cdots s_k$, za $s_i(z) = b_i + 1/z$, $b_i \geq 1$ in $k > 2$. Poglejmo si kaj velja za velikosti odvodov. Zaradi $|s_i'(z)| = |-1/z^2| \leq 1$ za vse i, lahko po verižnem pravilu za odvajanje kompozituma sklepamo $|S'(z)| = |s_k'(w_{k-1}) \cdot s_{k-1}'(w_{k-2}) \cdots s_1'(z)| \leq |s_k'(w_{k-1})| \cdot 1$, kjer so $w_i = s_i \cdots s_1(z)$ in zadnji neenačaj velja, ker so odvodi v splošnem po velikosti največ 1. Ker je $k > 2$ velja, da je $w_{k-1} > w_2 \geq a_1$. Dobili smo torej, da je $S'(z) \leq 1/a_1 < 1$. Ker je S gladka funkcija, po Lagrangevem izreku velja $|S(z_1) - S(z_2)| = |S'(\xi)(z_1 - z_2)| \leq |z_1 - z_2|/a_1$, kjer je ξ neka točka med z_1 in z_2. To pomeni, da je S skrčitev (na polnem metričnem prostoru $\mathbb{R}$) in ima zato negibno točko ζ znotraj svoje kodomene $(1, \infty)$. Ker je odvod po velikosti manjši od 1 je ta negibna točka privlačna.
Definirajmo še $\tilde{S} = s_k \cdots s_1$. Analogni postopek nas pripelje do ugotovitve,

da ima $\tilde{S}$ privlačno negibno točko $\tilde{\zeta} \in (1, \infty)$. Naj bo $\sigma(z) = -1/z$. Tedaj velja

$$\delta s_i(z) = -\frac{1}{b + \frac{1}{z}} = \frac{1}{-\frac{1}{z} - b} = s_i^{-1}\delta(z),$$

za vse z in i. Posledično velja tudi $\delta S = \tilde{S}^{-1}\delta$. Ko v zadnjo enačbo vstavimo $\delta(\tilde{\zeta})$ vidimo, da je to negibna točka kompozituma S. Ker $\delta(\tilde{\zeta}) \in (-1, 0)$ je to od ζ različna negibna točka in je zato odbojna. □

Združimo zdaj vse v dokazu Galois-evega izreka, ki ga zaradi preglednosti še enkrat zapišimo.

Izrek (Galois-ev izrek)**.** *Za* $x = \overline{[b_0, \ldots, b_{k-1}]}$ *velja* $\overline{[b_{k-1}, \ldots, b_0]} = -\frac{1}{x^*}$.

Dokaz. Naj bodo $s_i = b_i + 1/z, \ i = 0, 1, 2, \ldots$ funkcije z $b_i \geq 1$ za vse i. Po definiciji velja

$$[b_0, b_1, b_2, \ldots] = \lim_{n \to \infty} s_0 \cdots s_n(\infty),$$

kjer smo za argument v kompozitumu vstavili ∞, saj je to limitna točka argumentov. Predpostavimo da je zaporedje $b_0, b_1, b_2, \ldots$ periodično s periodo k in označimo $S = s_0 \cdots s_{k-1}$. Privlačno negibno točko S označimo z $\zeta > 1$. Naj bo

$$K = \{\infty, s_0(\infty), s_0 s_1(\infty), \ldots, s_0 \cdots s_{k-2}(\infty)\}.$$

S pomočjo leme sklepamo, da je $K \subset (1, \infty)$ in ker leži odbojna negibna točka S na intervalu $(-1, 0)$ velja $S^n(z) \to \zeta$ za vse $z \in K$. To bi lahko ekvivalentno povedali kot $s_0 \cdots s_n(\infty) \to \zeta$ ko $n \to \infty$. Torej je

$$\overline{[b_0, \ldots, b_{k-1}]} = [b_0, \ldots, b_{k-1}] = \zeta.$$

Ker je ζ enak periodičnemu verižnemu ulomku, je kvadratni iracional in po trditvi 3 velja, da je $S \in \Gamma$. Iz te trditve sledi tudi, da je druga (odbojna) negibna točka S enaka ζ^*.

Kot v lemi zdaj obrnimo periodo zaporedja na $b_{k-1}, \ldots, b_0$ in označimo privlačno negibno točko kompozituma $\tilde{S} = s_{k-1} \cdots s_0$ z $\tilde{\zeta}$. Po enakem premisleku kot zgoraj velja

$$\overline{[b_{k-1}, \ldots, b_0]} = \tilde{\zeta}.$$

Zdaj še enkrat uporabimo lemo pri sklepu, da je tudi $-1/\tilde{\zeta}$ odbojna negibna točka S. Ker ima S natanko eno odbojno odbojno negibno točko, torej velja $-1/\tilde{\zeta} = \zeta^*$. To nas pripelje do željenega rezultata

$$\overline{[b_{k-1}, \ldots, b_0]} = -\frac{1}{\zeta^*}.$$

□

Angleško-slovenski slovar strokovnih izrazov

algebraic conjugate algebraična konjugirana vrednost
Banach fixed-point theorem Banachovo skrčitveno načelo
continued fraction verižni ulomek
eventually periodic sčasoma periodičen
extended complex plane razširjena kompleksna ravnina
extended real axis razširjena realna os
fixed point negibna točka
general linear group splošna linea grupa
loxodromic isometries loksodromične izometrije
modular group modularna grupa
Möbius map Möbiusova transformacija
periodic periodičen
projective linear group projektivna linearna grupa
quadratic irrational kvadratno iracionalno število
Riemman sphere Riemmanova sfera

Literatura

[1] A. F. Beardon, *Möbius Maps and Periodic Continued Fractions*, Mathematics Magazine **88** (2015) 272–277.

[2] Zapiski predavanj predmeta proseminar B, profesorja dr. Igorja Klepa (Univerza v Ljubljani, Fakulteta za matematiko in fiziko, študijsko leto 2018/2019).

[3] *Hyperbolic geometry*, v: Wikipedia, The Free Encyclopedia, [ogled 27. 2. 2020], dostopno na `https://en.wikipedia.org/wiki/Hyperbolic_geometry`.

[4] *Poincaré half-plane model*, v: Wikipedia, The Free Encyclopedia, [ogled 27. 2. 2020], dostopno na `https://en.wikipedia.org/wiki/Poincar%C3%A9_half-plane_model`.

[5] G. H. Hardy in E. M. Wright, *An Introduction to the Theory of Numbers, Fifth edition*, Oxford Science Pub, Slarendon Press, Oxford, 1979.

[6] R. Schwartz, *Mobius Transformations and Circles*, October 8, 2007, dostopno na `https://www.math.brown.edu/~res/MFS/handout5.pdf`.