Řetězové zlomky s předepsanou periodou

Martin Kuděj

Abstrakt. Článek se zabývá řetězovými zlomky algebraických čísel stupně 2. Jsou ukázány jejich základní vlastnosti, včetně potřebné teorie. Ta je poté použita k nalezení tvaru řetězového zlomku druhých odmocnin z přirozených čísel, které nejsou čtverce, a jejich symetrické posloupnosti $(a_1, \ldots, a_k)$. Dále, pro danou symetrickou posloupnost přirozených čísel $(a_1, \ldots, a_k)$ jsou charakterizována všechna přirozená čísla, jejichž druhé odmocniny mají řetězový zlomek právě s touto symetrickou posloupností. Tato přirozená čísla jsou popsána jako funkční hodnoty jistého kvadratického polynomu.

1. Úvod

Řetězové zlomky patří mezi základní objekty teorie čísel. Obecně slouží k aproximaci iracionálních čísel racionálními, jde o tzv. dobré aproximace. Zaměříme se na řetězové zlomky odmocnin z přirozených čísel. Z nich lze vyčíst užitečné informace, například nám umožní nalézt všechna řešení Pellovy rovnice, které se budeme věnovat v kapitole 7.

Odvodíme, že každý řetězový zlomek druhé odmocniny z přirozeného čísla N, které není čtvercem, je tvaru $\sqrt{N} = [a_0, \overline{a_1, \ldots, a_k, 2a_0}]$, kde $a_0 = \lfloor \sqrt{N} \rfloor$ a $(a_1, \ldots, a_k)$ je symetrická posloupnost přirozených čísel. C. Friesen ve svém článku [2] zkoumal, jaké symetrické posloupnosti přirozených čísel $(a_1, \ldots, a_k)$ lze tímto způsobem získat a explicitně popsal všechna přirozená čísla N, jejichž druhá odmocnina má řetězový zlomek daného tvaru. C. Friesen dále ukázal, že pro každé celé nezáporné číslo k existuje nekonečně mnoho bezčtvercových N, jejichž druhá odmocnina má řetězový zlomek se symetrickou částí periody o délce k.

Nejprve zavedeme pojem řetězového zlomku (libovolného reálného čísla) a připomeneme jeho základní vlastnosti. Ve větách 22, 25 a 26 charakterizujeme periodické a ryze periodické řetězové zlomky pomocí čísel, které těmto řetězovým zlomkům přísluší, kvůli čemuž připomeneme pojem algebraického čísla stupně 2 (nad tělesem $\mathbb{Q}$) a zavedeme pojem tzv. redukované kvadratické iracionality, přičemž tato teorie byla čerpána z učebnice C. D. Oldse [5].

Následně se již zaměříme na řetězové zlomky druhých odmocnin z přirozených čísel a ve větě 27 odvodíme jejich tvar s využitím prostředků získaných v kapitole 4. Poté reprodukujeme důkaz C. Friesena z článku [2] o popisu všech přirozených čísel N takových, že $\sqrt{N} = [a_0, \overline{a_1, \ldots, a_k, 2a_0}]$ pro danou symetrickou posloupnost přirozených čísel $(a_1, \ldots, a_k)$. Nutné a postačující podmínky na tuto posloupnost zaručující, že vhodná N vskutku existují, shrneme ve větě 33. Podotkněme, že samotný důkaz v článku [2] je relativně stručný, v tomto článku pak dojde k jeho důkladnému rozpracování a rovněž i k částečnému zjednodušení kolem odhadu (42). C. Friesen ve

Bc. Martin Kuděj, Katedra algebry, MFF UK, Praha, Sokolovská 83, 186 75 Praha 8, e-mail: mudej@centrum.cz

svém článku navíc ukázal, že existuje nekonečně mnoho bezčtvercových přirozených čísel N výše popsaného typu, otázce bezčtvercovosti se však věnovat nebudeme.

Nakonec využijeme znalost řetězového zlomku $\sqrt{N}$ k nalezení všech řešení speciální diofantické rovnice, tzv. Pellovy rovnice $x^2 - Ny^2 = B$ pro $B = \pm 1$, která popisuje všechny invertibilní prvky okruhu $\mathbb{Z}[\sqrt{N}]$.

Dalším zdrojem k teorii řetězových zlomků je [8].

2. Zavedení řetězových zlomků

V této sekci zavedeme pojem řetězového zlomku a další pojmy s ním související. Důkazy většiny tvrzení z této sekce lze nalézt v 1. kapitole skript [3].

Symbolem $\mathbb{N}_0$ budeme dále značit množinu nezáporných celých čísel.

Definice 1. Buď k $\in \mathbb{N}_0$. *Konečným řetězovým zlomkem délky k* rozumíme číslo tvaru

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_k}}}},$$

kde $\{a_i\}_{i=0}^{k}$ je posloupnost reálných čísel, přičemž pokud $i > 0$, tak $a_i > 0$. Pro konečný řetězový zlomek budeme používat značení $[a_0, a_1, \ldots, a_k]$.

Tvrzení 2 ([3], tvrzení 1.5)**.** *Nechť $a_0 \in \mathbb{Z}$ a $\{a_i\}_{i=1}^{\infty}$ je posloupnost přirozených čísel. Pak pro každé $k \in \mathbb{N}_0$ platí*

$$[a_0, a_1, \ldots, a_n] = \frac{k_{n+1}(a_0, a_1, \ldots, a_n)}{k_n(a_1, a_2, \ldots, a_n)},$$

kde $\{k_i\}_{i=-1}^{\infty}$ je posloupnost polynomů definovaných následovně:

$$\begin{aligned} k_{-1} &= 0, \\ k_0 &= 1, \\ k_i(x_1, \ldots, x_i) &= x_i k_{i-1}(x_1, \ldots, x_{i-1}) + k_{i-2}(x_1, \ldots, x_{i-2}) \quad \text{pro } i \geqq 1. \end{aligned}$$

Dále položme $p_n := k_{n+1}(a_0, a_1, \ldots, a_n)$ a $q_n := k_n(a_1, a_2, \ldots, a_n)$, tj. platí

$$\begin{aligned} p_{-1} &= 1, & q_{-1} &= 0, \\ p_0 &= a_0, & q_0 &= 1, \\ p_n &= a_n p_{n-1} + p_{n-2}, & q_n &= a_n q_{n-1} + q_{n-2}. \end{aligned}$$

Definice 3. Polynom k_i z předchozího tvrzení nazveme *i-tým řetězovým polynomem.*

Lemma 4 ([3], tvrzení 1.6)**.** *Pro každé celé $n \geqq -1$ a pro všechna $a_1, a_2, \ldots, a_n \in \mathbb{R}$ platí $k_n(a_1, a_2, \ldots, a_n) = k_n(a_n, a_{n-1}, \ldots, a_1)$.*

I pro účely tohoto článku je občas dobré pracovat obecně s řetězovými polynomy, ve velké většině případů budeme však mít fixní posloupnost přirozených čísel $(a_1, \ldots, a_n)$ a budou nás zajímat funkční hodnoty řetězových polynomů přesně v bodě určeném touto posloupností, a tak pro jednoduchost zavedeme následující značení.

Definice 5. Nechť $a_0 \in \mathbb{Z}$ a $\{a_i\}_{i=1}^{\infty}$ je posloupnost přirozených čísel. Položme $p_n := k_{n+1}(a_0, a_1, \ldots, a_n)$ a $q_n := k_n(a_1, a_2, \ldots, a_n)$, tj. platí:

$$\begin{aligned} p_{-1} &= 1, & q_{-1} &= 0, \\ p_0 &= a_0, & q_0 &= 1, \\ p_n &= a_n p_{n-1} + p_{n-2}, & q_n &= a_n q_{n-1} + q_{n-2}. \end{aligned}$$

Pokud je $a_0 \in \mathbb{N}$, pak jsou posloupnosti $\{p_i\}_{i=-1}^{\infty}$ a $\{q_i\}_{i=-1}^{\infty}$ rostoucí, a navíc platí (pro libovolné $a_0 \in \mathbb{Z}$)

$$[a_0, a_1, \ldots, a_n] = \frac{p_n}{q_n}. \tag{1}$$

Tvrzení 6 ([3], tvrzení 1.7)**.** *Pro $k \in \mathbb{N}_0$ platí $p_{k-1}q_k - p_k q_{k-1} = (-1)^k$.*

Důsledek 7. *Pokud $k \in \mathbb{N}_0$, pak (p_k, q_k), (p_k, p_{k-1}), (q_k, q_{k-1}) jsou dvojice nesoudělných čísel.*

Důkaz. Pokud jakákoliv z těchto dvojic čísel má nějakého společného dělitele d, potom díky tvrzení 6 d dělí i číslo 1, či -1. To ovšem nutně znamená, že d je rovno ± 1. □

Věta 8 ([3], tvrzení 1.4 a věta 1.9)**.** *Buď α reálné číslo. Položme $\alpha_0 := \alpha$, $a_0 := \lfloor \alpha_0 \rfloor$. Pokud $i \in \mathbb{N}$ a $a_{i-1} \neq \alpha_{i-1}$, potom rekurentně definujme*

$$\alpha_i := \frac{1}{\alpha_{i-1} - a_{i-1}}, \quad a_i := \lfloor \alpha_i \rfloor.$$

Pak nastává jedna ze dvou možností:

- *Existuje $k \in \mathbb{N}_0$ takové, že $\alpha_k = a_k$. Potom α je racionální a $\alpha = [a_0, a_1, \ldots, a_k]$.*
- *Pro každé $k \in \mathbb{N}_0$ je $a_k \neq \alpha_k$. Potom α je iracionální, existuje limita posloupnosti $\lim_{n\to\infty}[a_0, a_1, \ldots, a_n]$ a tato limita je rovna právě číslu α. Budeme též psát $\alpha = [a_0, a_1, a_2, \ldots]$. V tomto případě navíc pro každé $m \in \mathbb{N}_0$ platí*

$$[a_0, a_1, \ldots, a_{2m}] < \alpha < [a_0, a_1, \ldots, a_{2m+1}]. \tag{2}$$

Všimněme si, že v předchozí větě máme $a_1, a_2, \ldots \in \mathbb{N}$. Jelikož $a_0 = \lfloor \alpha \rfloor$, dostáváme jednoduchý odhad

$$a_0 \leqq \alpha \leqq a_0 + 1. \tag{3}$$

Definice 9. Řetězový zlomek $[a_0, \ldots, a_k]$ (pro maximální možné k) resp. $[a_0, a_1, \ldots]$ z věty 8 se nazývá *řetězovým zlomkem čísla α* nebo též *rozvojem čísla α do řetězového zlomku*.

Rozvoj každého iracionálního čísla α do řetězového zlomku je určen jednoznačně.

Definice 10. Nechť α je reálné číslo a $[a_0, a_1, a_2, \ldots]$ je jeho řetězový zlomek. Pak zlomek p_k/q_k z definice 5 nazveme *k-tým sblíženým zlomkem* čísla $[a_0, a_1, \ldots]$.

Tvrzení 11 ([3], tvrzení 1.8)**.** *Nechť $\alpha = [a_0, a_1, \ldots, a_k, \beta]$, kde $\alpha > 0$, $\beta > 1$. Pak*

$$\alpha = \frac{\beta p_k + p_{k-1}}{\beta q_k + q_{k-1}}. \tag{4}$$

Platí též obrácené tvrzení, které však ve skriptech [3] není ani vysloveno, proto je zde uvedeme i s důkazem.

Lemma 12. *Mějme $\alpha > 0$, $\beta > 1$, $a_0 \in \mathbb{N}_0$. Potom pro každé $m \in \mathbb{N}_0$ a libovolná přirozená čísla $a_1, a_2, \ldots, a_m$ rovnost $\alpha = \frac{\beta p_m + p_{m-1}}{\beta q_m + q_{m-1}}$ implikuje $\alpha = [a_0, a_1, \ldots, a_m, \beta]$.*

Důkaz. Důkaz provedeme matematickou indukcí podle m. Pro $m = 0$ máme

$$\alpha = \frac{\beta p_0 + p_{-1}}{\beta q_0 + q_{-1}} = \frac{\beta a_0 + 1}{\beta} = a_0 + \frac{1}{\beta} = [a_0, \beta].$$

Platí-li tvrzení pro $m - 1 \in \mathbb{N}_0$, ukažme platnost tvrzení pro m:

$$\alpha = \frac{\beta p_m + p_{m-1}}{\beta q_m + q_{m-1}} = \frac{p_m + \frac{p_{m-1}}{\beta}}{q_m + \frac{q_{m-1}}{\beta}} = \frac{p_{m-1}\left(a_m + \frac{1}{\beta}\right) + p_{m-2}}{q_{m-1}\left(a_m + \frac{1}{\beta}\right) + q_{m-2}}$$

a použitím indukčního předpokladu dostáváme

$$\alpha = \left[a_0, a_1, \ldots, a_{m-1}, a_m + \frac{1}{\beta}\right] = [a_0, a_1, \ldots, a_m, \beta].$$

□

Na závěr této sekce představíme periodické řetězové zlomky.

Definice 13. Nechť α je iracionální číslo a $[a_0, a_1, a_2, \ldots]$ je nekonečný řetězový zlomek čísla α. Říkáme, že tento zlomek je *periodický* s *periodou* $(a_k, \ldots, a_l)$, pokud existují $k, l \in \mathbb{N}_0$, $k \leqq l$ taková, že

$$\forall i \in \{0, 1, \ldots, l-k\}\ \forall j \in \mathbb{N} \quad a_{k+i} = a_{k+i+j(l-k+1)}.$$

Takový řetězový zlomek značíme $[a_0, a_1, \ldots, a_{k-1}, \overline{a_k, a_{k+1}, \ldots, a_l}]$. Řetězový zlomek $[a_0, a_1, \ldots, a_{k-1}, \overline{a_k, a_{k+1}, \ldots, a_l}]$ je *ryze (čistě) periodický*, pokud $k = 0$, neboli $\alpha = [\overline{a_0, a_1, \ldots, a_l}]$.

Tato definice se možná jeví na první pohled trochu komplikovaná, nicméně jde o něco velice intuitivního, neboť obdobným způsobem značíme periodický desetinný rozvoj racionálního čísla, např. $0{,}\overline{12} = 0{,}121\ 212\ 121\ 2\ldots$ Tedy pokud máme $\alpha = [\overline{a_0, a_1}]$, tak platí, že $a_0 = a_2 = a_4 = \ldots$, a podobně $a_1 = a_3 = a_5 = \ldots$

Index k z předchozí definice 13 odpovídající „začátku periody“, stejně jako index l představující „konec periody“, není jednoznačně určen. Jde o podobnou situaci jako u reálných čísel, kde platí např. $0{,}\overline{12} = 0{,}\overline{121\ 2} = 0{,}12\overline{12}$ apod. Proto budeme vždy uvažovat indexy k, l co nejmenší možné.

Příklad. Buď $\alpha = \sqrt{2}$. Potom $a_0 = 1$ a pro každé $i \in \mathbb{N}$ platí $\alpha_i = 1 + \sqrt{2}$, $a_i = 2$. Proto $\alpha = [1, \overline{2}]$.

Ne všechna iracionální čísla mají rozvoj do periodického řetězového zlomku, například $\pi = [3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, 1, 1, 2, \ldots]$. Číslo α má periodický řetězový zlomek, právě když je α algebraické číslo stupně 2 (tento pojem bude zaveden v definici 16), a navíc platí, že α má rozvoj do ryze periodického zlomku (s kladným a_0), právě když je α je tzv. *redukovaná kvadratická iracionalita* (tento pojem bude zaveden v definici 18), což dokážeme v kapitole 4.

3. Algebraická čísla stupně 2

V této sekci připomeneme některé vlastnosti iracionálních čísel, jejichž řetězové zlomky budeme později v kapitole 4 zkoumat. Půjde o algebraická čísla stupně 2, speciálně tzv. redukované kvadratické iracionality.

Definice 14. Buď N přirozené číslo. Řekneme, že N je *čtverec*, pokud je N druhou mocninou nějakého přirozeného čísla, tedy pokud existuje $k \in \mathbb{N}$, $k^2 = N$. V opačném případě říkáme, že N *není čtverec*.

Dále řekneme, že N je *bezčtvercové*, pokud není dělitelné žádným čtvercem různým od 1, neboli pro všechna $k, l \in \mathbb{N}$, $k > 1$, $N \neq k^2 l$.

Je známo, že přirozené číslo N není čtverec, právě když $\sqrt{N}$ je iracionální.

Nyní zavedeme pojem algebraického čísla stupně 2 a pojem ireducibilního polynomu, který využijeme čistě z technických důvodů. Jde o známé pojmy z algebry, čtenář se o nich může dozvědět více v učebnici [7].

Definice 15. Buď $f \in \mathbb{Q}[x]$ polynom stupně alespoň 1. Řekneme, že f je *ireducibilní* polynom, pokud pro každé dva polynomy $g, h \in \mathbb{Q}[x]$, $f = gh$ platí, že g nebo h je konstantní polynom.

Definice 16. Buď α reálné číslo. Řekneme, že α je *algebraické stupně* 2, je-li kořenem ireducibilního polynomu f stupně 2 s celočíselnými koeficienty.

Nejprve poukážeme na to, že libovolné algebraické číslo stupně 2 je nutně iracionální. Pokud by totiž bylo racionální, pak by jistě bylo kořenem nějakého polynomu stupně 1 s celočíselnými koeficienty, a tak by příslušný polynom f dle definice 16 nebyl ireducibilní. Dále si všimneme, že polynom f je jednoznačně určen až na násobek. Je více způsobů, jak toto dokázat, např. pomocí tzv. *minimálního polynomu* a jeho jednoznačnosti, z čehož plyne jednoznačnost polynomu f, ale s racionálními koeficienty. Poté stačí vynásobit koeficienty tohoto polynomu společným násobkem jmenovatelů koeficientů tohoto polynomu. Podrobnosti k minimálnímu polynomu lze nalézt v učebnici [7] na straně 135.

Předpokládejme, že α je algebraické číslo stupně 2 kořenem polynomu s celočíselnými koeficienty $ax^2 + bx + c$, $a \neq 0$, přičemž bez újmy na obecnosti $a > 0$ (jinak bychom vynásobili celý polynom -1). Použitím vzorce pro řešení kvadratické rovnice dostaneme

$$\alpha = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Číslo $b^2 - 4ac$ je přirozené a není to čtverec, neboť α je iracionální. Tudíž

$$\alpha = \frac{u \ \pm \sqrt{N}}{v} = e + f\sqrt{N}, \tag{5}$$

kde $u = -b \in \mathbb{Z}$, $v = 2a \in \mathbb{N}$, $N = b^2 - 4ac \in \mathbb{N}$ není čtverec, $e, f \in \mathbb{Q}$, $f \neq 0$.

Pokud navíc přidáme předpoklad, že N je bezčtvercové (viz definice 14), tak racionální čísla e, f i přirozené N jsou určena jednoznačně. Tím je zaručena korektnost následující definice.

Definice 17. Je-li $\alpha = e + f\sqrt{N}$, kde e, $f \in \mathbb{Q}$, $N \in \mathbb{N}$, N není čtverec, pak definujeme *sdružené číslo* (též *konjugát*) čísla α jako $\alpha' := e - f\sqrt{N}$.

Je-li takovéto $\alpha = e + f\sqrt{N}$ kořenem polynomu $ax^2 + bx + c$, pak $\alpha' = e - f\sqrt{N}$ je kořenem téhož polynomu. Pouze pro $f = 0$ platí, že α není algebraické číslo stupně 2, a právě pro tato α platí $\alpha = \alpha'$.

Pro libovolná a, b, c, kde $c \neq 0$, lze snadno dokázat vztahy

$$(a+b)' = a' + b', \quad (a-b)' = a' - b', \quad (ab)' = a'b', \quad \left(\frac{a}{c}\right)' = \frac{a'}{c'}.$$

Nyní konečně můžeme zavést slíbený pojem redukované kvadratické iracionality.

Definice 18. Buď α algebraické číslo stupně 2. Řekneme, že α je *redukovaná kvadratická iracionalita*, pokud $\alpha > 1$ a zároveň $\alpha' \in (-1, 0)$.

Příklad. Nechť N je přirozené číslo, které není čtvercem, a uvažujme $a_0 = \lfloor \sqrt{N} \rfloor$. Potom $a_0 + \sqrt{N}$ je redukovaná kvadratická iracionalita.

Redukovaná kvadratická iracionalita α je nutně algebraické číslo stupně 2, tedy jde o iracionální číslo, které je kořenem nějakého kvadratického polynomu s celočíselnými koeficienty. Zaměříme se na vyjádření $\alpha = \frac{u \pm \sqrt{N}}{v}$, kde $u \in \mathbb{Z}$, $v \in \mathbb{N}$, $N \in \mathbb{N}$ není čtverec (bez újmy na obecnosti předpokládáme $v > 0$, jinak bychom zlomek rozšířili -1). Tento zápis je v prvé řadě pouhou aplikací vzorce pro řešení kvadratické rovnice, ale též nám umožní charakterizovat redukované kvadratické iracionality pouze pomocí koeficientů u, v, čehož využijeme v důkazu věty 25. Potřebujeme ještě jednoznačnost těchto koeficientů u, v v rovnosti (5), pokud máme N fixní, neboli skutečnost, že $\frac{u_1 \pm \sqrt{N}}{v_1} = \frac{u_2 \pm \sqrt{N}}{v_2}$ implikuje $u_1 = u_2$, $v_1 = v_2$, což je velice jednoduché si rozmyslet.

Uvědomíme si ještě jednu věc: Pokud $\alpha = \frac{u - \sqrt{N}}{v}$ jako výše, tak α nemůže být redukovanou kvadratickou iracionalitou. V opačném případě by platilo $\frac{u - \sqrt{N}}{v} > 1$, $\frac{u + \sqrt{N}}{v} \in (-1, 0)$. První nerovnost nutně implikuje $u > 0$, pak ale $\frac{u + \sqrt{N}}{v} > 0$ a máme spor. Proto jsou všechny redukované kvadratické iracionality tvaru

$$\frac{u + \sqrt{N}}{v}, \text{ kde } u \in \mathbb{Z}, \ v \in \mathbb{N}, N \in \mathbb{N} \text{ není čtverec.} \tag{6}$$

Podobně lze ověřit, že pokud by koeficient v byl záporný, pak výraz $\frac{u \pm \sqrt{N}}{v}$ může být redukovanou kvadratickou iracionalitou jen v případě, kdy je v čitateli rozdíl. Přesně tento případ nastane v důkazu tvrzení 23.

Jak konkrétně je třeba volit koeficienty u, v, aby se vskutku jednalo o redukovanou kvadratickou iracionalitu, uvádí následující tvrzení, jehož důkaz je přenechán čtenáři.

Tvrzení 19. *Nechť α je algebraické číslo stupně 2 ve tvaru (6). Pak α je redukovaná kvadratická iracionalita, právě když platí následující nerovnosti:*

$$0 < u < \sqrt{N}, \tag{7}$$

$$\sqrt{N} - u < v < u + \sqrt{N}. \tag{8}$$

Právě dokázané tvrzení má následující důležitý důsledek.

Důsledek 20. *Pro každé $N \in \mathbb{N}$, které není čtvercem, existuje konečně mnoho redukovaných kvadratických iracionalit tvaru* (6).

4. Řetězové zlomky algebraických čísel stupně 2

Našim cílem je charakterizovat redukované kvadratické iracionality pomocí jejich řetězových zlomků. Dokážeme, že právě redukované kvadratické iracionality mají ryze periodický řetězový zlomek s kladným prvním členem. Tuto charakterizaci dokážeme ve větách 22 a 25. Nejprve odvodíme jednu jednoduchou vlastnost posloupností $\{p_i\}_{i=0}^{\infty}$ a $\{q_i\}_{i=0}^{\infty}$ zavedených v definici 5:

Lemma 21. *Mějme* $[a_0, a_1, \ldots, a_k] = \dfrac{p_k}{q_k}$ *jako v rovnosti* (1). *Pak platí:*

$$[a_k, a_{k-1}, \ldots, a_0] = \frac{p_k}{p_{k-1}}, \tag{9}$$

$$[a_k, a_{k-1}, \ldots, a_1] = \frac{q_k}{q_{k-1}}, \quad \text{pokud } k \in \mathbb{N}. \tag{10}$$

Důkaz. Dokážeme pouze první rovnost (9), důkaz rovnosti (10) je obdobný. Důkaz se opírá pouze o rekurentní definici $p_k, p_{k-1}, \ldots$, respektive $q_k, q_{k-1}, \ldots$:

$$\frac{p_k}{p_{k-1}} = \frac{a_k p_{k-1} + p_{k-2}}{p_{k-1}} = a_k + \frac{p_{k-2}}{p_{k-1}} = a_k + \frac{1}{\dfrac{p_{k-1}}{p_{k-2}}}.$$

Stejným způsobem pokračujeme pro p_{k-1}/p_{k-2}. Je jasné, že tímto postupem tvoříme nějaký řetězový zlomek. Postup opakujeme tak dlouho, dokud nedojdeme k rovnosti

$$\frac{p_k}{p_{k-1}} = \left[a_n, a_{n-1}, \ldots, a_1, \frac{p_0}{p_{-1}}\right].$$

Jelikož $p_0 = a_0$, $p_{-1} = 1$, opravdu jsme odvodili rovnost (9). □

Nyní již dokážeme první část slíbené charakterizace redukovaných kvadratických iracionalit pomocí jejich řetězových zlomků.

Věta 22. *Buď $k \in \mathbb{N}_0$ a mějme přirozená čísla $a_0, a_1, \ldots, a_k$. Označme*

$$\alpha := [\overline{a_0, a_1, \ldots, a_k}].$$

Potom α je redukovaná kvadratická iracionalita a navíc platí rovnost

$$\frac{-1}{\beta} = \alpha', \ \text{kde } \beta := [\overline{a_k, a_{k-1}, \ldots, a_0}]. \tag{11}$$

Důkaz. Jelikož $a_0 \in \mathbb{N}$, tak z nerovností (3) a iracionality α plyne, že $\alpha > 1$; analogicky platí i $\beta > 1$. Rozlišíme 2 případy:

1. $k = 0$: V tomto případě platí $\alpha = [\overline{a_0}]$. Z věty o aritmetice limit a periodicity tohoto řetězového zlomku snadno nahlédneme, že $\alpha = [a_0, \alpha] = a_0 + 1/\alpha$. Tuto rovnost upravíme vynásobením α:

$$\alpha^2 - a_0\alpha - 1 = 0. \tag{12}$$

Právě jsme ukázali, že α je algebraické číslo stupně 2. Nyní se podívejme na β. V našem případě platí $\alpha = \beta = [\overline{a_0}]$, a proto bude platit rovnost $\beta^2 - a_0\beta - 1 = 0$. Vynásobíme tuto rovnost zlomkem $-1/\beta^2$ a dostaneme

$$-1 - a_0\left(\frac{-1}{\beta}\right) + \frac{1}{\beta^2} = \left(\frac{-1}{\beta}\right)^2 - a_0\left(\frac{-1}{\beta}\right) - 1 = 0. \tag{13}$$

Z rovností (12) a (13) plyne, že jak α, tak $-1/\beta$ jsou kořeny kvadratické rovnice $x^2 - a_0x - 1 = 0$. Číslo α je ale kladné, $-1/\beta$ je záporné, jedná se tedy o dva různé kořeny této rovnice, proto platí rovnost (11). Konečně $\beta > 1$, proto $\alpha' = -1/\beta \in (-1, 0)$ a z definice jsme tedy ověřili, že α je redukovaná kvadratická iracionalita.

2. $k > 0$: V tomto případě budeme postupovat obdobně, jen si pro nalezení příslušné kvadratické rovnice budeme muset pomoci lemmatem 21.

Nechť $\alpha = [\overline{a_0, a_1, \ldots, a_k}]$. Potom $\alpha = [a_0, a_1, \ldots, a_k, \alpha]$ a použitím tvrzení 11 dostáváme $\alpha = \frac{\alpha p_k + p_{k-1}}{\alpha q_k + q_{k-1}}$. Vynásobíme jmenovatelem a upravíme:

$$q_k\alpha^2 - (p_k - q_{k-1})\alpha - p_{k-1} = 0. \tag{14}$$

Tedy α je opět algebraické číslo stupně 2. Nyní zkoumejme číslo $\beta = [\overline{a_k, a_{k-1}, \ldots, a_0}]$. Nechť r_n/s_n je n-tý sblížený zlomek čísla β. Speciálně se podíváme na k-tý a $(k-1)$-ní sblížený zlomek, pro které dle rovnosti (1) platí

$$[a_k, a_{k-1}, \ldots, a_1] = \frac{r_{k-1}}{s_{k-1}}, \quad [a_k, a_{k-1}, \ldots, a_0] = \frac{r_k}{s_k}. \tag{15}$$

Již zmíněné lemma 21 dává

$$[a_k, a_{k-1}, \ldots, a_0] = \frac{p_k}{p_{k-1}}, \quad [a_k, a_{k-1}, \ldots, a_1] = \frac{q_k}{q_{k-1}}. \tag{16}$$

Porovnáním rovností (15) a (16) dostáváme $p_k/p_{k-1} = r_k/s_k$ a $q_k/q_{k-1} = r_{k-1}/s_{k-1}$. Všechny tyto zlomky jsou sblížené zlomky, proto jsou dle důsledku 7 v základním tvaru, takže musí nastat rovnost příslušných čitatelů a jmenovatelů, tj.

$$p_k = r_k, \quad p_{k-1} = s_k, \quad q_k = r_{k-1}, \quad q_{k-1} = s_{k-1}. \tag{17}$$

S výrazem β uděláme podobné úpravy jako s výrazem α: Platí $\beta = [a_k, a_{k-1}, \ldots, a_0] = [\overline{a_k, a_{k-1}, \ldots, a_0}, \beta]$, tedy opět použijeme tvrzení 11 a dostáváme, spolu s použitím vztahu (17), rovnost

$$\beta = \frac{\beta r_k + r_{k-1}}{\beta s_k + s_{k-1}} = \frac{\beta p_k + q_k}{\beta p_{k-1} + q_{k-1}}. \tag{18}$$

Opět se zbavíme jmenovatele, upravíme: $\beta^2 p_{k-1} + \beta(q_{k-1} - p_k) - q_k = 0$. Stejně jako v minulém případě i zde nás zajímá výraz $\frac{-1}{\beta}$, tedy vydělíme celou rovnost $-\beta^2$. Po úpravě

$$q_k\left(\frac{-1}{\beta}\right)^2 - (p_k - q_{k-1})\left(\frac{-1}{\beta}\right) - p_{k-1} = 0. \tag{19}$$

Odvodili jsme rovnosti (14) a (19), ze kterých plyne, že α a $\frac{-1}{\beta}$ jsou dva různé kořeny rovnice $q_kx^2 - (p_k - q_{k-1})x - p_{k-1} = 0$, tedy platí vztah (11).

Zbytek důkazu je analogický jako v případě $k = 0$, tedy $\alpha > 1$, a zároveň $\alpha' = -1/\beta \in (-1, 0)$, takže i v tomto případě je α redukovaná kvadratická iracionalita. □

Všimněme si, že ve druhé části důkazu, kdy $k > 0$, vzniklý polynom s kořenem α dává po dosazení $k = 0$ přesně polynom, který jsme odvodili v první části důkazu. Jediným důvodem, proč jsme se s případem $k = 0$ museli vypořádat zvlášť, byly předpoklady lemmatu 21, které pro $k = 0$ nedává smysl, neboť $q_{-1} = 0$ a nulou dělit nemůžeme.

Platí též obrácené tvrzení k větě 22, které dokážeme ve větě 25.

Nyní se již podíváme na rozvoj redukované kvadratické iracionality α do řetězového zlomku. Prozkoumáme začátek tohoto rozvoje α_1 v následujícím tvrzení. Tvrzení pak opakovaně využijeme na začátku důkazu klíčové věty 25. Pro účely věty 26 bude nutné se podívat na rozvoj v obecnějším případě, kdy α je algebraické číslo stupně 2.

Tvrzení 23. *Buď α algebraické číslo stupně 2. Nechť $\alpha = a_0 + 1/\alpha_1$, kde $a_0 = \lfloor \alpha \rfloor$ a $\alpha_1 \in \mathbb{R}$. Potom α_1 je též algebraické číslo stupně 2. Je-li navíc $\alpha = \frac{u+\sqrt{N}}{v}$ redukovaná kvadratická iracionalita jako ve vztahu (6), potom α_1 je též redukovaná kvadratická iracionalita tvaru $\alpha_1 = \frac{u_1+\sqrt{N}}{v_1}$ pro vhodná $u_1, v_1 \in \mathbb{Z}$.*

Důkaz. Nechť α je algebraické číslo stupně 2, kořen polynomu $ax^2 + bx + c$. Potom $a\alpha^2 + b\alpha + c = 0$. Dosadíme $\alpha = a_0 + 1/\alpha_1$ a po úpravě dostaneme

$$\alpha_1^2(aa_0^2 + ba_0 + c) + \alpha_1(2aa_0 + b) + a = 0. \tag{20}$$

Zřejmě je α_1 iracionální číslo a zároveň polynom z (20) je opravdu kvadratický, protože kvadratický člen $aa_0^2 + ba_0 + c$ je nenulový, což plyne z toho, že polynom $ax^2 + bx + c$ má dva iracionální kořeny α a α', tedy a_0 nemůže být kořenem tohoto polynomu. Proto α_1 je nutně algebraické číslo stupně 2.

Pokud je navíc α redukovaná kvadratická iracionalita tvaru (6), tj. platí $u = -b$, $v = 2a$, $N = b^2 - 4ac$, pak spočítáme diskriminant kvadratické rovnice (20), jejímž řešením je α_1: $(2aa_0 + b)^2 - 4a(aa_0^2 + ba_0 + c) = b^2 - 4ac = N$, tedy diskriminant je stejný jako u rovnice, jejímž řešením je α, což jsme chtěli ukázat. Proto

$$\alpha_1 = \frac{-(2aa_0 + b) \pm \sqrt{N}}{2(aa_0^2 + ba_0 + c)}. \tag{21}$$

Jelikož $2(aa_0^2 + ba_0 + c) < 0$, tak správná volba znaménka v čitateli bude minus, proto je potřeba volit $u_1 = 2aa_0 + b$, $v_1 = -2(aa_0^2 + ba_0 + c)$. Nyní stačí ukázat nerovnosti $\alpha_1 > 1$, $\alpha_1' \in (-1, 0)$. To však plyne ze vztahu $\alpha = a_0 + 1/\alpha_1$.

Jelikož $1/\alpha_1 = \alpha - a_0 \in (0, 1)$, tak jistě $\alpha_1 > 1$. K důkazu, že $\alpha_1' \in (-1, 0)$, stačí ukázat $-1/\alpha_1' > 1$: Máme $-1/\alpha_1' = a_0 - \alpha' > 1$, protože $a_0 \in \mathbb{N}$ a $\alpha' < 0$. Tedy jsme ukázali, že α_1 je redukovaná kvadratická iracionalita a důkaz je hotov. □

Opakovaným použitím předchozího tvrzení 23 dostáváme následující důsledek.

Důsledek 24. *Buď α algebraické číslo stupně 2. Uvažujme libovolné $n \in \mathbb{N}$ a $\alpha = [a_0, a_1, \ldots, a_{n-1}, \alpha_n]$ podobně jako v definici 9. Potom α_n je algebraické číslo*

stupně 2 a bylo-li číslo α redukovanou kvadratickou iracionalitou, potom je i α_n redukovanou kvadratickou iracionalitou.

Tvrzení 23 naznačuje, jakým způsobem se rozvíjí redukovaná kvadratická iracionalita do řetězového zlomku. Jak tento rozvoj opravdu vypadá, dokážeme v následující větě 25, která představuje obrácení věty 22.

Věta 25. *Nechť α je redukovaná kvadratická iracionalita. Potom řetězový zlomek čísla α je ryze periodický.*

Důkaz. Nejprve se podíváme na posloupnost reálných čísel $\{\alpha_i\}_{i=0}^{\infty}$ z definice 9. Podle důsledku 24 jsou všechna čísla α_i redukované kvadratické iracionality tvaru $\frac{u_i+\sqrt{N}}{v_i}$ pro vhodná u_i, $v_i \in \mathbb{N}$. Podle důsledku 20 ovšem platí, že takových α_i je konečně mnoho.

Existuje tedy právě jedna volba indexů k, l, $0 \leqq k < l$, taková, že α_0, α_1, …, α_k, …, α_{l-1} jsou po dvou různá, a navíc $\alpha_l = \alpha_k$ (značení odpovídá definici periodického řetězového zlomku). Matematickou indukcí se poté snadno ukáže, že pro každé $i \in \mathbb{N}$ je $\alpha_{k+i} = \alpha_{l+i}$. Dále platí implikace $\alpha_i = \alpha_j \Rightarrow a_i = a_j$ pro libovolná i, j, čímž jsme ukázali, že $[a_0, a_1, \ldots, a_{k-1}, \overline{a_k, a_{k+1}, \ldots, a_{l-1}}]$, tedy řetězový zlomek čísla α je od nějakého místa periodický. Nyní zbývá ukázat, že $k = 0$, tedy že jde o ryze periodický řetězový zlomek. To ukážeme sporem.

Ukážeme, že $\alpha_k = \alpha_l$ implikuje $\alpha_{k-1} = \alpha_{l-1}$, čímž bychom pro $k > 0$ dostali spor s tím, že $\alpha_0, \alpha_1, \ldots, \alpha_k, \ldots, \alpha_{l-1}$ jsou navzájem různá. Pro větší přehlednost označíme $\beta_i = -1/\alpha_i'$ podobně jako ve větě 22. Upravíme:

$$\begin{aligned} \alpha_{k-1} &= a_{k-1} + \frac{1}{\alpha_k}, & \alpha_{l-1} &= a_{l-1} + \frac{1}{\alpha_l}, \\ \alpha_{k-1}' &= a_{k-1} - \beta_k, & \alpha_{l-1}' &= a_{l-1} - \beta_l, \\ \beta_k &= a_{k-1} + \frac{1}{\beta_{k-1}}, & \beta_l &= a_{l-1} + \frac{1}{\beta_{l-1}}. \end{aligned}$$

Z rovnosti $\alpha_k = \alpha_l$ okamžitě plyne $\beta_k = \beta_l$. Z toho, že α_{k-1} je redukovaná kvadratická iracionalita, plyne $\frac{1}{\beta_{k-1}} \in (0, 1)$, a tak $\lfloor \beta_k \rfloor = a_{k-1}$, analogicky $\lfloor \beta_l \rfloor = a_{l-1}$. Z toho dostáváme, že $a_{k-1} = a_{l-1}$, z čehož dále snadno plyne požadovaná rovnost $\alpha_{k-1} = \alpha_{l-1}$. Proto nutně $k = 0$ a řetězový zlomek čísla α je ryze periodický, což jsme chtěli ukázat. □

Dokázali jsme, že právě redukované kvadratické iracionality mají ryze periodický řetězový zlomek s kladným prvním koeficientem. Důsledkem této ekvivalence je charakterizace všech periodických řetězových zlomků:

Věta 26. *Reálné číslo α je algebraické číslo stupně 2, právě když řetězový zlomek čísla α je periodický.*

Důkaz. $\Leftarrow$: Nechť $\alpha = [a_0, a_1, \ldots, a_{k-1}, \overline{a_k, \ldots, a_l}]$ pro nějaké indexy $k, l \in \mathbb{N}_0$, $k \leqq l$. Označme $\beta = [\overline{a_k, \ldots, a_l}]$. Potom β má ryze periodický řetězový zlomek, tedy podle věty 22 je číslo β redukovaná kvadratická iracionalita; speciálně jde o algebraické číslo stupně 2. Zároveň platí vztah $\alpha = [a_0, a_1, \ldots, a_{k-1}, \beta]$. Nyní použijeme tvrzení 11

a dostaneme rovnost $\alpha = \frac{\beta p_k + p_{k-1}}{\beta q_k + q_{k-1}}$. Z této rovnosti je patrné, že $\alpha \in \mathbb{Q}(\beta)$, kde $\mathbb{Q}(\beta)$ je nejmenší nadtěleso tělesa $\mathbb{Q}$ obsahující číslo β. Z toho již snadno dostáváme, že α je kořenem nějakého kvadratického polynomu, a jelikož je α iracionální, tak α je algebraické číslo stupně 2, což jsme chtěli ukázat.

$\Rightarrow$: Nechť je α algebraické číslo stupně 2. Uvažujme nějaké $n \in \mathbb{N}$ (přesná volba n bude specifikována později) a pišme $\alpha = [a_0, a_1, \ldots, a_n, \alpha_{n+1}]$. Podle důsledku 24 je α_{n+1} algebraické číslo stupně 2 a navíc zřejmě $\alpha_{n+1} > 1$. Ukážeme, že vhodná volba indexu n též zajistí, že $\alpha'_{n+1} \in (-1, 0)$, což bude znamenat, že číslo α_{n+1} je redukovaná kvadratická iracionalita, tedy dle věty 25 má číslo α_{n+1} ryze periodický řetězový zlomek, z čehož již snadno plyne periodičnost řetězového zlomku čísla α.

Podle tvrzení 11 platí rovnost $\alpha = \frac{\alpha_{n+1} p_n + p_{n-1}}{\alpha_{n+1} q_n + q_{n-1}}$. Podívejme se na sdružená čísla obou stran této rovnosti (ta si budou pochopitelně rovna). Po úpravě dostaneme

$$\alpha'_{n+1} = -\frac{\alpha' q_{n-1} - p_{n-1}}{\alpha' q_n - p_n} = -\frac{q_{n-1}}{q_n} \frac{\alpha' - c_{n-1}}{\alpha' - c_n}, \tag{22}$$

kde $c_k := p_k/q_k$ pro vhodná k. Podívejme se podrobněji na posloupnost

$$\frac{\alpha' - c_{n-1}}{\alpha' - c_n}, \quad n \in \mathbb{N}. \tag{23}$$

Z věty 8 plyne, že pro $n \to \infty$ konverguje zlomek z rovnosti (23) k 1, a navíc díky nerovnostem (2) dostáváme, že pokud k-tý člen posloupnosti určené (23) byl menší než 1, tak člen a_{k+1} stejné posloupnosti bude větší než 1, člen a_{k+2} bude opět menší než 1 apod. Z toho ovšem plyne, že pro vhodné (a dostatečně velké) $n \in \mathbb{N}$ platí $\frac{\alpha' - c_{n-1}}{\alpha' - c_n} \in (0, 1)$. Dále snadno nahlédneme, že $q_{n-1}/q_n \in (0, 1)$, a tak podle rovnosti (22) dostáváme vztah $\alpha'_{n+1} \in (-1, 0)$. To však přesně znamená, že α_{n+1} je redukovaná kvadratická iracionalita, tedy dle věty 25 má α_{n+1} ryze periodický řetězový zlomek. Nechť tedy platí $\alpha_{n+1} = [\overline{a_{n+1}, a_{n+2}, \ldots, a_l}]$. Po dosazení dostáváme $\alpha = [a_0, a_1, \ldots, a_n, \overline{a_{n+1}, a_{n+2}, \ldots, a_l}]$, tedy α má periodický řetězový zlomek, což jsme chtěli ukázat. □

5. Řetězové zlomky druhých odmocnin z přirozených čísel

Dalším důsledkem charakterizace redukovaných kvadratických iracionalit pomocí jejich řetězových zlomků jsou vlastnosti řetězového zlomku pro $\sqrt{N}$, kde N je přirozené číslo, které není čtvercem.

Věta 27. *Nechť N je přirozené číslo, které není čtvercem, a $\sqrt{N} = [a_0, a_1, a_2, \ldots]$. Pak existuje $k \in \mathbb{N}_0$ tak, že $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$, kde navíc $a_i = a_{k+1-i}$ pro každé $i \in \{1, 2, \ldots, k\}$, tedy posloupnost přirozených čísel $(a_1, a_2, \ldots, a_k)$ je symetrická.*

Důkaz. Nechť $\sqrt{N} = [a_0, a_1, a_2, \ldots]$. Potom jistě $\sqrt{N} + a_0 = [2a_0, a_1, a_2, \ldots]$. $\sqrt{N} + a_0$ je zřejmě redukovaná kvadratická iracionalita, a tak podle věty 25 je řetězový zlomek $[2a_0, a_1, a_2, \ldots]$ ryze periodický.

Pokud $\sqrt{N} + a_0 = [\overline{2a_0}]$, pak jednoduchou úpravou dostaneme $\sqrt{N} = [a_0, \overline{2a_0}]$ (toto odpovídá případu $k = 0$) a jsme hotovi.

Pokud existuje $k \in \mathbb{N}$ tak, že $\sqrt{N} + a_0 = [\overline{2a_0, a_1, a_2, \ldots, a_k}]$, pak jistě $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$. Zbývá ukázat symetrii posloupnosti $(a_1, a_2, \ldots, a_k)$.

Označme $\alpha := a_0 + \sqrt{N}$ a $\beta = -1/\alpha'$. Podle věty 22 platí

$$\beta = [\overline{a_k, a_{k-1}, \ldots, a_1, 2a_0}]. \tag{24}$$

Vztah $\beta = -1/\alpha'$ nám však umožní určit řetězový zlomek čísla β i jinak:

$$-\alpha' = \sqrt{N} - a_0 = [0, \overline{a_1, a_2, \ldots, a_k, 2a_0}] = \frac{1}{[\overline{a_1, a_2, \ldots, a_k, 2a_0}]},$$

z čehož snadno dostáváme

$$\beta = [\overline{a_1, a_2, \ldots, a_k, 2a_0}]. \tag{25}$$

Rozvoj čísla β do řetězového zlomku je však jednoznačný, a proto porovnáním koeficientů řetězových zlomků v rovnostech (24) a (25) dostaneme požadované vztahy $a_i = a_{k+1-i}$ pro každé $i \in \{1, 2, \ldots, k\}$, tedy posloupnost $(a_1, a_2, \ldots, a_k)$ je symetrická, což jsme chtěli ukázat. □

Uveďme několik příkladů, jak vypadají řetězové zlomky druhých odmocnin z přirozených čísel (další příklady lze nalézt v [5] na straně 116):

N	řetězový zlomek $\sqrt{N}$
1	$[1]$
2	$[1, \overline{2}]$
3	$[1, \overline{1, 2}]$
4	$[2]$
5	$[2, \overline{4}]$
6	$[2, \overline{2, 4}]$
7	$[2, \overline{1, 1, 1, 4}]$

Pokud je pro nějaké $k \in \mathbb{N}$ posloupnost přirozených čísel $(a_1, \ldots, a_k)$ symetrická, pak platí následující identita.

Lemma 28. *Nechť $a_0 \in \mathbb{N}$ a $(a_1, \ldots, a_k)$ je symetrická posloupnost přirozených čísel. Potom $p_k = a_0 q_k + q_{k-1}$.*

Důkaz. V tomto důkazu využijeme řetězové polynomy z definice 3 a jejich vlastnosti z lemmatu 4. Pro $k = 1$ je tvrzení zřejmé. Pro $k > 1$ upravujme

$$\begin{aligned} p_k = k_{k+1}(a_0, \ldots, a_k) &= k_{k+1}(a_k, \ldots, a_0) = a_0 k_k(a_k, \ldots, a_1) + k_{k-1}(a_k, \ldots, a_2) = \\ &= a_0 k_k(a_1, \ldots, a_k) + k_{k-1}(a_1, \ldots, a_{k-1}), \end{aligned}$$

kde v poslední rovnosti jsme u prvního sčítance použili lemma 4 a u druhého sčítance symetrii posloupnosti $(a_1, \ldots, a_k)$. Platnost vztahu $q_i = k_i(a_1, \ldots, a_i)$ pro každé $i \in \mathbb{N}$ pak dokazuje tvrzení. □

Bez předpokladu symetrie posloupnosti $(a_1, \ldots, a_k)$ lemma 28 neplatí. Protipříkladem je nesymetrická posloupnost přirozených čísel (1, 2).

Ve větě 27 jsme odvodili, že pro každé přirozené číslo N, které není čtvercem, platí

$$\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}] \tag{26}$$

pro nějaké $k \in \mathbb{N}_0$, kde posloupnost $(a_1, a_2, \ldots, a_k)$ je symetrická nebo prázdná $(k = 0)$. Jelikož $a_0 = \lfloor \sqrt{N} \rfloor$, má smysl zabývat se otázkou, zdali pro danou symetrickou posloupnost $(a_1, a_2, \ldots, a_k)$ (či prázdnou posloupnost pro případ $k = 0$) existují $N \in \mathbb{N}$ taková, že $\sqrt{N}$ má právě řetězový zlomek tvaru (26).

Skutečnost je taková, že pro danou symetrickou posloupnost tato N buď vůbec neexistují, nebo jich existuje nekonečně mnoho, dokonce bezčtvercových, což dokázal C. Friesen v článku [2]. Nutné a postačující podmínky, za jakých tato čísla N existují, odvodíme v této a následující kapitole 6. Za jistých předpokladů pro symetrickou posloupnost $(a_1, a_2, \ldots, a_k)$ vyjádříme všechna přirozená N, pro která platí rovnost (26), jako funkční hodnoty jistého kvadratického polynomu s celočíselnými koeficienty, které závisejí pouze na hodnotách q_k a q_{k-1}, definovaných rekurentním vzorcem z definice 5.

Degenerovanému případu $k = 0$, kdy symetrická posloupnost je prázdná, neboli $\sqrt{N} = [a_0, \overline{2a_0}]$, se budeme věnovat zvlášť.

5.1. Případ $k = 0$

Nejprve charakterizujeme všechna přirozená N, pro která platí rovnost

$$\sqrt{N} = [a_0, \overline{2a_0}], \tag{27}$$

kde a_0 je dolní celá část čísla $\sqrt{N}$.

Pokud N je takové přirozené číslo, pak lze snadno nahlédnout rovnost

$$\sqrt{N} = a_0 + \frac{1}{\sqrt{N} + a_0},$$

která je ekvivalentní vztahu

$$N = a_0^2 + 1. \tag{28}$$

Opačně lze též ověřit, že čísla tvaru $a_0^2 + 1$ mají přesně požadovaný řetězový zlomek, což lze ukázat přímo z definice posloupnosti $\{a_i\}$. Shrňme získané poznatky do věty.

Věta 29. *Existuje nekonečně mnoho přirozených čísel N splňujících rovnost* (27). *Taková čísla N jsou právě hodnoty kvadratického polynomu $x^2 + 1$ pro $x = 1, 2, 3, \ldots$*

5.2. Obecný případ $k \in \mathbb{N}$

Nyní pro libovolnou symetrickou posloupnost $(a_1, a_2, \ldots, a_k)$ přirozených čísel charakterizujeme všechna přirozená čísla N taková, že $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$ a prozkoumáme, zda taková čísla N vůbec existují.

Buďte N, k přirozená čísla a uvažujme symetrickou posloupnost přirozených čísel $(a_1, a_2, \ldots, a_k)$ takovou, že platí rovnost $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$. Potom platí

též rovnost $\sqrt{N} = [a_0, a_1, \ldots, a_k, \sqrt{N} + a_0]$. Nyní použijeme vztah (4) z tvrzení 11 a dostaneme

$$\sqrt{N} = \frac{(\sqrt{N} + a_0)p_k + p_{k-1}}{(\sqrt{N} + a_0)q_k + q_{k-1}}. \tag{29}$$

Tuto rovnost lze ekvivalentně přepsat jako

$$Nq_k + \sqrt{N}(a_0 q_k + q_{k-1}) = \sqrt{N}p_k + a_0 p_k + p_{k-1}. \tag{30}$$

Snadno si lze rozmyslet, že předchozí rovnost je ekvivalentní rovnosti celočíselných částí a koeficientů u $\sqrt{N}$. Tedy dostáváme dvě rovnosti,

$$p_k = a_0 q_k + q_{k-1}, \tag{31}$$

$$Nq_k = a_0 p_k + p_{k-1}. \tag{32}$$

I toto byla ekvivalentní úprava. Následující rovnost platí díky tvrzení 6 a jednoduché úpravě:

$$p_{k-1} = \frac{p_k q_{k-1} + (-1)^k}{q_k}. \tag{33}$$

Nakonec vezmeme rovnost (32) a dosadíme za p_k, p_{k-1} ze vztahů (31) a (33), vzniklý výraz upravíme a dostaneme

$$(N - a_0^2)q_k - 2a_0 q_{k-1} = \frac{q_{k-1}^2 + (-1)^k}{q_k}. \tag{34}$$

Nyní ale není vůbec jasné, že šlo o ekvivalentní úpravu v tom smyslu, že pro každé $a_0 \in \mathbb{N}$ a symetrickou posloupnost přirozených čísel $(a_1, \ldots, a_k)$ z rovnosti (34) plyne, že $\sqrt{N}$ má požadovaný tvar řetězového zlomku (26). K tomu ovšem stačí ukázat, že platí všechny rovnosti od (29) až po (33). Upozorněme, že v tomto konkrétním případě nedefinujeme posloupnosti $\{p_i\}$, $\{q_i\}$ jakožto čitatele a jmenovatele nějakých sblížených zlomků, neboť nemáme žádný řetězový zlomek (naopak jej chceme určit). Proto definujeme posloupnosti $\{p_i\}$, $\{q_i\}$ v této úvaze rekurentními vzorci z definice 5. S rovností (33) není problém, ta plyne z tvrzení 11, stejně jako s rovností (31), která plyne z lemmatu 28.

Z rovností (31) a (33) již lze úpravou rovnosti (34) snadno odvodit i rovnost (32), což nás přesně zajímalo. Nicméně platnost rovností (31) a (32) je ekvivalentní s platností rovnice (30) a ta je ekvivalentní se vztahem (29).

Nyní zbývá dojít od vztahu (29) zpět k rovnosti (26), tj. zjistit, proč z rovnosti $\sqrt{N} = \frac{(\sqrt{N}+a_0)p_k+p_{k-1}}{(\sqrt{N}+a_0)q_k+q_{k-1}}$ plyne $\sqrt{N} = [a_0, \overline{a_1, \ldots, a_k, 2a_0}]$. Jde o důsledek lemmatu 12, kde volíme $\alpha = \sqrt{N}$, $\beta = \sqrt{N} + a_0$ a $m = k$. Tím jsme vskutku odvodili rovnost (26), jak jsme si přáli.

Shrňme získané výsledky do následujícího tvrzení.

Tvrzení 30. *Buďte N, k, a_0 přirozená čísla a uvažme symetrickou posloupnost přirozených čísel $(a_1, a_2, \ldots, a_k)$. Potom $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$, právě když platí*

$$(N - a_0^2)q_k - 2a_0 q_{k-1} = \frac{q_{k-1}^2 + (-1)^k}{q_k}.$$

Připomeňme, že naším cílem je určit všechna přirozená N taková, že pro danou symetrickou posloupnost přirozených čísel $(a_1, \ldots, a_k)$, $k \in \mathbb{N}$, platí vztah (26). Díky tvrzení 30 stačí zjistit, pro která přirozená N platí rovnost (34).

Prvním naivním přístupem je vyjádřit číslo N přímo z rovnosti (34), což ovšem nevede k ničemu zajímavému, protože dané vyjádření N by záviselo na q_k, q_{k-1}, k, což je v pořádku, ale také by záviselo na a_0, nicméně z tvrzení 30 plyne, že $a_0 = \lfloor \sqrt{N} \rfloor$, což činí toto vyjádření irelevantním.

To znamená, že chceme vyjádřit a_0 pouze v závislosti na q_k, q_{k-1} a k.

6. Lineární diofantické rovnice a dokončení důkazu o řetězovém zlomku $\sqrt{N}$

Podle důsledku 7 jsou čísla q_k, q_{k-1} nesoudělná. Pak ovšem z rovnosti (34) plyne, že dvojice přirozených čísel $(N - a_0^2, 2a_0)$ je řešením lineární diofantické rovnice

$$xq_k - yq_{k-1} = \frac{q_{k-1}^2 + (-1)^k}{q_k}. \tag{35}$$

Není těžké ukázat, jak vypadají všechna řešení lineární diofantické rovnice, tedy všechny uspořádané dvojice celých čísel (x, y) splňující $ax - by = c$ pro dané celočíselné parametry a, b, c, kde navíc čísla a, b jsou nesoudělná. Množinu všech řešení této rovnice popisuje následující lemma.

Lemma 31. *Buďte a, b nesoudělná celá čísla, $c \in \mathbb{Z}$. Nechť (x_1, y_1) je řešení rovnice*

$$ax - by = c. \tag{36}$$

Potom (x_2, y_2) je řešením rovnice (36)*, právě když existuje $k \in \mathbb{Z}$ takové, že*

$$x_2 = x_1 + kb, \quad y_2 = y_1 + ka. \tag{37}$$

Důkaz. Je zřejmé, že (x_2, y_2) vyhovující podmínce (37) též splňují rovnost (36).

Opačně, mějme nějaké řešení (x_2, y_2) rovnice (36), tedy nechť $ax_2 - by_2 = c$. Jelikož (x_1, y_1) též řeší tuto rovnici, tak odečtením vzniklých rovnic dostáváme vztah

$$a(x_2 - x_1) = b(y_2 - y_1). \tag{38}$$

Napišme si Bézoutovu rovnost pro nesoudělná a, b, tedy nechť $ra + sb = 1$ pro nějaká r, $s \in \mathbb{Z}$. Tuto rovnost vynásobíme rozdílem $x_2 - x_1$ a dosazením vztahu (38) dostáváme $rb(y_2 - y_1) + sb(x_2 - x_1) = x_2 - x_1$. To ovšem znamená, že b dělí $x_2 - x_1$, neboli existuje $k \in \mathbb{Z}$ tak, že $x_2 - x_1 = kb$, což přesně znamená $x_2 = x_1 + kb$. Nakonec dosadíme vztah $x_2 - x_1 = kb$ do (38) a dostaneme $y_2 - y_1 = ka$. To odpovídá vztahu $y_2 = y_1 + ka$ a lemma je dokázáno. □

Poznamenejme, že z Bézoutovy rovnosti rovněž plyne, že řešení (x_1, y_1) vždy existuje. Zároveň zde využíváme předpokladu nesoudělnosti čísel a, b: Pokud by čísla a, b byla soudělná, tak by řešení (x_1, y_1) nemuselo existovat.

Přímým výpočtem lze ověřit, že v našem případě je jedním řešením diofantické rovnice (35) dvojice čísel

$$x = \frac{(-1)^k z(q_{k-1}^2 + (-1)^k)}{q_k}, \quad y = (-1)^k z q_{k-1}, \tag{39}$$

kde $z := \frac{q_{k-1}^2+(-1)^k}{q_k}$ je pravá strana rovnice (35). Všechna řešení této rovnice jsme popsali v lemmatu 31, jehož aplikací (a dosazením za z) dostáváme

$$N - a_0^2 = \frac{(-1)^k(q_{k-1}^2+(-1)^k)^2}{q_k^2} + mq_{k-1}, \tag{40}$$

$$2a_0 = (-1)^k q_{k-1}\frac{q_{k-1}^2+(-1)^k}{q_k} + mq_k,\ m \in \mathbb{Z}. \tag{41}$$

Tady však ztrácíme kontrolu nad znaménkem a celočíselností a_0. V předpokladech tvrzení 30 máme $a_0 \in \mathbb{N}$, což však rovností (41) není zaručené. V prvé řadě požadujeme $a_0 > 0$, což lze zajistit tím, že množinu vhodných $m \in \mathbb{Z}$ zdola omezíme, neboť pro některá m záporná jistě nastane $a_0 < 0$. Vhodná m specifikujeme dle rovnice (41):

$$2a_0 = (-1)^k q_{k-1}\frac{q_{k-1}^2+(-1)^k}{q_k} + mq_k = \frac{1}{q_k}\left((-1)^k q_{k-1}(q_{k-1}^2+(-1)^k) + mq_k^2\right).$$

Protože $q_k > 0$, stačí volit právě taková $m \in \mathbb{Z}$, pro která platí nerovnost

$$mq_k^2 > (-1)^{k-1}q_{k-1}(q_{k-1}^2+(-1)^k). \tag{42}$$

Pokud označíme m_0 nejmenší $m \in \mathbb{Z}$, které vyhovuje nerovnosti (42), potom všechna vhodná m lze zřejmě charakterizovat podmínkou $m \geqq m_0$.

Nyní tedy platí $a_0 > 0$. Dále potřebujeme zajistit, aby a_0 bylo vůbec celočíselné. Ověříme, pro která $m \in \mathbb{Z}$ vyhovující vztahu (42) současně platí, že pravá strana rovnosti (41) je sudá, tedy

$$(-1)^k q_{k-1}\frac{q_{k-1}^2+(-1)^k}{q_k} + mq_k \quad \text{je sudé.} \tag{43}$$

V tuto chvíli je třeba rozlišit 4 případy podle parity čísel q_{k-1}, q_k a $\frac{q_{k-1}^2+(-1)^k}{q_k}$. Tyto případy budeme odlišovat až do konce této sekce a budeme se držet jejich značení „případ i“ pro $i = 1, 2, 3, 4$.

1. q_{k-1} je sudé. Potom z nesoudělnosti čísel q_{k-1} a q_k plyne, že q_k je nutně liché, což znamená, že (42) platí, právě když je m sudé.

2. q_{k-1} je liché, q_k je liché. Poté snadno nahlédneme, že zlomek $\frac{q_{k-1}^2+(-1)^k}{q_k}$ je nutně sudý, a tak (42) platí, právě když je m sudé.

3. q_{k-1} je liché, q_k je sudé a $\frac{q_{k-1}^2+(-1)^k}{q_k}$ je sudé. V tomto případě dokonce platí, že libovolné m zajišťuje platnost (42).

4. q_{k-1} je liché, q_k je sudé a $\frac{q_{k-1}^2+(-1)^k}{q_k}$ je liché. V tomto případě bude pravá strana rovnosti (41) pro libovolné $m \in \mathbb{Z}$ lichá, a tak (42) nebude platit nikdy.

Jiný případ zjevně nastat nemůže.

Diskuzi o paritách čísel q_{k-1}, q_k a $\frac{q_{k-1}^2+(-1)^k}{q_k}$ shrneme do tabulky, přičemž se nadále budeme držet značení $z := \frac{q_{k-1}^2+(-1)^k}{q_k}$:

Číslo případu	q_{k-1}	q_k	z	Vhodná m vyhovující (43)
1	sudé	liché	liché	sudá
2	liché	liché	sudé	sudá
3	liché	sudé	sudé	libovolná
4	liché	sudé	liché	žádná

Volba vhodných m nám tedy zajistí, aby z rovnosti (41) vyplynulo, že $a_0 \in \mathbb{N}$. Pro případ, kdy z i q_{k-1} jsou lichá, jsme se přesvědčili o tom, že z rovnosti (41) plyne, že a_0 nikdy přirozené nebude. Spolu s použitím tvrzení 30 jsme tak odvodili tuto další ekvivalentní podmínku k původní rovnosti (26), která se zároveň ukáže být charakterizací případů, kdy pro symetrickou posloupnost $(a_1, a_2, \ldots, a_k)$ existuje N přirozené, vyhovující rovnosti (26), tedy $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$.

Popišme právě odvozené výsledky v následujícím tvrzení.

Tvrzení 32. *Buďte N, k, a_0 přirozená čísla a uvažujme symetrickou posloupnost přirozených čísel $(a_1, a_2, \ldots, a_k)$. Potom rovnost*

$$\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$$

je ekvivalentní s rovnostmi (40) a (41), *tedy po označení $z := \frac{q_{k-1}^2 + (-1)^k}{q_k}$ se vztahy*

$$N - a_0^2 = (-1)^k z^2 + m q_{k-1},\ m \in \mathbb{Z}, \tag{44}$$

$$2a_0 = (-1)^k q_{k-1} z + m q_k,\ m \in \mathbb{Z}. \tag{45}$$

Navíc výroky na obou stranách ekvivalence platí, právě když nastane jeden z případů 1, 2, 3 *v tabulce výše, tedy právě když buď q_{k-1} nebo z je sudé, m vyhovuje odhadu* (42) a *v případech* 1, 2 *je navíc m sudé.*

V případě 4 *ani jeden z výroků na obou stranách ekvivalence neplatí.*

Připomeňme, že pokud se chceme ujistit, který z případů 1, 2, 3, 4 nastane, stačí se podívat na příslušnou symetrickou posloupnost $(a_1, a_2, \ldots, a_k)$, ze které odvodíme hodnoty q_k, q_{k-1} i z a jejich paritu.

Zjistili jsme, že pokud q_{k-1} i z jsou lichá, tak neexistuje žádné $N \in \mathbb{N}$ splňující rovnost (26). V ostatních případech jsme již jen krůček od vyjádření čísla N v závislosti na q_{k-1}, q_k a k, neboť budeme moci vyjádřit a_0 z rovnosti (45) a poté bude stačit dosadit do vztahu (44). To znamená, že pro danou symetrickou posloupnost přirozených čísel $(a_1, a_2, \ldots, a_k)$ dokážeme popsat všechna přirozená čísla N s řetězovým zlomkem $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$ jakožto funkční hodnoty kvadratického polynomu s celočíselnými koeficienty, které závisejí pouze na posloupnosti $(a_1, a_2, \ldots, a_k)$, za předpokladu, že q_{k-1} či z je sudé (což odpovídá případům 1, 2, 3 z diskuze v minulé části). Nejprve se budeme věnovat zvlášť případu 3 (q_{k-1} liché, q_k a z sudá) a poté najednou vyšetříme případy 1 (q_{k-1} sudé, q_k i z lichá) a 2 (q_{k-1} i q_k lichá, z sudé), neboť se v nich bude postupovat zcela analogicky. Jak jsme již avizovali, budeme vycházet z rovností (45) a (44), což motivuje k tomu, proč zkoumat případ 3 zvlášť. V případě 3 je m libovolné, zatímco v případech 1, 2 musí být m sudé. Současně v každém z případů 1, 2, 3 musí být m dostatečně velké, tato podmínka je popsána nerovností (42).

Případ 3: Nechť $(a_1, a_2, \ldots, a_k)$ je symetrická posloupnost přirozených čísel taková, že q_{k-1} je liché, z i q_k jsou sudá. Přeznačíme $x := m$, obdobně přeznačíme $x_0 := m_0$, kde m_0 bylo nejmenší m vyhovující odhadu (42). V tomto případě tedy $x \geqq x_0$. Vydělením obou stran rovnice (45) dvěma dostaneme

$$a_0 = (-1)^k q_{k-1} \frac{z}{2} + x \frac{q_k}{2},$$

a tedy

$$a_0^2 = q_{k-1}^2 \frac{z^2}{4} + (-1)^k q_{k-1} z \frac{q_k}{2} x + \frac{q_k^2}{4} x^2. \tag{46}$$

Dosazením rovnosti (46) do rovnice (44) a po drobné úpravě dostáváme

$$N = \frac{q_k^2}{4} x^2 + q_{k-1} \frac{3 + (-1)^k q_{k-1}^2}{2} x + \frac{z^2}{4}(4(-1)^k + q_{k-1}^2), \tag{47}$$

kde $x \in \mathbb{Z}$, $x \geqq x_0$. Jde o kvadratický polynom $e_1 x^2 + f_1 x + g_1$, kde

$$e_1 = \frac{q_k^2}{4}, \quad f_1 = q_{k-1} \frac{3 + (-1)^k q_{k-1}^2}{2}, \quad g_1 = \frac{z^2}{4}(4(-1)^k + q_{k-1}^2), \tag{48}$$

což jsme chtěli odvodit.

Dá se ukázat, že případ 3 může nastat jedině tehdy, pokud je k liché (index k uvažujeme co nejmenší možný, aby byl určen jednoznačně). K tomu je potřeba spočítat diskriminant polynomu (47), který vyjde $(-1)^{k-1}$, čili $b_1^2 - 4a_1c_1 = (-1)^{k-1}$. Z toho plyne $b_1^2 \equiv (-1)^{k-1} \bmod 4$. Nicméně žádný čtverec nemůže být kongruentní s -1 modulo 4, proto k nemůže být sudé.

Případy 1, 2: Pokud máme takovou symetrickou posloupnost $(a_1, a_2, \ldots, a_k)$, že buď je q_{k-1} sudé, nebo je q_{k-1} liché, ale zároveň je i q_k liché, pak bychom postupovali analogicky jako u předchozího případu 3, pouze máme navíc podmínku, aby vhodná m byla nejen dostatečně velká, ale i sudá. Označme $x := m/2$, $x_0 := m_0/2$, kde opět m_0 je nejmenší (v tomto případě sudé) m vyhovující odhadu (42). Došli bychom ke kvadratickému polynomu

$$N = q_k^2 x^2 + q_{k-1}(3 + (-1)^k q_{k-1}^2) x + \frac{z^2}{4}(4(-1)^k + q_{k-1}^2), \tag{49}$$

kde $x \in \mathbb{Z}$, $x \geqq x_0$. To je kvadratický polynom $e_2 x^2 + f_2 x + g_2$, kde

$$e_2 = q_k^2, \quad f_2 = q_{k-1}(3 + (-1)^k q_{k-1}^2), \quad g_2 = \frac{z^2}{4}(4(-1)^k + q_{k-1}^2), \tag{50}$$

což nás přesně zajímalo. Dále si všimněme, že platí

$$4e_1 = e_2, \quad 2f_1 = f_2, \quad g_1 = g_2. \tag{51}$$

Tím jsme zobecnili výsledek z odstavce 5.1 pro libovolné $k \in \mathbb{N}_0$, popsali jsme všechna přirozená čísla N s řetězovým zlomkem

$$\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}],$$

kde $a_0 = \lfloor \sqrt{N} \rfloor$ a $(a_1, a_2, \ldots, a_k)$ je symetrická posloupnost přirozených čísel, jako funkční hodnoty kvadratického polynomu s celočíselnými koeficienty, závisejícími pouze na q_{k-1}, q_k a k, tedy pouze na symetrické posloupnosti $(a_1, a_2, \ldots, a_k)$.

Shrňme získané výsledky do následující věty, kde budeme nadále používat značení $z := \frac{q_{k-1}^2 + (-1)^k}{q_k}$. Odhad (42), popisující dolní mez pro vhodná m, lze ekvivalentně přepsat jako

$$mq_k > (-1)^{k-1} z q_{k-1}. \tag{52}$$

Současně připomeňme, že pomocí posloupnosti $\{a_i\}_{i=1}^k$ definujeme posloupnost $\{q_i\}_{i=1}^k$ rekurentně tak, jak je popsáno v definici 5.

Věta 33. *Mějme přirozená čísla N, a_0, k a uvažujme symetrickou posloupnost přirozených čísel $(a_1, a_2, \ldots, a_k)$. Potom platí:*

1. *Pokud je q_{k-1} sudé (případ 1) nebo je q_{k-1} i q_k liché a zároveň z sudé (případ 2), pak $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$, právě když platí (49) pro $x \geqq x_0 = m_0/2$, kde m_0 je nejmenší m vyhovující odhadu (52).*

2. *Pokud je q_{k-1} liché a současně jsou obě čísla q_k i z sudá (případ 3), pak $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$, právě když platí (47) pro $x \geqq x_0 = m_0$, kde m_0 je nejmenší m vyhovující odhadu (52).*

3. *Pokud q_{k-1} i z jsou lichá, pak neexistuje žádné $N \in \mathbb{N}$ vyhovující vztahu $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$.*

Ve větách 29 až 33 jsme nepředpokládali, že N není čtverec. Případ, kdy N je čtvercem, odpovídá tomu, že $\sqrt{N}$ je racionální a má tedy konečný řetězový zlomek, což je případ triviální a nezajímavý. Za zmínku ovšem stojí fakt, že hodnoty polynomů z rovností (49) a (47) nikdy nejsou čtverce.

Ilustrujme větu 33 na dvou jednoduchých příkladech.

Příklad. Nejprve uvažujme jednoprvkovou symetrickou posloupnost $a_1 = 1$. Potom $k = q_{k-1} = q_k = 1$, $z = 0$. Vidíme, že zde se jedná o případ 2, tedy podle věty 33 po dosazení do (49) a (52) pro každé $N \in \mathbb{N}$ je $\sqrt{N} = [a_0, \overline{1, 2a_0}]$, kde $a_0 = \lfloor \sqrt{N} \rfloor$, právě když platí $N = x^2 + 2x$ pro $x \geqq x_0 = m_0/2$, kde m_0 je nejmenší sudé m vyhovující vztahu (52). Vidíme, že v tomto příkladu je $m_0 = 2$, tedy volíme právě $x \in \mathbb{N}$. Vskutku lze i přímým výpočtem ověřit rovnost, že pro každé $x \in \mathbb{N}$ je $[x, \overline{1, 2x}] = \sqrt{x(x+2)}$.

Obecně lze říct, že volba $k = 1$, kdy posloupnost $(a_1, \ldots, a_k)$ je jednoprvková, nám zajistí, že nastane právě jeden z případů 2, 3, protože potom je $q_{k-1} = 1$ liché a zároveň $z = 0$ je sudé. Zda nastane případ 2, nebo 3 záleží na paritě $q_k = a_1$. Lichá q_k dají případ 2, sudá q_k dají případ 3.

Příklad. Uvažujme symetrickou dvouprvkovou posloupnost $a_1 = a_2 = 1$. Potom $k = q_k = 2$, $z = q_{k-1} = 1$. Vidíme, že jsme narazili na případ 4, proto dle věty 33 neexistuje žádné $N \in \mathbb{N}$ splňující $\sqrt{N} = [a_0, \overline{1, 1, 2a_0}]$, kde $a_0 = \lfloor \sqrt{N} \rfloor$. Dokonce lze výpočtem relativně snadno ukázat, že pro libovolné $x \in \mathbb{N}$ platí $[x, \overline{1, 1, 2x}] = \frac{\sqrt{4x^2+4x+2}}{2}$.

C. Friesen dále ve svém článku [2] ukázal, že pro danou symetrickou posloupnost $(a_1, a_2, \ldots, a_k)$, pro kterou je q_{k-1} nebo z sudé (tedy nastane jeden z případů 1,

2 nebo 3), existuje nekonečně mnoho bezčtvercových přirozených čísel N, pro která platí rovnost (26). V této sekci jsme dokázali o něco slabší tvrzení. Netvrdíme totiž nic o tom, zda nalezená čísla N z věty 33 jsou bezčtvercová.

7. Pellova rovnice

V závěrečné kapitole tohoto článku využijeme znalost řetězového zlomku čísla $\sqrt{N}$, kde N je přirozené číslo, pro nalezení všech celočíselných řešení (x, y) Pellovy rovnice

$$x^2 - Ny^2 = B, \tag{53}$$

kde $B \in \{-1, 1\}$.

7.1. Historie

Již staří Řekové zkoumali Pellovu rovnici pro $N = 2$ v souvislosti s číslem $\sqrt{2}$. Pokud máme řešení (x, y) Pellovy rovnice $x^2 - 2y^2 = \pm 1$, potom x/y je aproximací čísla $\sqrt{2}$. Později Archimédés aproximoval $\sqrt{3}$ zlomkem $1351/780$ a na řešení jisté složité Pellovy rovnice také vede Archimédova úloha o dobytku, o které si lze více přečíst v [1].

Pellovy rovnice byly také zkoumány ve staré Indii: slavný indický matematik Brahmagupta kolem roku 600 n. l. prohlásil, že za matematika se může považovat ten, kdo umí vyřešit rovnici $x^2 - 29y^2 = 1$ (minimálním řešením této rovnice je (9 801, 1 820), přičemž pojem minimálního řešení formálně zavedeme v definici 34). Sám Brahmagupta také přišel s myšlenkou, jak generovat různá řešení Pellovy rovnice pomocí násobení prvků okruhu $\mathbb{Z}[\sqrt{N}]$, kterou podrobněji představíme.

Pellovým rovnicím se také intenzivně věnoval Pierre de Fermat, obzvlášť rovnici $x^2 - 61y^2 = 1$. Minimální řešení této Pellovy rovnice je (1 766 319 049, 226 153 980).

Teorii řešení Pellovy rovnice pomocí řetězových zlomků čísla $\sqrt{N}$, kterou si představíme v části 7.2, vybudoval Joseph-Louis Lagrange v 18. století. Další podrobnosti lze nalézt v [6].

7.2. Řešení Pellovy rovnice

Dá se velice snadno ukázat, že pokud N je čtverec, pak rovnice $x^2 - Ny^2 = -1$ nemá žádné řešení a rovnice $x^2 - Ny^2 = 1$ má právě dvě řešení, a sice $(1, 0)$ a $(-1, 0)$. Tato řešení také nazýváme *triviálními řešeními Pellovy rovnice* $x^2 - Ny^2 = 1$ (triviální řešení uvažujeme i v případě, kdy N není čtverec).

Jestliže N není čtverec, je situace daleko zajímavější, neboť pokud řešení těchto rovnic existují, tak jich je nekonečně mnoho:

Brahmaguptův nápad, jak generovat různá řešení Pellovy rovnice, odpovídá násobení prvků okruhu $\mathbb{Z}[\sqrt{N}]$. Myšlenka, jak řešit Pellovu rovnici $x^2 - Ny^2 = \pm 1$, je rozklad $x^2 - Ny^2 = (x - y\sqrt{N})(x + y\sqrt{N})$. Přesněji, máme-li dvě řešení (x_1, y_1) a (x_2, y_2) Pellovy rovnice $x^2 - Ny^2 = 1$, potom řešením téže rovnice je i (x_3, y_3), kde platí $x_3 + y_3\sqrt{N} = (x_1 + y_1\sqrt{N})(x_2 + y_2\sqrt{N})$. K tomu si stačí uvědomit, že platí $x_3 - y_3\sqrt{N} = (x_1 - y_1\sqrt{N})(x_2 - y_2\sqrt{N})$, a proto

$$\begin{aligned} x_3^2 - Ny_3^2 &= (x_3 - y_3\sqrt{N})(x_3 + y_3\sqrt{N}) = \\ &= (x_1 - y_1\sqrt{N})(x_2 - y_2\sqrt{N})(x_1 + y_1\sqrt{N})(x_2 + y_2\sqrt{N}) = \\ &= (x_1^2 - Ny_1^2)(x_2^2 - Ny_2^2) = 1. \end{aligned}$$

Demonstrujme tuto myšlenku na konkrétním příkladě.

Příklad. Uvažme rovnici $x^2 - 3y^2 = 1$. Jedním řešením této rovnice je dvojice $(2, 1)$. Dále platí $(2 + \sqrt{3})(2 + \sqrt{3}) = 7 + 4\sqrt{3}$, tedy dvojice $(7, 4)$ by také měla být řešením této rovnice, což je vskutku pravda.

Obecně lze říci, že řešení rovnice $x^2 - Ny^2 = \pm 1$ tvoří grupu invertibilních prvků okruhu $\mathbb{Z}[\sqrt{N}]$.

Ukazuje se, že všechna řešení Pellovy rovnice, pokud existují, lze popsat pomocí jediného řešení této rovnice, tzv. minimálního řešení. Nejprve si všimněme, že pokud máme nějaké řešení (a, b) Pellovy rovnice, poté řešením téže Pellovy rovnice jsou i uspořádané dvojice $(a, -b)$, $(-a, -b)$, $(-a, b)$. Proto můžeme bez újmy na obecnosti předpokládat, že $a > 0$, $b > 0$.

Definice 34. Nechť $B \in \{-1, 1\}$ a N je přirozené číslo, které není čtvercem. Předpokládejme, že existuje netriviální řešení rovnice (53). Řekneme, že netriviální řešení (a, b) splňující $a > 0$, $b > 0$ je *minimální řešení* rovnice (53), pokud pro každé řešení (a', b') splňující $a' > 0$, $b' > 0$ platí $a + b\sqrt{N} \leqq a' + b'\sqrt{N}$.

Jak toto minimální řešení nalézt a jak je využít k nalezení všech řešení dané Pellovy rovnice ukážeme v následující větě. Její důkaz nebudeme provádět.

Věta 35. *Nechť $N \in \mathbb{N}$ není čtverec a $\sqrt{N} = [a_0, \overline{a_1, a_2, \ldots, a_k, 2a_0}]$ pro minimální index k. Uvažujme posloupnosti $\{p_n\}_{n=0}^{\infty}$ a $\{q_n\}_{n=0}^{\infty}$ z definice 5. Pro celočíselná (x, y) označme*

$$M(x, y) := \{(c, d) \mid c + d\sqrt{N} = \pm(x + y\sqrt{N})^m \mid m \in \mathbb{Z}\}$$

Pak platí:

1. *Je-li k liché, pak rovnice (53) pro $B = -1$ nemá žádné řešení a pro $B = 1$ je minimální řešení rovno (p_k, q_k). Všechna řešení Pellovy rovnice jsou popsána množinou $M(p_k, q_k)$.*

2. *Je-li k sudé, pak rovnice (53) pro $B = -1$ má minimální řešení (p_k, q_k) a pro $B = 1$ má minimální řešení (p_{2k+1}, q_{2k+1}). Pro množinu $M(p_k, q_k)$ platí, že volbou lichých exponentů m dostaneme všechna řešení Pellovy rovnice pro $B = -1$ a volbou sudých exponentů dostaneme všechna řešení Pellovy rovnice pro $B = 1$. Zároveň platí vztah $(p_k + q_k\sqrt{N})^2 = p_{2k+1} + q_{2k+1}\sqrt{N}$, a proto množina všech řešení pro $B = 1$ je $M(p_{2k+1}, q_{2k+1})$.*

Část důkazu této věty, pojednávající o existenci minimálních řešení, lze nalézt ve [5], sekce 4.8.

Ilustrujme větu na dvou příkladech.

Příklad. Uvažujme rovnici $x^2 - 2y^2 = B$, $B \in \{-1, 1\}$. Tedy platí $N = 2$ a jelikož $\sqrt{2} = [1, \overline{2}]$, tak máme $k = 0$. Proto podle věty 35 máme pro $B = -1$ minimální řešení $(p_0, q_0) = (1, 1)$ a pro $B = 1$ je minimální řešení $(p_1, q_1) = (3, 2)$.

Všechna řešení rovnice $x^2 - 2y^2 = -1$ jsou právě prvky množiny

$$M(1, 1) = \{(c, d) \mid c + d\sqrt{2} = \pm(1 + \sqrt{2})^m \mid m \in \mathbb{Z}\}.$$

Všechna řešení rovnice $x^2 - 2y^2 = 1$ jsou právě prvky množiny

$$M(3, 2) = \{(c, d) \mid c + d\sqrt{2} = \pm(3 + 2\sqrt{2})^m \mid m \in \mathbb{Z}\}.$$

Příklad. Nyní budeme uvažovat rovnici $x^2 - 3y^2 = B$, $B \in \{-1, 1\}$. Zde máme $N = 3$ a $\sqrt{3} = [1, \overline{1, 2}]$, tedy $k = 1$. Proto pro $B = 1$ máme minimální řešení $(p_1, q_1) = (2, 1)$ a všechna řešení rovnice $x^2 - 3y^2 = 1$ jsou právě prvky množiny

$$M(2, 1) = \{(c, d) \mid c + d\sqrt{3} = \pm(2 + \sqrt{2})^m \mid m \in \mathbb{Z}\}.$$

Rovnice $x^2 - 3y^2 = -1$ nemá řešení, což lze také nahlédnout přímo, pohledem na zbytky modulo 3.

S rostoucím N mohou hodnoty vyskytující se v minimálním řešení Pellovy rovnice pro $B = 1$ velice rychle růst. Například pro $N = 13$ máme $\sqrt{13} = [3, \overline{1, 1, 1, 1, 6}]$. Pro rovnici $x^2 - 13y^2 = -1$ máme minimální řešení $(p_4, q_4) = (18, 5)$ a minimální řešení rovnice $x^2 - 13y^2 = 1$ je $(p_9, q_9) = (649, 180)$. Další příklady byly uvedeny v odstavci 7.1.

Poděkování. Rád bych poděkoval Mgr. Vítězslavu Kalovi, Ph.D., za veškerou jeho ochotu a čas věnovaný individuálním konzultacím, dále za cenné připomínky a rady při psaní mé bakalářské práce [4] a tohoto článku, který vznikl její úpravou. Dále bych chtěl velice poděkovat recenzentovi tohoto článku a doc. RNDr. Antonínu Slavíkovi, Ph.D., za upozornění na stylistické i matematické nedostatky původní verze článku.

L i t e r a t u r a

[1] Bártlová, T.: *Archimédova úloha o dobytku.* In: Z. Halas: Archimédés. Několik pohledů do jeho života a díla. MatfyzPress, Praha, 2012, 99–107.

[2] Friesen, C.: *On continued fractions of given period.* Proc. Amer. Math. Soc. 103 (1988), 9–14.

[3] Kala, V.: *Teorie čísel* [online].
Dostupné z: `http://karlin.mff.cuni.cz/~kala/1920 tc/TC skripta.pdf`

[4] Kuděj, M.: *Řetězové zlomky s předepsanou periodou.* Bakalářská práce. MFF UK, 2020.

[5] Olds, C. D.: *Continued fractions.* Random House, New York, 1963.

[6] Pell's equation. Dostupné z: `https://en.wikipedia.org/w/index.php?title=Pell%27s_equation&oldid=987446034`

[7] Stanovský, D.: *Základy algebry.* MatfyzPress, Praha, 2010.

[8] Vít, P.: *Řetězové zlomky.* Mladá fronta, Praha, 1982.