Univerzita Karlova v Praze, Filozofická fakulta
Katedra logiky

David Jurenka

NEROZHODNUTELNOST STRUKTURY RACIONÁLNÍCH ČÍSEL

Diplomová práce

Vedoucí práce: RNDr. Vítězslav Švejdar, CSc.

2006

Prohlašuji, že jsem diplomovou práci vypracoval samostatně s využitím uvedených pramenů a literatury.

V Praze 9. září 2006

David Jurenka

Obsah

1. Úvod

Otázka rozhodnutelnosti, tj. otázka, zda existuje algoritmus, který by byl schopen rozhodnout o platnosti každé prvořádové predikátové formule, se dostala na výsluní pozornosti matematiků ve dvacátých letech minulého století. Spolu s ní byla zkoumána i rozhodnutelnost druhořádových formulí a obecně jakéhokoli matematického tvrzení. Souhrnně byly tyto otázky označovány jako Hilbertův *Entscheidungsproblem* a ještě roku 1930 Hilbert věřil v jejich kladné řešení. Roku 1936 však Alonzo Church ukázal, že samotná predikátová logika prvního řádu je nerozhodnutelná, a téhož roku pak Alan Turing představil dnes již klasický nerozhodnutelný problém, problém zastavení. Oba při tom ve svých pracech využili myšlenek, které formuloval Kurt Gödel ve svém důkazu neúplnosti aritmetiky.

V otázce rozhodnutelnosti základních aritmetických struktur přinesl první významný výsledek Mojżesz Presburger, který roku 1929 dokázal rozhodnutelnost přirozených čísel s operací sčítání a konstantami 0 a 1. Nicméně hned následujícího roku vyplynulo z Gödelových výsledků, že tatáž struktura včetně operace násobení již rozhodnutelná být nemůže. Tím byla zároveň vyřešena i otázka rozhodnutelnosti čísel celých, neboť pojem přirozeného čísla je v této struktuře definovatelný (viz kapitolu 4.2), a tak je možno v celých číslech reformulovat každou aritmetickou sentenci. Na druhou stranu roku 1939 dokázal Alfred Tarski, že teorie reálných čísel připouští eliminaci kvantifikátorů a příslušná struktura reálných čísel je rozhodnutelná. Racionální čísla, ležící alespoň co do inkluze mezi čísly přirozenými a reálnými, tak zůstala posledním elementárním číselným oborem s otevřenou otázkou rozhodnutelnosti. Negativní odpověď na tuto otázku pak přinesla ve své disertaci Julia Robinson [Rob49], která ukázala formuli definující množinu přirozených čísel ve struktuře racionálních čísel s nulou, jedničkou, sčítáním a násobením.

Robinsonové konstrukce je nicméně založena na platnosti několika teorémů z vysoce pokročilé teorie čísel. Cílem této práce je tedy dokázat co možná největší část této konstrukce za použití pouze elementárních prostředků, naznačit možné způsoby důkazu zbylé neelementární části a popsat k tomu potřebný matematický aparát.

V kapitolách 2–5 budou vybudovány jednotlivé prerekvizity potřebné k celkovému důkazu. Vlastní konstrukce pak je obsahem 6. kapitoly, při-

čemž dvě lemmata zde budou dokázána jen z poloviny. Následující kapitola pak nabízí několik možností doplnění důkazu a závěrečná 8. kapitola naznačuje jednu z cest k plnému důkazu nerozhodnutelnosti struktury racionálních čísel.

2. Struktury a definovatelnost

2.1. Struktury

Strukturou nad jazykem $\mathcal{L}$ (zkráceně $\mathcal{L}$-strukturou) budeme rozumět libovolnou neprázdnou množinu spolu s realizacemi všech relačních a operačních symbolů jazyka $\mathcal{L}$ na této množině. V dalším textu budou hrát ústřední roli struktury nad tzv. *aritmetickým jazykem* $\{0, 1, +, \cdot\}$, tj. struktury s konstantami nula a jedna a s operacemi sčítání a násobení. Jde zejména o čtyři základní číselné obory: strukturu $\mathbb{N}$ přirozených čísel včetně nuly, $\mathbb{Z}$ celých, $\mathbb{Q}$ racionálních a $\mathbb{R}$ reálných čísel. Důležitou úlohu v tomto textu textu pak také sehrají dvě obecné třídy struktur nad aritmetickým jazykem.

Definice. *Strukturu nad aritmetickým jazykem nazveme* obor integrity, *jestliže operace sčítání a násobení jsou komutativní a asociativní,* 0 *je neutrální prvek vůči sčítání,* 1 *vůči násobení, násobení je distributivní vzhledem ke sčítání a dále platí následující tři formule.*

$$0 \neq 1$$
$$(\forall a)(\exists b)a + b = 0$$
$$(\forall a, b)(a \cdot b = 0 \rightarrow (a = 0 \vee b = 0))$$

Definice. *Strukturu nad aritmetickým jazykem nazveme* těleso, *jestliže je oborem integrity a navíc v ní platí formule*

$$(\forall a \neq 0)(\exists b)a \cdot b = 1.$$

Inverzní prvek k prvku a vzhledem ke sčítání budeme značit $-a$ a inverzní prvek vzhledem k násobení a^{-1} nebo $\frac{1}{a}$.

Z definic je patrné, že struktury $\mathbb{Q}$ a $\mathbb{R}$ jsou tělesa, $\mathbb{Z}$ těleso není, ale je oborem integrity, a $\mathbb{N}$ není ani obor integrity. Dalším klasickým příkladem těles jsou např. struktury $\mathbb{Z}/p\mathbb{Z}$ zbytkových tříd vzhledem k prvočíselnému modulu p.

2.2. Definovatelnost ve strukturách

Definice. *Nechť* $\mathbb{S}$ *je* $\mathcal{L}$*-struktura s nosičem* S. *Množina* $A \subseteq S$ *je* definovatelná *v* $\mathbb{S}$, *existuje-li* $\mathcal{L}$*-formule s jednou volnou proměnnou* $\varphi(x)$ *taková, že*

$$(\forall a \in S)(a \in A \equiv \mathbb{S} \models \varphi[a]).$$

Jako příklady definovatelných množin a příslušných definujících formulí lze uvést formuli $(\exists y)x = y \cdot y$, která definuje nezáporná čísla v $\mathbb{R}$, formuli $(\forall y)x + y = y$ definující množinu $\{0\}$ ve strukturách $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ i $\mathbb{R}$ či formuli

$$(\exists y)(x = 1 + 1 + y) \,\&\, (\forall y)((\exists z)(x = z \cdot y) \rightarrow (y = 1 \vee y = x)),$$

která definuje prvočísla v $\mathbb{N}$.

V dalším textu bude předvedena formule definující přirozená čísla ve struktuře celých čísel a zejména pak formule, která definuje přirozená čísla v číslech racionálních a která bude pointou celé práce.

3. Kvadratická rezidua

3.1. Úmluva o značení

V tomto textu bude používáno klasické aritmetické značení. $a \equiv b \pmod{m}$ označuje skutečnost, že a je kongruentní s b modulo m, tj. že existuje celé číslo n takové, že platí $a = b + nm$. $a \mid b$ vyjadřuje, že b je dělitelné číslem a, tedy že existuje celé číslo n tak, že platí $b = na$. Tradičním symbolem (a, b) budeme označovat největší společný dělitel čísel a a b, tedy speciálně formule $(a, b) = 1$ vyjadřuje fakt, že čísla a a b jsou nesoudělná. Stejným symbolem budou nicméně značeny i vektory a otevřené intervaly. Z kontextu však bude vždy zřejmé, v jakém smyslu je tento symbol právě užit, a nemělo by tak nikdy dojít k nedorozumění. Pod pojmem prvočíslo budeme rozumět vždy pouze přirozené číslo větší než 1, jehož dělitelem je pouze jednička a ono samo, a to i když se budeme pohybovat v oboru celých čísel nebo v jakékoli jiné číselné struktuře.

Ve zbytku této kapitoly vybudujeme pojem kvadratického rezidua a dokážeme několik souvisejících vět, které povedou k zákonu kvadratické reciprocity.

Tato kapitola spolu s kapitolou následující vychází ze značné míry z knihy [Dav52].

3.2. Kvadratická rezidua

Při studiu kvadratických reziduí se budeme zabývat otázkou, pro jaká a při pevně zvoleném prvočíselném modulu p existuje x takové, že platí

$$x^2 \equiv a \pmod{p},$$

tedy jaká čísla jsou čtverci v tělese $\mathbb{Z}/p\mathbb{Z}$. Taková čísla označíme za *kvadratická rezidua*, přičemž z praktických důvodů se v našich úvahách nebudeme zabývat číslem 0, ačkoli je pro něj výše uvedená rovnice triviálně řešitelná.

Následující tabulka uvádí jako příklad všechny možné čtverce modulo 11.

x	1	2	3	4	5	6	7	8	9	10
x^2	1	4	9	5	3	3	5	9	4	1

Je tedy patrné, že čísla 1, 3, 4, 5 a 9 jsou kvadratickými rezidui modulo 11, naproti tomu čísla 2, 6, 7, 8 a 10 jsou tzv. *kvadratická nonrezidua*.

Při studiu řešitelnosti výše uvedené rovnice se ukáže jako velmi užitečný nástroj pojem *primitivního kořene*. Připomeňme tedy nejprve několik základních pojmů. *Řádem* nenulového prvku a v tělese $\mathbb{Z}/p\mathbb{Z}$ budeme rozumět nejmenší přirozené číslo n takové, že $a^n \equiv 1 \pmod{p}$. Je zřejmé, že řád každého čísla je nejvýše $p-1$, neboť dle malé Fermatovy věty $a^{p-1} \equiv 1 \pmod{p}$. Navíc pro každé prvočíslo existuje tzv. *primitivní kořen*, jehož řád je právě $p-1$. Primitivní kořeny budeme dále označovat písmenem g. Lze snadno nahlédnout, že v posloupnosti $g, g^2, g^3, \ldots, g^{p-1} \pmod{p}$ jsou každá dvě čísla různá, jelikož g^{p-1} je prvním prvkem z této posloupnosti, který je kongruentní s 1. Zároveň se v ní nemůže objevit číslo 0. Tato posloupnost je tedy permutací posloupnosti $1, 2, \ldots, p-1$ a každému číslu $1 \leqslant a \leqslant p-1$ tak můžeme jednoznačně přidělit *index* α vzhledem k primitivnímu kořenu g tak, aby platilo $a \equiv g^\alpha \pmod{p}$. Jedním z primitivních kořenů pro prvočíslo 11 je 2, a rezidua modulo 11 tak můžeme oindexovat následujícím způsobem.

číslo	1	2	3	4	5	6	7	8	9	10
index	10	1	8	2	4	9	7	3	6	5

Za pomoci indexů nyní můžeme zredukovat operaci násobení modulo p na pouhé sčítání (indexy tak fungují podobně jako logaritmy). Mějme čísla a s indexem α a b s indexem β. Pak $a \cdot b \equiv g^\alpha \cdot g^\beta \equiv g^{\alpha+\beta} \pmod{p}$. Platí tedy zřejmě, že index součinu je součtem indexů, případně se liší o $p-1$. Chceme-li tedy vynásobit libovolný počet čísel, stačí vzít jejich indexy (v literatuře lze nalézt tabelované hodnoty), sečíst je, vhodným odečtením násobku $p-1$ zmenšit výsledek tak, aby ležel v intervalu $[1, p-1]$ (tím se výsledná hodnota nezmění, neboť $g^{p-1} \equiv 1$), a na závěr vyhledat číslo, jehož je takto získaný výsledek indexem.

Když se nyní vrátíme k našemu problému s řešitelností rovnice $x^2 \equiv a \pmod{p}$, pak jestliže označíme index čísla a jako α a index x jako ξ, můžeme tuto kongruenci přepsat do tvaru $(g^\xi)^2 \equiv g^\alpha \pmod{p}$. Dvě čísla jsou kongruentní, mají-li stejné indexy, tedy se stačí zabývat otázkou, pro jaká ξ se 2ξ a α liší o násobek $p-1$, čili kongruencí

$$2\xi \equiv \alpha \pmod{p-1}. \tag{1}$$

Otázka kvadratických reziduí modulo 2 je triviální. 1 je kvadratickým reziduem a nulou se dle shora uvedené úmluvy nezabýváme. Předpokládejme tedy v dalším textu, že p je liché prvočíslo. Pak $p-1$ je sudé, a rovnice (1)

tak může mít řešení, pouze pokud je α také sudé. Nechť tedy dále $\alpha = 2\beta$. Pak můžeme rovnici upravit na tvar

$$\xi \equiv \beta \pmod{\tfrac{p-1}{2}},$$

čímž získáme právě jedno řešení $\pmod{\frac{p-1}{2}}$, a právě dvě řešení modulo $p-1$ (to druhé lze získat přičtením $\frac{p-1}{2}$). Shrneme-li celou úvahu, ukázali jsme, že nenulové číslo a je kvadratickým reziduem pro liché prvočíslo p právě tehdy, když má sudý index. Rovnice $x^2 \equiv a \pmod{p}$ má pak právě dvě řešení, jejichž indexy se navíc liší o $\frac{p-1}{2}$. Dále ze vztahu $x^2 \equiv (-x)^2$ je zřejmé, že se tato dvě řešení liší pouze ve znaménku.

Dokázali jsme tedy, že z čísel $1, 2, \ldots, p-1$ je právě polovina (ta se sudými indexy) kvadratickými rezidui a druhá polovina jsou nonrezidua. Navíc platí, že součin dvou reziduí nebo dvou nonreziduí je kvadratickým reziduem, neboť součet dvou indexů se stejnou paritou vzhledem k sudému modulu $p-1$ je sudým číslem. Naopak součin rezidua s nonreziduem dává nonreziduum. Právě tato multiplikativní vlastnost zřejmě vedla Adriena-Marie Legendra k zavedení následujícího symbolu vyjadřujícího kvadratický charakter čísla a vzhledem k prvočíselnému modulu p.

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{je-li } a \text{ kvadratickým reziduem} \pmod{p} \\ -1, & \text{je-li } a \text{ kvadratickým nonreziduem} \pmod{p} \end{cases}$$

Jiným způsobem, jak vyjádřit tuto definici, je položit $\left(\frac{a}{p}\right) = (-1)^{\alpha}$, kde α je indexem a. Výše uvedenou multiplikativní vlastnost pak můžeme vyjádřit jako

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

Poznamenejme, že často bývá technicky výhodné dodefinovat Legendrův symbol i pro a soudělná s p a položit v tomto případě $\left(\frac{a}{p}\right) = 0$.

3.3. Eulerovo kritérium

Podle malé Fermatovy věty platí pro každé a nesoudělné s p kongruence $a^{p-1} \equiv 1 \pmod{p}$. Jelikož je $p-1$ sudé, můžeme rovnici přepsat jako $(a^{\frac{p-1}{2}})^2 \equiv 1 \pmod{p}$. A dále vzhledem k tomu, že jedinými řešeními rovnice $x^2 \equiv 1 \pmod{p}$ jsou čísla 1 a -1, musí pro každé takové a platit

$$a^{\frac{p-1}{2}} \equiv 1, \quad \text{nebo} \quad a^{\frac{p-1}{2}} \equiv -1 \pmod{p}.$$

Leonhard Euler pak dokázal, že rozdíl mezi těmito dvěma možnostmi přesně koresponduje se skutečností, zda a je nebo není kvadratickým reziduem, tedy že platí

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}.$$

Důkaz tohoto tzv. *Eulerova kritéria* už je poměrně snadný. Je-li a kvadratickým reziduem, pak $a \equiv g^{2\beta} \pmod{p}$, kde g je primitivní kořen prvočísla p, a tedy $a^{\frac{p-1}{2}} \equiv g^{(p-1)\beta} \equiv 1^{\beta} \equiv 1 \pmod{p}$. Je-li naopak a nonreziduem, a jeho index α je tedy lichý, pak $\alpha\frac{p-1}{2}$ nemůže být násobkem $p-1$, a proto $a^{\frac{p-1}{2}} \equiv g^{\alpha\frac{p-1}{2}} \not\equiv 1$, pročež musí být kongruentní s -1.

Na základě právě dokázaného kritéria můžeme snadno rozhodnout, pro která prvočísla je -1 kvadratickým reziduem. Platí totiž $\left(\frac{-1}{p}\right) \equiv (-1)^{\frac{p-1}{2}}$, a číslo -1 tedy bude reziduem právě pro ta prvočísla p, pro něž je $\frac{p-1}{2}$ sudé. To nastává pro prvočísla tvaru $4k+1$. Naopak pro prvočísla tvaru $4k+3$ bude -1 kvadratickým nonreziduem.

3.4. Gaussovo lemma

Další jednoduché pravidlo pro zjišťování kvadratického charakteru libovolného $a \not\equiv 0$ vzhledem k modulu p nabízí tzv. *Gaussovo lemma*, které nyní vyslovíme a dokážeme.

Vezměme čísla

$$a, 2a, 3a, \ldots, \tfrac{p-1}{2}a$$

a odečtením vhodného násobku p je zredukujme tak, aby ležela v intervalu $(-\frac{p}{2}, \frac{p}{2})$. Označme výslednou množinu R a nechť ν je počet záporných čísel v této množině. Gaussovo lemma pak tvrdí, že $\left(\frac{a}{p}\right) = (-1)^{\nu}$, tzn. a je kvadratické reziduum, je-li ν sudé, a nonreziduum, je-li liché.

Důkaz je poměrně snadný. Je zřejmé, že každé z čísel z množiny R je rovno jednomu z čísel $\pm1, \pm2, \pm3, \ldots, \pm\frac{p-1}{2}$, neboť tak bylo R definováno. Zároveň platí, že každé z čísel $1, 2, \ldots, \frac{p-1}{2}$ se v R vyskytuje nejvýše jednou, a to buď s kladným, nebo záporným znaménkem. Neboť pokud by se vyskytlo dvakrát se stejným znaménkem, znamenalo by to, že pro nějaká dvě různá m a n z intervalu $\left[1, \frac{p-1}{2}\right]$ platí $ma \equiv na \pmod{p}$, tedy $(m-n)a \equiv 0 \pmod{p}$. Což není možné, neboť $\mathbb{Z}/p\mathbb{Z}$ je obor integrity a a i $(m-n) \not\equiv 0$. Podobně kdyby se nějaké z čísel vyskytlo v R zároveň s kladným i záporným znaménkem, pak $ma \equiv -na$, čili $(m+n)a \equiv 0$, což opět nemůže nastat, jelikož m i $n \leqslant \frac{p-1}{2}$. Množina R je tedy rovna množině $\pm1, \pm2, \pm3, \ldots, \pm\frac{p-1}{2}$,

v níž má každé z čísel jedno určité znaménko. Pokud vzájemně vynásobíme prvky obou množin, dostaneme

$$a \cdot 2a \cdot 3a \cdot \ldots \cdot \tfrac{p-1}{2}a \equiv \pm 1 \cdot \pm 2 \cdot \pm 3 \cdot \ldots \cdot \pm \tfrac{p-1}{2} \pmod{p}.$$

Obě strany pak můžeme vydělit $\frac{p-1}{2}!$, čímž získáme

$$a^{\frac{p-1}{2}} \equiv \pm 1 \cdot \pm 1 \cdot \pm 1 \cdot \ldots \cdot \pm 1 = (-1)^{\nu},$$

kde ν je počet negativních znamének v množině R. Tím je na základě Eulerova kritéria lemma dokázáno.

Podobně jako jsme v minulé kapitole na základě Eulerova kritéria určili, pro která prvočísla je -1 kvadratickým reziduem, můžeme nyní na základě Gaussova lemmatu rozhodnout o kvadratickém charakteru čísla 2. Pokud vezmeme $a = 2$, dostáváme posloupnost

$$2, 4, 6, \ldots, p-1.$$

Nyní máme určit, kolik z těchto čísel bude záporných po zredukování do intervalu $(-\frac{p}{2}, \frac{p}{2})$ odečítáním násobků p. Všechna čísla jsou z intervalu $[1, p-1]$, a je tedy zřejmé, že záporná budou po redukci právě ta, která jsou větší než $\frac{p}{2}$. Stačí nám tedy pouze zjistit, kolik sudých čísel $2x$ splňuje nerovnost $\frac{p}{2} < 2x < p$, tedy ekvivalentně, pro kolik přirozených čísel x platí $\frac{p}{4} < x < \frac{p}{2}$. Prvočíslo p je jednoho z tvarů $8k + r$, kde $r = 1, 3, 5,$ nebo 7. Dosazením do rovnice tedy získáváme

$$2k + \frac{r}{4} < x < 4k + \frac{r}{2}.$$

Vzhledem k tomu, že nás zajímá, pouze zda je počet x vyhovujících této nerovnosti sudý, nebo lichý, můžeme z obou stran nerovnice odstranit sudá čísla $2k$ a $4k$, neboť tato změna paritu počtu řešení neovlivní. Dostáváme tedy $\frac{r}{4} < x < \frac{r}{2}$. Tato nerovnost nemá žádné řešení pro $r = 1$, jedno řešení má pro $r = 3, 5$ a dvě řešení pro $r = 7$. V prvním a posledním případě tedy dvojka je kvadratickým reziduem a ve dvou prostředních případech reziduem není. Celkově tak můžeme uzavřít, že 2 je kvadratickým reziduem pro prvočísla tvaru $8k \pm 1$ a nonreziduem pro prvočísla typu $8k \pm 3$.

Zcela stejnou úvahou můžeme stanovit, že 3 je kvadratickým reziduem pro prvočísla tvaru $12k \pm 1$ a nonreziduem pro prvočísla tvaru $12k \pm 5$ nebo že číslo 5 je kvadratickým reziduem pro prvočísla tvaru $20k \pm 1$ a $20k \pm 9$ a naopak nonreziduem pro prvočísla typu $20k \pm 3$ a $20k \pm 7$.

Mezi právě uvedenými nutnými a postačujícími podmínkami pro to, aby číslo a (zde čísla 2, 3 a 5) bylo kvadratickým reziduem modulo p, lze vysledovat jeden společný rys. Pakliže jsou prvočísla p a q kongruentní (mod $4a$), pak $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$. Navíc a má stejný kvadratický charakter pro p a q i v případě, kdy $p \equiv -q \pmod{4a}$. Tato vlastnost platí obecně a je předmětem následujícího lemmatu.

Lemma. *Nechť a je libovolné přirozené číslo. Vyjádřeme p ve tvaru $4ak + r$, kde pro zbytek r platí $0 < r < 4a$. Pak kvadratický charakter a (mod p) je stejný pro všechna p, jejichž hodnoty zbytku r se rovnají, a tento charakter je navíc stejný i pro prvočísla se zbytky r a $4a - r$.*

Toto lemma dokážeme postupem zcela obdobným tomu, který jsem použili při vyšetřování kvadratického charakteru čísla 2.

Vezměme tedy dle Gaussova lemmatu posloupnost

$$a, 2a, 3a, \ldots, \tfrac{p-1}{2}a$$

a ptejme se, kolik z těchto čísel bude po redukci do intervalu $\left(-\frac{p}{2}, \frac{p}{2}\right)$ záporných. Zjevně to budou právě ta čísla, která nyní leží mezi $\frac{p}{2}$ a p, mezi $\frac{3}{2}p$ a $2p$ a tak dále. Položme b rovno tomu nejmenšímu přirozenému číslu, pro něž platí $\frac{p-1}{2}a < bp$ (chceme, aby bp majorizovalo shora uvedenou posloupnost). Číslo b bude rovno $\frac{a}{2}$, nebo $\frac{a-1}{2}$ podle toho, zda je a sudé, nebo liché. Posledním intervalem, kterým je se třeba zabývat, je tak zřejmě interval od $(b - \frac{1}{2})p$ do bp. Otázka tedy zní, kolik násobků čísla a leží v intervalech

$$\left(\tfrac{p}{2}, p\right), \left(\tfrac{3}{2}p, 2p\right), \ldots, \left((b - \tfrac{1}{2})p, bp\right).$$

Tyto intervaly můžeme vzít otevřené, neboť žádné z jejich hraničních čísel nemůže být samo násobkem čísla a. Levé meze totiž vůbec nejsou celými čísly a pravé meze jsou všechny tvaru mp, kde $m \leqslant \frac{a}{2}$. Pokud nyní celý řádek vydělíme číslem a, převedeme tak problém na otázku kolik celých čísel se nachází v intervalech

$$\left(\tfrac{p}{2a}, \tfrac{p}{a}\right), \left(\tfrac{3p}{2a}, \tfrac{2p}{a}\right), \ldots, \left(\tfrac{(2b-1)p}{2a}, \tfrac{bp}{a}\right).$$

Nyní vyjádříme p ve tvaru $4ak + r$. Jelikož se ve všech jmenovatelích vyskytuje a, lze nahlédnout, že nahrazení p výrazem $4ak + r$ je totéž, jako kdybychom nahradili všechny výskyty p číslem r, pouze s tím rozdílem, že

k hraničním bodům intervalů jsou přičtena jistá sudá čísla. Jelikož nás ale zajímá pouze parita počtu celých čísel v těchto intervalech, můžeme tato sudá čísla ignorovat. Dostáváme tak

$$\left(\tfrac{r}{2a}, \tfrac{r}{a}\right), \left(\tfrac{3r}{2a}, \tfrac{2r}{a}\right), \ldots, \left(\tfrac{(2b-1)r}{2a}, \tfrac{br}{a}\right). \tag{2}$$

Označíme-li si jako ν počet celých čísel v těchto intervalech, pak dle Gaussova lemmatu je a kvadratickým reziduem, právě když je ν sudé. Tato skutečnost zřejmě nezávisí na původní hodnotě p, ale pouze na r. Tím je dokázána první část lemmatu.

Nyní zvažme situaci, když by zbytek prvočísla p byl $4a - r$ místo r. Dospěli bychom pak místo (2) k intervalům

$$\left(2 - \tfrac{r}{2a}, 4 - \tfrac{r}{a}\right), \left(6 - \tfrac{3r}{2a}, 8 - \tfrac{2r}{a}\right), \ldots, \left(4b - 2 - \tfrac{(2b-1)r}{2a}, 4b - \tfrac{br}{a}\right). \tag{3}$$

Nyní si již jen stačí všimnout, že parita počtu celých čísel v libovolném i-tém intervalu z (3) je stejná jako u i-tého intervalu z (2). Vezměme jako příklad první interval $\left(2 - \frac{r}{2a}, 4 - \frac{r}{a}\right)$. Odečtením irelevantních sudých čísel dostaneme interval $\left(-\frac{r}{2a}, -\frac{r}{a}\right)$ (vhodnějším zápisem by samozřejmě bylo $\left(-\frac{r}{a}, -\frac{r}{2a}\right)$, neboť $-\frac{r}{2a} > -\frac{r}{a}$). Zrcadlovým převrácením tohoto intervalu na číselné přímce podle počátku pak získáme interval, který bude zřejmě obsahovat stejný počet celých čísel a který navíc bude identický prvnímu intervalu z (2). Tím je dle Gaussova lemmatu celé tvrzení dokázáno.

3.5. Kvadratická reciprocita

Zákon kvadratické reciprocity, který v této podkapitole dokážeme, je klasickým a nepostradatelným nástrojem při výpočtu hodnot Legendrova symbolu. Zároveň jde o jednu z nejslavnějších vět celé teorie čísel. Pro libovolná různá lichá prvočísla p a q dává do souvislosti hodnoty $\left(\frac{p}{q}\right)$ a $\left(\frac{q}{p}\right)$. Říká, že tato čísla jsou vždy stejná až na případ, kdy p i q jsou tvaru $4k + 3$, kdy jsou opačná. Formálně tedy můžeme tento teorém vyslovit následovně.

Věta (zákon kvadratické reciprocity). *Nechť p a q jsou různá lichá prvočísla. Pak*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}.$$

Exponent na pravé straně rovnice je lichý (a $\left(\frac{p}{q}\right)$ a $\left(\frac{q}{p}\right)$ jsou tedy různá čísla) pouze v případě, kdy $\frac{p-1}{2}$ i $\frac{q-1}{2}$ jsou lichá, což nastává právě tehdy, když jsou obě prvočísla tvaru $4k + 3$.

Zákon kvadratické reciprocity nyní již snadno odvodíme z výsledku ze závěru předchozí podkapitoly o kvadratickém charakteru libovolného pevně zvoleného čísla vůči různým prvočíselným modulům.

Předpokládejme nejprve, že $p \equiv q \pmod 4$. Existuje tedy a takové, že $p = q + 4a$ a $q = p - 4a$. Pak platí

$$\left(\frac{p}{q}\right) = \left(\frac{q+4a}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{a}{q}\right).$$

Podobně platí

$$\left(\frac{q}{p}\right) = \left(\frac{p-4a}{p}\right) = \left(\frac{-4a}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{a}{p}\right).$$

Pokud obě rovnice vynásobíme, získáváme

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{a}{p}\right)\left(\frac{a}{q}\right).$$

Jelikož čísla p a q nechávají stejný zbytek po dělení číslem $4a$, budou dle lemmatu z předchozí podkapitoly $\left(\frac{a}{p}\right)$ a $\left(\frac{a}{q}\right)$ stejné, a jejich součin tak bude 1. Jelikož -1 je kvadratickým reziduem pouze pro prvočísla tvaru $4k+1$, výraz

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \left(\frac{-1}{p}\right)$$

bude roven číslu 1, pakliže p (a s ním i q) je tvaru $4k+1$, a číslu -1 v případě, že p a q jsou tvaru $4k+3$.

Nechť naopak $p \not\equiv q \pmod 4$. V tom případě $p \equiv -q \pmod 4$, a existuje tedy a, pro něž platí $p = 4a - q$ a $q = 4a - p$. Pak platí

$$\left(\frac{p}{q}\right) = \left(\frac{4a-q}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{a}{q}\right)$$
$$\left(\frac{q}{p}\right) = \left(\frac{4a-p}{p}\right) = \left(\frac{4a}{p}\right) = \left(\frac{a}{p}\right).$$

Výrazy $\left(\frac{a}{p}\right)$ a $\left(\frac{a}{q}\right)$ si jsou nyní opět rovny, neboť hodnoty p a q modulo $4a$ se liší pouze ve znaménku. Tedy $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$.

Ačkoli v dalším textu právě dokázaný zákon nikdy nepoužijeme ke konkrétním výpočtům, ilustrujme si jeho sílu např. při výpočtu hodnoty $\left(\frac{31}{103}\right)$.

$$\left(\frac{31}{103}\right) = -\left(\frac{103}{31}\right) = -\left(\frac{10}{31}\right) = -\left(\frac{2}{31}\right)\left(\frac{5}{31}\right)$$

A dále jelikož $31 \equiv -1 \pmod 8$, a 2 je tak kvadratickým reziduem modulo 31, dostáváme

$$-\left(\frac{2}{31}\right)\left(\frac{5}{31}\right) = -\left(\frac{5}{31}\right) = -\left(\frac{31}{5}\right) = -\left(\frac{1}{5}\right) = -1.$$

31 je tedy kvadratickým nonreziduem modulo 103.

4. Sumy čtverců

V této kapitole se budeme zabývat otázkou, která přirozená čísla jsou reprezentovatelná jako suma dvou, resp. čtyř čtverců, tedy otázkou, pro jaká n mají rovnice

$$n = x^2 + y^2$$
$$n = a^2 + b^2 + c^2 + d^2$$

řešení v oboru celých čísel. Odpověď na první otázku bude v dalším textu několikrát využita, řešení druhé otázky pak povede k formuli definující množinu přirozených čísel v $\mathbb{Z}$.

4.1. Součet dvou čtverců

Je snadné ukázat, že některá čísla nemohou být reprezentovatelná jako součet dvou čtverců. Druhá mocnina každého sudého čísla je kongruentní s 0 modulo 4, druhá mocnina každého lichého čísla pak s 1. $x^2 + y^2$ tedy může být kongruentní pouze s 0, 1 nebo 2 (mod 4). Proto je patrné, že čísla tvaru $4k + 3$ reprezentovat nelze.

V této úvaze můžeme jít ještě o něco dále. Nechť n je reprezentovatelné jako $x^2 + y^2$ a nechť p je prvočíslo tvaru $4k+3$, které dělí n. Pak $x^2 + y^2 \equiv 0 \pmod{p}$, tedy $x^2 \equiv -y^2 \pmod{p}$. Jestliže $p \nmid y$, můžeme rovnici dále upravit na $x^2(y^{-1})^2 \equiv -1 \pmod{p}$. Jelikož je ale p tvaru $4k + 3$, -1 nemůže být kvadratickým reziduem. Proto nutně $p \mid y$, a potažmo tak i $p \mid x$, a tedy $p^2 \mid n$ a celou rovnici tak můžeme vydělit p^2. Platí tedy $n = p^2 n'$, a pokud je n' stále dělitelné p, můžeme celou úvahu opakovat, dokud nedojdeme k nejvyšší mocnině p, která dělí n a která tak musí být sudá. Je-li tedy n reprezentovatelné jako suma dvou čtverců, každý prvočíselný dělitel tvaru $4k + 3$ čísla n se musí vyskytovat v jeho prvočíselném rozkladu právě na sudou mocninu.

Tato podmínka zahrnuje i podmínku předchozí, neboť prvočísla tvaru $4k+3$ na sudou mocninu jsou kongruentní s 1 (mod 4), stejně jako prvočísla tvaru $4k + 1$ na libovolnou mocninu. Jejich součin spolu s libovolnou mocninou dvojky tedy nemůže být kongruentní s 3 modulo 4. Každé číslo tvaru $4k + 3$ tedy musí ve svém rozkladu obsahovat nějaké prvočíslo tvaru $4k + 3$ na lichou mocninu.

Ve zbytku podkapitoly ukážeme, že právě získaná podmínka pro reprezentovatelnost součtem dvou čtverců už je postačující. Směřujeme tedy k důkazu následujícího tvrzení.

Věta (o součtu dvou čtverců). *Libovolné přirozené číslo lze zapsat ve tvaru součtu dvou čtverců právě tehdy, když každé prvočíslo tvaru* $4k+3$, *které je jeho dělitelem, se v jeho rozkladu objevuje právě na sudou mocninu.*

Naším východiskem bude tzv. *Brahmagupta–Fibonacciho identita*

$$(a^2+b^2)(c^2+d^2) = (ac+bd)^2 + (ad-bc)^2,$$

která zaručuje, že číslo, které vznikne postupným vynásobením čísel reprezentovatelných jako sumy dvou čtverců, je samo součtem dvou čtverců.

Každé číslo n, v jehož rozkladu se prvočísla tvaru $4k+3$ vyskytují pouze na sudou mocninu, lze získat jako součin, jehož každý jednotlivý člen je buď 2, nebo prvočíslo tvaru $4k+1$, nebo prvočíslo tvaru $4k+3$ umocněné na druhou. Pokud ukážeme, že čísla ze všech těchto tří skupin lze vyjádřit jako součin dvou čtverců, můžeme na základě výše uvedené identity uzavřít, že i n je reprezentovatelné jako součet dvou čtverců. Číslo 2 lze snadno zapsat jako 1^2+1^2, stejně tak druhou mocninu prvočísla p tvaru $4k+3$ lze zapsat jako p^2+0^2. Stačí tedy pouze ukázat, že pro každé prvočíslo p tvaru $4k+1$ existují celá čísla x a y taková, že $p = x^2+y^2$.

Důkaz bude proveden ve dvou krocích. Nejprve ukážeme, že existuje nějaký násobek čísla p reprezentovatelný jako z^2+1, a ve druhém kroku pak z tohoto odvodíme, že i p samo lze zapsat jako x^2+y^2.

Otázka, zda existují m a z taková, aby platilo $mp = z^2+1$, je ekvivalentní řešitelnosti rovnice

$$z^2+1 \equiv 0 \pmod{p}.$$

Z minulé kapitoly víme, že pro prvočísla tvaru $4k+1$ je číslo -1 kvadratickým reziduem, a proto takové z zřejmě existuje. Fixujme jedno takové z z intervalu $\left[-\frac{p-1}{2}, \frac{p-1}{2}\right]$ a k němu příslušné m tak, že $mp = z^2+1$. Platí tedy

$$m = \frac{z^2+1}{p} < \frac{\frac{p^2}{4}+1}{p} < p.$$

Pro účely dalšího postupu můžeme tento výsledek poněkud oslabit a vyjít pouze z toho, že máme $m < p$, pro něž platí

$$mp = x^2+y^2. \tag{1}$$

Myšlenka dalšího důkazu je taková, že pokud $m > 1$, pak můžeme najít $m' < m$, $m' \neq 0$ se stejnými vlastnostmi. Po konečném počtu kroků tak dorazíme až k jedničce, což znamená, že i samo p je reprezentovatelné jako součet dvou čtverců.

Vyjděme z rovnice (1) a vezměme čísla u, v z intervalu $\left(-\frac{m}{2}, \frac{m}{2}\right]$ (interval bereme zprava uzavřený, neboť m může být sudé) taková, že

$$\begin{aligned} u &\equiv x \pmod{m} \\ v &\equiv y \pmod{m}. \end{aligned} \tag{2}$$

Pak platí

$$u^2 + v^2 \equiv x^2 + y^2 \equiv 0 \pmod{m},$$

čili existuje r takové, že

$$mr = u^2 + v^2. \tag{3}$$

Číslo r nemůže být nula, neboť pak by i hodnoty u a v byly nula, čili x a y by byly dělitelné číslem m. Rovnice (1) by pak byla dělitelná m^2, což by byl spor se skutečností, že p je prvočíslo. Dále platí

$$r = \frac{u^2 + v^2}{m} \leqslant \frac{\frac{m^2}{4} + \frac{m^2}{4}}{m} = \frac{m}{2} < m.$$

Pokud nyní vzájemně vynásobíme rovnice (1) a (3) a použijeme Brahmagupta–Fibonacciho identitu, získáme

$$m^2 rp = (x^2 + y^2)(u^2 + v^2) = (xu + yv)^2 + (xv - yu)^2. \tag{4}$$

Nyní je důležité si všimnout, že jak $xu + yv$, tak $xv - yu$ jsou dělitelné číslem m, neboť dle (2) platí

$$\begin{aligned} xu + yv &\equiv x^2 + y^2 \equiv 0 \pmod{m} \\ xv - yu &\equiv xy - yx \equiv 0 \pmod{m}. \end{aligned}$$

Rovnici (4) tedy můžeme vydělit číslem m^2 a získáme tak

$$rp = x'^2 + y'^2$$

pro nějaká celá čísla x' a y', což spolu s faktem, že $r < m$ a $r \neq 0$, stačí, jak bylo řečeno výše, k dokončení důkazu.

4.2. Věta o čtyřech čtvercích

Zatímco ne každé přirozené číslo lze zapsat jako součet dvou čtverců, u sumy čtyř čtverců již žádné omezující podmínky nejsou, a platí tedy následující věta.

Věta (o čtyřech čtvercích). *Každé přirozené číslo lze zapsat jako součet čtyř čtverců. Tedy rovnice*

$$n = a^2 + b^2 + c^2 + d^2$$

má celočíselné řešení pro každé přirozené číslo n.

Tuto větu dokázal Joseph-Louis Lagrange kolem roku 1770. My zde však podáme o něco snazší Eulerův důkaz, jehož postup bude navíc zcela obdobný tomu z předcházející podkapitoly.

Tentokrát vyjdeme z identity

$$\begin{aligned}(a^2 + b^2 + c^2 + d^2)(w^2 + x^2 + y^2 + z^2) = {} & (aw + bx + cy + dz)^2 \\ & + (ax - bw - cz + dy)^2 \\ & + (ay + bz - cw - dx)^2 \\ & + (az - by + cx - dw)^2,\end{aligned}$$

kterou objevil Euler a která – podobně jako Brahmagupta–Fibonacciho identita v minulém případě – redukuje problém na otázku reprezentovatelnosti prvočísel. V předchozí kapitole bylo ukázáno, že číslo 2 a prvočísla tvaru $4k + 1$ lze vyjádřit jako součet dvou, a tím pádem snadno i čtyř čtverců, neboť za třetí a čtvrtý čtverec lze vzít 0^2. K dokázání věty tedy stačí ukázat, že i každé prvočíslo p tvaru $4n + 3$ lze zapsat jako součet čtyř čtverců.

Stejně jako v předchozí kapitole nejprve dokážeme, že existuje násobek mp čísla p, kde $0 < m < p$, který lze zapsat jako sumu čtyř čtverců, a následně ukážeme, že toto m lze zredukovat na jedničku.

Pro důkaz první části stačí dokázat, že rovnice

$$x^2 + y^2 + 1 = 0 \pmod{p} \tag{5}$$

má řešení, neboť pak stačí vzít příslušná x a y z intervalu $\left[-\frac{p-1}{2}, \frac{p-1}{2}\right]$ a platí $mp = x^2 + y^2 + 1^2 + 0^2$, kde

$$m = \frac{x^2 + y^2 + 1^2}{p} < \frac{\frac{p^2}{4} + \frac{p^2}{4} + 1}{p} = \frac{p}{2} + \frac{1}{p} < p.$$

Rovnici (5) můžeme přepsat do tvaru

$$x^2 + 1 = -y^2 \pmod{p}.$$

Z minulé kapitoly víme, že -1 je kvadratické nonreziduum pro každé prvočíslo tvaru $4k+3$. Číslo $-y^2$ tedy bude také kvadratickým nonreziduem a navíc každé nonreziduum lze zapsat v tomto tvaru. Naproti tomu jakékoli nenulové x^2 je reziduem. Stačí tedy nalézt takové reziduum r a nonreziduum n, že $n = r+1$. Vyjděme tedy od jedničky, která je reziduem, a postupně procházejme čísla $1, 2, \ldots p-1$, dokud nenarazíme na první nonreziduum, kterých je $\frac{p-1}{2}$. Předchůdce prvního nonrezidua je zjevně reziduum a tím je první část důkazu hotova.

Nyní již jen stačí ukázat, že jestliže mp je reprezentovatelné jako

$$mp = a^2 + b^2 + c^2 + d^2, \tag{6}$$

kde $1 < m < p$, pak existuje nenulové m' menší než m, které má tutéž vlastnost. Vezměme čísla e, f, g, h z intervalu $\left(-\frac{m}{2}, \frac{m}{2}\right]$ taková, že

$$\begin{aligned} e &\equiv a \pmod{m} \\ f &\equiv b \pmod{m} \\ g &\equiv c \pmod{m} \\ h &\equiv d \pmod{m}. \end{aligned} \tag{7}$$

Stejně jako minule platí

$$e^2 + f^2 + g^2 + h^2 \equiv a^2 + b^2 + c^2 + d^2 \equiv 0 \pmod{m},$$

čili existuje r takové, že

$$mr = e^2 + f^2 + g^2 + h^2. \tag{8}$$

Číslo r opět nemůže být nula, neboť pak by i hodnoty e, f, g, h byly nula, čili a, b, c, d by byly dělitelné číslem n, a celá rovnice (6) by pak byla dělitelná m^2, což by byl spor se skutečností, že p je prvočíslo.

Dále platí

$$r = \frac{e^2 + f^2 + g^2 + h^2}{m} \leqslant \frac{\frac{m^2}{4} + \frac{m^2}{4} + \frac{m^2}{4} + \frac{m^2}{4}}{m} = m.$$

My však potřebujeme, aby r bylo ostře menší než m. Rovnost $r = m$ nastává pouze v případě, kdy m je sudé a všechna čísla e, f, g, h jsou rovna $\frac{m}{2}$. Tedy

dle (7) $a = ml + \frac{m}{2}$ pro nějaké l. Tudíž $a^2 = m^2l^2 + m^2l + \frac{m^2}{4}$, což znamená $a^2 \equiv \frac{m^2}{4} \pmod{m^2}$. Totéž pak platí i pro b, c a d. Ve výsledku tedy

$$a^2 + b^2 + c^2 + d^2 \equiv \frac{m^2}{4} + \frac{m^2}{4} + \frac{m^2}{4} + \frac{m^2}{4} \equiv 0 \pmod{m^2}$$

a rovnice (6) by byla dělitelná m^2, což by opět bylo ve sporu s prvočíselností p. Platí tedy $0 < r < m$.

Pokud nyní vzájemně vynásobíme rovnice (6) a (8) a použijeme Eulerovu identitu ze začátku podkapitoly, získáme

$$\begin{aligned} m^2rp = (a^2 + b^2 + c^2 + d^2)(e^2 + f^2 + g^2 + h^2) &= (ae + bf + cg + dh)^2 \\ &+ (af - be - ch + dg)^2 \\ &+ (ag + bh - ce - df)^2 \\ &+ (ah - bg + cf - de)^2 \end{aligned} \tag{9}$$

Všechny čtyři sčítance na pravé straně rovnice (9) jsou dělitelné číslem m^2, neboť dle (7) platí

$$\begin{aligned} ae + bf + cg + dh &\equiv a^2 + b^2 + c^2 + d^2 \equiv 0 \pmod{m} \\ af - be - ch + dg &\equiv ab - ba - cd + dc \equiv 0 \pmod{m} \\ ag + bh - ce - df &\equiv ac + bd - ca - db \equiv 0 \pmod{m} \\ ah - bg + cf - de &\equiv ad - bc + cb - da \equiv 0 \pmod{m}, \end{aligned}$$

a tedy může rovnici (9) číslem m^2 vydělit a získat tak

$$rp = a'^2 + b'^2 + c'^2 + d'^2,$$

kde a', b', c', d' jsou celá čísla, což spolu s faktem, že $r < m$ a $r \neq 0$, dokončuje důkaz.

Ukázali jsme tedy, že každé nezáporné celé číslo lze vyjádřit jako součet čtyř celočíselných čtverců. Vzhledem ke skutečnosti, že číslo vyjádřitelné jako součet čtverců je jistě nezáporné, získali jsem tak i formuli, která definuje množinu přirozených čísel ve struktuře celých čísel s operacemi sčítání a násobení. Pro každé celé číslo x platí, že patří do množiny přirozených čísel, právě když splňuje formuli

$$(\exists a, b, c, d)x = a^2 + b^2 + c^2 + d^2.$$

5. Kvadratické formy

V tomto oddílu zavedeme třídu polynomů označovanou jako *kvadratické formy*. Nás budou především zajímat kvadratické formy definované nad strukturou $\mathbb{Q}$ (tzv. racionální kvadratické formy), nicméně vzhledem k tomu, že nás to nebude stát žádné úsilí navíc, budeme se v této kapitole zabývat kvadratickými formami obecně nad libovolným tělesem.

Tato kapitola zavádí základní pojmy potřebné při důkazu neelementární části věty o nerozhodnutelnosti racionálních čísel. Při důkazu elementární části použijeme z této kapitoly pouze závěrečné lemma, navíc jen jeho speciální případ, který je snadno dokazatelný bez jakýchkoli definic. Zjednodušený důkaz tohoto lemmatu proto bude na příslušném místě podán v poznámce pod čarou, a tuto kapitolu je tak možné přeskočit a případně se k ní vrátit před čtením rozšiřující 8. kapitoly.

Definice. Kvadratickou formou n proměnných nad tělesem $\mathbb{F}$ *rozumíme libovolnou funkci danou předpisem*

$$Q(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j,$$

kde koeficienty a_{ij} *jsou prvky* $\mathbb{F}$.

Ilustrujme si nejprve tuto definici na několika příkladech. Reálná funkce $f(x, y) = x^2 + 2xy + y^2$ přiřazující libovolným dvěma číslům čtverec jejich součtu je příkladem kvadratické formy dvou proměnných nad tělesem $\mathbb{R}$. Klasickou euklidovskou metriku přiřazující dvěma bodům v rovině jejich vzájemnou vzdálenost definovanou jako

$$\mathrm{dist}([x_1, y_1], [x_2, y_2]) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

pak lze chápat jako odmocninu z kvadratické formy Q_{dist} čtyř proměnných dané předpisem

$$Q_{\mathrm{dist}}(x_1, x_2, y_1, y_2) = x_1^2 - 2x_1x_2 + x_2^2 + y_1^2 - 2y_1y_2 + y_2^2.$$

Definici kvadratických forem pak vyhovuje např. i funkce tří proměnných $Q(x, y, z) = 0$, ačkoli je zřejmé, že tato kvadratická forma je poněkud „degenerovaná“ (tento pojem záhy zpřesníme).

Při prozkoumání definice a pohledu na ilustrační příklady si lze všimnout, že kvadratické formy jsou právě takové polynomy, jejichž každý člen

(monom) obsahuje právě dva výskyty proměnných. Všechny členy tak mají stejný, druhý stupeň. Polynomy, jejichž všechny členy mají stejný stupeň, jsou označovány jako *homogenní*, a kvadratické formy lze tedy ekvivalentně definovat jako homogenní polynomy druhého stupně.

Uvažujeme-li koeficienty a_{ij} jako čtvercovou matici A typu $n \times n$ a vektor proměnných $\mathbf{x} = (x_1, \ldots, x_n)$ jako matici typu $n \times 1$, můžeme kvadratickou formu $Q(x_1, \ldots, x_n)$ vyjádřit jako součin matic $\mathbf{x}^T \cdot A \cdot \mathbf{x}$, kde $\mathbf{x}^T$ je transponovaná matice k matici $\mathbf{x}$, tj. jako součin

$$\begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} \cdot \begin{bmatrix} a_{11} \ldots a_{1n} \\ \vdots \\ a_{n1} \ldots a_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Matice koeficientů příslušná k výše uvedené kvadratické formě $Q_{dist}(x_1, x_2, y_1, y_2)$ by tedy měla následující tvar.

$$\begin{bmatrix} 1 & -2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Obraťme nyní pozornost k obecnému tělesu $\mathbb{F}$, nad nímž jsme na začátku kapitoly definovali kvadratické formy. Jedním ze základních třídících znaků pro tělesa je jejich takzvaná *charakteristika*. Charakteristika tělesa $\mathbb{F}$ je definována jako to nejmenší přirozené číslo n, pro něž platí

$$\mathbb{F} \models \overbrace{1 + 1 + \cdots + 1}^{n\text{-krát}} = 0.$$

Pro tělesa, ve kterých nelze sčítáním jedniček získat nulu, se charakteristika pokládá rovna 0. Takovými tělesy jsou například struktury $\mathbb{Q}$ a $\mathbb{R}$. Pro ilustraci tohoto pojmu si ještě uvědomme, že charakteristikou tělesa zbytkových tříd $\mathbb{Z}/p\mathbb{Z}$ je číslo p a že žádné těleso nemůže mít charakteristiku 1, neboť musí dle definice tělesa splňovat formuli $0 \neq 1$.

Pakliže těleso $\mathbb{F}$ není charakteristiky 2, tj. neplatí v něm $1+1=0$, existuje k číslu $1+1$ (kteréžto budeme dále značit jednoduše jako 2) inverzní prvek $\frac{1}{2}$. Z vlastností sčítání a násobení je pak zřejmé, že pro každá různá i, j platí

$$a_{ij}x_ix_j + a_{ji}x_jx_i = \tfrac{1}{2}(a_{ij} + a_{ji})x_ix_j + \tfrac{1}{2}(a_{ij} + a_{ji})x_ix_j.$$

Příslušnou matici koeficientů A tedy můžeme předpisem

$$b_{ij} = \begin{cases} a_{ij} & \text{pro } i = j \\ \frac{1}{2}(a_{ij} + a_{ji}) & \text{pro } i \neq j \end{cases}$$

převést na matici B, která bude symetrická podle tzv. *hlavní diagonály* (tvořené prvky $a_{11}, a_{22}, \ldots, a_{nn}$) a která bude odpovídat téže kvadratické formě. Aplikujeme-li tento postup na matici koeficientů formy Q_{dist}, získáváme matici

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Zároveň je zřejmé, že každá symetrická matice řádu n odpovídá jedné kvadratické formě n proměnných. V dalším textu se tedy omezíme pouze na tělesa charakteristiky různé od 2, čímž získáme (až na přejmenování proměnných) vzájemně jednoznačnou korespondenci mezi kvadratickými formami n proměnných a symetrickými maticemi řádu n.

V transformaci kvadratických forem do jednoduššího tvaru pak lze jít ještě o krok dále. Je možné ukázat, že každou formu lze pomocí lineární substituce proměnných převést do tzv. *kanonického tvaru*, tj. takového tvaru, kdy všechny koeficienty a_{ij} pro $i \neq j$ jsou rovny nule. Formu v kanonickém tvaru pak lze vyjádřit jako

$$Q(x_1, \ldots, x_n) = a_1 x_1^2 + a_2 x_2^2 + \cdots + a_n x_n^2$$

a jí odpovídající symetrická matice bude mít na všech pozicích mimo hlavní diagonálu nuly. Takovéto matice se označují jako *diagonální*. Např. racionální formu $x^2 + xy + y^2$ lze zapsat jako $x'^2 + 3y'^2$, kde $x' = x + \frac{1}{2}y$ a $y' = \frac{1}{2}y$. Tato transformace však již není ekvivalentní v tom smyslu, že by Q a Q′, kde Q′ je transformovaná forma Q, byly stejnými funkcemi (v právě uvedeném příkladu přiřazuje např. původní forma dvojici $(1, 1)$ hodnotu 3, kdežto transformovaná forma má na stejném argumentu hodnotu 4). Nicméně základní charakteristiky kvadratických forem, tj. determinant (viz dále) a zejména pak obor hodnot, budou mít Q a Q′ stejné.

Právě obor hodnot kvadratické formy Q bude hlavním předmětem našeho zájmu. Budeme zkoumat, zda určitá forma Q tzv. *reprezentuje* nějaké dané číslo r, tj. jestli existuje vektor $\mathbf{x}$ takový, aby $Q(\mathbf{x}) = r$. Jinými slovy, budeme zkoumat řešitelnost rovnice $Q(\mathbf{x}) = r$.

Jestliže Q a transformovaná Q′ tedy mají stejný obor hodnot, je otázka řešitelnosti $Q(\mathbf{x}) = r$ ekvivalentní otázce řešitelnosti $Q'(\mathbf{x}) = r$. V dalším textu už proto budeme bez újmy na obecnosti pracovat pouze s kanonickými formami.

Determinant kvadratické formy je definován jako determinant jí příslušející matice koeficientů. Pro kanonickou formu $Q(x_1, \ldots, x_n) = a_1x_1^2 + a_2x_2^2 + \cdots + a_nx_n^2$ je tedy determinantem číslo $\prod_{i=1}^{n} a_i$, čili součin prvků na hlavní diagonále matice příslušné ke Q.

Kvadratickou formu následně nazveme *degenerovanou*, je-li její determinant roven nule. V opačném případě nazveme formu *nedegenerovanou*. Kvadratická forma je tedy degenerovaná právě tehdy, je-li některý z jejích koeficientů roven nule, což znamená, že hodnota této funkce je nezávislá na hodnotě jedné ze vstupních proměnných.

Ústřední roli při vyšetřování reprezentovatelnosti čísel kvadratickými formami ve smyslu $Q(\mathbf{x}) = r$ bude hrát číslo 0 a zkoumání řešitelnosti rovnice $Q(\mathbf{x}) = 0$. Je zřejmé, že jedno takové řešení získáme vždy, vezmeme-li za $\mathbf{x}$ nulový vektor $(0, 0, \ldots, 0)$. Toto řešení proto prohlásíme za triviální a budeme se dále zabývat pouze netriviální řešitelností takovéto rovnice.

Kvadratické formy $Q(\mathbf{x})$, pro něž existuje netriviální řešení rovnice $Q(\mathbf{x}) = 0$, označíme jako *izotropní*. Příslušný nenulový vektor $\mathbf{x}$, který je řešením této rovnice, pak budeme nazývat *izotropním vektorem* pro formu Q.

Izotropní kvadratické formy jsou tedy právě ty formy, které netriviálně reprezentují nulu. Lze snadno nahlédnout, že každá degenerovaná forma je izotropní. (Nechť např. koeficient $a_2 = 0$, pak vektor $(0, 1, 0, 0 \ldots, 0)$ je izotropní pro danou formu.) Je tedy přirozené, že nás v dalším textu bude zajímat izotropie pouze u nedegenerovaných forem. Pro nedegenerované kvadratické formy navíc platí následující lemma.

Lemma. *Nechť Q je nedegenerovaná kvadratická forma nad tělesem* $\mathbb{F}$. *Jestliže Q je izotropní, pak reprezentuje všechny prvky z nosiče* $\mathbb{F}$ *(je tzv.* univerzální*).*

Důkaz tohoto lemmatu je ryze numerický. Nechť $Q(x_1, x_2, \ldots, x_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ je nedegenerovaná kvadratická forma a nechť $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ je nenulový vektor takový, že $Q(\mathbf{y}) = 0$. Bez újmy na obecnosti můžeme předpokládat, že $y_1 \neq 0$. Zaveďme nyní novou proměnnou t a dosaďme do Q za x_1 výraz $y_1(1 + t)$ a za x_i pro $i = 2, \ldots, n$ výraz $y_i(1 - t)$. Dostáváme tak funkci Q′ jedné proměnné definovanou jako

$$Q'(t) = a_1(y_1^2(1+2t+t^2)) + a_2(y_2^2(1-2t+t^2)) + \cdots + a_n(y_n^2(1-2t+t^2)).$$

Z pravé strany nyní můžeme vytknout výraz $(1 + t^2)$, čili

$$Q'(t) = (1+t^2)(a_1y_1^2+a_2y_2^2+\cdots+a_ny_n^2)+(2ta_1y_1^2-2ta_2y_2^2-\ldots-2ta_ny_n^2).$$

První člen je roven nule, neboť předpokládáme $Q(\mathbf{y}) = 0$, a z druhého můžeme vytknout 2t. Dostáváme tak

$$Q'(t) = 2t(a_1y_1^2 - a_2y_2^2 - \ldots - a_ny_n^2).$$

Jelikož $a_1y_1^2 + a_2y_2^2 + \cdots + a_ny_n^2 = 0$, je zřejmé, že $-a_2y_1^2 - \ldots - a_ny_n^2 = a_1y_1^2$, a tedy ve výsledku

$$Q'(t) = 4ta_1y_1^2.$$

Zvolíme-li nyní libovolné m prvek $\mathbb{F}$, pak $Q'\left(\frac{m}{4a_1y_1^2}\right) = m$. Podíváme-li se zpět na způsob, jakým byla Q' definována, je patrné, že pak i

$$Q\left(y_1(1+\tfrac{m}{4a_1y_1^2}), y_2(1-\tfrac{m}{4a_1y_1^2}), \ldots, y_n(1-\tfrac{m}{4a_1y_1^2})\right) = m,$$

a Q je tedy univerzální.

Na základě právě dokázaného lemmatu již není obtížné ukázat, že při studiu reprezentovatelnosti libovolného nenulového čísla m formou n proměnných se lze omezit na otázku izotropie jisté formy $n + 1$ proměnných.

Lemma. *Nedegenerovaná kvadratická forma* $Q(x_1, x_2, \ldots x_n)$ *reprezentuje nenulové číslo* m *tehdy a jen tehdy, když kvadratická forma*

$$Q(x_1, x_2, \ldots x_n) - mx_{n+1}^2$$

je izotropní.

Důkaz již je poměrně snadný. Mějme řešení rovnice $Q(x_1, x_2, \ldots x_n) = m$. Pak stačí převést číslo m na levou stranu, čímž získáme $Q(x_1, x_2, \ldots x_n) - m1^2 = 0$, a $(x_1, x_2, \ldots x_n, 1)$ je tedy příslušný izotropní vektor. Nechť naopak $Q(x_1, x_2, \ldots x_n) - mx_{n+1}^2 = 0$ má netriviální řešení. Jestliže $x_{n+1} \neq 0$, můžeme převést mx_{n+1}^2 na pravou stranu a následně vydělit celou rovnici výrazem x_{n+1}^2 a získat tak reprezentaci čísla m. Je-li naproti tomu $x_{n+1} = 0$, pak $(x_1, x_2, \ldots x_n)$ nemůže být nulový vektor a navíc platí $Q(x_1, x_2, \ldots x_n) = 0$. Forma Q je tím pádem izotropní a dle předchozího lemmatu reprezentuje libovolné m.

6. Definovatelnost přirozených čísel v $\mathbb{Q}$

6.1. Úmluva o značení

V této a následující kapitole budeme pracovat se smíšenými aritmetickými formulemi hovořícími o celých i racionálních číslech. Přijměme tedy konvenci, že v takovýchto formulích budou pro celá čísla používány proměnné psané malými písmeny, velká písmena pak budou výlučně používána pro označení racionálních čísel. Formule $(\forall x)(\exists Y)x = Y$ tedy například vyjadřuje, že každé celé číslo je racionální.

Dále budeme-li hovořit o základním tvaru racionálního čísla M, budeme vždy mít na mysli uspořádanou dvojici celých čísel n a d, kde $M = \frac{n}{d}$, $(n, d) = 1$ a $d > 0$. Zmíníme-li pouze „nechť $M = \frac{n}{d}$“, budeme tím vždy myslet základní tvar čísla M.

6.2. Úvodní lemmata

Nejprve zformulujme dvě lemmata, z nichž bude vycházet celá konstrukce vedoucí k definici přirozených čísel v $\mathbb{Q}$, kterážto bude podána v následující podkapitole.

Lemma 1. *Nechť* p *je prvočíslo kongruentní s* 3 (mod 4) *a* m *je libovolné nenulové přirozené číslo. Pak* m *je reprezentovatelné jako*

$$m = X^2 + Y^2 - pZ^2, \tag{1}$$

právě když m *není tvaru* $m = ks^2$, *kde* $k \equiv p \pmod 8$, *ani tvaru* $m = pks^2$, *kde* $\left(\frac{k}{p}\right) = 1$.

Lemma 2. *Nechť* p *a* q *jsou lichá prvočísla,* $p \equiv 1 \pmod 4$ *a* $\left(\frac{q}{p}\right) = -1$ *a nechť* m *je libovolné nenulové přirozené číslo. Pak* m *je reprezentovatelné jako*

$$m = X^2 + qY^2 - pZ^2, \tag{2}$$

právě když m *není tvaru* $m = pks^2$, *kde* $\left(\frac{k}{p}\right) = -1$, *ani tvaru* $m = qks^2$, *kde* $\left(\frac{k}{q}\right) = -1$.

Obě lemmata nyní převedeme do tvaru vhodnějšího k důkazu. Nejprve převedeme levé strany obou ekvivalencí (tedy zda je m reprezentovatelné jistými kvadratickými formami) do průhlednějšího tvaru a posléze provedeme

podobné zjednodušující, nicméně ekvivalentní úpravy i s pravými stranami (tj. zda je m jednoho z uvedených tvarů). Dokážeme-li pak, že levá a pravá strana jsou po těchto úpravách ekvivalentní, získáme z tranzitivity ekvivalence i důkazy lemmat 1 a 2.

Obraťme tedy pozornost nejprve na levé strany právě vyslovených lemmat. Z minulé kapitoly víme, že otázka reprezentovatelnosti čísla ternárními racionálními formami je ekvivalentní otázce izotropie jistých kvaternárních forem. Řešitelnost rovnic (1) a (2) je tak ekvivalentní netriviální řešitelnosti rovnic

$$X^2 + Y^2 - pZ^2 - mW^2 = 0 \tag{3}$$

$$X^2 + qY^2 - pZ^2 - mW^2 = 0.^1 \tag{4}$$

Navíc je možné si všimnout, že tyto rovnice jsou řešitelné v oboru racionálních čísel, právě když jsou řešitelné i v oboru celých čísel, neboť každé celočíselné řešení je i racionální, a na druhou stranu máme-li racionální řešení např. rovnice (3) vektorem $\left(\frac{x_n}{x_d}, \frac{y_n}{y_d}, \frac{z_n}{z_d}, \frac{w_n}{w_d}\right)$, pak můžeme celou rovnici vynásobit číslem $x_d^2 y_d^2 z_d^2 w_d^2$, tedy druhými mocninami všech jmenovatelů, a získat tak

$$(x_n^2 y_d^2 z_d^2 w_d^2) + (y_n^2 x_d^2 z_d^2 w_d^2) - p(z_n^2 x_d^2 y_d^2 w_d^2) - m(w_n^2 x_d^2 y_d^2 z_d^2) = 0,$$

což je řešení v oboru celých čísel a bude netriviální právě tehdy, když původní racionální řešení bylo netriviální. Totéž samozřejmě platí i pro (4). Rovnice tedy můžeme přepsat do tvaru

$$x^2 + y^2 - pz^2 - mw^2 = 0 \tag{5}$$

[1] Poznámka pro ty, co předchozí kapitolu přeskočili. Máme-li řešení rovnice $m = X^2 + Y^2 - pZ^2$, můžeme od jejích obou stran odečíst m a získat tak $X^2 + Y^2 - pZ^2 - m1^2 = 0$. Vektor $(X, Y, Z, 1)$ je pak zřejmě netriviálním řešení rovnice (3). Mějme naopak netriviální řešení (X, Y, Z, W) rovnice (3). Jestliže $W \neq 0$, můžeme mW^2 převést na pravou stranu a následně celou rovnici vydělit W^2, čímž získáme reprezentaci čísla m ve tvaru (1). Jestliže $W = 0$, pak alespoň jedno z čísel X, Y, Z není nula – nechť je to např. X – a zároveň platí $X^2 + Y^2 - pZ^2 = 0$. Pak $(X^2(1 + \frac{m}{4X^2})^2) + (Y^2(1 - \frac{m}{4X^2})^2) - p(Z^2(1 - \frac{m}{4X^2})^2)$ je hledaná reprezentace čísla m. Pokud totiž roznásobíme závorky, získáme $(X^2(1 + \frac{m}{2X^2} + \frac{m^2}{16X^4})) + (Y^2(1 - \frac{m}{2X^2} + \frac{m^2}{16X^4})) - p(Z^2(1 - \frac{m}{2X^2} + \frac{m^2}{16X^4}))$, dále můžeme vytknout výraz $(1 + \frac{m^2}{16X^4})$ a získáváme tak $(1 + \frac{m^2}{16X^4})(X^2 + Y^2 - pZ^2) + \frac{m}{2X^2}(X^2 - Y^2 + pZ^2)$. Jelikož předpokládáme $X^2 + Y^2 - pZ^2 = 0$, je první sčítanec roven nule a zároveň $-Y^2 + pZ^2 = X^2$. Konečně tedy $\frac{m}{2X^2}(X^2 + X^2) = m$.

Zcela stejnou úvahu lze použít pro ukázání ekvivalence netriviální řešitelnosti (4) s řešitelností (2), přičemž pro důkaz reprezentovatelnosti čísla m lze použít výraz $(X^2(1 + \frac{m}{4X^2})^2) + q(Y^2(1 - \frac{m}{4X^2})^2) - p(Z^2(1 - \frac{m}{4X^2})^2)$.

$$x^2 + qy^2 - pz^2 - mw^2 = 0. \tag{6}$$

Poslední úvaha, kterou provedeme, je redukce čísla m. Pokud m není tzv. *square-free*, tzn. pokud je tvaru $m = m'n^2$, kde n je nějaký netriviální dělitel, můžeme v rovnicích nahradit m jeho square-free faktorem m' a otázka řešitelnosti výsledné rovnice bude ekvivalentní řešitelnosti rovnice původní. Nechť je totiž opět například rovnice (5) řešitelná vektorem (x, y, z, w), čili platí

$$x^2 + y^2 - pz^2 - m'n^2w^2 = 0.$$

Pak (x, y, z, nw) je řešením pro rovnici, v níž je koeficient m nahrazen svou square-free částí m'. A naopak nechť máme řešení (x, y, z, w) pro

$$x^2 + y^2 - pz^2 - m'w^2 = 0.$$

Pak můžeme celou rovnici vynásobit n^2 a získáme řešení (xn, yn, zn, w) pro původní rovnici s koeficientem m.

Celou právě provedenou sérii ekvivalentních úprav nyní můžeme shrnout tak, že chceme-li se zabývat reprezentovatelností celých čísel ve smyslu (1) a (2), stačí se omezit pouze na netriviální řešitelnost celočíselných rovnic

$$x^2 + y^2 - pz^2 - mw^2 = 0$$
$$x^2 + qy^2 - pz^2 - mw^2 = 0,$$

přičemž stačí uvažovat pouze square-free čísla m, jelikož každé číslo je reprezentovatelné, právě když je jeho square-free část reprezentovatelná. Navíc u řešení těchto rovnic se stačí omezit pouze na tzv. *primitivní řešení*, tj. taková, pro něž $(x, y, z, w) = 1$. Máme-li totiž řešení x, y, z, w, kde všechna čtyři čísla mají společný faktor, pak můžeme celou rovnici vydělit čtvercem tohoto faktoru.

Zaměřme nyní pozornost na podmínky reprezentovatelnosti, tak jak je stanovují lemmata 1 a 2, tedy na pravé strany obou ekvivalencí. I zde uděláme patřičné úpravy, abychom se mohli v dalším textu omezit výlučně na square-free čísla.

Lemma 1 praví, že libovolné číslo m je reprezentovatelné, není-li ani jednoho z tvarů $m = ks^2$, kde $k \equiv p \pmod 8$, nebo $m = pks^2$, kde $\left(\frac{k}{p}\right) = 1$. Nechť $m = m'n^2$, kde m' je jeho square-free část. Ukážeme, že m není ani jednoho z výše uvedených tvarů právě tehdy, když $m' \not\equiv p \pmod 8$ a zároveň m' není tvaru $m' = pk$, kde $\left(\frac{k}{p}\right) = 1$.

Nechť m' je kongruentní s $p \pmod 8$, pak m je evidentně tvaru $m = ks^2$, kde $k \equiv p \pmod 8$, neboť stačí vzít za k číslo m' a za s číslo n. Stejně tak je-li m' tvaru $m' = pk$, kde $\left(\frac{k}{p}\right) = 1$, pak $m = m'n^2$ je zřejmě tvaru $m = pks^2$, kde $\left(\frac{k}{p}\right) = 1$. Stačí se tedy zabývat opačnými implikacemi.

Nechť $m = ks^2$, kde $k \equiv p \pmod 8$. Chceme ukázat, že pak i square-free část m' čísla m je kongruentní s $p \pmod 8$. Nechť $k = k'l^2$, kde k' je square-free část čísla k, pak zjevně $m' = k'$. p je kongruentní s 3 (mod 4), tedy s 3 nebo 7 (mod 8), v obou případech je však liché modulo 8. Z rovnice $m'l^2 \equiv p \pmod 8$ pak plyne, že l^2 nemůže být sudé. Jelikož jediné čtverce modulo 8 jsou 0, 1 a 4, musí být l^2 kongruentní s 1. Tím pádem $m' \equiv p \pmod 8$.

Dále nechť m je tvaru $m = pks^2$, kde $\left(\frac{k}{p}\right) = 1$. Je zřejmé, že $p \nmid k$, neboť k je kvadratickým reziduem modulo p. Square-free část m' čísla m tak bude rovna pk', kde k' je square-free část čísla k. Nechť tedy $k = k'l^2$. Pak

$$\left(\frac{k}{p}\right) = \left(\frac{k'l^2}{p}\right) = \left(\frac{k'}{p}\right)\left(\frac{l^2}{p}\right) = \left(\frac{k'}{p}\right),$$

a m' je tedy tvaru $m' = pk'$, kde $\left(\frac{k'}{p}\right) = 1$.

Zcela obdobnou úvahu pak lze provést i pro obě dvě podmínky z lemmatu 2.

Celou věc tedy můžeme shrnout tak, že lemmata 1 a 2 lze reformulovat v následující ekvivalentní podobě.

Lemma 3. *Nechť p je prvočíslo kongruentní s 3 (mod 4) a m je libovolné kladné square-free číslo. Pak celočíselná rovnice*

$$x^2 + y^2 - pz^2 - mw^2 = 0 \tag{5}$$

je netriviálně řešitelná, právě když $m \not\equiv p \pmod 8$ a m není tvaru $m = pk$, kde $\left(\frac{k}{p}\right) = 1$.

Lemma 4. *Nechť p a q jsou lichá prvočísla, $p \equiv 1 \pmod 4$ a $\left(\frac{q}{p}\right) = -1$ a nechť m je libovolné kladné square-free číslo. Pak celočíselná rovnice*

$$x^2 + qy^2 - pz^2 - mw^2 = 0 \tag{6}$$

je netriviálně řešitelná, právě když m není tvaru $m = pk$, kde $\left(\frac{k}{p}\right) = -1$, ani tvaru $m = qk$, kde $\left(\frac{k}{q}\right) = -1$.

Přistupme nyní k důkazu těchto dvou lemmat. Omezíme se nicméně pouze na důkaz nutnosti příslušných podmínek, tj. ukážeme, že je-li m jednoho ze ze dvou zmíněných „kritických" tvarů, pak daná rovnice nemůže mít řešení. Důkaz postačitelnosti pro obě lemmata vyžaduje vybudování pokročilého matematického aparátu a několik vysoce netriviálních vět z teorie čísel. Postup důkazu bude nicméně naznačen v 8. kapitole.

Začněme tedy s lemmatem 3 a uvažujme situaci, kdy $m \equiv p \pmod 8$. p je kongruentní s 3 modulo 4, a je tedy kongruentní s 3 nebo 7 modulo 8. Se stejným číslem je pak kongruentní i m. Je-li rovnice (5) řešitelná v oboru celých čísel, pak zjevně musí být řešitelná modulo libovolné n. Stačí nám tedy ukázat jeden konkrétní modul, pro který rovnice nemůže mít řešení. Postupujme tedy sporem. Nechť x, y, z, w jsou primitivním celočíselným řešením rovnice (5) a nechť např. $p \equiv m \equiv 3 \pmod 8$ (pro sedmičku by byl postup identický). Uvažujme kongruenci

$$x^2 + y^2 - 3z^2 - 3w^2 \equiv 0 \pmod 8.$$

Druhá mocnina každého lichého čísla je kongruentní s jedničkou modulo 8. Na druhou stranu čtverec sudého čísla může být kongruentní s 0 nebo 4. Jelikož předpokládáme, že řešení je primitivní, je zřejmé, že všechna čísla nemohou být sudá. Alespoň jedno z čísel x^2, y^2, z^2, w^2 tedy musí být 1. Pokud by zbylé čtverce byly sudé, lze nahlédnout, že pak by součet levé strany byl nutně lichý, a nemohl by tak být kongruentní s nulou. Proto alespoň dvě z čísel x^2, y^2, z^2, w^2 jsou kongruentní s 1. Rozeberme jednotlivé tři možné případy.

Nechť $z^2 \equiv w^2 \equiv 1$. Pak lze výše uvedenou kongruenci upravit na $x^2 + y^2 \equiv 6 \pmod 8$, což vzhledem ke skutečnosti, že součet dvou čtverců modulo 8 může být pouze 0, 1, 2, 4 nebo 5, není možné.

Pakliže $x^2 \equiv y^2 \equiv 1$, pak $2 \equiv 3(z^2 + w^2) \pmod 8$. A tato kongruence není ze stejného důvodu řešitelná.

Nechť tedy konečně $x^2 \equiv z^2 \equiv 1$, pak $y^2 - 3w^2 \equiv 2 \pmod 8$ a probráním možných hodnot y^2 a w^2 se lze přesvědčit, že i tato kongruence není řešitelná, a rovnice (5) tedy nemůže mít primitivní řešení v oboru celých čísel.

Uvažme nyní druhou podmínku, tj. situaci kdy $m = pk$, kde $\left(\frac{k}{p}\right) = 1$. Rovnici (5) pak můžeme přepsat do tvaru

$$x^2 + y^2 = p(z^2 + kw^2).$$

Postupujme opět sporem a předpokládejme, že máme takové nenulové číslo n, které je reprezentovatelné zároveň jako $n = x^2 + y^2$ a $n = p(z^2 + kw^2)$, přičemž navíc $(x, y, z, w) = 1$. Je zřejmé, že $p \mid n$. Z kapitoly 4.1 víme, že n je reprezentovatelné ve tvaru $n = x^2 + y^2$, tj. jako suma dvou čtverců, tehdy a jen tehdy, když každé prvočíslo tvaru $4k+3$ se v jeho rozkladu vyskytuje na sudou mocninu. $p \equiv 3 \pmod 4$ a zároveň zjevně $p \mid n$, proto nutně i $p^2 \mid n$, což znamená, že $p \mid z^2 + kw^2$. Platí tak

$$z^2 + kw^2 \equiv 0 \pmod p.$$

Tedy $z^2 \equiv -kw^2 \pmod p$, a jestliže $p \nmid w$, můžeme tuto kongruenci dále upravit na $z^2(w^{-1})^2 \equiv -k \pmod p$. To by znamenalo, že $\left(\frac{-k}{p}\right) = 1$. Protože ale jednička je kvadratické nonreziduum modulo každé prvočíslo tvaru $4k + 3$, platí, že

$$\left(\frac{-k}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{k}{p}\right) = -\left(\frac{k}{p}\right),$$

což by byl spor s předpokladem, že $\left(\frac{k}{p}\right) = 1$. Z toho plyne, že $p \mid w$, a tedy i $p \mid z$. Pokud $w = pw'$ a $z = pw'$, platí, že $n = p(z^2+kw^2) = p^3(z'^2+kw'^2)$. Pokud by p ještě dále dělilo $z'^2 + kw'^2$, mohli bychom příslušnou úvahu s kvadratickým charakterem k dále opakovat a postupně odebírat další faktory p^2. p se tedy musí v rozkladu čísla n vyskytovat na lichou mocninu, což je spor s předpokladem, že n je reprezentovatelné jako suma dvou čtverců.

Důkaz nutnosti obou podmínek v lemmatu 4 se ponese ve velmi podobném duchu jako důkaz druhé části důkazu předchozího. To ostatně již dává tušit tvar těchto podmínek. Nechť tedy nejprve $m = pk$, kde $\left(\frac{k}{p}\right) = -1$. Zároveň předpokládáme, že $p \equiv 1 \pmod 4$ a $\left(\frac{q}{p}\right) = -1$. Ze zákona kvadratické reciprocity pak plyne, že i $\left(\frac{p}{q}\right) = -1$. Rovnici (6) můžeme přepsat do tvaru

$$x^2 + qy^2 = pz^2 + pkw^2.$$

Uvažujme opět takové nenulové n, které je reprezentovatelné výrazy na obou stranách rovnice. Navíc můžeme předpokládat, že $(x, y, z, w) = 1$. Platí tedy $x^2 + qy^2 = n$ a zřejmě $p \mid n$, což znamená, že $x^2 + qy^2 \equiv 0 \pmod p$. Pakliže $p \nmid y$, můžeme rovnici dále upravit na tvar

$$-x^2(y^{-1})^2 \equiv q \pmod p.$$

Vzhledem k tomu, že p je tvaru $4k+1$, je -1 kvadratickým reziduem, a tedy i q by bylo kvadratickým reziduem, což je spor s předpokladem $\left(\frac{q}{p}\right) = -1$. p tedy musí být dělitelem y, a tím pádem i čísla x. Nechť $x = px'$ a $y = py'$. Celou rovnici pak můžeme vydělit číslem p a získáváme

$$px'^2 + qpy'^2 = z^2 + kw^2,$$

přičemž $n' = \frac{n}{p}$ je reprezentovatelné výrazy na obou stranách rovnice. Tedy $n' = z^2 + kw^2$ a zřejmě opět $p \mid n'$. Tedy $z^2 + kw^2 \equiv 0 \pmod{p}$. Jestliže $p \nmid w$ můžeme rovnici dále upravit na

$$-z^2(w^{-1})^2 \equiv k \pmod{p}.$$

Podobně jako před okamžikem q by nyní číslo k bylo kvadratickým reziduem modulo p, a to by byl spor s předpokladem $\left(\frac{k}{p}\right) = -1$. p tedy dělí w i z, a je tak společným dělitelem všech čísel x, y, z, w, čímž dostáváme spor s předpokladem, že $(x, y, z, w) = 1$, tj. že tato čísla byla primitivním řešením rovnice (6).

Situaci, kdy $m = qk$, kde $\left(\frac{k}{q}\right) = -1$, už můžeme pojednat jen heslovitě. Rovnici (6) můžeme zapsat jako

$$x^2 - pz^2 = -qy^2 + qkw^2,$$

a máme tedy n takové, že $x^2 - pz^2 = n$ a $q \mid n$. Čili $x^2 - pz^2 \equiv 0 \pmod{q}$, tj. $x^2 \equiv pz^2 \pmod{q}$. Předpoklad $q \nmid z$ vede k $x^2(z^{-1})^2 \equiv p \pmod{q}$, a tedy ke sporu s $\left(\frac{p}{q}\right) = -1$. Proto $q \mid z$ i $q \mid x$. Dělme celou rovnici číslem q a získáváme

$$qx'^2 - pqz'^2 = -y^2 + kw^2.$$

Máme tedy $n' = \frac{n}{q}$, $n' = -y^2 + kw^2$ a zřejmě $q \mid n'$. Tedy $y^2 \equiv kw^2 \pmod{q}$ a předpoklad $p \nmid w$ vede k $y^2(w^{-1})^2 \equiv k \pmod{q}$ a ke sporu s $\left(\frac{k}{q}\right) = -1$. Tedy q vedle x a z dělí i w a y dostáváme spor s vzájemnou nesoudělností čísel x, y, z, w.

6.3. Robinsonové věta

Nyní již můžeme přistoupit k vlastnímu důkazu Robinsonové věty, která bude hrát ústřední úlohu při hledání formule definující množinu přirozených čísel v $\mathbb{Q}$. Nejprve ale dokažme tři pomocná lemmata.

Lemma 5. *Nechť* p *je prvočíslo kongruentní s* $3 \pmod 4$. *Pak pro libovolné racionální číslo* M, *jehož základním tvarem je* $M = \frac{n}{d}$, *platí*

$$(\exists X, Y, Z)(2 + pM^2 = X^2 + Y^2 - pZ^2) \equiv (2 \nmid d \,\&\, p \nmid d).$$

Levá strana ekvivalence vypovídá o řešitelnosti jisté rovnice. Nechť je řešitelná, a existují tedy taková X, Y, Z, pro něž platí

$$2 + p\frac{n^2}{d^2} = X^2 + Y^2 - pZ^2.$$

Tuto rovnici nyní můžeme vynásobit číslem d^2 a získáváme tak

$$2d^2 + pn^2 = d^2X^2 + d^2Y^2 - pd^2Z^2,$$

což když položíme $X' = dX$, $Y' = dY$ a $Z' = dZ$, lze přepsat do tvaru

$$2d^2 + pn^2 = X'^2 + Y'^2 - pZ'^2.$$

To jinými slovy znamená, že číslo $2d^2 + pn^2$ je reprezentovatelné ve tvaru (1) ve smyslu lemmatu 1 z úvodu této kapitoly. Na druhou stranu pokud lze číslo $2d^2 + pn^2$ zapsat jako

$$2d^2 + pn^2 = X^2 + Y^2 - pZ^2,$$

získáme vydělením rovnice číslem d^2

$$2 + pM^2 = \frac{X^2}{d^2} + \frac{Y^2}{d^2} - \frac{Z^2}{d^2} = X'^2 + Y'^2 - pZ'^2.$$

Otázka platnosti formule z levé strany ekvivalence je tedy ekvivalentní reprezentovatelnosti čísla $2d^2 + pn^2$ ve smyslu (1). Pro důkaz lemmatu 5 nám tedy stačí ukázat, že podmínky z jeho pravé strany, čili $2 \nmid d$ a $p \nmid d$, jsou ekvivalentní podmínkám reprezentovatelnosti čísla $m = 2d^2 + pn^2$ dle lemmatu 1, tzn. že m není tvaru $m = ks^2$, kde $k \equiv p \pmod 8$, ani tvaru $m = pks^2$, kde $\left(\frac{k}{p}\right) = 1$.

Nechť tedy d není sudé a $p \nmid d$. Jelikož p nedělí d, nedělí ani $m = 2d^2 + pn^2$, a m tedy zcela jistě nelze vyjádřit ve tvaru $m = pks^2$.

Dále předpokládejme pro spor, že $m = 2d^2 + pn^2$ je vyjádřitelné ve tvaru ks^2, kde $k \equiv p \pmod 8$. Pak by samozřejmě musela platit i kongruence

$$2d^2 + pn^2 \equiv ks^2 \pmod 4.$$

Jelikož d je liché, platí $d^2 \equiv 1 \pmod 4$. Zároveň předpokládáme, že $p \equiv -1 \pmod 4$. Navíc $k \equiv p \pmod 8$, a tedy i $k \equiv p \pmod 4$, tzn. $k \equiv -1 \pmod 4$. Rovnici tedy můžeme přepsat jako

$$2 - n^2 \equiv -s^2 \pmod 4,$$

což můžeme dále upravit na

$$n^2 - s^2 \equiv 2 \pmod 4.$$

Čísla n^2 a s^2 však mohou mít coby čtverce modulo 4 hodnotu pouze 0 nebo 1. Jejich rozdíl tak nemůže být 2. Číslo m tedy není ani tvaru $m = ks^2$, kde $k \equiv p \pmod 8$.

V důkazu druhého směru pak stačí ukázat, že jestliže $p \mid d$ nebo je d sudé, pak $m = 2d^2 + pn^2$ je jednoho ze dvou „kritických tvarů".

Nechť tedy $p \mid d$, tj. $d = pr$. Pak $m = 2p^2r^2 + pn^2 = p(2pr^2 + n^2)$. Položme $k = 2pr^2 + n^2$. p nemůže dělit n, jelikož pak by n a d byla soudělná, což by byl spor, že $M = \frac{n}{d}$ je v základním tvaru. p tedy nedělí ani $k = 2pr^2 + n^2$ a můžeme se ptát na kvadratický charakter k modulo p.

$$\left(\frac{k}{p}\right) = \left(\frac{2pr^2 + n^2}{p}\right) = \left(\frac{n^2}{p}\right) = 1$$

Číslo m je tedy tvaru $m = pkl^2$, kde $\left(\frac{k}{p}\right) = 1$.

Nechť d je sudé, tj. $d = 2r$. Pak $m = 8r^2 + pn^2$, a tedy $m \equiv pn^2 \pmod 8$. Navíc jistě $2 \nmid n$, neboť d je sudé a $(n, d) = 1$. Proto $n^2 \equiv 1 \pmod 8$ a $m \equiv p \pmod 8$. Tedy $m = ml^2$, kde $m \equiv p \pmod 8$.

Lemma 6. *Nechť p a q jsou lichá prvočísla,* $p \equiv 1 \pmod 4$ *a* $\left(\frac{q}{p}\right) = -1$. *Pak pro libovolné racionální číslo M, jehož základním tvarem je* $M = \frac{n}{d}$, *platí*

$$(\exists X, Y, Z)(2 + pqM^2 = X^2 + qY^2 - pZ^2) \equiv (p \nmid d \,\&\, q \nmid d).$$

Postup důkazu bude obdobný tomu z lemmatu 5. Má-li rovnice na levé straně ekvivalence řešení, můžeme jej opět vynásobit číslem d^2 a získat tak

$$2d^2 + pqn^2 = d^2X^2 + qd^2Y^2 - pd^2Z^2 = X'^2 + qY'^2 - pZ'^2,$$

což znamená, že číslo $m = 2d^2 + pqn^2$ je reprezentovatelné ve smyslu (2) lemmatu 2. Naopak reprezentovatelnost čísla $m = 2d^2 + pqn^2$ ve smyslu (2) vede k řešitelnosti rovnice z levé strany ekvivalence (příslušnou rovnici stačí

vydělit číslem d^2). Na základě lemmatu 2 tedy stačí dokázat, že $p \nmid d$ a zároveň $q \nmid d$ právě tehdy, když přirozené číslo $m = 2d^2 + pqn^2$ není tvaru $m = pks^2$, kde $\left(\frac{k}{p}\right) = -1$, ani tvaru $m = qks^2$, kde $\left(\frac{k}{q}\right) = -1$.

Jedna strana je zcela evidentní. Neboť jestliže $p \nmid d$ a $q \nmid d$, pak p ani q nemohou být děliteli čísla $m = 2d^2 + pqn^2$, a m tak nelze zapsat ve tvaru $m = pks^2$, resp. $m = qks^2$.

Naopak nechť $p \mid d$, tedy $d = pr$. Pak $m = 2p^2r^2 + pqn^2 = p(2pr^2 + qn^2)$. Položme $k = 2pr^2 + qn^2$. $p \nmid n$, neboť $p \mid d$ a $(n, d) = 1$. Tedy také $p \nmid k$. Pro kvadratický charakter čísla k platí

$$\left(\frac{k}{p}\right) = \left(\frac{2pr^2 + qn^2}{p}\right) = \left(\frac{qn^2}{p}\right) = \left(\frac{q}{p}\right)\left(\frac{n^2}{p}\right) = \left(\frac{q}{p}\right) = -1,$$

a m je tedy tvaru pkl^2, kde $\left(\frac{q}{p}\right) = -1$.

Nakonec nechť $q \mid d$, čili $d = qr$. Pak $m = 2q^2r^2 + pqn^2 = q(2qr^2 + pn^2)$. Položme $k = 2qr^2 + pn^2$. Jelikož $q \mid d$, q nedělí n, a tedy ani k. Pro $\left(\frac{k}{q}\right)$ pak platí

$$\left(\frac{k}{q}\right) = \left(\frac{2qr^2 + pn^2}{q}\right) = \left(\frac{pn^2}{q}\right) = \left(\frac{p}{q}\right)\left(\frac{n^2}{p}\right) = \left(\frac{p}{q}\right) = \left(\frac{q}{p}\right) = -1,$$

přičemž v předposledním kroku jsme využili skutečnosti, že $p \equiv 1 \pmod 4$, a tedy dle zákona kvadratické reciprocity mají $\left(\frac{p}{q}\right)$ a $\left(\frac{q}{p}\right)$ stejnou hodnotu. m je tedy tvaru qkl^2, kde $\left(\frac{q}{p}\right) = -1$.

Lemma 7. *Pro každé liché prvočíslo p existuje liché prvočíslo q takové, že* $\left(\frac{q}{p}\right) = -1$.

Důkaz tohoto lemmatu je poměrně snadný. Vezměme libovolné kvadratické nonreziduum s modulo p. (Připomeňme, že kvadratických nonreziduí modulo p je právě $\frac{p-1}{2}$.) Jedno z čísel s a $s + p$, které je zřejmě také kvadratickým nonreziduem, je jistě liché. Vezměme jeho prvočíselný rozklad. V něm musí být alespoň jedno prvočíslo, které je kvadratickým nonreziduem. Kdyby totiž všechna čísla v rozkladu byla rezidui, byl by reziduem i jejich součin.

Zaveďme nyní nový predikát $\mathcal{R}$, který bude mít centrální úlohu ve zbytku této kapitoly. Definujme ho tak, že pro každé M, A, B platí

$$\mathcal{R}(M, A, B) \equiv (\exists X, Y, Z)(2 + ABM^2 = X^2 + BY^2 - AZ^2).$$

V původním nezbohaceném aritmetickém jazyce můžeme formuli $\mathcal{R}(M,A,B)$ formulovat jako $(\exists X,Y,Z)(1+1+ABMM+AZZ=XX+BYY)$. Ve světle této definice se objasňuje smysl dvou předchozích lemmat. Lemma 5 stanovuje, pro která racionální čísla M platí $\mathcal{R}(M,p,1)$ při nějakém pevně zvoleném prvočísle p tvaru $4k+3$. Naproti tomu lemma 6 se zabývá platností formule $\mathcal{R}(M,p,q)$ pro pevně zvolená prvočísla p a q, kde $p \equiv 1 \pmod 4$ a $\left(\frac{q}{p}\right) = -1$.

Při hledání formule definující přirozená čísla v $\mathbb{Q}$ se můžeme inspirovat ve způsobu, jakým se přirozená čísla definují v teorii množin. Zde je množina ω všech přirozených čísel definována jako průnik všech induktivních množin, tj. takových množin, které obsahují prázdnou množinu a s každým prvkem obsahují i jeho následníka, kde následníkem x rozumíme množinu $x \cup \{x\}$. Budeme-li prázdnou množinu značit 0 a operaci následníka jako $+1$, pak můžeme formuli $\varphi(x)$ definující přirozená čísla v teorii množin zapsat jako

$$(\forall y)((0 \in y \,\&\, (\forall z)(z \in y \to z+1 \in y)) \to x \in y).$$

Aritmetická formule definující přirozená čísla v $\mathbb{Q}$ pak bude mít částečně tutéž strukturu. Nahraďme množinový predikát náležení právě zavedeným predikátem $\mathcal{R}$ a kvantifikujme místo přes všechny množiny y přes všechny dvojice A, B racionálních čísel. Na místě volné proměnné x pak raději použijme velké N, abychom vyhověli typografické konvenci pro aritmetické formule z úvodu této kapitoly. Získáváme tak formuli

$$(\forall A,B)((\mathcal{R}(0,A,B) \,\&\, (\forall M)(\mathcal{R}(M,A,B) \to \mathcal{R}(M+1,A,B))) \to \mathcal{R}(N,A,B)).$$

Tato formule sice ještě není onou definicí přirozených čísel, ke které směřujeme, platí však následující věta.

Věta (Robinson). *Pro každé racionální číslo* N *platí, že splňuje formuli*

$$(\forall A,B)((\mathcal{R}(0,A,B) \,\&\, (\forall M)(\mathcal{R}(M,A,B) \to \mathcal{R}(M+1,A,B))) \to \mathcal{R}(N,A,B)) \tag{7}$$

právě tehdy, když je celým číslem.

Jeden směr důkazu, a sice že každé celé číslo tuto formuli splňuje, je zřejmý. Nechť N je přirozené číslo a A, B jsou racionální čísla taková, že je splněn antecedent formule (7). Pak z indukce plyne, že je splněn i konsekvent, a tedy každé přirozené číslo splňuje (7). Podíváme-li se dále na definici

predikátu $\mathcal{R}$, je možné si všimnout, že M se v ní objevuje pouze na druhou mocninu. Zřejmě tedy platí $(\forall M, A, B)(\mathcal{R}(M, A, B) \equiv \mathcal{R}(-M, A, B))$, a jelikož každé přirozené číslo splňuje formuli (7), splňuje ji i každé celé číslo.

Přistupme nyní k důkazu opačného směru, tedy k důkazu, že každé racionální číslo splňující (7) je celé. Nechť N je tedy takové racionální číslo, které splňuje formuli (7).

Zvolme za A libovolné prvočíslo p tvaru $4k + 3$ a za B vezměme číslo 1. Z lemmatu 5 víme, že pro libovolné $M = \frac{n}{d}$ formule $\mathcal{R}(M, p, 1)$ platí právě tehdy, když $2 \nmid d$ a $p \nmid d$. Zřejmě tedy platí $\mathcal{R}(0, p, 1)$, neboť jmenovatel čísla 0 je 1, a pro každé M také platí formule $\mathcal{R}(M, p, 1) \to \mathcal{R}(M + 1, p, 1)$, neboť čísla M a $M + 1$ v základním tvaru mají stejné jmenovatele. Pakliže totiž $M = \frac{n}{d}$, pak $M + 1 = \frac{n+d}{d}$. Jelikož M je v základním tvaru, pak $(n, d) = 1$. Žádný dělitel čísla d tedy není dělitelem čísla n, a tedy ani $n + d$. $\frac{n+d}{d}$ je proto základním tvarem čísla $M + 1$. Pro parametry $A = p$ a $B = 1$ je tedy splněn antecedent implikace (7), a musí tak platit i závěr, tj. $\mathcal{R}(N, p, 1)$. Z lemmatu 5 nyní plyne, že jmenovatel čísla N není dělitelný číslem dvě, a jelikož jsme číslo p zvolili libovolně, ani žádným prvočíslem tvaru $4k + 3$.

Ve druhém kroku zvolme za A libovolné prvočíslo p tvaru $4k + 1$ a za B k němu vezměme liché prvočíslo q takové, že $\left(\frac{q}{p}\right) = -1$. Existenci takového q garantuje lemma 7. Z lemmatu 6 dále víme, že libovolné $M = \frac{n}{d}$ splňuje formuli $\mathcal{R}(M, p, q)$, právě když p ani q nedělí d. Opět tedy zřejmě platí formule $\mathcal{R}(0, p, q)$ i formule $(\forall M)(\mathcal{R}(M, p, q) \to \mathcal{R}(M + 1, p, q))$. Jelikož N splňuje formuli (7) i její antecedent, splňuje nutně i závěr, tedy $\mathcal{R}(N, p, q)$. A jelikož bylo p zvoleno libovolně, můžeme na základě lemmatu 6 odvodit, že jmenovatel čísla N není dělitelný žádným prvočíslem tvaru $4k + 1$. V minulém odstavci jsme ukázali, že nemůže být dělitelný ani číslem dva či jakýmkoli prvočíslem tvaru $4k + 3$. Jmenovatel čísla N tedy musí být 1, a N je celé číslo.

6.4. Definovatelnost přirozených čísel

Je patrné, že je nyní zcela na místě označit si formuli (7) z předchozí věty jako $\mathrm{Int}(N)$ a predikát Int pak chápat jako „být celým číslem“. V kapitole 4.2 jsme navíc ukázali, že libovolné celé číslo je přirozené právě tehdy, když jej lze zapsat jako součet čtyř čtverců. Kombinací těchto výsledků už můžeme snadno sestavit formuli definující množinu přirozených čísel ve struktuře $\mathbb{Q}$: Libovolné racionální číslo N je přirozené tehdy a jen tehdy, když splňuje

Robinsonové formuli Int(N) a zároveň je součtem čtyř čtverců. Příslušná formule tedy vypadá takto.

$$\mathrm{Int}(N)\ \&\ (\exists A, B, C, D)(N = A^2 + B^2 + C^2 + D^2)$$

7. Poznámka

V konstrukci z předchozí kapitoly jsme u dvou výchozích lemmat vynechali důkazy jednoho směru ekvivalence. Příslušná tvrzení jsou totiž důsledkem hlubokého číselněteoretického teorému, který bude představen v následující kapitole.

Tato kapitola má za cíl zdůraznit místa, na kterých bylo těchto dvou nedokázaných implikací užito, a zároveň vzhledem k tomu, že ani jedna z nich nebyla ve výsledné větě aplikována v plné síle, také formulovat tvrzení o něco slabší, která by nicméně postačovala k úplnému důkazu Robinsonové věty.

Předně tedy připomeňme ony dvě zmíněné nedokázané implikace.

Tvrzení 1 (Implikace $\Leftarrow$ lemmatu 1). *Nechť* p *je prvočíslo kongruentní s* $3 \pmod 4$. *Pak každé nenulové přirozené číslo* m, *které není tvaru* $m = ks^2$, *kde* $k \equiv p \pmod 8$, *ani tvaru* $m = pks^2$, *kde* $\left(\frac{k}{p}\right) = 1$, *je reprezentovatelné ve tvaru*

$$m = X^2 + Y^2 - pZ^2.$$

Tvrzení 2 (Implikace $\Leftarrow$ lemmatu 2). *Nechť* p *a* q *jsou lichá prvočísla,* $p \equiv 1 \pmod 4$ *a* $\left(\frac{q}{p}\right) = -1$. *Pak každé nenulové přirozené číslo* m, *které není tvaru* $m = pks^2$, *kde* $\left(\frac{k}{p}\right) = -1$, *ani tvaru* $m = qks^2$, *kde* $\left(\frac{k}{q}\right) = -1$, *je reprezentovatelné ve tvaru*

$$m = X^2 + qY^2 - pZ^2.$$

Poznamenejme, že jsme tato dvě tvrzení dále převedli do následující ekvivalentní podoby.

Tvrzení 3 (Implikace $\Leftarrow$ lemmatu 3). *Nechť* p *je prvočíslo kongruentní s* $3 \pmod 4$. *Pak pro každé kladné square-free číslo* m, *které není kongruentní s* $p \pmod 8$ *a zároveň není tvaru* $m = pk$, *kde* $\left(\frac{k}{p}\right) = 1$, *má rovnice*

$$x^2 + y^2 = pz^2 + mw^2$$

netriviální řešení.

Tvrzení 4 (Implikace ⇐ lemmatu 4). *Nechť* p *a* q *jsou lichá prvočísla,* $p \equiv 1 \pmod 4$ *a* $\left(\frac{q}{p}\right) = -1$. *Pak pro každé kladné square-free číslo* m, *které není tvaru* $m = pk$, *kde* $\left(\frac{k}{p}\right) = -1$, *ani tvaru* $m = qk$, *kde* $\left(\frac{k}{q}\right) = -1$, *má rovnice*

$$x^2 + qy^2 = pz^2 + mw^2$$

netriviální řešení.

V následném důkazu lemmatu 5 jsme vyšli ze skutečnosti, že pro libovolné prvočíslo p typu $4k + 3$ je pro každé $M = \frac{n}{d}$ formule $\mathcal{R}(M, p, 1)$ ekvivalentní reprezentovatelnosti čísla $2d^2 + pn^2$ ve smyslu lemmatu 1. Podobně důkaz lemmatu 6 vycházel z ekvivalence formule $\mathcal{R}(M, p, q)$ a reprezentovatelnosti čísla $2d^2 + pqn^2$ ve smyslu lemmatu 2 pro libovolné $M = \frac{n}{d}$, prvočíslo $p \equiv 1 \pmod 4$ a liché prvočíslo q, pro něž $\left(\frac{q}{p}\right) = -1$. Jelikož poloviny lemmat 1 a 2 nebyly dokázány, nejsou plným důkazem podloženy ani tyto dvě části lemmat 5 a 6.

Tvrzení 5 (Implikace ⇐ lemmatu 5). *Nechť* p *je prvočíslo kongruentní s* $3 \pmod 4$ *a* $M = \frac{n}{d}$ *je libovolné racionální číslo. Pak*

$$(2 \nmid d \,\&\, p \nmid d) \to \mathcal{R}(M, p, 1).$$

Tvrzení 6 (Implikace ⇐ lemmatu 6). *Nechť* p *a* q *jsou lichá prvočísla,* $p \equiv 1 \pmod 4$, $\left(\frac{q}{p}\right) = -1$ *a* $M = \frac{n}{d}$ *je libovolné racionální číslo. Pak*

$$(p \nmid d \,\&\, q \nmid d) \to \mathcal{R}(M, p, q).$$

Ve finální konstrukci reprezentované Robinsonové větou však nebylo ani jedno z lemmat 5 a 6 použito v plné síle. Pouze jsme ukázali, že jistě platí $\mathcal{R}(0, p, 1)$, resp. $\mathcal{R}(0, p, q)$, a dále pro každé M také implikace $\mathcal{R}(M, p, 1) \to \mathcal{R}(M + 1, p, 1)$, resp. $\mathcal{R}(M, p, q) \to \mathcal{R}(M + 1, p, q)$. Na základě těchto faktů jsme za předpokladu, že N splňovalo Robinsonové definující formuli $\mathrm{Int}(N)$, došli k závěru, že N splňuje i formule $\mathcal{R}(N, p, 1)$ a $\mathcal{R}(N, p, q)$, pročež jeho jmenovatel nemohl být dělitelný žádným prvočíslem, a N tak muselo být celým číslem. Tento závěrečný úsudek se však zakládal na plně dokázaných směrech lemmat 5 a 6.

Nedokázané části lemmat 1 a 2 tak bylo použity pouze k dokázání čtyř výše uvedených formulí $\mathcal{R}(0, p, 1)$, $\mathcal{R}(0, p, q)$, $(\forall M)(\mathcal{R}(M, p, 1) \to \mathcal{R}(M + 1, p, 1))$, $(\forall M)(\mathcal{R}(M, p, q) \to \mathcal{R}(M + 1, p, q))$. Jádro neelementární části konstrukce tedy spočívá v následujících čtyřech tvrzeních (pro větší přehlednost je v nich predikát $\mathcal{R}$ rozepsán dle své definice).

Tvrzení 7 ($\mathcal{R}(0,p,1)$). *Nechť* p *je prvočíslo kongruentní s* 3 (mod 4). *Pak platí*

$$(\exists X,Y,Z)(2 = X^2 + Y^2 - pZ^2).$$

Tvrzení 8 ($\mathcal{R}(0,p,q)$). *Nechť* p *a* q *jsou lichá prvočísla,* $p \equiv 1 \pmod 4$, $\left(\frac{q}{p}\right) = -1$. *Pak platí*

$$(\exists X,Y,Z)(2 = X^2 + qY^2 - pZ^2).$$

Tvrzení 9 ($(\forall M)(\mathcal{R}(M,p,1) \to \mathcal{R}(M+1,p,1))$). *Nechť* p *je prvočíslo kongruentní s* 3 (mod 4). *Pak pro každé* M *platí, že je-li rovnice*

$$2 + pM^2 = X^2 + Y^2 - pZ^2$$

řešitelná, pak je řešitelná i rovnice

$$2 + p(M+1)^2 = X^2 + Y^2 - pZ^2.$$

Tvrzení 10 ($(\forall M)(\mathcal{R}(M,p,q) \to \mathcal{R}(M+1,p,q))$). *Nechť* p *a* q *jsou lichá prvočísla,* $p \equiv 1 \pmod 4$, $\left(\frac{q}{p}\right) = -1$. *Pak pro každé* M *platí, že je-li rovnice*

$$2 + pqM^2 = X^2 + qY^2 - pZ^2$$

řešitelná, pak je řešitelná i rovnice

$$2 + pq(M+1)^2 = X^2 + qY^2 - pZ^2.$$

Navíc je patrné, že tvrzení 7 je triviálně dokazatelné, neboť stačí vzít za X, Y, Z čísla 1, 1, 0.

Ke zkompletování důkazu korektnosti celé konstrukce by tedy postačovalo dokázat pouze tvrzení 8–10. Na ně se totiž redukuje užití neelementárních částí lemmat 1 a 2 (uvedených výše jako tvrzení 1 a 2).

V následující závěrečné kapitole, jejímž cílem je představit aparát potřebný k dokončení důkazu, se nicméně zaměříme přímo na tvrzení 1 a 2. Navíc bude představena věta, jejímž jsou lemmata 1 a 2 pouhým velmi speciálním případem.

8. p-adická čísla a Hasse–Minkowského věta

Na konci 19. století zavedl Kurt Hensel do teorie čísel nový prostředek, tzv. p-*adická čísla*, který umožňoval vyšetřovat klasické, ryze aritmetické problémy prostředky matematické analýzy. Ukažme si zde tedy příslušnou analytickou motivaci.

8.1. Absolutní hodnota a metriky

Absolutní hodnota $|.|$ pro racionální čísla je dobře známá funkce, jejímž smyslem je určovat *velikost* jednotlivých čísel. Jde o zobrazení $|.| : \mathbb{Q} \longrightarrow \mathbb{R}_+$ definované jednoduchým předpisem

$$|X| = \begin{cases} X & \text{pro } X \geqslant 0 \\ -X & \text{pro } X < 0. \end{cases}$$

Je snadno patrné, že tato funkce splňuje následující tři předpoklady.

(i) $|X| = 0$, *právě když* $X = 0$

(ii) $|X \cdot Y| = |X| \cdot |Y|$

(iii) $|X + Y| \leqslant |X| + |Y|$

Zároveň je možné souhlasit s názorem, že tyto tři vlastnosti by měla splňovat jakákoli funkce, která by měla rozumným způsobem stanovovat velikost racionálních čísel. Definujme tedy pojem *zobecněné absolutní hodnoty* jako libovolnou funkci $|.| : \mathbb{Q} \longrightarrow \mathbb{R}_+$ splňující právě uvedené podmínky *(i)–(iii)*. Nejjednodušším příkladem takového zobrazení je funkce přiřazující nule nulu a všem ostatním číslům jedničku. Výpovědní hodnota této funkce je však velmi malá, a proto se touto tzv. *triviální absolutní hodnotou* dále nebudeme zabývat. Velkého významu jsou však tzv. p-*adické absolutní hodnoty*.

Fixujme libovolné prvočíslo p. Každé nenulové celé číslo m pak lze jednoznačně vyjádřit jako $m = p^n m'$, kde $p \nmid m'$. Můžeme tedy zavést funkci $\nu_p(x)$, která každému nenulovému celému číslu x přiděluje počet, kolikrát se prvočíslo p vyskytuje v jeho faktorizaci. Aby byla tato funkce totální, položme dále $\nu_p(0) = +\infty$ (důvodem pro tuto definici je skutečnost, že nulu lze jistě vydělit číslem p, přičemž výsledkem bude opět 0, a tu lze dále dělit číslem p a takto až do nekonečna). Funkci ν_p lze tedy chápat jako *míru dělitelnosti prvočíslem* p a navíc ji lze rozšířit i pro libovolná racionální čísla.

Nechť $M = \frac{n}{d}$. Pak položíme $\nu_p(M) = \nu_p(n) - \nu_p(d)$. Lze snadno ověřit, že pro tuto definici není podstatné, aby $\frac{n}{d}$ bylo základním tvarem čísla M, neboť společné násobky p se vyruší a ostatní faktory jsou irelevantní. Hodnota $\nu_p(M)$ tedy bude kladná, pakliže bude čitatel dělitelnější číslem p více než jmenovatel, a záporná v opačném případě. Lze také nahlédnout, že funkce ν_p pro každé nenulové racionální číslo M je určena formulí

$$M = p^{\nu_p(M)} \frac{n'}{d'}, \qquad \text{kde } p \nmid n' \text{ a } p \nmid d'.$$

Jestliže nyní položíme

$$|X|_p = p^{-\nu_p(X)},$$

získáme tzv. p-*adickou absolutní hodnotu* pro racionální čísla. Pakliže interpretujeme hodnotu $p^{-\infty}$ jako 0, je možné se přesvědčit, že $|0|_p = 0$ a že $|.|_p$ dále splňuje všechny výše uvedené podmínky *(i)–(iii)*. Získáváme tak pro každé prvočíslo p nestandardní, tzv. p-*adickou* absolutní hodnotu $|.|_p$. Tu klasickou pak budeme tradičně označovat jako $|.|_\infty$ a budeme ji nazývat *absolutní hodnotou v nekonečnu.*

Pojem *metriky*, tedy funkce přiřazující libovolným dvěma číslům jejich vzájemnou vzdálenost, bývá na racionálních číslech definován jednoduchým způsobem jako

$$d(X, Y) = |X - Y|.$$

Pro libovolná dvě čísla tedy vezmeme jejich rozdíl a pomocí absolutní hodnoty stanovíme jeho velikost. Pokud však použijeme nestandardní absolutní hodnotu, dostaneme i nestandardní metriku. Vedle klasické euklidovské metriky, vycházející z klasické absolutní hodnoty $|.|_\infty$, tedy můžeme uvažovat i p-*adické metriky*, vyjdeme-li z právě představených p-adických absolutních hodnot.

8.2. Lokální tělesa

Těleso reálných čísel bývá obvykle definováno jako zúplnění racionálních čísel, tak aby každá cauchyovská posloupnost měla limitu. Pakliže však místo klasické euklidovské metriky použijeme metriku p-adickou, získáme jinou topologii na $\mathbb{Q}$, a tedy i jinou množinu cauchyovských posloupností. Stejnou procedurou zúplňování jako reálná čísla tak můžeme získat pro každé prvočíslo p těleso $\mathbb{Q}_p$ coby zúplnění racionálních čísel vzhledem k p-adické

metrice. V souladu s naší předchozí konvencí, kdy jsme značili klasickou absolutní hodnotu $|.|_\infty$, budeme často označovat $\mathbb{R}$ jako $\mathbb{Q}_\infty$. Všechna tělesa $\mathbb{Q}_p$ a $\mathbb{Q}_\infty$ pak budeme souhrnně označovat jako $\mathbb{Q}_v$ a budeme je nazývat *lokálními* tělesy, kdežto původní struktura $\mathbb{Q}$ bude označována jako *globální*.

8.3. Lokálně-globální princip

Při zkoumání řešitelnosti libovolné polynomiální rovnice nad tělesem racionálních čísel je zřejmé, že řešitelnost v $\mathbb{Q}$ implikuje řešitelnost v každém lokálním tělese $\mathbb{Q}_v$. Důvodem k tomu je ta prostá skutečnost, že globální těleso $\mathbb{Q}$ je podstrukturou všech svých zúplnění, a racionální řešení tak bude řešením ve všech lokálních tělesech. Pro některé rovnice se však stává, že platí i opačný směr, tedy že řešitelnost ve všech lokálních tělesech implikuje řešitelnost v tělese globálním. Takováto situace bývá označována jako tzv. *lokálně-globální (též Hasseho) princip*.

Zřejmě nejdůležitější skupinou polynomů, pro něž Hasseho princip platí, jsou kvadratické formy libovolného počtu proměnných. Lokálně-globální vlastnost pro tuto třídu dokázal Helmut Hasse v článku [Has23]. Tento slavný výsledek bývá označován jako *Hasse–Minkowského věta* a šlo tehdy o první velmi významnou aplikaci p-adických těles. Na základě tohoto teorému se totiž můžeme při zkoumání izotropie libovolné racionální kvadratické formy zaměřit pouze na otázku izotropie téže formy nad lokálními zúplněními tělesa $\mathbb{Q}$. Otázka izotropie kvadratické formy nad libovolným lokálním tělesem je však poměrně snadno rozhodnutelná a navíc vzhledem k tomu, že se při aplikaci Hasse–Minkowského věty na libovolnou kvadratickou formu stačí zabývat izotropií pouze nad konečným počtem lokálních těles (ve zbylých bude totiž izotropní triviálně), představuje tento teorém algoritmus schopný v konečném počtu kroků rozhodnout o izotropii libovolné racionální kvadratické formy.

Připomeňme, že jsme v 5. kapitole ukázali, že otázka reprezentovatelnosti libovolného racionálního čísla libovolnou n-ární formou je ekvivalentní otázce izotropie jisté $n+1$-ární formy. Na základě Hase–Minkowského věty tedy lze rozhodovat i o reprezentovatelnosti racionálních čísel racionálními formami.

Rozhodování o izotropii libovolné kvadratické formy má různý, často značně odlišný charakter podle počtu proměnných příslušné formy. Zaměřme se postupně na formy jedné, dvou, tří, čtyř a více než čtyř proměnných.

Kvadratická forma jedné proměnné zřejmě izotropní být nemůže, neboť rovnice $AX^2 = 0$, kde $A \neq 0$, má pouze triviální řešení $X = 0$.

Při řešení izotropie binární formy $AX^2 + BY^2 = 0$ můžeme předpokládat, že X i $Y \neq 0$. Pokud by totiž např. $Y = 0$, dostali bychom rovnici $AX^2 = 0$, která je řešitelná pouze pro $X = 0$, nás však zajímají pouze netriviální řešení rovnice $AX^2 + BY^2 = 0$. Tu tedy můžeme přepsat do tvaru $\frac{X^2}{Y^2} = -\frac{B}{A}$ a následně rozhodnout o její řešitelnosti na základě pouhého prozkoumání, zda je $-\frac{B}{A}$ čtvercem v $\mathbb{Q}$, či nikoli.

I otázku izotropie ternárních forem je možné řešit bez odvolání na Hasse–Minkowského teorém. Větu stanovující nutné a postačující podmínku pro netriviální řešitelnost rovnic tvaru $AX^2 + BY^2 + CZ^2 = 0$ dokázal bez použití konceptu p-adických čísel již Legendre.

Naopak pro kvadratické formy pěti a více proměnných bylo dokázáno, že jsou izotropní nad všemi lokálními tělesy $\mathbb{Q}_p$. Pro jejich izotropii nad tělesem racionálních čísel tak stačí, aby byly izotropní i nad $\mathbb{R}$. Otázka izotropie nad $\mathbb{R}$ je však pro formy libovolného počtu proměnných triviální. Máme-li rovnici $A_1X_1^2 + A_2X_2^2 + \ldots + A_nX_n^2 = 0$, pak pro její netriviální řešitelnost v oboru reálných čísel stačí, aby koeficienty $A_1, A_2, \ldots, A_n$ nebyly všechny zároveň kladné, resp. záporné. Nicméně i tato věta, stanovující, že racionální kvadratická forma pěti a více proměnných je izotropní nad $\mathbb{Q}$, právě když je izotropní nad $\mathbb{R}$, byla dokázána A. Meyerem ještě před zavedením p-adických čísel.

Hlavní význam Hasse–Minkowského teorému tedy leží ve stanovení podmínek izotropie kvaternárních forem. Poznamenejme, že právě tato otázka je předmětem nedokázaných částí lemmat 1 a 2. Při důkazu lokálně-globálního principu pro formy čtyř proměnných se vychází z důkazu téže vlastnosti pro ternární formy. Zcela integrální roli v důkazu pro kvaternární formy však také hraje *Dirichletova věta o aritmetických posloupnostech*. Dirichletova věta zaručuje, že v každé posloupnosti přirozených čísel $a, a + b, a + 2b, a + 3b, \ldots$, kde a a b jsou nesoudělná, se vyskytuje nekonečné množství prvočísel, tedy jinými slovy, že existuje nekonečné množství prvočísel kongruentních s a modulo b (pro důkaz Hasse–Minkowského nicméně postačuje pouze jedno takové prvočíslo). Ačkoli je Dirichletova věta poměrně jednoduše formulovatelné číselněteoretické tvrzení, její důkaz je velmi netriviální a je v něm využíváno mnoha analytických prostředků (funkcí spojitých proměnných, limit, nekonečných sum).

8.4. Odkazy

Velice čtivým úvodem do teorie p-adických čísel je kniha [Gou97], na důkaz Hasse–Minkowského věty zde však nedojde. Klasické důkazy této věty lze nalézt např. v knihách [B–Š66], [Cas78] či v práci [Gam06]. První jmenovaná kniha přitom obsahuje i důkaz Dirichletovy věty. Důkazy obou těchto teorémů lze také nalézt v knize [Ser73]. Zde je Hasse–Minkowského věta vyvrcholením první poloviny knihy, Dirichletova věta je pak jedním z hlavních výsledků druhé části.

Dosud jsme výslovně nezmínili, že se při studiu netriviální řešitelnosti libovolné racionální rovnice tvaru $A_1X_1^2 + A_2X_2^2 + \ldots + A_nX_n^2 = 0$ lze omezit na celá čísla, neboť stačí rovnici vynásobit jmenovateli všech koeficientů $A_1, A_2, \ldots, A_n$ a výsledná rovnice bude netriviálně řešitelná v oboru racionálních čísel, právě když bude mít netriviální celočíselné řešení (tato úvaha byla provedena na začátku 6. kapitoly). Navíc řešitelnost v lokálním p-adickém tělese je v úzkém vztahu s řešitelností téže rovnice modulo všechny mocniny prvočísla p. Hasse–Minkowského větu lze tedy ekvivalentně vyslovit také tak, že libovolná celočíselná rovnice $a_1x_1^2 + a_2x_2^2 + \ldots + a_nx_n^2 = 0$ je netriviálně řešitelná, právě když je netriviálně řešitelná modulo p^r pro každé prvočíslo p a všechna přirozená čísla r a když je zároveň netriviálně řešitelná v oboru reálných čísel. Důkaz takto formulované věty (který se tedy přímo neodvolává na koncept p-adických čísel) lze nalézt v knize [Mor69].

Na závěr ještě poznamenejme, že obecný důkaz Hasse–Minkowského věty je konstruktivní a umožňuje na základě kombinace lokálních řešení příslušné kvadratické formy sestrojovat globální, racionální řešení. Algoritmus takové konstrukce byl také implementován do matematického softwarového balíku Magma a je popsán v článku [S–P04].

Literatura

[B–Š66] Borevič, Zenon Ivanovič; Šafarevič, Igor Rostislavovič. *Number Theory.* Academic Press, New York, 1966.

[Cas78] Cassels, John William Scott. *Rational Quadratic Forms.* Academic Press, New York, 1978.

[Dav52] Davenport, Harold. *The Higher Arithmetic. An Introduction to the Theory of Numbers.* Cambridge University Press, Cambridge, 1952.

[Gam06] Gamzon, Adam. *The Hasse–Minkowski Theorem.* Senior Honors Thesis, University of Connecticut, Storrs, 2006.

[Gou97] Gouvêa, Fernando Quadros. p-*adic Numbers: An Introduction. Second Edition.* Springer-Verlag, New York, 1997.

[Has23] Hasse, Helmut. Über die Darstellbarkeit von Zahlen durch quadratische Formen im Körper der rationalen Zahlen. *Journal für die reine und angewandte Mathematik*, sv. 152 (1923), s. 129–148.

[Mor69] Mordell, Louis Joel. *Diophantine Equations.* Academic Press, London, 1969.

[Rob49] Robinson, Julia. Definability and Decision Problems in Arithmetic. *The Journal of Symbolic Logic*, sv. 14, č. 2 (červen 1949), s. 98 až 114.

[Ser73] Serre, Jean-Pierre. *A Course in Arithmetic.* Springer-Verlag, New York, 1973.

[S–P04] Schicho, Josef; Pilníková, Jana. *Algorithms for Solving Rational Quadratic Forms.* Preprint, 2004. Dostupný online na `http://www.sfb013.uni-linz.ac.at/~sfb/reports/2004/pdf-files/rep_04-32_pilnikova.pdf`