Teorie nejen čísel 1 – Součiny

Milý příteli,
vítáme Tě na začátku letošního seriálu, v němž se naučíme vařit teorii čísel z roztodivných a mnohdy exotických ingrediencí. Jako předkrm vhodíme do hrnce celá čísla a ukážeme si, jak z nich ukuchtit něco jednoduchého na zasycení. Po zbytek našich kuchařských lekcí však již uvidíme, co lze získat, když do hrnce přidáme trochu koření a necháme vše bublat nebo když si pokrm rozkrájíme na menší sousta.

Seriál pro Tebe letos píší Fíla Čermák a Matěj Doležálek. Vyvstanou-li Ti při čtení na mysli jakékoliv otázky či nejasnosti, neváhej se na nás obrátit, například nám můžeš napsat mail na adresy *filip.cermak2@gmail.com* a *matej@prase.cz*.

O co půjde?

Pod honosným názvem *teorie čísel*[1] se skrývá ta oblast matematiky, jež zkoumá hlavně přirozená, celá anebo racionální čísla. Pracuje tedy například s dělitelností, s prvočísly nebo s diofantickými[2] rovnicemi. Tak říkáme rovnicím, v jejichž řešení navíc požadujeme, aby neznámé nabývaly jen celočíselných (někdy dokonce jen přirozených) hodnot.

V tomto seriálu se podíváme na to, jak vést podobné úvahy o dělitelnosti, o speciálních vlastnostech některých „čísel“ a diofantických rovnicích i s jinými objekty než s celými čísly. Typicky tak kupříkladu k celým číslům přidáme nějaký další exotičtější objekt a uvidíme, co nám vznikne a jaké to bude mít vlastnosti. Jindy se naopak rozhodneme něco zapomenout, aby nám zbylo jenom to důležité.

Proč se vůbec o něco takového zajímat

Důvodů, proč se snažit používat nějakou „teorii čísel“ i na jiné struktury než celá čísla, je hned několik. Jednak je to přirozená matematická snaha zobecňovat to, co již známe – tím si můžeme lépe rozmyslet, které vlastnosti jsou na celých čísel výjimečné a které naprosto tuctové. Dále nám ale počítání s novými objekty a strukturami může přinést i nějaká nová pozorování o celých číslech, která bychom jinak získávali jen obtížně.

Nejčastěji nám přidání nových objektů pomůže s rozkládáním výrazů na součin. V celých číslech dovedeme snadno rozložit výrazy jako třeba $a^2 - b^2 = (a - b)(a + b)$, $a^2 + 2ab + b^2 = (a + b)^2$ nebo $ab - a - b + 1 = (a - 1)(b - 1)$. Díky tomu pak mnohé rovnice, které řešíme v celých číslech, dovedeme rozlousknout vhodnou úpravou na součin.

[1] Zastarale též *aritmetika*. S tímto označením se ještě lze setkat v některých odvozených pojmech. V moderním významu se pod pojmem *aritmetika* rozumí spíše základní početní úkony.

[2] Diofantos z Alexandrie (asi 3. století n. l.), řecký matematik, proslul svým spisem *Aritmetika* o řešení (diofantických) rovnic a dalších úloh.

Příklad. (motivační) V diofantické rovnici $ab = a + b$ (tedy hledáme řešení, v němž jsou a, b celá čísla) úpravou do tvaru

$$ab - a - b + 1 = 1,$$
$$(a-1)(b-1) = 1$$

hned zjistíme, že dvě celá čísla $a-1$, $b-1$ dávají v součinu 1. Potom ale nutně buďto $a-1 = 1$ a zároveň $b-1 = 1$ (a tedy $a = b = 2$), anebo $a-1 = b-1 = -1$ (a tedy $a = b = 0$).

Všimněme si, že úprava na součin nám okamžitě omezila možné hodnoty $a-1$, $b-1$. I kdybychom bývali na pravé straně získali jinou hodnotu než 1, stačilo by nám projít konečně mnoho možností, neboť každé celé číslo má jen konečně mnoho dělitelů.

S tímto přístupem ale mnohdy narazíme – výrazy jako $a^2 + b^2$, $a^2 + ab + b^2$, $a^2 - 2b^2$ nebo

$$a^4 + a^3b + a^2b^2 + ab^3 + b^4$$

se nám v celých číslech na součin rozložit nepodaří. Řešením této nesnáze pak může být právě to, že si celá čísla rozšíříme na nějakou „větší“ strukturu, v níž už daný výraz rozložit půjde. Bohužel se však ne vždy budeme moci spolehnout na všechny vlastnosti celých čísel, na něž jsme zvyklí. Klíčové pro nás tedy bude zjišťovat, které hezké vlastnosti celých čísel se při různých způsobech rozšíření zachovají a které již platit přestanou.

Jak seriál číst

Seriál obsahuje celou řadu úkolů na procvičení. Ty, jež jsou označeny „Cvičení“, po Tobě většinou žádají nějak jednoduše použít zavedené pojmy či tvrzení, anebo rozmyslet si nějaký jejich jednoduchý důsledek. Vykřičníkem jsou označena ta cvičení, jež považujeme za obzvláště důležitá a která později využíváme. Cvičení s hvězdičkou jsou pak ta, která z rozličných důvodů považujeme spíše jen za zajímavosti. Mohou být o něco obtížnější a jejich znalost určitě není třeba k pochopení zbytku seriálu. Posledním druhem úkolů jsou pak „Úlohy“, což jsou (alespoň na první pohled) obyčejné příklady, v nichž lze tvrzení a postupy, které si v seriálu ukážeme, s úspěchem využít.

Na konci dílu nalezneš návody k některým cvičením a všem úlohám a také řešení všech cvičení. Doporučujeme tedy zkusit si souběžně se čtením seriálu řešit cvičení, především ta vykřičníková. Pokud se Ti to nebude dařit, nezoufej, zkus si přečíst návod, a pokud si stále nebudeš vědět rady, nahlédni do řešení. Úlohy jsou dosti nezávislé na zbytku seriálu, ničemu tedy nevadí, když jich vyřešíš jenom pár nebo pokud se k nim vrátíš až po dočtení celého dílu. Hvězdičková cvičení můžeš při prvním čtení s klidným svědomím přeskakovat a vrátit se k nim až později, budeš-li chtít.

Ještě než začneme, zmiňme, že některé části seriálu mohou být náročnější na pochopení. Pokud se tedy například zasekneš na některém důkazu, můžeš jej zkusit přeskočit a později se k němu vrátit. Dokud budeš rozumět definovaným pojmům a vědět, co tvrzení říká, nemělo by Tě vynechání důkazů později příliš omezovat.

Úmluva. (značení množin) V seriálu značíme množinu přirozených čísel $\mathbb{N} = \{1, 2, 3, \dots\}$. Nulu nepovažujeme za přirozené číslo. Celá čísla značíme $\mathbb{Z}$, racionální čísla (zlomky) jsou $\mathbb{Q}$ a konečně $\mathbb{R}$ představuje množinu reálných čísel, tedy čehokoliv z reálné přímky: racionální i iracionální čísla jako $\sqrt{2}$, π a mnohá další. Komplexní čísla, která si v průběhu seriálu zavedeme, budeme značit $\mathbb{C}$.

Dále budeme někdy množiny psát notací $\{\text{výraz} : \text{podmínky}\}$. Tím myslíme množinu všech možných hodnot výrazu za splnění podmínek. Například $\{2n : n \in \mathbb{N}\}$ je množina všech sudých přirozených čísel, zatímco $\left\{\frac{a}{b} : a \in \mathbb{Z}, b \in \mathbb{N}\right\}$ je jenom jinak zapsané $\mathbb{Q}$.

Celá čísla

Celá čísla[3] snad není třeba představovat příliš okázale – jedná se zkrátka o čísla

$$\ldots, \quad -3, \quad -2, \quad -1, \quad 0, \quad 1, \quad 2, \quad 3, \quad \ldots,$$

tedy o nulu, přirozená čísla a jejich záporná dvojčata. S celými čísly umíme provádět některé základní operace: lze je sčítat, odčítat a násobit. Při dělení by však výsledkem nemuselo být celé číslo (a dělení nulou není definováno vůbec).

Shrňme některé základní vlastnosti celých čísel a počítání s nimi. Většina z nich má svůj vznešený matematický název, není však příliš důležité si všechny tyto názvy pamatovat:

(1) Sečtením i vynásobením dvou celých čísel dostaneme opět celé číslo, tedy pro $a, b \in \mathbb{Z}$ platí $a + b \in \mathbb{Z}$ i $a \cdot b \in \mathbb{Z}$. Říkáme, že celá čísla jsou na sčítání i násobení *uzavřená*.

(2) Při sčítání i násobení dvou celých čísel nezáleží na pořadí neboli pro celá a, b platí

$$a + b = b + a, \qquad a \cdot b = b \cdot a.$$

Říkáme, že sčítání i násobení celých čísel je *komutativní*.

(3) Při sčítání, resp. násobení tří celých čísel nezáleží na tom, jak tyto operace uzávorkujeme neboli pro celá a, b, c platí

$$a + (b + c) = (a + b) + c, \qquad a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$

Říkáme, že sčítání i násobení celých čísel je *asociativní*.

(4) Násobí-li celé číslo součet dvou dalších, pak lze „roznásobit závorku“ (anebo naopak vytknout ze součtu součinů společný činitel) neboli

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

Říkáme, že násobení je *distributivní* ke sčítání.

(5) V $\mathbb{Z}$ existují výjimečné prvky 0 a 1, které pro každé $a \in \mathbb{Z}$ splňují

$$a + 0 = a, \qquad a \cdot 1 = a, \qquad a \cdot 0 = 0.$$

Říkáme, že 0 je *neutrální prvek* vzhledem ke sčítání, 1 je *neutrální prvek* vzhledem k násobení a 0 je *absorpční prvek* vzhledem k násobení.

(6) Pro každé celé a existuje nějaké celé číslo $-a$, které splňuje $a + (-a) = 0$. Říkáme, že $-a$ je *inverzní prvek* k a vzhledem ke sčítání. Jinými slovy tato vlastnost říká, že celá čísla lze odčítat.

(7) Pro celá a, b platí $a \cdot b = 0$ jenom tehdy, když je $a = 0$ nebo $b = 0$.

Dělitelnost

Když už umíme čísla sčítat, odčítat a násobit, můžeme se podívat i na jejich dělení, které je všem určitě dobře známé. Pro jistotu to ale lehce zopakujeme.

Definice. Číslo a *dělí* číslo b (značíme $a \mid b$), pokud existuje nějaké číslo k takové, že $b = ka$.

Dělitelnost má spoustu hezkých vlastností, které si můžeš dokázat v následujícím cvičení.

Cvičení(!) 1. Mějme nenulová celá a, b, c, d. Potom platí:

(i) Každé a splňuje $1 \mid a$ i $a \mid 0$.

[3] Celá čísla značíme $\mathbb{Z}$, z německého *Zahlen*, v překladu „čísla“ nebo „počty“.

(ii) Když $a \mid b$ a zároveň $b \mid c$, pak i $a \mid c$.

(iii) Když $a \mid b$, pak i $a \mid bc$ pro libovolné c.

(iv) Když $a \mid c$ a zároveň $b \mid d$, pak i $ab \mid cd$.

(v) Když $a \mid c$ a zároveň $a \mid d$, pak i $a \mid c + d$.

(vi) Pokud $a \mid b$ a zároveň $b \mid a$, pak $|a| = |b|$.

Definice. Pokud $a \mid b$ a zároveň $b \mid a$, říkáme, že čísla a, b jsou *asociovaná*, značíme $a \parallel b$. *Jednotkami* jsou ta čísla, která dělí 1.

Jednotkami v $\mathbb{Z}$ jsou tedy čísla 1 a -1. Jak také vidíme z bodu (vi) předchozího cvičení, pokud jsou čísla a a b asociovaná, pak je $a = \pm b$ a naopak.

Cvičení 2. Rozmysli si, že pokud je a asociované s b a b asociované s c, pak je i a asociované s c.

Doteď jsme se bavili o dělitelnosti obecně, ale již ze základní školy každý dobře ví, že v dělitelnosti jsou důležitá třeba prvočísla a jiná podobná zvířátka, tak se na ně pojďme podívat zblízka.

Definice. *Ireducibilním* nazvěme číslo q, které není jednotkou a splňuje, že když $q = ab$, pak jedno z čísel a či b je jednotka.[4]

Definice. *Prvočíslo* je takové číslo p, které není jednotkou a splňuje, že pokud $p \mid ab$, tak i $p \mid a$ nebo $p \mid b$.

Tyto definice se dají samozřejmě rozšířit na součin více čísel. Všimni si také, že při takovéto definici považujeme za prvočísla i záporná čísla -2, -3, -5, -7 a podobně.

Cvičení(!) 3. (i) Nechť je q ireducibilní prvek. Pokud $q = a_1 a_2 \cdots a_n$, potom všechny a_i až na jedno jsou jednotky.

(ii) Nechť je p prvočíslo. Pokud $p \mid a_1 a_2 \cdots a_n$, potom p dělí nějaké z a_i.

V celých číslech nám definice ireducibilních čísel a prvočísel splývají a to se nám líbí, avšak obecně to platit nemusí a taky neplatí, nicméně odstrašující příklady si ukážeme raději až později v kapitole o komutativních okruzích. Předtím než si ukážeme důkaz, že v celých číslech jsou ireducibilní prvky prvočísly, budeme chtít vědět, že pro každou dvojici máme největšího společného dělitele. Druhou implikaci však zvládneme už teď.

Tvrzení. *Každé prvočíslo je ireducibilní.*

Důkaz. Nechť $p = ab$, potom z definice prvočísla víme, že $p \mid a$ nebo $p \mid b$. Z druhé strany však platí $a, b \mid ab \mid p$, tedy určitě $a \parallel p$ nebo $b \parallel p$. To už znamená, že ten druhý činitel musí být jednotkou, takže p je ireducibilní. □

Dělení se zbytkem a Eukleidův[5] algoritmus

Už jsme si ukázali dělení beze zbytku, avšak teď nás čeká i to se zbytkem. Pravděpodobně si jej alespoň matně pamatuješ z prvního stupně, rovnou tedy zformulujeme trochu formálně, o co se jedná.

Tvrzení. (dělení se zbytkem) *Pro libovolná celá čísla a, $b \neq 0$ existují celá čísla q, r taková, že $a = qb + r$ a zároveň $|r| < |b|$.*

Platnost tohoto tvrzení lze snadno nahlédnout. Každé $|b|$-té číslo je násobek b, takže některé z čísel $a, a-1, \ldots, a-|b|+1$ je násobek b. Stačí nám tedy od a odečíst nějaké $r \in \{0, \ldots, |b|-1\}$ a

[4] Ekvivalentně: pokud $q \parallel ab$, pak jedno z čísel a nebo b je s ním asociované.

[5] Eukleidés, někdy též Euklid (asi 325–260 př. n. l.), řecký matematik, položil *Základy* geometrie svým stejnojmenným spisem. Některé části *Základů* se však věnují i teorii čísel jako např. existenci nekonečně mnoha prvočísel či právě Eukleidovu algoritmu.

získáme tím qb. Předtím než pomocí dělení se zbytkem zavedeme *Eukleidův algoritmus*, zadefinujme si největšího společného dělitele, kterého nám tento algoritmus pomůže nalézt.

Definice. Číslo d je *společný dělitel* čísel a, b, pokud $d \mid a$ a zároveň $d \mid b$. *Největším společným dělitelem* (zkráceně NSD) čísel a, b pak myslíme největší takové přirozené číslo, které je společným dělitelem a, b. NSD čísel a, b značíme (a, b).

Teď už víme, co největší společný dělitel je, ale jak ho spočteme? No Eukleidův algoritmus je tu pro nás! Standardní Eukleidův algoritmus slouží ke spočtení největšího společného dělitele. My však nebudeme žádná ořezávátka – raději si přiděláme trochu práce a vyrobíme rovnou i takzvané *Bézoutovy*[6] *koeficienty* x, y, což jsou čísla, pro něž platí $ax+by = (a, b)$. Zároveň tím, že nalezneme algoritmus, který tyto koeficienty najde, dokážeme jejich existenci.[7]

Jak bychom ale mohli začít? Podívejme se na to prvně v přirozených číslech, protože tam je to nejjednodušší. Můžeme si všimnout, že $(a, b) = (b, a - b)$. To lehce ověříme třeba tak, že pokud má a i b společného dělitele, pak dělí i $a - b$. Naopak pokud existuje nějaký společný dělitel d čísel b a $a - b$, pak děli i $d \mid b + (a - b) = a$. Z toho tedy už plyne, že největší společní dělitelé těchto dvojic se rovnají.

Spočítat $(1005, 1000)$ může zprvu vypadat děsivě, ale když si řekneme, že tento NSD je totožný s $(1005, 5)$, už jsme si možnosti značně omezili. Tady vidíme myšlenku, na které je celý Eukleidův algoritmus založený. Budeme stále zmenšovat danou dvojici a u toho zachovávat stejného největšího společného dělitele, až dorazíme na tak malé číslo, že už bude jasné, co tím největším společným dělitelem je.

Ještě si dovolme malou poznámku. Všimněme si, že pokud $a \geq b$, pak se nám vyplatí b odečíst nejvícekrát, co to jde. To tedy znamená $(a, b) = (b, a-b) = (b, a-2b) = \ldots = (b, a-qb)$, kde $a-qb$ chceme nezáporné. To nám ale už určitě připomíná naše dělení se zbytkem.

Pokud se tedy vrátíme k našemu příkladu s čísly 1005 a 1000, dostaneme že $(1005, 1000) = (1000, 5) = (5, 0) = 5$.

Nyní si to jen zformulujeme pořádně a s přidanou hodnotou Bézoutových koeficientů.

Definice. (rozšířený Eukleidův algoritmus) Mějme jako vstup dvě nenulová celá čísla a, b. Zavedeme posloupnosti $\{a_n\}$, $\{x_n\}$, $\{y_n\}$ následovně:

(i) Počáteční hodnoty nastavme jako

$$a_1 = a, \qquad a_2 = b, \qquad x_1 = 1, \qquad y_1 = 0, \qquad x_2 = 0, \qquad y_2 = 1.$$

(ii) Dále postupně pro $n = 2, 3, \ldots$ použijme dělení se zbytkem na vyjádření $a_{n-1} = qa_n + r$ za splnění $|r| < |a_n|$. Další členy posloupností pak definujme

$$a_{n+1} = r = a_{n-1} - qa_n, \qquad x_{n+1} = x_{n-1} - qx_n, \qquad y_{n+1} = y_{n-1} - qy_n.$$

(iii) Skončeme, jakmile pro nějaké N dostaneme $a_N = 0$.

Za okamžik si ukážeme, že z takto definovaného algoritmu získáme na pozici $N - 1$ největší společný dělitel spolu s Bézoutovými koeficienty. Na první pohled však nemusí být zcela průhledné, co že to náš algoritmus dělá. Tak si ho pojďme ukázat na příkladě a poté i dokázat, že opravdu počítá, co chceme.

Příklad. Najděme největší společný dělitel čísel 13 a 5:

Posloupnost $\{a_n\}$:	13	$\to$	5	$\to$	3	$\to$	2	$\to$	1	$\to$	0.
Posloupnost $\{x_n\}$:	1	$\to$	0	$\to$	1	$\to$	-1	$\to$	2	$\to$	-5.
Posloupnost $\{y_n\}$:	0	$\to$	1	$\to$	-2	$\to$	3	$\to$	-5	$\to$	13.

[6] Étienne Bézout (1730–1783), francouzský matematik.

[7] Pozor, Bézoutovy koeficienty nejsou jednoznačné. Platí totiž například $a(x - b) + b(y + a) = ax + by = (a, b)$.

Ve sloupečcích si můžeme všimnout, že platí $a_n = x_n a + y_n b$. Ve výsledku jsme spočetli, že NSD je 1 spolu s vyjádřením $1 = 2 \cdot 13 - 5 \cdot 5$.

Tvrzení. *Rozšířený Eukleidův algoritmus najde NSD čísel a, b. Platí $a_{N-1} = (a, b)$ a navíc pro $x = x_{N-1}$, $y = y_{N-1}$ máme vyjádření $(a, b) = ax + by$.*

Důkaz. Uvažujme podle definice rozšířeného Eukleidova algoritmu index N takový, že $a_N = 0$, ale $a_{N-1} \neq 0$. Tato situace určitě někdy nastane, jelikož díky dělení se zbytkem máme vždy $|a_n| > |a_{n+1}|$. Označme $D(u, v)$ množinu všech společných dělitelů u, v. Dokažme si, že pro každé n platí $D(a_{n-1}, a_n) = D(a_n, a_{n+1})$. Tím už bude zřejmé, že

$$(a, b) = (a_1, a_2) = \cdots = (a_{N-1}, a_N) = (a_{N-1}, 0),$$

což je až na znaménko zjevně a_{N-1}. Uvažujme jakéhokoliv společného dělitele d_1 čísel a_{n-1}, a_n. Podle bodu (ii) definice rozšířeného Eukleidova algoritmu pro nějaké číslo q platí

$$a_{n+1} = a_{n-1} - q a_n.$$

Když tedy $d_1 \mid a_{n-1}$ a zároveň $d \mid a_n$, pak určitě i $d_1 \mid a_{n+1}$. Stejně tak ale uvedenou rovnost můžeme přepsat jako $a_{n-1} = a_{n+1} + q a_n$, čímž obdobně získáme, že každý společný dělitel d_2 čísel a_n, a_{n+1} je i dělitelem a_{n-1}. Dohromady jsme tak ukázali, že $D(a_{n-1}, a_n) = D(a_n, a_{n+1})$.

Tím jsme již napůl vyhráli, víme, že $(a, b) = a_{N-1}$. Dokažme dále indukcí, že pro každé n platí $a_n = x_n a + y_n b$. Pro $n = 1$ a $n = 2$ to platí díky nastavení počátečních hodnot $\{x_n\}$, $\{y_n\}$. Nyní nechť rovnost platí pro $n-1$ a n a dokažme ji pro $n+1$. Opět pro číslo q získané z dělení se zbytkem s použitím indukčního předpokladu dostaneme

$$\begin{aligned} a_{n+1} &= a_{n-1} - q a_n = (a x_{n-1} + b y_{n-1}) - q(a x_n + b y_n) = \\ &= (x_{n-1} - q x_n) \cdot a + (y_{n-1} - q y_n) \cdot b = x_{n+1} a + y_{n+1} b. \end{aligned}$$

Indukcí pak toto platí pro všechna n, takže speciálně $(a, b) = a_{N-1} = x_{N-1} a + y_{N-1} b$, jak jsme chtěli. □

Úloha 1. Dokaž, že pro každé přirozené n je zlomek $\frac{21n+17}{5n+4}$ v základním tvaru, tedy že čitatel a jmenovatel jsou nesoudělní.

Pomocí Eukleidova algoritmu jsme již NSD nalezli, takže si pojďme s jeho pomocí dokázat, že každý ireducibilní prvek je také prvočíslo.

Tvrzení. *Je-li celé číslo p ireducibilní, pak je to také prvočíslo.*

Důkaz. Mějme p ireducibilní a předpokládejme, že platí $p \mid ab$ pro nějaká nenulová a, b. Chceme dokázat, že p musí dělit jedno z a, b.

Pro spor předpokládejme, že $p \nmid a$ a zároveň $p \nmid b$. Jelikož je p ireducibilní, tak jsou jeho děliteli pouze ± 1 a $\pm p$. Proto když $p \nmid a$, tak už musí být $(a, p) = 1$. Existují Bézoutovy koeficienty x_1, y_1 splňující $x_1 a + y_1 p = 1$. Obdobně je $(b, p) = 1$, takže existují Bézoutovy koeficienty x_2, y_2 splňující $x_2 b + y_2 p = 1$. Znásobením těchto dvou rovností získáme

$$x_1 x_2 ab + x_1 y_2 ap + y_1 x_2 bp + y_1 y_2 p^2 = 1.$$

Na levé straně jsou ab, ap, bp i p^2 násobky p, takže i celá levá strana je násobkem p. To ale znamená $p \mid 1$, což je spor, protože ireducibilní prvek není jednotkou. To znamená, že určitě muselo nastat $p \mid a$ nebo $p \mid b$, jak jsme chtěli. □

Proč je prvočíselný rozklad jednoznačný?!

Prvně dokažme, že jsou celá čísla součinem ireducibilních a jednotky (zatím bez jednoznačnosti).

Lemma. *Každé nenulové celé číslo n je součinem ireducibilních čísel a jednotky.*

Důkaz. Čísla 1 a -1 jsou jednotky, tedy součin jednotky a žádných ireducibilních čísel. Dále postupujme silnou indukcí podle velikosti čísel v absolutní hodnotě: nechť pro všechna nenulová $|n| \leq N$ platí, že jsou součinem ireducibilních čísel a jednotky. Poté musí pro $|n| = N+1$ být n buďto samo ireducibilní, nebo platí $n = k\ell$, kde $|k|, |\ell| > 1$. Pak ale musí $|k|$ i $|\ell|$ být menší než $N+1$, tedy menší nebo rovna N. Potom z indukce víme, že k i ℓ jsou součinem ireducibilních čísel a jednotky. To samé tedy platí pro n, neboť nám stačí zapsat činitele z k a z ℓ za sebe a vynásobit spolu příslušné jednotky. □

Nyní už máme všechen potřebný aparát k tomu, abychom odpověděli na otázku z nadpisu sekce. Víme, že ireducibilní čísla jsou totéž co prvočísla a že každé celé číslo lze rozložit na součin ireducibilních čísel. Zbývá tak dokázat jednoznačnost.

Tvrzení. (Základní věta aritmetiky) *Každé nenulové celé číslo n se dá jednoznačně (až na pořadí a změny znamének) rozložit na součin prvočísel.*

Důkaz. Mějme tedy $\pm p_1 \cdots p_r = n = \pm q_1 \cdots q_s$, kde p_i a q_i jsou prvočísla. Potom bychom chtěli dokázat, že $r = s$ a prvočísla napravo lze spárovat s těmi nalevo tak, aby spárovaná prvočísla byla asociovaná neboli se lišila nanejvýš znaménkem.

Použijme znovu silnou indukci podle absolutní hodnoty. Pro $|n| = 1$ je tvrzení triviální. Dále předpokládejme, že tvrzení platí pro všechna n splňující $|n| \leq N$. Jelikož platí $p_1 \cdots p_r = q_1 \cdots q_s$, musí p_1 z definice prvočísla dělit některý činitel napravo. BÚNO[8] nechť $p_1 \mid q_1$. Jelikož je q_1 ireducibilním prvkem, tak je $p_1 = \pm q_1$. Pokrácením těchto prvočísel dostaneme, že $\pm p_2 \cdots p_r = \pm q_2 \cdots q_s$. Přitom však $|p_2 \cdots p_r| = \left|\frac{n}{p}\right| < |n| = N+1$, takže $|p_2 \cdots p_r| \leq N$. Tento součin už je tedy z indukčního předpokladu jednoznačný. Platí tudíž $r - 1 = s - 1$ neboli $r = s$. Dále musí zbývající prvočísla být spárována do asociovaných dvojiček, takže spolu s $p_1 \parallel q_1$ už je celý rozklad jednoznačně určen. □

Poznámka. Určitě všichni víme, že se normálně nepíše $24 = 2 \cdot 2 \cdot 2 \cdot 3$, nýbrž $24 = 2^3 \cdot 3$. Tento zkrácený zápis si můžeme dovolit nyní i my. Když v jednoznačném rozkladu spojíme všechna navzájem asociovaná prvočísla do jedné mocniny, dostaneme rozklad ve tvaru $n = \pm p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, kde pro $i \neq j$ platí $p_i \nmid p_j$. Potom platí $b \mid a$ právě tehdy, když $b = \pm p_1^{\beta_1} \cdot \cdots \cdot p_k^{\beta_k}$ pro nezáporná celá β_i splňující $\alpha_i \geq \beta_i$.

Úloha 2. Najdi všechny dvojice přirozených čísel x, y takové, že $x + y + 1 \mid 2xy$ a zároveň $x + y - 1 \mid x^2 + y^2 - 1$.

Nyní bychom se chtěli podívat, k čemu nám jednoznačný rozklad je.

Definice. Řekneme, že celá čísla a, b jsou *nesoudělná*, pokud jsou jejich společnými děliteli pouze jednotky neboli když $(a, b) = 1$.

Tvrzení. (mocniny a nesoudělnost) *Nechť pro nesoudělná a, b platí $ab = c^n$. Potom existují celá čísla k, ℓ splňující*

$$a = \pm k^n, \qquad b = \pm \ell^n$$

a zároveň $k\ell = \pm c$.

Důkaz. Zapišme si rozklad na prvočísla $c = \pm p_1^{\alpha_1} \cdot \cdots \cdot p_r^{\alpha_r}$, kde p_i jsou navzájem neasociovaná. Potom víme, že rozklad c^n je $\pm p_1^{n\alpha_1} \cdot \cdots \cdot p_r^{n\alpha_r}$. Jelikož jsou a, b nesoudělná, nemůže se v jejich rozkladu vyskytovat stejné prvočíslo. Proto pokud $p_i \mid a$, pak už se v a vyskytuje p_i ve stejné mocnině jako v c^n. Stačí tedy definovat k jako součin těch $p_i^{\alpha_i}$, pro která $p_i \mid a$. Totéž provedeme pro b. Tím jsme naše tvrzení dokázali. □

[8] Zkratka pro „bez újmy na obecnosti“. Myslíme tím, že se nabízí několik možností, které jsou ale stejné až na nějaký nepodstatný detail, takže si prostě vybereme jen jednu a pokračujeme.

Příklad. V celých číslech vyřešme rovnici $x^2 + x = y^3$.

Řešení. Upravíme rovnici do tvaru $x(x+1) = y^3$. Jelikož x a $x+1$ jsou nesoudělná, musí být oba činitelé jednotka krát třetí mocnina. Platí však, že 1 i -1 lze zapsat jako třetí mocniny celého čísla, takže můžeme BÚNO „vložit“ jednotku do mocniny a předpokládat $x = a^3$, $x+1 = b^3$ pro nějaká $a, b \in \mathbb{Z}$. Z toho máme

$$1 = b^3 - a^3 = (b-a)(b^2 + ba + a^2).$$

Musí platit $b > a$, takže už určitě $b - a = 1$. Následně má být

$$1 = b^2 + ba + a^2 = (a+1)^2 + (a+1)a + a^2 = 3a^2 + 3a + 1,$$

takže $3a(a+1) = 0$. To znamená $a = 0$ nebo $a = -1$, z čehož dopočítáme řešení $(x, y) = (0, 0)$ a $(x, y) = (-1, 0)$.

Poznámka. Zatím jsme tvrzení o nesoudělnosti a mocninách použili v případě, kdy jsou v rovnici $ab = c^n$ čísla a, b nesoudělná. Co když to ale není tak snadné? Nechť jsou a, b soudělná nějakým prvočíslem p. Pak i $p \mid c$, takže můžeme zapsat $a = pa_0$, $b = pb_0$, $c = pc_0$ a upravit rovnici na

$$a_0 b_0 = p^{n-2} c^n.$$

Takto lze pokračovat, až na levé straně dostaneme nesoudělné činitele. Na pravé straně nám sice mohou zbýt nějaké mocniny prvočísel, ale to příliš nevadí – z nesoudělnosti činitelů nalevo si každá tato mocnina bude muset vybrat, zda dělí a_0, nebo b_0. Nějaká soudělnost a, b nám tedy typicky jenom přidělá práci s rozebíráním několika případů. V následujících cvičeních si můžeš vyzkoušet, jak v takovém případě postupovat.

Cvičení 4. Vyřeš v celých číslech rovnici $x^2 = y^3 + 16$.

Úloha 3. Najdi všechna celočíselná řešení rovnice $n(n+1)(n+2)(n+3)(n+4) = k^2$.

Úloha 4. Najdi všechny dvojice přirozených prvočísel (p, q), pro něž je $p^2 + 5pq + 4q^2$ druhou mocninou celého čísla.

Modulíme

Nyní už máme spoustu znalostí o dělitelnosti v celých číslech. Pojďme si ještě ale zavést jeden zápis, který je velmi častý, dost možná jej již znáš a ve světe teorie čísel nám usnadní práci.

Definice. Skutečnost, že $n \mid a - b$, značíme $a \equiv b \pmod{n}$ a říkáme, že a je *kongruentní* s b *modulo* n. Množině všech čísel $a + xn$ pro $x \in \mathbb{Z}$ říkáme *zbytková třída*. Množinu všech zbytkových tříd značíme $\mathbb{Z}_n$.

Možná jsi již s kongruencemi někdy pracoval(a), ale nikdy není na škodu si zopakovat, proč v nich některé věci platí.

Doporučujeme nad kongruencemi nepřemýšlet tak, že se jedná o vztah mezi celými čísly z množiny $\mathbb{Z}$, nýbrž tak, že počítáme se zbytkovými třídami z množiny $\mathbb{Z}_n$. Jak totiž uvidíme v následujícím cvičení, výsledky běžných operací typicky nezáleží na tom, se kterým zástupcem z dané zbytkové třídy je provedeme. Chytrým vybíráním si tak často lze ušetřit práci: například $99^2 \equiv (-1)^2 \equiv 1 \pmod{100}$ je snazší než $99^2 \equiv 9801 \equiv 1 \pmod{100}$.

Cvičení(!) 5. Pokud $a \equiv b \pmod{n}$, $c \equiv d \pmod{n}$ a k je nějaké číslo, pak platí i:

(i) $a + k \equiv b + k \pmod{n}$,
(ii) $a \cdot k \equiv b \cdot k \pmod{n}$,
(iii) $a + c \equiv b + d \pmod{n}$,
(iv) $a \cdot c \equiv b \cdot d \pmod{n}$.

Poslední věc, která nám schází, je krácení. S tím je ale trochu potíž, protože nemůžeme vždy jenom zkrátit čísla na obou stranách kongruence: kongruence $2 \equiv 10 \pmod 8$ je pravdivá, ale „zkrácením“ dvojky bychom dostali $1 \equiv 5 \pmod 8$, což neplatí. Proto musíme zkrátit i modul a získáme platnou kongruenci $1 \equiv 5 \pmod 4$. Nyní si to pojďme dokázat obecně.

Cvičení 6. Nechť $ac \equiv bc \pmod n$.

(i) Pokud navíc $(n, c) = 1$, pak $a \equiv b \pmod n$.

(ii) Obecněji bez nároků na c platí $a \equiv b \pmod{\frac{n}{(n,c)}}$.

Abychom si trochu s kongruencemi pohráli, zkusíme si nějaké to lehčí cvičení.

Cvičení 7. Urči, jaký zbytek dává 5^{20} po dělení 26.

Cvičení 8. Urči paritu, poslední cifru a zbytek po dělení sedmi čísla $N = 22 \cdot 31 + 11 \cdot 17 + 13 \cdot 19$.

Cvičení(!) 9. Rozmysli si, že x^2 dává po dělení čtyřmi zbytek 0, je-li x sudé, či 1, je-li x liché.

Nyní už umíme s kongruencemi nějak pracovat, ještě jednou se pojďme zaměřit na dělení. Přesněji na to, kdy lze zbytkovou třídou $a \bmod n$ dělit.

Tvrzení. *Pokud je a nesoudělné s n, pak existuje číslo x, takové že $ax \equiv 1 \pmod n$. Toto číslo zapisujme jako $\frac{1}{a}$.*

Důkaz. Kongruence $ax \equiv 1 \pmod n$ je ekvivalentní s dělitelností $n \mid ax - 1$, což je znovu ekvivalentní s $nk = ax - 1$ neboli $n(-k) + ax = 1$. Díky tomu, že už známe Eukleidův algoritmus, však víme, že pro nesoudělná a, n dovedeme najít takové Bézoutovy koeficienty $-k$, x. □

Vzhůru do komplexních čísel

Nyní se ale posuňme o kus dál. V úlohách, které jsme viděli, nám často pomohlo rozložit nějaký výraz na součin. Úskalím je tomu ale fakt, že ne všechny výrazy jde na součin rozložit. Například $a^2 + b^2$ se nám v celých či reálných číslech nepodaří rozložit na součin dvou výrazů s a, b. Abychom toto zklamání napravili, přidáme si nová čísla. Konkrétně, prohlašme, že existuje nějaké „číslo“ i (bývá nazýváno *imaginární jednotkou*), které splňuje $i^2 = -1$. Takové i určitě nenajdeme mezi reálnými čísly, nenechme se však tímto odradit. Prostě si vymyslíme takové i a řekneme, že se naučíme počítat s čísly tvaru $a + bi$, kde a, b mohou být reálná čísla (později se omezíme na celá).

Čísla $a + bi$ nazýváme *komplexní* a jejich množinu značíme $\mathbb{C}$. Hned si rozmysleme, že počítání s nimi není nic děsivého. Chceme-li sečíst $a + bi$ s $c + di$, jednoduše spočteme

$$(a + bi) + (c + di) = (a + c) + (b + d)i.$$

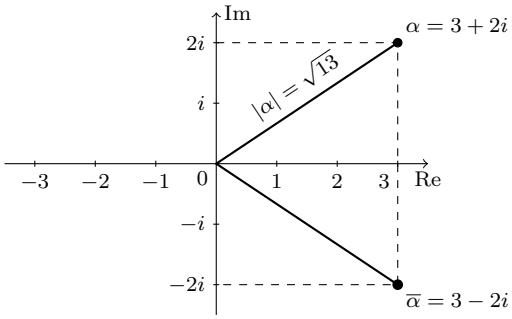
Odečítání provedeme obdobně. Budeme-li chtít komplexní čísla násobit, zkrátka roznásobíme závorky a dostaneme

$$(a + bi) \cdot (c + di) = ac + bci + adi + bdi^2.$$

Když však využijeme $i^2 = -1$, což je jediná podmínka, kterou po i vůbec požadujeme, získáme $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$. Takto definované operace s komplexními čísly mají všechny obvyklé vlastnosti, na které jsme zvyklí z reálných čísel: nezáleží v nich na pořadí, můžeme roznásobovat závorky a tak podobně. My tyto vlastnosti nebudeme dokazovat příliš dopodrobna. Pokud se s komplexními čísly potkáváš poprvé, můžeme Tě ujistit, že z hlediska upravování výrazů zde vše funguje stejně jako v reálných číslech.[9]

[9] Pokud se o nich chceš dozvědět více, doporučujeme PraSečí seriál o komplexních číslech: *https://prase.cz/archive/30/9.pdf*.

Abychom mohli vymyslet dělení komplexních čísel, nakreslíme si nejprve obrázek. Stejně jako si reálná čísla kreslíme na reálnou přímku, budeme komplexní čísla kreslit do *komplexní* (nebo též *Gaussovy*[10]) *roviny*. Komplexní číslo $\alpha = a + bi$ zde zakreslíme na pozici se souřadnicí a na *reálné ose* a b na *imaginární ose*. Také nazýváme a *reálnou* a b *imaginární částí* komplexního čísla α.



Pro komplexní číslo $\alpha = a + bi$ v komplexní rovině definujeme jeho *komplexně sdružené číslo*[11] jako to, které odpovídá bodu překlopenému podle reálné osy, tedy obrácení znaménka imaginární části. Značíme jej $\overline{\alpha} = a - bi$ (čteme „alfa s pruhem“). Všimněme si, že

$$\alpha \cdot \overline{\alpha} = (a + bi)(a - bi) = a^2 - (bi)^2 = a^2 + b^2.$$

Díky Pythagorově větě je toto druhá mocnina délky úsečky spojující α v komplexní rovině s bodem 0. To nám umožňuje definovat absolutní hodnotu komplexního čísla jako $|\alpha| = \sqrt{\alpha\overline{\alpha}} = \sqrt{a^2 + b^2}$. Rozmysli si, že pro reálné číslo se tato definice shoduje s běžnou absolutní hodnotou.

S pomocí komplexně sdruženého čísla už můžeme zavést dělení. Pokud je komplexní číslo $a + bi$ nenulové, pak $a^2 + b^2 > 0$, takže můžeme při dělení rozšířit komplexně sdruženým číslem a spočíst

$$\frac{c + di}{a + bi} = \frac{(c + di)(a - bi)}{(a + bi)(a - bi)} = \frac{(ac + bd) + (ad - bc)i}{a^2 + b^2} = \frac{ac + bd}{a^2 + b^2} + \frac{ad - bc}{a^2 + b^2}\,i.$$

Gaussova celá čísla

Již jsme si zavedli komplexní čísla s reálnými složkami. Abychom se s nimi mohli pustit do nějaké teorie čísel, bylo by záhodno si je omezit na něco podobného celým číslům. Přirozenou volbou je povolit v reálné i imaginární části pouze celá čísla. Dostaneme takzvaná *Gaussova (celá) čísla*, jimiž myslíme komplexní čísla $a + bi$ pro $a, b \in \mathbb{Z}$. Jejich množinu značíme $\mathbb{Z}[i]$. Je snadné si rozmyslet, že součet i součin dvou Gaussových čísel je opět Gaussovo číslo, takže $\mathbb{Z}[i]$ představuje rozumnou „komplexní verzi $\mathbb{Z}$“ (splňuje vlastnosti (1) až (7), které měla celá čísla). Později v kapitole o komutativních okruzích uvidíme, že v komplexních číslech existují i jiné struktury s podobnými vlastnostmi.

Definice. Pro $\alpha, \beta \in \mathbb{Z}[i]$ řekneme, že α *dělí* β (značíme $\alpha \mid \beta$), pokud existuje $\gamma \in \mathbb{Z}[i]$ splňující $\beta = \alpha\gamma$.

Tato definice dělitelnosti říká v podstatě totéž, co definice dělitelnosti pro celá čísla. Stejně tak pro nenulové α platí $\alpha \mid \beta$, právě když je $\frac{\beta}{\alpha}$ Gaussovo celé číslo. Pojďme si dělitelnost vyzkoušet.

[10] Carl Friedrich Gauss (1777–1855), německý matematik přezdívaný *kníže matematiků*, přispěl k rozvoji mnoha oblastí matematiky i dalších věd. Z jeho díla zmiňme knihu *Disquisitiones Arithmeticae*, kterou položil základy moderní teorie čísel. Jeho jméno v seriálu ještě několikrát potkáme.

[11] V angličtině se setkáme s pojmem *complex conjugate*.

Cvičení 10. Rozhodni, zda $9+3i \mid 12-6i$.

V celých číslech jsme ve spoustě důkazů využívali absolutní hodnotu. Ta nám sloužila k tomu, abychom uměli nějakým způsobem určit velikost čísel. Využívali jsme ji například všude, kde jsme potřebovali nějak indukovat podle velikosti, tedy v důkazu lemmatu o rozkladu na ireducibilní prvky nebo (trochu skrytě) v Eukleidově algoritmu. Abychom tedy s Gaussovými čísly mohli pracovat velmi podobně jako s čísly celými, bylo by skvělé, kdybychom na nich měli něco podobného jako absolutní hodnotu. Absolutní hodnota Gaussova celého čísla však nemusí být celé číslo, například $|3+2i| = \sqrt{13}$. To by nám mohlo činit určité nepříjemnosti, proto budeme radši pracovat s druhou mocninou absolutní hodnoty.

Definice. *Normu* Gaussova celého čísla $\alpha = a+bi$ definujeme jako $N(\alpha) = \alpha\overline{\alpha} = a^2+b^2$.

Norma nám nejenom říká, jak je Gaussovo číslo velké, ale jak si ukážeme, chová se také hezky k násobení. Všimni si, že každé celé číslo a je i Gaussovo celé číslo a platí pro něj platí $N(a) = a^2$.

Tvrzení. *Pro $\alpha, \beta \in \mathbb{C}$ platí $\overline{(\alpha\beta)} = \overline{\alpha} \cdot \overline{\beta}$.*

Důkaz. Prostě to rozpočítejme. Nechť $\alpha = a+bi$, $\beta = c+di$. Potom

$$\alpha\beta = (a+bi)(c+di) = (ac-bd)+(ad+bc)i,$$
$$\overline{\alpha} \cdot \overline{\beta} = (a-bi)(c-di) = (ac-bd)+(-ad-bc)i,$$

takže skutečně $\overline{(\alpha\beta)} = \overline{\alpha} \cdot \overline{\beta}$. □

Důsledek. (multiplikativita normy) *Norma součinu dvou Gaussových čísel je součin jejich norem. Říkáme, že norma je multiplikativní.*[12]

Důkaz. Pro $\alpha, \beta \in \mathbb{Z}[i]$ máme $N(\alpha\beta) = \alpha\beta \cdot \overline{(\alpha\beta)} = \alpha\beta \cdot \overline{\alpha}\overline{\beta} = \alpha\overline{\alpha} \cdot \beta\overline{\beta} = N(\alpha) \cdot N(\beta)$. □

Cvičení 11. Pokud lze přirozená čísla m, n obě vyjádřit ve tvaru x^2+y^2 (pro $x, y \in \mathbb{Z}$), pak lze v tomto tvaru vyjádřit i mn.

Cvičení(!) 12. Pokud $\alpha \mid \beta$, pak i $N(\alpha) \mid N(\beta)$. Opačné tvrzení však neplatí.

Poznámka. Můžeme si všimnout, že je to dobrý aparát k tomu, abychom zjistili, kdy čísla dělitelná nejsou.

Cvičení 13. Rozhodni, zda je $14-3i$ násobkem $3-i$.

Nyní si rozmysleme, jak vypadají jednotky v Gaussových celých číslech. Stejně jako v $\mathbb{Z}$ jednotkami myslíme ta čísla, jež dělí 1.

Tvrzení. *Jednotky jsou právě čísla ± 1 a $\pm i$.*

Důkaz. Určitě vidíme, že tato čísla jednotkami jsou, jelikož $\pm 1 \cdot (\pm 1) = 1$ a $\pm i \cdot (\mp i) = 1$.

Nyní se podívejme, zda nemohou existovat další. Pokud $a+bi \mid 1$, pak nutně platí i $N(a+bi) \mid N(1) = 1$. Jinými slovy $a^2+b^2 = N(a+bi) = 1$. Celá čísla a, b, která splňují naši rovnici, jsou ale pouze ta, která nám dávají jednotky ± 1, $\pm i$. □

Vidíme, že jednotky jsou ta čísla, jejichž norma je jedna, což je dle jejich názvu také velmi přirozené.

Abychom dovršili průzkum vlastností $\mathbb{Z}[i]$, dokážeme si, že i Gaussova celá čísla se rozkládají jednoznačně na součin „prvočísel“. Budeme postupovat stejnou cestou jako v celých číslech. Definice ireducibilního čísla a prvočísla ponechme stejné – ireducibilní číslo nelze rozložit na součin dvou

[12] Obecně říkáme, že funkce f je multiplikativní, pokud splňuje $f(ab) = f(a) \cdot f(b)$. Někdy se tímto označením míní i funkce, která toto splňuje jen pro *nesoudělná* a, b, v seriálu však tento význam používat nebudeme.

nejednotek, zatímco prvočíslo dělí součin jen tehdy, když dělí jeden z činitelů. Budeme si ale muset upravit některé důkazy a rozmyslet si, že vše funguje tak, jak má, i tady.

Náš postup k důkazu Základní věty aritmetiky se odvíjel od dělení se zbytkem.

Tvrzení. *Pro $\alpha, \beta \in \mathbb{Z}[i]$ existují $\gamma, \delta \in \mathbb{Z}[i]$ taková, že $\alpha = \beta\gamma + \delta$ a zároveň $N(\delta) < N(\beta)$.*

V Gaussových číslech to ale není tak snadné. Nechť $\alpha = 27 - 23i$ a $\beta = 8 + i$. Zkusme je vydělit normálně:

$$\frac{\alpha}{\beta} = \frac{27 - 23i}{8 + i} \doteq 2{,}97 - 3{,}25i.$$

V celých číslech bychom si vzali největší násobek β, který je ještě menší než α, tedy podíl $\frac{\alpha}{\beta}$ zaokrouhlený dolů. Kdybychom toto provedli zvlášť v reálné a imaginární složce, bylo by to $2 - 4i$. Potom by ale $\delta = (27 - 23i) - (2 - 4i)(8 + i) = 7 + 7i$. Vidíme, že norma tohoto zbytku je vetší než norma β, což je něco, čeho se chceme vyvarovat. Není tedy nějaká lepší možnost?

Podívejme se na to prvně zase zpátky v číslech celých. Pokud budeme využívat naše tvrzení pro čísla 33 a 7, dostaneme $33 = 7{\cdot}4{+}5$. Ale také bychom místo toho největšího menšího násobku 7 mohli vzít nejbližší násobek (ať už je vyšší, nebo nižší). V tomto případě bychom zvolili $33 = 7 \cdot 5 - 2$. Za cenu toho, že máme záporný zbytek, jsme ho zmenšili (v absolutní hodnotě) nanejvýš na polovinu absolutní hodnoty původního dělitele.

Toho využijeme i zde v Gaussových číslech. Když si vybereme v každé složce číslo nejbližší, pak dostáváme $\delta = (27 - 23i) - (3 - 3i)(8 + i) = -2i$. Tato hodnota už naše tvrzení splňuje, teď už si to pojďme jen dokázat obecně.

Důkaz. Zapíšeme si obecně podíl $\frac{\alpha}{\beta}$, který chceme vyjádřit se zbytkem, jako

$$\frac{\alpha}{\beta} = \frac{\alpha\overline{\beta}}{N(\beta)} = \frac{k + \ell i}{N(\beta)}.$$

Využijeme toho, že v $\mathbb{Z}$ už dělit umíme a umíme to dokonce tak, že jsou zbytky dost malé. Tedy nechť $k = N(\beta)q_1 + x_1$ a $\ell = N(\beta)q_2 + x_2$, kde $|x_1| \leq \frac{N(\beta)}{2}$, $|x_2| \leq \frac{N(\beta)}{2}$. Nyní si to už jenom rozepišme:

$$\frac{\alpha}{\beta} = \frac{k + \ell i}{N(\beta)} = q_1 + q_2 i + \frac{x_1 + x_2 i}{N(\beta)}.$$

Zvolme $\gamma = q_1 + q_2 i$. Potom $\delta = \alpha - \beta\gamma = \frac{x_1 + x_2 i}{\overline{\beta}}$. Potom tedy

$$N(\delta) = N\left(\frac{x_1 + x_2 i}{\overline{\beta}}\right) = \frac{N(x_1 + x_2 i)}{N(\beta)} \leq \frac{\frac{N(\beta)^2}{4} + \frac{N(\beta)^2}{4}}{N(\beta)} = \frac{N(\beta)}{2} < N(\beta). \qquad \square$$

Už tedy umíme dělit se „správným“ zbytkem. V důkazu funkčnosti Eukleidova algoritmu jsme využívali pouze toho, že algoritmus skončí, což bylo způsobeno klesající absolutní hodnotou. S normou namísto absolutní hodnoty bychom mohli téhož využít i zde. Dostali bychom tak rozšířený Eukleidův algoritmus pro Gaussova čísla. Na jeho základě už lze dokázat, že v $\mathbb{Z}[i]$ je každé ireducibilní číslo prvočíslem, z čehož už by vyplynul jednoznačný rozklad Gaussových čísel na prvočísla. Změní se pouze obor čísel, se kterým pracujeme, tedy co pro nás budou prvočísla či jednotky. Jedinou vadou na krase je, že zatím nevíme, jak taková Gaussova prvočísla vypadají.

Nebudeme se s těmito důkazy nyní trápit do detailu, protože si je později ukážeme v obecnější verzi. Namísto Eukléidova algoritmu si ukážeme Bézoutovo lemma, které přináší stejné důsledky, pouze bez algoritmického pohledu na věc.

I bez znalosti Gaussových prvočísel nám jednoznačný rozklad může být užitečný – plyne z něj totiž tvrzení o mocninách a nesoudělnosti.

Úloha 5. Najdi všechny dvojice (x, y) celých čísel takových, že splňují $x^2 = y^3 - 1$.

Komutativní okruhy

Dosud jsme viděli několik struktur, v nichž dovedeme sčítat a násobit s podobnými pravidly jako v celých číslech. Konkrétní podobu těmto strukturám podobným $\mathbb{Z}$ dáme definováním komutativního okruhu.

Definice. *Binární operací* $*$ definovanou na množině M rozumíme zobrazení, které dvojici prvků $a, b \in M$ přiřazuje právě jeden prvek $a * b$ z množiny M.

Neformálně řečeno: binární operace je cokoliv, co dvěma věcem jednoznačně přiřazuje nějakou třetí věc, přičemž výsledek této operace značíme tak, že mezi dva původní argumenty napíšeme znak pro příslušnou operaci.

Příklad. Uveďme některé běžné příklady binárních operací:

- Sčítání „+“, odčítání „−“, násobení „·“ celých čísel nebo dělení „:“ nenulových racionálních čísel. Povšimni si, že při odčítání nebo dělení dvou čísel záleží na pořadí.
- Budiž X nějaká množina. Pro dvě její podmnožiny $A, B \subseteq X$ definujeme
  - jejich *sjednocení* $A \cup B$, jež obsahuje ty prvky, které jsou alespoň v jedné z A, B,
  - jejich *průnik* $A \cap B$, jež obsahuje ty prvky, které jsou v obou A, B,
  - jejich *symetrickou diferenci* $A \oplus B$, jež obsahuje ty prvky, které jsou v právě jedné z množin A, B.
- Uvažme funkce z $\mathbb{R}$ do $\mathbb{R}$. Pro dvě takové funkce f, g definujeme jejich složení $g \circ f$ jako funkci h předepsanou $h(x) = g(f(x))$ pro každé $x \in \mathbb{R}$. V této operaci záleží na pořadí – např. pro $f(x) = 2x$ a $g(x) = x^2$ dostaneme

$$(g \circ f)(x) = (2x)^2 = 4x^2, \qquad \text{ale} \qquad (f \circ g)(x) = 2x^2.$$

Definice. *Komutativním okruhem* nazýváme množinu[13] R (té říkáme *nosná množina* okruhu) spolu s binárními operacemi $+$ a $\cdot$ (nazýváme je „sčítání“ a „násobení“) definovanými na R, které splňují:

(1) Množina R je uzavřená na sčítání i násobení neboli pro libovolná $a, b \in R$ platí

$$a + b \in R, \qquad a \cdot b \in R.$$

(2) Sčítání i násobení jsou komutativní neboli pro libovolná $a, b \in R$ platí

$$a + b = b + a, \qquad a \cdot b = b \cdot a.$$

(3) Sčítání i násobení jsou asociativní neboli pro libovolná $a, b, c \in R$ platí

$$a + (b + c) = (a + b) + c, \qquad a \cdot (b \cdot c) = (a \cdot b) \cdot c.$$

(4) Násobení je distributivní na sčítání neboli pro libovolná $a, b, c \in R$ platí

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

(5) V R existují prvky 0 a 1, které pro každé $a \in R$ splňují

$$a + 0 = 0 + a = a, \qquad a \cdot 1 = 1 \cdot a = a, \qquad a \cdot 0 = 0.$$

(6) Pro každé $a \in R$ existuje nějaký prvek $-a \in R$, který splňuje $a + (-a) = 0$.

[13] Obecný okruh budeme často značit R z německého *(Zahl)ring*.

Kdybychom chtěli být zcela formální a přesní, pak bychom měli vždy, když chceme hovořit o nějakém komutativním okruhu, přesně říci, s jakými operacemi $+$ a $\cdot$ pracujeme, či dokonce které prvky jsou 0 a 1 nebo co je $-a$ pro každé $a \in R$. Komutativní okruh by tak měl být určen šesticí $(R, +, \cdot, 0, 1, -)$. Velmi často je však z kontextu zřejmé, o jakých operacích hovoříme, takže budeme říkat pouze „(komutativní) okruh R“.[14]

Dále si v komutativních okruzích budeme dovolovat běžné zjednodušování zápisu, jaké známe z celých čísel, budeme psát násobení jako ab namísto $a \cdot b$, odčítání jako $a - b$ namísto $a + (-b)$ a mocnění jako a^n namísto $\underbrace{a \cdot a \cdots a}_{n\text{-krát}}$ pro přirozené číslo n.

Poznámka. Možná Tě napadlo, proč říkáme „komutativní“ okruh. Samotným slovem *okruh* se nazývá struktura, u které na rozdíl od komutativního okruhu nepředpokládáme, že násobení je komutativní, takže se může stát, že ab je něco jiného než ba. Takové okruhy však v seriálu nebudeme potkávat.[15] I když tedy v seriálu řekneme jenom „okruh“, bude se takřka vždy jednat o komutativní okruh.

Cvičení(!) 14. Dokaž, že v okruhu existuje právě jedna nula (neutrální prvek vzhledem ke sčítání) a právě jedna jednička (neutrální prvek vzhledem k násobení).

Cvičení(*) 15. Podmínku o komutativitě sčítání můžeme v definici (i nekomutativního) okruhu vypustit: pokud R s operacemi $+$, $\cdot$ splňují pouze podmínky (1), (3) až (6), pak už pro libovolná $a, b \in R$ platí $a + b = b + a$.

Cvičení(*) 16. Podmínku $0 \cdot a$ můžeme v definici okruhu vypustit – pokud R s operacemi $+$, $\cdot$ splňují všechny ostatní podmínky z definice okruhu, pak už pro libovolné $a \in R$ platí $0 \cdot a = 0$.

Příklady (komutativních) okruhů

Viděli jsme abstraktní definici okruhu, nyní si ukážeme konkrétní příklady. Mnohé z nich nás budou v seriálu ještě nějakou dobu provázet.

Příklad. Celá, racionální, reálná i komplexní čísla tvoří komutativní okruhy s obvyklým sčítáním a násobením. Nula a jednička jsou v nich přesně ty, na které jsme zvyklí.

Příklad. Přirozená čísla, lichá celá čísla či sudá celá čísla netvoří s obvyklým sčítáním a násobením okruh. Přirozeným číslům chybí 0, lichým číslům chybí 0 a nejsou uzavřená na sčítání, zatímco sudým číslům chybí 1.

Příklad. Pro přirozené číslo n tvoří množina $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ zbytkových tříd mod n komutativní okruh se sčítáním a násobením zbytkových tříd. Nulou je zde zbytková třída 0, jedničkou zbytková třída 1.

Příklad. Nechť je X množina. Množina všech jejích podmnožin[16] $\mathcal{P}(X)$ spolu s operacemi symetrické diference (jako sčítání) a průniku (jako násobení) tvoří okruh. K tomu si rozmysli, že pro $A, B, C \in \mathcal{P}(X)$ platí

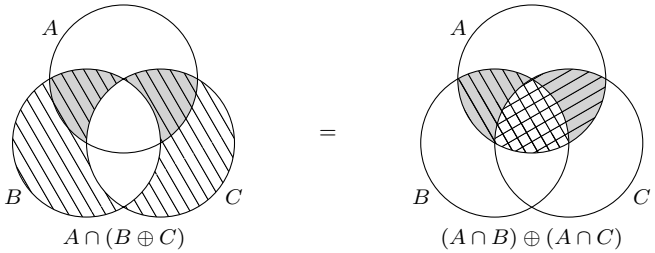
$$A \cap (B \oplus C) = (A \cap B) \oplus (A \cap C)$$

(viz Vennovy diagramy níže). Jedničkou je celá množina X, neboť $X \cap A = A$ pro každou $A \subseteq X$. Nulou je prázdná množina $\emptyset$, neboť $\emptyset \oplus A = A$ pro každou $A \subseteq X$. Odčítání je v tomto okruhu totéž jako sčítání, neboť platí $A \oplus A = \emptyset$ pro každou $A \subseteq X$.

[14] A slibujeme, že kdykoliv si vymyslíme okruh s neobvyklými operacemi, tak na to upozorníme.

[15] Kdyby Tě však zajímaly, doporučujeme jako příklady na seznámení s nekomutativními okruhy tzv. *kvaterniony* nebo čtvercové matice řádu n.

[16] Říká se jí *potenční množina*.



$A \cap (B \oplus C)$ $(A \cap B) \oplus (A \cap C)$

Příklad. Libovolná jednoprvková množina $\{a\}$ tvoří spolu s operacemi sčítání a násobení definovanými jako

$$a + a = a, \qquad a \cdot a = a$$

komutativní okruh. O okruhu s jednoprvkovou nosnou množinou říkáme, že je *triviální*. Všimni si, že v tomto okruhu jsou „nula“ a „jednička“ ten samý prvek.

Cvičení(!) 17. Budiž R komutativní okruh s více než jedním prvkem. Dokaž, že potom v R platí $0 \neq 1$.

Příklad. Gaussova celá čísla $\mathbb{Z}[i]$ tvoří komutativní okruh se sčítáním a násobením komplexních čísel.

Příklad. Nechť $\omega = \frac{1+i\sqrt{3}}{2}$. Potom množina $\mathbb{Z}[\omega] = \{a + b\omega : a, b \in \mathbb{Z}\}$ tvoří komutativní okruh, neboť platí $(a+b\omega)(c+d\omega) = (ac-bd)+(ad+bc+bd)\omega$. Číslům tohoto tvaru se říká *Eisensteinova*[17].

Příklad. Nechť je S množina všech funkcí $f\colon \mathbb{R} \to \mathbb{R}$. Potom S tvoří okruh se sčítáním a násobením funkcí definovaným jako

$$(f + g)(x) = f(x) + g(x), \qquad (f \cdot g)(x) = f(x) \cdot g(x).$$

Nulou a jedničkou jsou konstantní funkce $f_0(x) = 0$ a $f_1(x) = 1$.

Cvičení 18. Nahlédni, že se stejnými operacemi jako v předchozím příkladu tvoří okruh i množina S_2 všech *sudých* funkcí, tedy funkcí $f\colon \mathbb{R} \to \mathbb{R}$ splňujících $f(-x) = f(x)$ pro každé $x \in \mathbb{R}$.

Příklad. Mějme čtyřprvkovou množinu $T = \{0, 1, a, b\}$ a definujme na ní sčítání a násobení pomocí následujících tabulek:

$+$	0	1	a	b
0	0	1	a	b
1	1	0	b	a
a	a	b	0	1
b	b	a	1	0

$\cdot$	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	b	1
b	0	b	1	a

Potom T tvoří s operacemi $+$, $\cdot$ okruh. Všimni si, že ač je to čtyřprvkový okruh, je různý od $\mathbb{Z}_4$.

Příklad. Uvažme všechna racionální čísla, jejichž jmenovatelé jsou mocniny dvou, tedy

$$\mathbb{Z}\left[\frac{1}{2}\right] = \left\{\frac{a}{2^n} : a \in \mathbb{Z}, n \in \mathbb{N}_0\right\}.$$

[17] Gotthold Eisenstein (1823–1852), německý matematik. Ani jeho jméno v seriálu ještě nevidíme naposledy.

Se sčítáním a násobením racionálních čísel tvoří tato množina okruh.

Možná už si všímáš obecného pravidla: když napíšeme $R[m]$, kde m je nějaké „číslo“, myslíme tím nejmenší takový okruh, pro který $R \subseteq R[m]$ a zároveň $m \in R[m]$. Praktičtější je ale představa, že v $R[m]$ prostě leží všechny výrazy $a_n m^n + \cdots + a_1 m + a_0$ pro nezáporné celé n a prvky $a_0, a_1, \ldots, a_n \in R$. V závislosti na tom, co je m zač, se mohou jednotlivá m^k pro $k = 0, 1, 2, \ldots$ postupně opakovat (jako když např. $m = i$ nebo $m = \omega$), anebo se může jednat o navzájem různé prvky $R[m]$ (jako když např. $m = \frac{1}{2}$).

Speciální vlastnosti okruhů

Víme již, co je okruh, je tedy na čase zkoumat, jaké vlastnosti mohou jednotlivé okruhy či jejich prvky mít. Naším cílem bude dopracovat se k „prvočíslům“ v okruzích. Začněme ale u vlastností nuly a jedničky.

Definice. Komutativní okruh nazveme *oborem integrity*, pakliže pro libovolné jeho prvky a, b platí $ab = 0$, pouze pokud $a = 0$ nebo $b = 0$.

Ekvivalentní pohled na obory integrity je, že jsou to přesně ty okruhy, v nichž můžeme v rovnicích krátit nenulové činitele:

Tvrzení. *V oboru integrity pro $a \neq 0$ platí, že pokud $ab = ac$, pak i $b = c$.*

Důkaz. Převedením ac na levou stranu dostaneme $ab - ac = 0$ a poté vytknutím a získáme $a(b - c) = 0$. Následně z definice oboru integrity platí $a = 0$ nebo $b - c = 0$. Předpokládáme však $a \neq 0$, takže už nutně $b - c = 0$ neboli $b = c$. □

Když R není obor integrity, může se stát, že $ab = ac$, i když $b \neq c$. Příkladem budiž okruh $\mathbb{Z}_4$, v němž platí $2 \cdot 1 \equiv 2 \cdot 3 \pmod 4$.

Cvičení 19. Rozmysli si, které z následujících okruhů jsou obory integrity:

(i) $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, (ii) $\mathbb{Z}[i]$, $\mathbb{Z}[\omega]$, $\mathbb{Z}[\frac{1}{2}]$, (iii) triviální okruh,

(iv) $\mathbb{Z}_n$ v závislosti na přirozeném n,

(v) okruh $\mathcal{P}(X)$ v závislosti na množině X,

(vi) čtyřprvkový okruh $T = \{0, 1, a, b\}$ z předposledního příkladu.

Obory integrity nám říkají, že můžeme krátit rovnice, neznamená to ale, že můžeme dělit. Abychom se blíže podívali na to, čím lze dělit, zavedeme si dělitelnost a několik dalších pojmů.

Definice. V komutativním okruhu R pro $a, b \in R$ říkáme, že a *dělí* b (značíme $a \mid b$), pokud existuje $c \in R$ takové, že $ac = b$.

Cvičení(!) 20. Pokud $a \mid b$ a zároveň $a \mid c$, pak už $a \mid bd + ce$ pro libovolná d, e.

Cvičení 21. V okruhu $\mathcal{P}(X)$ platí $A \mid B$, právě pokud $B \subseteq A$.

Definice. V komutativním okruhu R nazveme $u \in R$ *jednotkou*, pokud $u \mid 1$.

Tedy u je jednotka, pokud existuje $v \in R$ takové, že $uv = 1$.

Cvičení 22. Najdi množiny všech jednotek v následujících okruzích:

(i) $\mathbb{Z}$, (ii) $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, (iii) $\mathbb{Z}[i]$, $\mathbb{Z}[\omega]$, (iv) $\mathbb{Z}[\frac{1}{2}]$,

(v) $\mathcal{P}(X)$ v závislosti na množině X.

Cvičení 23. Pokud okruh R není triviální, pak 0 není jednotkou.

Definice. V okruhu R řekneme, že jsou prvky $a, b \in R$ *asociované*, pokud platí $a \mid b$ a zároveň $b \mid a$. Tuto skutečnost budeme značit $a \parallel b$.

Asociované prvky se „chovají stejně“ vzhledem k dělitelnosti. Když $a \parallel b$, pak pro všechna c platí

$$a \mid c \iff b \mid c \qquad \text{a zároveň} \qquad c \mid a \iff c \mid b.$$

Cvičení(!) 24. V oboru integrity jsou a, b asociované, právě pokud existuje jednotka u taková, že $a = ub$.

Ireducibilní prvky a prvočinitelé

Je na čase, abychom se podívali na zobecnění prvočísel. Podobně jako jsme učinili dříve, rozlišíme dvě vlastnosti, které budeme zkoumat: to, jak se prvky rozkládají na součin a jak dělí součiny. Většinou už odteď budeme pracovat jen v oborech integrity, i když některé pojmy definujeme v obecném komutativním okruhu.

Definice. Buď R komutativní okruh a uvažme jeho nenulový prvek p, který není jednotkou. Řekneme, že p je *ireducibilní* (v R), pokud pro každá $a, b \in R$ splňující $ab = p$ platí, že jedno z a, b je jednotka. Řekneme, že p je *prvočinitel* (v R), pokud pro libovolná $a, b \in R$ platí, že když $p \mid ab$, pak $p \mid a$ nebo $p \mid b$.

Vlastnosti ireducibilních prvků a prvočinitelů se jednoduchou indukcí zobecňují pro součiny libovolně mnoha činitelů. Když je součin $a_1 \cdots a_n$ roven ireducibilnímu prvku, pak jsou všichni činitelé krom jednoho jednotkami. Naproti tomu pokud je tento součin násobkem prvočinitele p, pak už musí p dělit některý z činitelů.

Cvičení(!) 25. V oboru integrity jsou prvočinitelé vždy ireducibilní.

Dle našich zkušeností ze $\mathbb{Z}$ a $\mathbb{Z}[i]$ bychom rádi, aby všechny ireducibilní prvky byly prvočiniteli. Často to však neplatí.

Příklad. (varovný) Uvažme okruh $\mathbb{Z}[\sqrt{-3}] = \{a + b\sqrt{-3} : a, b \in \mathbb{Z}\}$, kde $\sqrt{-3}$ je (komplexní) číslo splňující $(\sqrt{-3})^2 = -3$ (kdyby Tě zápis se záporným číslem pod odmocninou zneklidňoval, můžeš si představovat, že $\sqrt{-3} = i\sqrt{3}$). Tento okruh je obor integrity (jelikož $\mathbb{Z}[\sqrt{-3}] \subseteq \mathbb{C}$) a platí v něm

$$2 \cdot 2 = 4 = (1 + \sqrt{-3})(1 - \sqrt{-3}).$$

Ukážeme, že 2 je ireducibilní. K tomu využijeme normu $N(\alpha) = \alpha\overline{\alpha}$, kterou snadno vyjádříme jako $N(x + y\sqrt{-3}) = x^2 + 3y^2$. Kdyby 2 nebyla ireducibilní, měli bychom $2 = ab$, kde a ani b nejsou jednotky. Norma je multiplikativní, tedy $N(a)N(b) = N(2) = 4$. Přitom prvky s normou 1 jsou jednotky, takže potřebujeme $N(a), N(b) > 1$, proto už nutně $N(a) = N(b) = 2$. Pokud ale $a = x + y\sqrt{-3}$ (kde $x, y \in \mathbb{Z}$), dostáváme $2 = x^2 + 3y^2$.

Ukážeme, že tato rovnice nemá v $\mathbb{Z}$ řešení. Když bude $y \neq 0$, tak už $3y^2 \geq 3 > 2$. Tedy nutně $y = 0$, takže by mělo platit $x^2 = 2$, tato rovnice ale nemá v $\mathbb{Z}$ řešení. Hledané prvky a, b tedy neexistují a 2 je ireducibilní. Přitom však není prvočinitelem – dělí součin $(1 + \sqrt{-3})(1 - \sqrt{-3})$, ale nedělí ani jeden z činitelů, neboť ty mají obě složky liché.

Ukážeme si však, jak najít obory, v nichž „prvočísla fungují“, tedy kde všechny ireducibilní prvky jsou prvočinitelé. K tomu si zobecníme postup, který jsme upotřebili v $\mathbb{Z}$. Tam jsme využili vlastnosti dělení se zbytkem a Eukleidova algoritmu, který počítá největšího společného dělitele dvou čísel. Začněme tedy tím, že si tento poslední pojem definujeme i v obecném okruhu.

Definice. Buď R komutativní okruh. Prvek $d \in R$ nazveme *společným dělitelem* prvků $a, b \in R$, pokud $d \mid a$ a zároveň $d \mid b$. Prvek $g \in R$ nazveme *největším společným dělitelem* a, b, pokud současně platí:

(i) Prvek g je společným dělitelem a, b.

(ii) Libovolný společný dělitel d prvků a, b dělí také g.

Řečeno jednodušeji: největší společný dělitel je společný dělitel, kterého všichni společní dělitelé dělí. Všimni si, že tímto není definován jeden jediný největší společný dělitel a, b. Když například největšího společného dělitele přenásobíme jednotkou, dostaneme něco, co je opět největším společným dělitelem. Taky ale nemáme zaručeno, že největší společný dělitel vůbec existuje.

Příklad. (varovný) Uvažme znovu okruh $\mathbb{Z}[\sqrt{-3}]$ a ukažme, že prvky 4 a $2+2\sqrt{-3}$ nemají největšího společného dělitele. Pro spor nechť mají nějakého největší společného dělitele g. Nejprve pozorujme, že 2 i $1+\sqrt{-3}$ jsou společnými děliteli, takže určitě $2 \mid g$ a zároveň $1+\sqrt{-3} \mid g$. Vzhledem k $2 \mid g$ tedy $g = 2a$ pro nějaké $a \in \mathbb{Z}[\sqrt{-3}]$. Také má platit $g \mid 4$, tedy $2a \mid 4$ neboli $a \mid 2$. Z předchozího příkladu už víme, že 2 je ireducibilní, takže každý jeho dělitel je buďto asociovaný s 2 samotnou, nebo s 1. Pokud $a \parallel 2$, pak máme $4 \mid g$, přitom ale $4 \nmid 2+2\sqrt{-3}$, takže toto nelze. Pokud $a \parallel 1$, pak $g \mid 2$, takže i $1+\sqrt{-3} \mid 2$. Když ale tuto dělitelnost rozšíříme pomocí $1-\sqrt{-3}$, dostaneme

$$4 = (1+\sqrt{-3})(1-\sqrt{-3}) \mid 2(1-\sqrt{-3}) = 2-2\sqrt{-3}.$$

Dělitelnost $4 \mid 2-2\sqrt{-3}$ ale neplatí, takže a nemůže být asociováno ani s 1. Vyčerpali jsme všechny možnosti a v každé jsme nalezli spor, takže největší společný dělitel 4 a $2+2\sqrt{-3}$ nemůže existovat.

Cvičení 26. Buď R obor integrity. Pokud jsou oba prvky $g_1, g_2 \in R$ největšími společnými děliteli nějakých $a, b \in R$, pak už jsou g_1 a g_2 asociované.

Eukleidovské a gaussovské obory

Vidíme tedy, že s největšími společnými děliteli bychom měli zacházet opatrně. Jak jsme již ale slíbili, za určitých podmínek můžeme na většinu z těchto obav zapomenout. Motto několika následujících stran zní: pokud umíme dělit s malým zbytkem, pak jsou ireducibilní prvky a prvočinitelé totéž a rozklad na ně je až na jednotky jednoznačný. Po příkladech, které jsme již dříve viděli, není příliš těžké si rozmyslet, co od takového dělení se zbytkem chtít – nějaké ohodnocení prvků oboru integrity, ve kterém lze libovolné nenulové prvky vydělit se zbytkem tak, aby ohodnocení zbytku bylo nižší než ohodnocení dělitele.

Definice. Obor integrity R nazveme *eukleidovským oborem*, pokud existuje funkce $d\colon R \to \mathbb{N}_0$ splňující pro všechna $a, b \in R$, $b \neq 0$, následující podmínky:

(i) Platí $d(a) = 0$, právě když $a = 0$.

(ii) Platí $d(a) \leq d(ab)$.

(iii) Existují $q, r \in R$ taková, že $a = bq + r$ a zároveň $d(r) < d(b)$.

Funkci d budeme říkat *eukleidovská funkce*.

Eukleidovská funkce nám tedy dává ohodnocení prvků oboru integrity, při kterém pouze nula dává nulu, násobky nejsou nikdy menší než dělitelé a umíme dělit se zbytkem tak, aby zbytek byl menší než dělitel.

Cvičení(*) 27. Podmínku (ii) bychom mohli v předchozí definici vynechat. Pokud totiž pro R existuje funkce d splňující (i) a (iii), pak už je funkce $\tilde{d}$ definovaná jako

$$\tilde{d}(a) = \min_{b \neq 0} d(ab)$$

eukleidovskou funkcí.

Nyní v několika krocích dokážeme, že v eukleidovském oboru jsou ireducibilní prvky prvočinitelé a že v nich lze jednoznačně rozkládat na prvočinitele. Začneme Bézoutovou identitou.

Tvrzení. (Bézoutova identita) *Budiž R eukleidovský obor. Pro libovolná nenulová $a, b \in R$ pak existuje $g \in R$ takové, že*

(i) *g je největší společný dělitel a, b,*

(ii) *prvky tvaru $xa + yb$ jsou právě násobky g neboli $\{xa + yb : x, y \in R\} = \{xg : x \in R\}$.*

Důkaz. Pojmenujme $S = \{xa + yb : x, y \in R\}$ a nechť je d eukleidovská funkce. Zvolme g jako nenulový prvek S, pro který je hodnota $d(g)$ minimální ze všech nenulových $g \in S$. Tady využíváme, že hodnoty $d(x)$ (pro $x \neq 0$) jsou přirozená čísla – každá množina přirozených čísel má totiž minimum.

Nejprve ukážeme, že všechny prvky S jsou násobky g. Nechť pro spor existuje nějaké $s \in S$ takové, že $g \nmid s$. Z definice eukleidovského oboru máme $q, r \in R$ taková, že $s = gq + r$ a zároveň $d(r) < d(g)$. Kdyby $r = 0$, znamenalo by to $g \mid s$, takže určitě $r \neq 0$. Nahlédněme také, že $r \in S$, neboť pokud $g = x_1a + y_1b$ a $s = x_2a + y_2b$, pak

$$r = s - gq = (x_2 - qx_1)a + (y_2 - qy_1)b \in S.$$

Získali jsme tedy nenulový prvek S, v němž funkce d nabývá menší hodnoty než v g. To je spor, protože g jsme zvolili tak, aby $d(g)$ bylo minimální.

Ukažme nyní, že g je největší společný dělitel a, b. Víme již, že g dělí všechny prvky S. Když ale zvolíme $x = 1$, $y = 0$, dostaneme $a \in S$, zatímco $x = 0$, $y = 1$ dá $b \in S$. Takže g je jejich společný dělitel. Zároveň když je m společný dělitel a, b, pak určitě i

$$m \mid x_1a + y_1b = g.$$

Ověřili jsme tedy, že g je společný dělitel, jehož všichni společní dělitelé dělí, takže je to největší společný dělitel. □

Důsledek. *V eukleidovském oboru je každý ireducibilní prvek p prvočinitelem.*

Důkaz. Mějme $a, b \in R$ taková, že $p \mid ab$. Pro spor předpokládejme, že p nedělí a ani b. Podle právě dokázaného tvrzení mají a, p nějakého společného dělitele g. Ten může jako dělitel ireducibilního p být buďto asociovaný s 1, nebo s p. Kdyby $p \parallel g$, pak by bylo i $p \mid a$, což (z předpokladu sporu) neplatí. Takže g je asociováno s 1, tudíž BÚNO $g = 1$. Z Bézoutovy identity existují x_1, y_1 taková, že $x_1a + y_1p = 1$. Zcela analogicky je 1 největším společným dělitelem b, p, takže existují x_2, y_2 splňující $x_2b + y_2p = 1$. Když dvě získané rovnosti vynásobíme, dostaneme

$$(x_1a + y_1p)(x_2b + y_2p) = 1,$$
$$x_1x_2ab + x_1y_2ap + x_2y_1bp + y_1y_2p^2 = 1.$$

Předpokládáme $p \mid ab$, takže každý sčítanec na levé straně je násobkem p. Celá levá strana je tak násobkem p, takže jsme dostali $p \mid 1$, což je spor – ireducibilní prvek nemůže být ze své definice jednotkou. Náš předpoklad, že existují a, b splňující $p \mid ab$ a zároveň $p \nmid a, b$, tak byl chybný, takže p musí být prvočinitel. □

Víme už tedy, že všechny ireducibilní prvky jsou v eukleidovském oboru prvočinitelé. Tento výsledek ale ještě vylepšíme na zobecnění Základní věty aritmetiky – ukážeme, že prvky eukleidovského oboru lze „jednoznačně“ rozložit na součin ireducibilních prvků. Tato vlastnost není jen výsadou eukleidovských oborů, pročež má své vlastní označení.

Definice. Řekneme, že obor R je *gaussovský*[18], pokud v něm lze každý nenulový prvek rozložit na součin ireducibilních prvků jednoznačně až na pořadí a asociovanost. Formálněji: každý prvek $a \in R$ se dá zapsat jako $a = u \cdot p_1 \cdots p_n$ pro nějaké ireducibilní prvky[19] $p_1, \ldots, p_n$ a jednotku u, a pokud jsou

$$a = u \cdot p_1 \cdots p_n = v \cdot q_1 \cdots q_m$$

dva takové rozklady na součin ireducibilních prvků, potom $n = m$ a posloupnost $q_1, \ldots, q_n$ se dá přeuspořádat na posloupnost $q'_1, \ldots, q'_n$ takovou, že $p_j \parallel q'_j$ pro $j = 1, 2, \ldots, n$.

[18] V angličtině se setkáme s označením *unique factorization domain*, často zkracováno jako *UFD*.

[19] Může být $n = 0$, v takovém případě je a jednotka.

Poznámka. Když už máme rozklad na ireducibilní prvky, můžeme ho ještě mírně zkrášlit tím, že navzájem asociované ireducibilní prvky „sloučíme“ do jedné mocniny. Dostaneme tak

$$a = u \cdot p_1^{\alpha_1} \cdots p_n^{\alpha_n},$$

kde pro $i \neq j$ platí $p_i \nmid p_j$. V tomto zápisu je zřejmé, jak vypadají dělitelé a – mají v rozkladu pouze ireducibilní prvky $p_1, \ldots, p_n$ (ne nutně všechny), přičemž exponent u p_i je nanejvýš α_i, a libovolnou jednotku.

Cvičení(!) 28. Nahlédni, že v gaussovském oboru musí

(i) každý ireducibilní prvek být prvočinitelem,

(ii) každé dva prvky mít největší společný dělitel.

K důkazu, že eukleidovské obory jsou gaussovské, se propracujeme v několika krocích. Nejprve ukážeme, že každý nenulový prvek lze zapsat jako *nějaký* součin ireducibilních prvků a jednotky. Posléze využijeme toho, že ireducibilní prvky jsou prvočinitelé, abychom ukázali, že každé dva takové rozklady jsou stejné až na pořadí a přenásobení ireducibilních prvků jednotkami.

Lemma. *Nechť je R eukleidovský obor s eukleidovskou funkcí d. Pro nenulová $a, b \in R$ platí $d(a) = d(ab)$ právě tehdy, když je b jednotka.*

Důkaz. Dokazujeme ekvivalenci, ukážeme tedy dva směry implikace. Nejprve nechť je b jednotka. Potom existuje $c \in R$ tak, že $bc = 1$. Z vlastnosti (ii) máme $d(a) \leq d(ab)$, ale také

$$d(ab) \leq d(ab \cdot c) = d(a),$$

takže dohromady $d(a) = d(ab)$.

Nyní nechť naopak $d(a) = d(ab)$. Použijeme vlastnost (iii) na dvojici a, ab. To nám říká, že musí existovat $q, r \in R$ tak, že $a = abq + r$ a zároveň $d(r) < d(ab) = d(a)$. Z toho ale $a(1 - bq) = r$, takže $1 - bq \neq 0$ by vlastností (ii) znamenalo $d(a) \leq d(a(1 - bq)) = d(r)$, což je spor s $d(r) < d(a)$. Určitě tak musí být $1 - bq = 0$ neboli $bq = 1$, což už značí, že b je jednotka. □

Tvrzení. *V eukleidovském oboru R lze každé nenulové $a \in R$ zapsat jako součin jednotky a ireducibilních prvků.*

Důkaz. Nechť je d eukleidovská funkce. Pokud je a jednotka, pak tvrzení platí triviálně – prostě zapíšeme $a = a$, tzn. nepoužijeme v součinu žádné ireducibilní prvky. Nadále předpokládejme, že a není jednotka.

Budeme postupovat silnou matematickou indukcí vzhledem k $d(a)$. To znamená, že budeme předpokládat, že tvrzení platí pro všechna $d(a) \leq n$, a dokážeme jeho platnost i pro $d(a) = n + 1$. Mějme tedy nějaké a takové, že $d(a) = n + 1$ a rozlišme dva případy. Pokud je a ireducibilní, pak máme vyhráno, protože potom $a = a$ je rozklad na součin jediného ireducibilního prvku. Pokud a není ireducibilní, platí $a = bc$ pro nějaká $b, c \in R$, která nejsou jednotky. Z předcházejícího lemmatu tak platí

$$d(b) < d(bc) = d(a) = n + 1,$$

takže určitě $d(b) \leq n$. Obdobně $d(c) \leq n$. Z indukčního předpokladu se tedy b i c dají rozložit na součin ireducibilních prvků a jednotky. Máme tedy

$$b = u \cdot p_1 \cdots p_k \qquad \text{a} \qquad c = v \cdot q_1 \cdots q_\ell$$

pro nějaké jednotky u, v a ireducibilní prvky $p_1, \ldots, p_k, q_1, \ldots, q_\ell$. Z toho pak máme rozklad

$$a = (uv) \cdot p_1 \cdots p_k \cdot q_1 \cdots q_\ell,$$

takže platnost tvrzení je pro dokázána pro $d(a) = n + 1$. Silnou indukcí je tak tvrzení dokázáno pro všechna nenulová $a \in R$. □

Tvrzení. *Každý eukleidovský obor R je gaussovský.*

Důkaz. Víme už, že každý nenulový prvek R lze rozložit na součin ireducibilních prvků a jednotky. Víme také, že všechny ireducibilní prvky v R jsou prvočinitelé. Mějme tedy jednotky u, v, nezáporná celá čísla m, n a ireducibilní prvky $p_1, \ldots, p_n, q_1, \ldots, q_m \in R$, pro něž

$$u \cdot p_1 \cdots p_n = v \cdot q_1 \cdots q_m.$$

Ukážeme, že potom jsou tyto dva rozklady stejné až na pořadí činitelů a asociovanost.

BÚNO nechť $n \geq m$. Pokud nyní $n = 0$, pak máme v rozkladu na obou stranách jen jednotku a není co dokazovat. Jinak vezměme ireducibilní prvek p_1. Jedná se o prvočinitel, který dělí levou stranu, takže musí dělit i některý činitel v součinu na pravé straně. Nemůže být $p_1 \mid v$, protože pak by p_1 byla jednotka. Pro nějaké $j \in \{1, \ldots, m\}$ tedy $p_1 \mid q_j$. Jelikož q_j je ireducibilní, každý jeho dělitel je buďto asociovaný s 1, nebo s q_j. Přitom p_1 není jednotka, takže už musí být p_1 a q_j asociované prvky. Platí tedy $q_j = v_1 p_1$ pro nějakou jednotku v_1. Nyní tak můžeme zapsat $q_1' = q_j$ a rovnost dvou rozkladů pokrátit na

$$u \cdot p_2 \cdots p_n = (vv_1) \cdot q_1 \cdots q_{j-1} q_{j+1} \cdots q_m$$

a proces opakovat: vezmeme p_2, to dělí nějaké q_k na pravé straně, takže $q_2' = q_k$ a opět pokrátíme.

Takto pokračujeme, dokud nepokrátíme všechna $q_1, \ldots, q_m$ na pravé straně (předpokládáme $n \geq m$, takže nám nedojdou ireducibilní prvky na levé straně). Kdyby $n > m$, pak by nám nyní na levé straně zbyly nějaké ireducibilní činitele, které by ale dělili jednotku na pravé straně, což není možné. Určitě tedy $n = m$. Zároveň jsme v procesu krácení získali pořadí ireducibilních prvků $q_1', \ldots, q_n'$, které jsou asociované po řadě s $p_1, \ldots, p_n$. Tím máme dokázáno, že R je gaussovský obor. □

Pozor! Tvrzení neplatí naopak: gaussovský obor vůbec nemusí být eukleidovský.

Příklady eukleidovských oborů

Dokázali jsme spoustu vlastností eukleidovských oborů a několik jsme jich už viděli ($\mathbb{Z}$, $\mathbb{Z}[i]$, $\mathbb{Z}[\omega]$). Eukleidovské obory nemusí být vždy jednoduché rozpoznat, protože obecně vůbec nevíme, jak by měla vypadat vhodná eukleidovská funkce. Ukážeme si proto alespoň několik příkladů. Budeme v nich volit takové funkce, které se chovají hezky k součinu prvků (ještě o něco hezčeji než jenom $d(a) \leq d(ab)$).

Nyní se podívejme na některé další obory, které dovedeme sestrojit v komplexních číslech. Zde může být dobrým tipem zvolit za eukleidovskou funkci normu definovanou jako $N(a) = a \cdot \overline{a}$. Ta totiž splňuje, že hodnoty 0 nabývá pouze v nule, a když navíc zvolíme obor tak, aby hodnoty normy byly celočíselné, pak už pro $b \neq 0$ platí $N(b) \geq 1$, takže i

$$N(ab) = N(a)N(b) \geq N(a) \cdot 1 = N(a).$$

Když tedy v takovémto oboru budeme chtít jako eukleidovskou funkci použít takto definovanou normu, stačí nám ověřit, že nabývá pouze celočíselných hodnot a že dovedeme dělit se zbytkem.

Příklad. Již jsme viděli, že $\mathbb{Z}$ a $\mathbb{Z}[i]$ jsou eukleidovské, neboť jsme v nich zprovoznili dělení se zbytkem. V $\mathbb{Z}$ jsme použili eukleidovskou funkci $d(a) = |a|$, v $\mathbb{Z}[i]$ posloužilo $d(a) = N(a)$.

Příklad. $\mathbb{Z}[\sqrt{-2}]$ je eukleidovský obor s normou jako eukleidovskou funkcí.

Řešení. Pro $x, y \in \mathbb{Z}$ vyjádříme normu jako $N(x + y\sqrt{-2}) = x^2 + 2y^2$, což je celé číslo. Ukažme tedy, jak budeme dělit se zbytkem. Mějme $a, b \in \mathbb{Z}[\sqrt{-2}]$, kde $b \neq 0$. Chceme zvolit $q \in \mathbb{Z}[\sqrt{-2}]$ tak, že

$$N(a - qb) = N(r) < N(b).$$

Vydělme $\frac{a}{b} = \frac{a\cdot\overline{b}}{N(b)} = x + y\sqrt{-2}$, kde $x, y \in \mathbb{Q}$, protože $\overline{b}$ je prvek $\mathbb{Z}[\sqrt{-2}]$, takže celý čitatel $a\overline{b}$ je prvek $\mathbb{Z}[\sqrt{-2}]$. Můžeme tedy využít multiplikativity normy a naši „cílovou" nerovnost ekvivalentně upravit na

$$N\left(\frac{a}{b} - q\right) < 1.$$

Nyní stačí zvolit $q = x_0 + y_0\sqrt{-2}$ tak, že zaokrouhlíme x, y na nejbližší celé číslo (nižší či vyšší podle toho, které je blíž). Tím bude zaručeno $|x - x_0|, |y - y_0| \leq \frac{1}{2}$, což nám dá

$$N\left(\frac{a}{b} - q\right) = (x - x_0)^2 + 2(y - y_0)^2 \leq \left(\frac{1}{2}\right)^2 + 2 \cdot \left(\frac{1}{2}\right)^2 \leq \frac{3}{4} < 1.$$

Tím už je dokázáno, že norma je eukleidovská funkce v oboru $\mathbb{Z}[\sqrt{-2}]$.

Příklad. Uvažme komplexní číslo $\alpha = \frac{1+i\sqrt{7}}{2}$, které splňuje $\alpha^2 = \alpha - 2$. Potom je obor $\mathbb{Z}[\alpha] = \{x + y\alpha : x, y \in \mathbb{Z}\}$ eukleidovský s normou jako eukleidovskou funkcí.

Řešení. Platí $\alpha + \overline{\alpha} = 1$ a $\alpha\overline{\alpha} = \frac{1-(-7)}{4} = 2$, takže pro $x, y \in \mathbb{Z}$ má prvek $x + y\alpha \in \mathbb{Z}[\alpha]$ normu

$$(x + y\alpha)(x + y\overline{\alpha}) = x^2 + xy \cdot (\alpha + \overline{\alpha}) + y^2 \cdot \alpha\overline{\alpha} = x^2 + xy + 2y^2,$$

což je celé číslo. Stejně jako v předchozím příkladu použijeme trik s dělením, takže budeme pro $\frac{a}{b} = x + y\alpha$, kde $x, y \in \mathbb{Q}$, hledat taková $x_0, y_0 \in \mathbb{Z}$, že

$$N\Big((x - x_0) + (y - y_0)\alpha\Big) < 1.$$

Na tento problém můžeme nahlížet geometricky. Nerovnice $x^2 + xy + 2y^2 < 1$ určuje v kartézské rovině vnitřek nějaké elipsy.[20] To, že odečítáme $x_0 + y_0\alpha$, odpovídá tomu, že se souřadnice nějakého zadaného (x, y) snažíme posunout o celá čísla tak, aby výsledek spadl dovnitř této elipsy.

S tím už se lze vypořádat celkem jednoduše. Nejprve zaokrouhlíme y na nejbližší celé číslo y_0. Tím bude rozdíl $y - y_0$ ležet mezi $-\frac{1}{2}$ a $\frac{1}{2}$. Když elipsu protneme s přímkou $y = \frac{1}{2}$, budou x-ové souřadnice průsečíků řešeními rovnice $x^2 + \frac{x}{2} - \frac{1}{2} = 0$. Těmi jsou $x_{1,2} = \frac{-\frac{1}{2} \pm \sqrt{\frac{1}{4} + 4\cdot\frac{1}{2}}}{2}$, takže tyto dva průsečíky jsou od sebe vzdáleny

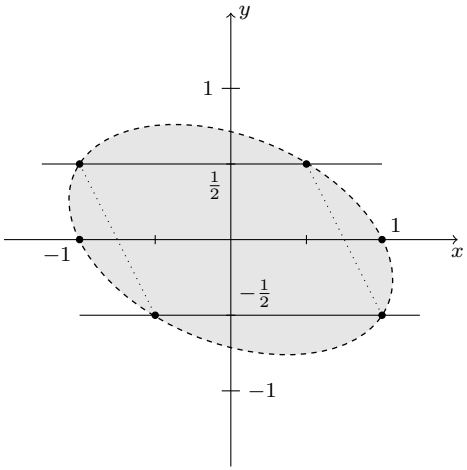
$$\sqrt{\frac{1}{4} + 4 \cdot \frac{1}{2}} = \sqrt{\frac{9}{4}} = \frac{3}{2} > 1.$$

Důležité je, že toto je ostře víc než 1. Na přímce $y = -\frac{1}{2}$ máme další dva průsečíky, středově souměrné podle počátku s prvními dvěma. Tyto čtyři průsečíky jsou vrcholy rovnoběžníku, který leží uvnitř této elipsy (vyjma vrcholů, ty leží na elipse), protože elipsa je konvexní útvar.[21] Tento rovnoběžník má výšku přesně 1 a šířku (délku příslušné strany) ostře větší než 1. Tyto vlastnosti už zaručují, že se do něj dovedeme trefit vhodnou volbou x_0, y_0 – nejprve se výběrem y_0 trefíme mezi $-\frac{1}{2}$ a $\frac{1}{2}$ a následně budeme schopni vybrat x_0, protože se trefujeme do intervalu delšího než 1.

Pro formální správnost postupu jenom musíme rovnoběžník o velmi malý kousek zmenšit, aby ležel v elipse celý, včetně vrcholů. Vrcholy leží přímo na elipse, takže jakkoliv malé zmenšení nám postačí. Je tedy zřejmé, že tohle dovedeme provést tak, abychom zachovali šířku větší než 1.

[20] Že tato nerovnice skutečně určuje elipsu, zde nebudeme dokazovat. Pokud si tedy lámeš hlavou nad tím, proč by tomu tak mělo být, prosíme Tě, abys nám to prostě věřil(a) :-).

[21] Je to totiž jenom splácnutá kružnice.



Cvičení 29. Z dřívějška už víme, že v $\mathbb{Z}[\sqrt{-3}]$ nejsou ireducibilní prvky vždy prvočinitelé, takže to určitě není eukleidovský obor. Rozmysli si, proč zde norma nevyhovuje jako eukleidovská funkce.

K čemu nám je jednoznačný rozklad

Ukažme si konečně, k čemu je nám jednoznačný rozklad na ireducibilní prvky dobrý – umožňuje nám získávat užitečné informace z úpravy rovnice na součin.

Definice. Nechť je R komutativní okruh. Řekneme, že prvky $a, b \in R$ jsou *nesoudělné*, pokud je 1 jejich největším společným dělitelem.

Tvrzení. (mocniny a nesoudělnost) *Nechť je R gaussovský obor. Pokud pro nesoudělná $a, b \in R$, $c \in R$ a přirozené k platí $ab = c^k$, pak už platí*

$$a = ud^k, \qquad b = ve^k,$$

pro prvky $d, e \in R$ a jednotky u, v splňující de $\parallel c$.

Důkaz. Budiž $c = w \cdot p_1^{\gamma_1} \cdots p_n^{\gamma_n}$ rozklad c na součin ireducibilních prvků a jednotky a nechť jsou $p_1, \ldots, p_n$ navzájem neasociované. Potom má prvek c^k rozklad

$$c^k = \left(w^k\right) \cdot p_1^{k\gamma_1} \cdots p_n^{k\gamma_n}.$$

Jelikož a, b jsou nesoudělná, nemohou mít ve svých rozkladech žádné společné ireducibilní prvky. Každé p_i se tak musí vyskytovat v rozkladu právě jednoho z a, b. Z jednoznačnosti rozkladu se v tomto rozkladu už musí vyskytovat v mocnině $p_i^{k\gamma_i}$. Pokud tedy d zavedeme jako součin všech těch $p_i^{\gamma_i}$, pro která $p_i \mid a$, pak můžeme rozklad a zapsat jako $a = ud^k$ pro nějakou jednotku u. Analogicky máme i $b = ve^k$. □

Poznámka. (schovávání jednotky do mocniny) Drobnou nepříjemností v používání tohoto tvrzení je jednotka, kterou může být mocnina přenásobená. Pokud má R mnoho jednotek, může být poměrně namáhavé je vyzkoušet všechny, a pokud jich má dokonce nekonečně mnoho, jedná už se o podstatný problém, který je třeba řešit nějak „chytřeji“ než rozebráním všech možností. Často si ale lze ušetřit práci tím, že jednotku u „BÚNO“ vložíme do mocniny d^k. Pokud totiž víme, že v R lze *každá* jednotka vyjádřit jako k-tá mocnina, pak lze prostě vzít $u = u_0^k$ a následně $a = (u_0 d)^k$. Obecněji můžeme jednotky, které procházíme, zredukovat tak, abychom měli zastoupeny „všechny jednotky až na přenásobení k-tými mocninami jednotek“.

Příklad. Najdi všechna celočíselná řešení rovnice $x^2 = y^3 - 2$.

Řešení. Využijeme obor $\mathbb{Z}[\sqrt{-2}]$, o kterém již víme, že je eukleidovský, a tedy i gaussovský. Rovnost upravíme na

$$x^2 + 2 = y^3,$$
$$(x + \sqrt{-2})(x - \sqrt{-2}) = y^3.$$

Rozmysleme si nyní, jaké společně dělitele mohou mít $x + \sqrt{-2}$ a $x - \sqrt{-2}$. Budiž d nějaký jejich společný dělitel. Potom d dělí i

$$(x + \sqrt{-2}) - (x - \sqrt{-2}) = 2\sqrt{-2}.$$

Když vezmeme normu, tak určitě $N(d) \mid N(2\sqrt{-2}) = 8$. Musí tedy být $N(d) \in \{1, 2, 4, 8\}$. Ukážeme, že $2 \nmid N(d)$, takže už $N(d) = 1$ a d je tak jednotka. Nechť pro spor $2 \mid N(d)$. Vzhledem k $d \mid x \pm \sqrt{-2}$ pak už dostaneme (v $\mathbb{Z}$)

$$2 \mid N(x \pm \sqrt{-2}) = x^2 + 2 = y^3,$$

takže y je sudé. Zároveň také $2 \mid x^2$, takže x je také sudé. Jenže potom už jsou x^2 i y^3 násobky čtyř. Když se nyní na původní rovnici podíváme mod 4, dostaneme $0 \equiv 2 \pmod 4$, což je spor. Určitě $2 \nmid N(d)$, a d je tedy jednotka.

Tímto jsme dokázali, že $x + \sqrt{-2}$ a $x - \sqrt{-2}$ jsou nesoudělná. Když nyní použijeme tvrzení, dostaneme, že $x + \sqrt{-2} = t^3$ pro nějaké $t \in \mathbb{Z}[\sqrt{-2}]$ – jednotku můžeme vložit do mocniny, protože jedinými jednotkami v $\mathbb{Z}[\sqrt{-2}]$ jsou 1 a -1, přičemž obě jsou třetími mocninami. Když tedy zapíšeme $t = a + b\sqrt{-2}$ pro $a, b \in \mathbb{Z}$, dostaneme

$$\begin{aligned} x + \sqrt{-2} = (a + b\sqrt{-2})^3 &= a^3 + 3a^2b\sqrt{-2} + 3a(b\sqrt{-2})^2 + (b\sqrt{-2})^3 = \\ &= a^3 + 3a^2b\sqrt{-2} - 6ab^2 - 2b^3\sqrt{-2} = \\ &= \left(a^3 - 6ab^2\right) + \left(3a^2b - 2b^3\right)\sqrt{-2}. \end{aligned}$$

V tomto tvaru se naše rovnice v $\mathbb{Z}[\sqrt{-2}]$ skládá ze dvou rovnic v $\mathbb{Z}$. Když se podíváme jen na koeficienty imaginárních složek, máme rovnici $1 = b\left(3a^2 - 2b^2\right)$. Pravá strana je násobkem b, ale levá strana je 1, takže b musí být jednotka, tedy 1 nebo -1. Z toho už si můžeme být jisti, že $b^2 = 1$. Upravíme tedy na $\pm 1 = 3a^2 - 2$. Rozlišme dva případy podle toho, jestli je na levé straně 1, nebo -1.

(i) Pokud $b = -1$, dostaneme rovnici $1 = 3a^2$. Tento případ tak nemá řešení, jelikož $3 \nmid 1$.

(ii) Pokud $b = 1$, obdržíme $3 = 3a^2$ neboli $a^2 = 1$. To znamená $a = \pm 1$. Když se v rovnici výše podíváme tentokrát na reálné složky, dostaneme

$$x = a\left(a^2 - 6b^2\right) = \pm 1 \cdot (1 - 6).$$

Řešeními tak mohou být pouze $x = \pm 5$. Snadno vyzkoušíme, že tyto hodnoty dávají řešení původní rovnice a oběma odpovídá $y = 3$.

Jedinými celočíselnými řešeními $x^2 = y^3 - 2$ jsou tedy $(x, y) = (\pm 5, 3)$.

Shrňme si, co náš postup obnášel. Nejprve jsme rovnici upravili na součin ve vhodném oboru, poté jsme se zabývali možnými společnými děliteli činitelů. Zde jsme měli štěstí a podařilo se nám vyloučit všechny až na jednotky – kdyby se toto nepodařilo, museli bychom rozebrat možné hodnoty největšího společného dělitele. Jakmile jsme si byli jisti, že máme součin *nesoudělných* činitelů, bylo již možno použít tvrzení, což nám s trochou vytrvalosti a rozebírání jednotek dalo jednodušší rovnice v $\mathbb{Z}$.

Úloha 6. Najdi všechna celočíselná řešení rovnice $x^2 = y^5 - 1$.

Gaussova celá čísla pod drobnohledem

Vezměme nyní vše, co jsme si dokázali o eukleidovských a gaussovských oborech, a podívejme se důkladněji na $\mathbb{Z}[i]$ a jeho využití. Gaussova celá čísla, potažmo komplexní čísla vůbec jsme původně zavedli proto, abychom uměli rozkládat na součin výrazy jako $a^2 + b^2$. Nemělo by tedy příliš překvapit, že $\mathbb{Z}[i]$ nám přijde k užitku tam, kde najdeme nějaký součet dvou čtverců[22].

Pythagorejské trojice

Jak ví každé školou povinné dítko, v pravoúhlém trojúhelníku je obsah čtverce nad přeponou roven součtu obsahů čtverců nad odvěsnami neboli

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

při obvyklém značení délek stran. Některá z těchto dítek si také možná pamatují, že jeden obzvláště hezký takový trojúhelník se pyšní délkami stran 3, 4 a 5, či možná znají trojúhelník $(5, 12, 13)$. Ta nejbystřejší pak dokonce vymyslí i to, že každý trojúhelník s délkami stran $2n + 1$, $2n^2 + 2n$, $2n^2 + 2n + 1$ je pravoúhlý. Takovémuto popisu však stále uniká spousta pravoúhlých trojúhelníků s celočíselnými délkami stran – pojďme je najít všechny.

Definice. Trojici (a, b, c) přirozených čísel nazveme *pythagorejskou trojicí*, pokud $a^2 + b^2 = c^2$. Pythagorejskou trojici (a, b, c) nazveme *primitivní*, pokud jsou a, b, c nesoudělná.

Máme-li pythagorejskou trojici (a, b, c) a g je největší společný dělitel a, b, c, pak je $\left(\frac{a}{g}, \frac{b}{g}, \frac{c}{g}\right)$ primitivní pythagorejská trojice, takže libovolnou pythagorejskou trojici lze získat z nějaké primitivní přenásobením. Dále si všimněme, že když je (a, b, c) primitivní pythagorejská trojice, tak už jsou a, b, c i *po dvou* nesoudělná. Kdyby totiž např. a, b byla soudělná, pak by nějaké prvočíslo p dělilo obě z nich. Potom ale i

$$p \mid a^2 + b^2 = c^2,$$

takže nutně $p \mid c$, což znamená, že (a, b, c) není primitivní.

Dále se podívejme na to, jak je to v primitivní pythagorejské trojici (a, b, c) s paritou. Jelikož členy trojice jsou po dvou nesoudělné, určitě je nanejvýš jeden z nich sudý. Vzpomeňme si, že x^2 dává mod 4 zbytek 0, když je x sudé, a zbytek 1, když je x liché. Kdyby tedy a, b i c byla lichá čísla, dostaneme $1 + 1 \equiv 1 \pmod 4$, což není možné. Kdyby bylo c sudé, dostaneme $1 + 1 \equiv 0 \pmod 4$, což také neplatí. Musí tedy být sudé jedno z a, b – BÚNO nechť je a liché a b sudé.

Po této přípravě se konečně můžeme na rovnici $a^2 + b^2 = c^2$ podívat v $\mathbb{Z}[i]$ a rozložit levou stranu na

$$(a + bi)(a - bi) = c^2.$$

Podívejme se na společně dělitele závorek na levé straně. Když nějaké $d \in \mathbb{Z}[i]$ dělí obě $a \pm bi$, pak nutně dělí i čísla

$$2a = (a + bi) + (a - bi),$$
$$2b = i \cdot (a - bi) - i \cdot (a + bi).$$

Vzetím normy je $N(d)$ v celých číslech dělitelem $N(2a) = 4a^2$ a $N(2b) = 4b^2$. Obdobně díky $d \mid a + bi$ platí

$$N(d) \mid N(a + bi) = a^2 + b^2 = c^2,$$

[22] V kontextu teorie čísel míníme slovem čtverec druhou mocninu.

což je liché číslo, takže i $N(d)$ je liché. Z toho už je $N(d)$ společným dělitelem a^2 a b^2, což jsou nesoudělná čísla, takže už musí být $N(d) = 1$. V $\mathbb{Z}[i]$ jsou tedy $a + bi$, $a - bi$ nesoudělná.

Na rovnici tak můžeme hodit tvrzení o mocninách a nesoudělnosti. Pro nějaká $x, y \in \mathbb{Z}$ a jednotku $u \in \mathbb{Z}[i]$ tak máme $a+bi = u(x+yi)^2$. Všimni si, že zde nemůžeme jednotku „schovat“ do mocniny, protože např. i je jednotka, kterou nelze vyjádřit jako druhou mocninu. Nahlédněme ale, že násobení jednotkami $u \in \{1, i, -1, -i\}$ nám může jenom prohazovat reálnou a imaginární složku či obracet jejich znaménka. Když se tedy domluvíme, že zanedbáme prohození a, b a budeme je hledat pouze jako přirozená čísla, můžeme „BÚNO“ předpokládat $u = 1$. Potom už snadno dopočítáme

$$a + bi = (x + yi)^2 = (x^2 - y^2) + 2xyi,$$

takže (primitivní) pythagorejské trojice v jistém smyslu odpovídají čtvercům (druhým mocninám) v $\mathbb{Z}[i]$.[23] Aby nyní a, b byla nenulová, budeme předpokládat, že x, y jsou přirozená a navíc $x > y$. Dále dopočítáme

$$c^2 = N(a + bi) = N\left((x + yi)^2\right) = N(x + yi)^2 = (x^2 + y^2)^2,$$

takže $c = x^2 + y^2$.

Doplňme ještě pár podmínek pro x, y. Aby a, b byla nesoudělná, musí i x, y být nesoudělná. Kdyby navíc obě x, y byla lichá, pak je $a = x^2 - y^2$ sudé, takže a, b by byla soudělná, tedy musí *právě jedno* z x, y být sudé. Obecnou (ne nutně primitivní) pythagorejskou trojici pak zpět z primitivní získáme přenásobením všech tří a, b, c nějakým přirozeným k. Shrňme:

Tvrzení. (parametrizace pythagorejských trojic) *Je-li (a, b, c) pythagorejská trojice, pak lze (po možné záměně a, b) vyjádřit*

$$a = k(x^2 - y^2), \qquad b = 2kxy, \qquad c = k(x^2 + y^2),$$

kde x, y, k jsou přirozená čísla, $x > y$, čísla x, y jsou nesoudělná a právě jedno z nich je sudé.

Úloha 7. Pokud přirozená čísla a, b, c splňují $a^2 + b^2 = c^2$, pak je $\frac{1}{2}(c - a)(c - b)$ čtverec přirozeného čísla.

Součty čtverců

Pythagorejské trojice byly odpovědí na otázku, kdy je součet čtverců opět čtvercem. Položme si ale obecnější otázku: která přirozená čísla umí být součtem dvou čtverců, tedy pro která n lze vyjádřit $n = x^2 + y^2$ pro $x, y \in \mathbb{Z}$.

Jako dříve toto přeformulujeme pomocí $\mathbb{Z}[i]$. Výraz $x^2 + y^2$ je jenom norma $x + yi$, což můžeme zvolit zcela libovolně, takže nás zajímá, jaká přirozená čísla se vyskytují jako normy prvků $\mathbb{Z}[i]$. Dále si rozmysleme, že nám vesměs stačí zodpovědět, která přirozená čísla se vyskytují jako normy *prvočinitelů* v $\mathbb{Z}[i]$. Každý prvek $\mathbb{Z}[i]$ se totiž rozkládá (jednoznačně) na součin prvočinitelů a jednotky – když

$$x + iy = up_1 \cdots p_n,$$

kde $p_1, \ldots, p_n$ jsou nějací prvočinitelé a u je jednotka, pak multiplikativitou normy

$$x^2 + y^2 = N(x + yi) = N(p_1) \cdots N(p_n).$$

Jakmile tedy budeme vědět, jaké hodnoty mohou mít normy prvočinitelů, dostaneme možné normy všech prvků $\mathbb{Z}[i]$ jejich vzájemným násobením.

Lemma. *Nechť je*[24] *$\pi \in \mathbb{Z}[i]$ prvočinitel. Potom pro nějaké prvočíslo $p \in \mathbb{Z}$ platí $\pi \mid p$.*

[23] Hezkou vizualizaci lze shlédnout v tomto videu (anglicky): *https://youtu.be/QJYmyhnaaek*.

[24] Řecké písmenko π zde pro nás nemá nic společného s konstantou 3,14159 . . .

Důkaz. Rozložme přirozené číslo $N(\pi)$ na součin prvočísel $p_1 \cdots p_n$. Jelikož $N(\pi) = \pi \cdot \overline{\pi}$, platí $\pi \mid N(\pi) = p_1 \cdots p_n$. Předpokládáme ale, že π je prvočinitel, takže když dělí součin, musí dělit i některého z činitelů. Pro nějaké $j \in \{1, \ldots, n\}$ tedy platí $\pi \mid p_j$, jak jsme chtěli. □

V důsledku tohoto lemmatu je norma každého prvočinitele rovna p nebo p^2 pro nějaké prvočíslo p. Zároveň musí každé prvočíslo p mít nějakého prvočinitele, který je dělí, takže se nám stačí podívat na to, jaké prvočinitele které prvočíslo „vyrábí".

Rozmysleme si také, že každému prvočíslu $p \in \mathbb{Z}$ takto přísluší buďto prvočinitelé s normou p, anebo prvočinitelé s normou p^2 – ne obojí. Kdybychom měli prvočinitele π, ρ splňující $N(\pi) = p$, $N(\rho) = p^2$, plynulo by z toho

$$\pi \mid p \mid p^2 = \rho\overline{\rho}.$$

Z toho nutně $\pi \mid \rho$ nebo $\pi \mid \overline{\rho}$, BÚNO nechť $\pi \mid \rho$. Jenže ρ je (jakožto prvočinitel) ireducibilní, takže už musí π a ρ být asociované neboli $\rho = u\pi$ pro nějakou jednotku u. Z toho $N(\rho) = N(u)N(\pi) = 1 \cdot p$, což je spor.

Stačí nám tedy zkoumat jenom to, která prvočísla p lze vyjádřit jako normu, tzn. jako součet dvou čtverců. Zde nám pomůže podívat se na problém mod p.

Tvrzení. *Prvočíslo $p \in \mathbb{N}$ lze vyjádřit jako součet dvou čtverců, právě pokud má kongruence $x^2 + 1 \equiv 0 \pmod{p}$ řešení.*

Důkaz. Chceme dokázat ekvivalenci, dokažme tedy dva směry implikace. Nechť nejprve máme vyjádření $p = a^2 + b^2$. Kdyby $p \mid b$, pak už nutně musí platit i $p \mid a^2$, takže $p \mid a$, tudíž

$$p^2 \mid a^2 + b^2 = p,$$

což je spor. Když se tedy na $p = a^2 + b^2$ podíváme mod p, můžeme v kongruenci $a^2 + b^2 \equiv 0 \pmod{p}$ vydělit nenulovým b^2 a dostat $x^2 + 1 \equiv 0 \pmod{p}$, kde $x \equiv \frac{a}{b} \pmod{p}$.

Nechť naopak existuje celé číslo x takové, že $p \mid x^2 + 1$. Ukážeme, že p není v $\mathbb{Z}[i]$ prvočinitelem. Vskutku platí

$$p \mid (x + i)(x - i),$$

ale přitom zcela zřejmě $p \nmid x \pm i$, protože imaginární složka je ± 1, což není násobek žádného prvočísla. Víme tedy, že p není prvočinitel, nemůže tedy být ani ireducibilní. To znamená, že existují nějaká $\alpha, \beta \in \mathbb{Z}[i]$, z nichž ani jedno není jednotka, která splňují $\alpha\beta = p$. Obě normy $N(\alpha)$, $N(\beta)$ nyní dělí $N(p) = p^2$. Kdyby ale $N(\alpha)$ bylo p^2, značilo by to $N(\beta) = 1$, což nelze, neboť β není jednotka. Určitě tedy $N(\alpha) = N(\beta) = p$, čímž jsme vyjádřili p jako normu Gaussova celého čísla neboli jako součet dvou čtverců. □

Zbývá nám tedy zkoumat jen řešitelnost $x^2 \equiv -1 \pmod{p}$. To lze činit mnoha způsoby, my si zde ukážeme jeden kombinatorický argument.

Tvrzení. *Pro prvočíslo $p \in \mathbb{N}$ má kongruence $x^2 \equiv -1 \pmod{p}$ řešení, právě pokud $p = 2$ nebo $p \equiv 1 \pmod{4}$.*

Důkaz. Vypořádejme se nejprve s dvojkou jako se speciálním případem. Prvočíslo $p = 2$ dovedeme vyjádřit jako $2 = 1^2 + 1^2$a zároveň máme kongruenci $1^2 \equiv -1 \pmod{2}$, takže dokazované tvrzení platí.

Nadále tedy předpokládejme, že p je liché. Provedeme následující trik: všech $p - 1$ nenulových zbytkových tříd mod p rozdělíme do skupinek

$$\left\{x, -x, \frac{1}{x}, -\frac{1}{x}\right\}.$$

Pozor, nemusí se vždy jednat o čtveřičky, neboť pro vhodné x mohou některé z výrazů x, $-x$, $\frac{1}{x}$, $-\frac{1}{x}$ být rovny – na tom bude důkaz založen. Rozmysleme si nejprve, že takto korektně rozdělíme

zbytkové třídy do skupinek, tedy že $x \equiv a$ nám vytvoří stejnou skupinku jako $x \equiv -a$ atp. To lze ověřit buďto projitím všech možností, anebo si prostě stačí rozmyslet, jak spolu interagují úpravy $x \mapsto -x$ a $x \mapsto \frac{1}{x}$.

Vytvořili jsme tedy skupinky a každá zbytková třída leží jednoznačně v jedné z nich. Očekávali bychom, že „typicky" bude mít skupinka čtyři prvky. Podívejme se tedy na to, jak mohou vypadat skupinky s méně prvky. Protože nezáleží, od kterého prvku x skupinku vytváříme, budeme BÚNO zkoumat, kdy je x totožné s dalším výrazem. Jelikož je p liché, nemůže nám nastat $x \equiv -x$, neboť kongruenci $2x \equiv 0 \pmod{p}$ splňuje jenom $x \equiv 0$. Každá skupinka má tedy aspoň 2 prvky. Kdyby nastalo $x \equiv \frac{1}{x}$, máme z toho kongruenci

$$\begin{aligned} x^2 - 1 &\equiv 0 \pmod{p}, \\ (x-1)(x+1) &\equiv 0 \pmod{p}, \end{aligned}$$

což znamená buďto $x \equiv 1$, nebo $x \equiv -1$. Toto nám tedy nastává pouze ve dvouprvkové skupince $\{1, -1\}$. Naproti tomu $x \equiv -\frac{1}{x}$ by nám vedlo na kongruenci $x^2 \equiv -1 \pmod{p}$, což nás přesně zajímá. Zároveň bychom tím dostali dvouprvkovou skupinku, neboť potom i $-x \equiv \frac{1}{x}$. Také si povšimněme, že toto by nám dalo jenom jednu dvouprvkovou skupinku, protože když už nějaké c splňuje $c^2 \equiv -1$, pak pro libovolné $x^2 \equiv -1$ můžeme rozložit

$$\begin{aligned} x^2 + 1 &\equiv 0 \pmod{p}, \\ x^2 - c^2 &\equiv 0 \pmod{p}, \\ (x-c)(x+c) &\equiv 0 \pmod{p}, \end{aligned}$$

tedy $x \equiv \pm c \pmod{p}$.

Shrňme: $p-1$ zbytkových tříd se nám dělí do skupinek. Vždy máme dvouprvkovou skupinku $\{1, -1\}$, následně se *může* vyskytnout nanejvýš jedna další dvouprvková skupinka a všechny ostatní skupinky jsou čtyřprvkové. Pokud tedy tato druhá dvouprvková skupinka existuje, dává celkový počet prvků rozdělených do skupinek mod 4 zbytek $2 + 2 \equiv 0$ neboli $p \equiv 1 \pmod{4}$. Obdobně když tato druhá dvouprvková skupinka neexistuje, dává $p-1$ zbytek 2 mod 4. Nicméně tato druhá dvouprvková skupinka existuje právě tehdy, když má $x^2 \equiv -1 \pmod{p}$ řešení. Tím je důkaz hotov. □

S tímto tvrzením je naše lopota završena. Shrňme, jak to tedy je s prvočiniteli a s normami v $\mathbb{Z}[i]$.

Věta. *Prvočinitelé v $\mathbb{Z}[i]$ jsou dvou druhů:*

(i) *Prvky s prvočíselnou normou $p = 2$ nebo $p \equiv 1 \pmod{4}$, např. $1 + i$, $2 - i$, $-2 - 3i$ atp.*

(ii) *Prvočísla $p \equiv 3 \pmod{4}$ a jejich asociované prvky, např. 3, -7, $11i$ atp. Jejich norma je pak p^2.*

Jelikož norma obecného Gaussova celého čísla je jenom součin norem prvočinitelů, plyne z tohoto, že normy nabývají těch hodnot, které dostaneme násobením dvojky, prvočísel tvaru $4k+1$ a druhých mocnin prvočísel tvaru $4k+3$. Jinými slovy:

Důsledek. (čísla tvaru $x^2 + y^2$) *Ve tvaru $n = x^2 + y^2$ se dají vyjádřit právě ta přirozená čísla n, v jejichž prvočíselném rozkladu se všechna prvočísla tvaru $4k+3$ vyskytují v sudých mocninách.*

Obecně bychom si z tohoto měli odnést, že vyjadřování ve tvaru $n = x^2 + y^2$ je spjato s rozkladem n na prvočísla, resp. s rozkladem $x + yi$ na prvočinitele. Zkus si to vyzkoušet na následujících cvičeních.

Cvičení 30. Nechť je $p \equiv 3 \pmod{4}$ prvočíslo. Pro $a, b \in \mathbb{Z}$ pak platí, že když $p \mid a^2 + b^2$, pak i $p \mid a, b$.

Cvičení 31. Nechť je $p = a^2 + b^2 = c^2 + d^2$ prvočíslo. Pak už jsou dvojice (a, b) a (c, d) až na pořadí a změnu znamének stejné, tzn. $\{|a|, |b|\} = \{|c|, |d|\}$.

Cvičení 32. Nechť je $\pi \in \mathbb{Z}[i]$ prvočinitel. Pokud $N(\pi) = 2$, pak jsou π, $\overline{\pi}$ asociované, stejně tak pokud je $N(\pi) = p^2$ pro prvočíslo $p \equiv 3 \pmod 4$. Naopak pokud $N(\pi) = p \equiv 1 \pmod 4$, pak π, $\overline{\pi}$ nejsou asociované.

Úloha 8. Najdi všechna řešení rovnice $4xy - x - y = z^2$ v přirozených číslech.

Úloha 9. Najdi všechna celočíselná řešení rovnice $x^4 = 4 + y^2 + z^2$.

Úloha 10. Dokaž, že rovnice $3^n = x^2 + y^2 + 1$ má nekonečně mnoho řešení (n, x, y) v přirozených číslech.

Úloha 11. Najdi všechna celočíselná řešení rovnice $x^2 = y^3 + 7$.

Úloha 12. Najdi všechna celočíselná řešení rovnice $x^7 + 7 = y^2$.

Úloha 13. Najdi všechna celočíselná řešení rovnice $x^3 - x^2 + 8 = y^2$.

Závěr

První díl tímto došel svého konce. Víme již, proč funguje prvočíselný rozklad, že dobrou cestou k jednoznačnému rozkladu v komutativních okruzích je dělení se zbytkem i že vyzbrojeni Gaussovými celými čísly se nemusíme zaleknout součtů čtverců. Děkujeme Ti, že ses dočetl(a) až sem. V příštím díle si opět rozšíříme svůj katalog okruhů o pár nových kousků, podíváme se na zub jednotkám v podobě Pellovy rovnice a rozmyslíme si, jak modulit v roztodivných okruzích.

Do té doby na viděnou a hodně zdaru s úlohami první soutěžní série!

Návody ke cvičením

1. Vždy si vše rozepiš definicí dělitelnosti.

2. Použij vlastnost (ii) dělitelnosti prvně z a do c přes b a poté naopak.

3. Použij indukci.

4. Rozlož $(x - 4)(x + 4) = y^3$. Pro liché x použij tvrzení o nesoudělnosti a mocninách. Pro sudé x už musí x i y být násobek čtyř, vhodnou úpravou získáš rovnici, kterou jsme již viděli.

5. Rozepiš z definice kongruence. Neboj se sčítat/odečítat dělitelnosti. Všimni si také, že (i) a (ii) jsou jenom speciální případy (iii) a (iv).

6. (i) Rozepiš z definice kongruence. (ii) Vytkni největšího společného dělitele.

7. Zkus druhou mocninu 5, není to už nějaký hezký zbytek? Pokud Ti nepřijde dostatečně hezký, zkus čtvrtou mocninu.

8. Spočti $N \pmod 2$, $\pmod{10}$, $\pmod 7$.

10. Znovu zlomek rozšiř číslem sdruženým ke jmenovateli. Alternativně můžeš zkusit zapsat $12-6i$ jako $(9 + 3i) \cdot (a + bi)$ a vyřešit soustavu rovnic.

12. Využij definici dělitelnosti a multiplikativity normy.

13. No, co ta norma?

14. Pokud existují dvě různé nuly, co je jejich součtem?

15. Roznásob $(1+1)(a+b)$ dvěma různými způsoby.

16. Využij $0 = 0 + 0$.

17. Ukaž, že když $0 = 1$, pak už $0 = a$ pro každé $a \in R$.

25. Uvaž $p = ab$ a využij definici prvočinitele.

27. Nechť $\tilde{d}(b) = d(bc)$, pak využij podmínku (iii) na dvojici (a, bc).

28. Použij jednoznačný rozklad.

32. Vezmi dělitelnost $\overline{\pi} \mid \pi$ a rozšiř ji na $p \mid \pi^2$.

Návody k úlohám

1. Eukleidův algoritmus. Nelam si hlavu s tím, aby úpravy nutně snižovaly hodnotu pro každé n, prostě to „zmenšuj“ jako výraz.

2. Ukaž nejprve, že i $x + y - 1 \mid 2xy$.

3. Vypořádej se s možnými společnými děliteli jednotlivých činitelů – není jich mnoho.

4. Rozlož $p^2 + 5pq + 4q^2 = (p+q)(p+4q)$. Pro $p \neq q$ je NSD závorek buďto 1, nebo 3. Připrav se na rozebírání případů.

5. Rozlož v $\mathbb{Z}[i]$ a podívej se na možné společné dělitele činitelů. Pomůže Ti rozmyslet si, že x nemůže být liché – využij $x^2 \equiv 1 \pmod 4$.

6. Rozlož v $\mathbb{Z}[i]$. Nesoudělnost závorek ukaž s pomocí mod 4.

7. Dosaď parametrizaci pythagorejských trojic.

8. Uprav levou stranu na součin a koukni se, co dostaneš na pravé.

9. Převeď čtyřku, rozlož na součin a najdi zakázaného dělitele. Pokud je x sudé, vyděl nejprve celou rovnici čtyřmi.

10. Vol $n = 2^k$ a rozlož $3^{2^k} - 1$ na součin.

11. Přičti jedničku a najdi zakázaného dělitele.

12. Nahlédni, že x je liché, přičti 121 a najdi zakázaného dělitele.

13. Až vyloučíš $2 \mid x$, přičti x^2 a najdi zakázaného dělitele.

Řesení cvičení

1. (i) Platí $a = 1 \cdot a$, $0 = a \cdot 0$.

(ii) Máme $b = ka$, $c = \ell b$, takže $c = (k\ell)a$. Obdobně se dokáží (iii) až (v).

(vi) Pokud $a = 0$, pak už i $b = 0$, neboť jen nula je násobkem nuly. Dále nechť $a, b \neq 0$. Když $b = ka$, $a = \ell a$, dostaneme $a = k\ell a$, tedy $1 = k\ell$, takže $|k| = |\ell| = 1$, z čehož už $|a| = |b|$.

2. Stručně: máme $a \mid b$ a zároveň $b \mid c$, takže $a \mid c$. Analogicky $c \mid a$, takže $a \parallel c$.

3. (i) Indukujeme podle n. Pro $n = 2$ se jedná o definici ireducibility. Dále použijeme definici ireducibilního prvku na $q = (a_1 a_2 \cdots a_{n-1})a_n$. Nyní je jednou možností, že a_n je jednotka a $\pm(a_1 a_2 \cdots a_{n-1}) = q$, pak jsme hotovi z indukčního předpokladu. Pokud by naopak $(a_1 a_2 \cdots a_{n-1})$ byla jednotka, tak všechna a_i jsou pro $i \leq n-1$ jednotky.

(ii) Velmi podobně jako (i).

4. Jak už víme z návodu, rozebereme si dva případy. Pokud $d \mid x - 4$, $d \mid x + 4$, pak i $d \mid 8$. Pokud je x liché, jsou však oba výrazy také liché, takže můžeme použít tvrzení o nesoudělnosti a mocninách (jelikož víme, že ± 1 jsou třetí mocniny, můžeme je „schovat“ do mocniny). $x - 4 = a^3$, $x + 4 = b^3$ neboli $(b-a)(a^2 + ab + b^2) = b^3 - a^3 = 8$. Jelikož druhá závorka je vždy kladná (a^2 nebo b^2 je v absolutní hodnotě vetší roven ab). Tudíž stačí vyzkoušet pouze kladné dělitele $b - a$.

Z našich čtyř možností má pouze jediná celočíselné řešení. A to ta, když $(b-a)=2$ s řešením $(a,b)=(0,2),(-2,0)$. Jelikož by nám však z těchto a, b vyšlo sudé x, můžeme i tuto dvojici vyloučit.

Informaci z návodu lehce ověříme. Díky tomu, že můžeme čísla zapsat ve tvaru $x=4x'$ a $y=4y'$ získáváme $x'^2=4y'^3+1$ neboli $(x'-1)(x'+1)=4y'^3$. Jelikož je NSD těchto dvou závorek nanejvýš 2 a zároveň je pravá strana sudá, je to právě číslo 2. Potom už můžeme zase použít naše známé tvrzení o nesoudělnosti a získáme dvě rovnice. $x'-1=2a^3$ a $x'+1=2b^3$. Použijeme známý trik z jejich odečtením z čehož získáme $(b-a)(a^2+ab+b^2)=b^3-a^3=1$. Jak už víme z minulého případu, stačí nám rozebrat jen jedna možnost a to ta, kde jsou obě závorky na levé straně rovny 1. To nám dá dvojice $(a,b)=(0,1),(-1,0)$.

Celkově máme tedy dvojice původních $(x,y)=(4,0)$, $(-4,0)$.

5. (i) $n \mid a-b=a+k-b-k$. (ii) $n \mid a-b$, tedy i $n \mid k\cdot(a-b)$.

(iii) $n \mid (a-b)+(c-d)=(a+c)-(b+d)$. (iv) $n \mid (a-b)\cdot c-(c-d)\cdot b=ac-bd$.

6. (i) Máme tedy $n \mid c(a-b)$. Jelikož máme jednoznačný rozklad a n s c jsou nesoudělné, tak také platí, že $n \mid a-b$.

(ii) V obecné variantě si označme $g=(n,c)$, potom $n=gu$ a $c=gv$ a $n\cdot k=c\cdot(a-b)$. Po dosazení $guk=gv\cdot(a-b)$ neboli $uk=v(a-b)$, tedy i $\frac{n}{(n,c)} \mid a-b$.

7. Platí $5^2=25\equiv -1$, tedy $5^{20}=(5^2)^{10}\equiv(-1)^{10}\equiv 1 \pmod{26}$.

8. Jak už jsme si řekli, stačí si vybrat jen vhodné reprezentanty. První tedy N modulo 2. $N=22\cdot 31+11\cdot 17+13\cdot 19\equiv 0\cdot 1+1\cdot 1+1\cdot 1\equiv 0 \pmod 2$. Stejně tak to uděláme modulo 10. $22\cdot 31+11\cdot 17+13\cdot 19\equiv 2\cdot 1+1\cdot -3+3\cdot -1\equiv 6 \pmod{10}$. A poslední modulo, které musíme prozkoumat je modulo 7, tedy $22\cdot 31+11\cdot 17+13\cdot 19\equiv 3\cdot -3+4\cdot 3+-1\cdot -2\equiv 5 \pmod 7$.

9. Jednou možností je prostě to vyzkoušet pro $x=0,1,2,3$. Alternativně můžeme nahlédnout $(2n)^2=4\cdot n^2$, zatímco $(2n+1)^2=4\cdot(n^2+n)+1$.

10. $12-6i=(9+3i)(1-i)$, takže ano.

11. Přirozené číslo lze vyjádřit jako x^2+y^2, právě když je normou nějakého Gaussova čísla. Máme $m=N(\alpha)$, $n=N(\beta)$, takže multiplikativita normy nám snadno dá $mn=N(\alpha)N(\beta)=N(\alpha\beta)$.

12. Z definice máme $\alpha k=\beta$. Podívejme se, jak vypadají normy těchto čísel: $N(\alpha)N(k)=N(\alpha k)=N(\beta)$. Tím pádem $N(\alpha) \mid N(\beta)$. Dělitelnost norem však není postačující k dělitelnosti samotných čísel: kupříkladu $N(2+i) \mid N(3+i)$, ale $2+i \nmid 3+i$.

13. Stačí se nám tedy podívat na normu: $N(3-i)=10 \nmid 205=N(14-3i)$. Jelikož se nedělí normy, nemůžou se dělit ani odpovídající Gaussova čísla.

14. Nechť existují dva prvky 0, $0'$ splňující $0+a=a=0'+a$ pro každé a. Dosazením $a=0'$ máme $0+0'=0'$, zároveň ale dosazením $a=0$ máme $0'+0=0$. Dohromady tak $0'=0$.

Analogicky pokud máme dvě jedničky 1, $1'$, pak dostaneme $1'=1\cdot 1'=1'\cdot 1=1$.

15. Uvažme $(1+1)(a+b)$. Když roznásobíme nejprve pravou závorku, dostaneme

$$(1+1)(a+b)=(1+1)a+(1+1)b=a+a+b+b.$$

Když naopak roznásobíme nejprve levou závorku, dostaneme

$$(1+1)(a+b)=1(a+b)+1(a+b)=a+b+a+b.$$

Dohromady taky $a+a+b+b=a+b+a+b$ neboli $a+b=b+a$.

16. Využijeme $0=0+0$ a distributivity násobení. Dostaneme

$$0\cdot a=(0+0)a=0\cdot a+0\cdot a,$$

takže odečtením $0 \cdot a$ z obou stran dostaneme $0 = 0 \cdot a$.

17. Pro každé $a \in R$ je $0 \cdot a = 0$, zatímco $1 \cdot a = a$. Kdyby tedy platilo $0 = 1$, pak dohromady $0 = 0 \cdot a = 1 \cdot a = a$. Takže každý prvek R je nula neboli R má jen jeden prvek.

19. (i) Všechny tyto okruhy jsou podmnožinami $\mathbb{C}$, takže stačí ukázat, že $\mathbb{C}$ je oborem integrity. V $\mathbb{C}$ umíme dělit nenulovými prvky, takže kdyby pro $a, b \in \mathbb{C}$ platilo $ab = 0$ a zároveň $a, b \neq 0$, dostaneme $a = \frac{ab}{b} = \frac{0}{b} = 0$, což je spor. Z toho je $\mathbb{C}$ obor integrity, takže i $\mathbb{Z}$, $\mathbb{Q}$ a $\mathbb{R}$ jsou obory integrity.

(ii) Stejně jako v předchozím odstavci jsou $\mathbb{Z}[i]$, $\mathbb{Z}[\omega]$ i $\mathbb{Z}\left[\frac{1}{2}\right]$ podmnožinami $\mathbb{C}$ (a jejich sčítání a násobení je jen zúžení sčítání a násobení v $\mathbb{C}$), takže už musí být obory integrity.

(iii) Triviální okruh je oborem integrity. Definice oboru integrity hovoří o tom, že pro $ab = 0$ musí být $a = 0$ nebo $b = 0$. V triviálním okruhu je ale každý prvek roven 0, takže závěr dokazovaného tvrzení platí automaticky.

(iv) Když $n = 1$, pak máme triviální okruh $\mathbb{Z}_1$, tedy obor integrity. Když je n složené číslo, můžeme ho zapsat jako $n = ab$ pro nějaká $a, b > 1$, načež $ab \equiv n \pmod{n}$, ale $a, b \not\equiv 0 \pmod{n}$, takže $\mathbb{Z}_n$ není obor integrity. Konečně když $n = p$ je prvočíslo, tak $ab \equiv 0 \pmod{p}$ znamená $p \mid ab$, takže $p \mid a$ nebo $p \mid b$, což znamená $a \equiv 0$ nebo $b \equiv 0 \pmod{p}$, takže $\mathbb{Z}_p$ je obor integrity. V souhrnu je tedy $\mathbb{Z}_n$ obor integrity, právě když n není složené.

(v) Pro $|X| = 0$ má $\mathcal{P}(X)$ jediný prvek $\emptyset$, takže je to triviální okruh. Pro $|X| = 1$ máme $\mathcal{P}(X) = \{\emptyset, X\}$. Pro $A, B \in \mathcal{P}(X)$ obě různé od $\emptyset$ tedy nutně $A \cap B = X \cap X = X \neq \emptyset$, takže $\mathcal{P}(X)$ je obor integrity. Pro $|X| \geq 2$ zvolme dva různé prvky $a, b \in X$. Obě množiny $\{a\}, \{b\}$ jsou neprázdné, ale přitom $\{a\} \cap \{b\} = \emptyset$, takže $\mathcal{P}(X)$ není obor integrity. V souhrnu je tedy $\mathcal{P}(X)$ obor integrity, právě pokud $|X| \leq 1$.

(vi) Když se podíváme na tabulku násobení v okruhu T, ověříme projitím všech jejích políček, že jediné součiny, které dávají výsledek 0, jsou ty, v nichž je 0 jedním z činitelů. Tedy T je obor integrity.

20. Z definice dělitelnosti $a \mid b$ znamená, že $b = ka$ pro nějaké k. Obdobně $c = \ell a$. Z toho už $bd + cd = (kd + \ell e)a$ neboli $a \mid bd + cd$.

21. Když $B \subseteq A$, tak vskutku $B = A \cap B$. Když $B = A \cap C$ pro nějakou C, tak z definice průniku musí být $B \subseteq A$.

22. (i) $\{1, -1\}$.

(ii) V těchto okruzích lze dělit každým nenulovým prvkem, takže $\mathbb{Q} \setminus \{0\}$, $\mathbb{R} \setminus \{0\}$, $\mathbb{C} \setminus \{0\}$.

(iii) Jsou to přesně prvky s normou 1 (v $\mathbb{Z}[\omega]$ definujeme normu pro $\alpha = a + b\omega$ jako $N(\alpha) = \alpha\overline{\alpha} = a^2 + ab + b^2$). Tedy jednotkami jsou $\{\pm 1, \pm i\}$, resp. $\{\pm 1, \pm\omega, \pm\omega^2\}$.

(iv) Pokud je čitatel násobkem nějakého lichého prvočísla, pak se ho násobením „nezbavíme“ a nevyrobíme tak jedničku. Zbývají tedy jen plus nebo mínus mocniny dvojky (včetně těch se zápornými exponenty). Ty ale jednotkami jsou, neboť $\pm 2^a \cdot (\pm 2^{-a}) = 1$. Množinou všech jednotek v $\mathbb{Z}\left[\frac{1}{2}\right]$ je tedy $\{\pm 2^a : a \in \mathbb{Z}\}$.

(v) Jedinou jednotkou je X. Podle dřívějšího cvičení již víme, že aby nějaká podmnožina $A \subseteq X$ dělila X, musí být její nadmnožinou, což znamená $A = X$.

23. Víme již, že když R není triviální, pak $0 \neq 1$. Potom kdyby $0 \mid 1$, pak pro nějaké $a \in R$ platí $1 = 0 \cdot a = 0$, což je spor.

24. Máme dokázat ekvivalenci, ukážeme tedy dvě implikace. Nechť nejprve $a = ub$. Potom zároveň $a = bv$, kde u, v jsou jednotky splňující $uv = 1$. To znamená $a \mid b$, $b \mid a$. Nyní nechť naopak $a \parallel b$. To znamená, že pro nějaká u, v platí $au = b$ a zároveň $a = bv$. Dohromady tedy $a = bv = auv$, takže $1 = uv$ neboli je u jednotka, jak jsme chtěli.

25. Budiž p prvočinitel a nechť $p = ab$. Potom $p \mid ab$, takže z definice prvočinitele máme $p \mid a$ nebo $p \mid b$. BÚNO předpokládejme $p \mid a$. Zároveň ale $a \mid p$, takže a, p jsou asociované, tedy $p = au$ pro nějakou jednotku u, takže už $au = ab$ neboli $u = b$, a b je tak jednotka.

26. Z definice je největší společný dělitel společný dělitel a zároveň každý společný dělitel dělí největšího společného dělitele. To znamená $g_1 \mid g_2$ a zároveň $g_2 \mid g_1$, takže g_1 a g_2 jsou asociované.

27. Je zřejmé, že $\tilde{d}$ splňuje (i). Abychom nahlédli (ii), všimněme si, že $\tilde{d}(a)$ je minimum z d na množině nenulových násobků a. Ale každý násobek ab je i násobek a, takže $\tilde{d}(ab)$ je jenom minimum z nějaké podmnožiny nenulových násobků a. Minimum podmnožiny je větší nebo rovno minimu celé množiny, takže $\tilde{d}(a) \leq \tilde{d}(ab)$.

Zaměřme se nyní na (iii). Jsou dána $a, b \in R$. Z definice $\tilde{d}$ existuje $c \in R$ takové, že $\tilde{d}(b) = d(bc)$. Nyní s funkcí d využijeme vlastnost (iii) na (a, bc). To nám dá $a = b(cq) + r$ pro $d(r) < d(bc)$. Přitom ale z definice $\tilde{d}(r) \leq d(r \cdot 1)$, takže $\tilde{d}(r) \leq d(r) < d(bc) = \tilde{d}(b)$.

28. (i) Ireducibilní prvek p dělí nenulový prvek c právě tehdy, když se p nachází (až na asociovanost) v rozkladu c na součin ireducibilních prvků. Když tedy p nedělí ani a, ani b, pak se nevyskytuje ani v jednom z jejich rozkladů, takže se nevyskytuje ani v rozkladu ab. To už značí, že když $p \mid ab$, pak $p \mid a$ nebo $p \mid b$.

(ii) Zapišme zkrášlené rozklady prvků a, b, přičemž použijme pro každou „rodinku“ asociovaných ireducibilních prvků v obou rozkladech jen jednoho zástupce. Můžeme tedy napsat

$$a = u \cdot p_1^{\alpha_1} \cdots p_n^{\alpha_n}, \qquad b = v \cdot p_1^{\beta_1} \cdots p_n^{\beta_n},$$

kde u, v jsou jednotky, $p_1, \ldots, p_n$ jsou navzájem neasociované ireducibilní prvky a exponenty $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ jsou nezáporná celá čísla – může se jednat o nuly, pokud se příslušný ireducibilní prvek vůbec nenachází v rozkladu. Nyní už je zjevné, že

$$g = p_1^{\min\{\alpha_1, \beta_1\}} \cdots p_n^{\min\{\alpha_n, \beta_n\}}$$

je největším společným dělitelem a, b. Tento prvek totiž určitě dělí obě a, b (má menší nebo rovné exponenty) a zároveň každý společný dělitel dělí g (nemůže mít větší žádný exponent).

29. Analogicky s důkazem eukleidovskosti $\mathbb{Z}[\sqrt{-2}]$ bychom se dostali na ekvivalentní formulaci: pro daná $x, y \in \mathbb{Q}$ najít $x_0, y_0 \in \mathbb{Z}$ takové, že $(x - x_0)^2 + 3(y - y_0)^2 < 1$. Pro $x = y = \frac{1}{2}$ však taková x_0, y_0 neexistují.

30. Platí $p \mid (a + bi)(a - bi)$ a p je prvočinitel, takže už dělí jedno z $a \pm bi$, tudíž dělí obě a, b.

31. Nechť $\pi = a + bi$, $\rho = c + di$. Jelikož je jejich norma prvočíslo, musí to být prvočinitelé. Máme $\pi\overline{\pi} = \rho\overline{\rho}$. Levá strana je násobkem π, takže i $\pi \mid \rho\overline{\rho}$. To znamená, že buďto $\pi \mid \rho$, načež $\pi \parallel \rho$, nebo $\pi \mid \overline{\rho}$, načež $\pi \parallel \overline{\rho}$. Jenže násobení jednotkou jenom prohazuje reálnou a imaginární složku a mění jejich znaménka, zatímco komplexní sdružení jenom mění znaménko imaginární složky. To znamená, že složky π, ρ jsou tak stejné až na prohození a změny znamének.

32. Pro $\pi \parallel p \equiv 3 \pmod 4$ je tvrzení zřejmé. Dále jsou prvočinitelé normy 2 jenom $1 + i$ a asociované prvky. Přitom $1 + i = (1 - i) \cdot i$, takže tyto dva komplexně sdružené prvky jsou asociované.

Mějme nyní $\pi = a + bi$ a $N(\pi) = p \equiv 1 \pmod 4$. Kdyby $\pi \parallel \overline{\pi}$, pak $\overline{\pi} \mid \pi$, z čehož rozšířením nutně i

$$p = \pi\overline{\pi} \mid \pi^2 = (a^2 - b^2) + 2abi.$$

Jenže a, b musí být nesoudělná, takže a i b je nesoudělné s $a^2 + b^2 = p$. Z lichosti p je i 2 nesoudělné s p, takže dohromady je $2ab$ nesoudělné s p. To znamená, že určitě $p \nmid 2ab$, takže ani $p \nmid \pi^2$. Prvky π, $\overline{\pi}$ tak nejsou asociované.

Teorie nejen čísel 2 – Jednotky

Milý příteli,
vítej u druhé části našeho povídání o tom, že teorii čísel lze potkat i mimo obor celých čísel. V prvním díle jsme se s cílem rozkládat věci na součiny ponořili do světa komutativních okruhů – struktur, kde dovedeme sčítat a násobit podle obvyklých pravidel. Naučili jsme se rozeznávat některé speciální prvky okruhů – ireducibilní prvky, prvočinitele či jednotky. Přidáváním koření v podobě dalších užitečných vlastností jsme pak dostávali různé příchutě okruhů: Pokud jsme v rovnicích mohli krátit, jednalo se o obor integrity. Dělení se zbytkem nám dávalo eukleidovský obor, zatímco jednoznačný rozklad na prvočinitele značil gaussovský obor. Ukázali jsme si, že dělení se zbytkem už zaručuje jednoznačný rozklad (ale ne naopak!) a že jednoznačný rozklad se může hodit k řešení diofantických rovnic. Získané dovednosti jsme si poté vyzkoušeli na zkoumání součtů dvou čtverců skrze vlastnosti okruhu $\mathbb{Z}[i]$.

V tomto díle si katalog okruhů rozšíříme o dvě významné rodinky: *reálné kvadratické okruhy* a *konečná tělesa*. Ač jsou tyto dvě skupiny na první pohled zcela odlišné, společná jim je jedna vlastnost: lze říci mnoho zajímavého o jejich jednotkách, tedy prvcích, které dělí jedničku, a v důsledku toho jimi lze dělit. Ukážeme si, že u obou dovedeme v jistém smyslu všechny jednotky vyrobit z jedné „základní“, a následně toho využijeme při řešení úloh. Stejně jako minule uvidíme, že tyto nové okruhy nám pomohou řešit úlohy zadané čistě v řeči celých čísel, nad kterými bychom bez nich jen tápali.

Jak seriál číst

Obdobně jako v prvním díle na Tebe opět čeká spousta cvičení a úloh. K obojímu nalezneš na konci seriálu návody a ke cvičením též řešení. Opět také některá cvičení odlišujeme co do významu: cvičení s vykřičníkem jsou opravdu důležitá a často je později využíváme, zatímco cvičení s hvězdičkou jsou různá rozšiřující zamyšlení a zajímavosti, které k dalšímu pochopení seriálu nejsou nutné.

V tomto dílu jsme hvězdičkou označili i dvě celé kapitolky – o existenci netriviální jednotky v reálných kvadratických okruzích a o charakteristice v konečných tělesech. Ty mohou být i v porovnání se zbytkem seriálu náročnější, takže pokud se na nich zasekneš, neboj se je přeskočit a vrátit se k nim později. Ani jedna není nutná k vyřešení soutěžních úloh, ani k chápání zbytku seriálu. Přesto si však myslíme, že dávají dobrou představu o příslušných rodinkách okruhů a stojí za to je (zkusit) přečíst.

Reálné kvadratické okruhy

Zavzpomínejme krátce na první díl. V něm jsme všechny vybudované pojmy a nástroje, jako prvočinitele, ireducibilní prvky, jednotky, eukleidovské a gaussovské okruhy, použili na spoustu okruhů, z nichž většina měla jeden společný rys – vznikly tak, že jsme k celým číslům přidali nějaké komplexní číslo. Přesněji, přidávali jsme k $\mathbb{Z}$ nějakou odmocninu ze *záporného* čísla (třeba i nebo $\sqrt{-2}$), anebo o chlup obecněji komplexní číslo splňující nějakou kvadratickou rovnici (například $\alpha = \frac{1+\sqrt{-7}}{2}$ splňující $\alpha^2 - \alpha + 2 = 0$). To nás přivádí k prvnímu tématu druhého dílu. Co kdy-

bychom k celým (anebo racionálním) číslům přidávali opět čísla splňující nějakou kvadratickou rovnici – co třeba nějaké odmocniny z *kladných* čísel? Ukážeme si, že takto vzniklé okruhy se od $\mathbb{Z}[i]$ a komplexních okruhů z prvního dílu v některých ohledech o mnoho neliší, zatímco v jiných jsou úplně jiné.

Definice. Nechť je d přirozené číslo, které není čtverec[1]. Potom definujeme okruhy

$$\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} : a, b \in \mathbb{Z}\}, \qquad \mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d} : a, b \in \mathbb{Q}\}.$$

Jejich prvkům pak budeme říkat *(reálná) kvadratická čísla.*

Poznámka. Jak jsme zmínili v prvním díle, zápisem $R[m]$ obecně značíme, že k okruhu R přidáme nějaké m a vezmeme si všechno, co se z m a prvků R dá vyrobit (klidně opakovaným) sčítáním a násobením. Množina $\{a+b\sqrt{d} : a, b \in \mathbb{Q}\}$ by tedy měla být zapsána spíše $\mathbb{Q}[\sqrt{d}]$. Zápis $R(m)$ obecně značí, že bereme vše, co se z m a prvků R dá vyrobit sčítáním, násobením a *dělením.* V tuto chvíli by se mohlo zdát, že okruh $\mathbb{Q}(\sqrt{d})$ bude obsahovat něco navíc oproti $\mathbb{Q}[\sqrt{d}]$, ale jak zanedlouho ukážeme, v $\mathbb{Q}[\sqrt{d}]$ dovedeme dělit, takže $\mathbb{Q}(\sqrt{d}) = \mathbb{Q}[\sqrt{d}]$. Pokud bychom tak všude psali $\mathbb{Q}[\sqrt{d}]$ místo $\mathbb{Q}(\sqrt{d})$, nic by se nezměnilo. Většinou se však používá $\mathbb{Q}(\sqrt{d})$, čehož se budeme držet i my.

Již v definici jsme prohlásili, že se jedná o okruhy, slušelo by se tedy ověřit, že jsou tyto množiny skutečně uzavřené na sčítání a násobení. Uzavřenost na sčítání je zřejmá – prostě sečteme odpovídající složky. Při násobení využijeme $(\sqrt{d})^2 = d$, což je prvek $\mathbb{Z}$ i $\mathbb{Q}$. Pro celá čísla a_1, b_1, a_2, b_2 máme

$$(a_1 + b_1\sqrt{d})(a_2 + b_2\sqrt{d}) = a_1b_1 + a_1b_2\sqrt{d} + a_2b_1\sqrt{d} + a_2b_2d = (a_1b_1 + da_2b_2) + (a_1b_2 + a_2b_1)\sqrt{d},$$

což je zase prvek $\mathbb{Z}[\sqrt{d}]$. Pro $\mathbb{Q}(\sqrt{d})$ bychom postupovali úplně stejně. Díky tomu, že $\mathbb{Q}(\sqrt{d})$ i $\mathbb{Z}[\sqrt{d}]$ jsou podmnožiny $\mathbb{R}$, se musí jednat o obory integrity: součin nenulových prvků v těchto okruzích nemůže být nula, a v rovnicích tak můžeme krátit.

Proč je v definici podmínka, aby d nebylo čtvercem? Kdyby $d = a^2$, pak by $\sqrt{d} = a$ bylo celé číslo, takže bychom jeho přidáním k $\mathbb{Z}$, resp. ke $\mathbb{Q}$ nezískali nic navíc. Naopak když d není čtverec, pak už je zaručeno, že přidáním $\sqrt{d}$ skutečně něco navíc dostaneme.

Tvrzení. *Pokud přirozené číslo d není čtverec v $\mathbb{Z}$, pak $\sqrt{d} \notin \mathbb{Q}$.*

Důkaz. Pro spor nechť $\sqrt{d} \in \mathbb{Q}$. Potom je $\sqrt{d}$ rovno nějakému zlomku $\frac{a}{b}$ v základním tvaru (tedy a, b jsou nesoudělná). Umocněním na druhou pak

$$d = \frac{a^2}{b^2}.$$

To znamená, že $b^2 \mid a^2$, jelikož d je celé číslo. Jenže a^2 je nesoudělné s b^2, takže v dělitelnosti nehraje roli, z čehož $b^2 \mid 1$. To však znamená $b^2 = 1$, takže v rovnosti výše zbude $d = a^2$. Jinými slovy, d musel být čtverec, což je spor. □

Cvičení(*) 1. Nahlédni, pro jaká přirozená čísla d_1, d_2, jež nejsou čtverce, mohou být množiny $\mathbb{Q}(\sqrt{d_1})$ a $\mathbb{Q}(\sqrt{d_2})$ totožné.

Reálná kvadratická čísla leží někde na reálné přímce, nicméně často se zase hodí představovat si je spíše „dvojrozměrně“ jako dvojice (a, b) celých, resp. racionálních čísel odpovídajících prvku $a + b\sqrt{d}$. Taková představa by měla zásadní vadu na kráse, kdyby dvěma takovýmto dvojicím mohlo odpovídat stejné reálné číslo. Díky tomu, že a, b bereme pouze z množiny $\mathbb{Q}$ (anebo z její podmnožiny $\mathbb{Z}$), nemůže něco takového nastat.

[1] Připomeňme, že slovem „čtverec“ míníme druhou mocninu nějakého prvku z daného oboru. Není-li řečeno jinak, míníme čtverec v $\mathbb{Z}$.

Tvrzení. *Nechť jsou a_1, b_1, a_2, b_2 racionální čísla. Potom z $a_1+b_1\sqrt{d} = a_2+b_2\sqrt{d}$ plyne $a_1 = a_2$ a $b_1 = b_2$.*

Důkaz. Rovnost upravme na $(a_1 - a_2) = (b_2 - b_1)\sqrt{d}$. Kdyby nyní $b_2 - b_1$ bylo nenulové číslo, pak bychom jím mohli vydělit a dostat $\sqrt{d} = \frac{a_1-a_2}{b_2-b_1}$, což by znamenalo, že $\sqrt{d}$ je racionální. To by byl spor s předchozím tvrzením, takže určitě $b_2 - b_1 = 0$. Z toho pak i $a_1 - a_2 = 0$, takže musí být (a_1, b_1) a (a_2, b_2) totožné dvojice. □

Cvičení 2. Průnikem množin $\mathbb{Q}$ a $\mathbb{Z}[\sqrt{d}]$ je $\mathbb{Z}$.

I když tedy reálné kvadratické číslo leží na reálné přímce, schovává se v něm informace o dvojici racionálních čísel. S jejich pomocí nyní podobně, jako tomu bylo v případě komplexních čísel, zavedeme sdružená čísla a normu.

Definice. K číslu $\alpha = a+b\sqrt{d} \in \mathbb{Q}(\sqrt{d})$ definujeme jeho *sdružené* číslo $\overline{\alpha} = a-b\sqrt{d}$ a jeho *normu* $N(\alpha) = \alpha\overline{\alpha} = a^2 - db^2$.

Stejně jako v $\mathbb{Z}[i]$ a dalších komplexních okruzích se i zde sdružená čísla chovají hezky ke sčítání a násobení, v důsledku čehož je norma multiplikativní. Rozdílem je však to, že norma může být záporná – například $N(1 + \sqrt{2}) = -1$. Stále také platí, že norma celého (případně racionálního) čísla je prostě jeho čtverec.

Cvičení(!) 3. Nechť $\alpha, \beta \in \mathbb{Q}(\sqrt{d})$. Potom platí

$$\overline{(\alpha + \beta)} = \overline{\alpha} + \overline{\beta}, \qquad \overline{(\alpha\beta)} = \overline{\alpha} \cdot \overline{\beta} \qquad \text{a} \qquad N(\alpha\beta) = N(\alpha) \cdot N(\beta).$$

Cvičení(!) 4. Jediný prvek $\mathbb{Q}(\sqrt{d})$, který má normu 0, je sama 0.

Cvičení(!) 5. V $\mathbb{Q}[\sqrt{d}]$ lze dělit. Přesněji, pro $\alpha, \beta \in \mathbb{Q}[\sqrt{d}]$, $\beta \neq 0$ platí i $\frac{\alpha}{\beta} \in \mathbb{Q}[\sqrt{d}]$.

Později v seriálu uvidíme, že ačkoliv je většinou lepší zacházet s reálným kvadratickým číslem jako s dvojicí, někdy se přece jenom hodí, že je to jedno reálné číslo. Abychom mezi těmito pohledy uměli jednoduše přecházet, vyplatí se umět získat z $\alpha \in \mathbb{Q}(\sqrt{d})$ jeho racionální a iracionální složku. To však s pomocí sdružených čísel není příliš těžké.

Tvrzení. (vzorec pro (i)racionální složku) *Nechť $\alpha = a + b\sqrt{d} \in \mathbb{Q}(\sqrt{d})$. Potom platí*

$$a = \frac{\alpha + \overline{\alpha}}{2}, \qquad b = \frac{\alpha - \overline{\alpha}}{2\sqrt{d}}.$$

Důkaz. Stačí dosadit podle definice. □

Cvičení 6. Kvadratické číslo α je racionální, právě pokud $\alpha = \overline{\alpha}$.

Tímto jsme se v $\mathbb{Z}[\sqrt{d}]$ a $\mathbb{Q}(\sqrt{d})$ náležitě zabydleli a máme k dispozici většinu vymožeností, na něž jsme zvyklí ze $\mathbb{Z}[i]$. Mohli bychom si tak položit otázku, jak je to v reálných kvadratických okruzích s prvočiniteli, ireducibilními prvky a jednotkami. Jednotky jsou zajímavou kapitolou samy o sobě, proto je ještě na moment odložíme. Říci něco o tom, zda funguje rozklad na prvočinitele, je obecně těžké, ale v některých konkrétních případech lze vymyslet, jak v $\mathbb{Z}[\sqrt{d}]$ dělit se zbytkem, což nám jednoznačný rozklad zaručí. Opět tak budeme gaussovskost oboru (jednoznačný rozklad) dokazovat skrze eukleidovskost (dělení se zbytkem). Podobně jako v $\mathbb{Z}[i]$ a komplexních okruzích bývalo dobrým nápadem zkusit jako eukleidovskou funkci normu, tak i zde se norma hodí. Jelikož ale norma může být záporná, typicky ji pro potřeby dělení se zbytkem budeme dávat do absolutní hodnoty.

Příklad. $\mathbb{Z}[\sqrt{2}]$ je eukleidovský obor.

Řešení. Dokážeme, že $|N(x)|$ vyhovuje jako eukleidovská funkce. Na to potřebujeme ukázat, že kdykoliv jsou dána $\alpha, \beta \in \mathbb{Z}[\sqrt{2}]$, $\beta \neq 0$, pak dovedeme nalézt takové $\gamma \in \mathbb{Z}[\sqrt{2}]$, že $|N(\alpha - \gamma\beta)| <$

$|N(\beta)|$. Díky tomu, že se norma (stejně jako absolutní hodnota) chová hezky k násobení, můžeme tuhle nerovnost vydělit $|N(\beta)|$ a ekvivalentně upravit na

$$\left|N\left(\frac{\alpha}{\beta}-\gamma\right)\right|<1.$$

Chceme tedy, aby γ byla „blízko“ $\frac{\alpha}{\beta}$, tak ji podobně, jako když jsme dokazovali eukleidovskost $\mathbb{Z}[i]$, položíme jako zaokrouhlení $\frac{\alpha}{\beta}$. Vydělme tedy $\frac{\alpha}{\beta}=x+y\sqrt{2}\in\mathbb{Q}(\sqrt{2})$ a zvolme x_0 a y_0 jako ta celá čísla, která jsou nejblíž k (racionálním) číslům x a y. To zaručí, že $|x-x_0|$ i $|y-y_0|$ jsou nanejvýš $\frac{1}{2}$, takže když zvolíme $\gamma=x_0+y_0\sqrt{2}$, pak už máme

$$\left|N\left(\frac{\alpha}{\beta}-\gamma\right)\right|=\left|N\Big((x-x_0)+(y-y_0)\sqrt{2}\Big)\right|=\left|(x-x_0)^2-2(y-y_0)^2\right|\leq$$
$$\leq|x-x_0|^2+2|y-y_0|^2\leq\frac{1}{4}+2\cdot\frac{1}{4}=\frac{3}{4}<1.$$

Během předchozí úpravy jsme použili tzv. *trojúhelníkovou nerovnost.* Když vyšetřujeme výraz tvaru $|A+B|$, tak mohou A, B mít opačná znaménka a navzájem se vyrušovat, anebo mohou mít stejné znaménko a táhnout za jeden provaz. Když tedy dáme do absolutních hodnot A a B každé zvlášť, tak jenom zaručíme, že táhnou za jeden provaz, takže celkovou hodnotu výrazu možná o něco zvýšíme, ale určitě nesnížíme. V našem případě bylo $A=(x-x_0)^2$, $B=-2(y-y_0)^2$, takže jsme absolutní hodnotu rozdílu odhadli součtem.

V souhrnu jsme tak skutečně ukázali, že dovedeme zvolit γ tak, aby dělení se zbytkem fungovalo, takže $\mathbb{Z}[\sqrt{2}]$ je eukleidovský obor.

Cvičení 7. Dokaž, že $\mathbb{Z}[\sqrt{3}]$ je eukleidovský obor.

Úloha 1. Dokaž, že $\mathbb{Z}[\sqrt{7}]$ je eukleidovský obor.

Cvičení(*) 8. Nechť $d\equiv 1 \pmod 4$ není čtverec. Rozmysli si, že $\left\{a+b\cdot\frac{1+\sqrt{d}}{2}:a,b\in\mathbb{Z}\right\}$ je množina uzavřená na násobení, a tvoří tak okruh.

Pellova[2] rovnice aneb jednotky v $\mathbb{Z}[\sqrt{d}]$

Při našem průzkumu okruhů $\mathbb{Z}[\sqrt{d}]$ jsme dosud nezmínili, jak v nich vypadají jednotky. V nich se totiž reálná kvadratická čísla výrazně liší od těch komplexních. Zatímco v Gaussových celých číslech $\mathbb{Z}[i]$ jsme měli čtyři jednotky 1, i, -1 a $-i$ a v Eisensteinových číslech $\mathbb{Z}[\omega]$ šest jednotek ±1, $\pm\omega$, $\pm\omega^2$, v okruhu $\mathbb{Z}[\sqrt{d}]$ budeme mít jednotek nekonečně mnoho.

Nejprve nahlédněme souvislost jednotek s normou.

Tvrzení. *Číslo $\omega\in\mathbb{Z}[\sqrt{d}]$ je jednotka právě tehdy, když $|N(\omega)|=1$.*

Důkaz. Nechť je nejprve ω jednotka. To znamená, že existuje nějaké $\omega'\in\mathbb{Z}[\sqrt{d}]$ takové, že $\omega\omega'=1$. Vzetím norem pak $N(\omega)\cdot N(\omega')=1$. Máme tedy dvě celá čísla, jejichž součin je 1, takže jsou to buď dvě jedničky, nebo dvě mínus jedničky. V obou případech však $|N(\omega)|=1$.

Naopak nechť je $N(\omega)=\pm1$ neboli $\omega\cdot\overline{\omega}=\pm1$. To znamená, že ω dělí ±1, což nehledě na znaménko dělí 1. Celkově tedy $\omega\mid 1$, takže ω je jednotka. □

Z tohoto plyne, že mocněním jednotky dostaneme zase jednotku. To není nic nového, když v $\mathbb{Z}[i]$ budeme brát mocniny i, dostaneme postupně všechny čtyři jednotky v tomto okruhu. V reálných kvadratických číslech však narazíme na to, že se nám toto mocnění nemusí zacyklit. Jednotky

[2] John Pell (1611–1685), anglický matematik, přišel k „vlastnictví“ Pellovy rovnice spíše omylem: její vyřešení mu chybně připsal Leonhard Euler a stejně jako spousta jiných mylných označení se i toto ujalo. Ve skutečnosti bylo řešení Pellovy rovnice nalezeno už o staletí dříve v Indii.

v $\mathbb{Z}[\sqrt{d}]$ jsou totiž reálná čísla, a když mocníme reálné číslo ω, tak budeme buďto dostávat větší a větší (pokud $|\omega| > 1$), anebo menší a menší (pokud $|\omega| < 1$) hodnoty. To ale znamená, že to budou navzájem různá čísla. Nekonečně mnoho jednotek bychom tedy takovýmto mocněním nevyrobili jen tehdy, kdybychom začali s něčím, co má absolutní hodnotu 0 (to jednotka nebude), anebo 1 (to mohou být 1 a -1, což jsou jednotky).

V souhrnu tedy: když najdeme jednotku, která není ± 1, pak už z ní mocněním vyrobíme nekonečně mnoho jednotek. Například v $\mathbb{Z}[\sqrt{2}]$ máme $N(1+\sqrt{2}) = -1$, takže je to jednotka, pročež jsou jednotkami i její mocniny

$$(1+\sqrt{2})^2 = 3+2\sqrt{2}, \qquad (1+\sqrt{2})^3 = 7+5\sqrt{2}, \qquad (1+\sqrt{2})^4 = 17+12\sqrt{2} \qquad \text{a tak dále.}$$

Nekonečně mnoho jednotek by nám mohlo činit obtíže, například kdybychom chtěli řešit nějakou rovnici tvaru $ab = c^n$. Když víme, že a, b jsou nesoudělná, tak v gaussovkém oboru chceme vyvodit, že i a, b musí samy být n-té mocniny. Jenže tohle platí s jedním háčkem – jsou to n-té mocniny *přenásobené jednotkami*. Bez znalosti jednotek bychom tedy v $\mathbb{Z}[\sqrt{d}]$ nemohli ani použít nástroje z prvního dílu.

Pojďme tedy hledat jednotky. Aby $x + y\sqrt{d}$ bylo jednotkou, má jeho norma být buďto 1, nebo -1. První z těchto možností odpovídá rovnici

$$x^2 - dy^2 = 1,$$

která si svým významem vysloužila vlastní jméno. Říká se jí *Pellova rovnice* a její řešení odpovídají jednotkám v $\mathbb{Z}[\sqrt{d}]$ s normou 1 (ještě by mohly existovat jednotky s normou -1). Takto přeformulovaný problém hledání jednotek už můžeme potkat často potkat v úlohách: mnohé diofantické rovnice se dají vhodnou úpravou dostat do tvaru $x^2 - dy^2 = 1$ a v tu chvíli na ně můžeme vrhnout mašinérii okruhu $\mathbb{Z}[\sqrt{d}]$ a jeho jednotek, kterou si postupně rozkryjeme.

Dvě řešení Pellovy rovnice dovedeme zaručit vždy: jsou jimi $y = 0$ a $x = 1$, resp. $x = -1$. To odpovídá tomu, že 1 a -1 mají normu jedna, a jsou tak jednotkami, což příliš nepřekvapí. Protože jsou tato řešení celkem nezajímavá, nazýváme je *triviální*. Obdobně taky říkáme, že 1 a -1 jsou triviální jednotky v $\mathbb{Z}[\sqrt{d}]$. Obecně budeme v povídání o Pellově rovnici ztotožňovat dvojici (x, y) a číslo $x + y\sqrt{d}$, takže třeba klidně řekneme, že $x + y\sqrt{d}$ je řešení Pellovy rovnice.

Obecněji se rovnicím $x^2 - dy^2 = c$, kde c je nějaké celé číslo, říká „rovnice Pellova typu“. Speciálně pro $c = -1$, což popisuje jednotky s normou -1, budeme této rovnici říkat „záporná Pellova rovnice“.

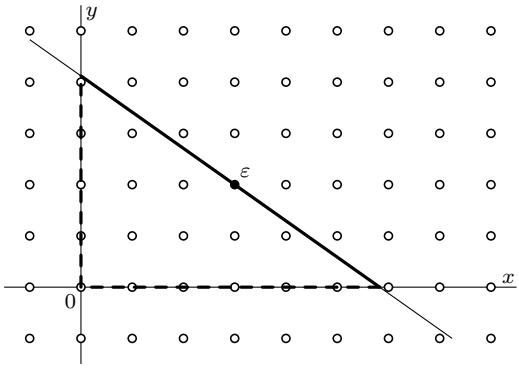
Cvičení 9. Jaká by byla řešení Pellovy rovnice, kdybychom povolili, aby $d = a^2$ byl čtverec celého čísla?

Za chvíli si ukážeme, že vždy existují i nějaké další, *netriviální* jednotky. Začneme ale tvrzením o jednotkách v $\mathbb{Z}[\sqrt{d}]$, které je jak snazší na dokázání, tak i mnohem užitečnější při řešení úloh. Ukážeme, že všechny jednotky se dají vyrobit z té „nejmenší“. Nejprve si pořádně zadefinujme, co tím přesně budeme myslet. Budeme u toho předpokládat, že vůbec nějaké netriviální jednotky existují, avšak tento předpoklad se později ukáže býti automaticky splněn, neboť existenci netriviální jednotky budeme mít zaručenu vždy, když d není čtverec.

Definice. Předpokládejme, že v $\mathbb{Z}[\sqrt{d}]$ existují netriviální jednotky. *Fundamentální jednotkou* potom rozumíme takovou jednotku $x+y\sqrt{d}$, které má kladné složky, a navíc je mezi všemi takovými jednotkami hodnota $x + y\sqrt{d}$ jakožto reálného čísla nejmenší.

Nad touto definicí se stojí zato pořádně zarazit, jelikož na první pohled není ani zřejmé, že vždy dává dobrý smysl. Některé nekonečné množiny totiž mezi sebou žádný nejmenší prvek mít nemusí: když například vezmeme převrácené hodnoty přirozených čísel $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5} \ldots$, pak mezi nimi žádné nejmenší číslo nenajdeme, protože pro každé $\frac{1}{n}$ je $\frac{1}{n+1}$ ještě menší. Množina jednotek v $\mathbb{Z}[\sqrt{d}]$ přitom dost možná nekonečná bude, takže tady máme zavařeno na pořádný problém.

Podívejme se na definici geometricky. Čísla $x+y\sqrt{d}$ si představme jako celočíselné body v rovině, to jest jako body s celočíselnými souřadnicemi (x, y). Podmínka $x, y > 0$ říká, že se díváme jenom na body nad osou x, napravo od osy y. Dále si vezměme nějakou netriviální jednotku $\varepsilon = x_0+y_0\sqrt{d}$. BÚNO nechť jsou x_0, y_0 obě kladná – vždy můžeme namísto ε vzít $-\varepsilon$, což obrátí znaménka obou složek, či $\overline{\varepsilon}$, což obrátí znaménko y_0 (anebo $-\overline{\varepsilon}$ obrátí znaménko x_0). Když takto máme jednu zvolenou jednotku, pak už tu nejmenší, pokud tedy existuje jedna nejmenší, stačí hledat mezi těmi, které jsou menší než ε. K podmínce $x, y > 0$ nám tak přibude $x + y\sqrt{d} \leq \varepsilon$. Rovnice $x + y\sqrt{d} = \varepsilon$ přestavuje nějakou přímku se záporným sklonem a nerovnost $x+y\sqrt{d} \leq \varepsilon$ povoluje jenom body pod touto přímkou. Dohromady tak podmínky vymezí trojúhelník (jeho odvěsny do něj nezahrnujeme, zatímco přeponu obsahující ε už ano).



V něm je jakožto v omezené množině jen konečně mnoho celočíselných bodů, což odpovídá konečně mnoha prvkům $\mathbb{Z}[\sqrt{d}]$. Z těch si vybereme ty, které jsou jednotkami, a vezmeme už bez problému minimum z konečně mnoha čísel (určitě v trojúhelníku máme jednotku ε, takže nebereme minimum z prázdné množiny). Nejmenší jednotka v trojúhelníku bude zároveň také menší než ostatní jednotky mimo trojúhelník, ale stále nad osou x a napravo od osy y, takže to celkově bude fundamentální jednotka, kterou hledáme. Tím je rozmyšleno, že definice fundamentální jednotky dává smysl.

S fundamentální jednotkou řádně zadefinovanou si už dokážeme, že z ní dovedeme vyrobit všechny ostatní.

Věta. *Nechť v $\mathbb{Z}[\sqrt{d}]$ existují netriviální jednotky a ω je fundamentální jednotka. Potom se každá jednotka dá zapsat jako $\pm\omega^n$ pro nějaké $n \in \mathbb{Z}$.*

Před důkazem samotné věty si vyrobíme drobné lemma.

Lemma. *Pro jednotku $x + y\sqrt{d}$ platí $x, y > 0$ právě tehdy, když $x + y\sqrt{d} > 1$.*

Důkaz lemmatu. Jedna implikace je zřejmá: pokud $x, y > 0$, znamená to $x, y \geq 1$, takže i

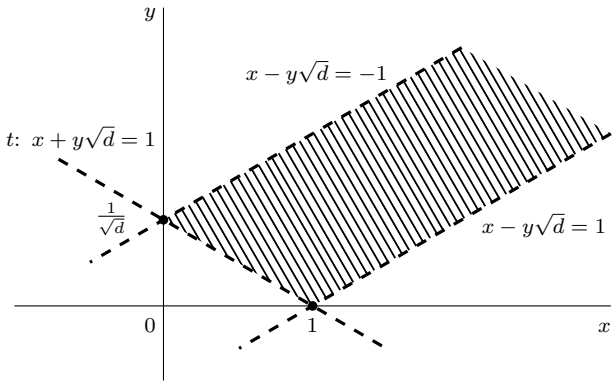
$$x + y\sqrt{d} > x \geq 1,$$

neboť $\sqrt{d}$ je kladné číslo. Nadále tedy naopak předpokládejme, že $x + y\sqrt{d}$ je jednotka větší než 1, a vyvoďme $x, y > 0$.

Rovnice $x + y\sqrt{d} = 1$ určuje v rovině přímku t, která protíná osu x v bodě $(1, 0)$ a osu y v bodě $\left(0, \frac{1}{\sqrt{d}}\right)$, takže nerovnost $x + y\sqrt{d} > 1$ určuje polorovinu nad přímkou t. Tím, že číslo $x + y\sqrt{d}$ je větší než 1, musí i jeho absolutní hodnota být větší než 1. Také předpokládáme, že $x + y\sqrt{d}$ je jednotka, což je ekvivalentní s $|x^2 - dy^2| = 1$. To upravíme na

$$1 = |x^2 - dy^2| = |x + y\sqrt{d}| \cdot |x - y\sqrt{d}| > 1 \cdot |x - y\sqrt{d}|,$$

takže $|x - y\sqrt{d}| < 1$ neboli $-1 < x - y\sqrt{d} < 1$. Přímka $x - y\sqrt{d} = 1$ je jenom přímka t převrácená podle osy x, zatímco $x - y\sqrt{d} = -1$ je t převrácená podle osy y.



Jednotka, která je větší než 1, tedy musí odpovídat bodu v polorovině nad t a zároveň v pruhu mezi rovnoběžkami $x - y\sqrt{d} = \pm 1$. Z obrázku je ale vidět, že celá tato oblast leží nad osou x a napravo od osy y, takže $x, y > 0$. □

Důkaz věty. Jelikož je ω fundamentální jednotka, máme $|N(\omega)| = 1$. Vzhledem k $\omega\overline{\omega} = \pm 1$ je $\overline{\omega} = \pm\omega^{-1}$, takže i ω^{-1} je prvek $\mathbb{Z}[\sqrt{d}]$, a tedy jednotka. Mocněním jednotky dostaneme opět jednotku, neboť

$$|N(\omega^n)| = |N(\omega)|^n = 1^n = 1,$$

takže všechny $\pm\omega^n$ budou jednotky. Povolujeme přitom i záporné n, protože z definice jednotky $\omega \mid 1$, takže $\omega^{-1} \in \mathbb{Z}[\sqrt{d}]$. Zbývá tak dokázat, že žádné další jednotky neexistují.

Nechť pro spor nějaká jednotka ε nejde vyjádřit ve tvaru $\pm\omega^n$. BÚNO nechť má ε obě složky kladné – kdyby nemělo, vzali bychom namísto něj $-\varepsilon$, $\overline{\varepsilon}$ nebo $-\overline{\varepsilon}$. Potom podle lemmatu $\varepsilon > 1$. Stejně tak $\omega > 1$, takže můžeme interval $(1, \infty)$ rozdělit na menší intervaly pomocí posloupnosti čísel $1 = \omega^0, \omega^1, \omega^2, \ldots$ Předpokládáme, že se ε žádnému z nich nerovná, takže leží ostře mezi některými dvěma. To dává nerovnosti

$$\omega^n < \varepsilon < \omega^{n+1}$$

pro nějaké n. Vše přenásobíme ω^{-n}, čímž dostaneme

$$1 < \varepsilon\omega^{-n} < \omega.$$

Číslo $\varepsilon\omega^{-n}$ je zase jednotka v $\mathbb{Z}[\sqrt{d}]$, protože je to jenom součin dvou jednotek. Také je to jednotka větší než 1, takže podle lemmatu má obě složky kladné. Přitom je ale menší než ω. To je spor s tím, že ω má být fundamentální, tedy „nejmenší“ jednotka. Úspěšně jsme vyvodili spor, takže takové ε, které by se nedalo vyjádřit jako $\pm\omega^n$, neexistuje. □

Poznámka. Z lemmatu, které jsme použili v důkazu věty, nám také plyne, jak bude $\pm\omega^n$ vypadat v závislosti na znaménku n a znaménku $\pm$. Především jednotky s oběma složkami kladnými budou vždy odpovídat ω^n pro přirozená n. Pro záporná n už bude situace záviset na normě ω, pro $N(\omega) = 1$ bude $\omega^{-1} = \overline{\omega}$, zatímco pro $N(\omega) = -1$ platí $\omega^{-1} = -\overline{\omega}$. Znaménko minus pak jenom obě složky obrátí.

Definice. Předpokládejme, že Pellova rovnice pro nějaké d má netriviální řešení. *Fundamentálním řešením* potom rozumíme takové řešení $x + y\sqrt{d}$, které má $x, y > 0$ a navíc je mezi všemi takovými řešeními hodnota $x + y\sqrt{d}$ jakožto reálného čísla nejmenší.

Fundamentální řešení Pellovy rovnice tak plní stejnou roli jako fundamentální jednotka, jenom jsme se omezili pouze na ty jednotky, které mají normu 1. Nepřekvapí tedy, že stejně jako všechny jednotky lze vyrobit z té fundamentální, tak i všechna řešení Pellovy rovnice jde vyrobit z toho fundamentálního.

Nechť je ω fundamentální jednotka a ω_1 fundamentální řešení Pellovy rovnice. Kdyby $N(\omega) = 1$, pak už to znamená, že všechny jednotky mají normu 1, žádné jednotky s normou -1 neexistují a jednotkami jsou právě řešení Pellovy rovnice. Potom je tedy $\omega = \omega_1$.

Kdyby naopak $N(\omega) = -1$, pak potom stejně $N(\omega^2) = 1$. Kdyby nyní $\omega_1 \neq \omega^2$, byl by to spor, protože jakožto jednotka musí ω_1 být rovno nějakému ω^n, a kdyby $n > 2$, pak by ω^2 bylo menší řešení Pellovy rovnice než ω_1. V tomto případě je také každé řešení Pellovy rovnice rovno $\pm\omega_1^n$. Musí se totiž jednat o jednotku, takže lze vyjádřit jako $\pm\omega^m$. Jeho norma však má být 1, takže $1 = N(\pm\omega^m) = (-1)^m$, což znamená, že m je sudé neboli $m = 2n$, takže $\pm\omega^m = \pm\omega^{2n} = \pm\omega_1^n$. Naopak mocniny $\pm\omega^\ell$ pro liché ℓ budou mít normu $(-1)^\ell = -1$. Shrňme:

Věta. *Nechť je ω fundamentální jednotka a ω_1 fundamentální řešení Pellovy rovnice. Potom:*

(i) *Pokud $N(\omega) = 1$, pak je $\omega_1 = \omega$ a jednotky s normou -1 neexistují.*

(ii) *Pokud $N(\omega) = -1$, pak $\omega_1 = \omega^2$. Všechny jednotky jsou tvaru $\pm\omega^n$, přičemž pro sudá n se jedná o řešení Pellovy rovnice, zatímco pro lichá n o jednotky s normou -1.*

V obou případech jsou řešeními Pellovy rovnice právě všechna $\pm\omega_1^n$ pro $n \in \mathbb{Z}$.

Příklad. V $\mathbb{Z}[\sqrt{5}]$ je fundamentálním řešením Pellovy rovnice $9 + 4\sqrt{5}$. Fundamentální jednotkou je však $2 + \sqrt{5}$ a platí $9 + 4\sqrt{5} = (2 + \sqrt{5})^2$.

Úloha 2. Dokaž, že záporná Pellova rovnice $x^2 - 34y^2 = -1$ nemá celočíselné řešení.

Nabízí se ještě otázka, jak fundamentální jednotku či fundamentální řešení Pellovy rovnice najít. Zde však trochu zklameme: v seriálu se nebudeme zabývat[3] čímkoliv lepším, než že prostě vyzkoušíme nějaká malá y a budeme doufat, že $dy^2 \pm 1$ bude nějaký čtverec x^2. Pokud tímto najdeme nějakou jednotku, pak už skutečně najdeme tu nejmenší, neboť fundamentální jednotka $x + y\sqrt{d}$ má mezi jednotkami s $x, y > 0$ zároveň i nejmenší x a nejmenší y. Při hledání fundamentální jednotky tímto způsobem je však třeba nemít smůlu, neboť i pro některá malá d může být fundamentální jednotka veliká: například pro $d = 61$ je fundamentální jednotkou $29718 + 3805\sqrt{61}$.

Cvičení 10. Pro některé speciální tvary d si lze nějaké řešení Pellovy rovnice hned tipnout. Najdi nějaký předpis pro x a y závislý na a, tak aby $x + y\sqrt{d}$ bylo řešení Pellovy rovnice pro

(i) $d = a^2 - 1$, (ii) $d = a^2 - 2$, (iii) $d = a^2 + 2$, (iv) $d = a^2 + 1$.

Pellova rovnice a řešení diofantických rovnic

Příklad. Najdi všechny dvojice přirozených čísel $n < m$ takové, že

$$1 + 2 + \cdots + n = (n + 1) + (n + 2) + \cdots + m.$$

Řešení. Použijeme vzorec $1 + \cdots + n = \frac{n(n+1)}{2}$. Na obě strany přičteme $1 + \cdots + n$, čímž rovnici upravíme do tvaru $2 \cdot \frac{n(n+1)}{2} = \frac{m(m+1)}{2}$. Obě strany vynásobíme osmi a potom přeuspořádáme na

$$(4m^2 + 4m + 1) + 1 = 2(4n^2 + 4n + 1)$$

neboli $(2m + 1)^2 - 2(2n + 1)^2 = -1$. Substitucí $x = 2m + 1$, $y = 2n + 1$ je toto záporná Pellova rovnice v $\mathbb{Z}[\sqrt{2}]$, kde je fundamentální jednotkou $\omega = 1 + \sqrt{2}$. Ta má normu -1, takže budeme brát

[3]Pokud by ses chtěl(a) dozvědět o tom, jak hledat řešení Pellovy rovnice pomocí řetězových zlomků, odkážeme Tě na tento příspěvek: *https://prase.cz/library/PellADL/PellADL.pdf*.

její liché mocniny. Zároveň chceme, aby n, m byla přirozená, takže určitě $x, y > 0$, pročež můžeme brát jenom $x + y\sqrt{2} = \omega^\ell$ pro kladná lichá ℓ. Jelikož $n, m \geq 1$, tak však dokonce $x, y \geq 3$, což pro $\ell = 1$ nedostaneme, uvažujeme tedy dokonce jenom lichá $\ell \geq 3$ (pro $\ell = 3$ máme $x = 7$, $y = 5$, takže už jsme v suchu). Zároveň také musíme splnit, že x, y jsou lichá. To je ale splněno pro každé řešení záporné Pellovy rovnice $x^2 - 2y^2 = -1$. Když se na ni podíváme modulo 2, pak uvidíme, že x je liché. Potom $x^2 \equiv 1 \pmod 4$, takže už $2y^2 \equiv -2 \equiv 2 \pmod 4$, což by pro sudé y neplatilo, takže y je také liché.

V souhrnu jsme ukázali, že validní řešení dostaneme právě pro každé liché $\ell \geq 3$. Když tedy ještě zapíšeme $\ell = 2k + 1$ pro přirozené k, pak už skrze $m = \frac{x-1}{2}$, $n = \frac{y-1}{2}$ dopočítáme

$$m = \frac{1}{2}\left(\frac{\omega^\ell + (\overline{\omega})^\ell}{2} - 1\right) = \frac{(1+\sqrt{2})^{2k+1} + (1-\sqrt{2})^{2k+1} - 2}{4},$$
$$n = \frac{1}{2}\left(\frac{\omega^\ell - (\overline{\omega})^\ell}{2\sqrt{2}} - 1\right) = \frac{(1+\sqrt{2})^{2k+1} - (1-\sqrt{2})^{2k+1} - 2}{4}.$$

Řešení obnášelo spoustu drobných technických detailů, ale základní myšlenka je jednoduchá – převést na Pellovu rovnici, vzít mocniny fundamentální jednotky (či fundamentálního řešení) a rozmyslet, které z nich dávají smysluplná řešení původní úlohy. Také si můžeme všimnout toho, co se obvykle stane, když něco vyřešíme Pellovou rovnicí: dostaneme předpis, který vypadá dost složitě (vyrábíme celé číslo pomocí mocnění nějakých výrazů s odmocninami), a na první pohled není vidět, co se v něm děje, ale přesto nám umožňuje všechna řešení popsat a dosazením nějakého ℓ si jakékoliv z nich spočítat.

Cvičení 11. Řeš v celých číslech rovnici $x^2 + y^2 - 1 = 4xy$.

Cvičení 12. Najdi třetí nejmenší přirozené číslo n, pro něž je $\frac{n(n+1)}{2}$ čtverec.

Příklad. Dokaž, že když je $m = 2\sqrt{12n^2 + 1} + 2$ celé číslo pro nějaké $n \in \mathbb{N}$, pak už je m čtverec celého čísla.

Řešení. Aby m bylo celé, tak musí výraz pod odmocninou být čtverec, takže máme $12n^2 + 1 = x^2$ pro nějaké $x \in \mathbb{N}$ (určitě $12n^2 + 1 > 0$). To je jenom přeuspořádaná Pellova rovnice $x^2 - 12n^2 = 1$. Fundamentální řešení Pellovy rovnice pro $d = 12$ je $\omega = 7 + 2\sqrt{12}$ a $x + n\sqrt{12}$ je také nějaké řešení, takže $x + n\sqrt{12} = \pm\omega^k$ pro nějaké $k \in \mathbb{Z}$. Víme ale, že x i n jsou přirozená, tedy kladná, takže v předchozím vyjádření bude určitě $\pm = +$ a exponent k bude kladný.

Ze vzorce pro racionální složku nyní vyjádříme $x = \frac{\omega^k + (\overline{\omega})^k}{2} = \frac{\omega^k + \omega^{-k}}{2}$, následně pak vyjádříme

$$m = 2x + 2 = \omega^k + 2 + \omega^{-k}.$$

Když se pohybujeme kolem Pellovy rovnice, tak takovýto výraz zpravidla nemá daleko k tomu, aby byl čtvercem, neboť $2 = 2\omega^{\frac{k}{2}}\omega^{-\frac{k}{2}}$. Potřebovali bychom tudíž, aby ω^k a ω^{-k} byly čtverce nějakých A, B, neboť potom už budeme moci vyjádřit m jako $(A+B)^2$. Tady však využijeme toho, že okruh $\mathbb{Z}[\sqrt{12}]$ leží uvnitř $\mathbb{Z}[\sqrt{3}]$, jelikož $\sqrt{12} = 2\sqrt{3}$. Přitom v $\mathbb{Z}[\sqrt{3}]$ máme i jednotky, které v $\mathbb{Z}[\sqrt{12}]$ neleží. Jmenovitě $\omega_0 = 2 + \sqrt{3}$ splňuje $\omega = \omega_0^2$. Z toho už dostaneme

$$m = \omega^k + 2 + \omega^{-k} = \omega_0^{2k} + 2 \cdot \omega_0^k \cdot \omega_0^{-k} + \omega_0^{-2k} = (\omega_0^k + \omega_0^{-k})^2.$$

Zde ještě není vyhráno – co kdyby výraz uvnitř závorky nebyl celé číslo? Býti čtvercem v $\mathbb{Z}[\sqrt{3}]$ a býti čtvercem v $\mathbb{Z}$ totiž není totéž – viz třeba $3 = (\sqrt{3})^2$. Toto dořešíme následovně: ω_0^k má v $\mathbb{Z}[\sqrt{3}]$ normu 1, takže $\omega_0^{-k} = \overline{(\omega_0^k)}$. Číslo $\omega_0^k + \omega_0^{-k}$ je tak jenom dvojnásobek racionální složky ω_0^k, což je určitě celé číslo, a m je tak čtvercem v $\mathbb{Z}$, jak jsme chtěli.

Cvičení 13. Číslo $m = 2 + 2\sqrt{28n^2 + 1}$ je celé pro nějaké $n \in \mathbb{N}$. Dokaž, že m je čtverec celého čísla.

Úloha 3. Je dáno přirozené číslo n takové, že n^2 lze vyjádřit jako rozdíl dvou po sobě jdoucích třetích mocnin. Dokaž, že $2n-1$ je čtverec.

Úloha 4. (těžší) Dokaž, že existují rostoucí nekonečné posloupnosti $\{a_n\}$, $\{b_n\}$ přirozených čísel takové, že $a_n(a_n+1) \mid b_n^2+1$ pro každé n.

Existence netriviální jednotky*

Nastal čas splnit ten velký rest, který nám doposud zůstával – dokážeme, že vždy když je d přirozené a není to čtverec, má Pellova rovnice $x^2-dy^2=1$ netriviální řešení. Tento důkaz je trochu náročný, takže pokud se zasekneš, neboj se tento oddíl přeskočit a vrátit se k němu později.

Během důkazu na několika místech použijeme Dirichletův[4] princip[5]. To také způsobí, že nám tento důkaz nedá žádný dobrý algoritmus na to, jak netriviální řešení Pellovy rovnice najít – jenom dokážeme, že existuje.

Tvrzení. (Dirichletův princip) *Když rozdělíme $n+1$ předmětů do n přihrádek, pak v alespoň jedné přihrádce skončí alespoň dva předměty.*

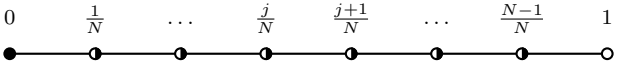
Jak to tak někdy v matematice bývá, i takto zřejmé a jednoduché tvrzení dovede být silným nástrojem a dokázat velké věci – malá ochutnávka přijde již za chvíli. Kromě obyčejného Dirichletova principu máme ještě jeho nekonečnou odrůdu.

Tvrzení. (nekonečný Dirichletův princip) *Když rozdělíme nekonečně mnoho předmětů do konečně mnoha přihrádek, pak v alespoň jedné přihrádce skončí nekonečně mnoho předmětů.*

S touto výzbrojí se již můžeme vrhnout na Pellovu rovnici. Přiblížíme se k ní postupně: začneme tím, jak se dá $\sqrt{d}$ aproximovat pomocí racionálních čísel. Intuice, proč by toto mohlo souviset s Pellovou rovnicí je následovná: když rovnici $x^2-dy^2=1$ upravíme na $d=\frac{x^2}{y^2}+\frac{1}{y^2}$, pak si můžeme říct, že $\frac{1}{y^2}$ bude typicky něco dost malého, takže to můžeme zahodit. Potom nám zbude $d\approx\frac{x^2}{y^2}$ neboli $\sqrt{d}\approx\frac{x}{y}$. To znamená, že když je $x+y\sqrt{d}$ řešení Pellovy rovnice, pak zlomek $\frac{x}{y}$ leží dost blízko $\sqrt{d}$.

Věta. (Dirichletova o diofantických aproximacích) *Nechť je dáno reálné číslo α a přirozené N. Potom existuje $q\in\{1,\dots,N\}$ a celé číslo p tak, že $|q\alpha-p|<\frac{1}{N}$.*

Důkaz. Půjdeme na to postupně. Pro každé $k\in\{1,\dots,N\}$ se podívejme na reálné číslo $k\alpha$ a zvolme celé číslo p_k jako $\lfloor k\alpha\rfloor$, tedy největší celé číslo, které nepřevyšuje $k\alpha$. Potom budou $k\alpha-p_k$ zbylé necelé části z $k\alpha$, takže budou všechny ležet někde v intervalu $\langle 0,1)$. Tenhle interval si rozřežeme na N přihrádek délky $\frac{1}{N}$, tedy na intervaly $\left\langle\frac{j}{N},\frac{j+1}{N}\right)$ pro $j=0,1,\dots,N-1$.



Nejprve koukněme na přihrádku $\langle 0,\frac{1}{N}\right)$. Kdyby v ní leželo nějaké $k\alpha-p_k$, pak by to znamenalo $|k\alpha-p_k|<\frac{1}{N}$, takže stačí zvolit $q=k$, $p=p_k$ a máme vyhráno. Nadále tedy předpokládejme, že v přihrádce $\left\langle 0,\frac{1}{N}\right)$ žádné z našich čísel neleží. Potom jsme ale do zbylých $N-1$ přihrádek museli rozmístit N předmětů $k\alpha-p_k$, takže z Dirichletova principu v některém $\left\langle\frac{j}{N},\frac{j+1}{N}\right)$ skončily dvě čísla $k_1\alpha-p_{k_1}$, $k_2\alpha-p_{k_2}$, BÚNO $k_1<k_2$. Jelikož ale tato čísla leží uvnitř intervalu délky $\frac{1}{N}$,

[4]Peter Gustav Lejeune Dirichlet (1805–1859), německý matematik, má kromě principu se svým jménem spojeny hned tři důležité věty v teorii čísel – jednu z nich si zde dokážeme.

[5]Též známý jako princip holubníku.

který neobsahuje svůj pravý koncový bod, tak jejich vzdálenost od sebe musí být ostře menší než $\frac{1}{N}$. Když tedy zvolíme $q = k_2 - k_1$ a $p = p_{k_2} - p_{k_1}$, dostaneme

$$\frac{1}{N} > |(k_2\alpha - p_{k_2}) - (k_1\alpha - p_{k_1})| = |(k_2 - k_1)\alpha - (p_{k_2} - p_{k_1})| = |q\alpha - p|,$$

jak jsme chtěli. Přitom q je rozdíl dvou přirozených čísel (větší mínus menší) z množiny $\{1, \ldots, N\}$, takže samo musí taky ležet někde v této množině. Tím jsme splnili vše, co q a p měla splňovat. □

Důsledek. *Nechť je α iracionální. Potom existuje nekonečně mnoho dvojic p, q takových, že* $\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}$.

Důkaz. Použijeme Dirichletovu větu. Když platí $|q\alpha - p| < \frac{1}{N}$ a zároveň $q \in \{1, \ldots, N\}$, takže $q \leq N$ neboli $\frac{1}{N} \leq \frac{1}{q}$, z čehož $\left|\alpha - \frac{p}{q}\right| < \frac{1}{qN} \leq \frac{1}{q^2}$. Každá dvojice, kterou dostaneme z Dirichletovy věty, tak vyhovuje důsledku.

Už tedy jenom potřebujeme těchto dvojic vyrobit nekonečně mnoho. Začněme tím, že si pro $N_1 = 1$ vyrobíme nějakou dvojici p_1, q_1. Jelikož předpokládáme, že α je iracionální, tak platí $\left|\alpha - \frac{p_1}{q_1}\right| \neq 0$, takže můžeme zvolit nějaké N_2 dost velké na to, aby $\frac{1}{N_2} < \left|\alpha - \frac{p_1}{q_1}\right|$. Pro tohle N_2 pak použijeme Dirichletovu větu, což nám dá nějakou dvojici p_2, q_2. Vzhledem k volbě N bude platit

$$\left|\alpha - \frac{p_2}{q_2}\right| < \frac{1}{q_2 N_2} \leq \frac{1}{N_2} < \left|\alpha - \frac{p_1}{q_1}\right|.$$

Zlomek $\frac{p_2}{q_2}$ tedy aproximuje α ostře lépe než $\frac{p_1}{q_1}$, takže to nemohou být stejné zlomky, a ani tedy stejné dvojice čísel. Takto postupujeme stále dál: vždy zvolíme N_{j+1} tak, aby $\frac{1}{N_{j+1}} < \left|\alpha - \frac{p_j}{q_j}\right|$, a touto volbou N_{j+1} vyrobíme další dvojici p_{j+1}, q_{j+1}. Každý další zlomek bude ležet blíž k α než ten předchozí, takže budou všechny tyto dvojice p, q navzájem různé a vyrobíme jich tak nekonečně mnoho. □

Důsledek věty nyní použijeme pro $\alpha = \sqrt{d}$. To bude fungovat, neboť předpokládáme, že d není čtverec, a už víme, že potom je $\sqrt{d}$ iracionální.

Lemma. (první) *Když* $\left|\sqrt{d} - \frac{p}{q}\right| < \frac{1}{q^2}$, *pak* $|N(p + q\sqrt{d})| < 2\sqrt{d} + 1$.

Důkaz. Označme si $\beta = -p + q\sqrt{d}$, potom chceme dokázat $|N(\beta)| < 2\sqrt{d} + 1$ (obrácení znaménka u p oproti znění lemmatu nemá na normu vliv). Dále máme předpoklad $|\beta| = q \cdot \left|\sqrt{d} - \frac{p}{q}\right| < \frac{1}{q}$. Chceme odhadnout výraz $|\beta\overline{\beta}|$, takže ještě chceme omezit $\overline{\beta}$. K tomu použijeme trojúhelníkovou nerovnost na

$$|\overline{\beta}| = |-p - q\sqrt{d}| = |-2q\sqrt{d} + (-p + q\sqrt{d})| \leq |-2q\sqrt{d}| + |-p + q\sqrt{d}| < 2q\sqrt{d} + \frac{1}{q}.$$

Následně už máme $|N(\beta)| = |\beta| \cdot |\overline{\beta}| < \frac{1}{q} \cdot \left(2q\sqrt{d} + \frac{1}{q}\right) = 2\sqrt{d} + \frac{1}{q^2} \leq 2\sqrt{d} + 1$, jak jsme chtěli. □

Lemma. (druhé) *Nechť je $k \neq 0$ celé číslo. Potom pokud $N(a + b\sqrt{d}) = N(x + y\sqrt{d}) = k$ a zároveň modulo k platí $a \equiv x$ a $b \equiv y$, pak už je $\frac{x+y\sqrt{d}}{a+b\sqrt{d}}$ řešení Pellovy rovnice.*

Důkaz. Lemma v podstatě říká dvě nezřejmé věci: podíl $\frac{x+y\sqrt{d}}{a+b\sqrt{d}}$ je prvek $\mathbb{Z}[\sqrt{d}]$ a zároveň má normu 1. To, že bude mít normu 1 je zjevné – multiplikativitou normy máme

$$N\left(\frac{x + y\sqrt{d}}{a + b\sqrt{d}}\right) = \frac{N(x + y\sqrt{d})}{N(a + b\sqrt{d})} = \frac{k}{k} = 1.$$

Zbývá se tedy zaměřit na to, aby zkoumaný podíl byl prvek $\mathbb{Z}[\sqrt{d}]$. Nu, zkusme prostě přímočaře vydělit s pomocí sdruženého čísla. Dostaneme

$$\frac{x+y\sqrt{d}}{a+b\sqrt{d}} = \frac{(x+y\sqrt{d})(a-b\sqrt{d})}{(a+b\sqrt{d})(a-b\sqrt{d})} = \frac{(xa-byd)+(ya-xb)\sqrt{d}}{a^2-db^2}.$$

Jmenovatel je nyní norma $a+b\sqrt{d}$, což je k. Modulo k pak pro složky čitatele platí

$$xa-byd \equiv a\cdot a - b\cdot bd \equiv k \equiv 0 \pmod{k}, \qquad ya-xb \equiv b\cdot a - a\cdot b \equiv 0 \pmod{k}.$$

To značí, že k dělí obě složky čitatele, takže $\frac{xa-byd}{k} + \frac{ya-xb}{k}\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$, jak jsme chtěli. □

Máme již všechny potřebné ingredience, zbývá už je jen naházet do hrnce.

Věta. *Když přirozené d není čtverec, pak má Pellova rovnice $x^2-dy^2=1$ netriviální řešení.*

Důkaz. Budeme postupovat takto: seženeme si hromadu prvků $\mathbb{Z}[\sqrt{d}]$ se stejnou normou k, které navíc mají stejné složky modulo k. Z těch nám potom druhé lemma vyrobí spoustu jednotek. Na to, abychom takovou hromádku prvků našli, použijeme důsledek Dirichletovy věty, první lemma a nekonečný Dirichletův princip.

Z důsledku Dirichletovy věty máme hromádku nekonečně mnoha dvojic p, q takových, že $\left|\sqrt{d}-\frac{p}{q}\right| < \frac{1}{q^2}$. Dále podle prvního lemmatu každá taková dvojice splňuje $N(p+q\sqrt{d}) < 2\sqrt{d}+1$. Přitom je ale tato norma celé číslo. Máme tedy nekonečnou hromádku čísel $p+q\sqrt{d}$, z nichž každé si vybere jedno z konečně mnoha celých čísel s absolutní hodnotou menší než $2\sqrt{d}+1$. Takových čísel je jen konečně mnoho, takže podle nekonečného Dirichletova principu pro nějaké celé číslo k, $|k| < 2\sqrt{d}+1$ má nekonečně mnoho z našich čísel $p+q\sqrt{d}$ normu k.

Dále chceme z této (stále nekonečné) hromádky čísel vybrat taková, aby všechny racionální složky dávaly stejný zbytek modulo k i všechny iracionální složky dávaly stejný zbytek modulo k. Číslu $p+q\sqrt{d}$ tak prostě přiřaďme dvojici zbytků ($p \bmod k, q \bmod k$). To je nějaká dvojice prvků $\mathbb{Z}_k$ a těch je $|\mathbb{Z}_k|^2 = k^2$, což je konečně mnoho. Opětovným použitím nekonečného Dirichletova principu tak musí být některé dvojici zbytků přiřazeno nekonečně mnoho z našich čísel $p+q\sqrt{d}$.

Máme tedy nekonečnou hromádku čísel, na které můžeme vrhnout druhé lemma. To nám vyrobí velikou spoustu jednotek v $\mathbb{Z}[\sqrt{d}]$. Přesněji, zvolme pevně jedno $a+b\sqrt{d}$ z naší výsledné hromádky. Potom můžeme za $x+y\sqrt{d}$ volit postupně všechna ostatní čísla z hromádky a vždy dostaneme jednotku $\frac{x+y\sqrt{d}}{a+b\sqrt{d}}$ s normou 1. Navíc díky tomu, že jednotlivá $x+y\sqrt{d}$ jsou navzájem různá, zatímco $a+b\sqrt{d}$ je stále stejné, budou takto získané jednotky navzájem různé. Tímto tedy vyrobíme nekonečně mnoho různých jednotek s normou 1 v $\mathbb{Z}[\sqrt{d}]$. Přitom ale triviální jednotky existují jen dvě (1 a -1), takže takto určitě vyrobíme alespoň jedno (resp. dokonce nekonečně mnoho) netriviální řešení Pellovy rovnice. Tím je důkaz hotov. □

Cvičení(*) 14. Nechť je $p \equiv 1 \pmod 4$ prvočíslo. Potom záporná Pellova rovnice $x^2-py^2=-1$ má řešení.

Úloha 5. Nechť je $x+y\sqrt{d}$ fundamentální jednotka v $\mathbb{Z}[\sqrt{d}]$. Dokaž, že pro každé $q \in \mathbb{N}$ existuje $n \in \mathbb{N}$ takové, že číslo $a+b\sqrt{d} = (x+y\sqrt{d})^n$ splňuje $q \mid b$.

Pellova rovnice a dělitelnost

Na závěr našeho povídání o reálných kvadratických okruzích a Pellově rovnici se podíváme na to, co se děje, když při mocnění nějakého $x+y\sqrt{d}$ chceme roznásobit a spočítat racionální a iracionální složku. Toto jsme dosud při vyjadřování jednotek ve tvaru $\pm\omega^n$ taktně obcházeli, avšak ukazuje se, že když trochu nahlédneme pod pokličku, získáme užitečné informace o dělitelnosti v situacích kolem Pellovy rovnice a jednotek v $\mathbb{Z}[\sqrt{d}]$.

Na to si nejprve pořídíme nástroj na roznásobení mocniny součtu $(A+B)^n$, přitom pro nás A, B budou libovolné prvky v nějakém komutativním okruhu R a n bude přirozené číslo. Když rozepíšeme

$$(A+B)^n = \underbrace{(A+B)\cdot(A+B)\cdots(A+B)}_{n\text{-krát}}$$

a budeme roznásobovat, určitě dostaneme nějaký výraz s členy tvaru A^aB^b. Každý takový člen přitom vznikne tak, že si postupně z každé z n závorek vybereme buďto A, nebo B a vynásobíme dohromady, co jsme dostali. Tím pádem mohou vzniknout jen ty členy A^aB^b, v nichž jsou $a, b \in \{0, 1, \ldots, n\}$ a zároveň $a+b=n$, takže budeme mít jen členy ve tvaru $A^{n-k}B^k$.

Kolik celkem takových členů $A^{n-k}B^k$ dostaneme? Každý vznikne z toho, že si při roznásobování v k závorkách vybereme B, zatímco ve zbylých vybereme A. Koeficient u $A^{n-k}B^k$ ve výsledném roznásobení mocniny tak musí být roven počtu způsobů, jak si mezi n nerozlišitelnými předměty (závorkami $(A+B)$) zvolit k z nich (ty, z nichž vezmeme B a nikoliv A). Takovýto počet se v různých oblastech matematiky vyskytuje celkem často, proto pro něj máme značení

$$\binom{n}{k},$$

čemuž říkáme *kombinační číslo* a čteme jej „n nad k“. Kombinační čísla splňují spoustu krásných kombinatorických identit a vztahů, my je však nebudeme potřebovat – postačíme si pouze s definicí „kolika způsoby lze mezi n nerozlišitelnými předměty zvolit k z nich“.

Nahlédli jsme tedy následující:

Věta. (binomická) *V libovolném komutativním okruhu platí*

$$(A+B)^n = \binom{n}{0}A^n + \binom{n}{1}A^{n-1}B + \cdots + \binom{n}{k}A^{n-k}B^k + \cdots + \binom{n}{n-1}AB^{n-1} + \binom{n}{n}B^n.$$

My zde binomickou větu použijeme na rozepsání $(x+y\sqrt{d})^n$. Když nás bude zajímat racionální a iracionální složka výsledku, prostě posbíráme ty členy, ve kterých se $\sqrt{d}$ umocnilo na racionální číslo, a ty, kde zůstalo iracionální $\sqrt{d}$.

Tvrzení. *Nechť v $\mathbb{Z}[\sqrt{d}]$ platí rovnost $a+b\sqrt{d} = (x+y\sqrt{d})^n$, přičemž n je přirozené číslo. Potom $y \mid b$. Pokud je n sudé, pak dokonce $xy \mid b$, zatímco když je n liché, pak $x \mid a$.*

Důkaz. Rozepišme podle binomické věty

$$(x+y\sqrt{d})^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y\sqrt{d} + \binom{n}{2}x^{n-2}y^2d + \cdots + \binom{n}{n}y^n(\sqrt{d})^n.$$

Když budeme procházet členy zleva doprava, tak každý druhý dostane $\sqrt{d}$ v mocnině s lichým exponentem, takže bude iracionální, zatímco zbylé budou mít $\left(\sqrt{d}\right)^{2k}$, z čehož nám vyjde racionální číslo. Má-li tedy být $a+b\sqrt{d} = (x+y\sqrt{d})^n$, můžeme posbírat

$$\begin{aligned} a &= \binom{n}{0}x^n + \binom{n}{2}x^{n-2}y^2d + \binom{n}{4}x^{n-4}y^4d^2 + \cdots, \\ b\sqrt{d} &= \binom{n}{1}x^{n-1}y\sqrt{d} + \binom{n}{3}x^{n-3}y^3d\sqrt{d} + \cdots \end{aligned}$$

Když u $b\sqrt{d}$ pokrátíme všechny $\sqrt{d}$, dostaneme dvě rovnosti v celých číslech, jelikož všechna $\binom{n}{k}$ jsou celá. Důležité je, že nevíme přesně, jakým členem která rovnost končí – pro liché n je $(y\sqrt{d})^n$ iracionální, takže se připočte do $b\sqrt{d}$, zatímco pro sudé je racionální, a přibude tak k a.

Pro b je jedna věc jistá – první člen v

$$b = \binom{n}{1} x^{n-1}y + \binom{n}{3} x^{n-3}y^3 d + \cdots$$

je násobkem y, a když pokračujeme zleva doprava k dalšímu členu, exponent u y se jenom zvýší, takže všechny jednotlivé členy napravo budou násobky y nehledě na to, jakým členem končíme. Tímto posledním členem by mohl být $\binom{n}{n-1} xy^{n-1}d^{\frac{n-2}{2}}$, ten do b přibude pro sudé n. Kdyby tomu tak bylo, pak i člen nejvíce napravo, tedy s nejmenším exponentem u x, byl násobkem xy. Když potom budeme procházet členy směrem doleva, budeme zmenšovat exponent u y a zvětšovat ten u x, až skončíme u členu $\binom{n}{1} x^{n-1}y$, který je stále dělitelný xy. To znamená, že všechny členy budou násobky xy, takže i $xy \mid b$.

Když bude naopak n liché, pak bude člen $\binom{n}{n-1} xy^{n-1}d^{\frac{n-1}{2}}$ (změnil se exponent u d, jelikož v a jsme nekrátili $\sqrt{d}$) posledním členem ve vyjádření a. To ale znamená, že jsem cestou zleva doprava šli od členu s x^n až do členu s xy^{n-1}, takže všechny byly násobky x. To znamená, že v tomto případě $x \mid a$, jak jsme chtěli. □

Cvičení(!) 15. Nechť je $x_0 + y_0\sqrt{d}$ fundamentální jednotka. Potom pro každou jednotku $x + y\sqrt{d}$ platí $x_0y_0 \mid xy$. Totéž platí, když namísto jednotek uvážíme řešení Pellovy rovnice.

Cvičení 16. Pokud celá čísla x, y splňují $x^2 - 37y^2 = 1$, pak $876 \mid xy$.

Příklad. Je dáno přirozené číslo n takové, že $3n+1$ i $4n+1$ jsou čtverce. Dokaž, že n je násobek sedmi.

Řešení. Označme si $a^2 = 3n+1$, $b^2 = 4n+1$ a BÚNO $a, b > 0$. Jakmile budeme něco vědět o a, b, pak dopočteme $n = b^2 - a^2 = (b-a)(a+b)$. Také dovedeme odečíst takové násobky a^2, b^2, abychom vyrušili n, konkrétně

$$4a^2 - 3b^2 = 1.$$

Chceme tedy, aby $2a + b\sqrt{3}$ bylo řešení Pellovy rovnice pro $d = 3$. V $\mathbb{Z}[\sqrt{3}]$ je fundamentálním řešením $\omega = 2 + \sqrt{3}$, takže budeme mít $2a + b\sqrt{3} = \omega^k$ pro nějaké $k \in \mathbb{N}$ (chceme $a, b > 0$, takže před ω^k nemusíme uvažovat $\pm$). Označme $x + y\sqrt{3} = \omega^k$. Potřebujeme, aby x vyšlo sudé – nahlédněme, že to nastane právě pro lichá k. Jelikož 2 je racionální složka ω, tak podle tvrzení pro sudé k vyjde sudé y, zatímco pro lichá k bude x sudé. Přitom x, y musí být nesoudělná, jelikož $x^2 - 3y^2 = 1$, takže pro sudá k je nutně x liché. To značí, že v této úloze potřebujeme uvažovat jen lichá k.

Zapišme $k = 2\ell - 1$ pro nějaké $\ell \in \mathbb{N}$. Platí $\omega^2 = (2+\sqrt{3})^2 = 7 + 4\sqrt{3}$. My chceme stran a a b něco říci o dělitelnosti sedmi, takže přítomnost 7 jako racionální části ω^2 může být poněkud návodná. Upravme

$$\begin{aligned} 2a + b\sqrt{3} &= (2+\sqrt{3})^{2\ell-1}, \\ (2a + b\sqrt{3})(2+\sqrt{3}) &= (2+\sqrt{3})^{2\ell}, \\ (4a+3b) + (2a+2b)\sqrt{3} &= (7+4\sqrt{3})^{\ell}. \end{aligned}$$

Pokud je ℓ sudé, pak podle tvrzení bude 7 jakožto racionální složka na pravé straně dělit $2a + 2b$ jakožto iracionální složku na levé straně, takže $7 \mid a+b$. Pokud bude ℓ naopak liché, pak $7 \mid 4a+3b$. Jenže $4a+3b \equiv 4a+3b-7b \equiv 4(a-b) \pmod 7$, takže toto znamená $7 \mid a-b$. Díky $n = (b-a)(a+b)$ tak platí $7 \mid n$ nehledě na paritu ℓ. Tím je úloha vyřešena.

Úloha 6. Ukaž, že v předchozím příkladu platí i $8 \mid n$, takže $56 \mid n$.

Úloha 7. Je dáno přirozené číslo n takové, že $2n+1$ i $3n+1$ jsou čtverce. Dokaž, že n je násobek čtyřiceti.

Modulíme, tentokráte obecně

V druhé polovině tohoto dílu se podíváme znovu a obecněji na to, jak dělit se zbytkem. V prvním díle jsme si ukázali, jak modulit v celých číslech, a nyní provedeme totéž v obecném okruhu.

Definice. Budiž R komutativní okruh. Pro $a, b, m \in R$ řekneme, že a je *kongruentní* s b *modulo* m (značíme $a \equiv b \pmod{m}$), pokud $m \mid a - b$.

V této definici používáme zcela obecnou definici dělitelnosti – zápis $m \mid a - b$ značí prostě to, že existuje $c \in R$ splňující $a - b = c \cdot m$. Obdobně jako v celých číslech si dále ukážeme, že modulení se chová hezky ke sčítání i násobení.

Tvrzení. *Nechť modulo m platí $a \equiv b$ a zároveň $c \equiv d$. Potom platí i*

$$a + c \equiv b + d \pmod{m} \qquad \text{a} \qquad ac \equiv bd \pmod{m}.$$

Důkaz. Z definice kongruence jsou $a - b$ i $c - d$ násobky m. Potom je ale násobkem m i výraz $(a - b) + (c - d) = (a + c) - (b + d)$, což znamená $a + c \equiv b + d \pmod{m}$. Dále je i každý násobek $a - b$ a $c - d$ stále násobkem m, takže

$$m \mid (a - b)c + b(c - d) = ac - bc + bc - bd = ac - bd$$

neboli $ac \equiv bd \pmod{m}$. □

Cvičení 17. Modulení se chová hezky ke sčítání a násobení, ale k jiným operacím už se hezky chovat nemusí. Rozmysli, že v $\mathbb{Z}[\sqrt{d}]$ (pro kladná i záporná d) obecně $a \equiv b \pmod{m}$ neimplikuje $\overline{a} \equiv \overline{b} \pmod{m}$. Pro $m \in \mathbb{Z}$ už to však platí.

Díky tomu, že se modulení chová hezky ke sčítání a násobení, můžeme v kongruencích provádět úpravy podobně jako v rovnicích, i když tyto úpravy nemusí vždy být ekvivalentní. Můžeme také „zapomenout“, přesně s jakými čísly počítáme – důležitý je jenom jejich zbytek po dělení m.

Definice. Budiž R okruh a $m \in R$. Pro $a \in R$ množinu

$$a \bmod m = \{a + xm : x \in R\}$$

nazýváme *zbytkovou třídou* prvku a modulo m. Množinu všech zbytkových tříd modulo m značíme $R/(m)$.

Víme, že při počítání (sčítání a násobení) modulo m nezáleží na tom, který konkrétní prvek ze zbytkové třídy si vezmeme, na což můžeme nahlížet tak, že počítáme přímo se zbytkovými třídami. Nehledě na to, které a ze zbytkové třídy $a_0 \bmod m$ a b ze zbytkové třídy $b_0 \bmod m$ vezmeme, součet $a + b$ bude patřit do stejné zbytkové třídy. Analogicky platí totéž pro součin ab.

Definice. Na množině $R/(m)$ definujeme sčítání a násobení zbytkových tříd pomocí

$$(a \bmod m) + (b \bmod m) = (a + b) \bmod m, \qquad (a \bmod m) \cdot (b \bmod m) = ab \bmod m.$$

Množina $R/(m)$ s těmito operacemi tvoří okruh a okruhy tohoto tvaru nazýváme *faktorokruhy*.

Poznámka. V definici jsme důsledně zapisovali zbytkovou třídu z okruhu $R/(m)$ jako $a \bmod m$, abychom ji odlišili od prvku a z okruhu R. Běžně však budeme počítání v $R/(m)$ zapisovat jako kongruence, takže místo $(7 \bmod 5) \cdot (-3 \bmod 5) = -21 \bmod 5 = 4 \bmod 5$ zapíšeme jenom $7 \cdot (-3) \equiv -21 \equiv 4 \pmod{5}$. Důležitá je změna pohledu: 7 a -3 zde nejsou celá čísla 7 a -3,

ale jejich zbytkové třídy modulo 5. Nebydlí již v okruhu $\mathbb{Z}$, který má nekonečně mnoho prvků, ale v okruhu $\mathbb{Z}/(5)$, který má jen pět prvků.

Příklad. Počítání modulo m jsem již viděli na celých číslech. Pro nenulové $m \in \mathbb{Z}$ je $\mathbb{Z}/(m)$ jenom jiný zápis pro $\mathbb{Z}_m$. Tato množina má $|m|$ prvků, které můžeme reprezentovat čísly $0, 1, \ldots, m-1$. Z prvního dílu také víme, že faktorokruh $\mathbb{Z}/(m)$ je oborem integrity, právě když je m prvočíslo nebo ± 1.

Příklad. Zbytkové třídy modulo 0 jsou vždy prostě jednoprvkové množiny $a \bmod 0 = \{a\}$, takže faktorokruh $R/(0)$ se chová úplně stejně jako původní R.

Obecně může mít faktorokruh $R/(m)$ nekonečně mnoho prvků i pro $m \neq 0$, nicméně okruhy, se kterými jsme doteď pracovali, nám takto divoké faktorokruhy nedávají. Příklady nekonečných faktorokruhů však potkáme v příštím dílu.

Cvičení 18. Rozmysli si, že když je $u \in R$ jednotka, pak má $R/(u)$ jen jeden prvek (je to tedy triviální okruh).

Příklad. Pro celé číslo $m \neq 0$ má faktorokruh $\mathbb{Z}[\sqrt{d}]/(m)$ přesně m^2 prvků. Tyto zbytkové třídy můžeme reprezentovat čísly $a + b\sqrt{d}$ pro $a, b \in \{0, 1, \ldots, m-1\}$.

Příklad. Ukažme, že faktorokruh $\mathbb{Z}[i]/(2+i)$ má pět prvků. Zaprvé $2 + i \mid N(2+i) = 5$, takže stačí jenom určit, které z prvků $a + bi$, $a, b \in \{0, 1, 2, 3, 4\}$ jsou si kongruentní. Modulo $2 + i$ však platí $i \equiv -2$, takže $a + bi \equiv a - 2b$. Ale $a - 2b$ je celé číslo, takže libovolný prvek $\mathbb{Z}[i]$ je modulo $2+i$ kongruentní nějakému celému číslu. Každá zbytková třída modulo $2 + i$ se tedy dá reprezentovat jedním z čísel 0, 1, 2, 3, 4. Ta jsou však navzájem nekongruentní (stačí uvážit normu rozdílu), takže $\mathbb{Z}[i]/(2+i)$ má přesně pět prvků.

Úloha 8. Pro nenulové $\alpha \in \mathbb{Z}[\sqrt{d}]$ (i pro záporná d) má faktorokruh $\mathbb{Z}[\sqrt{d}]/(\alpha)$ přesně $|N(\alpha)|$ prvků.

Kromě počtu prvků $R/(m)$ se můžeme zajímat i o vlastnosti tohoto faktorokruhu. Následující tvrzení nám bude oslím můstkem do další kapitoly.

Tvrzení. *Nechť $p \in R$ není jednotka. Potom je faktorokruh $R/(p)$ oborem integrity, právě když je p prvočinitel v R.*

Důkaz. Stačí si důsledně rozmyslet definice jednotlivých pojmů a rozklíčovat, že výroky „$R/(p)$ je obor integrity“ a „p je prvočinitel“ říkají doslova to samé. $R/(p)$ je oborem integrity, pokud

$$ab \equiv 0 \pmod{p}$$

nastává právě tehdy, když je a nebo b kongruentní nule modulo p. Jinými slovy, $p \mid ab$ jen tehdy, když $p \mid a$ nebo $p \mid b$, což je však přesně definice prvočinitele. □

V tvrzení požadujeme, aby p nebyla jednotka, neboť modulení jednotkou dá triviální okruh, což je obor integrity, ačkoliv jednotky za prvočinitele nepovažujeme.

Konečná tělesa

Před chvíli jsme si ukázali, že prvočísla jsou hezká v tom smyslu, že nám umí vyrábět obory integrity, tedy okruhy, ve kterých dovedeme krátit v rovnicích.

Ale krácení by se dalo zlepšit. Třeba libovolné dělení by bylo ještě lepší. Ano, ono krácení a dělení je k sobě velmi blízko, ale není to to samé. Například v celých číslech krátit můžeme ($2a = 2b$ je to stejné jako $a = b$), ale dělit již nemůžeme, protože například $2x = 1$ není v celých číslech to stejné jako $x = \frac{1}{2}$, jelikož $\frac{1}{2}$ už je číslo racionální, nikoli celé.

Vidíme, že náš protipříklad na to, že krácení a dělení není to stejné, pracuje s nekonečným oborem integrity. To není náhodou. Kdybychom se omezili pouze na konečné obory, už bychom mohli i beztrestně dělit.

Definice. *Tělesem* rozumíme okruh s alespoň dvěma prvky, kde jsou všechny nenulové prvky jednotkami.

Poznámka. Ekvivalentně řečeno: pro každé $x \in T \setminus \{0\}$ existuje (právě jedno) $y \in T$ takové, že $x \cdot y = 1$. Pro nás je spíše přirozené, že existuje zlomek $\frac{1}{x}$, který leží v tomto tělese, protože jsme na zlomky zvyklí z $\mathbb{Q}$ a používali jsme je i v okruzích $\mathbb{Z}_p$.

To, aby těleso mělo alespoň dva prvky, požadujeme prostě proto, že jednoprvkové (triviální) těleso by často odporovalo tvrzením, která platí pro „rozumná“ tělesa. *Konečným tělesem* rozumíme těleso s konečným počtem prvků. Tělesa se často značí K, F nebo T.[6]

Příklad. $\mathbb{Q}$, $\mathbb{R}$ i $\mathbb{C}$ jsou tělesa, neboť v nich umíme dělit kterýmkoliv nenulovým prvkem.

Příklad. Pro prvočíslo p je $\mathbb{Z}_p$ těleso. Pro $a \not\equiv 0 \pmod p$ je totiž 1 největším společným dělitelem a a p, takže máme Bézoutovy koeficienty x, y splňující $xa + yp = 1$, z čehož $xa \equiv 1 \pmod p$.

To, že nám prvočísla vyrobí konečné těleso, není náhodou. $\mathbb{Z}_p = \mathbb{Z}/(p)$ má konečně mnoho prvků a jedná se o obor integrity (p je prvočinitel v $\mathbb{Z}$). Dokážeme si, že tyto dvě podmínky stačí k tomu, aby okruh byl tělesem.

Příklad. Okruh $\mathbb{Z}$ je oborem integrity, ale není tělesem. Má přitom nekonečně mnoho prvků. Naproti tomu $\mathbb{Q}$ je také nekonečný obor integrity a tělesem je.

Tvrzení. *Každý konečný obor integrity je též konečným tělesem.*

Důkaz. Z příkladu vidíme, že musíme nějak využít konečnosti. Budiž R konečný obor integrity a a jeho nenulový prvek. Ukažme, že k němu najdeme prvek b takový, že $a \cdot b = 1$.

Vezměme si nekonečnou posloupnost $a, a^2, a^3, a^4, \dots$ Z konečnosti R se v ní někdy musí nějaký prvek zopakovat, takže pro nějaká $m < n$ nastane $a^m = a^n$. To znamená, že $0 = a^n - a^m = a^m(a^{n-m} - 1)$. Z definice oboru integrity je tak a^m nebo $a^{n-m} - 1$ nulové. Pokud $a^m = 0$, tak můžeme definici oboru integrity aplikovat znovu – pokud je součin $a \cdots a$ nulový, je nulový alespoň jeden činitel, z čehož už nutně plyne, že samotné a je nulové, což odporuje předpokladu.

Proto nutně platí, že mocnina a^{m-n} je rovna 1. Pokud tedy zvolíme $b = a^{n-m-1}$, pak splníme $a \cdot b = a \cdot a^{n-m-1} = a^{n-m} = 1$. □

Důsledek. *Když je $p \in R$ prvočinitel a zároveň je $R/(p)$ konečná množina, pak už je $R/(p)$ těleso.*

Důkaz tvrzení lze vést více způsoby. My jsme si vybrali tento, protože jak za chvíli uvidíme, nejmenší exponent r splňující $a^r = 1$ je sám o sobě zajímavý a využívá se v mnoha úlohách. Důsledek tvrzení dává dobrou představu, v jaké podobě nejčastěji konečná tělesa potkáme – jako faktorokruhy modulo prvočinitel.

Příklad. V $\mathbb{Z}[i]$ je $2+i$ prvočinitelem, takže $\mathbb{Z}[i]/(2+i)$ je těleso. Dříve jsme už viděli, že má pět prvků, které můžeme reprezentovat jako 0, 1, 2, 3 a 4. Za povšimnutí stojí, že díky $2+i \mid N(2+i) = 5$ se tento okruh v podstatě chová stejně jako obyčejné $\mathbb{Z}_5$.

Příklad. Taktéž 3 je prvočinitelem v $\mathbb{Z}[i]$, takže $\mathbb{Z}[i]/(3)$ je těleso, neboť má $3^2 = 9$ prvků. Ty můžeme zapsat jako 0, 1, 2, i, $1+i$, $2+i$, $2i$, $1+2i$, $2+2i$. Můžeme ověřit, že ke každému $a \in \mathbb{Z}[i]/(3)$ lze nalézt $\frac{1}{a}$, kupříkladu k $a = 2 + i$ máme $(2 + i)(1 + i) \equiv 1 + 3i \equiv 1 \pmod 3$.

Cvičení(!) 19. Každé těleso je i obor integrity. To značí, že schopnost dělit je silnější než jen krátit v rovnicích.

---

[6] Po řadě z německého *Körper*, anglického *field* a (překvapivě) českého *těleso*. V češtině se lze občas potkat i s označením „pole“, my jej však používat nebudeme.

Z toho už například plyne, že pro složené číslo n okruh $\mathbb{Z}_n$ není těleso, protože není ani obor integrity.

Cvičení 20. Připomeň si, proč $\mathbb{Z}_n$ pro složené n není obor integrity.

Ukažme si některé věty, které nám v konečných tělesech pomohou se zjednodušováním výrazů.

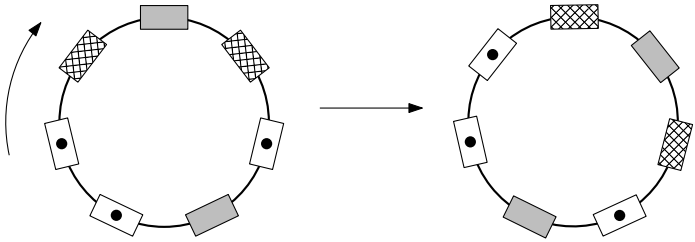
Věta. (malá Fermatova[7]) *Mějme konečné těleso s n prvky. Potom pro každý nenulový prvek a tohoto tělesa platí, že $a^{n-1} = 1$.*

Ukážeme si dva důkazy. První je názorně kombinatorický, nicméně provedeme jej jen pro tělesa $\mathbb{Z}_p$, zatímco druhý funguje pro libovolné konečně těleso.

Důkaz. (korálkový) Představme si, že si chceme vyrobit náhrdelník z korálků. Jak takový náhrdelník vypadá? Jelikož jsme matematici a máme tolik rádi prvočísla, řekli jsme si, že náš náhrdelník bude tvořen p korálky spojenými do kruhu (pro prvočíslo p).

Mít jednobarevný náhrdelník je ale nuda, a tak si každý korálek obarvíme jednou z a barviček. Nyní nás jako matematiky zajímá, kolik různých náhrdelníků takto můžeme vyrobit. To je ale jednoduše a^p, protože na každém z p korálků si vybíráme z a barev.

Po nějaké době, co takový náhrdelník nosíme, si však uvědomíme, že se nám na krku pořád točí, a tudíž jsou náhrdelníky, které se liší jenom pootočením, vlastně stejné. To ale znamená, že jsme některé náhrdelníky započítali ve více otočeních.



Které to jsou? Jednobarevný náhrdelník (těch je přesně a) jsme dvakrát nezapočítali. Naproti tomu náhrdelník, který není jednobarevný, jsme započítali právě p-krát, neboť p je prvočíslo.

Proč tomu tak je? Nechť je pro spor nějaký nejednobarevný náhrdelník započítán méně než p-krát. Pro nějaké k tedy otočením o $k < p$ korálků dostaneme stejný náhrdelník (stejné barvy na jednotlivých pozicích). Toto otočení nyní můžeme provádět opakovaně a vždy tím dostaneme stejný náhrdelník. Díky $k < p$ a prvočíselnosti p jsou k a p nesoudělná, takže pro nějaké x platí $xk \equiv 1 \pmod{p}$. To odpovídá tomu, že když x-krát zopakujeme otočení o k korálků, dostaneme ve výsledku otočení o jeden korálek. Nyní však i toto otočení o jediný korálek vyrobí stále stejný náhrdelník, takže všechny korálky mají stejnou barvu, což je spor.

Všechny náhrdelníky kromě jednobarevných jsme tedy původně započítali p-krát. Celkový počet náhrdelníků, u nichž nehledíme na otočení, je potom $\frac{a^p - a}{p} + a$, neboť z původního počtu a^p bylo a náhrdelníků jednobarevných a pro zbylých $a^p - a$ jsme započítali jen jedno z p otočení. Počet náhrdelníků ale musí být celočíselný, takže $p \mid a^p - a$. Pokud je navíc a nesoudělné s p, pak kongruenci $a^p \equiv a \pmod{p}$ zkrátíme na $a^{p-1} \equiv 1 \pmod{p}$. □

Důkaz. (obecný) Očíslujme si všech $n-1$ nenulových prvků tělesa jako $x_1, \ldots, x_{n-1}$. Dále každý z nich přenásobme nenulovým prvkem a, pro který budeme chtít větu dokázat. Ukažme, že přenásobení tyto prvky jenom zamíchá mezi sebou. Díky tomu, že jich máme jen konečně mnoho, si stačí rozmyslet, že pro $x_i \neq x_j$ bude i $ax_i \neq ax_j$ a že žádné ax_i nebude nulové. Pro spor nechť

[7]Pierre de Fermat (1607–1665) byl francouzský právník a ve volném čase také matematik. K frustraci budoucích generací k většině tvrzení, která objevil, nezanechal žádné důkazy.

$ax_i = ax_j$ pro nějaká $i \neq j$. Jelikož je a nenulové, můžeme jím vydělit, čímž dostaneme $x_i = x_j$, což je spor. Obdobně kdyby $ax_i = 0$, znamenalo by to $x_i = 0$, což neplatí.

Z toho už nutně plyne, že posloupnost $ax_1, ax_2, \ldots, ax_{n-1}$ se od $x_1, x_2, \ldots, x_{n-1}$ liší jen pořadím svých prvků. Když tedy všechny prvky jedné posloupnosti vynásobíme dohromady a s druhou provedeme totéž, určitě dostaneme stejný součin. To nám dává rovnost

$$x_1 \cdot x_2 \cdots x_{n-1} = ax_1 \cdot ax_2 \cdots ax_{n-1} = a^{n-1} \cdot x_1 \cdot x_2 \cdots x_{n-1}.$$

V součinu $x_1 \cdot x_2 \cdots x_{n-1}$ se vyskytují jen nenulové prvky, takže je sám nenulový (těleso je i obor integrity), tudíž jím můžeme obě strany vydělit Tím už dostáváme kýženou rovnost $a^{n-1} = 1$. □

Cvičení 21. Urči zbytek 29^{25} po dělení 11.

Cvičení 22. Urči zbytek $(11 + 7i)^{18}$ po dělení 3.

Úloha 9. Urči zbytek $2^{20} + 3^{30} + 4^{40} + 5^{50} + 6^{60}$ po dělení 7.

Úloha 10. (těžší) Urči, která přirozená čísla x jsou nesoudělná se všemi členy posloupnosti $a_n = 2^n + 3^n + 6^n - 1$.

Úloha 11. Najdi všechna celočíselná řešení rovnice $x^5 + 2 = 101^y$.

Mějme na paměti, že malá Fermatova věta platí pro konečná *tělesa*, takže pro faktorokruhy, které nejsou konečnými tělesy, taková identita platit nemusí. Když například budeme v $\mathbb{Z}$ modulit složeným číslem $9 = 3 \cdot 3$, narazíme na $2^8 \equiv 4 \not\equiv 1 \pmod 9$. Někdy dokonce mocněním nenulového prvků můžeme dostat nulu, jako třeba $3^2 \equiv 0 \pmod 9$.

Ve skutečnosti i pro složené moduly máme identitu podobnou malé Fermatově větě, jenom s jiným exponentem a s omezením na základ mocniny. Uvedeme ji bez důkazu a budeme se dále věnovat konečným tělesům.

Věta. (Eulerova[8]) *Nechť $\varphi(n)$ značí počet jednotek v okruhu $\mathbb{Z}_n$. Pak pro jednotku $a \in \mathbb{Z}_n$ platí $a^{\varphi(n)} \equiv 1 \pmod n$.*

Zatímco malá Fermatova věta říká, co získáme mocněním jednoho prvku, ta následující zase pomáhá tam, kde vynásobíme (skoro) všechny prvky tělesa.

Věta. (Wilsonova[9]) *Buď T těleso s $n - 1$ nenulovými prvky $a_1, a_2, \ldots, a_{n-1}$. Potom platí*

$$a_1 a_2 \cdots a_n = -1,$$

kde -1 značí prvek splňující $(-1) + 1 = 0$.

Důkaz. Využijeme toho, že v tělese je každý nenulový prvek jednotkou, takže pro $a \in T$ existuje (jednoznačně určené) $\frac{1}{a}$. Můžeme tak spárovat a s $\frac{1}{a}$ a v rámci součinu je znásobit na jedničku.

Párování a s $\frac{1}{a}$ se nám rozbije, pokud $a = \frac{1}{a}$ neboli $a^2 = 1$. To však značí $0 = a^2 - 1 = (a-1)(a+1)$, takže to může nastat jen pro $a = 1$ nebo $a = -1$. Všechny zbylé prvky se nám tudíž podaří spárovat, takže zbude $a_1 a_2 \cdots a_n = 1 \cdot (-1) = -1$. Kdyby platilo $1 = -1$ (například v $\mathbb{Z}_2$), zbude nám jen jeden nespárovaný prvek $1 = -1$, takže součin stále vyjde -1. □

Úloha 12. Je dáno $n > 1$. Urči, jaké $x \in \mathbb{Z}_n$ v závislosti na n splňuje $(n-1)! \equiv x \pmod n$.

---

[8] Leonhard Euler (1707–1783), švýcarský matematik, výrazně rozvinul takřka všechna odvětví tehdejší matematiky, některá další svou prací založil a k tomu také zavedl či ustálil mnoho prvků moderní matematické notace. Výsledky a koncepty po něm pojmenované lze v matematice potkat prakticky kdekoliv.

[9] John Wilson (1741–1793) byl anglický matematik. Jak už to tak bývá, Wilsonova věta byla známa již dlouho před ním.

Charakteristika*

V této sekci se podíváme na vlastnost tělesa naznačující, kterému $\mathbb{Z}_p$ (anebo $\mathbb{Q}$) se toto těleso tak nějak podobá.

Definice. *Charakteristika* tělesa T je nejmenší číslo c takové, že součet c jedniček v T je roven 0. Pokud takové c neexistuje, pak říkáme, že charakteristika je 0.

Cvičení 23. Rozmysli si, která známe tělesa s charakteristikou 0.

Tvrzení. *Každé konečné těleso má nenulovou charakteristiku.*

Důkaz. Stačí použít podobnou techniku jako v důkazu, že konečný obor integrity je těleso. Namísto mocnin si vezměme nekonečnou posloupnost

$$1, \qquad 1+1, \qquad 1+1+1, \qquad 1+1+1+1, \ldots$$

Jelikož máme konečné těleso, víme, že nějaké dva prvky musí být v této posloupnosti stejné. Tedy součet m jedniček je roven součtu n jedniček pro nějaká $m < n$. Potom je jejich rozdíl nula a zároveň součet $n-m$ jedniček, tedy $\underbrace{1+1+\cdots+1}_{(n-m)\text{-krát}} = 0$. Charakteristika tělesa tedy není 0. □

Cvičení(!) 24. Nechť má těleso T charakteristiku $c > 0$. Pak je součet n jedniček nulový, právě když $c \mid n$.

Tvrzení. *Každé těleso, které má nenulovou charakteristikou, ji má dokonce prvočíselnou.*

Důkaz. Nechť je pro spor charakteristika složené číslo $c = ab$, kde a i b jsou vetší než 1. Poté si vezměme prvky představující součty a a b jedniček. Vytknutím závorek potom dostáváme, že

$$0 = \underbrace{1+1+\ldots+1}_{c\text{-krát}} = \underbrace{(1+1+\ldots+1)}_{a\text{-krát}} \cdot \underbrace{(1+1+\ldots+1)}_{b\text{-krát}}.$$

Víme také, že každé těleso je obor integrity, a tudíž alespoň jedna ze závorek na pravé straně musí být nulová. To je ale spor s tím, že c je nejmenší počet jedniček, které se posčítají na nulu, protože a i b jsou obě menší. □

Charakteristika je velmi užitečná vlastnost, protože se dá ukázat, že konečné těleso s charakteristikou p má p^k prvků pro nějaké přirozené k. Lze také dokázat, že pro každé prvočíslo p a přirozené číslo k existuje konečné těleso s p^k prvky. Dá se dokonce dokázat i to, že struktura konečného tělesa je v jistém smyslu jednoznačně určena tím, kolik má prvků, takže všechna tělesa se stejným počtem prvků jsou „stejná“. Poslední dvě tvrzení však poněkud přesahují možnosti našeho seriálu, proto se omezíme na to první.

Tvrzení. *Nechť je prvočíslo p charakteristikou konečného tělesa T. Pak má T právě p^k prvků pro nějaké přirozené k.*

Důkaz. Budiž n počet prvků T. Dokážeme toto: když prvočíslo q dělí n, potom je $q = p$, tedy jiné prvočíslo než charakteristika nemůže dělit počet prvků. Z toho už vyplyne, že n musí být mocnina p, tedy p^k a zároveň $n > 1$, takže $k > 0$.

Mějme tedy nějaké prvočíslo $q \mid n$. Provedeme trik podobný počítání náhrdelníků v důkazu malé Fermatovy věty. Podíváme se, kolik existuje uspořádaných q-tic $(a_0, a_1, \ldots, a_{q-1})$ prvků T takových, že $a_0 + \cdots + a_{q-1} = 0$. Na jednu stranu hned nahlédneme, že jich je n^{q-1}. Vždy si totiž můžeme zcela libovolně zvolit čísla $a_1, \ldots, a_{q-1}$ a zbývající číslo a_0 je pak jednoznačně určeno jako $a_0 = -(a_1 + \cdots + a_{q-1})$. Při výběru každého jednotlivého a_i máme n možností a tyto možnosti jsou nezávislé, což dá celkem n^{q-1} způsobů, jak vybrat prvních $q-1$ čísel.

Nyní tyto q-tice začneme rotovat. Orotováním q-tice $(a_0, a_1, \ldots, a_{q-1})$ o j míst budeme myslet to, že z ní vyrobíme q-tici $(a_j, a_{1+j}, \ldots, a_{q-1+j})$, přičemž se na indexy díváme, jako by byly modulo

q, tedy a_q je totéž co a_0 atp. Takovýmto orotováním nezměníme to, že součet celé q-tice je nula. Rotováním o $j = 0, 1, \ldots, q-1$ tak vyrobíme q nějakých q-tic – budou však navzájem různé. Jsou-li některé dvě stejné, nechť je to BÚNO rotace o 0 a o nějaké $0 < j < q$. To potom znamená, že

$$a_0 = a_j = a_{2j} = a_{3j} = \cdots$$

Jenže q je prvočíslo, takže v posloupnosti $0, j, 2j, 3j \ldots$ se vyskytnou všechny zbytky modulo q.

Z toho plyne, že q-tice složené ze stejných čísel $(a, a, \ldots, a)$ jsou výjimečné. Kolik jich existuje? Pro všechny ostatní q-tice, které obsahují nějaké různé prvky, už máme všech q rotací odlišných, takže počet q-tic s různými prvky je násobek q. Zároveň je ale počet všech q-tic roven n^{q-1}, což je podle předpokladu $q \mid n$ také násobek q. Tím pádem musí i počet q-tic tvaru $(a, a, \ldots, a)$ být násobkem q. Tento počet však není nula, protože máme q-tici $(0, 0, \ldots, 0)$, jejíž součet určitě je 0. Aby tedy počet těchto q-tic byl násobek q, musí existovat ještě nějaká další q-tice $(a, a, \ldots, a)$ jejíž součet je 0 a zároveň $a \neq 0$. Potom však v rovnosti

$$\underbrace{a + a + \cdots + a}_{q\text{-krát}} = 0$$

stačí vynásobit prvkem $\frac{1}{a}$ a dostaneme

$$\underbrace{1 + 1 + \cdots + 1}_{q\text{-krát}} = 0.$$

To podle dřívějšího cvičení znamená $p \mid q$. Jenže p i q jsou prvočísla, takže $p = q$, což jsme přesně chtěli dokázat. □

Příklad. Pro prvočíslo p je $\mathbb{Z}_p$ konečné těleso s charakteristikou p a s p^1 prvky.

Příklad. Okruh $\mathbb{Z}[i]/(3)$ je konečné těleso s charakteristikou 3 a s $9 = 3^2$ prvky.

Příklad. Vyrobme si těleso s charakteristikou 2 a s $8 = 2^3$ prvky. Vezmeme okruh $\mathbb{Z}_2$ a přidáme k němu prvek α, kterému přiřkneme vlastnost $\alpha^3 = \alpha + 1$. Každý prvek okruhu $\mathbb{Z}_2[\alpha]$ potom budeme moci zapsat jako $a + b\alpha + c\alpha^2$ pro $a, b, c \in \mathbb{Z}_2$, a vždy když dva vynásobíme, tak výsledek bude opět možno vyjádřit pomocí 1, α a α^2, neboť každé α^3 přepíšeme na $\alpha+1$ (obdobně s vyššími mocninami). Celkově tak tento okruh bude mít osm prvků

$$0, \qquad 1, \qquad \alpha, \qquad 1+\alpha, \qquad \alpha^2, \qquad 1+\alpha^2, \qquad \alpha+\alpha^2, \qquad 1+\alpha+\alpha^2.$$

Že se jedná o těleso, není na první pohled zřejmé, ale je tomu tak, neboť každé $x \in \mathbb{Z}_2[\alpha]$ k sobě má y takové, že $xy = 1$. Například pro $x = 1 + \alpha^2$ vyhoví $y = \alpha$, neboť

$$(1 + \alpha^2) \cdot \alpha = \alpha + \alpha^3 = \alpha + (\alpha + 1) = (1 + 1)\alpha + 1 = 1.$$

Primitivní prvek

Podívejme se znovu a důkladněji na mocnění v konečných tělesech. Malá Fermatova věta nám dává jistotu, že v n-prvkovém tělese nám umocnění nenulového prvku na $n-1$ vyrobí jedničku. Nestačil by ale nějaký menší exponent?

Definice. *Řádem* nenulového prvku a rozumíme nejmenší přirozené číslo r takové, že $a^r = 1$.

Ekvivalentně nám řád říká, kolik různých prvků se vyskytne v posloupnosti $a^1, a^2, a^3, \ldots$

Tvrzení. *Nechť je r řád prvku a. Potom $a^n = 1$, právě když $r \mid n$.*

Důkaz. Když $r \mid n$, pak zjevně $a^n = 1^{\frac{n}{r}} = 1$. Pokud naopak $r \nmid n$, pak skrze dělení se zbytkem platí $n = rq + x$ pro nějaké $0 < x < r$. Následně platí $a^n = a^{x+rq} = a^x \cdot 1^q = a^x$, což nemůže být jedna, neboť to by byl vzhledem k $0 < x < r$ spor s definicí řádu. □

Cvičení(!) 25. Mějme řád r nenulového prvku a nějakého n-prvkového tělesa. Dokaž, že $r \mid n-1$.

Cvičení 26. Najdi nejmenší liché prvočíslo p, které dělí $89^8 + 1$.

Cvičení(!) 27. Budiž r řádem a. Potom když $k \mid r$, pak má a^k řád $\frac{r}{k}$.

Spousta prvků může mít řád menší než $n-1$. Zaměříme se na ty, které jej mají největší možný.

Definice. *Primitivním prvkem* tělesa s n prvky rozumíme takový prvek g, který má řád $n-1$.

Poznámka. Ekvivalentně můžeme říci, že v posloupnosti $g^1, g^2, \ldots, g^{n-1}$ vystupuje každý nenulový prvek tělesa právě jednou. Z tohoto důvodu se primitivnímu prvku též někdy říká *generátor*, a proto jej často značíme g.

Příklad. V tělese $\mathbb{Z}_7$ je 3 primitivním prvkem, neboť

$$3^1 \equiv 3, \qquad 3^2 \equiv 2, \qquad 3^3 \equiv 6, \qquad 3^4 \equiv 4, \qquad 3^5 \equiv 5, \qquad 3^6 \equiv 1.$$

Příklad. V tělese $\mathbb{Z}[i]/(3)$ je $1+i$ primitivním prvkem, neboť

$$(1+i)^1 \equiv 1+i, \qquad (1+i)^2 \equiv 2i, \qquad (1+i)^3 \equiv 1+2i, \qquad (1+i)^4 \equiv 2,$$
$$(1+i)^5 \equiv 2+2i, \qquad (1+i)^6 \equiv i, \qquad (1+i)^7 \equiv 2+i, \qquad (1+i)^8 \equiv 1.$$

Věta. *Každé konečné těleso obsahuje primitivní prvek.*

Důkaz je trochu techničtější, proto se k němu propracujeme postupně přes několik lemmat.

Lemma. (první) *Rovnice $x^n = 1$ má v konečném tělese nanejvýš n řešení.*

Ve třetím díle seriálu si dokážeme o něco obecnější tvrzení. Nyní si ukážeme alespoň náznak důkazu, proč by něco takového mělo platit.

Náznak důkazu. Převedeme 1 na levou stranu a budeme pracovat s $x^n - 1 = 0$. Myšlenka důkazu je, že kdykoliv bude a řešením rovnice, vytkneme na levé straně činitel $(x-a)$. Jak to provedeme? Když je a řešením rovnice, znamená to $a^n - 1 = 0$, takže můžeme levou stranu naší rovnice přepsat jako

$$x^n - 1 = (x^n - 1) - 0 = (x^n - 1) - (a^n - 1) = x^n - a^n = (x-a)(x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1}).$$

Tím jsme úspěšně vytknuli $(x-a)$. Co když bude nyní nějaké $b \neq a$ dalším řešením? Inu, pokud má při dosazení $x = b$ býti $(x-a)(x^{n-1} + \cdots + a^{n-1}) = 0$, pak musí alespoň jedna ze dvou závorek vyjít nula. Jenže díky $b - a \neq 0$ to nemůže být ta první, musí to být tedy ta druhá. To znamená, že b je řešením nové rovnice vzniklé z druhé závorky, ve které máme o jedna menší největší exponent u x. Dále opakujeme stejný trik: když je b řešením, upravíme

$$x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1} =$$
$$= (x^{n-1} + x^{n-2}a + \cdots + xa^{n-2} + a^{n-1}) - (b^{n-1} + b^{n-2}a + \cdots + ba^{n-2} + a^{n-1}).$$

Toto však stačí přeuspořádat tak, že dáme k sobě odpovídající mocniny x^k a b^k. Dostaneme tak součet výrazů tvaru $(x^k - b^k) \cdot$ něco. V každém z těchto členů dovedeme vytknout $x - b$, takže toto můžeme vytknout z výrazu jako celku. Tím původní rovnici přepíšeme do tvaru $(x-a)(x-b)$(nějaký výraz s x^{n-2}) $= 0$.

Takto pokračujeme dál a dál. V každém kroku vytkneme dvojčlen odpovídající jednomu řešení a snížíme tím největší exponent u x ve „zbývajícím výrazu“ o jedna. Tím pádem nebudeme moci

vytknout více než n členů: po n vytknutých závorkách by nám zbyla jenom nějaká konstanta, ze které už žádné $x - d$ vytknout nepůjde. □

Zbylá dvě lemmata se budou zabývat tím, jak získat prvek s řádem, který je násobkem řádů dalších prvků.

Lemma. (druhé) *Pokud mají a a b řády m a n, kde m, n jsou nesoudělná, pak má ab řád mn.*

Důkaz. Podívejme se na řád r prvku ab a dokažme, že je roven mn. Jelikož m a n jsou řády a a b, platí $ab^{mn} = (a^m)^n(b^n)^m = 1^n 1^m = 1$. Každý exponent, který v mocnině vyrobí jedničku, musí být násobek řádu, takže $r \mid mn$.

Nyní dokažme opačnou dělitelnost. Víme, že $1 = (ab)^r = a^r b^r$. Umocněním obou stran rovnosti na m dostáváme $1 = 1^m = a^{rm} b^{rm} = (a^m)^r b^{rm} = b^{rm}$. Řád b tak musí dělit rm, tedy $n \mid rm$. Jenže m je s n nesoudělné, takže v dělitelnosti nehraje roli, pročež $n \mid r$. Analogicky umocněním $1 = a^r b^r$ na n dostaneme $m \mid r$. Dohromady je r násobkem n i m, takže z jejich nesoudělnosti i $mn \mid r$. Spojením dělitelností $r \mid mn$ a $mn \mid r$ pak máme $r = mn$. □

V obecném případě se nám ale stane, že řády nemusí být nesoudělné. Proto si druhé lemma vylepšíme, abychom pro každá a, b byli schopni najít prvek, jehož řád je dělitelný řádem a i b.

Lemma. (třetí) *Pro každé dva prvky a a b s řády m a n umíme najít prvek, jehož řád je nejmenší společný násobek m a n.*

Důkaz. Nechť $\operatorname{nsn}(m,n)$ značí nejmenší společný násobek. Upravme si zadané prvky tak, abychom získali nesoudělné řády – pak budeme moci použít druhé lemma. Nejprve si z m, n vyrobme nějaké jejich dělitele m', n', kteří budou nesoudělní a přitom splní $m' \cdot n' = \operatorname{nsn}(m,n)$. Nechť jsou $p_1, \ldots, p_k$ všechna prvočísla, která dělí m nebo n. Dále nechť jsou $m = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ a $n = p_1^{\beta_1} \cdots p_k^{\beta_k}$ prvočíselné rozklady našich řádů. Pro každé $i = 1, \ldots, k$ se podíváme na mocniny p_i v rozkladech m a n. U toho řádu, který má tuto mocninu větší, ji ponecháme, zatímco tomu s menší mocninou ji celou sebereme[10] – a takto vytvoříme m', n'. Formálněji, nechť

$$\alpha_i' = \begin{cases} \alpha_i, & \text{pokud } \alpha_i \geq \beta_i, \\ 0, & \text{pokud } \alpha_i < \beta_i, \end{cases} \qquad \beta_i' = \begin{cases} 0, & \text{pokud } \alpha_i \geq \beta_i, \\ \beta_i, & \text{pokud } \alpha_i < \beta_i \end{cases}$$

a následně položme $m' = p_1^{\alpha_1'} \cdots p_k^{\alpha_k'}$ a $n' = p_1^{\beta_1'} \cdots p_k^{\beta_k'}$. Tím docílíme

$$m' \cdot n' = p_1^{\max(\alpha_1, \beta_1)} \cdots p_k^{\max(\alpha_k, \beta_k)} = \operatorname{nsn}(m,n),$$

neboť jsme zachovali největší mocninu každého p_i. Zároveň budou m' a n' nesoudělná, neboť nesdílejí žádná prvočísla ve svých prvočíselných rozkladech.

Nyní ještě vyrobme prvky a', b', které budou mít řády m', n'. Rozmysleme si, že stačí vzít $a' = a^{m/m'}$, $b' = b^{n/n'}$. Z konstrukce m' platí $m' \mid m$, takže m/m' je přirozené číslo, které je také dělitelem m. Pro získání a' tak jenom a mocníme na nějaký dělitel jeho řádu. Podle dřívějšího cvičení už víme, že řád výsledku bude jen původní řád vydělený exponentem, takže a' má řád m'. Obdobně má b' řád n'. V této situaci už můžeme použít druhé lemma, takže $a'b'$ má řád $m'n' = \operatorname{nsn}(m,n)$. □

Skloubením prvního a třetího lemmatu už nyní svedeme dokázat, že primitivní prvek existuje v každém tělese.

Důkaz věty. Nechť má uvažované těleso n prvků. Ty nenulové z nich si očíslujme $a_1, \ldots, a_{n-1}$ a nechť mají řády $r_1, \ldots, r_{n-1}$. Podle třetího lemmatu můžeme vzít kterékoliv dva prvky a nalézt prvek, jehož řád je násobkem obou řádů původních dvou prvků. Opakovaným použitím tohoto

[10] V případě rovnosti příslušnou mocninu ponecháme třeba v m'.

lemmatu (nejprve na první dva prvky, potom na výsledek a na třetí prvek atd.) tak dovedeme najít nějaký prvek g s řádem m, který je násobkem všech r_i.

Ukážeme $m = n-1$, což už bude znamenat, že g je primitivní prvek. Zaprvé m je řádem nějakého prvku tělesa, takže určitě $m \mid n-1$, z čehož $m \leq n-1$. Z druhé strany se podívejme na rovnici $x^m = 1$. Řád každého a_i dělí m, takže $a_i^m = 1$. To je $n-1$ různých řešení rovnice $x^m = 1$, takže podle prvního lemmatu nutně $n-1 \leq m$. Dohromady tak $m = n-1$, takže g je primitivní prvek. □

Poznámka. Dokázali jsme, že existuje alespoň jeden primitivní prvek, což ale neznamená, že existuje právě jeden – typicky jich máme více. Pro použití primitivního prvku nám to nijak nevadí, jde nám jenom o to, abychom dovedli nenulové prvky tělesa vypsat jako geometrickou posloupnost $g^1, g^2, \ldots, g^{n-1}$.

Využití primitivního prvku

Důkaz byl trochu pracnější, ale uvidíme, že primitivní prvek je opravdu silný nástroj. Tvrzení, která by jinak byla nejasná či pracná, se často stávají přímočaře nahlédnutelnými, když si nenulové prvky tělesa zapíšeme jako $g^1, g^2, \ldots, g^{n-1}$.

Tvrzení. (geometrická řada) *Pokud v tělese máme prvek $a \neq 1$, pak $1+a+a^2+\cdots+a^k = \frac{a^{k+1}-1}{a-1}$.*

Důkaz. Stačí obě strany vynásobit $a-1$, potom se na levé straně většina členů vyruší a zbude platná rovnost. □

Pokud řada začíná členem a, je třeba vytknout a, tedy $a + a^2 + \cdots + a^k = a \cdot \frac{a^k-1}{a-1}$.

Příklad. Nechť je p prvočíslo. Najdi všechna přirozená k, pro která p dělí $1^k + \cdots + (p-1)^k$.

Řešení. Nechť je g primitivní prvek v $\mathbb{Z}_p$. Čísla $1, \ldots, p-1$ můžeme (v nějakém jiném pořadí) zapsat jako $g^1, g^2, \ldots, g^{p-1}$. Uvažovaný součet se tím (modulo p) upraví na

$$g^k + g^{2k} + g^{3k} + \cdots + g^{(p-1)k}.$$

Pokud $g^k \not\equiv 1$, pak toto vzorcem pro součet geometrické řady vyjde

$$g^k \cdot \frac{g^{(p-1)k}-1}{g^k-1} \equiv g^k \cdot \frac{1^k-1}{g^k-1} \equiv 0 \pmod{p}.$$

Naopak pokud $g^k \equiv 1$, pak máme v součtu $p-1$ jedniček, takže dostaneme $p-1 \equiv -1 \not\equiv 0$. Zbývá tedy jenom rozhodnout, kdy $g^k \equiv 1$. To však z definice primitivního prvku nastává, právě když $p-1 \mid k$. Zadaný součet je tak násobkem p, právě když $p-1 \nmid k$.

Úloha 13. Rozhodni, zda lze tabulku 10×10 vyplnit čísly $1, 2, \ldots, 100$ a zvolit $A, B \in \mathbb{Z}_{101}$ tak, aby platilo:

(i) Součin prvků libovolného řádku dává po dělení 101 zbytek A.

(ii) Součet prvků libovolného sloupce dává po dělení 101 zbytek B.

Cvičení 28. Dokaž Wilsonovu větu pomocí primitivního prvku pro těleso s lichým počtem prvků.

Definice. Nechť je T těleso. O množině $S \subseteq T$ řekneme, že je *multiplikativní*, pokud je neprázdná, $0 \notin S$ a zároveň je S uzavřená na násobení, tedy kdykoliv $a, b \in S$, pak i $ab \in S$.

Příklad. V tělese $\mathbb{Z}_7$ máme multiplikativní množiny $\{1\}$, $\{1, 6\}$, $\{1, 2, 4\}$ a $\{1, 2, 3, 4, 5, 6\}$.

Tvrzení. *Nechť je T konečné n-prvkové těleso s primitivním prvkem g. Potom jde jeho každou multiplikativní podmnožinu S popsat pomocí vhodného $m \mid n-1$ jako*

$$S = \{g^m, g^{2m}, g^{3m}, \ldots, g^{(n-2)m}, g^{(n-1)m}\}.$$

Posléze pro nenulové $a \in T$ platí $a \in S$, právě pokud $a^{\frac{n-1}{m}} = 1$, a množina S tak má $\frac{n-1}{m}$ prvků.

Důkaz. Vypišme si všechny nenulové prvky T v pořadí $g^1, g^2, g^3, \ldots, g^{n-1}$. Nechť je m nejmenší exponent, pro který $g^m \in S$. Postupným násobením g^m dovedeme vyrobit i $g^{2m}, g^{3m}, \ldots$, takže z definice multiplikativní množiny musí i tyto prvky ležet v S. Dále také platí $g^{-1} = g^{n-2}$, jelikož $g \cdot g^{n-2} = g^{n-1} = 1$, takže i $g^{-m} = g^{(n-2)m} \in S$. Z toho už $g^{xm} \in S$ pro libovolné celé číslo x.

Ukažme, že jiné prvky než mocniny g^m už v S neleží. Uvažujme nějaké $g^k \in S$ a dokažme $m \mid k$. Nechť je d největší společný dělitel m a k. Pak existují Bézoutovy koeficienty x, y takové, že $xm + yk = d$. BÚNO nechť $y > 0$ (vždycky můžeme x snížit o $n-1$ a y zvýšit o m). Potom máme

$$g^d = g^{xm+yk} = g^{xm} \cdot \left(g^k\right)^y.$$

Přitom g^{xm} je prvek S a g^k je prvek S, takže na pravé straně máme jenom součin několika prvků S, což znamená i $g^d \in S$. Jenže m má být nejmenší exponent s touto vlastností, takže $d \geq m$. Ale $d \mid m$, takže $d \leq m$. Proto už nutně $d = m$, takže m je dělitel k, jak jsme chtěli.

Speciálně i $(g^m)^{n-1} = 1 = g^{n-1}$ je prvkem S, takže i $m \mid n-1$, jak jsme chtěli. Zbývá tak dokázat, že $a \in S$, právě pokud $a^{\frac{n-1}{m}} = 1$. Zapišme $a = g^b$ pro nějaké b, pak je rovnost $a^{\frac{n-1}{m}} = 1$ ekvivalentní $g^{\frac{b(n-1)}{m}} = 1$. Jenže g má řád $n-1$, takže toto je ekvivalentní $n-1 \mid \frac{b(n-1)}{m}$ neboli $\frac{b}{m} \in \mathbb{Z}$ neboli $m \mid b$. To ale přesně odpovídá tomu, že $a = g^b = g^{xm} \in S$.

Konečně díky tomu, že každá m-tá mocnina g leží v S a v celém T máme $n-1$ prvků $g^1, g^2, \ldots, g^{n-1}$, už musí S mít $\frac{n-1}{m}$ prvků. □

Cvičení 29. Nechť je S multiplikativní množina v konečném tělese. Potom pokud má S více než jeden prvek, pak je součet všech prvků S roven 0.

Cvičení(!) 30. Mějme n-prvkové těleso T a přirozené číslo k. Zobrazení $f(x) = x^k$ je na T prosté[11], právě když je k nesoudělné s $n-1$.

Úloha 14. (těžší) Posloupnost $\{a_n\}$ je definována předpisem $a_1 = 1$ a pro další členy $a_{k+1} = a_k^3 + 1$. Dokaž, že pro každé prvočíslo p tvaru $3\ell + 2$ (pro $\ell \in \mathbb{N}$) je nějaké a_n násobkem p.

Kvadratické zbytky

Na závěr se podívejme na velmi konkrétní příklad multiplikativní množiny prvků, které se dají zapsat jako druhé mocniny. Během seriálu jsme již několikrát využili, že výraz x^2 většinou nemůže modulo nějaké pevně zvolené n nabývat úplně libovolných hodnot: například modulo 4 dávají čtverce celých čísel jenom zbytky 0 a 1. Podívejme se, jak to vypadá v konečných tělesech.

Definice. O prvku a konečného tělesa T říkáme, že je to *kvadratický zbytek*, pokud pro něj existuje prvek x splňující $x^2 = a$. V opačném případě říkáme, že a je *kvadratický nezbytek*.

Pozorování. *Nenulové kvadratické zbytky tvoří v tělese T multiplikativní množinu.*

Důkaz. Nechť je S množina nenulových kvadratických zbytků. Z $a, b \in S$ plyne $a = x^2$, $b = y^2$ pro nějaká $x, y \in T$, takže $ab = (xy)^2$. Zároveň z nenulovosti a, b je i ab nenulové, takže $ab \in S$. □

Důsledek. (Eulerovo kritérium) *Mějme konečné těleso T s n prvky, kde n je liché. Potom pro nenulový prvek a platí*

$$a^{\frac{n-1}{2}} = \begin{cases} 1, & \text{je-li } a \text{ kvadratický zbytek,} \\ -1, & \text{je-li } a \text{ kvadratickým nezbytkem.} \end{cases}$$

Důkaz. Nechť je g primitivní prvek v T. Pro jaké m tvoří multiplikativní množinu kvadratických zbytků právě $g^m, g^{2m}, g^{3m}, \ldots$? Pro $m = 2$ budou všechny exponenty sudé, takže dostaneme samé

[11] Zobrazení f je *prosté*, zobrazuje-li různé vzory na různé obrazy, tedy $x \neq y \implies f(x) \neq f(y)$.

kvadratické zbytky. Naopak když je $a = x^2$ nenulový kvadratický zbytek, potom $x = g^k$ pro nějaké k, takže $a = g^{2k}$. Zároveň z předpokladu $2 \mid n-1$, takže vše funguje a nenulové kvadratické zbytky jsou přesně $g^2, g^4, g^6 \ldots$

Z tvrzení o multiplikativních množinách to znamená, že $a \neq 0$ je kvadratickým zbytkem, právě když $a^{\frac{n-1}{2}} = 1$. Naopak když je a kvadratickým nezbytkem, tak $a^{\frac{n-1}{2}}$ není 1, ale stále platí $a^{n-1} = 1$, což upravíme na $\left(a^{\frac{n-1}{2}} - 1\right)\left(a^{\frac{n-1}{2}} + 1\right) = 0$. První závorka není nulová, takže je nulová ta druhá, tedy $a^{\frac{n-1}{2}} = -1$. □

Obvykle se výrazu $a^{\frac{n-1}{2}}$ pro prvek a v n-prvkovém tělese T říká *Legendreův*[12] *symbol* a značí se $\left(\frac{a}{T}\right)$. Pro $T = \mathbb{Z}_p$ se ujalo jednoduší značení $\left(\frac{a}{p}\right)$. Ukažme si nyní, jaké úlohy se tímto dají vyřešit.

Cvičení(!) 31. Pro liché n je v n-prvkovém tělese přesně $\frac{n+1}{2}$ kvadratických zbytků (včetně 0).

Úloha 15. Nechť je T konečné těleso s lichým počtem prvků. Pak pro každé $a \in T$ existují $x, y \in T$ taková, že $a = x^2 + y^2$.

Cvičení 32. Pokud je n liché a větší než 3, pak je součet všech kvadratických zbytků v n-prvkovém tělese roven 0.

Cvičení 33. Modulo liché prvočíslo p je -1 kvadratický zbytek, právě když $p \equiv 1 \pmod 4$.

Cvičení(!) 34. Rozmysli si, že Legendreův symbol je multiplikativní, tedy $\left(\frac{ab}{T}\right) = \left(\frac{a}{T}\right) \cdot \left(\frac{b}{T}\right)$.

Příklad. Mějme liché prvočíslo p. Dokaž, že existuje celé číslo x splňující $p \mid x^2 - x + 3$, právě pokud existuje celé číslo y splňující $p \mid y^2 - y + 25$.

Řešení. Podívejme se na obě dělitelnosti jako na kongruence v $\mathbb{Z}_p$. Potom $0 \equiv x^2 - x + 3$, což je po přenásobení 4 ekvivalentní $0 \equiv 4x^2 - 4x + 12 \equiv (2x-1)^2 + 11$ neboli $-11 \equiv (2x-1)^2$. Když už najdeme $a \in \mathbb{Z}_p$ tak, aby $a^2 \equiv -11$, tak z lichosti p dopočteme $x \equiv \frac{a+1}{2}$, takže x splňující zadanou dělitelnost existuje, právě když je -11 kvadratický zbytek. Obdobně můžeme upravit druhou dělitelnost, abychom zde znovu dostali čtverec. Tedy $0 \equiv 4y^2 - 4y + 100 \equiv (2y-1)^2 + 99$, takže existence takového y je ekvivalentní tomu, že -99 je kvadratickým zbytkem.

Použijme nyní Legendreovy symboly. Pokud $p = 11$, tak jsou -11 i -99 nuly v $\mathbb{Z}_p$, tedy oba kvadratické zbytky. Obdobně když $p = 3$, tak jsou $-11 \equiv 1$ i $-99 \equiv 0$ kvadratické zbytky. Konečně když p není 11 ani 3, jsou -11 i -99 nenulové v $\mathbb{Z}_p$, takže z multiplikativity Legendreova symbolu $\left(\frac{-99}{p}\right) = \left(\frac{-11}{p}\right)\left(\frac{3^2}{p}\right) = \left(\frac{-11}{p}\right)$, neboť 3^2 je čtverec. Tím je příklad vyřešen.

Příklad. Rozhodni, zda má rovnice $x^2 = y^5 + 7$ celočíselné řešení.

Řešení. Nemá. Podívejme se na ni modulo 11. Tím se z y^5 stane Legendreův symbol $\left(\frac{y}{11}\right)$, takže bude nabývat jen hodnot 0 (když $y \equiv 0$) nebo ± 1. Pravá strana tak bude moci nabývat jen hodnot 6, 7 a 8. Snadno však ověříme, že ani jedno z těchto čísel není kvadratický zbytek v $\mathbb{Z}_{11}$, takže $x^2 \equiv y^5 + 7 \pmod{11}$ nemá řešení.

Úloha 16. Nechť p je prvočíslo tvaru $4k - 1$. Nyní pokud modulo p platí $a \equiv x^2$, pak $x \equiv \pm a^k$.

Úloha 17. Nechť je p prvočíslo tvaru $2^k + 1$. Potom je $g \in \mathbb{Z}_p$ primitivním prvkem, právě když je kvadratickým nezbytkem.

[12] Adrien-Marie Legendre (1752–1833), francouzský matematik.

Závěr

Zde končí druhý díl seriálu. Víme již, co dělat s Pellovou rovnicí i jak ji nalézt skrytou v úlohách, nezalekneme se modulení, ať už pracujeme v jakémkoli okruhu, a vidíme dovnitř struktury konečných těles pomocí primitivního prvku. Děkujeme Ti, že ses dočetl(a) až sem. V třetím díle budeme k okruhům místo nějakého konkrétního čísla přidávat úplně obecnou proměnnou. Tím získáme okruhy polynomů, u nichž si opět prozkoumáme, jak je to s dělením se zbytkem či s jednoznačným rozkladem a jak souvisí vlastnosti okruhu polynomů s vlastnostmi původního okruhu.

Mezitím se těšíme na viděnou a přejeme mnoho zdaru při řešení úloh druhé seriálové série!

Návody ke cvičením

1. Kdy je $\frac{\sqrt{d_1}}{\sqrt{d_2}}$ racionální? Až budeš chtít ukázat, že jiné dvojice to být nemohou, zkus ve správnou chvíli umocnit.

2. Porovnej jejich prvky v $\mathbb{Q}(\sqrt{d})$.

3. Stačí prostě všechno rozepsat.

4. Kdyby něco nenulového mělo normu 0, pak by $\sqrt{d}$ bylo racionální.

5. Rozšiř pomocí $\overline{\beta}$.

7. Postupuj stejně jako pro $\mathbb{Z}[\sqrt{2}]$, jen lépe odhadni $|x^2 - 3y^2|$ pro $|x|, |y| \leq \frac{1}{2}$. Trojúhelníková nerovnost nestačí, protože si nemůžeme dovolit rovnost!

8. Prostě to ověř, stačí se dívat jenom na samotné $\frac{1+\sqrt{d}}{2}$.

9. Rozlož na součin v $\mathbb{Z}$.

11. Uprav na čtverce.

12. Uprav na Pellovu rovnici s $d = 8$.

13. Postupuj jako v řešeném příkladu. Pokud Ti nepůjde nalézt fundamentální řešení pro $d = 28$, začni u $d = 7$.

14. Vezmi kladnou Pellovu rovnici a její fundamentální řešení. Uprav ji tak, aby šlo využít nesoudělné činitele a specifickou podobu prvočíselného rozkladu.

15. Každá jednotka je $\pm$ mocninou té fundamentální. Rozliš, zda tato mocnina má sudý či lichý exponent.

16. Speciální případ předchozího cvičení.

17. Rozepiš podle definice kongruence. Využij toho, že celé číslo m dělí $c_1 + c_2\sqrt{d}$, právě když dělí obě složky c_1, c_2.

18. Jednotka dělí cokoliv.

19. Krácení je jenom speciální případ dělení.

20. Najdi dva nenulové prvky, jejichž součin je nulový.

21. Použij malou Fermatovu větu.

22. Použij malou Fermatovu větu v tělese $\mathbb{Z}[i]/(3)$. Neroznásobuj binomickou větou.

24. Ze součtu n jedniček odebírej po c jedničkách.

25. Slož dohromady předchozí tvrzení a malou Fermatovu větu.

26. Rozmysli si, že 89 musí mít v $\mathbb{Z}_p$ řád 16.

27. Každá mocnina a^k je i mocninou a.

28. Zapiš prvky tělesa pomocí primitivního prvku.

29. Posčítej geometrickou řadu.

30. Zapiš prvky tělesa pomocí primitivního prvku. Exponenty jsou modulo $n-1$.

31. Použij tvrzení o multiplikativních množinách.

32. Kvadratické zbytky jsou multiplikativní množina – sečti geometrickou řadu. Toto cvičení je jen speciální případ jednoho předchozího.

33. Eulerovo kritérium.

34. Rozepiš si z definice Legendreova symbolu.

Návody k úlohám

1. Opět vyděl $\frac{\alpha}{\beta} = x + y\sqrt{7}$ a snaž se od x a y odečítat celá čísla tak, aby absolutní hodnota normy spadla pod 1. BÚNO stačí vyřešit $x, y \in \langle 0, \frac{1}{2}\rangle$. Funguje $\gamma = 1$ nebo $\gamma = -1 + \sqrt{7}$. Bacha na to, kde nastává rovnost.

2. Pokud se na rovnici budeš snažit použít kvadratické zbytky, neuspěješ. Radši využij fundamentální řešení kladné Pellovy rovnice, které je v tomto případě $35 + 6\sqrt{34}$

3. Uprav zadání do tvaru Pellovy rovnice a rozmysli, které exponenty dávají zpět platná řešení. Na výrazu, který dostaneš, nemusí být na první pohled vidět, že jde upravit na čtverec, ale skutečně jde. Hodí se $2 \cdot (2 + \sqrt{3}) = (1 + \sqrt{3})^2$.

4. Zvol si nějaké d, které není čtverec, a vztah $b^2 + 1 = d \cdot a(a+1)$ uprav vhodnou substucí na rovnici Pellova typu $N(x + y\sqrt{d}) =$ konstanta. Pokud najdeš jedno řešení, už jich máš nekonečně mnoho – stačí přenásobovat řešeními Pellovy rovnice. Dobře funguje třeba $d = 5$.

5. Využij $\mathbb{Z}[\sqrt{q^2 d}] \subset \mathbb{Z}[\sqrt{d}]$.

6. Využij čtyřku v $7 + 4\sqrt{3} = (2 + \sqrt{3})^2$. Je potřeba důsledně najít všechny činitele dvojky, které se nabízí. Pokud znáš kvadratické zbytky modulo 8, pak může být snazší prostě použít je.

7. Pokud $a^2 = 2n+1$, $b^2 = 3n+1$, pak substituce $a = x+3y$, $b = x+2y$ povede k Pellově rovnici s $d = 6$.

8. Nejprve si rozmysli, že toto platí tehdy, když jsou pro $\alpha = a + b\sqrt{d}$ složky a, b nesoudělné.

9. Použij malou Fermatovu větu.

10. Je to jenom $x = 1$. Hodí se $\frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 1$.

11. 101 je prvočíslo. Pro $y > 0$ získej kongruenci $x^5 \equiv -2 \pmod{101}$, umocni na 20 a použij malou Fermatovu větu.

12. Když je n prvočíslo, použij Wilsonovu větu. Pro složené n si rozmysli $x \equiv 0 \pmod{n}$.

13. Jde to. 101 je prvočíslo.

14. Podívej se na předpis posloupnosti jako na zobrazení v $\mathbb{Z}_p$. Všimni si, že 0 se posílá na 1.

15. Pro spor předpokládej, že rovnice nemá řešení, a najdi příliš mnoho kvadratických nezbytků.

16. Využij, že a je kvadratický zbytek, a připomeň si, co pro něj díky Legendreovu symbolu platí.

17. Když prvek není primitivní, jaký může mít řád? Porovnej s Eulerovým kritériem.

Řesení cvičení

1. Jsou to přesně ty dvojice, kde $\frac{d_1}{d_2} = q^2$ pro nějaké racionální q. Potom určitě $\sqrt{d_1} = q\sqrt{d_2} \in \mathbb{Q}(\sqrt{d_2})$ a obdobně $\sqrt{d_2} \in \mathbb{Q}(\sqrt{d_1})$, což značí $\mathbb{Q}(\sqrt{d_1}) = \mathbb{Q}(\sqrt{d_2})$. Naopak pokud $\mathbb{Q}(\sqrt{d_1}) = \mathbb{Q}(\sqrt{d_2})$, pak $\sqrt{d_1} = a + b\sqrt{d_2} \in \mathbb{Q}(\sqrt{d_2})$. Pokud $b = 0$, pak je $\sqrt{d_1}$ racionální, což je spor. Pro $b \neq 0$ předchozí vztah upravíme na

$$\begin{aligned} a &= \sqrt{d_1} - b\sqrt{d_2}, \\ a^2 &= d_1 - 2b\sqrt{d_1 d_2} + b^2 d_2, \\ \frac{a^2 - d_2 b^2 - d_1}{2b} &= \sqrt{d_1 d_2}, \end{aligned}$$

takže $\sqrt{d_1 d_2}$ je racionální. Pak je i $q = \sqrt{\frac{d_1}{d_2}} = \frac{\sqrt{d_1 d_2}}{d_2}$ racionální, neboli $\frac{d_1}{d_2} = q^2$.

2. Každý prvek průniku $\mathbb{Q} \cap \mathbb{Z}[\sqrt{d}]$ zároveň vyjádříme jako $q \in \mathbb{Q}$ a jako $a + b\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$. Když to formulujeme jako rovnost $q + 0 \cdot \sqrt{d} = a + b\sqrt{d}$ dvou prvků $\mathbb{Q}(\sqrt{d})$, pak podle předchozího tvrzení $q = a$ a $0 = b$. První rovnost ale znamená, že q je celé číslo, takže každý racionální prvek $\mathbb{Z}[\sqrt{d}]$ je celé číslo.

3. Ověříme takřka stejně jako pro komplexní čísla v prvním díle. Nechť $\alpha = a + b\sqrt{d}$, $\beta = x + y\sqrt{d}$. V případě sčítání je $\overline{(\alpha + \beta)} = \overline{\alpha} + \overline{\beta}$ zřejmé. V případě násobení prostě dosadíme a obdržíme

$$\begin{aligned} \alpha\beta &= (a + b\sqrt{d})(x + y\sqrt{d}) = (ax + byd) + (ay + bx)\sqrt{d}, \\ \overline{\alpha} \cdot \overline{\beta} &= (a - b\sqrt{d})(x - y\sqrt{d}) = (ax + byd) - (ay + bx)\sqrt{d}, \end{aligned}$$

takže $\overline{(\alpha\beta)} = \overline{\alpha} \cdot \overline{\beta}$. Z toho následně $N(\alpha\beta) = \alpha\beta \cdot \overline{\alpha}\,\overline{\beta} = \alpha\overline{\alpha} \cdot \beta\overline{\beta} = N(\alpha) \cdot N(\beta)$.

4. Nechť $\alpha = a + b\sqrt{d}$. Rovnost $N(\alpha) = 0$ upravíme na $a^2 = b^2 \cdot d$. Kdyby nyní $b \neq 0$, mohli bychom vydělit a dostali bychom $\sqrt{d} = \frac{a}{b}$, což je spor. Určitě tedy $b = 0$, takže i $a = 0$ a dohromady $\alpha = 0$.

5. Nechť $\alpha = a + b\sqrt{d}$, $\beta = x + y\sqrt{d}$. Pak

$$\frac{\alpha}{\beta} = \frac{\alpha\overline{\beta}}{\beta\overline{\beta}} = \frac{(a + b\sqrt{d})(x - y\sqrt{d})}{x^2 - dy^2} = \frac{ax - byd}{x^2 - dy^2} + \frac{bx - ay}{x^2 - dy^2}\sqrt{d}.$$

Jelikož $\beta \neq 0$, tak určitě $N(\beta) \neq 0$ čili nedělíme nulou a vše je v pořádku.

6. Pokud je α racionální, pak je zřejmě samo sobě sdružené. Naopak pokud $\alpha = \overline{\alpha}$, pak je jeho racionální složka rovna $\frac{\alpha + \overline{\alpha}}{2} = \alpha$, takže α je racionální.

7. Stejně jako v příkladu $\mathbb{Z}[\sqrt{2}]$ vydělíme a zaokrouhlíme. Potom pro $a = x - x_0$, $b = y - y_0$ chceme dokázat $|a^2 - 3b^2| < 1$, přičemž máme zaručeno $|a|, |b| \leq \frac{1}{2}$. Kdybychom jenom použili trojúhelníkovou nerovnost, dostaneme $|a^2 - 3b^2| \leq \frac{1}{4} + 3 \cdot \frac{1}{4} = 1$, což nám ale nestačí, protože potřebujeme ostrou nerovnost. Využijeme toho, že a^2 i b^2 jsou nezáporná čísla, takže díky omezení

absolutní hodnoty určitě obě leží v intervalu $\langle 0, \frac{1}{4}\rangle$. Výraz $a^2 - 3b^2$ (bez absolutních hodnot) tak bude nejmenší možný, když a^2 bude nejmenší možné a b^2 největší možné, takže $0 - 3 \cdot \frac{1}{4} = -\frac{3}{4}$. Naopak největší hodnoty výrazu dosáhneme při $a^2 = \frac{1}{4}$, $b^2 = 0$, tedy $\frac{1}{4} - 3 \cdot 0 = \frac{1}{4}$. Dohromady $a^2 - 3b^2 \in \langle -\frac{3}{4}, \frac{1}{4}\rangle$, z čehož určitě $|a^2 - 3b^2| \leq \frac{3}{4} < 1$.

8. Uzavřenost na sčítání je jasná. Označme $\alpha = \frac{1+\sqrt{d}}{2}$, potom $\alpha^2 = \frac{1+2\sqrt{d}+d}{4} = \alpha + \frac{d-1}{4} \in \mathbb{Z}[\alpha]$, jelikož $\frac{d-1}{4} \in \mathbb{Z}$. Když pak roznásobíme $(a_1 + b_1\alpha) \cdot (a_2 + b_2\alpha)$, tak už bude každý člen výsledného součtu prvkem $\mathbb{Z}[\alpha]$.

9. Když $d = a^2$, pak rozložíme na součin $x^2 - a^2y^2 = (x - ay)(x + ay) = 1$. Nyní je součin dvou celých čísel roven 1, takže jsou buď obě 1, nebo obě -1. Z první možnosti dostaneme $2x = (x-ay)+(x+ay) = 1+1 = 2$, takže $x = 1$. Obdobně z druhé možnosti $x = -1$ a v obou případech pak nutně $a^2y^2 = 0$, takže $y = 0$. Takže kdyby byl d čtverec, Pellova rovnice by měla jen triviální řešení.

10. (i) $a + \sqrt{d}$, (ii) $(a^2 - 1) + a\sqrt{d}$, $(a^2 + 1) + a\sqrt{d}$,

(iii) Díky $a^2 - 1^2 \cdot (a^2 + 1) = -1$ je $a + \sqrt{d}$ jednotka s normou -1. Její čtverec pak bude jednotka s normou 1, tedy máme řešení $(2a^2 + 1) + 2a\sqrt{d}$.

11. Rovnici upravíme na $(x-2y)^2-3y^2 = 1$, takže je to Pellova rovnice v $\mathbb{Z}[\sqrt{3}]$. Fundamentálním řešením je $\omega = 2 + \sqrt{3}$. Z její mocniny ω^n už dopočteme $y = \frac{\omega^n - \omega^{-n}}{2\sqrt{3}}$ a $x = \frac{\omega^{n+1} - \omega^{-n-1}}{2\sqrt{3}}$. Další řešení bychom dostali obrácením znamének u x i y zároveň.

12. Nechť $\frac{n(n+1)}{2} = y^2$. To upravíme na $4n^2 + 4n = 8y^2$ a následně $(2n + 1)^2 - 8y^2 = 1$. Pojmenujeme-li $x = 2n + 1$, pak pro každé řešení $x + y\sqrt{8}$ musí x být liché, takže nám z něj $n = \frac{x-1}{2}$ dá platné řešení. Stačí tedy vzít třetí nejmenší řešení $x + y\sqrt{8}$, to jest třetí mocninu fundamentálního řešení. Tím je $3 + \sqrt{8}$, takže $x + y\sqrt{8} = (3 + \sqrt{8})^3 = 99 + 35\sqrt{8}$, z čehož $n = 49$.

13. Stačí postupovat podobně jako v řešeném příkladu. Najít fundamentální řešení pro $d = 28$ může být trochu problém, můžeme se však podívat na $d = 7$. Tam je fundamentálním řešením $\omega_0 = 8 + 3\sqrt{7}$. Jeho iracionální složka je lichá, takže to vzhledem $\sqrt{28} = 2\sqrt{7}$ není prvek $\mathbb{Z}[\sqrt{28}]$. Zato $\omega_0^2 = 127 + 48\sqrt{7} = 127 + 24\sqrt{28} = \omega$ už je, takže je to fundamentální řešení pro $d = 28$. Potom už postupujeme úplně stejně jako v řešeném příkladu.

14. Nechť je $x + y\sqrt{p}$ fundamentální řešení Pellovy rovnice $x^2 - py^2 = 1$. Upravme ji do tvaru $x^2 - 1 = py^2$. Podle toho, zda je x sudé či liché, může na levé straně modulo čtyři být buďto -1, či 0. Na pravé straně pak podle parity y může díky $p \equiv 1 \pmod 4$ být buďto 0, nebo 1. Aby tedy mohla nastat rovnost, určitě je x liché a y sudé.

V upravené Pellově rovnici tak můžeme vydělit čtyřmi, čímž dostaneme $\frac{x+1}{2} \cdot \frac{x-1}{2} = p\left(\frac{y}{2}\right)^2$. Na levé straně jsou nesoudělná celá čísla $\frac{x+1}{2}$, $\frac{x-1}{2}$ (jejich rozdíl je 1), zatímco na pravé straně máme číslo, v jehož prvočíselném rozkladu mají všechna prvočísla vyjma p sudý exponent. Jelikož činitelé nalevo jsou nesoudělní, musí jeden z nich být a^2 a druhý pb^2 pro nějaká $a, b \in \mathbb{N}$ (nemůže se zde vyskytnout nula, jelikož pak by i y bylo nulové). Kdyby platilo $\frac{x+1}{2} = a^2$ a $\frac{x-1}{2} = pb^2$, pak dostaneme $a^2 - pb^2 = 1$. To by bylo opět řešení Pellovy rovnice, jenže by muselo být menší než $x + y\sqrt{p}$, které je fundamentální. To by byl spor, takže musí být $\frac{x+1}{2} = pb^2$ a $\frac{x-1}{2} = a^2$, což znamená $a^2 - pb^2 = -1$. Záporná Pellova rovnice tedy má řešení, jak jsme chtěli.

15. Víme, jak vyrobit všechna řešení z toho fundamentálního, takže $x + y\sqrt{d} = \pm(x_0 + y_0\sqrt{d})^n$ pro nějaké $n \in \mathbb{Z}$. Když obrátíme znaménko $\pm$ nebo znaménko exponentu n, jenom tím obrátíme znaménko jedné nebo obou složek x, y, což na dělitelnost $x_0y_0 \mid xy$ nebude mít vliv. BÚNO tedy předpokládejme $x + y\sqrt{d} = (x_0 + y_0\sqrt{d})^n$ pro přirozené n.

Nyní když je n liché, tak podle předchozího tvrzení jak x_0, tak y_0 dělí y, z čehož určitě $x_0y_0 \mid xy$. Když je naopak n sudé, tak stále y_0 dělí y, ale x_0 tentokrát dělí x, což však stále implikuje $x_0y_0 \mid xy$. Dohromady tak $x_0y_0 \mid xy$ platí vždy.

Pro fundamentální jednotku a obecnou jednotku bychom postupovali stejně: důležité je jenom to, že máme (třeba s pomocí nějakého BÚNO) $x + y\sqrt{d} = (x_0 + y_0\sqrt{d})^n$ pro $n \in \mathbb{N}$.

16. Stačí nalézt fundamentální řešení $73 + 12\sqrt{37}$, načež díky $876 = 73 \cdot 12$ stačí vzít předchozí obecné cvičení. Jak na toto řešení přijít? Třeba díky $37 = 6^2 + 1$ máme jednotku $6 + \sqrt{37}$ s normou -1, která už musí být fundamentální, jelikož žádná jiná jednotka nemůže mít menší iracionální část. Fundamentální řešení Pellovy rovnice je pak čtvercem fundamentální jednotky.

17. Když znění cvičení rozepíšeme podle definice kongruence, pak chceme říci, že $m \mid a - b$ neimplikuje $m \mid \overline{a} - \overline{b} = \overline{(a-b)}$. Naproti tomu $\overline{m} \mid \overline{(a-b)}$ by z $m \mid a - b$ plynulo, ale tím, že provedeme sdružení jen na pravé straně dělitelnosti, nemusíme dostat pravdivé tvrzení: například v $\mathbb{Z}[i]$ máme $2 + i \mid 2 + i$, ale $2 + i \nmid 2 - i$.

Nyní nechť je však $m \in \mathbb{Z}$ a budiž $a - b = c_1 + c_2\sqrt{d}$ pro nějaká $c_1, c_2 \in \mathbb{Z}$. Potom je $m \mid c_1 + c_2\sqrt{d}$ ekvivalentní $m \mid c_1$ a zároveň $m \mid c_2$. Ale $m \mid \overline{(a-b)} = c_1 - c_2\sqrt{d}$ je totéž, jelikož $m \mid c_2$ a $m \mid (-c_2)$ je ekvivalentní.

18. Jednotka u dělí 1 a 1 dělí každý prvek R, takže i u dělí všechno. Potom ale pro každé $a \in R$ máme $u \mid a - 0$, takže $a \equiv 0 \pmod{u}$. Každý prvek je tak kongruentní nule, takže máme jedinou zbytkovou třídu, v níž leží každý prvek okruhu.

19. Nechť je T těleso a mějme $a, b, c \in T$, kde c je nenulové. Víme, že c je jednotka, takže existuje $\frac{1}{c}$. Potom pokud $ac = bc$, pak přenásobíme dostaneme $a = a \cdot 1 = a \cdot \left(c \cdot \frac{1}{c}\right) = bc \cdot \frac{1}{c} = b \cdot 1 = b$, čímž je hotovo.

20. Z definice toho, že n je složené, existuje jeho rozklad $n = ab$, kde a i b jsou větší než 1 a menší než n. Z toho už nutně plyne, že $a, b \not\equiv 0 \pmod{n}$ a zároveň $pq \equiv n \equiv 0 \pmod{n}$.

21. Základ mocniny můžeme zmenšit modulo 11 pomocí $29 \equiv 7 \pmod{11}$. Z malé Fermatovy věty platí $7^{10} = 1$, takže exponent můžeme snížit modulo 10. Dostaneme tak $29^{25} \equiv 7^5$, následně zakončíme třeba trikem $7^5 \equiv (-4)^5 \equiv -4^5 \equiv -\left(2^2\right)^5 \equiv -2^{10} \equiv -1 \pmod{11}$, kde v poslední kongruence opět používáme malou Fermatovu větu.

22. Těleso $\mathbb{Z}[i]/(3)$ má 9 prvků, takže exponent snížíme modulo 8. Dostaneme $(11 + 7i)^{18} \equiv (2 + i)^2 \equiv 3 + 4i \equiv i \pmod{3}$.

23. Třeba racionální čísla. V $\mathbb{Q}$ umíme nenulovými prvky dělit, takže je to těleso, avšak suma libovolně mnoha jedniček nikdy není 0. Stejně tak nadtělesa racionálních čísel jako $\mathbb{R}$, $\mathbb{C}$ nebo kterékoliv $\mathbb{Q}(\sqrt{d})$ mají charakteristiku 0.

24. Nechť $x \times 1$ značí součet x jedniček. Pro spor nechť $c \nmid n$, potom dělení se zbytkem dává $n = cq + r$, kde $0 < r < c$. Máme $n \times 1 = 0$ a zároveň $c \times 1 = 0$. Z $n \times 1$ tak můžeme odečítat $c \times 1$ a vždy dostaneme nulu, takže $r \times 1 = (n - qc) \times 1 = 0$. Přitom ale $0 < r < c$, takže to je spor s definicí charakteristiky.

25. Z malé Fermatovy věty platí $a^{n-1} = 1$, takže musí $n - 1$ být násobek řádu a.

26. Nechť je r řád 89 v $\mathbb{Z}_p$. Vztah $p \mid 89^8 + 1$ upravme na $89^8 \equiv -1 \pmod{p}$, což značí $r \nmid 8$. Umocněním na druhou pak $89^{16} \equiv 1 \pmod{p}$, z čehož $r \mid 16$. Z toho je r dělitel 16, který není dělitelem 8, takže už $r = 16$. Těleso $\mathbb{Z}_p$ má p prvků, takže vzhledem k $r \mid p - 1$ hledáme prvočísla se zbytkem 1 modulo 16. Prvním takovým je 17, jenže spočítáme, že

$$89^8 + 1 \equiv 4^8 + 1 \equiv 16^4 + 1 \equiv (-1)^4 + 1 \equiv 2 \pmod{17}.$$

Další čísla se zbytkem 1 modulo 16 jsou 33, 49, 65 a 81, což nejsou prvočísla. Další se tedy nabízí teprve $p = 97$, což vyjde díky

$$89^8 + 1 \equiv (-8)^8 + 1 \equiv 64^4 + 1 \equiv (-33)^4 + 1 \equiv 1089^2 + 1 \equiv 22^2 + 1 \equiv 484 + 1 \equiv 0 \pmod{97}.$$

27. Nechť je s řád a^k. Víme, že $\left(a^k\right)^{\frac{r}{k}} = a^r = 1$, takže $s \leq \frac{r}{k}$. Zároveň kdyby $s < \frac{r}{k}$, znamenalo by to $a^{ks} = \left(a^k\right)^s = 1$ a zároveň $ks < r$, což by byl spor s tím, že r je řád a.

28. Nechť je g primitivní prvek. Wilsonova věta říká, že součin všech nenulových prvků konečného tělesa je roven -1. Tedy součin nenulových prvků tělesa je

$$a_1 a_2 \cdots a_{n-1} = g^0 \cdot g^1 \cdots g^{n-2} = g^{\frac{(n-1)(n-2)}{2}} = \left(g^{\frac{n-1}{2}}\right)^{n-2} = (-1)^{n-2} = -1.$$

Rozeberme postupně všechny rovnosti. V první prostě zapíšeme prvky tělesa pomocí primitivního prvku. Druhá rovnost je jen součet exponentů a použití vzorce $1 + \cdots + (n-2) = \frac{(n-1)(n-2)}{2}$. Poté nahlédněme $g^{\frac{n-1}{2}} = -1$. Z malé Fermatovy věty máme $\left(g^{\frac{n-1}{2}}\right)^2 = g^{n-1} = 1$, takže $\left(g^{\frac{n-1}{2}} - 1\right)\left(g^{\frac{n-1}{2}} + 1\right) = 0$. Z definice primitivního prvku nemůže nastat $g^{\frac{n-1}{2}} = 1$, takže $g^{\frac{n-1}{2}} = -1$. Poslední rovnost plyne lichostí $n-2$.

29. Podle tvrzení máme $S = \left\{g^m, g^{2m}, \ldots, g^{\frac{n-1}{m}\cdot m}\right\}$. Navíc jelikož $|S| > 1$, pak určitě $m < n-1$, takže $g^m \neq 1$. Součet prvků S je potom jenom

$$g^m + g^{2m} + \cdots + g^{(\frac{n-1}{m}-1)m} + g^{\frac{n-1}{m}\cdot m} = g^m \cdot \frac{g^{(n-1)m} - 1}{g^m - 1} = g^m \cdot \frac{1^m - 1}{g^m - 1} = 0.$$

30. Nechť je g primitivní prvek. Nehledě na k platí $0^k = 0$ a pro $x \neq 0$ je x^k nenulové, takže nulu nadále nemusíme řešit. Prostost f je ekvivalentní tomu, že se každé $a \in T$ objeví jako obraz nějakého x. Pokud je k nesoudělné s $n-1$, tak máme Bézoutovy koeficienty u, v splňující $uk + v(n-1) = 1$. Pro $a = g^r$ potom $f(g^{ru}) = g^{ruk} = g^{r(uk+v(n-1))} = g^r = a$, takže už je f skutečně prosté.

Pokud naopak mají k a $n-1$ nějakého společného dělitele $d > 1$, pak už každé $f(x)$ bude ležet v multiplikativní množině $\{g^d, g^{2d}, g^{3d}, \ldots\}$, která neobsahuje všechny prvky T. Zobrazení f tak nemůže být prosté.

31. Je-li g primitivní prvek, jsou kvadratickými zbytky právě $g^2, g^4, g^6, \ldots$, takže jich je $\frac{n-1}{2}$. Poté ještě přičteme 1 za nulu, která je vždy kvadratickým zbytkem.

32. Nula součet nezmění, takže se stačí dívat na nenulové kvadratické zbytky. Ty tvoří multiplikativní množinu o $\frac{n-1}{2} > 1$ prvcích, takže už se podle dřívějšího cvičení o součtu multiplikativní množiny musí kvadratické zbytky posčítat na 0.

33. Platí $\left(\frac{-1}{p}\right) \equiv (-1)^{\frac{p-1}{2}}$. Dále bude -1 kvadratickým zbytkem, právě když uvedený Legendreův symbol vyjde jedna, což nastane právě tehdy, když bude $\frac{p-1}{2}$ sudé, tedy když $4 \mid p-1$.

34. $\left(\frac{ab}{T}\right) = (ab)^{\frac{n-1}{2}} = a^{\frac{n-1}{2}} b^{\frac{n-1}{2}} = \left(\frac{a}{T}\right)\left(\frac{b}{T}\right)$.

Teorie nejen čísel 3 – Polynomy

Milý příteli,
vítej u závěrečného dílu letošního seriálu. Minule jsme si ukázali sílu jednotek pomocí Pellovy rovnice, jejíž řešení se dají všechna vyrobit z jednoho fundamentálního, i na konečných tělesech, jejichž (nenulové) prvky se dají všechny vyrobit z jednoho primitivního.

V tomto díle se zaměříme na to, jak si upéct okruhy *polynomů* a jak potom budou chutnat v závislosti na tom, z jakého těsta je pečeme. Při jejich zkoumání se vrátíme ke kořenům.[1] Proto oprášíme znalosti z prvního dílu, především to, že když už „funguje“ dělení s nějakým malým zbytkem, tak už „fungují“ i obdoby prvočísel a jednoznačný rozklad na ně. Na závěr potom změníme žánr a jako detektivové se naučíme skládat velkou modulární skládačku ze spousty menších kousků.

Jak seriál číst

Tak jako v předchozích dílech na Tebe i nyní čeká množství cvičení a úloh na procvičení. Na konci seriálu nalezneš návody ke cvičením i k úlohám a také řešení cvičení. Tak jako dříve přisuzujeme některým cvičení mírně odlišný význam: vykřičníky označují obzvláště důležitá cvičení, která mnohdy využíváme v některých pozdějších důkazech, zatímco cvičení s hvězdičkou se zabývají nějakou zajímavostí nebo náročnější myšlenkou, která není k dalšímu chápání seriálu nutná.

Okruhy polynomů

Doposud jsme v seriálu vyráběli nové, větší okruhy z těch menších tak, že jsme k nim přidali nový prvek. Ten jsme vždy zvolili tak, aby splňoval nějakou rovnost: když například k celým číslům přidáváme imaginární jednotku i, máme $i^2 = -1$, takže jakýkoliv výraz s vysokými mocninami i dovedeme zjednodušit, a k zapsání libovolného prvku $\mathbb{Z}[i]$ nám tak stačí tvar $a + bi$, kde $a, b \in \mathbb{Z}$. Podobně když jsme v minulém díle vyráběli osmiprvkové těleso, přidali jsme k $\mathbb{Z}_2$ prvek α splňující $\alpha^3 = \alpha + 1$, což nám dovolilo vystačit pro zapsání prvků $\mathbb{Z}_2[\alpha]$ s tvarem $a\alpha^2 + b\alpha + c$, kde $a, b, c \in \mathbb{Z}_2$. Tentokráte to zkusíme jinak: budeme přidávat prvek x, po kterém nebudeme požadovat vůbec nic. Dostaneme tak okruhy, jejichž prvky jsou nějaké výrazy s neznámou.

Úmluva. Není-li řečeno jinak, bude v tomto dílu R vždy značit nějaký (libovolný) komutativní okruh.

Definice. *Polynomem*[2] *nad* R rozumíme výraz tvaru

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

kde n je nezáporné celé číslo a *koeficienty* $a_0, a_1, \ldots, a_n$ jsou prvky R. Koeficientu a_0 říkáme *absolutní člen* a polynom, který má všechny koeficienty kromě a_0 nulové, nazveme *konstantní*.

[1] A to ve více než jednom významu ...

[2] Nebo též česky *mnohočlenem*.

Polynomy sčítáme tak, že sečteme koeficienty u jednotlivých mocnin, kupříkladu

$$(ax^2 + bx + c) + (dx + e) = ax^2 + (b + d)x + (c + e).$$

Při násobení zase roznásobíme všechny členy, součiny mocnin x znásobíme jako $x^k \cdot x^\ell = x^{k+\ell}$ a výsledky posčítáme, například

$$\begin{aligned}(ax^2 + bx + c) \cdot (dx + e) &= adx^3 + aex^2 + bdx^2 + bex + cdx + ce = \\ &= (ad)x^3 + (ae + bd)x^2 + (be + cd)x + ce.\end{aligned}$$

Polynomy nad R tvoří okruh, který značíme $R[x]$.

Poznámka. Symbol x v zápisu polynomů představuje nějakou úplně obecnou proměnnou, která nemá žádnou konkrétní hodnotu. Okruh $R[x]$ tedy vzniká tím, že k R přidáme prvek x, kterému ale nepřiřkneme vůbec žádné specifické vlastnosti. Později uvidíme, že nám toto umožňuje za x dosadit libovolnou hodnotu. V tomto dílu seriálu pro nás však samotné x bude vždy značit proměnnou v zápisu polynomu a pro konkrétní hodnoty vždy použijeme jiné písmenko.

V zapisování polynomů si budeme dovolovat některá zjednodušení: členy s nulovým koeficientem budeme vynechávat, namísto $(-a)x$ budeme psát $-ax$ a s koeficienty 1 se nebudeme obtěžovat (tedy místo $1x$ napíšeme jen x).

Poznámka. Každý prvek $a \in R$ se vyskytuje i v $R[x]$ jako konstantní polynom a, takže můžeme uvažovat, že R je podmnožinou množiny $R[x]$, resp. dokonce i podokruhem okruhu $R[x]$.

Cvičení 1. Okruh $R[x]$ je triviální, právě pokud je triviální R. Připomeňme, že okruhu říkáme *triviální*, pokud má jen jeden prvek.

Polynomy jsou *konečné* výrazy tvořené násobením a sčítáním proměnné x a nějakých konstant z okruhu R. Nekonečné výrazy jako například

$$1 + x + x^2 + x^3 + x^4 + x^5 + \cdots,$$

kde máme nenulový koeficient u nekonečně mnoha mocnin x^k, za polynomy nepovažujeme. U každého polynomu tak máme nějaký největší exponent mocniny, která má ještě nenulový koeficient.

Definice. Mějme nenulový polynom $f \in R[x]$ zapsaný ve tvaru $f = a_n x^n + \cdots + a_1 x + a_0$, kde n je nezáporné celé číslo. *Stupněm* polynomu f rozumíme největší k takové, že $a_k \neq 0$. Stupeň f značíme $\deg f$ a pro $f = 0$ dodefinováváme $\deg 0 = -\infty$.

Poznámka. Konstantní jsou ty polynomy, které mají stupeň 0 nebo $-\infty$. Polynomům stupňů 1, 2 a 3 říkáme po řadě *lineární*, *kvadratické* a *kubické*.

Stupeň je tedy exponent nejvyšší mocniny x, která „je“ v daném polynomu. Proč stupeň nulového polynomu definujeme jako mínus nekonečno? Když se domluvíme, že

$$-\infty + n = -\infty \qquad \text{a} \qquad -\infty + (-\infty) = -\infty,$$

pak platí následující:

Tvrzení. *Nechť je R obor integrity. Potom pro $f, g \in R[x]$ platí* $\deg(fg) = \deg f + \deg g$.

Důkaz. Nejprve předpokládejme, že f, g jsou oba nenulové. Nechť je $f = a_n x^n + \cdots + a_0$ a $g = b_m x^m + \cdots + b_0$, kde $a_n, b_m \neq 0$. Potom když roznásobíme

$$(a_n x^n + \cdots + a_0) \cdot (b_m x^m + \cdots + b_0),$$

pak dostaneme hromadu členů, v nichž nejvyšší zastoupená mocnina x bude x^{n+m}. Ta však vznikne pouze z $a_n x^n \cdot b_m x^m$ a z žádného jiného součinu. Koeficient při x^{n+m} tak ve výsledném polynomu bude $a_n b_m$, což z předpokladu $a_n, b_m \neq 0$ znamená, že tento koeficient je nenulový, jelikož pracujeme v oboru integrity. Žádné vyšší mocniny x v součinu nevzniknou, takže už $\deg(fg) = n + m = \deg f + \deg g$.

Zbývá vyřešit případy, kdy je alespoň jeden z f, g nulový. Potom je ale i $fg = 0$, takže $\deg(fg) = -\infty$. I jeden z f, g je ale nulový, takže $\deg f + \deg g$ je součet $-\infty$ a buďto celého čísla, nebo dalšího $-\infty$. V obou případech je tento součet $-\infty$, takže dokazované tvrzení platí. □

Příklad. Předpoklad, že R je obor integrity, nemůžeme v předchozím tvrzení vypustit. Když např. v $\mathbb{Z}_4[x]$ vezmeme $f = g = 2x + 1$, pak $\deg f = \deg g = 1$, ale $fg = 4x^2 + 4x + 1 = 1$, takže $\deg(fg) = 0$.

Cvičení(!) 2. Pro polynomy f, g platí $\deg(f + g) \leq \max\{\deg f, \deg g\}$, a pokud navíc $\deg f \neq \deg g$, pak už určitě nastává rovnost.

S pomocí stupně si zvládneme rozmyslet některé základní vlastnosti okruhů polynomů – kdy se jedná o obory integrity a posléze jak mohou vypadat jednotky.

Tvrzení. *Je-li R oborem integrity, pak jím je i $R[x]$.*

Důkaz. Mějme $f, g \in R[x]$ splňující $fg = 0$. Stupeň pravé strany je $-\infty$. Pokud f ani g nejsou nulové polynomy, pak jsou jejich stupně alespoň 0, čímž dostaneme

$$-\infty = \deg 0 = \deg(fg) = \deg f + \deg g \geq 0 + 0 = 0,$$

což je spor. □

Jak je to v okruzích polynomů s dělitelností? Z obecné definice dělitelnosti platí $f \mid g$, když existuje polynom h takový, že $fh = g$. Pokud pracujeme nad oborem integrity, tak rovnou víme, že bychom takové h měli hledat se stupněm $\deg g - \deg f$. Přímo z obecné definice bychom mohli rovnou také modulit, ale podrobněji se na modulení polynomů a některá jeho specifika podíváme o něco později.

Cvičení(!) 3. Nechť je R obor integrity. Pokud jsou $f, g \in R[x]$ nenulové polynomy, pak $f \mid g$ implikuje $\deg f \leq \deg g$.

Cvičení(!) 4. Mějme $a \in R$ a polynom $f \in R[x]$. Pak $a \mid f$, právě když a dělí (v okruhu R) všechny koeficienty f.

Cvičení 5. Rozhodni, zda platí následující dělitelnosti nad uvedenými okruhy:

(i) $x^2 + x + 1 \mid x^4 + x^2 + 1$ nad $\mathbb{Q}$,
(ii) $x + 1 \mid x^2 - \sqrt{2}$ nad $\mathbb{R}$,
(iii) $2x + 2 \mid x^2 - 1$ nad $\mathbb{Z}$,
(iv) $x^2 + 1 \mid x^5 + x^4 + x^3 + x^2 + x + 1$ nad $\mathbb{Z}_2$.

S pomocí stupně se také můžeme podívat na to, jaké existují v $R[x]$ jednotky, tedy polynomy, které dělí 1.

Tvrzení. *Pro obor integrity R jsou jednotkami v $R[x]$ pouze jednotky z R.*

Důkaz. Pokud je R triviální, pak je i $R[x]$ triviální a tvrzení platí (v obou je jednotkou jejich jediný prvek). Nadále předpokládejme, že R není triviální – potom ani $R[x]$ není triviální.

Pokud je u jednotkou v R, znamená to, že $uv = 1$ pro nějaké $v \in R$. Pak jsou ale u, v také prvky $R[x]$, takže u je jednotkou i v $R[x]$.

Dále řešme $uv = 1$ pro $u, v \in R[x]$. Pokud je jedno z u, v nula, pak dostaneme $0 = 1$ a rovnost neplatí ($R[x]$ není triviální). Dále když jedno z u, v nebude konstantní, pak

$$0 = \deg 1 = \deg(uv) = \deg u + \deg v \geq 1 + 0,$$

což je spor. Řečeno slovy: součin nenulových polynomů může být konstantní, jen když jsou oba činitelé rovněž konstantní. Takže řešení $uv = 1$ můžeme dostat jen tehdy, když jsou u i v konstantní, tedy když se jedná o jednotky z R. □

Příklad. Bez předpokladu, že R je obor integrity, by předchozí tvrzení neplatilo. Např. v $\mathbb{Z}_4[x]$ je $2x+1$ jednotkou, protože platí $(2x+1)(-2x+1) = -4x^2+1 = 1$.

V seriálu se polynomy nad okruhy, které nejsou obory integrity, příliš zabývat nebudeme, takže tvrzení plynoucí z této vlastnosti můžeme považovat za celkem základní poučky: násobení sčítá stupně a přidání x k oboru integrity R nám nerozbije krácení v rovnicích ani nepřidá žádné nové jednotky.

Dosazování a kořeny

Dosud jsme s polynomy jenom počítali jako v jakémkoliv jiném okruhu. Nyní si představíme nástroj, který už je specifickou vlastností polynomů – lze do nich dosazovat.

Definice. Máme-li v $R[x]$ polynom $f = c_n x^n + \cdots + c_1 x + c_0$, můžeme do něj *dosadit* nějaký jiný polynom $g \in R[x]$ tím, že „místo x napíšeme g“, tedy

$$f(g) = c_n g^n + \cdots + c_1 g + c_0.$$

Poznámka. Když se omezíme na dosazování pouze prvků $a \in R$, pak nám polynom dává zobrazení z R do R. Potom říkáme, že *f nabývá v bodě a hodnoty $f(a)$*. Samotný polynom však odlišujme od zobrazení, které takto definuje. Např. nad $\mathbb{Z}_2$ určují polynomy x^2 a x totožná zobrazení, jelikož

$$0^2 \equiv 0 \pmod 2 \qquad \text{i} \qquad 1^2 \equiv 1 \pmod 2.$$

Přesto se však jedná o odlišné polynomy a v $\mathbb{Z}_2[x]$ platí $x^2 \neq x$.

Příklad. Pro libovolný polynom f získáme dosazením nuly jeho absolutní člen. Naproti tomu $f(1)$ bude součet všech koeficientů.

Příklad. Když do polynomu f dosadíme polynom x, pak jsme jenom „místo x napsali x“, takže se nic nezmění. Zápisy f a $f(x)$ pro nás tedy představují stejný polynom. Pro některé konkrétní polynomy se nemusí nic změnit i pro jiná dosazení – kupříkladu $f = x^4 + x^2 + 1$ splňuje $f(-x) = f$.

Cvičení 6. Nechť je R obor integrity. Potom pro nekonstantní polynomy $f, g \in R[x]$ platí

$$\deg(f(g)) = \deg f \cdot \deg g.$$

Když už umíme do polynomů dosazovat, pojďme se podívat na nějaké hezké vlastnosti, které se k dosazování vážou.

Lemma. (rozdíl argumentů dělí rozdíl hodnot) *Pro polynom $f \in R[x]$ a libovolná $a, b \in R[x]$ platí $a - b \mid f(a) - f(b)$.*

Důkaz. Nechť $f(x) = c_n x^n + c_{n-1} x^{n-1} + \cdots + c_1 x + c_0$, pak můžeme vyjádřit

$$\begin{aligned} f(a) - f(b) &= \left(c_n a^n + c_{n-1} a^{n-1} + \cdots + c_1 a + c_0\right) - \left(c_n b^n + c_{n-1} b^{n-1} + \cdots + c_1 b + c_0\right) = \\ &= c_n \left(a^n - b^n\right) + c_{n-1} \left(a^{n-1} - b^{n-1}\right) + \cdots + c_1 (a - b) + c_0 (1 - 1). \end{aligned}$$

Když si nyní vezmeme člen $c_k(a^k - b^k)$, pak můžeme nahlédnout, že je opravdu dělitelný $a - b$, jelikož platí

$$a^k - b^k = (a - b)\left(a^{k-1} + a^{k-2} b + \cdots + a b^{k-2} + b^{k-1}\right).$$

Tato rovnost platí, protože když na pravé straně roznásobíme závorky, všechny smíšené členy tvaru $a^{k-j}b^j$ se navzájem vyruší a zbude jen $a^k - b^k$. Tím jsme hotovi, jelikož každý člen ve vyjádření $f(a) - f(b)$ je dělitelný $a - b$, pročež je tím dělitelný i celý součet. □

Jak za chvíli uvidíme, toto lemma se může hodit při řešení některých olympiádních úloh. V úlohách se nejčastěji hodí dosazovat za a, b konstanty (typicky celá čísla, pracujeme-li nad $\mathbb{Z}$). Všimni si ale, že lemma platí i tehdy, když jsou a, b nekonstantní polynomy – na použitých dělitelnostech se totiž nic nezmění. Z tohoto důvodu se lemma může hodit i tehdy, když chceme nahlédnout nějakou dělitelnost v okruhu polynomů.

Úloha 1. Nechť je $f \in \mathbb{Z}[x]$. Dokaž, že pokud pro nějaké celé číslo a platí[3] $f(f(\dots f(a)\dots)) = a$, kde f je použito k-krát za sebou, pak už dokonce $f(f(a)) = a$.

Definice. *Kořenem* polynomu $f \in R[x]$ rozumíme takový prvek $t \in R$, který splňuje $f(t) = 0$.

Příklad. Polynom $x^2 - 5x + 6$ má nad okruhem $\mathbb{Z}$ kořeny 2 a 3. Polynom $x^2 + 1$ nemá nad $\mathbb{R}$ žádný kořen, ale nad $\mathbb{C}$ už má dvojici kořenů i a $-i$. Polynom $x^3 - 2x$ nad $\mathbb{Z}_5$ má jen jeden kořen 0. Pro konstantní polynom 0 nad okruhem R je kořenem libovolný prvek R.

Poznámka. I když máme polynom f nad R, občas se může vyplatit podívat se na něj nad nějakým širším okruhem $S \supset R$ a hledat kořeny tam. Např. v předchozím příkladě je polynom $x^2 + 1$ celočíselný, ale kořeny má až teprve v komplexních číslech.

Cvičení(!) 7. Když $f \mid g$, pak je každý kořen f také kořenem g.

Cvičení 8. Dokaž, že libovolný kořen polynomu dělí jeho absolutní člen.

Cvičení 9. Mějme polynom $f \in \mathbb{Z}[x]$, který ve třech různých bodech $a, b, c \in \mathbb{Z}$ nabývá hodnoty 1 nebo -1. Ukaž, že potom nemůže mít celočíselný kořen.

Předchozí cvičení naznačují, že kořeny mohou být užitečné pro používání dosazovacího lemmatu a dovedou nám něco říci o příslušném polynomu. Je snadné odvodit, že jediným kořenem lineárního polynomu $ax + b$ je v tělese (třeba reálných číslech) číslo $t = \frac{-b}{a}$. Obdobně je známý vzoreček pro kořeny kvadratického polynomu, které mohou existovat nanejvýš dva. Jak je to ale obecně?

V minulém díle jsme už kořeny potkali při zkoumání těles – okruhů, v nichž je každý nenulový prvek jednotkou, jako jsou například racionální čísla $\mathbb{Q}$ nebo konečná tělesa $\mathbb{Z}_p$ pro prvočísla p. Nastínili jsme, že počet kořenů polynomu tvaru $x^n - 1$ je nad tělesem nanejvýš n, to jest stupeň daného polynomu. Nyní už se toho nebudeme bát, protože o polynomech víme více, a dokážeme si to obecněji.

Věta. *Pokud je R obor integrity, pak má nenulový polynom $f \in R[x]$ nejvýše $\deg f$ různých kořenů v R.*

Důkaz. Budeme postupovat indukcí podle stupně f. Máme nenulový polynom, takže $\deg f \geq 0$. Pokud $\deg f = 0$, pak je f nenulový konstantní, takže $f = c \neq 0$, kde $c \in R$. To ale znamená $f(t) = c \neq 0$ pro každé $t \in R$, takže f nemá žádný kořen. Skutečně tedy má nanejvýš $0 = \deg f$ kořenů.

Nyní postupujme indukcí. Pokud f nemá žádné kořeny, jsme hotovi. Pokud má nějaký kořen t, použijeme lemma. Za a vezmeme lineární polynom x a za b vezmeme kořen t. Pak vidíme, že

$$x - t \mid f(x) - f(t) = f - 0 = f.$$

Z definice dělitelnosti to znamená $f = (x - t) \cdot g$ pro nějaký polynom $g \in R[x]$. Předpokládáme, že R je obor integrity, tudíž platí

$$\deg f = \deg(x - t) + \deg g = 1 + \deg g$$

[3] Abychom dostali $f(a)$, dosadíme do polynomu a. Abychom dostali $f(f(a))$, dosadíme do něj $f(a)$. Takto můžeme pokračovat dále a dostávat „více f kolem a“.

neboli $\deg g = n-1$. Nyní z indukčního předpokladu víme, že g má nanejvýš $n-1$ kořenů a $x-t$ má navíc nanejvýš jeden nový[4] kořen t. Jelikož je R oborem integrity, musí kořen f být kořenem $x-t$ nebo kořenem g, takže f má celkově nanejvýš n kořenů. □

Příklad. (varovný) Bez předpokladu, že R je obor integrity, věta nemusí platit. Například nad okruhem $\mathbb{Z}_8$ má polynom x^2-1 hned čtyři různé kořeny 1, 3, 5 a 7.

Tvrzení, které jsme používali v minulém díle, totiž že $x^n = 1$ má nad tělesem T nanejvýš n řešení, je jen speciální případ právě dokázané věty, jelikož každé těleso už je i obor integrity. Z věty také snadno plynou dva užitečné důsledky (v obou stále předpokládáme, že R je obor integrity).

Důsledek. *Má-li polynom $f \in R[x]$ stupně nejvýše n alespoň $n+1$ různých kořenů, pak už $f = 0$.*

Důkaz. Kdyby $f \neq 0$, pak má z věty nanejvýš n různých kořenů. □

Důsledek. *Pokud se dva polynomy $f, g \in R[x]$ stupně nanejvýš n shodují v alespoň $n+1$ různých bodech, pak jsou už nutně totožné.*

Důkaz. Body, v nichž se f a g shodují, jsou kořeny polynomu $f-g$, který tak má alespoň $n+1$ kořenů. Taktéž má ale stupeň nanejvýš n, z čehož plyne $f-g=0$ neboli $f=g$. □

Omezení na počet kořenů, resp. na to, v kolika bodech se mohou dva polynomy shodovat, se často hodí i v úlohách – můžeme si hned nějakou zkusit.

Příklad. Jsou dána tři různá reálná čísla a, b, c. Zjednoduš polynom

$$f = \frac{(x-a)(x-b)}{(c-a)(c-b)} + \frac{(x-b)(x-c)}{(a-b)(a-c)} + \frac{(x-c)(x-a)}{(b-c)(b-a)}.$$

Řešení. Není třeba cokoliv roznásobovat. Když dosadíme a, dostaneme v prostředním zlomku $\frac{(a-b)(a-c)}{(a-b)(a-c)} = 1$, zatímco ve zbylých dvou bude některá závorka v čitateli 0. Máme tedy $f(a) = 1$ a obdobně $f(b) = f(c) = 1$. Polynom f a konstantní polynom 1 mají oba stupeň nanejvýš 2 a shodují se ve třech různých bodech, takže dostáváme rovnost polynomů $f = 1$.

Cvičení 10. Nenulový polynom $f \in \mathbb{R}[x]$ stupně n s navzájem různými reálnými kořeny $t_1, \ldots, t_n$ nazveme *vteřinový*, pokud $f(t_i+1) = 1$ pro každé $i = 1, 2, \ldots, n$. Najdi všechny vteřinové polynomy.

Úloha 2. Polynomy $f, g, h \in \mathbb{Q}[x]$ splňují $x^2+x+1 \mid f$ a $f(x) = g(x^3) + x \cdot h(x^3)$. Dokaž, že $x-1$ dělí g i h.

Úloha 3. Mějme polynom $f \in \mathbb{Z}[x]$ stupně $n > 1$ a označme $g(x) = f(f(\ldots(x)\ldots))$, kde f je aplikováno k-krát. Dokaž, že existuje nanejvýš n celočíselných řešení t rovnice $g(t) = t$.

Podíváme-li se znovu na důkaz věty o počtu kořenů, mohli bychom postup přeformulovat takto: z polynomu f stupně n postupně vytýkáme lineární dvojčleny $x-t$, až nakonec dostaneme vyjádření $f = (x-t_1)\cdots(x-t_r)\cdot g$, kde g je nějaký polynom bez kořenů. Pokud zrovna budeme mít štěstí a z f půjde vytknout n takových dvojčlenů, pak bude g jenom nějaká konstanta a obdržíme tvar $f = c(x-t_1)\cdots(x-t_n)$. Speciálně tak platí:

Pozorování. Má-li polynom f nad oborem integrity R stupeň n a také n *navzájem různých* kořenů $t_1, \ldots, t_n$, potom už $f = c_n(x-t_1)\cdots(x-t_n)$, kde c_n je koeficient u x^n.

Úloha 4. Nechť je $f \in \mathbb{R}[x]$ polynom stupně 2020 takový, že $f(k) = \frac{1}{k}$ pro $k \in \{1, \ldots, 2021\}$. Najdi hodnotu $f(2022)$.

Máme-li takové součinové vyjádření polynomu, je namístě položit si otázku: jak souvisí kořeny $t_1, \ldots, t_n$ (ty nemusí být nutně navzájem různé) s koeficienty f? Drobný příklad vztahu koeficientů a kořenů jsme už viděli v jednom ze cvičení (kořeny dělí absolutní člen), nyní si tento vztah poněkud upřesníme. Pro jednoduchost budeme uvažovat $c = 1$.

[4] Pokud je t i kořenem g, tak se nejedná o nový kořen a f jich má stejně mnoho jako g.

Tvrzení. (Viètovy[5] vztahy) *Mějme polynom $f = x^n + c_{n-1}x^{n-1} + \cdots c_1 x + c_0 \in R[x]$ vyjádřený ve tvaru $f = (x - t_1)\cdots(x - t_n)$ pro nějaké kořeny $t_1, \ldots, t_n \in R$. Potom roznásobením dostáváme vztahy*

$$\begin{aligned} c_0 &= (-1)^n \cdot t_1 \cdots t_n, \\ c_1 &= (-1)^{n-1} \cdot (t_1 \cdots t_{n-1} + \cdots + t_2 \cdots t_n), \\ &\vdots \\ c_{n-1} &= -(t_1 + \cdots + t_n). \end{aligned}$$

Slovy: c_{n-k} se rovná $(-1)^k$ krát součet všech součinů (neuspořádaných) k-tic z čísel $t_1, \ldots, t_n$.

Důkaz. Prostě roznásobíme – vyrobíme spoustu členů, z nichž každý vznikne tak, že si v každé ze závorek $(x - t_i)$ vybereme buďto x, nebo $-t_i$. Mocninu x^{n-k} potom budou obsahovat přesně ty členy, kde jsme si v $n - k$ závorkách vybrali x a v k závorkách $-t_i$. Když tyto členy pak posčítáme, dostaneme v koeficientu c_{n-k} přesně součty všech k-tic kořenů t_i přenásobené $(-1)^k$. □

Pracujeme-li nad tělesem, můžeme se snadno vyrovnat i s tím, když u x^n máme nějaký koeficient $c_n \neq 0$. Prostě můžeme celý polynom vydělit konstantou c_n a z hlediska kořenů se nic nezmění. Ve Viètových vztazích bychom potom prostě místo c_k psali $\frac{c_k}{c_n}$.

V případě kvadratického polynomu odpovídají Viètovy vztahy metodě řešení kvadratické rovnice pomocí rozkladu na součin dvou závorek namísto počítání diskriminantu.

Příklad. Mějme reálná čísla a, b, c, d, pro která platí

$$a + b = c + d \qquad \text{a} \qquad ab = cd.$$

Dokaž, že $\{a, b\} = \{c, d\}$, tedy že se tyto dvojice čísel mohou lišit jen pořadím.

Řešení. Uvažme Viètovy vztahy pro polynomy

$$f = (x - a)(x - b), \qquad g = (x - c)(x - d).$$

Víme, že pak platí, že lineární koeficienty jsou rovny $-(a + b)$ a $-(c + d)$ a absolutní členy jsou rovny ab a cd. Z toho už jasně vidíme, že jde o totožné kvadratické polynomy, tedy i jejich kořeny budou stejné.

Myšlenka tohoto příkladu se nám občas hodí i jinde než v algebře. Pokud například v geometrii najdeme dvě dvojice délek, které splňují uvedené rovnosti, pak také víme, že jsou (v nějakém pořadí) stejné. Také si můžeme všimnout, že polynomy nám dovedou pomoct i v situacích, které na první pohled žádný polynom neobnášejí.

Cvičení 11. Mějme pro dvě různá reálná čísla a, b rovnost $a^2 + 4a + 1 = b^2 + 4b + 1 = 0$. Urči hodnotu $\frac{a}{b} + \frac{b}{a}$.

Úloha 5. Nechť jsou a, b, c reálná čísla, taková že

$$a + b + c > 0, \qquad ab + bc + ca > 0, \qquad abc > 0.$$

Dokaž, že a, b, c jsou kladná.

Na závěr této kapitolky si ukážeme ještě jedno tvrzení, které může pomoci, když zkoumáme racionální kořeny celočíselných polynomů, tedy když se na $f \in \mathbb{Z}[x]$ podíváme nad $\mathbb{Q}$.

[5] François Viète (1540–1603), francouzský matematik, má se svým jménem spojen také vzorec pro π v podobě nekonečného součinu $\frac{2}{\pi} = \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{2+\sqrt{2}}}{2} \cdot \frac{\sqrt{2+\sqrt{2+\sqrt{2}}}}{2} \cdots$

Věta. (o racionálním kořenu) *Mějme polynom* $f(x) = c_n x^n + \cdots + c_0 \in \mathbb{Z}[x]$, *kde* $c_n \neq 0$, *a zkusme do něj dosazovat racionální čísla. Potom je-li zlomek v základním tvaru* $\frac{p}{q}$ *kořenem* f, *pak* $p \mid c_0$ *a* $q \mid c_n$.

Důkaz. Dosaďme $\frac{p}{q}$. Víme, že $c_n \left(\frac{p}{q}\right)^n + \cdots + c_0 = f\left(\frac{p}{q}\right) = 0$. Když tento vztah vynásobíme q^n, získáme

$$c_n p^n + c_{n-1} p^{n-1} q + \cdots + c_1 p q^{n-1} + c_0 q^n = 0.$$

Nyní zde vystupují jen celá čísla. To znamená, že všechny členy až na $c_n p^n$ jsou dělitelné q, tudíž i ono musí být. Předpokládáme, že $\frac{p}{q}$ je v základním tvaru, takže p, q jsou nesoudělná, pročež už z $q \mid c_n p^n$ plyne $q \mid c_n$. Analogicky jsou všechny členy krom $c_0 q^n$ dělitelné p, takže $p \mid c_0$. □

Věta o racionálním kořenu má spoustu fajn důsledků, které si můžeš zkusit nahlédnout.

Definice. Říkáme, že polynom $f = c_n x^n + \cdots + c_0$ je *monický*, pokud $c_n = 1$.

Cvičení(!) 12. Nahlédni, že je-li $f \in \mathbb{Z}[x]$ monický, pak už je každý racionální kořen polynomu f dokonce celé číslo.

Cvičení 13. Nechť je a racionální číslo takové, že $n = a^7 - 7a^5 + 5a^3 - 3a + 7$ je celé číslo. Potom musí a také být celým číslem.

Cvičení 14. Najdi všechny racionální kořeny polynomu $6x^3 - 11x^2 + 6x - 1$.

Úloha 6. Jsou dána celá čísla a, b, c taková, že

$$\frac{a}{b} + \frac{b}{c} + \frac{c}{a} \qquad \text{i} \qquad \frac{b}{a} + \frac{c}{b} + \frac{a}{c}$$

jsou také celá čísla. Dokaž, že $|a| = |b| = |c|$.

Ireducibilní polynomy

Podívejme se nyní, jaké prvky hrají v okruzích $R[x]$ roli prvočísel. V prvním díle jsme se ve vší obecnosti zabývali pojmy, které prvočísla zobecňují – ireducibilními prvky a prvočiniteli. Nyní se zaměříme na ireducibilní prvky v okruzích polynomů a ukážeme si některé jejich základní vlastnosti. Uvidíme, že obzvláště hezky vypadá situace v polynomech nad tělesy.

Definice. *Ireducibilním polynomem* nad R míníme polynom $f \in R[x]$, který je v tomto okruhu ireducibilním prvkem.

Jinými slovy: polynom je ireducibilní, pokud není nulou ani jednotkou a nedá se rozložit na součin dvou polynomů, které také nejsou jednotky.

Příklad. Dokažme, že polynom $f = x^2 - 2$ je nad $\mathbb{Q}$ ireducibilní. Pro spor nechť není, pak existují polynomy $g, h \in \mathbb{Q}[x]$, které nejsou jednotky a splňují $f = gh$. Určitě jsou nenulové, a jelikož je $\mathbb{Q}$ těleso, každý nenulový konstantní polynom je jednotka. Pak musí být $\deg g, \deg h \geq 1$, tedy vzhledem k $\deg f = 2$ dokonce $\deg g = \deg h = 1$. Když si nyní označíme koeficienty, máme $g = ax + b$, $h = cx + d$ a platí

$$x^2 - 2 = (ax + b)(cx + d) = acx^2 + (ad + bc)x + bd.$$

Porovnáním koeficientů tak $ac = 1$. BÚNO můžeme předpokládat $a = c = 1$, neboť jinak bychom namísto g, h prostě uvažovali $\frac{1}{a} \cdot g$, $\frac{1}{c} \cdot h$ a stále by platilo $\frac{g}{a} \cdot \frac{h}{c} = gh \cdot \frac{1}{ac} = f \cdot \frac{1}{1}$. Dále z koeficientů u x musí být $0 = d + b$, takže $b = -d$. Konečně z absolutních členů pak $-2 = bd = -d^2$. To však nemá řešení, protože $\sqrt{2}$ ani $-\sqrt{2}$ nejsou racionální čísla. Polynom f je tak opravdu ireducibilní.

V příkladu jsme postupovali přímo z definice. Místo toho jsme si mohli ušetřit práci tím, že bychom se zajímali o kořeny.

Cvičení(!) 15. Nechť je K těleso a polynom $f \in K[x]$ má stupeň nanejvýš 3. Potom pokud f není ireducibilní, pak má v K kořen.

Cvičení(!) 16. Nad tělesem je každý lineární polynom ireducibilní.

Příklad. (varovný) Máme-li několik v sobě obsažených okruhů, jako například $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$, pak jsou v sobě obsaženy i odpovídající okruhy polynomů, tedy $\mathbb{Z}[x] \subset \mathbb{Q}[x] \subset \mathbb{R}[x] \subset \mathbb{C}[x]$. To, zda je daný polynom ireducibilní, pak může záležet na tom, nad jakým okruhem se na něj díváme.

(i) Polynom $2x + 2$ se nad $\mathbb{Z}$ rozkládá na $2 \cdot (x + 1)$, takže není ireducibilní, protože 2 není nad $\mathbb{Z}$ jednotkou, avšak nad $\mathbb{Q}$ už ano, a $2x + 2$ tak ireducibilní je.

(ii) Polynom $x^2 - 2$ je ireducibilní nad $\mathbb{Q}$, jelikož nemá v $\mathbb{Q}$ kořen, ale nad $\mathbb{R}$ už se rozkládá na $(x + \sqrt{2})(x - \sqrt{2})$.

(iii) Polynom $x^2 + 1$ je ireducibilní nad $\mathbb{R}$, jelikož nemá v $\mathbb{R}$ kořen, ale v komplexních číslech se rozkládá na $(x + i)(x - i)$.

Hovoříme-li tedy o ireducibilitě, musíme specifikovat, nad jakým okruhem ji míníme.

Úloha 7. Jsou dána navzájem různá celá čísla $a_1, a_2, \ldots, a_n$ a polynom

$$f = (x - a_1)(x - a_2) \cdots (x - a_n) - 1.$$

Dokaž, že f je nad $\mathbb{Z}$ ireducibilní.

Úloha 8. Urči, pro která přirozená n lze zvolit n navzájem různých celých čísel $a_1, a_2, \ldots, a_n$ tak, aby polynom

$$f = (x - a_1)(x - a_2) \cdots (x - a_n) + 1$$

nebyl nad $\mathbb{Z}$ ireducibilní.

Obecně není jednoduché dokázat ireducibilitu nějakého daného polynomu. Pokud bychom např. spoléhali na řešení soustavy rovnic v koeficientech jako v příkladu s $x^2 - 2$ nad $\mathbb{Q}$, museli bychom pro polynomy vysokého stupně řešit obří soustavy nelineárních rovnic se spoustou proměnných. Toto by nemělo být příliš překvapivé – v celých číslech je taky výpočetně náročné rozhodnout, zda je nějaké dané číslo prvočíslem.

Nyní si ukážeme, že nad tělesy jsou ireducibilní polynomy automaticky prvočinitelé. Stejně jako dříve v seriálu to provedeme přes eukleidovskost. Připomeňme, že obor nazýváme *eukleidovský*, pokud v něm umíme dělit se zbytkem tak, aby zbytek byl v nějakém smyslu menší než dělitel. Formálněji požadujeme, aby existovala *eukleidovská funkce* d, která prvkům oboru přiřazuje nezáporná celá čísla tak, aby pro libovolná $a, b \in R$, $b \neq 0$, platilo

(i) $d(a) = 0$, právě když $a = 0$,

(ii) $d(a) \leq d(ab)$,

(iii) existují $q, r \in R$ taková, že $a = bq + r$ a zároveň $d(r) < d(b)$.

Z prvního dílu již víme, že eukleidovský obor je nutně i gaussovský, tedy že rozklad na ireducibilní prvky je v něm jednoznačný. Zde si rozmyslíme, že polynomy nad tělesem dovedeme dělit se zbytkem tak, aby zbytek měl menší stupeň než dělitel.

Tvrzení. *Pro libovolné těleso K je $K[x]$ eukleidovský obor.*

Důkaz. Jako eukleidovskou funkci použijeme stupeň, avšak abychom dostáli formální definici eukleidovské funkce, drobně si ho upravíme. Pro $f \in K[x]$ definujme

$$d(f) = \begin{cases} 1 + \deg f, & \text{pokud } f \neq 0, \\ 0, & \text{pokud } f = 0. \end{cases}$$

Ověřme, že toto vyhovuje jako eukleidovská funkce. Pro nenulový polynom f platí $\deg f \geq 0$, takže d nabývá nezáporných celočíselných hodnot a platí $d(f) = 0$, právě když $f = 0$. Dále pro $f, g \in K[x]$, kde $g \neq 0$, platí, že když $f = 0$, tak $d(f) = d(fg) = 0$, zatímco pro $f \neq 0$ máme

$$d(f) = 1 + \deg f \leq 1 + \deg f + \deg g = 1 + \deg(fg) = d(fg),$$

kde využíváme $\deg g \geq 0$. Dohromady tak vždy platí $d(f) \leq d(fg)$.

Tím jsou splněny podmínky (i), (ii) pro eukleidovskou funkci, zbývá tak dokázat, že pro $f, g \in K[x]$, kde $g \neq 0$, dovedeme zvolit polynomy $q, r \in K[x]$ tak, aby bylo $f = gq + r$ a zároveň $d(r) < d(g)$. Označme $n = \deg f$, $m = \deg g$. Budeme postupovat jako při dělení polynomů pod sebe: v jednom kroku vždy od f odečteme vhodný násobek g tak, abychom vyrušili nejvyšší mocninu x, která ještě zbývá, a postupně takto vyrušíme všechny mocniny s exponentem větším nebo rovným m. Pojďme si to rozmyslet podrobněji.

Pokud $n < m$, můžeme prostě zvolit $q = 0$, $r = f$. Nadále tedy předpokládejme $n \geq m$ a mějme

$$f = a_n x^n + \cdots + a_1 x + a_0, \qquad g = b_m x^m + \cdots + b_1 x + b_0.$$

Zvolíme q tak, abychom vyrušili členy $a_n x^n + \cdots + a_m x^m$ v polynomu f. Koeficienty q budeme volit postupně: stanovíme si pro polynom q stupeň $n - m$ a předpis

$$q = c_{n-m} x^{n-m} + \cdots + c_1 x + x_0,$$

kde $c_{n-m}, \ldots, c_1, c_0$ jsou zatím neurčené koeficienty. Postupně je určíme tak, aby $r = f - gq$ byl polynom stupně menšího než m. Využijeme přitom toho, že b_m je nenulový prvek tělesa K, takže už to je jednotka, tedy existuje $\frac{1}{b_m}$.

Nejprve zvolíme c_{n-m} tak, abychom vyrušil člen x^n – jednoduše položíme $c_{n-m} = \frac{a_n}{b_m}$. Tím zařídíme, že polynom

$$\begin{aligned} f - c_{n-m} x^{n-m} \cdot g &= \left(a_n x^n + a_{n-1} x^{n-1} + \cdots\right) - \left(\frac{a_n}{b_m} b_m x^n + \frac{a_n}{b_m} b_{m-1} x^{n-1} + \cdots\right) = \\ &= \left(a_n - \frac{a_n}{b_m} b_m\right) x^n + \left(a_{n-1} - \frac{a_n}{b_m} b_{m-1}\right) x^{n-1} + \cdots \end{aligned}$$

bude mít stupeň nanejvýš $n - 1$, jelikož koeficient u x^n vyjde $a_n - \frac{a_n}{b_m} b_m = a_n - a_n = 0$. Stejný postup ale můžeme opakovat: zvolíme c_{n-m-1} takovým způsobem, aby mezivýsledný polynom

$$f - (c_{n-m} x^{n-m} + c_{n-m-1} x^{n-m-1}) \cdot g$$

postrádal člen s x^{n-1}, tedy aby měl stupeň nanejvýš $n - 2$, atp. Tento proces se nám nemůže „rozbít“, dokud nám zbývají koeficienty ke stanovení, neboť vždy prostě zvolíme

$$c_{n-m-k} = \frac{1}{b_m} \cdot \big(\text{koeficient u } x^{n-k} \text{ v posledním mezivýsledku}\big).$$

V závěrečném kroku pomocí volby c_0 vyrušíme mocninu x^m, takže nakonec dostaneme, že $r = f - q \cdot g$ má stupeň menší než $m = \deg g$, tudíž $d(r) < d(g)$, jak jsme chtěli. □

Důsledek. *$K[x]$ je gaussovský obor, takže každý nenulový polynom nad K má jednoznačný*[6] *rozklad na součin ireducibilních polynomů. Dále pro libovolné nenulové polynomy $f, g \in K[x]$ existuje jejich největší společný dělitel h a také existují Bézoutovy koeficienty*[7] *$u, v \in K[x]$ splňující Bézoutovu identitu $uf + vg = h$.*

[6] Až na změny pořadí a přenásobení jednotkami.

[7] Tento pojem si nepleťme s koeficienty polynomu, což jsou prvky K – pracujeme s okruhem polynomů $K[x]$, takže Bézoutovy koeficienty jsou zde taktéž polynomy.

Příklad. Rozklady polynomů x^2-1 a x^3+x^2+x+1 na ireducibilní polynomy v $\mathbb{Q}[x]$ jsou

$$x^2-1=(x+1)(x-1), \qquad \text{a} \qquad x^3+x^2+x+1=(x+1)(x^2+1).$$

Jejich největší společný dělitel je tak $x+1$, což lze pomocí Bézoutových koeficientů $-\frac{x}{2}-\frac{1}{2}$ a $\frac{1}{2}$ vyjádřit jako $x+1=\left(-\frac{x}{2}-\frac{1}{2}\right)\cdot(x^2-1)+\frac{1}{2}\cdot(x^3+x^2+x+1)$.

Předpoklad, že K je těleso, nemůžeme z tvrzení o eukleidovskosti vypustit – bez něj by totiž b_m nemusela být jednotka, takže bychom neměli zaručeno, že vždy dovedeme vyrušit nejvyšší mocninu x v mezivýsledku.

Příklad. (varovný) $\mathbb{Z}[x]$ není eukleidovský obor. Kdyby byl, pak by v něm musela fungovat Bézoutova identita, tedy by existoval největší společný dělitel každých dvou $f,g\in\mathbb{Z}[x]$, který by navíc musel být vyjádřitelný jako $uf+vg$ pro nějaká $u,v\in\mathbb{Z}[x]$. Jako protipříklad zvolme $f=x$, $g=2$. Dělitelé 2 jsou jenom ±1 a ±2, přitom však $2\nmid x$. Společnými děliteli x a 2 jsou tak pouze ±1, takže 1 je jejich největší společný dělitel. Rovnice $ux+2v=1$ ale nemá řešení – každý polynom $ux+2v$ totiž má sudý absolutní koeficient, což 1 nesplňuje.

Ohledně společných dělitelů polynomů nám něco dovedou říct jejich kořeny. Ukažme si to nad $\mathbb{Q}$: pokud mají polynomy $f,g\in\mathbb{Q}[x]$ nějaký společný kořen $t\in\mathbb{Q}$, pak platí $x-t\mid f$ i $x-t\mid g$, takže $x-t$ je společný dělitel f a g. S pomocí eukleidovskosti si však dovedeme rozmyslet, že nám mohou pomoci i kořeny, které nejsou racionální.

Tvrzení. *Mějme polynomy $f,g\in\mathbb{Q}[x]$ takové, že nějaké $t\in\mathbb{C}$ je kořenem obou z nich. Potom jsou f, g v $\mathbb{Q}[x]$ soudělné*[8].

Důkaz. Pro spor nechť jsou f, g nesoudělné. Potom máme pro nějaká $u,v\in\mathbb{Q}[x]$ Bézoutovu identitu $uf+vg=1$. Dosazením t se však z této rovnosti stane

$$1=u(t)\cdot f(t)+v(t)\cdot g(t)=u(t)\cdot 0+v(t)\cdot 0=0,$$

což je spor. Určitě tedy musí f, g být soudělné. □

Důsledek. (kořeny chodí spolu) *Pokud je $f\in\mathbb{Q}[x]$ ireducibilní a sdílí nějaký komplexní kořen s $g\in\mathbb{Q}[x]$, pak už v okruhu $\mathbb{Q}[x]$ platí $f\mid g$. Speciálně je pak už každý (komplexní) kořen f taktéž kořenem g.*

Důkaz. Polynom f je ireducibilní, takže jeho děliteli jsou (až na přenásobení jednotkou) jen 1 a f. Z tvrzení nesmí být 1 největším společným dělitelem f a g, takže už jím musí být f, z čehož $f\mid g$. □

Gaussovo lemma

Podívejme se nyní blíže na okruh celočíselných polynomů $\mathbb{Z}[x]$. Ten sice není eukleidovský, ale ukážeme si, že je stále gaussovský. Uvidíme, že z hlediska rozkládání se na součin menších polynomů není příliš velký rozdíl v tom, jestli se na celočíselný polynom díváme v $\mathbb{Z}[x]$ nebo v $\mathbb{Q}[x]$. V $\mathbb{Q}[x]$ už máme jednoznačný rozklad na ireducibilní polynomy, který posléze přeneseme do $\mathbb{Z}[x]$. Abychom toho docílili, budeme se muset vypořádat s prvočíselnými konstantními polynomy, jelikož prvočíslo $p\in\mathbb{Z}$ je (konstantní) ireducibilní polynom v $\mathbb{Z}[x]$, ale v $\mathbb{Q}[x]$ se z něj stává jednotka.

Definice. Celočíselný polynom f nazýváme *primitivní*, pokud jej nedělí žádné přirozené číslo $d>1$.

Poznámka. Pokud $f=a_nx^n+\cdots+a_0\in\mathbb{Z}[x]$, pak je f primitivní právě tehdy, když jsou čísla $a_n,\ldots,a_1,a_0$ nesoudělná. Tím míníme, že celá $(n+1)$-tice nemá jiného společného dělitele než ±1, ačkoliv některé dvojice koeficientů spolu soudělné být mohou.

[8] Připomeňme, že prvky a, b okruhu R nazýváme *soudělné*, pokud mají společného dělitele, který není jednotka.

Příklad. Polynom $x^3 + 2x^2 + 3x + 4$ je primitivní, zatímco $2x^2 + 8x + 14$ není, jelikož všechny koeficienty jsou sudé. Polynom $6x^2+10x+15$ je zase primitivní – ačkoliv je každá dvojice koeficientů soudělná, všechny tři dohromady už jsou nesoudělné.

Když zkoumáme ireducibilitu polynomu, zajímá nás, jestli se dá rozložit na součin menších polynomů. V $\mathbb{Z}[x]$ však můžeme narazit na docela nezajímavé rozklady jako

$$2x + 2 = 2 \cdot (x + 1),$$

kde jenom vytkneme jedno celé číslo ze všech koeficientů. Následně takovýto polynom není ireducibilní v $\mathbb{Z}[x]$, ale v $\mathbb{Q}[x]$ najednou může být ireducibilní – vytknuté celé číslo 2 se v $\mathbb{Q}[x]$ stává jednotkou, takže rozklad jako $2 \cdot (x + 1)$ už není v rozporu s ireducibilitou. Primitivní polynomy jsou přesně ty, které tyto nezajímavé rozklady neumožňují.

Cvičení(!) 17. Když v $\mathbb{Z}[x]$ máme $f = gh$ a f je primitivní, pak jsou i oba polynomy g, h primitivní.

Cvičení(!) 18. Mějme polynom $f \in \mathbb{Q}[x]$. Pak existuje primitivní $g \in \mathbb{Z}[x]$ a racionální číslo k takové, že $f = k \cdot g$. Tento zápis je navíc jednoznačný až na přenásobení jednotkou.

Cvičení 19. Nekonstantní ireducibilní polynom nad $\mathbb{Z}$ musí být primitivní.

Nyní už si dovedeme rozmyslet *Gaussovo lemma* – několik tvrzení o vztahu rozkladů nad $\mathbb{Z}$ a $\mathbb{Q}$. První hovoří o prvočíslech, další o primitivních polynomech a poslední dává do souvislosti ireducibilitu nad $\mathbb{Z}$ a nad $\mathbb{Q}$. Ačkoliv to tak na první pohled nemusí vypadat, říkají jednotlivé verze vesměs totéž, jen každá v jiné formulaci, a proto se název *Gaussovo lemma* používá pro všechny tři.

Ačkoliv cílíme na souvislost $\mathbb{Z}[x]$ s $\mathbb{Q}[x]$, začneme souvislostí s okruhy $\mathbb{Z}_p[x]$.

Pozorování. (modulení konstantou) Mějme celočíselný polynom $f = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$. Máme-li dáno celé číslo m, můžeme se na koeficienty podívat modulo m a získat tím polynom

$$\hat{f} = (a_n \bmod m)x^n + \cdots + (a_0 \bmod m) \in \mathbb{Z}_m[x].$$

Sčítání polynomů odpovídá sčítání jednotlivých dvojic koeficientů, zatímco při násobení polynomů budou koeficienty výsledku rovny nějakému součtu součinů původních koeficientů. Modulení však zachovává sčítání i násobení celých čísel, takže v důsledku se tyto operace zachovají i na polynomech neboli pro $f, g \in \mathbb{Z}[x]$ dostaneme

$$\widehat{(f + g)} = \hat{f} + \hat{g}, \qquad \widehat{(f \cdot g)} = \hat{f} \cdot \hat{g}.$$

Pozor, sčítání, resp. násobení $\hat{f}$ a $\hat{g}$ zde už myslíme v okruhu $\mathbb{Z}_m[x]$, protože v tomhle okruhu tyto polynomy bydlí. Také si všimněme, že modulením se může stupeň snížit (pokud se některé koeficienty vysokých mocnin x vynulují), ale nemůže se zvýšit, tedy $\deg \hat{f} \leq \deg f$.

Lemma. (Gaussovo – verze o prvočíslech) *Každé prvočíslo $p \in \mathbb{Z}$ je prvočinitelem v $\mathbb{Z}[x]$.*

Důkaz. Dokažme podle definice prvočinitele: mějme polynomy $f, g \in \mathbb{Z}[x]$ tak, že $p \mid fg$, a dokažme, že $p \mid f$ nebo $p \mid g$. Označme $h = fg$ a podívejme se podle předchozího pozorování na celou věc modulo p. Máme $p \mid h$, takže zmodulený polynom $\hat{h}$ bude prostě 0. Tím dostáváme rovnost $0 = \hat{h} = \hat{f} \cdot \hat{g}$ (v okruhu $\mathbb{Z}_p[x]$, nikoliv už v $\mathbb{Z}[x]$).

Z prvočíselnosti p je $\mathbb{Z}_p$ obor integrity, takže i $\mathbb{Z}_p[x]$ je obor integrity. Tím pádem z $0 = \hat{f} \cdot \hat{g}$ plyne, že jeden z $\hat{f}$, $\hat{g}$ musí být 0. BÚNO nechť $\hat{f} = 0$. To značí, že zmodulením f se všechny koeficienty f vynulovaly, takže se muselo jednat o samé násobky p neboli $p \mid f$, jak jsme chtěli. □

Lemma. (Gaussovo – verze o primitivnosti) *Nechť jsou $f, g \in \mathbb{Z}[x]$ primitivní polynomy. Potom je také $f \cdot g$ primitivní.*

Důkaz. Pro spor nechť fg není primitivní. Pak je fg násobkem nějakého přirozeného čísla $n > 1$. Potom můžeme vzít nějaké prvočíslo $p \mid n$ a stále budeme mít $p \mid fg$. Jenže podle verze o prvočíslech je p prvočinitel v $\mathbb{Z}[x]$, takže z $p \mid fg$ plyne $p \mid f$ nebo $p \mid g$, což je spor s primitivností těchto polynomů. □

Lemma. (Gaussovo – verze o ireducibilitě) *Nechť je $f \in \mathbb{Z}[x]$ primitivní. Potom je f ireducibilní v $\mathbb{Z}[x]$, právě pokud je ireducibilní v $\mathbb{Q}[x]$.*

Důkaz. Dokážeme ekvivalentně, že f *není* ireducibilní v $\mathbb{Z}[x]$, právě když *není* ireducibilní v $\mathbb{Q}[x]$. Nechť se nejprve rozkládá $f = gh$ pro polynomy $g, h \in \mathbb{Z}[x]$, které nejsou jednotky. Máme f primitivní, takže g ani h nemohou být konstantní – kdyby třeba g bylo konstantní, pak to nemůže být ± 1 (to jsou jednotky), takže by f mělo přirozeného dělitele většího než 1. Oba polynomy g, h jsou tedy nekonstantní, což znamená, že nejsou jednotkami v $\mathbb{Q}[x]$. Vzhledem k $\mathbb{Z}[x] \subset \mathbb{Q}[x]$ tak rozklad $f = gh$ dokládá, že ani v $\mathbb{Q}[x]$ není f ireducibilní.

Nyní dokažme opačnou implikaci. Nechť $f = gh$ pro nekonstantní polynomy $g, h \in \mathbb{Q}[x]$ a dokažme, že ani v $\mathbb{Z}[x]$ není f ireducibilní. Podle dřívějšího cvičení můžeme k polynomům g, h zvolit racionální čísla $\frac{a_1}{b_1}$, $\frac{a_2}{b_2}$ tak, že $g_1 = \frac{a_1}{b_1} \cdot g$ i $h_1 = \frac{a_2}{b_2} \cdot h$ jsou primitivní celočíselné polynomy. Pak máme

$$g_1 h_1 = \frac{a_1}{b_1} g \cdot \frac{a_2}{b_2} h = \frac{a_1 a_2}{b_1 b_2} \cdot gh = \frac{a_1 a_2}{b_1 b_2} \cdot f,$$
$$B g_1 h_1 = A f,$$

kde jsme označili $A = a_1 a_2$, $B = b_1 b_2$. Nyní máme rovnost v $\mathbb{Z}[x]$, kde jsou f, g_1 i h_1 primitivní polynomy, ale naproti tomu nám přibyly konstanty A, B. Těch se však dovedeme zbavit. Uvažujme prvočíslo $p \mid A$. To dělí pravou stranu rovnosti, takže dělí i levou. Podle verze o prvočíslech je p prvočinitel, takže dělí některý činitel na levé straně. Ale g_1, h_1 jsou primitivní, takže je p dělit nemůže. Platí proto $p \mid B$, takže toto prvočíslo můžeme z A a B zkrátit. Obdobně když prvočíslo $q \mid B$, pak i $q \mid Af$, ale nemůže být $q \mid f$, takže $q \mid A$.

Toto můžeme opakovat a postupně takto vykrátit všechna prvočísla z A, B. Celkově se tedy A a B liší jen o jednotku, tedy $A = \pm B$. Zůstane nám tak rovnost $f = \pm g_1 h_1$, v níž máme celočíselné polynomy. Z předpokladu jsou g_1 i h_1 nekonstantní, takže to určitě nejsou jednotky v $\mathbb{Z}[x]$, čímž je dokázáno, že f není ireducibilní v $\mathbb{Z}[x]$. □

Cvičení(!) 20. Mějme $f, g \in \mathbb{Z}[x]$ a nechť je f primitivní. Potom když $f \mid g$ v okruhu $\mathbb{Q}[x]$, pak $f \mid g$ i v $\mathbb{Z}[x]$.

Důsledek. *$\mathbb{Z}[x]$ je gaussovský obor.*

Důkaz. Vezměme polynom $f \in \mathbb{Z}[x]$, nalezněme jeho rozklad na ireducibilní polynomy a ukažme, že je tento rozklad jednoznačný (až na pořadí a přenásobení jednotkami). Vytkněme nejprve z f největší celé číslo t, které jej dělí, takže mějme $f = c \cdot g$ pro $c \in \mathbb{Z}$ a primitivní $g \in \mathbb{Z}[x]$. Tento rozklad na $c \cdot g$ je (až na přenásobení ± 1) jednoznačný, protože c je prostě největší společný dělitel koeficientů f. V celých číslech pak rozložíme c na součin prvočísel $p_1 \cdots p_k$, což je jednoznačné.

Zbývá se tedy postarat o primitivní polynom g. Na něj se můžeme dívat jako na prvek $\mathbb{Q}[x]$, což je eukleidovský, a tedy i gaussovský obor, takže máme jednoznačný rozklad

$$g = g_1 \cdots g_\ell$$

na ireducibilní polynomy $g_1, \ldots, g_\ell \in \mathbb{Q}[x]$. Každý racionální polynom g_i však můžeme zapsat jako $\frac{a_i}{b_i} \cdot f_i$, kde $\frac{a_i}{b_i}$ je nějaký zlomek a f_i je primitivní celočíselný polynom. Přitom se ale f_i liší od g_i jen přenásobením konstantou, což je v $\mathbb{Q}[x]$ jednotka, takže když je g_i ireducibilní nad $\mathbb{Q}$, je i f_i ireducibilní nad $\mathbb{Q}$. Spolu s primitivností to podle Gaussova lemmatu ve verzi o ireducibilitě znamená, že f_i je ireducibilní nad $\mathbb{Z}$. Tím máme rozklad

$$g = \frac{a_1 \cdots a_\ell}{b_1 \cdots b_\ell} \cdot f_1 \cdots f_\ell,$$
$$(b_1 \cdots b_\ell) \cdot g = (a_1 \cdots a_\ell) \cdot f_1 \cdots f_\ell.$$

Toto je rovnost v $\mathbb{Z}[x]$ a polynomy g i $f_1, \ldots, f_\ell$ jsou primitivní, takže stejně jako v důkazu verze o ireducibilitě už musí být $b_1 \cdots b_\ell = \pm a_1 \cdots a_\ell$. BÚNO je znaménko $\pm$ v této rovnosti $+$, takže máme rozklad $g = f_1 \cdots f_\ell$ na ireducibilní polynomy nad $\mathbb{Z}$.

Nahlédněme, že tento rozklad je jednoznačný. Mějme dva rozklady $g = f_1 \cdots f_\ell = h_1 \cdots h_m$ na ireducibilní (primitivní) celočíselné polynomy. Z primitivnosti jsou všechny zúčastněné polynomy ireducibilní i nad $\mathbb{Q}$, takže zde máme dva rozklady g na ireducibilní polynomy v $\mathbb{Q}[x]$. Zde je ale rozklad jednoznačný, takže $f_1 \cdots f_\ell$ a $h_1 \cdots h_m$ se musí lišit jen pořadím činitelů a přenásobením jednotkami nad $\mathbb{Q}$, tedy konstantami. Aby se však zachovala celočíselnost a primitivnost polynomů, musí se jednat jen o přenásobení plus nebo minus jedničkou. Jedná se tedy jen o změnu pořadí a přenásobení jednotkami nad $\mathbb{Z}$, což znamená, že rozklad g na ireducibilní polynomy nad $\mathbb{Z}$ je jednoznačný.

Tím je dohromady pro původní f jednoznačně určen rozklad $f = p_1 \cdots p_k \cdot f_1 \cdots f_\ell$. □

Poznámka. Z předchozího důkazu plyne, že ireducibilní prvky, resp. prvočinitelé v $\mathbb{Z}[x]$ jsou následujících dvou druhů:

(i) prvočísla $p \in \mathbb{Z}$,

(ii) primitivní polynomy $f \in \mathbb{Z}[x]$, které jsou ireducibilní nad $\mathbb{Q}$.

Příklad. Rozložme polynom $f = 12x^6 - 60x^4 - 12x^2 + 60$ nad $\mathbb{Z}$ na ireducibilní polynomy. Nejprve můžeme ze všech koeficientů vytknout $12 = 2 \cdot 2 \cdot 3$. Následně si všimneme, že 1 i -1 jsou kořeny, takže dovedeme vytknout $(x-1)(x+1)$. Po vytknutí zbude polynom $x^4 - 4x^2 - 5$. Zde můžeme buďto hned uhodnout rozklad, anebo si všimneme, že máme jen sudé mocniny x, takže rozložíme polynom s poloviční exponenty $x^2 - 4x - 5$ díky zjevnému kořenu -1 na $(x+1)(x-5)$, což zpětným přepsáním na dvojnásobné exponenty dá rozklad $x^4 - 4x^2 - 5 = (x^2+1)(x^2-5)$. Dostaneme tak rozklad

$$f = 2 \cdot 2 \cdot 3 \cdot (x-1) \cdot (x+1) \cdot (x^2+1) \cdot (x^2-5).$$

Poslední dva kvadratické polynomy nemají racionální kořeny, takže jsou ireducibilní nad $\mathbb{Q}$, a díky své primitivnosti jsou tak ireducibilní i nad $\mathbb{Z}$.

Závěrem povídání o Gaussově lemmatu uveďme bez důkazu, že Gaussovo lemma lze formulovat obecněji pro libovolný gaussovský obor R namísto $\mathbb{Z}$. Můžeš si zkusit rozmyslet, že když všude místo $\mathbb{Z}$ napíšeme gaussovský obor R (a „prvočinitel v R“ místo „prvočíslo“), všechny důkazy stále fungují. Z toho pak plyne následující:

Tvrzení. *Je-li R gaussovský obor, pak už je gaussovský i $R[x]$.*

Neformálně řečeno: když funguje rozklad na „prvočísla“ v R, pak už funguje i rozklad na ireducibilní polynomy v $R[x]$. V důkazu tohoto tvrzení bychom opět postupovali takřka stejně jako v $\mathbb{Z}$, jenom bychom si museli zadefinovat „zlomky“, v nichž jsou čitatel a jmenovatel prvky obecného oboru integrity R. Tím bychom dostali jeho *podílové těleso* K, díky němuž bychom si mohli jednoznačný rozklad na ireducibilní polynomy v $K[x]$ půjčit a poupravit do $R[x]$, což je přesně to, co jsme dělali s $R = \mathbb{Z}$ a $K = \mathbb{Q}$.

Eisensteinovo kritérium

Gaussovo lemma nám dává jednoduchý vztah mezi ireducibilitou nad $\mathbb{Q}$ a nad $\mathbb{Z}$, ale stále nemáme žádný snadný způsob, jak ireducibilní polynomy rozeznat. K tomuto účelu existuje mnoho kritérií, která jsou postačujícími (ale ne nutnými) podmínkami ireducibility. My si zde ukážeme jedno z nich.

Tvrzení. (Eisensteinovo kritérium) *Mějme primitivní polynom $f = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$. Pokud nějaké prvočíslo p splňuje*

(i) $p \mid a_i$ *pro* $0 \le i \le n-1$, (ii) $p \nmid a_n$, (iii) $p^2 \nmid a_0$,

pak je f ireducibilní nad $\mathbb{Z}$.

Důkaz. Pro spor nechť lze rozložit $f = gh$ pro $g, h \in \mathbb{Z}[x]$, které nejsou jednotkami. Díky primitivnosti f nemohou g ani h být konstantní. Na situaci se podíváme modulo p a následně využijeme jednoznačného rozkladu na ireducibilní polynomy v $\mathbb{Z}_p[x]$. Tím zjistíme, že absolutní koeficienty polynomů g i h jsou násobky p, což dá spor s předpokladem (iii).

Zmodulený polynom $\hat{f} \in \mathbb{Z}_p[x]$ bude mít podle předpokladu (i) všechny koeficienty kromě a_n nulové, takže zbude $\hat{f} = a_n x^n$. Z předpokladu (ii) je a_n nenulové v $\mathbb{Z}_p$, takže $\deg \hat{f} = n$. Spolu se zmodulenými $\hat{g}, \hat{h} \in \mathbb{Z}_p[x]$ pak máme v okruhu $\mathbb{Z}_p[x]$ rovnost

$$a_n x^n = \hat{f} = \hat{g} \cdot \hat{h}.$$

Rozmysleme si, že $\hat{g}$ musí mít tvar $\hat{g} = bx^k$ pro $b \not\equiv 0$. Jelikož je $\mathbb{Z}_p$ těleso, funguje v $\mathbb{Z}_p[x]$ jednoznačný rozklad na ireducibilní polynomy. Ale lineární polynom x je určitě ireducibilní, takže máme rozklad

$$\hat{f} = a_n x^n = a_n \cdot \underbrace{x \cdots x}_{n\text{-krát}},$$

kde a_n je jednotka. Jinými slovy má $\hat{f}$ ve svém rozkladu jen jediný ireducibilní polynom x, ale zato jej má hned n-krát. Polynom $\hat{g}$ potom nemůže mít ve svém rozkladu jiné ireducibilní polynomy než x, takže $\hat{g} = bx^k$ pro nějaké k a $b \in \mathbb{Z}_p$. Nemůže přitom být $b = 0$, jelikož $\hat{f} \neq 0$. Z tohoto vyvodíme, že $\hat{g}$ má nulový absolutní člen, avšak k tomu si potřebujeme rozmyslet, jaký má stupeň.

Máme tedy $\hat{g} = bx^k$, obdobně pak i $\hat{h} = cx^\ell$ pro nenulové $c \in \mathbb{Z}_p$. Nahlédněme, že $k = \deg g$ a $\ell = \deg h$, tedy že se nám modulením nezmenšili stupně. Platí $k \leq \deg g$, $\ell \leq \deg h$ a zároveň

$$k + \ell = n = \deg f = \deg g + \deg h.$$

Kdyby $k < \deg g$ nebo $\ell < \deg h$, pak už by předchozí rovnost nemohla platit, takže určitě $k = \deg g$, $\ell = \deg h$. Víme, že g i h jsou nekonstantní, takže $k, \ell \geq 1$. Z toho plyne, že $\hat{g}$ i $\hat{h}$ mají nulové absolutní členy, takže absolutní člen $b_0 \in \mathbb{Z}$ polynomu g i absolutní člen $c_0 \in \mathbb{Z}$ polynomu h musely být násobky p. Z rozkladu $f = gh$ pak máme $a_0 = b_0 c_0$, což značí $p^2 \mid a_0$. To je spor s předpokladem (iii), žádný popsaný rozklad $f = gh$ tedy nemůže existovat, pročež je f ireducibilní. □

Cvičení 21. Pro prvočíslo p a přirozené n je polynom $x^n + p$ ireducibilní nad $\mathbb{Z}$.

Pozorování. (posunutí argumentu) Je-li $c \in \mathbb{Z}$, pak je f ireducibilní, právě když je ireducibilní $\tilde{f} = f(x + c)$. Když se rozkládá $f = gh$, pak i $\tilde{f} = g(x + c) \cdot h(x + c)$, a obdobně když $\tilde{f} = \tilde{g} \cdot \tilde{h}$, pak i $f = \tilde{f}(x - c) = \tilde{g}(x - c) \cdot \tilde{h}(x - c)$. Při zkoumáni ireducibility polynomu si tedy můžeme libovolně „posunout argument“ a nic se nezmění.

Příklad. Ukažme, že $f = x^4 + x^3 + x^2 + x + 1$ je ireducibilní nad $\mathbb{Z}$. Posunutím argumentu můžeme namísto f zkoumat

$$\begin{aligned} f(x+1) &= (x+1)^4 + (x+1)^3 + (x+1)^2 + (x+1) + 1 = \\ &= \big(x^4 + 4x^3 + 6x^2 + 4x + 1\big) + \big(x^3 + 3x^2 + 3x + 1\big) + \big(x^2 + 2x + 1\big) + (x+1) + 1 = \\ &= x^4 + 5x^3 + 10x^2 + 10x + 5, \end{aligned}$$

což je ireducibilní polynom skrze Eisensteinovo kritérium s $p = 5$.

Úloha 9. Pro prvočíslo $p \in \mathbb{N}$ je polynom $f = x^{p-1} + x^{p-2} + \cdots + x + 1$ je ireducibilní nad $\mathbb{Z}$.

Když spojíme Eisensteinovo kritérium s Gaussovým lemmatem, dostaneme informaci o ireducibilitě nad $\mathbb{Q}$. V této verzi budeme moci vypustit předpoklad o primitivnosti: pokud f není primitivní, můžeme vytknout $f = d \cdot f_1$ pro nějaké primitivní f_1, a přenásobení konstantou d, což je nad $\mathbb{Q}$ jednotka, na ireducibilitě nic nezmění.

Drobnou úpravou důkazu lze také získat zobecnění Eisensteinova kritéria, které zaručuje vysoký stupeň nějakého činitele v rozkladu na ireducibilní polynomy.

Tvrzení. (Eisensteinovo kritérium nad $\mathbb{Q}$) *Mějme polynom $f = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$ stupně n. Pokud nějaké prvočíslo p splňuje*

(i) $p \mid a_i$ *pro* $0 \le i \le n-1$, (ii) $p \nmid a_n$, (iii) $p^2 \nmid a_0$,

pak je f ireducibilní nad $\mathbb{Q}$.

Cvičení(*) 22. (rozšířený Eisenstein) Mějme polynom $f = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$ stupně n. Pokud pro nějaké prvočíslo p a přirozené k platí

(i) $p \mid a_i$ pro $0 \le i \le k-1$, (ii) $p \nmid a_k$, (iii) $p^2 \nmid a_0$,

pak je f násobkem nějakého ireducibilního polynomu stupně alespoň k.

Úloha 10. Dokaž, že pro každé přirozené číslo n je polynom $x^{n+1} + 5x^n + 3$ ireducibilní nad $\mathbb{Z}$.

Modulení polynomů

K důkazům Gaussova lemmatu a Eisensteinova kritéria se nám hodilo zmodulit celočíselný polynom f nějakým prvočíslem p, čímž jsme dostali polynom $\hat{f} \in \mathbb{Z}_p[x]$. Co když nás ale zajímá modulení nějakým nekonstantním polynomem? Pro jednoduchost budeme jako modulo uvažovat jen *monické* polynomy, tedy takové, které mají u nejvyšší mocniny x koeficient 1.

Příklad. Zkusme se nad $\mathbb{Z}$ podívat na $f = x^3 + 2x^2 + x + 1$ modulo $x - 1$. Upravíme f postupným odečítáním násobků $x - 1$ tak, abychom dostali polynom s co nejmenším stupněm. Podobně jako v důkazu eukleidovskosti $K[x]$ postupně dostaneme

$$\begin{aligned} f - x^2 \cdot (x-1) &= \big(x^3 + 2x^2 + x + 1\big) - \big(x^3 - x^2\big) = 3x^2 + x + 1, \\ f - \big(x^2 + 3x\big)(x-1) &= \big(3x^2 + x + 1\big) - \big(3x^2 - 3x\big) = 4x + 1, \\ f - \big(x^2 + 3x + 4\big)(x-1) &= 4x + 1 - (4x - 4) = 5. \end{aligned}$$

Dohromady tak $f \equiv 5 \pmod{x-1}$. Samozřejmě bychom při počítání mod $x - 1$ mohli používat i $f \equiv 4x + 1$ nebo $f \equiv 5x$ nebo jakýkoliv jiný polynom tvaru $f + g \cdot (x-1)$ pro $g \in \mathbb{Z}[x]$, ale nahrazení f za konstantní polynom 5 pravděpodobně povede k nejsnazšímu počítání.

Všimněme si, že v předchozím příkladu máme $f \equiv 5 = f(1)$. To není náhodou: když se na okruh $R[x]$ podíváme modulo $x - a$ pro nějaké $a \in R$, znamená to, že jsme dosud naprosto neurčité proměnné x nařídili vlastnost $x - a \equiv 0$ neboli $x \equiv a$. To povede k $f(x) \equiv f(a)$ pro každé $f \in R[x]$. Formálněji:

Tvrzení. *Pro $f \in R[x]$ a $a \in R$ platí $f \equiv f(a) \pmod{x-a}$.*

Důkaz. Rozdíl argumentů dělí rozdíl hodnot, takže $x - a \mid f(x) - f(a)$, tedy $f(x) - f(a) \equiv 0 \pmod{x-a}$, což přesně chceme. □

Z tohoto plyne, že všechny zbytkové třídy polynomů mod $x - a$ (prvky okruhu $R[x]/(x-a)$) můžeme reprezentovat nějakým prvkem původního okruhu R. Pokud pracujeme nad oborem integrity, pak už je toto vyjádření dokonce jednoznačné – kdyby jeden polynom f byl kongruentní dvěma různým $b_1, b_2 \in R$, pak už by nenulová konstanta $b_1 - b_2$ byla násobkem lineárního polynomu $x - a$, což není možné. V jistém smyslu se tedy okruh $R[x]/(x-a)$ „chová úplně stejně“ jako R. Pro úplnost uveďme trochu formálněji, co přesně myslíme tím, že se okruhy chovají úplně stejně.

Definice. Řekněme, že okruhy R a S jsou *izomorfní*, pokud lze jejich prvky navzájem popárovat[9] zobrazením $\varphi\colon R \to S$ splňujícím

$$\varphi(a + b) = \varphi(a) + \varphi(b), \qquad \varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$$

[9]Požadujeme tedy *bijekci*: každému prvku R musí odpovídat právě jeden prvek S a naopak.

pro libovolná $a, b \in R$. Takové zobrazení φ pak nazýváme *izomorfismus* a skutečnost, že R a S jsou izomorfní, značíme $R \simeq S$.

Jinými slovy: okruhy jsou izomorfní, pokud se liší jenom nějakým přejmenováním svých prvků a jinak je jejich vnitřní struktura úplně stejná, tzn. toto přejmenování zachovává sčítání i násobení. Například pro $R[x]/(x-a) \simeq R$ je izomorfismem přiřazení $\varphi(f) = f(a)$. Snadno si pak lze rozmyslet, že dosazení jednoho pevně zvoleného prvku se chová hezky ke sčítání i násobení polynomů.

Izomorfní okruhy jsou z hlediska okruhových vlastností (cokoliv, co hovoří jen o pojmech odvozených ze sčítání a násobení) opravdu naprosto totožné. Slibujeme, že na pochopení seriálu bohatě stačí jen intuitivní představa, že izomorfní okruhy jsou stejné až na přejmenování prvků.

Ukázali jsme si, jak modulit lineárním polynomem. Co však polynomy vyšších stupňů? Ukážeme si, že modulení *ireducibilním* polynomem je jako přidání jeho kořenu.

Příklad. Podívejme se na $\mathbb{Z}[x]$ modulo x^2+1. To odpovídá tomu, že proměnné x přiřkneme vlastnost $x^2+1 \equiv 0$ neboli $x^2 \equiv -1$, což je stejný vztah, jaký splňuje komplexní číslo i, které je kořenem x^2+1. Libovolný polynom $f \in \mathbb{Z}[x]$ můžeme modulo x^2+1 zjednodušit tak, že každou mocninu x^k pro $k \geq 2$ přepíšeme na $x^k \equiv -x^{k-2}$. Dokud je exponent větší než 1, můžeme toto opakovat a postupně dostat

$$x^k \equiv -x^{k-2} \equiv x^{k-4} \equiv -x^{k-6} \equiv \cdots,$$

takže každou mocninu x^k nakonec upravíme na $\pm x$ nebo ± 1. Každý prvek $\mathbb{Z}[x]/(x^2+1)$ tak dovedeme zapsat jako $ax+b$, kde $a, b \in \mathbb{Z}$ a x je prvek splňující $x^2 \equiv -1$. Tento zápis je navíc jednoznačný, protože kdyby jeden polynom $f \in \mathbb{Z}[x]$ byl kongruentní dvěma různým a_1x+b_1, a_2x+b_2, pak už by kvadratický polynom x^2+1 dělil nenulový polynom $(a_1x+b_1)-(a_2x+b_2)$, který má stupeň nanejvýš 1.

To vše nápadně připomíná Gaussova celá čísla $\mathbb{Z}[i]$. Všechno, co jsme řekli o $\mathbb{Z}[x]/(x^2+1)$, zůstane v platnosti pro $\mathbb{Z}[i]$, pokud jen budeme místo x psát i: každý prvek $\mathbb{Z}[i]$ také dovedeme zapsat jako $ai+b$, kde $a, b \in \mathbb{Z}$ a rovněž i je prvek splňující $i^2 = -1$. Dostaneme tedy párování zbytkových tříd $ax+b \bmod x^2+1$ s čísly $ai+b$.

Tím jsme nahlédli $\mathbb{Z}[x]/(x^2+1) \simeq \mathbb{Z}[i]$. Všimněme si, že polynomu $f \equiv ax+b$ odpovídá číslo $ai+b$, což je přesně jeho hodnota $f(i)$ v komplexním bodě i.

Příklad. Nahlédněme, že $\mathbb{Z}[x]$ modulo x^3-2 se chová stejně jako okruh

$$\mathbb{Z}[\sqrt[3]{2}] = \left\{ a\left(\sqrt[3]{2}\right)^2 + b\sqrt[3]{2} + c : a, b, c \in \mathbb{Z} \right\}.$$

K tomu si rozmyslíme, že $f \equiv g \pmod{x^3-2}$, právě když $f(\sqrt[3]{2}) = g(\sqrt[3]{2})$. Izomorfismem následně bude párování, které zbytkové třídě $f \bmod x^3-2$ přiřadí číslo $f(\sqrt[3]{2})$. V jednom směru: pokud $f \equiv g$, pak máme $f = g + h \cdot (x^3-2)$ pro nějaký polynom h, takže

$$f(\sqrt[3]{2}) = g(\sqrt[3]{2}) + h(\sqrt[3]{2}) \cdot \left(\left(\sqrt[3]{2}\right)^3 - 2\right) = g(\sqrt[3]{2}) + h(\sqrt[3]{2}) \cdot 0 = g(\sqrt[3]{2}).$$

V druhém směru: když $f(\sqrt[3]{2}) = g(\sqrt[3]{2})$, pak už je $\sqrt[3]{2}$ kořenem polynomu $f-g \in \mathbb{Q}[x]$. Polynom x^3-2 je ireducibilní (třeba Eisensteinovým kritériem) nad $\mathbb{Z}$, a proto i nad $\mathbb{Q}$ a sdílí kořen $\sqrt[3]{2}$ s polynomem $f-g$. Podle důsledku „kořeny chodí spolu“ už tedy musí dělitelnost $x^3-2 \mid f-g$ platit v okruhu $\mathbb{Q}[x]$, a tedy Gaussovým lemmatem i v $\mathbb{Z}[x]$, což přesně znamená $f \equiv g \pmod{x^3-2}$.

Úloha 11. Najdi všechna celá čísla k, pro něž existují celá čísla a, b taková, že polynomy $f = x^5 - kx - 1$ a $g = x^2 - ax - b$ mají společný komplexní kořen.

Argumentace, kterou jsme v příkladu použili pro ireducibilní polynom x^3-2 a okruh $\mathbb{Z}[\sqrt[3]{2}]$, platí obecně, čímž dostáváme:

Tvrzení. *Nechť je* $f = x^n + c_{n-1}x^{n-1} + \cdots + c_1x + c_0 \in \mathbb{Z}[x]$ *monický ireducibilní polynom s kořenem* $\alpha \in \mathbb{C}$. *Potom*

$$\mathbb{Z}[x]/(f) \simeq \mathbb{Z}[\alpha] = \left\{b_{n-1}\alpha^{n-1} + \cdots + b_1\alpha + b_0 : b_{n-1}, \ldots, b_1, b_0 \in \mathbb{Z}\right\},$$

přičemž polynomu g *odpovídá jeho hodnota* $g(\alpha)$.

Ještě si ukažme využití modulení polynomů nad konečnými tělesy $\mathbb{Z}_p$. Pomocí něho totiž dovedeme snáze konstruovat konečná tělesa libovolné velikosti. Pokud budeme chtít konečné těleso s p^n prvky, kde p je prvočíslo, stačí najít ireducibilní polynom stupně n nad $\mathbb{Z}_p$.

Tvrzení. *Nechť je* $f \in \mathbb{Z}_p[x]$ *ireducibilní polynom* a $\deg f = n$. *Potom je* $\mathbb{Z}_p[x]/(f)$ *konečné těleso s* p^n *prvky.*

Důkaz. Nechť $f = c_nx^n + \cdots + c_0$. BÚNO můžeme uvažovat $c_n = 1$, protože bychom kdykoliv mohli f přenásobit konstantou c_n^{-1} a nic by se nezměnilo. Když nějaký polynom zmodulíme f, pak dovedeme všechny mocniny x^k přepsat na polynom menšího stupně pomocí vztahu

$$x^n \equiv -\left(c_{n-1}x^{n-1} + \cdots + c_0\right).$$

Všechny prvky $\mathbb{Z}_p[x]/(f)$ tak dovedeme reprezentovat jako $a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ pro nějaká $a_{n-1}, \ldots, a_0 \in \mathbb{Z}_p$. Naopak jsou všechny tyto polynom navzájem nekongruentní – libovolný jejich rozdíl má stupeň nanejvýš $n-1$, takže nemůže být násobkem f. Prvků $\mathbb{Z}_p[x]/(f)$ je tak přesně p^n, protože každý z n koeficientů volíme z p prvků tělesa $\mathbb{Z}_p$. Z toho speciálně plyne, že okruh $\mathbb{Z}_p[x]/(f)$ je konečný. Z eukleidovskosti $\mathbb{Z}_p[x]$ je ireducibilní f také prvočinitel, takže $\mathbb{Z}_p[x]/(f)$ je obor integrity. Konečný obor integrity je už nutně těleso, takže jsme skutečně sestrojili konečné těleso s p^n prvky. □

Cvičení 23. Sestroj těleso se 125 prvky.

Cvičení(*) 24. Nechť je R eukleidovský obor a $p \in R$ prvočinitel. Dokaž, že $R/(p)$ je těleso, a to i tehdy, když je $R/(p)$ nekonečná množina. To lze například využít, když v $\mathbb{Q}[x]$ modulíme libovolným ireducibilním polynomem.

Čínská zbytková věta

V druhém díle jsme podrobně zkoumali konečná tělesa, což byly typicky faktorokruhy, které jsme dostali modulením pomocí prvočinitele. Stejně tak polynomy jsme se dosud odvážili modulit jen těmi ireducibilními. Co když ale budeme chtít pracovat modulo nějaký složený prvek? Ukážeme si, jak takové faktorokruhy rozkládat na menší kousky.

Definice. Nechť jsou R, S dva komutativní okruhy. Jejich *direktním součinem* $R \times S$ míníme následovný okruh:

(i) Jeho nosnou množinou je množina uspořádaných dvojic (r, s) pro $r \in R$, $s \in S$.

(ii) Jeho operace jsou pro $(a, b), (c, d) \in R \times S$ definovány jako

$$(a, b) + (c, d) = (a + c, b + d), \qquad (a, b) \cdot (c, d) = (a \cdot c, b \cdot d),$$

přičemž v první složce provádíme sčítání a násobení v R, zatímco v druhé složce provádíme sčítání a násobení v S.

Analogicky definujeme direktní součin $R_1 \times \cdots \times R_n$ nějakých n okruhů.

Pozorování. Nulou je v $R \times S$ prvek $(0, 0)$, jedničkou zase $(1, 1)$. Mínusy se taktéž aplikují na každou složku zvlášť, tedy $-(a, b) = (-a, -b)$.

Cvičení(*) 25. Nechť je X nějaká množina. Vzpomeňme si na okruh $\mathcal{P}(X)$ definovaný na množině podmnožin X, kde jako součet podmnožin $A, B \subseteq X$ figuruje jejich symetrická diference $A \oplus B$ (množina těch prvků, které jsou v právě jedné z A, B) a jako jejich součin figuruje průnik $A \cap B$. Rozmysli si, že pro n-prvkovou množinu X je okruh $\mathcal{P}(X)$ izomorfní okruhu $\mathbb{Z}_2^n = \underbrace{\mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2}_{n\text{-krát}}$.

Direktní součin je jenom formální popis toho, že si dva okruhy žijí každý po svém, ale zabalíme je do jedné krabičky a díváme se na ně zároveň. Hned si dokážeme Čínskou zbytkovou větu, jež říká, že za určitých podmínek lze $R/(k\ell)$ rozlomit na $R/(k)$ a $R/(\ell)$ neboli že R mod $k\ell$ se chová stejně jako R mod k a R mod ℓ zabalené dohromady.

Věta. (Čínská zbytková) *Nechť je R eukleidovský obor a $k, \ell \in R$ jsou nesoudělná. Potom pro libovolná $a, b \in R$ existuje $c \in R$, které zároveň splní obě kongruence*

$$c \equiv a \pmod{k}, \qquad c \equiv b \pmod{\ell}.$$

Všechna taková c jsou navíc navzájem kongruentní mod $k\ell$.

Důkaz. Využijeme toho, že v eukleidovském oboru platí Bézoutova identita. Největším společným dělitelem k a ℓ je podle předpokladu nesoudělnosti jednička. Existují tedy Bézoutovy koeficienty $u, v \in R$, které nám splní $uk + v\ell = 1$. Zmodulením této rovnosti mod k a mod ℓ dostáváme

$$v\ell \equiv 1 \pmod{k} \qquad \text{a zároveň} \qquad uk \equiv 1 \pmod{\ell}.$$

Potom můžeme jednoduše zvolit $c = b \cdot uk + a \cdot v\ell$, čímž bude zaručeno, že

$$c \equiv a \cdot v\ell \equiv a \cdot 1 \equiv a \pmod{k} \qquad \text{a zároveň} \qquad c \equiv b \cdot uk \equiv b \cdot 1 \equiv b \pmod{\ell}.$$

Zbývá si rozmyslet jednoznačnost c mod $k\ell$. Uvažujme dvě $c_1, c_2 \in R$, která splňují

$$c_1 \equiv c_2 \equiv a \pmod{k} \qquad \text{a zároveň} \qquad c_1 \equiv c_2 \equiv b \pmod{\ell}.$$

Potom máme $c_1 - c_2 \equiv a - a \equiv 0 \pmod{k}$, takže $k \mid c_1 - c_2$. Obdobně $\ell \mid c_1 - c_2$, z čehož dohromady nesoudělností k, ℓ plyne i $k\ell \mid c_1 - c_2$ neboli $c_1 \equiv c_2 \pmod{k\ell}$. □

Důsledek. *Při splnění podmínek věty je $R/(k\ell)$ izomorfní okruhu $R/(k) \times R/(\ell)$, přičemž izomorfismem je zde přiřazení $\varphi(a \bmod k\ell) = (a \bmod k,\ a \bmod \ell)$.*

Důkaz. Z věty víme, že φ je skutečně vzájemné párování prvků $R/(k\ell)$ a $R/(k) \times R/(\ell)$. Stačí si tedy rozmyslet, že φ zachovává operace. To je ale jasné, protože se v obou složkách jedná jen o modulení a modulení operace skutečně zachovává. □

Poznámka. Snadnou indukcí plyne, že věta platí i pro více (ale konečně mnoho) modul: jsou-li $k_1, \ldots, k_n \in R$ navzájem nesoudělné prvky, pak $R/(k_1 \cdots k_n) \simeq R/(k_1) \times \cdots \times R/(k_n)$.

Poznámka. Všimni si, že eukleidovskost v důkazu používáme jen na to, abychom mohli vzít Bézoutovu identitu $uk + v\ell = 1$. O něco obecněji bychom tedy Čínskou zbytkovou větu mohli formulovat pro prvky k, ℓ libovolného okruhu, pro které lze splnit $uk + v\ell = 1$. Třeba s okruhem $\mathbb{Z}[x]$ (ten není eukleidovský) tak stále můžeme dokázat např. $\mathbb{Z}[x]/(x^2+x) \simeq \mathbb{Z}[x]/(x) \times \mathbb{Z}[x]/(x+1)$, což je posléze izomorfní $\mathbb{Z} \times \mathbb{Z}$, jelikož $\mathbb{Z}[x]$ modulo monický lineární polynom se chová jako $\mathbb{Z}$.

Na praktické použití Čínské zbytkové věty v úlohách se hodí následující úhel pohledu: máme-li něco splnit modulo $k\ell$, pak se můžeme zcela odděleně starat o to, abychom to splnili mod k a mod ℓ. Stejně tak pokud je nějaká podmínka formulována modulo $k\ell$, jsou to jenom dvě nezávislé podmínky, jedna mod k, druhá mod ℓ.

Příklad. Dokaž, že pro libovolné přirozené n existují $a, b \in \mathbb{Z}$ splňující $4a^2 + 9b^2 \equiv 1 \pmod{n}$.

Řešení. Zapišme $n = 2^k \cdot \ell$, kde ℓ je liché číslo. Podle zbytkové věty nám stačí ukázat, že zadaná kongruence má řešení mod 2^k a mod ℓ. Modulo 2^k můžeme zvolit $a \equiv 0$, $b \equiv 3^{-1}$, čímž myslíme takový prvek, že $3^{-1} \cdot 3 \equiv 1$. Ten existuje, protože 3 a 2^k jsou nesoudělná čísla. Modulo ℓ zase zvolíme $b \equiv 0$ a $a \equiv 2^{-1}$, což můžeme, jelikož 2 a ℓ jsou nesoudělná. Poté už si jen ze zbytkové věty poručíme

$$a \equiv 0 \pmod{2^k}, \qquad b \equiv 3^{-1} \pmod{2^k},$$
$$a \equiv 2^{-1} \pmod{\ell}, \qquad b \equiv 0 \pmod{\ell}$$

a máme vyhráno.

Cvičení 26. Jsou dána dvě různá kladná prvočísla p, q. Dokaž, že $p^{q-1} + q^{p-1} \equiv 1 \pmod{pq}$.

Když už takto pomocí Čínské zbytkové věty lámeme úlohu mod n na menší kusy, často se vyplatí ji rozlámat, co nejvíc to jen jde. Pro n s prvočíselným rozkladem

$$n = p_1^{k_1} \cdots p_r^{k_r},$$

kde jednotlivá prvočísla $p_1, \ldots, p_r$ jsou navzájem různá, to znamená dívat se zvlášť na jednotlivé mocniny $p_i^{k_i}$.

Definice. *Prvočíselnou mocninou* rozumíme číslo tvaru p^k pro $k \in \mathbb{N}$ a prvočíslo $p \in \mathbb{N}$.

Cvičení(!) 27. Urči v závislosti na $n \in \mathbb{N}$, kolik z čísel $1, 2, \ldots, n$ je nesoudělných s n.

Úloha 12. Dokaž, že pro každé $n \in \mathbb{N}$ existuje n po sobě jdoucích přirozených čísel, z nichž žádné není prvočíselná mocnina.

Čínská zbytková věta dovede posloužit i tam, kde se o dvou nesoudělných modulech na první pohled nic neříká. Použijeme ji tak, že si prostě v navzájem nesoudělných modulech objednáme nějaké podmínky, které nám úlohu vyřeší, a Čínská zbytková věta nám bude černou skříňkou, která splnění těchto podmínek zařídí.

Příklad. Jsou dána přirozená čísla a, b taková, že pro libovolné $n \in \mathbb{N}$ platí $b^n + n \mid a^n + n$. Dokaž, že $a = b$.

Řešení. Zadání nám dává spoustu dělitelností a chce po nás dokázat rovnost. To obecně nemusí být vůbec snadné, ale zde budeme schopni ukázat, že každé prvočíslo p dělí rozdíl $a - b$. Každé nenulové celé číslo přitom má jenom konečně mnoho dělitelů, takže tohle už zaručí $a - b = 0$.

Nejprve si zadanou dělitelnost ekvivalentně upravme na $b^n + n \mid (a^n + n) - (b^n + n)$, tedy $b^n + n \mid a^n - b^n$. Nechť je dále dáno libovolné prvočíslo $p \in \mathbb{N}$. Rozmyslíme si, že dovedeme zvolit n tak, aby platilo $p \mid b^n + n$ a zároveň $a^n - b^n \equiv a - b \pmod{p}$. Zde přichází do hry zbytková věta. Když se na výrazy díváme mod p, pak se n, které přičítáme k b^n, chová jako zbytek mod p, zatímco všechna n v exponentech se chovají jako zbytky mod $p - 1$, protože z malé Fermatovy věty je $a \equiv a^p \equiv a^{2p-1} \equiv \cdots \pmod{p}$. Čísla p a $p - 1$ jsou zjevně nesoudělná (společný dělitel musí dělit i rozdíl), takže si zbytky v těchto modulech můžeme navolit zcela nezávisle. Objednejme si tedy $n \equiv 1 \pmod{p-1}$ a $n \equiv -b \pmod{p}$. Z první kongruence budeme mít mod p zaručeno $a^n \equiv a$, $b^n \equiv b$, takže zbude

$$b^n + n \equiv b + (-b) \equiv 0 \qquad \text{a zároveň} \qquad a^n - b^n \equiv a - b.$$

To bude znamenat $p \mid b^n + n \mid a^n - b^n$, tedy $0 \equiv a^n - b^n \equiv a - b \pmod{p}$, jak jsme chtěli, což zaručí $a - b = 0$.

Mohli jsme dosazovat přímo do $b^n + n \mid a^n + n$, avšak s $a^n - b^n$ na pravé straně je už před zformulováním důkazu jasné, že náš postup musí uspět: volbou $n \pmod{p-1}$ si dělence upravíme na $a - b \pmod{p}$ a volba $n \pmod{p}$ nám zbude na zařízení $p \mid b^n + n$.

Úloha 13. Uvažujme v rovině mřížové body $(a, b) \in \mathbb{Z}^2$ s celočíselnými souřadnicemi. Bod (a, b) je *viditelný*, pokud jsou a, b nesoudělná celá čísla. Dokaž, že pro libovolné $n \in \mathbb{N}$ existuje čtverec $n \times n$ mřížových bodů, z nichž žádný není viditelný.

Úloha 14. Dokaž, že existuje přirozené číslo n takové, že pro libovolné $a \in \mathbb{Z}$ nemá číslo $a^2 + a + n$ žádného (kladného) prvočíselného dělitele menšího než 2021.

Příklad. (Lagrangeova[10] interpolace) Představme si tuto situaci: jsme v tělese K a někdo nám prozradil hodnoty

$$b_0 = f(a_0), \qquad b_1 = f(a_1), \qquad \ldots, \qquad b_n = f(a_n)$$

[10] Joseph-Louis Lagrange (1736–1813), francouzský matematik italského původu, se podílel mj. i na vzniku a standardizaci metru a kilogramu.

pro nějaká navzájem různá $a_1, \ldots, a_n \in K$. Jak z nich odhalíme polynom $f \in K[x]$? Víme, že dosazení a_i je jako modulení polynomem $x - a_i$, takže se jedná o situaci z Čínské zbytkové věty: známe zbytky f modulo jednotlivá $x - a_i$, což jsou z různosti všech a_i nesoudělné polynomy. Z Čínské zbytkové věty je tak f jednoznačně určeno modulo $(x - a_0)(x - a_1) \cdots (x - a_n)$. To je polynom stupně $n+1$, takže všechny zbytkové třídy půjdou reprezentovat právě jedním polynomem stupně nejvýše n.

Pro tento polynom lze dokonce vymyslet explicitní, ač trochu komplikovaný předpis. Pro každé $k \in \{0, 1, \ldots, n\}$ si označme

$$f_k = \text{součin polynomů } \frac{x - a_j}{a_k - a_j} \text{ pro } 0 \leq j \leq n,\ j \neq k.$$

Tento předpis zařídí, že $f_k(a_k) = 1$, protože každý ze zlomků v součinu bude po dosazení $\frac{a_k - a_j}{a_k - a_j} = 1$. Naproti tomu dosazení a_j pro $j \neq k$ povede k tomu, že dostaneme v součinu člen $\frac{a_j - a_j}{a_k - a_j} = 0$, takže $f_k(a_j) = 0$. Nyní můžeme vzít

$$f = b_0 f_0 + b_1 f_1 + \cdots + b_n f_n,$$

což už skutečně zařídí $f(a_k) = b_k$ pro každé k, neboť $b_k f_k(a_k) = b_k$ a $b_j f_j(a_k) = 0$ pro $j \neq k$. Zároveň je každý f_k součinem n lineárních polynomů, takže $\deg f_k = n$, z čehož má i f stupeň nanejvýš n.

Příklad. Polynom $x^4 - 5x^2 + 6$ není nad $\mathbb{Q}$ ireducibilní, jelikož se rozkládá na $(x^2 - 2)(x^2 - 3)$. Oba polynomy $x^2 - 2$ a $x^2 - 3$ jsou ireducibilní, takže skloubením Čínské zbytkové věty s tvrzením o modulení ireducibilním polynomem dostaneme

$$\mathbb{Z}[x]/(x^4 - 5x^2 + 6) \simeq \mathbb{Z}[x]/(x^2 - 2) \times \mathbb{Z}[x]/(x^2 - 3) \simeq \mathbb{Z}[\sqrt{2}] \times \mathbb{Z}[\sqrt{3}].$$

Libovolnému polynomu f v tomto přiřazení odpovídá dvojice hodnot $\big(f(\sqrt{2}), f(\sqrt{3})\big)$.

Cvičení(*) 28. Nechť je T konečné těleso. Nahlédni, že libovolné zobrazeni $F\colon T \to T$ se dá reprezentovat polynomem, tedy existuje $f \in T[x]$ splňující $F(a) = f(a)$ pro každé $a \in T$.

Závěr

Seriál doputoval do svého úplného závěru. Víme již, že jak pracovat s výrazy s proměnnou pomocí okruhů polynomů, dovedeme do nich dosazovat a také rozpoznávat jejich kořeny. Nezalekneme se ani ireducibilních polynomů a máme dobrou představu, co to znamená modulit polynomy. Konečně pomocí Čínské zbytkové věty dovedeme velké děsivé problémy lámat na menší a přátelštější.

Doufáme, že Tě naše kuchařské lekce bavily. Snad se nám podařilo Tě přesvědčit o tom, že důvěrně známá prostředí, pojmy a myšlenky jako celá čísla, prvočísla a jednoznačnost rozkladu na ně se dají zobecňovat a že tyto obecnější pohledy nám dovedou dát hlubší znalosti, které bychom bez nich jen těžko nahlíželi. Algebraický pohled na teorii čísel, se kterým jsme v průběhu seriálu k věcem přistupovali, skýtá nepřeberné množství krás a tajů – věz, že to, co jsme si stihli ukázat, je jen špičkou ledovce.

Nyní nám nezbývá než se rozloučit. Seriál pro Tebe psali Fíla a Matěj, ale jak to tak bývá, nápomocna nám byla mnohá další PraSátka. Zde bychom rádi poděkovali Hedvice, Danilovi, Kubovi, Pavlovi a Radkovi. Ze všeho nejvíce však děkujeme a gratulujeme Tobě, že ses dočetl(a) až na konec, a přejeme hodně štěstí při řešení úloh třetí seriálové série!

Návody ke cvičením

1. Když je R triviální, pak není moc možností, jak zvolit koeficienty polynomů v $R[x]$.

2. Kdy se mohou nejvyšší mocniny x vyrušit?

3. Použij tvrzení o stupni součinu.

4. Rozepiš podle definice dělitelnosti.

5. Pozor na to, nad jakým okruhem pracujeme. Pokud dělitelnost platí, je jasné, jaký stupeň by měl mít podíl.

6. Kde vznikne nejvyšší mocnina x?

7. Rozepiš z definice dělitelnosti.

8. Máš možnost dosadit dvě čísla. Jedno bude zřejmě kořen a jedno číslo, kterým dostaneme jen absolutní člen.

9. Pro spor předpokládej, že existuje kořen $t \in \mathbb{Z}$, a vhodně použij lemma.

10. Podívej se na kořeny a stupeň polynomu $f(x+1) - f(x) - 1$.

11. Vyjádři pomocí Viètových vztahů pro polynom $x^2 + 4x + 1$.

12. Použij větu o racionálním kořenu.

13. Věta o racionálním kořenu.

14. Věta o racionálním kořenu dává jen několik možných zlomků, které by mohly být kořenem.

15. V rozkladu f na součin menších polynomů musí jeden činitel být lineární.

16. Jaké stupně mohou mít dělitelé lineárního polynomu?

17. Co kdyby nebyly?

18. Násobením vyrob celá čísla a poté poděl největším společným dělitelem.

19. Celočíselný dělitel dává rozklad na nejednotky.

20. Vezmi $g = fh$ pro $h \in \mathbb{Q}[x]$ a pomocí Gaussova lemmatu dokaž $h \in \mathbb{Z}[x]$.

22. Vezmi rozklad f na ireducibilní polynomy, zmodul a využij jednoznačný rozklad v $\mathbb{Z}_p[x]$.

23. $125 = 5^3$, takže stačí najít ireducibilní polynom stupně 3 nad $\mathbb{Z}_5$.

24. Použij Bézoutovu identitu.

25. Zapiš do k-té složky, zdali podmnožina obsahuje k-tý prvek X.

26. Podívej se zvlášť mod p a mod q, dostaneš známou větu.

27. Nejprve vyřeš pro prvočíselné mocniny a poté vynásob.

28. Interpolace.

Návody k úlohám

1. Vytvoř posloupnost $a_{i+1} = f(a_i)$, kde $a_0 = a$. Víme, že po nějakém počtu kroků (nejvýše k) se zase dostaneme k a_0. Použij na rozdíly dvou následujících členů lemma a následně využij zacyklení.

2. Dosaď kořeny x^2+x+1. Potom buďto prostě dopočítej $g(1)$, $h(1)$, anebo si všimni, že polynom $h(1) \cdot x + g(1)$ má moc kořenů.

3. Pokud jsou a, b dvě řešené $g(t) = t$, použij na ně opakovaně lemma o rozdílu argumentů, dokud se situace nezacyklí. Následně nahlédni, že pokud je řešení $g(t) = t$ příliš mnoho, už musí být f lineární.

4. Uvaž polynom $g = x \cdot f - 1$. Zadání mu dává spoustu kořenů, zbývá určit koeficient c_n.

5. Viètovy vztahy s kubickým polynomem. Zkus pro spor uvažovat, že je nějaký kořen nekladný.

6. Využij polynom $\left(x - \frac{a}{b}\right)\left(x - \frac{b}{c}\right)\left(x - \frac{c}{a}\right)$.

7. Předpokládej $f = gh$, jaké potom mohou být hodnoty $g(a_i)$, $h(a_i)$? Dokaž $g + h = 0$ a všimni si, že polynom f umí nabývat kladných hodnot.

8. Předpokládej $f = gh$ a dokaž $g - h = 0$. Zamysli se, zda umí f nabývat záporných reálných hodnot.

9. Posuň argument, posčítej kombinační čísla a použij Eisensteina s prvočíslem p. Hodí se vědět, že $\binom{p}{k}$ je násobek p pro $0 < k < p$.

10. Aplikuj rozšířeného Eisensteina. Pak už stačí říct, že daný polynom nemá kořen.

11. Rozliš, zda je g ireducibilní. Pokud je, pak už $f \equiv 0 \pmod{g}$.

12. Nechť je a začátek hledaného úseku. Vyber si $2n$ prvočísel a požaduj $a \equiv -i \pmod{p_i q_i}$ pro $i = 0, 1, \ldots, n-1$.

13. Poruč si pro každý bod hledaného čtverce prvočíslo, kterým mají být souřadnice soudělné.

14. Polynom $x^2 + x$ nad $\mathbb{Z}_p$ neumí nabývat všech hodnot mod p. Podle toho navol $n \pmod{p}$.

Řesení cvičení

1. Podle předchozí poznámky $R \subseteq R[x]$, takže když má $R[x]$ jediný prvek, pak i R má jediný prvek. Naopak když je R triviální, jeho jediný prvek je 0. V polynomu $f \in R[x]$ musí všechny koeficienty být prvky R, takže jsou všechny 0, pročež je $f = 0$ jediným prvkem $R[x]$.

2. Nechť je $n = \max\{\deg f, \deg g\}$. V f ani g nejsou nenulové koeficienty u žádných mocnin x vyšších než x^n, takže žádné vyšší mocniny nemohou vzniknout ani v součtu, tedy $\deg(f+g) \leq n$. Pokud $\deg f \neq \deg g$, pak BÚNO $\deg f = n > \deg g$. Koeficient u x^n tak bude v g nula, takže odpovídající koeficient v $f + g$ bude stejný jako v f, kde je nenulový, takže už $\deg(f+g) = n$.

3. Z definice dělitelnosti máme $g = f \cdot h$ pro nějaké $h \in R[x]$. Kdyby $h = 0$, pak $g = f \cdot 0 = 0$, což je spor, takže $h \neq 0$. Potom ale $\deg h \geq 0$, takže $\deg g = \deg(fh) = \deg f + \deg h \geq \deg f$.

4. Nechť je $f = b_n x^n + \cdots + b_0$. Pokud máme $b_i = a \cdot c_i$ pro každé i, pak $f = a \cdot g$ pro $g = c_n x^n + \cdots + c_0$, takže $a \mid f$. Naopak když $a \mid f$, značí $f = ag$ pro nějaké $g \in R[x]$ a každý koeficient f je a-krát příslušný koeficient g.

5. (i) Platí $(x^2 + x + 1)(x^2 - x + 1) = x^4 + x^2 + 1$, takže uvedená dělitelnost vskutku platí.

(ii) Dělitelnost neplatí. Kdyby platila, tak by vzhledem k $(x+1)(x-1) = x^2 - 1$ muselo i $(x^2 - \sqrt{2}) - (x^2 - 1) = 1 - \sqrt{2}$ být násobkem $x+1$. Jenže $\deg(x+1) = 1 > 0 = \deg(1 - \sqrt{2})$, což by byl spor.

(iii) Nad $\mathbb{Q}$ bychom sice měli $(2x+2) \cdot \frac{1}{2}(x-1) = x^2 - 1$, jenže $\frac{1}{2}(x-1)$ není prvek $\mathbb{Z}[x]$. Kdyby opravdu $2x + 2 \mid x^2 - 1$, znamenalo by to i $2 \mid x^2 - 1$, jenže polynom 2 nedělí koeficienty 1 ani -1. Dělitelnost tedy neplatí.

(iv) Platí $(x^2 + 1)(x^3 + x^2) = x^5 + x^4 + x^3 + x^2$, takže kdyby zadaná dělitelnost platila, měli bychom i $x^2 + 1 \mid x + 1$, což díky stupňům nemůže platit ($\mathbb{Z}_2$ je obor integrity).

6. Nechť $f = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$, kde $a_n \neq 0$, a $m = \deg g > 0$. Potom $f(g) = a_n g^n + a_{n-1} g^{n-1} + \cdots + a_0$. Jsme v oboru integrity, takže $\deg(g^k) = \deg(g^k) = k \cdot \deg g = km$. Nenulovostí a_n tak má sčítanec $a_n g^n$ stupeň nm, zatímco každý nižší sčítanec $a_k g^k$ má menší stupeň $km < nm$, jelikož m je kladné. Mocnina x^{nm} ze sčítance $a_n g^n$ se tak nemůže vyrušit, čímž už $\deg(f(g)) = nm$, jak jsme chtěli.

7. Z definice dělitelnosti máme $g = f \cdot h$ pro nějaký polynom h. Je-li t kořen f, pak dosazením $g(t) = f(t) \cdot h(t) = 0 \cdot h(t) = 0$.

8. Dosadíme kořen t a 0. Podle lemmatu pak dostáváme $a = a - 0 \mid f(a) - f(0) = 0 - c_0 = c_0$, kde c_0 je absolutní člen f.

9. Z našeho lemmatu víme, že $a - t \mid f(a) - f(t) = f(a) = \pm 1$. Analogicky pro b a c dostáváme, že $a - t$, $b - t$ i $c - t$ dělí plus nebo mínus jedničku. To v celých číslech znamená, že jsou plus mínus

jedna, takže z Dirichletova principu jsou alespoň dvě totožná, což je ale spor s tím, že jsou a, b, c různá.

10. Označme $g = f(x+1) - f(x)$. Každé t_i je pak kořenem polynomu $g-1$. Nahlédněme, že g bude mít stupeň přesně $n-1$. Nechť $f = a_n x^n + \cdots + a_1 x + a_0$. Potom se mocnina x^n v $a_n(x+1)^n - a_n x^n$ vyruší, zatímco člen x^{n-1} vzniklý z $a_n(x+1)^n + a_{n-1}(x+1)^{n-1} - a_n x^n - a_{n-1}x^{n-1}$ bude podle binomické věty přesně $a_n \binom{n}{1} x^{n-1}$, takže bude nenulový. Polynom $g-1$ má stupeň nanejvýš $n-1$, ale n různých kořenů, takže už platí $g = 1$. To má mít stupeň $n-1$, takže určitě $n = 1$. Pak už snadno dořešíme, že vteřinové polynomy jsou přesně $f = x + c$ pro libovolné $c \in \mathbb{R}$.

11. Polynom $x^2 + 4x + 1$ má dva různé kořeny a, b, takže z Viètových vztahů víme, že $ab = 1$ a $a + b = -4$. Z toho upravíme $\frac{a}{b} + \frac{b}{a} = \frac{a^2+b^2}{ab} = \frac{(a+b)^2-2ab}{ab} = \frac{16-2}{1} = 14$. Obecné bychom v úlohách podobného rázu vyjádřili hledanou hodnotu pomocí symetrických výrazů ve Viètových vztazích.

12. Je-li $\frac{p}{q}$ racionální kořen v základním tvaru, pak $q \mid c_n = 1$, takže jde opravdu o celé číslo.

13. Pokud chceme, aby $a^7 - 7a^5 + 5a^3 - 3a + 7 = n$, pak má $a = \frac{p}{q}$ být kořenem $x^7 - 7x^5 + 5x^3 - 3x + 7 - n$. Z věty o racionálním kořenu máme $q \mid 1$, což znamená, že a je celé číslo.

14. Stačí vyzkoušet všechny $\frac{p}{q}$ pro $p \mid 1$, $q \mid 6$, tedy $\pm\frac{1}{6}$, $\pm\frac{1}{3}$, $\pm\frac{1}{2}$ a ± 1. Zjistíme, že kořeny jsou 1, $\frac{1}{2}$ a $\frac{1}{3}$.

15. Platí $f = gh$ pro nějaké $g, h \in K[x]$, které nejsou jednotky. Všechny nenulové konstantní polynomy jsou jednotky (K je těleso), takže $\deg g, \deg h \geq 1$. Ale také $\deg g + \deg h = \deg(gh) = \deg f \leq 3$, takže jedno z g, h je lineární. BÚNO $\deg g = 1$. Potom máme $g = ax + b$ pro nějaká $a, b \in K$. Tento lineární dvojčlen má kořen $-\frac{b}{a}$, takže toto je i kořenem f jakožto násobku g.

16. Uvažujme polynom f stupně 1 nad tělesem K a nechť $f = gh$ pro $g, h \in K[x]$. Tyto polynomy jsou určitě nenulové a součet jejich stupňů je 1, takže jeden je lineární a jeden konstantní. Ale nenulový konstantní polynom nad tělesem je jednotka, takže f nelze rozložit na součin dvou nejednotek.

17. Nechť pro spor g není primitivní. Potom nějaké $d \in \mathbb{Z}$, které není jednotkou, dělí g. Potom ale určitě i $d \mid gh = f$, což je spor s primitivností f.

18. Koeficienty f jsou zlomky, vezměme tedy nějaké přirozené číslo N, které je násobkem všech jmenovatelů těchto zlomků. Polynom $h = N \cdot f$ pak určitě bude ležet v $\mathbb{Z}[x]$. Koeficienty h jsou celá čísla, můžeme tedy vzít jejich největšího společného dělitele d. BÚNO je d kladné, potom je d to největší přirozené číslo, které splňuje $d \mid h$. Polynom $g = \frac{1}{d} \cdot h = \frac{N}{d} \cdot f$ potom nemůže být dělitelný žádným přirozeným číslem kromě 1, neboť $t \mid g$ znamená $td \mid h$. Máme tedy vyhráno a k polynomu g vezmeme racionální číslo $q = \frac{N}{d}$.

Zbývá jednoznačnost. Nechť $f = k_1 g_1 = k_2 g_2$, potom $g_2 = \frac{k_1}{k_2} g_1$. Zapišme zlomek v základním tvaru $\frac{p}{q} = \frac{k_1}{k_2}$. Jmenovatel q musí být společným jmenovatelem koeficientů g_1, což je primitivní polynom, takže $q = \pm 1$. Posléze $g_2 = \pm p g_1$, takže primitivností g_2 i $p = \pm 1$, z čehož $k_1 = \pm k_2$, $g_1 = \pm g_2$.

19. Budiž uvažovaný polynom f. Pokud $a \mid f$ pro nějaké $a \in \mathbb{Z}$, pak máme $f = ag$ a z nekonstantnosti f je i g nekonstantní, takže není jednotkou. Rozklad $f = ag$ tak dosvědčuje, že f není ireducibilní.

20. Mějme $g = fh$ pro nějaké $h \in \mathbb{Q}[x]$. Přepišme h na násobek primitivního celočíselného polynomu, tedy $h = \frac{a}{b} \cdot h_1$, kde $a, b \in \mathbb{Z}$ a $h_1 \in \mathbb{Z}[x]$ je primitivní. Potom $b \cdot g = a \cdot f h_1$. Každé prvočíslo $p \mid b$ nyní dělí pravou stranu, ale nemůže dělit f ani h_1 (primitivní polynomy), tedy $p \mid a$. Takto můžeme postupně krátit, až dostaneme $b \mid a$. To znamená $\frac{a}{b} \in \mathbb{Z}$, takže ve skutečnosti bylo $h = \frac{a}{b} h_1$ celočíselné, což značí $f \mid g$ nad $\mathbb{Z}$.

21. Díky koeficientu 1 u x^n je daný polynom primitivní a s prvočíslem p jsou splněny všechny tři podmínky Eisensteinova kritéria, takže máme vyhráno.

22. Nechť je $f = g_1 \cdot g_2 \cdots g_r$ rozklad f na ireducibilní polynomy. Zmodulme na obou stranách prvočíslem p. Nalevo dostaneme $\hat{f} = (a_n \bmod p)x^n + \cdots + (a_k \bmod p)x^k$, tedy $x^k \mid \hat{f}$. Polynom x je ireducibilní nad $\mathbb{Z}_p$, takže ve zmodulené rovnosti $\hat{f} = \hat{g}_1 \cdots \hat{g}_r$ se musí tyto násobky x nějak rozdělit mezi činitele na pravé straně. Kdyby však x dělilo některé dva, BÚNO $x \mid \hat{g}_1, \hat{g}_2$, pak mají g_1 i g_2 absolutní členy, které jsou násobky p. Absolutní člen f je součinem absolutních členů všech g_i, takže toto by znamenalo $p^2 \mid a_0$, což je spor. Určitě tak celé x^k dělí jedno jediné $\hat{g}_i$, BÚNO $x^k \mid \hat{g}_1$. Nemůže být $\hat{g}_1 = 0$, neboť pak by bylo $p \mid g_1 \mid f$, což vzhledem k $p \nmid a_k$ neplatí. Takže je $\hat{g}_1$ nenulový, a tudíž $k \leq \deg \hat{g}_1 \leq \deg g_1$, jak jsme chtěli.

23. Stačí najít vhodný polynom nad $\mathbb{Z}_5$. Aby byl polynom stupně ≤ 3 nad tělesem ireducibilní, stačí, aby v něm neměl kořen. To splňuje třeba $x^3 + x + 1$, takže $\mathbb{Z}_5[x]/(x^3 + x + 1)$ je 125prvkové těleso. (Stejně dobře lze použít i mnoho jiných polynomů.)

24. Stačí dokázat, že libovolný prvek $a \in R$, který není 0 mod p, má k sobě inverzní prvek $b \in R$ splňují $ab \equiv 1 \pmod{p}$. K tomu využijme toho, že a, p jsou nesoudělné prvky, takže existují Bézoutovy koeficienty u, v splňující rovnost $ua + vp = 1$. Pak stačí vzít $b = u$.

25. BÚNO nechť $X = \{1, 2, \ldots, n\}$. Předepišme izomorfismus $\varphi\colon \mathcal{P}(X) \to \mathbb{Z}_2^n$ tak, že pro $A \subseteq X$ bude $\varphi(A) = (a_1, \ldots, a_n)$, kde

$$a_k = \begin{cases} 1, & \text{pokud } k \in A, \\ 0, & \text{pokud } k \notin A. \end{cases}$$

Nyní uvažme dvě $A, B \subseteq X$ a odpovídající $\varphi(A) = (a_1, \ldots, a_n)$, $\varphi(B) = (b_1, \ldots, b_n)$. Prvek k leží v $A \oplus B$, právě když leží v právě jedné z A, B. To odpovídá tomu, že ve dvojici a_k, b_k je jedno 0 a druhé 1, což znamená $a_k + b_k \equiv 1 \pmod 2$. Podobně když $k \notin A \oplus B$, tak $a_k \equiv b_k$, tedy $a_k + b_k \equiv 0$. Z tohoto tedy vidíme, že $\varphi(A) + \varphi(B) = \varphi(A \oplus B)$.

Obdobně prvek k leží v $A \cap B$, právě když leží v obou. To odpovídá tomu, že $a_k \equiv b_k \equiv 1$. Potom $a_k \cdot b_k \equiv 1$, zatímco v ostatních případech (kdy je jedno z a_k, b_k nula) už nutně $a_k \cdot b_k \equiv 0$. Z toho vidíme, že $\varphi(A) \cdot \varphi(B) = \varphi(A \cap B)$.

26. Jakožto různá prvočísla jsou p, q nesoudělná, takže můžeme použít Čínskou zbytkovou větu. Stačí tedy dokázat zvlášť dvě odpovídající kongruence mod p a mod q. Modulo p máme dokázat $q^{p-1} \equiv 1 \pmod{p}$, což je ale jen malá Fermatova věta, protože z různosti $q \not\equiv 0 \pmod{p}$. Obdobně pro modulo q, takže máme hotovo.

27. Nejprve vyřešme případ, kdy je $n = p^k$ prvočíselná mocnina. Potom jsou s n soudělné právě násobky p, kterých je mezi $1, 2, \ldots, p^k$ přesně p^{k-1}. Nesoudělných čísel tedy zbude $p^k - p^{k-1} = p^{k-1}(p-1)$.

Nyní mějme n s prvočíselným rozkladem $n = p_1^{k_1} \cdots p_r^{k_r}$. Nějaké $a \in \{1, 2, \ldots, n\}$ je nesoudělné s n, právě když je nesoudělné s každou prvočíselnou mocninou $p_i^{k_i}$. Podle zbytkové věty tedy číslo a nesoudělné s n odpovídá nějakému a_1 nesoudělnému s $p_1^{k_1}$, nějakému a_2 nesoudělnému s $p_2^{k_2}$ atd., takže jejich počty se znásobí. Celkem tedy dostaneme $p_1^{k_1-1}(p_1 - 1) \cdots p_r^{k_r-1}(p_r - 1)$ čísel nesoudělných s n.

28. Použijeme Lagrangeovu interpolaci. Za body a_i vezmeme úplně všechny prvky T a jako hodnoty $f(a_i)$ si poručíme $b_i = F(a_i)$. Z Lagrangeovy interpolace plyne, že existuje f splňující tyto požadavky, takže skutečně $F(a) = f(a)$ pro každé $a \in T$.