ZÁKLADY ALGEBRY PRO INFORMATIKY

DAVID STANOVSKÝ

stanovsk@karlin.mff.cuni.cz

Cíle:

- algebra (a teorie čísel) je užitečná
  - chytrá reprezentace dat: pomocí konečných těles (symetrické šifry, samoopravné kódy, sdílení tajemství, kryptoanalýza), modulární reprezentace založená na čínské věta o zbytcích
  - zdroj těžkých výpočetních problémů (diskrétní logaritmus v cyklických grupách, odmocňování modulo a RSA)
- základy abstraktního algebraického uvažování: abstraktní struktury a obecné vlastnosti těchto struktur
- grupy jako nástroj pro popis symetrií objektů
- konkrétní aplikace: kryptografie (RSA, diskrétní logaritmus a Diffie-Hellman, sdílení tajemství), samoopravné kódy (Reed-Salomon)

Date: 10. února 2021.

Obsah

I. Základy teorie čísel 4
1. Prvočíselné rozklady a největší společný dělitel 5
1.1. Dělitelnost a základní věta aritmetiky 5
1.2. Eukleidův algoritmus a Bézoutova rovnost 6
2. Počítání modulo číslo 8
2.1. Kongruence 8
2.2. Eulerova věta a kryptosystém RSA 9
2.3. Čínská věta o zbytcích 12

II. Základy teorie polynomů 15
3. Tělesa a obory integrity 16
3.1. Definice a příklady 16
3.2. Základní vlastnosti 19
3.3. Podílová tělesa 19
4. Polynomy 20
4.1. Obory polynomů 20
4.2. Hodnota polynomu a polynomiální zobrazení 22
4.3. Dělení polynomů se zbytkem 22
4.4. Kořeny a dělitelnost 23
5. Základní pojmy dělitelnosti 24
5.1. Dělitelnost a asociovanost 24
5.2. Největší společný dělitel 25
5.3. Ireducibilní prvky a rozklady 25
5.4. Dělitelnost v gaussovských oborech 27
6. Dělitelnost v oborech polynomů 29
6.1. Polynomy jedné proměnné nad tělesem 29
6.2. Polynomy nad oborem vs. nad jeho podílovým tělesem 30
6.3. Racionální kořeny a Eisensteinovo kritérium ireducibility 31
7. Abstraktní teorie dělitelnosti 32
7.1. Zobecnění základní věty aritmetiky 32
7.2. Eukleidův algoritmus a Bézoutova rovnost 34
8. Počítání modulo polynom 36
8.1. Čínská věta o zbytcích a interpolace 36
8.2. Faktorokruh modulo polynom 39
9. Konečná tělesa a jejich aplikace 41
9.1. Konečná tělesa a počítačová reprezentace dat 41
9.2. Sdílení tajemství 43
9.3. Samoopravné kódy 43
9.4. Vzájemně ortogonální latinské čtverce a návrh experimentů 45

III. Grupy 49
10. Pojem grupy 50
10.1. Definice a příklady 50
10.2. Mocniny a řád prvku 53
11. Podgrupy 54
11.1. Generátory 54
11.2. Lagrangeova věta 57

12. Působení grupy na množině 59
12.1. Abstraktní grupa jako grupa permutací 59
12.2. Burnsideova věta a počítání orbit 62
13. Cyklické grupy 65
13.1. Podgrupy, generátory, řády prvků 65
13.2. Multiplikativní grupy konečných těles jsou cyklické 68
13.3. Diskrétní logaritmus a kryptografie 68
Dodatek o permutacích 71
Základní vlastnosti permutací 71
Loydova patnáctka a generátory alternující grupy 72

Základy teorie čísel

1. Prvočíselné rozklady a největší společný dělitel

Cílem této sekce je shrnout základní vlastnosti oboru celých čísel z hlediska dělitelnosti: prvočíselné rozklady, Eukleidův algoritmus, kongruence, Eulerovu větu a čínskou větu o zbytcích. Většinu z těchto principů později zobecníme (například na polynomy či větší číselné obory), ale přesto je důležité začít tímto speciálním případem.

Většina teorie této sekce byla v nějaké formě známa již ve starověku a v moderní podobě byly formulovány Carlem Friedrichem Gaussem v jeho slavné knize *Disquisitiones Arithmeticae* z roku 1801, která položila základ moderní teorie čísel.

Základním matematickým objektem v této sekci bude množina přirozených čísel $\mathbb{N} = \{1, 2, 3, 4, \dots\}$, respektive množina celých čísel $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$, a standardní aritmetické operace $+, -, \cdot$ na těchto číslech. Je řada způsobů, jak čísla definovat: můžeme je vybudovat v rámci teorie množin, můžeme je zavést axiomaticky (za standardní axiomatizaci se považují Peanovy axiomy, které jsou založeny na principu matematické indukce). Ani jedním z těchto postupů se zabývat nebudeme a budeme vycházet ze základních středoškolských poznatků.

Pokud napíšeme „číslo“, myslíme v této sekci celé číslo.

1.1. Dělitelnost a základní věta aritmetiky.

Buď a, b celá čísla. Řekneme, že číslo b *dělí* číslo a, píšeme $b \mid a$, pokud existuje číslo q splňující $a = b \cdot q$. Pro každé a platí $\pm 1 \mid a$ a $\pm a \mid a$; tito dělitelé se nazývají *nevlastní*. Pokud b nedělí a, má smysl se ptát po zbytku po dělení.

Tvrzení 1.1 (dělení celých čísel se zbytkem)**.** *Buď $a, b \in \mathbb{Z}$, $b \neq 0$. Pak existuje právě jedna dvojice celých čísel q, r splňující*

$$a = q \cdot b + r \qquad a \qquad 0 \leq r < |b|.$$

Díky jednoznačnosti můžeme definovat *celočíselný podíl* $a \operatorname{div} b = q$ a *zbytek* $a \bmod b = r$. Je vidět, že $b \mid a$ právě tehdy, když $a \bmod b = 0$.

Důkaz. Pro jednoduchost předpokládejme $a, b > 0$, ostatní případy se vyřeší analogicky. Buď q největší číslo splňující $q \cdot b \leq a$ a položme $r = a - q \cdot b$. Zřejmě $0 \leq r < b$ a platí výše uvedený vztah.

Kdyby $a = q_1 b + r_1 = q_2 b + r_2$, pak $b(q_1 - q_2) = r_2 - r_1$, tedy $b \mid r_2 - r_1$, avšak $0 \leq |r_2 - r_1| < |b|$, takže jedinou možností je případ $r_2 - r_1 = 0$. Z toho plyne $r_1 = r_2$ i $q_1 = q_2$. □

Přirozené číslo $p > 1$, které má pouze nevlastní dělitele, se nazývá *prvočíslo*; ostatní přirozená čísla se nazývají *složená*. Zcela základním poznatkem teorie čísel je fakt, že každé číslo lze jednoznačně vyjádřit jako součin prvočísel.

Věta 1.2 (základní věta aritmetiky)**.** *Pro každé přirozené číslo $a \neq 1$ existují po dvou různá prvočísla $p_1, p_2, \dots, p_n$ a přirozená čísla $k_1, k_2, \dots, k_n$ splňující*

$$a = p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_n^{k_n}$$

(tomuto vyjádření se říká prvočíselný rozklad*). Tento zápis je jednoznačný až na pořadí činitelů.*

Tuto „samozřejmost“ jako první dokázal Eukleides v 4. století př.n.l. a v dnešní době ji každý středoškolák dobře zná, nicméně, přiznejme si, kdo z vás by to uměl dokázat? Tedy existenci rozkladu lze dokázat poměrně snadno indukcí. Tuto „samozřejmost“ jako první dokázal Eukleides v 4. století př.n.l. a v dnešní době ji každý

středoškolák dobře zná, nicméně, přiznejme si, kdo z vás by to uměl dokázat? Tedy existenci rozkladu lze dokázat poměrně snadno indukcí.

Důkaz Věty 1.2, existence rozkladu. Buď a nejmenší přirozené číslo, pro nějž neexistuje prvočíselný rozklad. To nemůže být prvočíslem, jinak bychom měli rozklad $a = a^1$. Čili a je složené a můžeme jej rozložit jako $a = b \cdot c$ pro nějaká $1 < b, c < a$. Podle indukčního předpokladu existuje prvočíselný rozklad jak pro b, tak pro c. Jejich složením získáme rozklad čísla a. □

Jednoduchým důsledkem existence prvočíselných rozkladů je fakt, že existuje nekonečně mnoho prvočísel. Kdyby jich bylo jenom konečně mnoho, označme je $p_1, \ldots, p_n$ a uvažujme číslo $p_1 \cdot p_2 \cdot \ldots \cdot p_n + 1$. Toto číslo není dělitelné žádným prvočíslem, přitom musí mít nějaký prvočíselný rozklad. Spor.

S důkazem jednoznačnosti je to složitější. Nástrojem nám bude Bézoutova rovnost, která vyjadřuje největší společný dělitel jako lineární kombinaci.

1.2. **Eukleidův algoritmus a Bézoutova rovnost.**

Největší společný dělitel celých čísel a a b je největší přirozené číslo c splňující zároveň $c \mid a$ a $c \mid b$, značíme jej $\mathrm{NSD}(a, b)$. Čísla a, b nazveme *nesoudělná*, pokud $\mathrm{NSD}(a, b) = 1$. Podobně, *nejmenší společný násobek* čísel a a b je nejmenší číslo c splňující zároveň $a \mid c$ a $b \mid c$, značíme jej $\mathrm{NSN}(a, b)$. Použitím základní věty aritmetiky je snadné nahlédnout, že

$$\mathrm{NSD}(a, b) \cdot \mathrm{NSN}(a, b) = a \cdot b$$

(vzorec nebudeme k důkazu této věty potřebovat). Vzhledem k tomu, že $\mathrm{NSD}(a, b) = \mathrm{NSD}(\pm a, \pm b)$, budeme se dále zabývat výpočtem NSD pro kladná čísla.

Jeden známý postup výpočtu NSD je pomocí prvočíselných rozkladů: např. $168 = 2^3 \cdot 3 \cdot 7$ a $396 = 2^2 \cdot 3^2 \cdot 11$, a tak vidíme, že $\mathrm{NSD}(168, 396) = 2^2 \cdot 3 = 12$. Problém je, že kdybychom neměli jednoznačnost rozkladů, kdyby se např. číslo 396 rozkládalo na součin úplně jiných prvočísel než $2, 3, 11$, dostali bychom z jiného rozkladu jiný výsledek, což je problém. Nejsme tedy schopni dokázat správnost této metody, aniž bychom dokončili důkaz základní věty aritmetiky. (Skutečným příkladem, že tato metoda v obecnosti nefunguje, je např. následující situace v oboru $\mathbb{Z}[\sqrt{5}]$: zde $4 = 2 \cdot 2 = (1 + \sqrt{5})(-1 + \sqrt{5})$. Z prvního rozkladu bychom vydedukovali $\mathrm{NSD}(2, 4) = 2$, z druhého $\mathrm{NSD}(2, 4) = 1$. Viz sekce 5.)

Druhým používaným postupem výpočtu NSD je *Eukleidův algoritmus*. Ten je založen na následujícím pozorování, nezávislém na základní větě aritmetiky.

Lemma 1.3. *Pro libovolná celá čísla a, b platí*

$$\mathrm{NSD}(a, b) = \mathrm{NSD}(b, a \bmod b).$$

Důkaz. Označme $q = a \operatorname{div} b$. Pak

$$a = b \cdot q + (a \bmod b),$$

a tedy dané číslo c dělí obě čísla a, b právě tehdy, když c dělí obě čísla $b, a \bmod b$. Čili obě dvojice mají stejné společné dělitele, mají tedy stejného i toho největšího. □

Algoritmus vezme daná čísla $a \geq b \geq 0$ a buduje posloupnost tak, že vždy vezme zbytek po dělení předposledního čísla posledním. Odpovědí je poslední nenulová hodnota. Formálně, inicializujeme $a_0 = a$, $a_1 = b$ a budujeme posloupnost předpisem

$$a_{i+1} = a_{i-1} \bmod a_i.$$

Pokud vyjde $a_{i+1} = 0$, odpovědí je a_i. Například pro $\mathrm{NSD}(168, 396)$ dostáváme posloupnost 396, 168, 60, 48, 12, 0, a tedy $\mathrm{NSD}(168, 396) = 12$. Z Lemmatu 1.3 vidíme, že

$$\mathrm{NSD}(a, b) = \mathrm{NSD}(a_0, a_1) = \mathrm{NSD}(a_1, a_2) = \ldots = \mathrm{NSD}(a_k, 0) = a_k,$$

což je poslední nenulová hodnota v posloupnosti.

Rozšířením Eukleidova algoritmu lze dokázat následující vlastnost.

Tvrzení 1.4 (Bézoutova rovnost)**.** *Pro každou dvojici celých čísel a, b existují celá čísla u, v (tzv.* Bézoutovy koeficienty*) splňující*

$$\mathrm{NSD}(a, b) = u \cdot a + v \cdot b.$$

Důkaz. Eukleidův algoritmus rozšíříme tak, že v každém kroku spočte u_i, v_i taková, že $a_i = u_i \cdot a + v_i \cdot b$. Inicializujeme $(u_0, v_0) = (1, 0)$ a $(u_1, v_1) = (0, 1)$. Protože $a_{i+1} = a_{i-1} \bmod a_i = a_{i-1} - q_i a_i$, položíme

$$(u_{i+1}, v_{i+1}) = (u_{i-1}, v_{i-1}) - q_i \cdot (u_i, v_i),$$

kde $q_i = a_{i-1} \operatorname{div} a_i$. Pokud $a_{i+1} = 0$, pak u_i, v_i jsou zřejmě Bézoutovy koeficienty pro $a_i = \mathrm{NSD}(a, b)$. □

Příklad. Pro $\mathrm{NSD}(168, 396)$ dostáváme posloupnosti

a_i	u_i	v_i
396	1	0
168	0	1
60	1	−2
48	−2	5
12	3	−7
0		

Tedy $\mathrm{NSD}(168, 396) = 3 \cdot 396 - 7 \cdot 168$.

Pomocí Bézoutovy rovnosti dokážeme jedno pomocné tvrzení.

Lemma 1.5. *Buď p prvočíslo a $a, b \in \mathbb{Z}$. Platí-li $p \mid a \cdot b$, pak $p \mid a$ nebo $p \mid b$.*

(Opět, kdybychom měli v ruce jednoznačnost prvočíselných rozkladů, bylo by tvrzení očividné. Avšak v oboru $\mathbb{Z}[\sqrt{5}]$ platí $2 \mid (1+\sqrt{5})(-1+\sqrt{5})$, přesto $2 \nmid \pm 1 + \sqrt{5}$; jednoznačnost tedy hraje roli.)

Důkaz. Předpokládejme, že $p \nmid a$. Pak $\mathrm{NSD}(a, p) = 1$, protože je p prvočíslo, a tedy podle Tvrzení 1.4 existují čísla u, v splňující $au + pv = 1$. Vynásobením obou stran rovnosti číslem b dostaneme $abu + pvb = b$. Jelikož p dělí oba sčítance na levé straně, dělí i b. □

Indukcí snadno odvodíme následující důsledek:

Lemma 1.6. *Buď p prvočíslo a $a_1, \ldots, a_n$ celá čísla. Platí-li $p \mid a_1 \cdot \ldots \cdot a_n$, pak $p \mid a_i$ pro alespoň jedno i.*

Nyní můžeme přistoupit k důkazu jednoznačnosti prvočíselných rozkladů z Věty 1.2.

Důkaz Věty 1.2, jednoznačnost rozkladu. Buď a nejmenší přirozené číslo s nejednoznačným prvočíselným rozkladem a uvažujme dva různé rozklady

$$a = p_1^{k_1} \cdot \ldots \cdot p_m^{k_m} = q_1^{l_1} \cdot \ldots \cdot q_n^{l_n}.$$

Protože $p_1 \mid a = q_1^{l_1} \cdot \ldots \cdot q_n^{l_n}$, musí existovat i takové, že $p_1 \mid q_i$. Ovšem q_i je prvočíslo, tedy $p_1 = q_i$. Pak ale uvažujme číslo $b = \frac{a}{p_1}$: to má také dva různé rozklady

$$b = p_1^{k_1-1} \cdot p_2^{k_2} \cdot \ldots \cdot p_m^{k_m} = q_1^{l_1} \cdot \ldots \cdot q_i^{l_i-1} \cdot \ldots \cdot q_n^{k_n},$$

ale přitom $b < a$, což je spor s minimalitou a. □

2. Počítání modulo číslo

2.1. **Kongruence.** Symbol $\equiv$ pro kongruenci, zavedený Gaussem ve zmiňované knize *Disquisitiones Arithmeticae*, usnadňuje zápis při počítání modulo dané číslo.

Definice. Buď a, b, m celá čísla, $m \neq 0$. Řekneme, že *a je kongruentní s b modulo m*, a zapisujeme

$$a \equiv b \pmod{m},$$

pokud $m \mid a - b$.

Předně si všimněte, že $a \equiv b \pmod{m}$ právě tehdy, když a a b dávají stejný zbytek po dělení m: napišme si $a = mq_1 + r_1$ a $b = mq_2 + r_2$, čili máme $a - b = m(q_1 - q_2) + (r_1 - r_2)$ a ihned vidíme, že $m \mid a - b$ právě tehdy, když $m \mid r_1 - r_2$, čili když jsou zbytky stejné.

Z uvedené interpretace je zřejmé, že relace „býti kongruentní modulo m“ je ekvivalence, tj. že pro všechna $a, b, c \in \mathbb{Z}$ platí

- $a \equiv a \pmod{m}$;
- pokud $a \equiv b \pmod{m}$, pak $b \equiv a \pmod{m}$;
- pokud $a \equiv b \pmod{m}$, a $b \equiv c \pmod{m}$, pak $a \equiv c \pmod{m}$.

Druhou základní vlastností je invariance vůči základním operacím.

Tvrzení 2.1 (vlastnosti kongruence)**.** *Nechť $a \equiv b \pmod{m}$ a $c \equiv d \pmod{m}$. Pak platí*

$$a + c \equiv b + d \pmod{m}, \quad a - c \equiv b - d \pmod{m}, \quad a \cdot c \equiv b \cdot d \pmod{m}$$

a pro každé přirozené k platí

$$a^k \equiv b^k \pmod{m}.$$

Důkaz. CVIČENÍ □

Obě vlastnosti lze interpretovat tak, že symbol $\equiv$ můžeme používat stejným způsobem, jako rovnítko: reflexivita říká, že z $a = b$ plyne také $a \equiv b$, symetrie říká, že zápis je platný zleva doprava i zprava doleva, a tranzitivita říká, že máme-li sérii po sobě jdoucích kongruencí, pak je číslo úplně vlevo kongruentní číslu úplně vpravo. Invariance vůči operacím pak umožňuje ve výpočtu nahrazovat navzájem kongruentní čísla. Ukážeme si to na jednoduchém příkladu.

Úloha. Spočtěte $77^{123} + 66^{321} \bmod 6$.

Řešení. Protože $66 \equiv 0$ a $77 \equiv -1 \pmod{6}$, můžeme psát

$$77^{123} + 66^{321} \equiv (-1)^{123} + 0^{321} = -1 + 0 \equiv 5 \pmod{6}.$$

Uvedený výraz tedy dává zbytek 5. □

Další důležitou vlastností je, že v kongruenci smíme krátit číslem, které je nesoudělné s modulem m. Naopak, jsou-li všechna tři čísla v kongruenci soudělná, celý výraz můžeme zjednodušit tím, že společný faktor vykrátíme na obou stranách *i v modulu*. Formálně tyto vlastnosti vyjadřuje následující tvrzení.

Tvrzení 2.2 (vlastnosti kongruence)**.** *Buď a, b, c, m celá čísla, $c, m \neq 0$. Pak*

(1) $a \equiv b \pmod{m} \Leftrightarrow ca \equiv cb \pmod{cm}$*;*
(2) *jsou-li c, m nesoudělná, pak $a \equiv b \pmod{m} \Leftrightarrow ca \equiv cb \pmod{m}$.*

Důkaz. CVIČENÍ □

Úloha. Najděte všechna $x \in \mathbb{Z}$ splňující (a) $6x \equiv 9 \pmod{21}$, (b) $10x \equiv 5 \pmod{21}$.

Řešení. Použijeme několikrát Tvrzení 2.2.

(a) Užitím (1) dostaneme ekvivalentní podmínku $2x \equiv 3 \pmod{7}$, a po přenásobení obou stran číslem 4, díky (2), ekvivalentní podmínku $x \equiv 5 \pmod{7}$. Řešením jsou všechna $x = 5 + 7k$, $k \in \mathbb{Z}$.

(b) Užitím (2) dostaneme ekvivalentní podmínku $2x \equiv 1 \pmod{21}$, a po přenásobení obou stran číslem 11, díky (2), ekvivalentní podmínku $x \equiv 11 \pmod{21}$. Řešením jsou všechna $x = 11 + 21k$, $k \in \mathbb{Z}$. □

2.2. **Eulerova věta a kryptosystém RSA.**

Pro motivaci připomeňme úlohu uvedenou za základními vlastnostmi kongruencí: řešení bylo snadné především proto, že $66 \equiv 0$ a $77 \equiv -1$, přičemž tato čísla se snadno umocňují. Zamyslete se nad následující úlohou.

Úloha. Zjistěte poslední cifru čísla 77^{123}.

Řešení. Jinými slovy, spočtěte $77^{123} \bmod 10$. Můžeme psát $77^{123} \equiv 7^{123} \equiv (-3)^{123} \pmod{10}$, ale bez další teorie nám nezbývá, než mocnit sedmičku nebo trojku. Např. pro sedmičku dostáváme

$$7^1 = 7,\ 7^2 = 49 \equiv 9,\ 7^3 \equiv 7 \cdot 9 \equiv 3,\ 7^4 \equiv 7 \cdot 3 \equiv 1,\ 7^5 \equiv 7 \cdot 1 = 7$$

a vidíme, že poslední cifry se opakují s periodou 4. Vzhledem k tomu, že $123 \bmod 4 = 3$, dostáváme $7^{123} \equiv 7^3 \equiv 3 \pmod{10}$. □

To, že zbytky modulo dané číslo vykazují periodu jako v předchozí úloze, není náhoda, nýbrž pravidlo, které se nazývá *Eulerova věta*. Délku periody udává tzv. Eulerova funkce.

Definice. *Eulerova funkce* $\varphi(n)$ značí pro přirozené číslo n počet čísel $k \in \{1, \ldots, n-1\}$ nesoudělných s číslem n, tj. splňujících $\mathrm{NSD}(k, n) = 1$.

Např. $\varphi(10) = 4$, neboť s desítkou nesoudělná jsou právě čísla $1, 3, 7, 9$. Pro libovolné prvočíslo p platí $\varphi(p) = p - 1$, protože s ním nesoudělná jsou všechna menší čísla.

Výpočet Eulerovy funkce přímo z definice by byl pro větší čísla pracný. Naštěstí existuje vzorec, pomocí něhož je snadné spočítat hodnotu $\varphi(n)$, pokud známe prvočíselný rozklad čísla n.

Tvrzení 2.3. *Je-li $n = p_1^{k_1} \cdot \ldots \cdot p_m^{k_m}$ prvočíselný rozklad čísla $n > 1$, pak*

$$\varphi(n) = p_1^{k_1 - 1}(p_1 - 1) \cdot \ldots \cdot p_m^{k_m - 1}(p_m - 1).$$

Příklad. $\varphi(4056) = \varphi(2^3 \cdot 3^1 \cdot 13^2) = 2^2 \cdot 1 \cdot 3^0 \cdot 2 \cdot 13^1 \cdot 12 = 1248$.

Vzorec se celkem snadno dokáže pomocí čínské věty o zbytcích, se kterou se seznámíme v příští části. Teď se podíváme na samotnou Eulerovu větu.

Věta 2.4 (Eulerova věta)**.** *Jsou-li a, m nesoudělná přirozená čísla, pak*

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

Leonhard Euler publikoval tuto větu v roce 1763. Již dříve, v roce 1736, pak dokázal speciální případ, kdy je m prvočíslo. Objev tohoto vztahu bývá připisován Pierre de Fermatovi, objevuje se v jednom z jeho dopisů z roku 1640, a proto bývá nazýván malá Fermatova věta.

Důsledek 2.5 (malá Fermatova věta)**.** *Je-li p prvočíslo a $p \nmid a$, pak*

$$a^{p-1} \equiv 1 \pmod{p}.$$

Označme pro účely této sekce

$$\Phi_m = \{k \in \{1, \ldots, m-1\} : \ \mathrm{NSD}(k, m) = 1\}.$$

Eulerovu funkci pak můžeme zapsat jako $\varphi(m) = |\Phi_m|$. K důkazu Eulerovy věty se nám bude hodit pomocné lemma.

Lemma 2.6. *Buď $f : X \to Y$ zobrazení mezi stejně velkými konečnými množinami. Je-li f prosté, pak je bijektivní.*

Důkaz. Nechť $n = |X| = |Y|$. Hodnoty, které zobrazení f přiřadí prvkům množiny X, jsou po dvou různé, takže obor hodnot zobrazení f má právě n prvků. Čili to musí být celé Y. □

Lemma 2.7. *Buďte a, m nesoudělná přirozená čísla a definujme zobrazení*

$$f_a : \Phi_m \to \Phi_m, \quad x \mapsto ax \bmod m.$$

Zobrazení f_a je dobře definované a je to bijekce.

Důkaz. Předně je třeba ověřit, že $ax \bmod m \in \Phi_m$. Jsou-li obě čísla a, x nesoudělná s m, pak je s m nesoudělné i číslo ax: kdyby existovalo prvočíslo p dělící m i ax, pak by p dělilo a nebo x (Lemma 1.5), spor s nesoudělností. Čili $1 = \mathrm{NSD}(ax, m) = \mathrm{NSD}(ax \bmod m, m)$ použitím Lemmatu 1.3.

Nyní dokážeme, že je zobrazení f_a prosté. Uvažujme $x, y \in \Phi_m$ taková, že $f_a(x) = f_a(y)$, tj. $ax \equiv ay \pmod{m}$. Podle Tvrzení 2.2 je $x \equiv y \pmod{m}$, tedy x i y dávají stejný zbytek po dělení m. Ovšem obě čísla jsou menší než m, takže musí být stejná.

Vzhledem k tomu, že je f_a zobrazením na konečné množině, musí být také na (Lemma 2.6). □

Důkaz Eulerovy věty. Uvažujme následující výpočet, kde f_a je zobrazení definované v předchozím lemmatu:

$$\prod_{b \in \Phi_m} b \ = \ \prod_{b \in \Phi_m} f_a(b) \ = \ \prod_{b \in \Phi_m} ab \bmod m \ \equiv \ \prod_{b \in \Phi_m} ab \ = \ a^{\varphi(m)} \cdot \prod_{b \in \Phi_m} b \pmod{m}.$$

První rovnost platí díky tomu, že v obou případech násobíme přes všechny prvky množiny Φ_m, pouze v různém pořadí. Označíme-li

$$c = \prod_{b \in \Phi_m} b,$$

právě jsme dokázali, že

$$c \equiv a^{\varphi(m)} \cdot c \pmod{m}.$$

Číslo c je nesoudělné s m (protože je součinem čísel nesoudělných s m), takže jím můžeme podle Tvrzení 2.2 krátit a dostáváme $1 \equiv a^{\varphi(m)} \pmod{m}$. □

Jiný důkaz Eulerovy věty uvidíme v sekci 11.2, kde ji dostaneme jako speciální případ Lagrangeovy věty aplikované na grupu $\mathbb{Z}_m^*$.

Úloha. Zjistěte poslední cifru čísla 77^{123}.

Řešení. Použijeme Eulerovu větu: protože $\varphi(10) = 4$ a $\mathrm{NSD}(77, 10) = 1$, platí

$$77^{123} \equiv 7^{123} = 7^{4 \cdot 30 + 3} \equiv (7^4)^{30} \cdot 7^3 \equiv 1^{30} \cdot 3 = 3 \pmod{10}.$$

(Z didaktických důvodů jsme vše detailně rozepsali, v praxi samozřejmě provedete většinu úvah zpaměti a budete psát rovnou $7^{123} \equiv 7^3 = 3$.) □

Úloha. Spočtěte $10^{10^{10}} \bmod 21$.

Řešení. Použijeme Eulerovu větu: protože $\varphi(21) = 12$ a $\mathrm{NSD}(10, 21) = 1$, stačí zjistit zbytek po dělení 10^{10} číslem 12. Avšak čísla $10, 12$ jsou soudělná. Protože $\mathrm{NSD}(10^{10}, 12) = 4$, výsledek bude dělitelný 4. Napišme

$$10^{10} = 2^{10} \cdot 5^{10} \equiv 4k \pmod{12},$$

podle Tvrzení 2.2 budeme řešit úlohu $2^8 \cdot 5^{10} \equiv k \pmod{3}$. Nyní znovu použijeme Eulerovu větu: protože $\varphi(3) = 2$ a všechna zmíněná čísla jsou nesoudělná, máme

$$k = 2^8 \cdot 5^{10} \equiv 2^0 \cdot 5^0 = 1 \pmod{3},$$

čili $10^{10} \equiv 4k = 4 \pmod{12}$, a tedy $10^{10^{10}} \equiv 10^4 = 4 \pmod{21}$. □

Poznámka. Lemma 2.7 říká, že pro každé a nesoudělné s m existuje právě jedno $b \in \{1, \ldots, m-1\}$ takové, že

$$a \cdot b \equiv 1 \pmod{m}.$$

Toto b lze najít dvěma způsoby:

- podle Eulerovy věty lze vzít $b = a^{\varphi(m)-1} \bmod m$,
- Eukleidovým algoritmem spočteme Bézoutovy koeficienty $1 = \mathrm{NSD}(a, m) = ua + vm$ a vezmeme $b = u \bmod m$.

Tato úloha má vyšší smysl: právě jsme popsali způsob nalezení inverzního prvku a^{-1} v okruhu $\mathbb{Z}_m$ (uvidíte později).

S Eulerovou větou je úzce spjata jedna významná aplikace: protokol *RSA* (pojmenovaný po trojici matematiků Rivest, Shamir, Adleman) na tzv. *šifrování s veřejným klíčem*. Problém je následující: Bob přijímá zprávy od řady klientů a je nepraktické, aby si s každým vyměňoval tajné heslo. Bob tedy publikuje tzv. *veřejný klíč*, pomocí něhož mu může každý poslat šifrovanou zprávu, a tajně bude držet *soukromý klíč*, pomocí něhož může pouze on zprávy dešifrovat. Popíšeme algoritmus, jak generovat klíče a jak šifrovat a dešifrovat zprávu.

Na začátku Bob zvolí dvě různá prvočísla p, q a spočte $N = pq$. Dále náhodně zvolí číslo e nesoudělné s $\varphi(N) = (p-1)(q-1)$ a pomocí Eukleidova algoritmu spočte číslo d splňující

$$de \equiv 1 \pmod{\varphi(N)}$$

(viz poznámka výše). Čísla N, e budou *veřejným klíčem* (ten Bob rozhlásí do světa), čísla d, p, q budou *soukromým klíčem* (ten bude držet v tajnosti).

Nyní popíšeme, jak může Alice poslat Bobovi zprávu. Pro jednoduchost budeme předpokládat, že zprávu tvoří nějaké přirozené číslo $0 < x < N$ nesoudělné s N. Alice vypočítá

$$y = x^e \bmod N$$

a výsledek pošle Bobovi. Bob, se znalostí soukromého klíče d, získá x výpočtem

$$x = y^d \bmod N.$$

Důkaz je snadný: platí $ed \equiv 1 \pmod{\varphi(N)}$, takže podle Eulerovy věty

$$y^d \equiv \left(x^e\right)^d = x^{ed} \equiv x^1 = x \pmod{N}.$$

Co vidí nepřítel? Hodnotu y a veřejný klíč N, e. Aby se dostal k hodnotě x, musel mít algoritmus, který z hodnoty $x^e \bmod N$ vypočte hodnotu x (tj. něco jako „e-tou odmocninu z x modulo N“). Očividným řešením je spočítat rozklad $N = pq$ a dále postupovat jako Bob, niceméně v současné době není znám algoritmus, který by uměl rozkládat velká čísla v rozumném čase (za dodržení jistých předpokladů, např. že rozklad neobsahuje mnoho prvočísel, že jsou tato prvočísla zhruba stejně velká atd.; při stavu současné techniky stačí volit prvočísla p, q s řádově tisíci binárních cifer). Žádný jiný efektivní postup na výpočet „odmocniny modulo N“ nebyl dosud nalezen.

2.3. **Čínská věta o zbytcích.**

Čínská věta o zbytcích hovoří o řešeních soustav lineárních kongruencí. Byla známa již starověkým Číňanům, je uvedena v knize Umění války autora Sun-c' ze 5. století př. n. l. Traduje se, že motivací čínské věty o zbytcích byl způsob, jakým Sun-c' počítal své vojáky. Věděl, že před bitvou měl 1000 vojáků, a chtěl je spočítat po bitvě. Velké množství lidí se špatně počítá, nechal je tedy seřadit do desetistupů, jedenáctistupů a třináctistupů a počítal, kolik mu jich zbyde mimo řady. Jinými slovy, zjistil, kolik je počet vojáků modulo 10, modulo 11 a modulo 13. Z čínské věty o zbytcích plyne, že z těchto zbytků lze jednoznačně zjistit celkový počet vojáků.

Věta 2.8 (čínská věta o zbytcích)**.** *Buď $m_1, \ldots, m_n$ po dvou nesoudělná přirozená čísla, označme $M = m_1 \cdot \ldots \cdot m_n$. Buď $u_1, \ldots, u_n$ libovolná celá čísla. Pak existuje právě jedno $x \in \{0, \ldots, M-1\}$, které řeší soustavu kongruencí*

$$x \equiv u_1 \pmod{m_1}, \quad \ldots, \quad x \equiv u_n \pmod{m_n}.$$

Důkaz. Nejprve dokážeme jednoznačnost řešení. Předpokládejme, že soustava má dvě řešení $x, y \in \{0, \ldots, M-1\}$, tj. pro každé i platí

$$x \equiv y \equiv u_i \pmod{m_i}.$$

Pak pro každé i

$$m_i \mid x - y$$

a protože jsou čísla m_i navzájem nesoudělná, dostáváme

$$M = m_1 \cdot \ldots \cdot m_n \mid x - y.$$

Ovšem obě čísla x, y, a tedy i jejich rozdíl, jsou menší než M, takže nutně $x - y = 0$, čili $x = y$.

Nyní dokážeme, že nějaké řešení vůbec existuje. Uvažujme zobrazení

$$f : \{0, \ldots, M-1\} \to \{0, \ldots, m_1 - 1\} \times \cdots \times \{0, \ldots, m_n - 1\}$$
$$x \mapsto (x \bmod m_1, \ldots, x \bmod m_n).$$

V předchozím odstavci jsme vlastně ukázali, že zobrazení f je prosté. Přitom definiční obor i obor hodnot této funkce mají stejnou velikost M (velikost kartézského součinu je součin velikostí činitelů), takže zobrazení f musí být podle Lemmatu 2.6 i na. Tedy ke každé n-tici $(u_1, \ldots, u_n)$ existuje právě jedno x, které se na něj zobrazuje, a to je hledaným řešením soustavy. □

Uvedený důkaz je zvláštní tím, že nedává žádný návod, jak řešení dané soustavy spočítat. Obecný postup (a tím i alternativní důkaz čínské věty o zbytcích) lze vypozorovat z řešení následující úlohy.

Úloha. Najděte všechna řešení soustavy kongruencí

$$x \equiv 2 \pmod 3, \quad x \equiv 1 \pmod 4, \quad x \equiv 3 \pmod 5.$$

Řešení. Z první rovnice vidíme, že $x = 3k + 2$, $k \in \mathbb{Z}$. Dosadíme do druhé rovnice a dostaneme $3k + 2 \equiv 1 \pmod 4$, tedy $k \equiv 1 \pmod 4$ a vidíme, že $k = 4l + 1$ a $x = 12l + 5$, $l \in \mathbb{Z}$. Dosadíme do třetí rovnice a dostaneme $12l + 5 \equiv 3 \pmod 5$, tedy $l \equiv 4 \pmod 5$, takže $l = 5m + 4$ a $x = 60m + 53$, $m \in \mathbb{Z}$. □

Čínská věta o zbytcích platí v mnohem obecnějším kontextu, později si ji ukážeme pro polynomy (sekce 8.1). Pro polynomy se jednoznačnost dokáže analogicky (horní mez je daná součtem stupňů polynomů m_i), ale s existencí je to o něco složitější, místo velikosti množiny je třeba argumentovat dimenzí jistého vektorového prostoru.

Čínská věta o zbytcích, pro čísla i pro polynomy, má zásadní aplikace ve výpočetní algebře: úlohu pro jeden „velký objekt“ (velké číslo, polynom velkého stupně) umožňuje rozdělit na větší množství úloh s „malými objekty“. To je výhodné nejen pro paralelizaci, ale také v tom, že pro modulární aritmetiku jsou k dispozici lepší algoritmy. Čtenáře odkazujeme např. do skript *Počítačová algebra*.

Na závěr pomocí čínské věty o zbytcích dokážeme vzorec na výpočet Eulerovy funkce, tj. vztah

$$\varphi(p_1^{k_1} \cdot \ldots \cdot p_m^{k_m}) = p_1^{k_1-1}(p_1 - 1) \cdot \ldots \cdot p_m^{k_m-1}(p_m - 1).$$

Důkaz Tvrzení 2.3. Dokážeme následující dvě vlastnosti:

(1) pro každé prvočíslo p platí $\varphi(p^k) = p^{k-1}(p-1)$;
(2) pro každá dvě nesoudělná čísla a, b platí $\varphi(ab) = \varphi(a) \cdot \varphi(b)$.

Uvedený vzorec snadno plyne z těchto dvou tvrzení: číslo n rozložíme na součin m po dvou nesoudělných mocnin $p_i^{k_i}$ a dostaneme

$$\varphi(n) \overset{(2)}{=} \varphi(p_1^{k_1}) \cdot \ldots \cdot \varphi(p_m^{k_m}) \overset{(1)}{=} p_1^{k_1-1}(p_1 - 1) \cdot \ldots \cdot p_m^{k_m-1}(p_m - 1).$$

(1) Zde je snadné spočítat *soudělná* čísla v intervalu $1, \ldots, p$: jsou to právě čísla $p, 2p, 3p, \ldots, p^{k-1} \cdot p$ a vidíme, že jich je p^{k-1}. Všechna zbylá čísla jsou nesoudělná, takže $\varphi(p^k) = p^k - p^{k-1} = p^{k-1}(p-1)$.

(2) Uvažujme zobrazení

$$f : \{0, \ldots, ab-1\} \to \{0, \ldots, a-1\} \times \{0, \ldots, b-1\}$$
$$x \mapsto (x \bmod a, x \bmod b).$$

Podle čínské věty o zbytcích je f bijekce. Dále uvažujme pouze restrikci f na množinu Φ_{ab}. To je prosté zobrazení, jehož definiční obor je množina Φ_{ab} velikosti $\varphi(ab)$. Stačí tedy dokázat, že jeho oborem hodnot je množina $\Phi_a \times \Phi_b$ – pak, díky prostosti, bude $\varphi(ab) = |\Phi_{ab}| = |\Phi_a \times \Phi_b| = |\Phi_a| \cdot |\Phi_b| = \varphi(a) \cdot \varphi(b)$, což chceme dokázat. Potřebujeme tedy ověřit, že

(a) f zobrazuje množinu Φ_{ab} do množiny $\Phi_a \times \Phi_b$, tj. že $\mathrm{NSD}(x, ab) = 1$ implikuje $\mathrm{NSD}(x \bmod a, a) = 1 = \mathrm{NSD}(x \bmod b, b)$;

(b) f zobrazuje množinu Φ_{ab} na tuto množinu, tj. že pokud $\mathrm{NSD}(u, a) = 1 = \mathrm{NSD}(v, b)$, pak to jediné x, které se zobrazuje na dvojici (u, v), splňuje $\mathrm{NSD}(x, ab) = 1$.

Obě části dokážeme zároveň: $\mathrm{NSD}(x, ab) = 1$ právě tehdy, když $\mathrm{NSD}(x, a) = 1 = \mathrm{NSD}(x, b)$ (uvažujte prvočíselného dělitele a použijte Lemma 1.5), což je právě tehdy, když $\mathrm{NSD}(x \bmod a, a) = 1 = \mathrm{NSD}(x \bmod b, b)$ použitím Lemmatu 1.3. □

Základy teorie polynomů

3. Tělesa a obory integrity

3.1. **Definice a příklady.**

Celá čísla nejsou jediným oborem, pro který má smysl studovat dělitelnost: typickým příkladem jsou polynomy, pro které možná už některé pojmy znáte (NSD, algoritmus dělení se zbytkem). Dělitelnost lze studovat v mnohem širším kontextu tak zvaných *oborů integrity*, mezi něž patří obor celých čísel, obory polynomů, ale také různá rozšíření celých čísel, např. *Gaussova celá čísla* (komplexní čísla s celočíselnými koeficienty).

V různých oborech pak platí různě silná tvrzení. Například existence a jednoznačnost prvočíselných rozkladů platí nejen pro celá čísla, ale také pro celočíselné polynomy či pro Gaussova celá čísla; přesto existují číselné obory, kde jednoznačnost neplatí. Eukleidův algoritmus na výpočet NSD založený na dělení se zbytkem funguje v oboru celých čísel, pro polynomy jedné proměnné nad tělesy, ale například pro polynomy více proměnných nastává s dělením problém. V sekci 7 si uděláme v uvedených vlastnostech pořádek. Nyní si definujeme matematické struktury, se kterými budeme dále pracovat. Nejobecnější strukturou je *komutativní okruh*, jehož axiomy vystihují základní aritmetické vlastnosti sčítání a násobení.

Definice. *Okruhem* $\mathbf{R}$ rozumíme pětici $(R, +, -, \cdot, 0)$, kde R je neprázdná množina, na které jsou definovány binární operace $+, \cdot$ (tj. zobrazení $R \times R \to R$), unární operace $-$ (tj. zobrazení $R \to R$) a prvek $0 \in R$ splňující pro každé $a, b, c \in R$ následující podmínky:

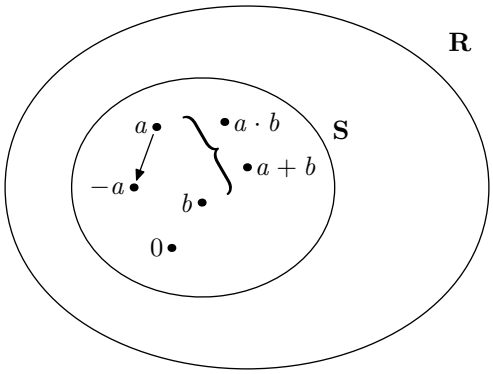
$$a + (b + c) = (a + b) + c, \qquad a + b = b + a, \qquad a + 0 = a,$$
$$a + (-a) = 0,$$
$$a \cdot (b \cdot c) = (a \cdot b) \cdot c,$$
$$a \cdot (b + c) = (a \cdot b) + (a \cdot c), \qquad (b + c) \cdot a = (b \cdot a) + (c \cdot a).$$

Okruh nazveme *komutativní*, pokud je komutativní také operace násobení, tj. $a \cdot b = b \cdot a$ pro všechna $a, b \in R$. *Okruhem s jednotkou* pak rozumíme okruh, ve kterém existuje prvek $1 \in R$ splňující $a \cdot 1 = a$ pro každé $a \in R$.

- Platí-li navíc podmínka $0 \neq 1$ a
$$\text{pokud } a, b \neq 0, \text{ pak } a \cdot b \neq 0,$$
nazýváme $\mathbf{R}$ *obor integrity*.
- Platí-li navíc podmínka $0 \neq 1$ a
$$\text{pro každé } a \neq 0 \text{ existuje } b \text{ splňující } a \cdot b = 1,$$
nazýváme $\mathbf{R}$ *těleso*. Z Tvrzení 3.2(5) plyne, že takové b je jednoznačně určeno, a značíme jej a^{-1}.

V zápise zpravidla vynecháváme závorky, násobení má vyšší prioritu než sčítání. Místo $a + (-b)$ píšeme $a - b$. Formálně odlišujeme mezi množinou R, tzv. *nosnou množinou*, a pěticí $\mathbf{R} = (R, +, -, \cdot, 0)$, která navíc obsahuje informaci o algebraické struktuře definované na R. Nebude-li výslovně uvedeno jinak, zápisem $\mathbf{R}$ rozumíme takto označenou pětici.

Příklad. Základní číselné obory $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ jsou komutativní okruhy vzhledem k standardním operacím sčítání, odčítání a násobení. Okruh $\mathbb{Z}$ je oborem integrity (ale nikoliv tělesem), okruhy $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ jsou tělesa. Další zajímavé příklady oborů integrity a těles se nacházejí mezi obory $\mathbb{Z}$ a $\mathbb{C}$, resp. mezi tělesy $\mathbb{Q}$ a $\mathbb{C}$, viz níže.



Obrázek 1. Podokruh **S** okruhu **R**.

Příklad. Důležitými příklady jsou konečné komutativní okruhy

$$\mathbb{Z}_n = (\{0, \ldots, n-1\}, +_{\bmod n}, -_{\bmod n}, \cdot_{\bmod n}, 0)$$

s operacemi modulo n. Všimněte si, že

$$\mathbb{Z}_n \text{ je těleso } \Leftrightarrow \mathbb{Z}_n \text{ je obor integrity } \Leftrightarrow n \text{ je prvočíslo.}$$

První implikace plyne z Tvrzení 3.3, každé těleso je oborem integrity. Druhá implikace: pokud by $n = k \cdot l$ bylo složené číslo, $k, l > 1$, pak by v $\mathbb{Z}_n$ platilo $k \cdot l = n \bmod n = 0$, čili by to nebyl obor integrity. A je-li n prvočíslo, pak je $a^{n-2} \bmod n$ inverzním prvkem pro $a \neq 0$, což plyne ihned z malé Fermatovy věty 2.5 (alternativně, inverzní prvek je koeficientem u z Bézoutovy rovnosti $1 = ua + vp$).

Příklad. Konečných komutativních okruhů je spousta, ale konečných těles mnoho není. Jak uvidíme v sekci 9.1, všechna konečná tělesa mají velikost mocniny prvočísla a pro každou mocninu prvočísla p^k existuje právě jedno konečné těleso s p^k prvky (právě jedno až na izomorfismus).

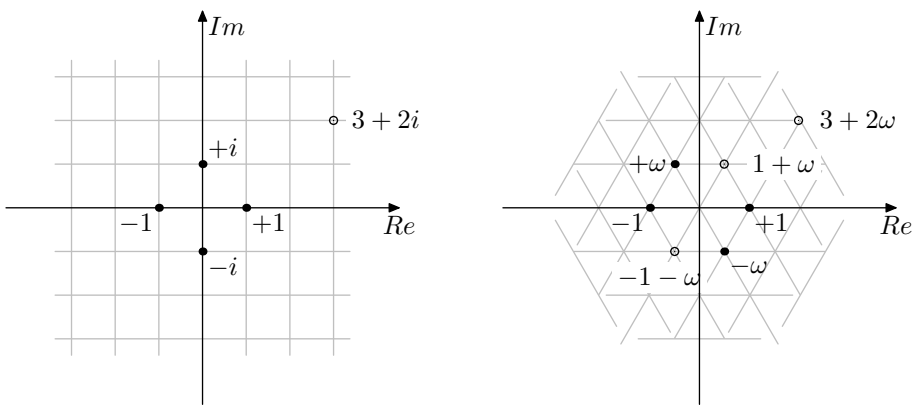
Příklad. Stěžejní roli hrají v algebře obory polynomů $\mathbf{R}[x]$, tj. formálních výrazů tvaru $a_0 + a_1 x + \ldots + a_n x^n$, kde koeficienty bereme z jistého komutativního okruhu **R**. Formální definici se dozvíte v sekci 4.

V tomto kurzu se budeme zabývat pouze komutativními okruhy. Přesto, některá tvrzení předpoklad komutativity nevyžadují a proto je dobré mít na paměti nějaký příklad.

Příklad. Základním příkladem nekomutativních okruhů jsou okruhy matic $\mathbf{M}_n(\mathbf{T})$ sestávající z čtvercových matic $n \times n$ nad tělesem **T** se standardními operacemi maticového sčítání a násobení.

K popisu oborů a těles mezi $\mathbb{Z}$ a $\mathbb{C}$ se hodí pojem podokruhu a podtělesa.

Definice. Buď **R** komutativní okruh a $S \subseteq R$ podmnožina taková, že $0 \in S$ a kdykoliv $a, b \in S$, pak také $-a \in S$, $a + b \in S$ a $a \cdot b \in S$ (říkáme, že množina S je *uzavřená* na uvedené operace). Vezmeme-li na množině S restrikce operací okruhu **R**, dostaneme také komutativní okruh, který označíme **S** (jsou-li všechny axiomy splněny na větší množině R, pak jistě i na její podmnožině S). Výsledky této konstrukce nazýváme *podokruhy* komutativního okruhu **R**, značíme $\mathbf{S} \leq \mathbf{R}$.



Obrázek 2. Gaussova a Eisensteinova celá čísla.

Je-li $\mathbf{R}$ obor integrity a platí-li navíc $1 \in S$, hovoříme o podoboru. Je-li $\mathbf{R}$ těleso a platí-li navíc podmínka $0 \neq a \in S \ \Rightarrow a^{-1} \in S$, hovoříme o *podtělesu*.

Příklad. Obor $\mathbb{Z}$ je podoborem tělesa $\mathbb{Q}$, které je podtělesem tělesa $\mathbb{R}$, které je podtělesem tělesa $\mathbb{C}$.

V algebraické teorii čísel se studují podobory a podtělesa tělesa $\mathbb{C}$, která vznikají „přidáním konečně mnoha prvků“ k oboru $\mathbb{Z}$, resp. tělesu $\mathbb{Q}$.

Příklad (Gaussova celá a racionální čísla)**.**

- Množina $\mathbb{Z}[i] = \{a + bi : \ a, b \in \mathbb{Z}\}$ tvoří podobor tělesa $\mathbb{C}$, zvaný *Gaussova celá čísla*. Uzavřenost na odčítání plyne ze vztahu $-(a + bi) = -a - bi$, na sčítání ze vztahu $(a + bi) + (c + di) = (a + c) + (b + d)i$ a na násobení ze vztahu $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$.
- Množina $\mathbb{Q}(i) = \{a+bi : \ a, b \in \mathbb{Q}\}$ tvoří podtěleso tělesa $\mathbb{C}$, zvané *Gaussova racionální čísla*. Je třeba navíc ověřit, že $(a + bi)^{-1} \in \mathbb{Q}(i)$, což plyne ze vztahu $\frac{1}{a+bi} = \frac{1}{a+bi} \cdot \frac{a-bi}{a-bi} = \frac{a}{a^2+b^2} + \frac{b}{a^2+b^2} i$.

Příklad (kvadratická rozšíření)**.** Obecněji, pro libovolné celé číslo s uvažujme *kvadratické rozšíření*

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

Není těžké nahlédnout, že množina na pravé straně je skutečně podoborem, resp. podtělesem, tělesa $\mathbb{C}$ (i když pozor, fakt, že je $\mathbb{Q}[\sqrt{s}]$ uzavřena na inverzy, není očividný!). V závislosti na s mají obory $\mathbb{Z}[\sqrt{s}]$ různé vlastnosti z hlediska dělitelnosti a budeme je občas používat jako protipříklady (například v oboru $\mathbb{Z}[\sqrt{5}]$ nejsou jednoznačné rozklady na prvočinitele).

Příklad (Eisensteinova čísla)**.** Množina $\mathbb{Z}[\omega] = \{a + b\omega : \ a, b \in \mathbb{Z}\}$, kde $\omega = e^{2\pi i/3}$ je komplexní třetí odmocnina z jedné, tvoří podokruh oboru $\mathbb{C}$, zvaný *Eisensteinova celá čísla*. Všimněte si, že $\omega^2 = -1 - \omega$, čili uzavřenost na násobení plyne ze vztahu $(a + b\omega)(c + d\omega) = ac + (ad + bc)\omega + bd\omega^2 = (ac - bd) + (ad + bc - bd)\omega$.

3.2. **Základní vlastnosti.**

V moderní matematice je zvykem definovat abstraktní struktury pomocí co nejmenší sady axiomů. V této části si odvodíme několik jednoduchých vlastností, které plynou z axiomů komutativních okruhů a v dalším textu je budeme zcela automaticky používat.

Tvrzení 3.1. *Buď $*$ asociativní operace na množině X a $a_1, \ldots, a_n \in X$. Hodnota výrazu $a_1 * \ldots * a_n$ nezávisí na uzávorkování.*

Důkaz. Těžší cvičení pro abnormálně zvídavé studenty. □

Tvrzení 3.2 (základní vlastnosti okruhů)**.** *Buď $\mathbf{R}$ okruh, $a, b, c \in R$. Pak*

(1) *pokud $a + c = b + c$, pak $a = b$;*
(2) $a \cdot 0 = 0$*;*
(3) $-(-a) = a, \quad -(a + b) = -a - b$*;*
(4) $-(a \cdot b) = (-a) \cdot b = a \cdot (-b), \quad (-a) \cdot (-b) = ab$*;*
(5) *je-li $\mathbf{R}$ oborem integrity, pokud $a \cdot c = b \cdot c$ a $c \neq 0$, pak $a = b$.*

Důkaz. (1) Je-li $a+c = b+c$, pak také $(a+c)+(-c) = (b+c)+(-c)$. Použitím axiomů dostaneme $(a+c)+(-c) = a+(c+(-c)) = a+0 = a$ a podobně $(b+c)+(-c) = b$, tedy $a = b$.

(2) Pomocí distributivity spočteme $0 + a \cdot 0 = a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$ a z (1) dostáváme $a \cdot 0 = 0$.

(3) Protože $0 = a + (-a) = -(-a) + (-a)$, z (1) dostáváme $a = -(-a)$. Protože $0 = (a + b) + (-(a + b))$ a zároveň $0 = a + (-a) + b + (-b) = (a + b) + (-a - b)$, z (1) dostáváme $-(a + b) = -a - b$.

(4) Protože $a \cdot b + (-a) \cdot b = (a + (-a)) \cdot b = 0 \cdot b = 0 = a \cdot b + (-(a \cdot b))$, z (1) dostáváme $-(a \cdot b) = (-a) \cdot b$. Druhou rovnost dokážeme analogicky a užitím předchozího $(-a) \cdot (-b) = -(a \cdot (-b)) = -(-(a \cdot b)) = a \cdot b$.

(5) Protože $a \cdot c = b \cdot c$, platí $0 = a \cdot c - b \cdot c = (a - b) \cdot c$. Tedy aspoň jeden z prvků c, $a - b$ musí být 0. Protože předpokládáme $c \neq 0$, musí být $a - b = 0$, tedy $a = b$. □

Tvrzení 3.3. *Každé těleso je oborem integrity.*

Důkaz. Kdyby existovaly $a, b \neq 0$ takové, že $a \cdot b = 0$, pak

$$b = (a^{-1} \cdot a) \cdot b = a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 = 0,$$

což je spor. V posledním kroku jsme použili Tvrzení 3.2(2)! □

3.3. **Podílová tělesa.**

Obor celých čísel lze přirozeně rozšířit do tělesa racionálních čísel způsobem, který je znám jako zlomky. Tuto myšlenku lze zobecnit na libovolný obor integrity, výsledkem bude tzv. *podílové těleso* tohoto oboru. Podílová tělesa se hodí v situacích, kdy dělení zjednoduší zadanou úlohu. Například, abychom k výpočtu NSD dvou polynomů mohli použít Eukleidův algoritmus.

Konstrukce. Buď $\mathbf{R}$ obor integrity a $M = R \setminus \{0\}$. Definujeme relaci $\sim$ na množině $R \times M$ předpisem

$$(a, b) \sim (c, d) \quad \iff \quad ad = bc.$$

Není těžké nahlédnout, že jde o ekvivalenci: reflexivita je zřejmá, symetrie plyne z komutativity násobení a tranzitivitu získáme následujícím výpočtem: je-li $(a, b) \sim$ $\sim (c, d) \sim (e, f)$, tedy $ad = bc$ a $cf = de$. Pak ale $adf = bcf = bde$, a krácením

prvkem $d \neq 0$ dostaneme $af = be$ (ke krácení potřebujeme předpoklad, že $\mathbf{R}$ je obor integrity!).

Blok $[(a,b)]_\sim$ této ekvivalence budeme nazývat *zlomek* a značit jej $\frac{a}{b}$. Uvažujme množinu Q všech zlomků a definujme na ní operace a konstanty

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}, \qquad -\frac{a}{b} = \frac{-a}{b}, \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}, \qquad 0 = \frac{0}{1}, \qquad 1 = \frac{1}{1}.$$

Jsou tyto operace dobře definované? Předně, aby jmenovatel součtu a součinu zůstal v M, potřebujeme, aby součin nenulových prvků byl nenulový, tj. aby byl R oborem integrity. Dále musíme dokázat, že pokud zvolíme jiné reprezentanty zlomků, výsledek operace zůstane stejný. Formálně, pokud $\frac{a}{b} = \frac{a'}{b'}$ a $\frac{c}{d} = \frac{c'}{d'}$, potřebujeme dokázat, že $\frac{a}{b} + \frac{c}{d} = \frac{a'}{b'} + \frac{c'}{d'}$, a podobně pro odčítání a násobení. Pro sčítání potřebujeme ověřit, že $\frac{ad+bc}{bd} = \frac{a'd'+b'c'}{b'd'}$, tedy že $(ad+bc)(b'd') = (a'd'+b'c')(bd)$. Roznásobíme a využijeme faktu, že $ab' = a'b$ a $cd' = c'd$. Odčítání a násobení se ověří podobně.

Definice. Výsledná struktura $\mathbf{Q} = (Q, +, -, \cdot, 0)$ z předchozí konstrukce se nazývá *podílové těleso* oboru $\mathbf{R}$.

Musíme ovšem dokázat, že to je skutečně těleso.

Tvrzení 3.4. *Buď $\mathbf{R}$ obor integrity a $\mathbf{Q}$ výsledek právě popsané konstrukce. Pak $\mathbf{Q}$ je těleso.*

Důkaz. Ověříme postupně všechny axiomy:

- Asociativita sčítání: $\frac{a}{b}+(\frac{c}{d}+\frac{e}{f}) = \frac{a}{b}+\frac{cf+de}{df} = \frac{adf+b(cf+de)}{bdf} = \frac{adf+bcf+bde}{bdf} = = \frac{ad+bc}{bd} + \frac{e}{f} = (\frac{a}{b} + \frac{c}{d}) + \frac{e}{f}$.
- Komutativita sčítání: $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd} = \frac{cb+da}{db} = \frac{c}{d} + \frac{a}{b}$.
- Nula: $\frac{a}{b} + \frac{0}{1} = \frac{a\cdot 1+b\cdot 0}{b\cdot 1} = \frac{a}{b}$.
- Odčítání: $\frac{a}{b} + \frac{-a}{b} = \frac{ab+(-ab)}{b^2} = \frac{0}{b^2} = 0$.
- Asociativita a komutativita násobení plyne okamžitě z týchž vlastností oboru $\mathbf{R}$.
- Jednotka: $\frac{a}{a} \cdot \frac{1}{1} = \frac{a\cdot 1}{a\cdot 1} = \frac{a}{a}$.
- Distributivita: $\frac{a}{b} \cdot (\frac{c}{d} + \frac{e}{f}) = \frac{acf+ade}{bdf} = \frac{abcf+abde}{b^2df} = \frac{ac}{bd} + \frac{ae}{bf}$.
- $0 = \frac{0}{1} \neq 1 = \frac{1}{1}$, protože $0 \cdot 1 \neq 1 \cdot 1$.
- Součin nenulových prvků: pokud $0 = \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$, pak $a = 0$ nebo $c = 0$, protože $\mathbf{R}$ je obor integrity.
- Inverz: Všimněte si, že $\frac{a}{b} = 0 = \frac{0}{1}$ právě tehdy, když $a \cdot 1 = b \cdot 0$, čili když $a = 0$. Tedy pro každé $\frac{a}{b} \neq 0$ platí $\frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ba} = \frac{1}{1}$.

□

Příklad. Těleso racionálních čísel $\mathbb{Q}$ je *definováno* jako podílové těleso oboru $\mathbb{Z}$.

4. Polynomy

4.1. Obory polynomů.

Stěžejním objektem v algebře komutativních okruhů jsou polynomy. V této sekci si ukážeme základní vlastnosti polynomů týkající se dělitelnosti a kořenů. Začneme ovšem definicí polynomu a polynomiálního okruhu.

V celé sekci bude $\mathbf{R}$ značit nějaký komutativní okruh s jednotkou.

Definice. *Polynomem proměnné x* nad okruhem $\mathbf{R}$ rozumíme výraz

$$a_0 + a_1x + a_2x^2 + \ldots + a_nx^n,$$

nebo zkráceně

$$\sum_{i=0}^{n} a_ix^i,$$

kde $a_0, \ldots, a_n \in R$ a $a_n \neq 0$. Prvky $a_0, \ldots, a_n$ nazýváme *koeficienty* a symbol x *proměnná*. (Implicitně se rozumí se $a_m = 0$ pro všechna $m > n$.) Číslo n nazýváme *stupeň polynomu*, značíme $\deg f$. Prvek a_n se nazývá *vedoucí koeficient* a a_0 *absolutní člen*. Polynom se nazývá *monický*, pokud je vedoucí člen 1. Je třeba speciálně dodefinovat *nulový polynom*; pro něj položíme $\deg 0 = -1$.

Na množině všech polynomů definujeme operace předpisy

$$\sum_{i=0}^{m} a_ix^i \; + \; \sum_{i=0}^{n} b_ix^i = \sum_{i=0}^{\max(m,n)} (a_i + b_i)x^i, \qquad - \sum_{i=0}^{m} a_ix^i = \sum_{i=0}^{m} (-a_i)x^i,$$

$$(\sum_{i=0}^{m} a_ix^i) \cdot (\sum_{i=0}^{n} b_ix^i) = \sum_{i=0}^{m+n} \big(\sum_{j+k=i} a_jb_k \big) x^i.$$

Jak si nyní ukážeme, dostaneme opět komutativní okruh, který budeme značit $\mathbf{R}[x]$.

Tvrzení 4.1. *Buď $\mathbf{R}$ komutativní okruh s jednotkou. Pak*

(1) *$\mathbf{R}[x]$ je komutativní okruh s jednotkou;*
(2) *je-li $\mathbf{R}$ oborem integrity, pak $\mathbf{R}[x]$ je také obor integrity a platí $\deg(fg) = \deg f + \deg g$ pro každé dva polynomy $f, g \neq 0$.*

Důkaz. Označme $f = \sum_{i=0}^{m} a_ix^i$, $g = \sum_{i=0}^{n} b_ix^i$, $h = \sum_{i=0}^{p} c_ix^i$. (1) Ověříme postupně všechny axiomy:

- Axiomy pro sčítání jsou očividné, sčítají se nezávisle koeficienty u jednotlivých mocnin, čili rovnosti pro polynomy ihned plynou z rovností v $\mathbf{R}$.
- Komutativita násobení plyne z toho, že vzorec je symetrický vzhledem k prohození písmen a a b.
- Jednotka: z definice součinu

  $$f \cdot 1 = (\sum_{i=0}^{n} a_ix^i) \cdot (1 + 0 + 0 + \ldots) = \sum_{i=0}^{n} (\sum_{j+k=i} a_jb_k) x^i,$$

  kde všechny b_i kromě b_0 jsou nulové, takže výsledkem je opět polynom f.
- Asociativita násobení: z jedné strany, $f \cdot (g \cdot h)$ je rovno

$$\left(\sum_{i=0}^{m} a_ix^i\right) \cdot \left(\left(\sum_{i=0}^{n} b_ix^i\right) \cdot \left(\sum_{i=0}^{p} c_ix^i\right)\right) = \left(\sum_{i=0}^{m} a_ix^i\right) \cdot \left(\sum_{i=0}^{n+p} \left(\sum_{k+l=i} b_kc_l\right) x^i\right)$$

$$= \sum_{i=0}^{m+n+p} \left(\sum_{j+k+l=i} a_jb_kc_l\right) x^i,$$

  a je vidět, že stejně vyjde i výpočet součinu $(f \cdot g) \cdot h$.
- Distributivita se prověří podobně, viz cvičení.

(2) Vedoucím koeficientem polynomu $f \cdot g$ je prvek a_mb_n, který je nenulový díky tomu, že $\mathbf{R}$ je oborem integrity. □

Pro libovolné polynomy f, g platí $\deg(f+g) \leq \max(\deg f, \deg g)$, ale rovnost nastat nemusí, například pro $g = -f$. Pokud $\mathbf{R}$ není oborem integrity, stupeň součinu nemusí být součtem stupňů, například v $\mathbb{Z}_4[x]$ platí $(2x+1)\cdot(2x+1) = 1$.

Induktivně lze definovat *obory polynomů více proměnných*: polynomem v proměnných $x_1, \ldots, x_m$ se rozumí polynom v proměnné x_m, jehož koeficienty jsou polynomy v proměnných $x_1, \ldots, x_{m-1}$. Ve zkratce,

$$\mathbf{R}[x_1, \ldots, x_m] = (\mathbf{R}[x_1, \ldots, x_{m-1}])[x_m].$$

Několikanásobnou aplikací Tvrzení 4.1 dostaneme, že polynomy více proměnných nad oborem integrity tvoří obor integrity. Díky distributivitě můžeme libovolný polynom více proměnných přepsat právě jedním způsobem do tzv. *distribuovaného tvaru*

$$\sum_{k_1,\ldots,k_m=0}^{n} a_{k_1,\ldots,k_m} x_1^{k_1} \cdots x_m^{k_m}$$

s koeficienty $a_{k_1,\ldots,k_m} \in R$. Alternativně bychom mohli obory plynomů více proměnných definovat tímto způsobem, ale museli bychom dokázat analogii Tvrzení 4.1 pro více proměnných.

4.2. **Hodnota polynomu a polynomiální zobrazení.**

Definice. Buď $\mathbf{R} \leq \mathbf{S}$ obory integrity a uvažujme polynom

$$f = a_0 + a_1 x + \ldots + a_n x^n \in R[x]$$

a prvek $u \in S$. *Hodnota polynomu* f po dosazení u je definována přepisem

$$f(u) = a_0 + a_1 u + \ldots + a_n u^n \in S,$$

přičemž v uvedeném vzorci provádíme všechny operace (mocnění, násobení i sčítání) v oboru $\mathbf{S}$. Zobrazení

$$S \to S, \qquad u \to f(u)$$

se nazývá *polynomiální zobrazení* dané polynomem f.

Například pro $\mathbf{R} = \mathbb{Z}$, $\mathbf{S} = \mathbb{C}$, $f = x^2 + x + 1$ a $u = i$ máme $f(i) = i$. Příslušné polynomiální zobrazení je dáno předpisem $u \mapsto u^2 + u + 1$ pro $u \in \mathbb{C}$.

Je třeba striktně rozlišovat mezi polynomem jako *výrazem* (tj. jeho zápisem ve formě „vzorečku") a odvozeným *polynomiálním zobrazením*, daným hodnotami po dosazení. Pozor, různé polynomy mohou dávat stejná polynomiální zobrazení! Například pro polynom $f = x^p \in \mathbb{Z}_p[x]$ platí díky malé Fermatově větě $f(u) = u$ pro každé $u \in \mathbb{Z}_p$, a tedy určuje stejné polynomiální zobrazení jako polynom $g = x$. (Jinak to pro konečné obory být ani nemůže, protože existuje nekonečně mnoho polynomů, ale pouze konečně mnoho zobrazení na konečné množině.)

4.3. **Dělení polynomů se zbytkem.**

Buď f, g polynomy z $\mathbf{R}[x]$. Řekneme, že g *dělí* f, píšeme $g \mid f$, pokud existuje polynom $h \in R[x]$ takový, že $f = gh$. Je-li $\mathbf{R}$ obor integrity a $g \mid f \neq 0$, pak $\deg g \leq \deg f$ díky Tvrzení 4.1. Pokud g nedělí f, má smysl se ptát po zbytku po dělení.

Tvrzení 4.2 (dělení polynomů se zbytkem)**.** *Buď* $\mathbf{R}$ *obor integrity,* $\mathbf{Q}$ *jeho podílové těleso,* $f, g \in R[x]$, $g \neq 0$. *Pak existuje právě jedna dvojice* $q, r \in Q[x]$ *splňující*

$$f = gq + r \quad a \quad \deg r < \deg g.$$

Navíc, je-li g *monický, pak* $q, r \in R[x]$.

Díky jednoznačnosti můžeme definovat $f \operatorname{div} g = q$ a $f \bmod g = r$. Je vidět, že $g \mid f$ právě tehdy, když $f \bmod g = 0$.

Důkaz. Existenci dokážeme tak, že popíšeme algoritmus, který podíl a zbytek najde. Na začátku vezmeme $q_0 = 0$, $r_0 = f$, a poté definujeme rekurzivně

$$q_{i+1} = q_i + \frac{l(r_i)}{l(g)} \cdot x^{\deg r_i - \deg g}, \qquad r_{i+1} = r_i - \frac{l(r_i)}{l(g)} \cdot x^{\deg r_i - \deg g} \cdot g,$$

kde $l(u)$ značí vedoucí koeficient polynomu u. Rekurzí pokračujeme do té doby, než bude $\deg r_i$ menší než $\deg g$. To jistě někdy nastane, protože v každém kroku zmenšíme stupeň zbytku, tj. $\deg r_{i+1} < \deg r_i$ pro včechna i. Indukcí snadno ověříme, že vztah $f = gq_i + r_i$ platí pro každé i, a tedy poslední dvojice q_i, r_i je hledaným podílem a zbytkem.

Z algoritmu je vidět, že je-li g monický, žádné zlomky se neobjeví a výsledkem budou polynomy z $\mathbf{R}[x]$.

Na závěr dokážeme jednoznačnost. Kdyby $f = gq_1 + r_1 = gq_2 + r_2$, pak $g(q_1 - q_2) = r_2 - r_1$, tedy $g \mid r_2 - r_1$. Přitom $\deg(r_2 - r_1) < \deg g$, tedy $r_2 - r_1 = 0$, čili $r_1 = r_2$. Z toho ihned plyne $q_1 = q_2$, protože $g \neq 0$ a jsme v oboru integrity. □

4.4. **Kořeny a dělitelnost.**

Definice. Buď $\mathbf{R} \leq \mathbf{S}$ obory integrity, $f \in R[x]$ a $a \in S$. Řekneme, že a je *kořen* polynomu f, pokud $f(a) = 0$.

Například $i \in \mathbb{C}$ je kořenem polynomu $x^2 + 1 \in \mathbb{Z}[x]$ v tělese $\mathbb{C}$. Ukážeme si, jak existence kořene souvisí s děliteli daného polynomu.

Tvrzení 4.3. *Buď $\mathbf{R}$ obor integrity, $f \in R[x]$ a $a \in R$. Pak a je kořen polynomu f právě tehdy, když $x - a \mid f$.*

Důkaz. ($\Leftarrow$) Předpokládejme, že $x - a \mid f$. Pak $f = (x - a) \cdot g$ pro nějaké $g \in R[x]$ a dosadíme-li do f prvek a, dostaneme

$$f(a) = (a - a) \cdot g(a) = 0 \cdot g(a) = 0.$$

($\Rightarrow$) Buďte $q, r \in R[x]$ podíl a zbytek při dělení polynomu f polynomem $x - a$ (dělíme monickým polynomem, čili nepotřebujeme podílové těleso). Tedy $f = (x - a) \cdot q + r$ a $\deg r < \deg(x - a) = 1$, čili r je konstantní polynom. Dosadíme-li prvek a, dostaneme

$$0 = f(a) = (a - a) \cdot q(a) + r(a) = 0 \cdot q(a) + r = r,$$

takže $r = 0$ a $x - a \mid f$. □

Z důkazu je vidět jedno důležité pozorování: je-li $f \in R[x]$ a $a \in R$, pak

$$f \bmod x - a = f(a).$$

Věta 4.4 (počet kořenů polynomu)**.** *Buď $\mathbf{R}$ obor integrity, $0 \neq f \in R[x]$ a $\deg f = n$. Pak má polynom f nejvýše n kořenů v $\mathbf{R}$.*

Důkaz. Budeme postupovat indukcí podle stupně polynomu f. Je-li $\deg f = 0$, tj. f je nenulový konstantní polynom, pak žádné kořeny nemá. Nyní předpokládejme, že tvrzení platí pro všechny polynomy stupně nejvýše n. Je-li $\deg f = n + 1$, pak jsou dvě možnosti. Buď polynom f nemá žádný kořen, v tom případě tvrzení platí. Nebo má polynom f nějaký kořen a a v tom případě jej lze podle předchozího lemmatu napsat jako $f = (x - a) \cdot g$ pro nějaký polynom g stupně n. Je-li b nějaký

jiný kořen, tj. $f(b) = (b-a)\cdot g(b) = 0$, pak, protože jde o obor integrity, musí být buď $b = a$ nebo $g(b) = 0$. Protože má polynom g nejvýše n kořenů, má polynom f nejvýše $n+1$ kořenů. □

Počet kořenů polynomu f samozřejmě může být menší než $\deg f$: např. polynom x^2+1 nemá nad $\mathbb{Z}$ žádný kořen a nad $\mathbb{Z}_2$ má jeden.

Poznámka. Věta 4.4 neplatí, není-li $\mathbf{R}$ oborem integrity, ale třeba jen komutativním okruhem. Předpoklad jsme použili v poslední fázi důkazu, když z $f(b) = (b-a)\cdot g(b) = 0$ plynulo $b-a = 0$ nebo $g(b) = 0$. Například polynom $2x \in \mathbb{Z}_4[x]$ má v $\mathbb{Z}_4$ dva kořeny $0, 2$ a polynom $x^2+x \in \mathbb{Z}_6[x]$ má v $\mathbb{Z}_6$ čtyři kořeny $0, 2, 3, 5$.

Poznámka. Věta 4.4 neplatí ani pro nekomutativní tělesa. Například v kvaternionech má polynom x^2+1 šest kořenů, $\pm i$, $\pm j$, $\pm k$.

5. Základní pojmy dělitelnosti

V této sekci definujeme základní pojmy jako dělitelnost, asociovanost, největší společný dělitel a ireducibilní rozklady. Tyto pojmy definujeme pro obecný obor integrity $\mathbf{R}$ a budeme je ilustrovat v následujících konkrétních případech: pro tělesa, pro obor $\mathbb{Z}$, pro obory polynomů a v některých případech pro kvadratická rozšíření celých čísel.

5.1. Dělitelnost a asociovanost.

Definice. Řekneme, že a *dělí* b v oboru $\mathbf{R}$ a píšeme $a \mid b$, pokud existuje $c \in R$ takové, že $b = ac$.

Pozor, při použití symbolu pro dělitelnost, i jakéhokoliv odvozeného pojmu, musíme vždy uvést (anebo musí být z kontextu zřejmé), v jakém oboru pracujeme! Například

- $3x+6 \mid x+2$ v oboru $\mathbb{Q}[x]$, protože $x+2 = \frac{1}{3}\cdot(3x+6)$; ale
- $3x+6 \nmid x+2$ v oboru $\mathbb{Z}[x]$, protože neexistuje $f \in \mathbb{Z}[x]$ splňující $x+2 = f\cdot(3x+6)$.

Definice. Řekneme, že prvky a a b jsou *asociované* a píšeme $a \parallel b$, pokud $a \mid b$ a $b \mid a$. Prvek a je invertibilní právě tehdy, když $a \parallel 1$. Prvek b splňující $ab = 1$ značíme a^{-1}.

Všimněte si, že relace dělitelnosti je reflexivní a tranzitivní: pokud $a \mid b$ a $b \mid c$, tedy pokud $b = ax$ a $c = by$ pro nějaká x, y, pak $c = axy$, tedy $a \mid c$. Z toho ihned plyne, že relace $\parallel$ je ekvivalencí.

Tvrzení 5.1 (asociovanost vs. invertibilní prvky)**.** *Buď $\mathbf{R}$ obor integrity a $a, b \in R$. Pak $a \parallel b$ právě tehdy, když existuje invertibilní prvek $q \in R$ takový, že $a = bq$.*

Důkaz. ($\Leftarrow$) Protože $a = bq$, platí $b \mid a$. Protože $b = aq^{-1}$, platí $a \mid b$.

($\Rightarrow$) Pokud $a = 0$, pak $b = 0$ a tvrzení platí. Uvažujme $a \neq 0$. Protože $b \mid a$, můžeme psát $a = bu$, a protože $a \mid b$, můžeme psát $b = av$, pro nějaká u, v. Tedy $a = bu = avu$ a krácením dostáváme $uv = 1$, čili $u, v \parallel 1$. □

Příklady.

- V tělese je každý nenulový prvek invertibilní. Tedy $a \parallel b$ pro každé $a, b \neq 0$.
- V oboru $\mathbb{Z}$ jsou invertibilní pouze prvky ± 1. Tedy $a \parallel b$ právě tehdy, když $a = \pm b$.

- V oboru $\mathbf{R}[x]$ jsou invertibilní právě polynomy stupně 0, jejichž koeficient je invertibilní v oboru $\mathbf{R}$. Nad tělesem jsou tedy invertibilní všechny nenulové polynomy stupně 0. Například v $\mathbb{Z}[x]$ je $f \parallel g$ právě tehdy, když $g = \pm f$, zatímco v $\mathbb{Q}[x]$ je $f \parallel g$ právě tehdy, když $g = cf$ pro nějaké $0 \neq c \in \mathbb{Q}$.

V obecných oborech integrity není možné definovat dělení se zbytkem, protože nemáme způsob, jak vyjádřit, že zbytek má být menší než dělitel. Přesto má smysl zavést symbol kongruence podmínkou

$$a \equiv b \pmod{m} \quad \iff \quad m \mid a - b.$$

Stejně jako pro čísla je snadné dokázat, že jde o ekvivalenci invariantní vzhledem k okruhovým operacím (důkaz Tvrzení 2.1 projde téměř doslova). S krácením je to složitější, k jeho důkazu jsme potřebovali prvočíselné rozklady; pohledem na důkaz je vidět, že analogie Tvrzení 2.2 platí v tzv. gaussovských oborech $\mathbf{R}$ (viz sekce 5.4).

5.2. **Největší společný dělitel.**

Definice. Řekneme, že $c = \mathrm{NSD}(a, b)$ (*největší společný dělitel*), pokud

(1) $c \mid a$ a $c \mid b$ (tj. c je společný dělitel);
(2) kdykoliv $d \mid a$ a $d \mid b$, pak $d \mid c$ (tj. c je největší takový).

(Všimněte si, že definice $\mathrm{NSD}(a, b)$ odpovídá definici infima množiny $\{a, b\}$ vzhledem k relaci $|$.) Prvky a, b nazýváme *nesoudělné*, pokud $\mathrm{NSD}(a, b) = 1$.

Operátor NSN označující *nejmenší společný násobek* se definuje analogicky.

S definicí NSD je potřeba zacházet opatrně: hodnota c není určena jednoznačně. Například, v oboru $\mathbb{Z}$ bude platit $\mathrm{NSD}(4, 6) = 2$, ale také $\mathrm{NSD}(4, 6) = -2$, obě čísla splňují definici. Na NSD je potřeba nahlížet jako na relaci „c vyhovuje definici největšího společného dělitele prvků a, b“.

Situace naštěstí není tak špatná: NSD je určen jednoznačně *až na asociovanost*. Z jedné strany, pokud $\mathrm{NSD}(a, b) = c_1$ a $\mathrm{NSD}(a, b) = c_2$, pak c_1 i c_2 jsou společní dělitelé a, b, a tedy oba musí být největší, tj. $c_1 \mid c_2$ a zároveň $c_2 \mid c_1$, tedy $c_1 \parallel c_2$. Na druhou stranu, pokud $\mathrm{NSD}(a, b) = c_1$ a $c_1 \parallel c_2$, pak c_2 jistě také splňuje podmínky největšího společného dělitele.

Další problém spočívá v tom, že existence NSD není ničím garantovaná. A skutečně, jsou obory, v nichž NSD pro některé dvojice prvků neexistuje. Například v oboru $\mathbb{Z}[\sqrt{5}]$ neexistuje NSD prvků

$$u = 4, \quad v = 2 + 2\sqrt{5}.$$

Čísla $r = 2$ a $s = 1 + \sqrt{5}$ jsou určitě společnými děliteli prvků u, v, protože $u = 2 \cdot 2 = (1 + \sqrt{5})(1 - \sqrt{5})$ a $v = 2 \cdot (1 + \sqrt{5})$, ale ani jeden není největší (tj. $r \nmid s$ a $s \nmid r$) a dá se dokázt, že neexistuje žádný prvek z, pro který by platilo $r, s \mid z$ a $z \mid u, v$.

5.3. **Ireducibilní prvky a rozklady.**

Pro každý prvek a platí, že $1 \mid a$ a $a \mid a$. Dělitel prvku a se nazývá *vlastní*, jestliže není asociovaný ani s 1, ani s a.

Definice. Prvek a se nazývá *ireducibilní*, pokud $a \neq 0$, $a \nparallel 1$ a a nemá vlastní dělitele. Jinými slovy, pokud pro každý rozklad $a = bc$ platí $b \parallel 1$ nebo $c \parallel 1$.

Příklad.

- V tělesech žádné ireducibilní prvky nejsou.
- V oboru $\mathbb{Z}$ jsou ireducibilní právě čísla $\pm p$, kde p je prvočíslo.

V oborech polynomů není obecně snadné určit, které polynomy jsou ireducibilní. V oboru $\mathbf{R}[x]$ jsou ireducibilní například

- ty polynomy stupně 0, které jsou ireducibilní jako prvky $\mathbf{R}$;
- ty polynomy stupně 1, které nejsou dělitelné žádným neinvertibilním prvkem $\mathbf{R}$ (např. polynom $2x + 2$ je ireducibilní v $\mathbb{Q}[x]$, ale nikoliv v $\mathbb{Z}[x]$).

Je-li f polynom z $\mathbf{R}[x]$ stupně ≥ 2, který má kořen $a \in R$, pak nemůže být ireducibilní, protože má vlastního dělitele $x - a$ (Tvrzení 4.3). Pozor, opačná implikace neplatí: např. polynom $x^4 + 2x^2 + 1$ je rozložitelný v oboru $\mathbb{Z}[x]$, přestože tam nemá kořen. Pro polynomy vyšších stupňů není žádné obecné pravidlo.

Příklad.

- V oboru $\mathbb{C}[x]$ jsou ireducibilní právě polynomy stupně 1, jak praví tzv. základní věta algebry.
- V oboru $\mathbb{R}[x]$ jsou ireducibilní právě polynomy stupně 1 a ty polynomy stupně 2, které nemají reálný kořen (viz cvičení).
- V oboru $\mathbb{Q}[x]$ jsou ireducibilní i některé polynomy vyšších stupňů, např. všechny polynomy $x^n - 2$, jak plyne z Eisensteinova kritéria (Tvrzení 6.10). Obecně není jednoduché určit, zda je daný polynom v $\mathbb{Q}[x]$ ireducibilní.

Příklad. Pro zajímavost, v Gaussových celých číslech $\mathbb{Z}[i]$ jdou rozkládat některá klasická prvočísla, například $5 = (2 + i)(2 - i)$. Dá se dokázat, že ireducibilní jsou následující prvky:

- $a + 0i$ a $0 + ai$ právě tehdy, když je $|a|$ prvočíslo a $|a| \equiv 3 \pmod 4$;
- $a + bi$, $b \neq 0$, právě tehdy, když $a^2 + b^2$ je prvočíslo.

Definice. *Ireducibilním rozkladem* prvku a rozumíme zápis

$$a \parallel p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_n^{k_n},$$

kde $p_1, \ldots, p_n$ jsou ireducibilní prvky, $p_i \nparallel p_j$ pro $i \neq j$, a $k_1, \ldots, k_n$ jsou přirozená čísla. Řekneme, že prvek *a má jednoznačný ireducibilní rozklad*, pokud má právě jeden rozklad *až na pořadí a asociovanost*, tj. jsou-li

$$a \parallel p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_n^{k_n} \parallel q_1^{l_1} \cdot q_2^{l_2} \cdot \ldots \cdot q_m^{l_m}$$

dva ireducibilní rozklady prvku a, pak $m = n$ a existuje permutace indexů π taková, že $p_i \parallel q_{\pi(i)}$ a $k_i = l_{\pi(i)}$ pro každé i.

Definice jednoznačnosti je motivována následujícím pozorováním: v oboru $\mathbb{Z}$ můžeme psát například $12 = 2^2 \cdot 3^1 = 3^1 \cdot (-2)^2 \parallel (-2)^2 \cdot (-3)^1$. Formálně vzato, jde o tři různé rozklady, ale přesto je rozumné je považovat za „totožné“: liší se pouze pořadím a volbou z navzájem asociovaných prvků. Svůj význam má v definici rozkladu také znaménko asociovanosti: prvek -4 v $\mathbb{Z}$ nelze vyjádřit jako druhá mocnina ireducibilního prvku, nicméně přesto má rozklad, $-4 \parallel 2^2$.

Je důležité, v kterém oboru rozkládáme (nutno explicitně zmínit, nebo to musí být zřejmé z kontextu). Např. polynom $2x^2 + 2$ je ireducibilní v $\mathbb{Q}[x]$, ale má ireducibilní rozklad $2^1 \cdot (x^2 + 1)^1$ v $\mathbb{Z}[x]$.

Příklad. V tabulce jsou uvedeny rozklady polynomů na součin ireducibilních v různých oborech:

	x^2+1	$2x^2+2$	x^2-2	x^4+2x^2+1
$\mathbb{Z}[x]$	ireducibilní	$2\cdot(x^2+1)$	ireducibilní	$(x^2+1)^2$
$\mathbb{Q}[x]$	ireducibilní	ireducibilní	ireducibilní	$(x^2+1)^2$
$\mathbb{R}[x]$	ireducibilní	ireducibilní	$(x-\sqrt{2})(x+\sqrt{2})$	$(x^2+1)^2$
$\mathbb{C}[x]$	$(x-i)(x+i)$	$(2x-2i)(x+i)$	$(x-\sqrt{2})(x+\sqrt{2})$	$(x-i)^2(x+i)^2$
$\mathbb{Z}_5[x]$	$(x+2)(x+3)$	$(x+2)(2x+1)$	ireducibilní	$(x+2)^2(x+3)^2$

Existence ani jednoznačnost rozkladů není ničím garantovaná. Následující příklady jsou spíše pro zajímavost, hlouběji se jimi zabývat nebudeme.

Příklad (obor bez rozkladů)**.** Uvažujme podokruh $\mathbf{R}$ oboru $\mathbb{Q}[x]$ sestávající z polynomů, jejichž absolutní člen je celé číslo. Polynom x není invertibilní, ale nemá ireducibilní rozklad: pokud $f \mid x$, tj. pokud existuje $g \in R$ takové, že $x = f \cdot g$, pak buď $f = a \in \mathbb{Z}$ a $g = \frac{1}{a}x$, nebo naopak. Přitom z těchto polynomů jsou ireducibilní pouze konstantní polynomy s prvočíselným koeficientem, protože každý polynom $\frac{1}{a}x$ má netriviální rozklad $\frac{1}{2a}x \cdot 2$. Součinem konstantních polynomů však nikdy nebude polynom x.

Příklad (obor s nejednoznačnými rozklady)**.** Uvažujme obor $\mathbb{Z}[\sqrt{5}]$. Prvek 4 lze rozložit dvěma způsoby:

$$4 = 2^2 = (1+\sqrt{5})(-1+\sqrt{5}).$$

Dá se dokázat, že jsou všechny tři prvky $2, \pm 1+\sqrt{5}$ ireducibilní a že nejsou asociované.

Definice. Obor integrity se nazývá *gaussovský*, pokud má každý neinvertibilní nenulový prvek jednoznačný rozklad na ireducibilní činitele.

Příklad. Řada oborů integrity je gaussovských:

- Tělesa jsou gaussovské obory, podmínka z definice je prázdná.
- Obor $\mathbb{Z}$ je gaussovský, jak říká základní věta aritmetiky (Věta 1.2).
- Obory polynomů nad tělesem jsou gaussovské. Pro polynomy jedné proměnné funguje podobný důkaz jako pro celá čísla; obecná varianta tohoto důkazu je předmětem následujících dvou sekcí (Věty 7.1 a 7.6). Pro polynomy více proměnných nebo pro polynomy nad $\mathbb{Z}$ můžeme použít Gaussovu větu (Věta 6.8).
- Některé obory $\mathbb{Z}[\sqrt{s}]$ jsou gaussovské, např. pro $s = -1, \pm 2, 3$, některé ne, např. pro $s = -3, 5$. Zabývat se jimi dále nebudeme.

5.4. Dělitelnost v gaussovských oborech.

Která z uvedených vlastností je nejsilnější? Přesvědčíme se, že z předpokladu existence a jednoznačnosti ireducibilních rozkladů rychle plyne řada užitečných důsledků. Stěžejní je následující pozorování o tom, jak vypadají dělitelé prvku s daným rozkladem.

Tvrzení 5.2. *Buď $\mathbf{R}$ gaussovský obor, $a, b \in R$ a uvažujme ireducibilní rozklad*

$$a \parallel p_1^{k_1} \cdot \ldots \cdot p_n^{k_n}.$$

Pak $b \mid a$ právě tehdy, když

$$b \parallel p_1^{l_1} \cdot \ldots \cdot p_n^{l_n}$$

pro nějaká $0 \le l_i \le k_i$.

Důkaz. ($\Leftarrow$) Pokud $a = qp_1^{k_1} \cdot \ldots \cdot p_n^{k_n}$ a $b = rp_1^{l_1} \cdot \ldots \cdot p_n^{l_n}$ pro nějaké invertibilní prvky $q, r \in R$, definujme $c = qr^{-1}p_1^{k_1-l_1} \cdot \ldots \cdot p_n^{k_n-l_n}$ a vidíme, že $a = bc$, tedy $b \mid a$.

($\Rightarrow$) Uvažujme prvek c takový, že $a = b \cdot c$ a uvažujme ireducibilní rozklady

$$b \parallel q_1^{s_1} \cdot \ldots \cdot q_u^{s_u}, \qquad c \parallel r_1^{t_1} \cdot \ldots \cdot r_v^{t_v}.$$

Složením těchto dvou rozkladů (vynecháme ty prvky r_i, které jsou asociované s některým q_j a jejich exponenty sečteme) dostaneme druhý ireducibilní rozklad prvku a:

$$a \parallel p_1^{k_1} \cdot \ldots \cdot p_n^{k_n} \parallel q_1^{s'_1} \cdot \ldots \cdot q_u^{s'_u} r_{i_1}^{t_{i_1}} \cdot \ldots \cdot r_{i_w}^{t_{i_w}}.$$

Z jednoznačnosti rozkladů plyne, že ke každému $i = 1, \ldots, u$ existuje jednoznačně určené $j \in \{1, \ldots, n\}$ takové, že $q_i \parallel p_j$, přičemž $s_i \leq s'_i = k_j$. Z toho vyplývá, že $b \parallel p_1^{l_1} \cdot \ldots \cdot p_n^{l_n}$ pro nějaká $0 \leq l_i \leq k_i$. □

Snadným důsledkem Tvrzení 5.2 je několik vlastností dobře známých z celých čísel. V gaussovských oborech existují NSD: stačí srovnat rozklady obou prvků a vzít největší společný podrozklad, např.

$$\begin{aligned}\mathrm{NSD}(540, 336) &= \mathrm{NSD}((-2)^2 \cdot 3^3 \cdot 5,\ 2^4 \cdot (-3) \cdot 7) = \\ &= \mathrm{NSD}(2^2 \cdot 3^3 \cdot 5^1 \cdot 7^0,\ 2^4 \cdot 3^1 \cdot 5^0 \cdot 7^1) = 2^2 \cdot 3^1 \cdot 5^0 \cdot 7^0 = 12.\end{aligned}$$

V gaussovských oborech platí analogie Lemmatu 1.5: pokud ireducibilní prvek p dělí součin ab, pak jej musíme nalézt v rozkladu aspoň jednoho z činitelů. A žádný prvek nemá nekonečnou posloupnost vlastních dělitelů, neboť při každém dělení ztrácíme z rozkladu aspoň jeden činitel. Důkaz následujícího důsledku je formálním vyjádřením právě uvedených myšlenek.

Důsledek 5.3 (dělitelnost v gaussovských oborech)**.** *Buď* $\mathbf{R}$ *gaussovský obor. Pak*

(1) *pro každé* $a, b \in R$ *existuje* $\mathrm{NSD}(a, b)$*;*
(2) *pro každý ireducibilní prvek* p *v* $\mathbf{R}$*, pokud* $p \mid ab$*, pak* $p \mid a$ *nebo* $p \mid b$*.*
(3) *neexistuje posloupnost* $a_1, a_2, a_3, \ldots \in R$ *taková, že* $a_{i+1} \mid a_i$ *a* $a_{i+1} \nmid a_i$*.*

Důkaz. (1) Uvažujme ireducibilní prvky $p_1, \ldots, p_n$, $p_i \nparallel p_j$ pro $i \neq j$, a $k_i, l_i \geq 0$ takové, že

$$a \parallel p_1^{k_1} \cdot \ldots \cdot p_n^{k_n}, \qquad b \parallel p_1^{l_1} \cdot \ldots \cdot p_n^{l_n}$$

(libovolné ireducibilní rozklady prvků a, b můžeme přepsat do této formy tak, že ze dvou asociovaných činitelů vybereme jeden a do rozkladu případně doplníme činitele v nulté mocnině.) Podle Tvrzení 5.2 $c \mid a, b$ právě tehdy, když $c \parallel p_1^{m_1} \cdot \ldots \cdot p_n^{m_n}$, kde $0 \leq m_i \leq k_i$ a $0 \leq m_i \leq l_i$, čili právě tehdy, když $0 \leq m_i \leq \min(k_i, l_i)$, pro všechna i. Největším z těchto společných dělitelů tedy bude ten, kde $m_i = \min(k_i, l_i)$.

(2) Uvažujme ireducibilní prvek p a prvky a, b takové, že $p \mid ab$. Podobně jako v (1), napišme $a \parallel p_1^{k_1} \cdot \ldots \cdot p_n^{k_n}$, $b \parallel p_1^{l_1} \cdot \ldots \cdot p_n^{l_n}$, kde $k_i, l_i \geq 0$, aspoň jeden nenulový. Pak $ab \parallel p_1^{k_1+l_1} \cdot \ldots \cdot p_n^{k_n+l_n}$ a podle Tvrzení 5.2 musí mít dělitel p rozklad, který obsahuje některé z prvků $p_1, \ldots p_n$. Z ireducibility plyne, že $p \parallel p_i$ pro některé i a tedy $p \mid a$ (pokud $k_i > 0$) nebo $p \mid b$ (pokud $l_i > 0$).

(3) Začneme obecnou úvahou. Každý nenulový neinvertibilní prvek a má, až na pořadí a volbu ireducibilních prvků v základu mocnin, jednoznačný rozklad $a \parallel p_1^{k_1} \cdot \ldots \cdot p_n^{k_n}$. Označme $\nu(a) = k_1 + \ldots + k_n$ a dodefinujme $\nu(a) = 0$ pro všechny invertibilní prvky a. Z jednoznačnosti rozkladů plyne, že číslo $\nu(a)$ je nezávislé na volbě rozkladu. Z Tvrzení 5.2 plyne, že pokud $u \mid v$ a $v \nmid u$, pak $\nu(u) < \nu(v)$.

Pro spor předpokládejme existenci takové posloupnosti $a_1, a_2, a_3 \dots$ Z úvah v předešlém odstavci plyne, že $\nu(a_1) > \nu(a_2) > \nu(a_3) > \dots$ je nekonečná klesající posloupnost nezáporných celých čísel, spor. □

Příklad. Uvažujme obor $\mathbb{Z}[\sqrt{5}]$. Prvek 2 je ireducibilní, ale tvrzení (2) pro něj neplatí: $2 \mid (\sqrt{5}-1)(\sqrt{5}+1)$, ale přitom $2 \nmid (\sqrt{5} \pm 1)$.

To, že jsou si příklady na neexistenci NSD, nejednoznačnost ireducibilního rozkladu a právě zmíněný protipříklad podobné, není náhoda. Podrobněji si to vysvětlíme v sekci 7.

6. Dělitelnost v oborech polynomů

6.1. Polynomy jedné proměnné nad tělesem.

Tuto sekci je nejlepší provést formou cvičení: nechť si každý student sám přijde na to, proč se některé typy polynomů chovají z hlediska dělitelnosti analogicky celým číslům. Návodem nechť je sekce 1, kde se analogická tvrzení dokazují pro obor $\mathbb{Z}$.

Cvičení 6.1. *Projděte si znovu popis Eukleidova algoritmu v sekci 1.2 a rozmyslete se, že analogický postup projde i pro obory polynomů jedné proměnné nad tělesem.*

Proč to neprojde pro obory polynomů více proměnných nebo pro obor $\mathbb{Z}[x]$? Které z dříve dokázaných vlastností polynomů jsou důležité, aby to fungovalo?

Na základě předchozího cvičení dokažte následující tvrzení.

Tvrzení 6.2 (Bézoutova rovnost)**.** *Buď $\mathbf{T}$ těleso. Pro každou dvojici polynomů $f, g \in T[x]$ existuje $NSD(f,g)$ a existují polynomy $r, s \in T[x]$ (tzv.* Bézoutovy koeficienty*) splňující*

$$\mathrm{NSD}(f,g) = r \cdot f + s \cdot g.$$

Problém s obory polynomů více proměnných i s oborem $\mathbb{Z}[x]$ je zásadní. Nejen, že neprojde Eukleidův algoritmus, ale dokonce ani neplatí Bézoutova rovnost.

Příklad. V oboru $\mathbb{Z}[x]$ neplatí Bézoutova rovnost. Platí $\mathrm{NSD}(x+1, x-1) = 1$, ale neexistují $r, s \in \mathbb{Z}[x]$ takové, že $r \cdot (x+1) + s \cdot (x-1) = 1$: pokud dosadíme číslo 1, vyjde nám $2r(1) = 1$, což není pro celočíselný polynom možné.

Příklad. V oboru $\mathbb{Q}[x,y]$ neplatí Bézoutova rovnost. Platí $\mathrm{NSD}(x,y) = 1$, ale neexistují $r, s \in \mathbb{Q}[x,y]$ takové, že $r \cdot x + s \cdot y = 1$, protože absolutní člen polynomu na levé straně je nutně nula.

Dobrá zpráva je, že i v těchto oborech existují NSD všech dvojic polynomů, ale není úplně snadné to dokázat.

Cvičení 6.3. *Projděte si znovu důkaz základní věty aritmetiky v sekci 1 a rozmyslete si, jak postup modifikovat tak, aby prošel i pro obory polynomů nad tělesem. Podle jakého parametru je třeba dělat indukci v důkazu existence, resp. jednoznačnosti ireducibilních rozkladů?*

Proč to neprojde pro obory polynomů více proměnných nebo pro obor $\mathbb{Z}[x]$? Které z dříve dokázaných vlastností polynomů jsou důležité, aby to fungovalo?

Na základě předchozího cvičení dokažte následující tvrzení.

Věta 6.4 (Analogie základní věty aritmetiky pro polynomy)**.** *Buď $\mathbf{T}$ těleso. Pak $\mathbf{T}[x]$ je gaussovský obor.*

Dobrá zpráva je, že také obory polynomů více proměnných nebo obor $\mathbb{Z}[x]$ jsou gaussovské, ale není úplně snadné to dokázat.

6.2. **Polynomy nad oborem vs. nad jeho podílovým tělesem.**
Důkaz gaussovskosti obecných polynomiálních oborů je založen na studiu vztahu dělitelnosti v oboru $\mathbf{R}[x]$ a v oboru $\mathbf{Q}[x]$, kde $\mathbf{Q}$ je podílové těleso oboru $\mathbf{R}$.

Definice. Polynom f nazveme *primitivní*, pokud jsou jeho koeficienty nesoudělné, tj. pokud c dělí všechny jeho koeficienty, pak $c \parallel 1$. Například x^2y+x není primitivní v oboru $(\mathbb{Z}[x])[y]$ (máme netriviálního dělitele x), ale je primitivní v oboru $(\mathbb{Z}[y])[x]$.

Technika, která umožňuje rozšířit výsledky sekce 6.1 na obor $\mathbb{Z}[x]$ či na polynomy více proměnných, je založená na tzv. *Gaussově lemmatu*, které říká, že součin primitivních polynomů je primitivní, za předpokladu, že obor koeficientů je gaussovský. Důkaz Gaussova lemmatu se učit nemusíte, ale pro úplnost jej uvedeme.

Lemma 6.5 (Gaussovo lemma)**.** *Buď $\mathbf{R}$ gaussovský obor a f, g primitivní polynomy z $\mathbf{R}[x]$. Pak fg je primitivní polynom.*

Důkaz. Označme $f = \sum_{i=0}^{n} a_i x^i$ a $g = \sum_{i=0}^{m} b_i x^i$ a předpokládejme, že fg není primitivní polynom. Díky existenci ireducibilních rozkladů existuje ireducibilní prvek $p \in R$, který dělí všechny koeficienty součinu fg. Zvolme nejmenší j takové, že $p \nmid a_j$, a nejmenší k takové, že $p \nmid b_k$ (protože jsou polynomy f, g primitivní, p nemůže dělit všechny jejich koeficienty). Podívejme se na $(j+k)$-tý koeficient polynomu fg:

$$c_{j+k} = a_0 b_{j+k} + \ldots + a_{j-1} b_{k+1} + a_j b_k + a_{j+1} b_{k-1} + \ldots + a_{j+k} b_0.$$

Protože $p \mid a_i$ pro všechna $i < j$, máme

$$p \mid a_0 b_{j+k} + \ldots + a_{j-1} b_{k+1}.$$

Protože $p \mid b_i$ pro všechna $i < k$, máme

$$p \mid a_{j+1} b_{k-1} + \ldots + a_{j+k} b_0.$$

Tedy p dělí všechny členy součtu vlevo i vpravo od $a_j b_k$. Tento člen naopak p dělitelný není, jak plyne z Důsledku 5.3(2). Čili $p \nmid c_{j+k}$, spor. □

Gaussovo lemma umožňuje dát do souvislosti dělitelnost v oborech $\mathbf{R}[x]$ a $\mathbf{Q}[x]$. Stěžejní je následující pozorování.

Lemma 6.6. *Buď $\mathbf{R}$ gaussovský obor, $\mathbf{Q}$ jeho podílové těleso a f, g primitivní polynomy z $\mathbf{R}[x]$. Pokud $f \mid g$ v $\mathbf{Q}[x]$, pak také $f \mid g$ v $\mathbf{R}[x]$.*

Důkaz. V čem je problém? První podmínka říká, že $g = fh$ pro nějaký polynom h z $\mathbf{Q}[x]$, zatímco druhá podmínka se dožaduje takového polynomu h z $\mathbf{R}[x]$. Pokud přenásobíme obě strany nejmenším společným násobkem jmenovatelů koeficientů v polynomu h, označme jej c, dostaneme rovnost $cg = f \cdot ch$, přičemž na pravé straně je součin dvou primitivních polynomů. Podle Gaussova lemmatu je primitivní i polynom cg, takžte musí být $c \parallel 1$, a tedy h byl ve skutečnosti polynom z $\mathbf{R}[x]$. □

Pomocí tohoto pozorování pak není těžké dokázat následující větu. Do detailů se pouštět nebudeme, jde v zásadě technickou a myšlenkově nijak zvlášť přínosnou věc.

Věta 6.7 (NSD a ireducibilita v oboru vs. podílovém tělese)**.** *Buď $\mathbf{R}$ gaussovský obor, $\mathbf{Q}$ jeho podílové těleso a f, g polynomy z $\mathbf{R}[x]$. Pak*

(1) $\mathrm{NSD}_{\mathbf{R}[x]}(f,g)$ *existuje a je roven součinu* $c\cdot h$*, kde* $c=\mathrm{NSD}_{\mathbf{R}}(c_f,c_g)$ *a* h *je primitivní polynom z* $\mathbf{R}[x]$ *splňující* $h=\mathrm{NSD}_{\mathbf{Q}[x]}(f/c_f,g/c_g)$*, přičemž* c_f *značí NSD koeficientů polynomu* f *a* c_g *značí NSD koeficientů polynomu* g*.*

(2) f *je ireducibilní v* $\mathbf{R}[x]$ *právě tehdy, když*
- $\deg f=0$ *a* f *je ireducibilní v* $\mathbf{R}$*; nebo*
- $\deg f>0$*,* f *je primitivní a ireducibilní v* $\mathbf{Q}[x]$*.*

Příklad. Uvažujme obor $\mathbb{Z}[x]$ a polynomy

$$f=4x^2+8x+4,\qquad g=-6x^2+6.$$

Pak $c=\mathrm{NSD}_{\mathbb{Z}}(4,-6)=2$, $h=\mathrm{NSD}_{\mathbb{Q}[x]}(x^2+2x+1,x^2-1)=x+1$, a tedy $\mathrm{NSD}_{\mathbb{Z}[x]}(f,g)=2\cdot(x+1)$.

Z bodu (2) plyne existence ireducibilních rozkladů v $\mathbf{R}[x]$. S jednoznačností to je trochu složitější, protože nemáme Bézoutovu rovnost. V sekci 7 si ukážeme, jak to obejít a jako okamžitý důsledek Vět 6.7 a 7.1 dostaneme následující větu.

Věta 6.8 (Gaussova věta)**.** *Je-li* $\mathbf{R}$ *gaussovský obor, pak je* $\mathbf{R}[x]$ *také gaussovský obor.*

Několikanásobnou aplikací Gaussovy věty ihned plyne, že polynomy libovolně mnoha proměnných nad gaussovským oborem (například obory $\mathbb{Z}[x,y,...]$ či $\mathbf{T}[x,y,...]$ pro libovolné těleso $\mathbf{T}$) jsou gaussovské.

6.3. Racionální kořeny a Eisensteinovo kritérium ireducibility.

Následující kritérium možná znáte ze střední školy, ale patrně jste si neuvědomili, že k jeho důkazu je potřeba vědět, že jsou v oborech polynomů jednoznačné ireducibilní rozklady! Konkrétně, jde o vlastnost (2) z Důsledku 5.3.

Tvrzení 6.9 (kritérium existence racionálního kořene)**.** *Buď* $\mathbf{R}$ *gaussovský obor a* $\mathbf{Q}$ *jeho podílové těleso. Má-li polynom* $f=\sum_{i=0}^{n}a_ix^i\in R[x]$ *kořen* $\frac{r}{s}\in Q$ *(předpokládáme* r,s *nesoudělná), pak* $r\mid a_0$ *a* $s\mid a_n$*.*

Důkaz. Dosaďme prvek $\frac{r}{s}$ do f. Protože $\sum_{i=0}^{n}a_i(\frac{r}{s})^i=0$, přenásobením prvkem s^n dostáváme

$$a_0s^n+a_1rs^{n-1}+a_2r^2s^{n-2}+\ldots+a_{n-1}r^{n-1}s+a_nr^n=0.$$

Protože r dělí všechny členy $a_1rs^{n-1},\ldots,a_nr^n$ i pravou stranu, musí dělit i první člen a_0s^n. Protože jsou r,s nesoudělné, musí r dělit a_0 (zde využíváme gaussovskost). Analogicky, protože s dělí všechny členy $a_0s^n,\ldots,a_{n-1}r^{n-1}s$, musí dělit i poslední člen a_nr^n, tedy $s\mid a_n$. □

Příklad. Najdeme všechny racionální kořeny polynomu $2x^5-3x^4+2x-3$. Podle Tvrzení 6.9 jsou jedinými kandidáty čísla ±1, ±3, $\pm\frac{1}{2}$ a $\pm\frac{3}{2}$. Dosazením zjistíme, že vyhovuje pouze číslo $-\frac{3}{2}$.

Příklad. Racionálními kořeny polynomu x^n-p, p prvočíslo, mohou být pouze čísla $\pm1,\pm p$ a ani jedno očividně nevyhovuje (pro $n\geq2$). Důsledkem je, že všechny odmocniny prvočísel $\sqrt[n]{p}$ jsou iracionální.

Podobným trikem se dokáže také Eisensteinovo kritérium ireducibility.

Tvrzení 6.10 (Eisensteinovo kritérium)**.** *Buď* $\mathbf{R}$ *gaussovský obor a* $f = \sum_{i=0}^{n} a_i x^i$ *primitivní polynom z* $\mathbf{R}[x]$*. Pokud existuje ireducibilní prvek* $p \in R$ *splňující* $p \mid a_0$*,* $p \mid a_1, \ldots, p \mid a_{n-1}$ *a* $p^2 \nmid a_0$*, pak je polynom* f *ireducibilní v* $\mathbf{R}[x]$*.*

Důkaz. Pro spor uvažujme rozklad $f = gh$, kde $g = \sum_{i=0}^{k} b_i x^i$ a $h = \sum_{i=0}^{l} c_i x^i$ jsou polynomy z $\mathbf{R}[x]$ stupně alespoň 1. Protože p dělí $a_0 = b_0 c_0$, platí $p \mid b_0$ nebo $p \mid c_0$ (Důsledek 5.3(2)), ale určitě ne oboje zároveň, protože $p^2 \nmid a_0$. Nechť je to bez újmy na obecnosti b_0. Podobně, protože $p \mid a_1 = b_0 c_1 + b_1 c_0$ a $p \nmid c_0$, musí $p \mid b_1$. Protože $p \mid a_2 = b_0 c_2 + b_1 c_1 + b_2 c_0$ a $p \nmid c_0$, musí $p \mid b_2$. Postupně zjistíme, že p dělí všechny koeficienty b_i, tedy $p \mid g \mid f$, což je spor s primitivitou. □

Příklad. Z Eisensteinova kritéria plyne ireducibilita polynomů $x^n - p$, p prvočíslo, v oboru $\mathbb{Z}[x]$ (a tedy i v $\mathbb{Q}[x]$).

7. Abstraktní teorie dělitelnosti

7.1. Zobecnění základní věty aritmetiky.

Cílem této sekce je dokončit analýzu vztahů mezi pojmy a vlastnostmi, se kterými jsme se setkali v sekci 5. Věta 7.1 v jistém smyslu zobecňuje základní větu aritmetiky. Její důkaz používal dvě základní ingredience: existenci NSD a Bézoutovu rovnost, a dále matematickou indukci, vycházející z uspořádání přirozených čísel (pro polynomy jsme indukci aplikovali na stupeň). Ovšem obecné obory integrity nemusí být smysluplně uspořádatelné (jak byste uspořádali třeba Gaussova celá čísla?), klasická indukce nám tedy nepomůže. A také bychom se měli obejít bez Bézouta, jinak postup nepůjde aplikovat například na obor $\mathbb{Z}[x]$. Postačující podmínky jsou zformulovány v následující větě, indukci nahrazuje podmínka (2).

Věta 7.1 (zobecněná základní věta aritmetiky)**.** *Buď* $\mathbf{R}$ *obor integrity. Pak* $\mathbf{R}$ *je gaussovský právě tehdy, když*

(1) *existuje NSD všech dvojic prvků;*
(2) *neexistuje posloupnost* $a_1, a_2, a_3, \ldots \in R$ *taková, že* $a_{i+1} \mid a_i$ *a* $a_{i+1} \nmid\!\mid a_i$*.*

Přímou implikaci jsme dokázali v Důsledku 5.3. Důkaz opačné implikace bude sledovat postup, kterým jsme dokázali základní větu aritmetiky v sekci 1.1, akorát některé kroky budou složitější.

Důkaz opět rozdělíme do dvou částí: nejprve dokážeme existenci rozkladů, což je poměrně snadné, a pak se budeme věnovat jednoznačnosti, která vyžaduje jistá technická lemmata.

Důkaz existence rozkladů. Pro spor uvažujme prvek a, který nemá ireducibilní rozklad, $0 \neq a \nmid\!\mid 1$. Rekurzí zkonstruujeme posloupnost, která protiřečí bodu (2).

- Položme $a_1 = a$. Tedy $a_1 \nmid\!\mid 1$ a nemá ireducibilní rozklad.
- Předpokládejme, že $a_i \nmid\!\mid 1$ a nemá ireducibilní rozklad. Speciálně, prvek a_i není sám ireducibilní, a tedy $a_i = b \cdot c$ pro nějaká $b, c \nmid\!\mid 1$. Kdyby b i c měly ireducibilní rozklad, pak by ho měl i a_i, takže aspoň jedno z nich ireducibilní rozklad nemá, označme jej a_{i+1}. Tedy a_{i+1} je vlastní dělitel a_i a nemá ireducibilní rozklad.

Tato posloupnost $a_1, a_2, \ldots$ protiřečí předpokladu (2). □

K důkazu jednoznačnosti se nám bude hodit ještě jedna analogie Lemmatu 1.5, tentokrát dokázaná pouze za předpokladu existence NSD. Protože obecně nemáme k dispozici Bézoutovu rovnost, budeme muset postupovat obezřetněji než v sekci

1.1. Důkaz následujícího lemmatu je technický, u zkoušky se na něj ptát nebudu a když si jej nepřečtete, o nic zásadního nepřijdete.

Lemma 7.2. *Buď* $\mathbf{R}$ *obor integrity a* $a, b, c \in R$ *takové, že existuje* $\mathrm{NSD}(a, b)$ *i* $\mathrm{NSD}(ac, bc)$*. Pak*

$$\mathrm{NSD}(ac, bc) = c \cdot \mathrm{NSD}(a, b).$$

Důkaz. Vzhledem k tomu, že NSD je definován *až na asociovanost*, stačí dokázat, že levá strana rovnosti dělí pravou a naopak. Označme $u = \mathrm{NSD}(ac, bc)$. Pro $c = 0$ tvrzení platí triviálně, předpokládejme tedy $c \neq 0$.

Nejprve dokážeme, že $u \mid c \cdot \mathrm{NSD}(a, b)$. Protože $u \mid ac$, existuje x s vlastností $ac = ux$. Protože $u \mid bc$, existuje y s vlastností $bc = uy$. Protože c je společný dělitel ac, bc, platí $c \mid u$, a tedy existuje z s vlastností $u = cz$. Dostáváme $ac = czx$ a $bc = czy$ a krácením získáme vztahy $a = zx$ a $b = zy$. Tedy z je společný dělitel a, b, tedy z dělí $\mathrm{NSD}(a, b)$, a tudíž $u = cz \mid c \cdot \mathrm{NSD}(a, b)$.

Naopak, protože $\mathrm{NSD}(a, b)$ dělí a i b, tak $c \cdot \mathrm{NSD}(a, b)$ dělí ac i bc, a tudíž musí dělit i jejich největšího společného dělitele. □

Lemma 7.3. *Uvažujme obor integrity* $\mathbf{R}$*, kde existují NSD všech dvojic prvků. Buď* p *ireducibilní prvek* $\mathbf{R}$*. Pokud* $p \mid ab$*, pak* $p \mid a$ *nebo* $p \mid b$*.*

Důkaz. Buď p ireducibilní a a, b taková, že $p \mid ab$. Předpokládejme, že $p \nmid a$. Z ireducibility plyne, že $\mathrm{NSD}(a, p) = 1$, a tedy podle Lemmatu 7.2

$$\mathrm{NSD}(ab, pb) = b \cdot \mathrm{NSD}(a, p) = b.$$

Ovšem p je společným dělitelem ab a pb, tedy $p \mid \mathrm{NSD}(ab, pb) = b$. □

Důkaz jednoznačnosti rozkladů. Sporem. Mezi všemi prvky s různými ireducibilními rozklady zvolme takové a, jehož rozklad je nejkratší, ve smyslu součtu exponentů u všech ireducibilních prvků v tomto rozkladu. Označme tento nejkratší rozklad $a \parallel p_1^{k_1} \cdot \ldots \cdot p_n^{k_n}$ a uvažujme nějaký jiný rozklad $a \parallel q_1^{l_1} \cdot \ldots \cdot q_m^{l_m}$. Protože je p_1 ireducibilní, podle Lemmatu 7.3 musí dělit některé q_i. Protože jsou všechna q_j ireducibilní a $p_1 \nparallel 1$, máme $p_1 \parallel q_i$. Pak ale

$$b = p_1^{k_1 - 1} p_2^{k_2} \cdot \ldots \cdot p_n^{k_n} \parallel q_1^{l_1} \cdot \ldots \cdot q_{i-1}^{l_{i-1}} \cdot q_i^{l_i - 1} q_{i+1}^{l_{i+1}} \cdot \ldots \cdot q_m^{l_m}$$

je prvek s kratším nejednoznačným rozkladem, spor. □

Věta 7.1 je zajímavá mimo jiné proto, že charakterizuje gaussovské obory dvěma zcela rozdílnými způsoby. Definice pomocí existence a jednoznačnosti rozkladů je čistě aritmetická, formulovaná jako vlastnost operace násobení v oboru $\mathbf{R}$. Naopak druhou stranu charakterizace lze formulovat čistě v jazyce relace dělitelnosti: podmínka (1) říká, že vzhledem k „uspořádání“ $\mid$ existují „infima“ všech dvouprvkových množin, a podmínka (2) říká, že v něm neexistuje nekonečný ostře klesající řetězec.

Na závěr si dokážeme Gaussovu větu.

Důkaz Věty 6.8. Použijeme charakterizaci z Věty 7.1. Existenci NSD v $\mathbf{R}[x]$ jsme dokázali ve Větě 6.7. Buď $f_1, f_2, f_3, \ldots$ nekonečná posloupnost vlastních dělitelů v $\mathbf{R}[x]$. Pak $\deg f_1 \geq \deg f_2 \geq \deg f_3 \geq \cdots \geq 0$, a tedy existuje n takové, že $\deg f_n = \deg f_{n+1} = \ldots$ Označíme-li u_i vedoucí koeficient polynomu f_i, pak $u_n, u_{n+1}, u_{n+2} \ldots$ tvoří nekonečnou posloupnost vlastních dělitelů v $\mathbf{R}$, spor. □

7.2. **Eukleidův algoritmus a Bézoutova rovnost.**

Svoji abstraktní variantu má také Eukleidův algoritmus. Teorii lze aplikovat jak na obory polynomů jedné proměnné nad tělesem (čímž získáme řešení cvičení ze sekce 6.1), tak například na Gaussova celá čísla $\mathbb{Z}[i]$ (důsledkem je například základní věta aritmetiky pro obor $\mathbb{Z}[i]$).

Základní vlastností, kterou potřebujeme k běhu Eukleidova algoritmu, je dělení se zbytkem, a k tomu potřebujeme nějak měřit „velikost prvku“.

Definice. Obor $\mathbf{R}$ se nazývá *eukleidovský*, pokud na něm existuje *eukleidovská norma*, tj. zobrazení

$$\nu : R \to \mathbb{N} \cup \{0\}$$

splňující

(0) $\nu(0) = 0$;
(1) pokud $a \mid b \neq 0$, pak $\nu(a) \leq \nu(b)$;
(2) pro všechna $a, b \in R$, $b \neq 0$, existují $q, r \in R$ taková, že

$$a = bq + r \quad \text{a} \quad \nu(r) < \nu(b).$$

Podmínka (2) říká, že pro každou dvojici $a, b \neq 0$ existuje „podíl“ q a „zbytek“ r (bez nároku na jejich jednoznačnost!), přičemž zbytek je „menší“ než prvek, kterým dělíme. Všimněte si, že $\nu(a) = 0$ právě tehdy, když $a = 0$: zbytek po dělení jakýmkoliv nenulovým prvkem musí mít menší normu, čili norma dělitele nemůže být 0.

Příklad. Řada gaussovských oborů je také eukleidovských:

- Tělesa jsou eukleidovské obory. Eukleidovskou normou je např. zobrazení $\nu(0) = 0$ a $\nu(a) = 1$ pro všechna $a \neq 0$.
- Obor $\mathbb{Z}$ je eukleidovský. Normou je absolutní hodnota, tj. $\nu(a) = |a|$.
- Obor $\mathbf{T}[x]$ je eukleidovský pro libovolné těleso $\mathbf{T}$. Normou je

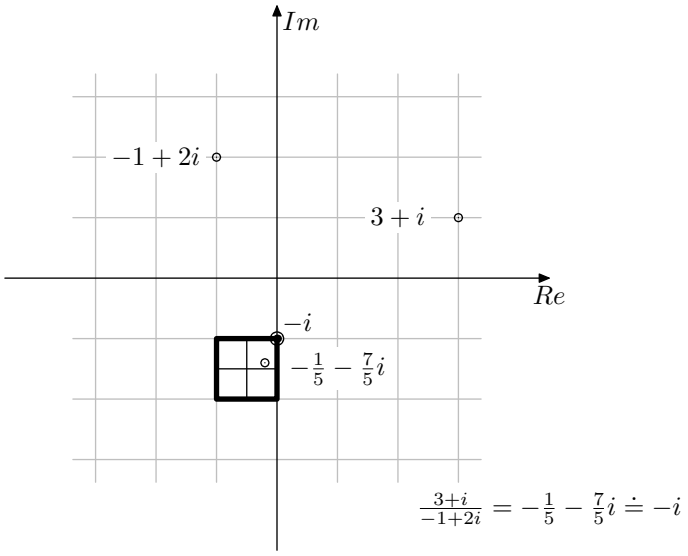
  $$\nu(f) = 1 + \deg f.$$

  Vlastnost (1) je zřejmá a vlastnost (2) plyne z Tvrzení 4.2. (Norma je nezáporná, takže ke stupni musíme přičíst 1.)
- Některé obory $\mathbb{Z}[\sqrt{s}]$ jsou eukleidovské, např. pro $s = -1, \pm 2, 3$, některé ne, např. pro $s = -3, 5$. V uvedených případech je normou

  $$\nu(a + b\sqrt{s}) = |a^2 - sb^2|.$$

  Vlastnost (1) platí pro každé s (cvičení), ale vlastnost (2) je netriviální. Například pro Gaussova celá čísla lze postupovat tak, že spočteme přesný podíl $\frac{a}{b} \in \mathbb{C}$, najdeme nejbližší prvek $\mathbb{Z}[i]$ k tomuto číslu, ten prohlásíme za podíl q a dopočteme zbytek $r = a - bq$ (viz obrázek). Jako cvičení si můžete zkusit dokázat, že skutečně $|r| < |b|$.

Existují gaussovské obory, které nejsou eukleidovské, například obor $\mathbb{Z}[x]$ nebo obory polynomů více proměnných (i nad tělesem). Rozebereme případ oboru $\mathbb{Z}[x]$. Zobrazení $\nu(f) = 1 + \deg f$ není eukleidovskou normou: například pro polynomy $3x$ a $2x$ neexistují $q, r \in \mathbb{Z}[x]$ splňující $3x = q \cdot 2x + r$ a $\deg r = 0$ — po dosazení nuly vidíme, že $r = 0$, a tedy musí platit $3x = 2qx$, ale takový polynom v $\mathbb{Z}[x]$ neexistuje. Pozor, z uvedeného neplyne, že obor $\mathbb{Z}[x]$ není eukleidovský! Pouze jsme dokázali, že toto konkrétní ν není eukleidovskou normou. Přímý důkaz, že žádné



Obrázek 3. Dělení se zbytkem v $\mathbb{Z}[i]$.

zobrazení $\mathbb{Z}[x] \to \mathbb{N} \cup \{0\}$ nesplňuje podmínky eukleidovské normy, by byl komplikovaný. Naštěstí, tento fakt ihned plyne z Věty 7.4: v eukleidovských oborech platí Bézoutova rovnost, ale v $\mathbb{Z}[x]$ ne, jak jsme si ukázali v minulé sekci.

Primárním důsledkem eukleidovské normy je Eukleidův algoritmus na výpočet NSD a Bézoutových koeficientů.

Eukleidův algoritmus. Buď $\mathbf{R}$ eukleidovský obor.

- `VSTUP:` $a, b \in R$, $\nu(a) \geq \nu(b)$.
- `VÝSTUP:` $\mathrm{NSD}(a,b)$ a $u, v \in R$ splňující $\mathrm{NSD}(a,b) = u \cdot a + v \cdot b$.
- $a_0 = a, \quad u_0 = 1, \quad v_0 = 0.$
  $a_1 = b, \quad u_1 = 0, \quad v_1 = 1.$
- pro $i = 2, 3, \ldots$ prováděj následující:
  zvol q, r tak, aby $a_{i-1} = a_i q + r$ a $\nu(r) < \nu(a_i)$, a definuj

$$a_{i+1} = r, \quad u_{i+1} = u_{i-1} - u_i q, \quad v_{i+1} = v_{i-1} - v_i q$$

  pokud $a_{i+1} = 0$, odpověz a_i, u_i, v_i

Věta 7.4 (správnost Eukleidova algoritmu)**.** *V eukleidovském oboru* $\mathbf{R}$ *najde Eukleidův algoritmus pro jakýkoliv vstup* $a, b \in R$ *hodnotu* $\mathrm{NSD}(a,b)$ *a tzv.* Bézoutovy koeficienty $u, v \in R$ *splňující*

$$\mathrm{NSD}(a,b) = u \cdot a + v \cdot b.$$

Důkaz. Vzhledem k tomu, že $\nu(a_0) \geq \nu(a_1) > \nu(a_2) > \nu(a_3) > \ldots \geq 0$, algoritmus se musí po konečně mnoha krocích zastavit; označme n číslo kroku, ve kterém se tak stane. Stačí dokázat následující dvě vlastnosti:

(1) NSD dvou po sobě jdoucích prvků se nemění, tj. pro každé $i = 1, \ldots, n$ platí $\mathrm{NSD}(a_{i-1}, a_i) = \mathrm{NSD}(a_i, a_{i+1})$;
(2) pro každé $i = 0, \ldots, n$ platí $a_i = u_i \cdot a + v_i \cdot b$.

Vzhledem k tomu, že $\mathrm{NSD}(u,0)=u$ pro každé u, algoritmus správně odpoví

$$a_n = \mathrm{NSD}(a_n,0) = \mathrm{NSD}(a_{n-1},a_n) = \ldots = \mathrm{NSD}(a_0,a_1) = \mathrm{NSD}(a,b).$$

Obě vlastnosti plynou z vyjádření

$$a_{i-1} = a_i q + a_{i+1}.$$

Pro důkaz (1) si stačí uvědomit, že dvojice a_{i-1}, a_i má stejné společné dělitele jako dvojice a_i, a_{i+1} (jde o analogii Lemmatu 1.3). Indukcí ověříme (2). Pro $i = 0, 1$ výrok platí z definice. Předpokládáme-li $a_{i-1} = u_{i-1}a + v_{i-1}b$ a $a_i = u_i a + v_i b$, pak

$$\begin{aligned} a_{i+1} &= a_{i-1} - a_i q = (u_{i-1}a + v_{i-1}b) - (u_i a + v_i b)\cdot q \\ &= (u_{i-1} - u_i q)\cdot a + (v_{i-1} - v_i q)\cdot b = u_{i+1}a + v_{i+1}b. \end{aligned}$$

□

Nyní již snadno dokážeme, že v eukleidovských oborech má každý prvek jednoznačný ireducibilní rozklad.

Lemma 7.5. *Buď* $\mathbf{R}$ *eukleidovský obor a* $a, b \in R$, $a, b \neq 0$. *Pokud* $a \mid b$ *a* $a \nmid b$, *pak* $\nu(a) < \nu(b)$.

Důkaz. Napišme

- $b = au$ pro nějaké $u \in R$,
- $a = bq + r$ pro nějaká $q, r \in R$, $\nu(r) < \nu(b)$.

Vzhledem k tomu, že $b \nmid a$, platí $r \neq 0$. Dosazením získáme vyjádření $r = a - bq = a - auq = a(1-uq)$, z kterého plyne, že $a \mid r$. Protože $r \neq 0$, dostáváme $\nu(a) \leq \nu(r) < \nu(b)$. □

Věta 7.6. *Eukleidovské obory jsou gaussovské.*

Důkaz. Podle Věty 7.1 stačí dokázat, že v eukleidovských oborech existují NSD a neexistují nekonečné posloupnosti vlastních dělitelů. První fakt jsme dokázali ve Větě 7.4. Druhý plyne bezprostředně z předešlého lemmatu: taková posloupnost by měla ostře klesající normu, což nelze. □

Důsledkem Věty 7.6 je, že Gaussova celá čísla či obory polynomů nad tělesem jsou gaussovské.

8. Počítání modulo polynom

8.1. Čínská věta o zbytcích a interpolace.

Čínská věta o zbytcích hovoří o tom, jak vypadají řešení soustav lineárních kongruencí. V sekci 2.3 jsme viděli její variantu pro obor celých čísel, nicméně tato věta platí daleko obecněji. V této sekci si ukážeme další speciální případ, pro polynomy.

Věta 8.1 (čínská věta o zbytcích pro polynomy)**.** *Buď* $\mathbf{T}$ *těleso. Buď* $m_1, \ldots, m_n \in T[x]$ *po dvou nesoudělné polynomy, označme* $d = \sum \deg m_i$. *Buď* $u_1, \ldots, u_n \in T[x]$ *libovolné polynomy. Pak existuje právě jeden polynom* $f \in T[x]$ *stupně* $< d$, *který řeší soustavu kongruencí*

$$f \equiv u_1 \pmod{m_1}, \quad \ldots, \quad f \equiv u_n \pmod{m_n}.$$

Důkaz. Nejprve dokážeme jednoznačnost. Předpokládejme, že soustava má dvě řešení f, g stupně $< d$, tj. pro každé i platí

$$f \equiv g \equiv u_i \pmod{m_i}.$$

Polynom $f - g$ je také stupně $< d$, je dělitelný každým m_i, a protože jsou polynomy m_i navzájem nesoudělné, dostáváme (díky gaussovskosti oboru $\mathbf{T}[x]$)

$$m_1 \cdot \ldots \cdot m_n \mid f - g.$$

Čili polynom stupně d dělí polynom stupně $< d$, což je možné pouze v tom případě, že $f - g = 0$, tj. $f = g$.

Nyní dokážeme, že nějaké řešení existuje. Označme

$$P_k = \{f \in T[x] : \deg f < k\}$$

a uvažujme tuto množinu jako vektorový prostor dimenze k nad tělesem $\mathbf{T}$ (jeho bází jsou polynomy $1, x, x^2, \ldots, x^{k-1}$, každý polynom stupně $< k$ je lineární kombinací těchto polynomů s koeficienty z $\mathbf{T}$). Označme $d_i = \deg m_i$ a uvažujme zobrazení

$$\begin{aligned} \varphi : P_d &\to P_{d_1} \times \cdots \times P_{d_n} \\ f &\mapsto (f \bmod m_1, \ldots, f \bmod m_n). \end{aligned}$$

Uvědomte si, že jde o homomorfismus vektorových prostorů, neboť $(f+g) \bmod m = (f \bmod m) + (g \bmod m)$ a $af \bmod m = a(f \bmod m)$ pro libovolné polynomy f, g, m a každé $a \in T$. V předchozím odstavci jsme ukázali, že zobrazení φ je prosté. Přitom definiční obor i obor hodnot mají stejnou dimenzi $d = \sum d_i$, a podle jisté věty z lineární algebry (v jistém smyslu jde o analogii Lemmatu 2.6) je prosté zobrazení mezi vektorovými prostory stejné konečné dimenze také *na*. Tedy ke každé n-tici $(u_1, \ldots, u_n)$ existuje právě jedno f, které se na něj zobrazuje, a to je hledaným řešením soustavy. □

Stejně jako pro celá čísla, i tento důkaz je nekonstruktivní a využívá konečnosti, tentokráte dimenze jistého vektorového prostoru.

Poznámka. Vzhledem k aplikacím je na místě uvést návod, jak se řešení hledá. V jednom kroku snížíme počet kongruencí o jedna, a tento krok opakujeme tak dlouho, než zbyde jedna kongruence, jejíž řešení je očividné. Podobný postup funguje i pro číselnou verzi.

Uvažujme soustavu dvou kongruencí. Z kongruence $f \equiv u_2 \pmod{m_2}$ vyjádříme $f = gm_2 + u_2$ pro nějaký polynom $g \in T[x]$ a dosadíme do první kongruence

$$f = gm_2 + u_2 \equiv u_1 \pmod{m_1}.$$

Označme $\widetilde{m_2}$ inverz polynomu m_2 modulo m_1, tj. takový polynom, pro který platí $m_2\widetilde{m_2} \equiv 1 \pmod{m_1}$. Ten najdeme pomocí Bézoutovy rovnosti: napíšeme $1 = \mathrm{NSD}(m_1, m_2) = um_1 + vm_2$ a vidíme, že $vm_2 \equiv 1 \pmod{m_1}$. Přenásobením původní kongruence polynomem $\widetilde{m_2}$ dostáváme

$$g \equiv gm_2\widetilde{m_2} \equiv (u_1 - u_2)\widetilde{m_2} \pmod{m_1},$$

řešením tedy je každý polynom $g = hm_1 + (u_1 - u_2)\widetilde{m_2}$, pro libovolné $h \in T[x]$. Zpětným dosazením dostaneme obecné řešení

$$f = gm_2 + u_2 = hm_1m_2 + (u_1 - u_2)\widetilde{m_2}m_2 + u_2$$

Původní dvojice kongruencí je tedy ekvivalentní jedné kongruenci $f \equiv u \pmod{m_1m_2}$, pro jisté u (pokud chceme řešení stupně $< d$, stačí vzít $u \bmod m_1m_2$). Podmínka

nesoudělnosti je zachovaná: jsou-li oba polynomy m_1, m_2 nesoudělné se všemi m_i, pak je s nimi nesoudělný i polynom $m_1 m_2$ (opět se využije gaussovskost).

Úloha. Najděte polynom $f \in \mathbb{Q}[x]$ stupně < 4 splňující

$$f \equiv 1 \pmod{x^2-1} \quad \text{a} \quad f \equiv x+1 \pmod{x^2+1}.$$

Řešení. Z druhé kongruence vyjádříme $f = g \cdot (x^2+1) + x + 1$ a dosadíme do první kongruence: budeme hledat $g \in \mathbb{Q}[x]$ splňující $g \cdot (x^2+1) \equiv -x \pmod{x^2-1}$. Všimněte si, že $\widetilde{x^2+1} = \frac{1}{2}$ (protože $x^2+1 \equiv 2$), takže dostaneme vyjádření $g \equiv -\frac{1}{2}x \pmod{x^2-1}$, čili řešením je každý polynom $g = h \cdot (x^2-1) - \frac{1}{2}x$, $h \in \mathbb{Q}[x]$. Zpětným dosazením dostaneme obecné řešení

$$f = h \cdot (x^2-1)(x^2+1) - \frac{1}{2}x(x^2+1) + (x+1), \ h \in \mathbb{Q}[x],$$

a hledaný polynom $f = -\frac{1}{2}x^3 + \frac{1}{2}x + 1$ dostaneme volbou $h = 0$. □

Buď $f \in T[x]$ a $a, u \in T$. Všimněte si, že

$$f(a) = u \ \Leftrightarrow f \equiv u \pmod{x-a}.$$

Skutečně, $x-a$ dělí polynom $f-u$ právě tehdy, když je a kořenem polynomu $f-u$, tedy když $f(a) - u = 0$.

Speciálním případem čínské věty o zbytcích je tedy *věta o interpolaci*: ta říká, že pokud předepíšeme hodnoty v n různých bodech, pak existuje právě jeden polynom stupně $< n$, který těchto hodnot v daných bodech nabývá.

Důsledek 8.2 (věta o interpolaci)**.** *Buď* $\mathbf{T}$ *těleso. Mějme po dvou různé body* $a_1, \ldots, a_n \in T$ *a libovolné hodnoty* $u_1, \ldots, u_n \in T$*. Pak existuje právě jeden polynom* $f \in T[x]$ *stupně* $< n$ *splňující* $f(a_i) = u_i$ *pro všechna* $i = 1, \ldots, n$*.*

Důkaz. Řešíme soustavu kongruencí $f \equiv u_i \pmod{x - a_i}$. □

Na rozdíl od obecné věty o zbytcích není těžké nalézt vzorec, který určuje řešení interpolační úlohy: je jím tzv. *Lagrangeův interpolační polynom*

$$f = \sum_{i=1}^{n} \left(u_i \cdot \prod_{j \neq i} \frac{x - a_j}{a_i - a_j} \right).$$

Dosazením do vzorce snadno zjistíme, že

$$f(a_k) = 0 + \ldots + 0 + u_k \cdot \prod_{j \neq k} \frac{a_k - a_j}{a_k - a_j} + 0 + \ldots + 0 = u_k.$$

Jednoznačnost pak plyne z věty o počtu kořenů vztažené na rozdíl $f - g$ dvou řešení.

Důsledek 8.3 (zobrazení na konečných tělesech jsou polynomiální)**.** *Buď* $\mathbf{T}$ *konečné těleso. Pak pro každé zobrazení* $\varphi : T \to T$ *existuje právě jeden polynom* $f \in T[x]$ *stupně* $< |T|$ *takový, že* $\varphi(a) = f(a)$ *pro každé* $a \in T$*.*

Důkaz. Interpolujme v bodě a hodnotou $\varphi(a)$ pro každé $a \in T$. □

Pro *nekonečná* tělesa nic takového platit nemůže, přesto polynomy hrají důležitou roli i v reálné analýze: každou spojitou reálnou funkci lze libovolně přesně aproximovat polynomiální funkcí, v různých smyslech. Například, lokální aproximaci (na okolí daného bodu) popisují Taylorovy polynomy, globální (na intervalu) třeba Weierstrassova věta, která říká, že pro každou spojitou reálnou funkci $\varphi : [u, v] \to \mathbb{R}$

na omezeném uzavřeném intervalu a pro každé $\varepsilon > 0$ existuje polynom $f \in \mathbb{R}[x]$ takový, že $|\varphi(a) - f(a)| < \varepsilon$ pro každé $a \in [u, v]$.

Cvičení 8.4. *Vymyslete analogii Lagrangeova vzorce na interpolaci polynomů více proměnných a dokažte analogii Důsledku 8.3 pro zobrazení $\varphi : T^n \to T$, pro libovolné $n \in \mathbb{N}$.*

8.2. **Faktorokruh modulo polynom.**

Připomeňme konstrukci okruhů $\mathbb{Z}_m$. Začali jsme s oborem celých čísel a uvažovali všechny možné zbytky po dělení m, tj. čísla $0, \ldots, m-1$, a na nich operace modulo m. Pokud bylo m prvočíslo, dostali jsme těleso. Podobný postup lze provést i s polynomy, dostaneme tzv. *faktorokruhy*. Aby se nepletla proměnná v polynomech s prvky faktorokruhu, obvykle se v konstrukci používá proměnná α.

Definice. Buď $\mathbf{T}$ těleso a zvolme polynom $m \in T[\alpha]$ stupně $n \geq 1$. *Faktorokruhem* $\mathbf{T}[\alpha]/(m)$ rozumíme množinu všech polynomů stupně $< n$ se standardními operacemi sčítání a odčítání a s operací násobení modulo m. Ve zkratce,

$$\mathbf{T}[\alpha]/(m) = (\{f \in T[\alpha] : \ \deg f < n\}, +, -, \odot, 0, 1),$$

kde $f \odot g = f \cdot g \bmod m$.

Předně je potřeba dokázat, že to je skutečně komutativní okruh. Axiomy obsahující pouze sčítání a odčítání jsou zřejmé, protože tyto operace jsou totožné jako v $\mathbf{T}[x]$. Pro úvahy s násobením je třeba si připomenout, že $f \equiv g \pmod m \Leftrightarrow f \bmod m = g \bmod m$, a že $f \bmod m \equiv f \pmod m$. Tímto způsobem lze všechny identity přeložit do kongruencí, kde je platnost zřejmá. Například pro asocitivitu dokazujeme

$$(f \odot g) \odot h = f \odot (g \odot h),$$

tj.

$$(f \cdot g \bmod m) \cdot h \bmod m = f \cdot (g \cdot h \bmod m) \bmod m,$$

což je ekvivalentní kongruenci

$$(f \cdot g) \cdot h \equiv f \cdot (g \cdot h) \pmod m,$$

což je pravda pro všechny polynomy f, g, h. Podobně můžeme ověřit distributivitu.

Příklad. Uvažujme faktorokruh $\mathbb{R}[\alpha]/(\alpha^2 + 1)$. Jeho prvky jsou polynomy $a + b\alpha$, $a, b \in \mathbb{R}$. Sčítání probíhá po složkách, tj. $(a + b\alpha) + (c + d\alpha) = (a + c) + (b + d)\alpha$. Násobení vypadá takto:

$$\begin{aligned}(a + b\alpha) \odot (c + d\alpha) &= ac + (ad + bc)\alpha + bd\alpha^2 \bmod (\alpha^2 + 1) \\ &= (ac - bd) + (ad + bc)\alpha.\end{aligned}$$

Všimněte si, že jsme dostali stejné vzorce jako pro sčítání a násobení komplexních čísel. Při ztotožnění symbolů i a α bychom mohli psát, že $\mathbb{R}[\alpha]/(\alpha^2 + 1) = \mathbb{C}$ (formálně, zobrazení $a + b\alpha \mapsto a + bi$ je izomorfismus). Vysvětlení je prosté: při počítání modulo α^2+1 vlastně zaměňujeme α^2 za -1, neboť $\alpha^2 \equiv -1 \pmod{\alpha^2+1}$. Čili pracujeme přesně s vlastností, která definuje komplexní jednotku.

Podobně lze nahlédnout, že faktorokruh $\mathbb{Q}[\alpha]/(\alpha^2 + 1)$ je v principu totožný s tělesem $\mathbb{Q}(i)$ (formálně: *izomorfní*, tj. jedno z druhého dostaneme přejmenováním prvků).

Příklad. Nad tělesy $\mathbb{Z}_p$ jsou vlastnosti faktorokruhu závislé na p.

- Faktorokruh $\mathbb{Z}_2[\alpha]/(\alpha^2+1)$ má čtyři prvky, ale není to těleso, dokonce ani obor integrity, protože
$$(\alpha+1)\odot(\alpha+1) = \alpha^2+1 \bmod (\alpha^2+1) = 0.$$
- Faktorokruh $\mathbb{Z}_3[\alpha]/(\alpha^2+1)$ má devět prvků. Je to těleso, ale na první pohled to vidět není.

Kdy dostaneme těleso vysvětluje následující tvrzení.

Tvrzení 8.5 (faktor podle ireducibilního prvku)**.** *Buď* $\mathbf{T}$ *těleso a* $m \in T[\alpha]$ *stupně* ≥ 1*. Následující tvrzení jsou ekvivalentní:*

(1) $\mathbf{T}[\alpha]/(m)$ *je těleso,*
(2) $\mathbf{T}[\alpha]/(m)$ *je obor integrity,*
(3) m *je ireducibilní prvek v* $\mathbf{T}[\alpha]$.

Důkaz. (1) $\Rightarrow$ (2) viz Tvrzení 3.3.

(2) $\Rightarrow$ (3). Pro spor předpokládejme, že v $\mathbf{T}[\alpha]$ existuje rozklad $m = f \cdot g$, kde $\deg f, \deg g < \deg m$. Pak ale v $\mathbf{T}[\alpha]/(m)$ platí $f \odot g = m \bmod m = 0$, spor.

(3) $\Rightarrow$ (1). Uvažujme polynom $f \neq 0$ stupně menšího než $\deg m$. Protože je m ireducibilní, platí $1 = \mathrm{NSD}(f,m) = uf + vm$ pro nějaké polynomy $u, v \in T[\alpha]$. Označme $\tilde{u} = u \bmod m$. Pak v $\mathbf{T}[\alpha]/(m)$ platí $\tilde{u} \odot f = \tilde{u}f \bmod m \equiv uf \equiv 1 \pmod m$, čili $\tilde{u}$ je hledaný inverzní prvek k f. □

V dalším textu budeme místo symbolu $\odot$ psát standardní symbol násobení; z kontextu bude vždy jasné, že jde o násobení ve faktorokruhu, tedy modulo m (stejně se používá standardní symbol pro násobení v okruzích $\mathbb{Z}_m$).

Právě popsaná konstrukce se používá ke konstrukci konečných těles, kterým se budeme věnovat v příští sekci. Teď si ukážeme jinou důležitou aplikaci: každé těleso lze rozšířit tak, aby v něm měl daný polynom kořen. Pro racionální polynomy to zní triviálně, každý racionální polynom má přece komplexní kořen, ale tento fakt je předmětem Základní věty algebry, a tu není vůbec jednoduché dokázat. Naopak, existence rozkladového nadtělesa je stěžejním krokem k jejímu důkazu. A pro konečná tělesa žádnou analogii použít nelze.

Tvrzení 8.6. *Buď* $\mathbf{T}$ *těleso a* $f \in T[x]$ *stupně* ≥ 1*. Pak existuje těleso* $\mathbf{S} \geq \mathbf{T}$*, kde má polynom* f *kořen.*

Důkaz. Buď m nějaký ireducibilní dělitel polynomu f, označme $m = \sum_{i=0}^{n} a_i x^i$. Stačí najít nadtěleso, kde má kořen polynom m, ten bude kořenem i pro f. Uvažujme faktorokruh $\mathbf{S} = \mathbf{T}[\alpha]/(m(\alpha))$. Podle Tvrzení 8.5 je $\mathbf{S}$ těleso. Vyhodnotíme-li v $\mathbf{S}$ polynom m na prvku α, dostaneme
$$m(\alpha) = \sum_{i=0}^{n} a_i(\alpha^i \bmod m(\alpha)) = \sum_{i=0}^{n-1} a_i\alpha^i + a_n(\alpha^n \bmod m(\alpha)),$$
ovšem $a_n\alpha^n \bmod m(\alpha) = -\sum_{i=0}^{n-1} a_i\alpha^i$, takže se to odečte na nulu. Prvek α je tedy kořenem obou polynomů m, f v nadtělese $\mathbf{S}$. □

Příklad.

- Pro polynom x^3-2 nad tělesem $\mathbb{Q}$ dostaneme těleso $\mathbb{Q}[\alpha]/(\alpha^3-2)$, které lze pomocí výše uvedených úvah ztotožnit s tělesem $\mathbb{Q}(\sqrt[3]{2})$.
- Pro polynom x^3-2 nad tělesem $\mathbb{Z}_7$ dostaneme těleso $\mathbb{Z}_7[\alpha]/(\alpha^3-2)$, což je těleso s 7^3 prvky, které jste ještě asi nepotkali.

Indukcí snadno dokážeme, že existuje také nadtěleso, kde má daný polynom všechny kořeny, tj. kde se rozkládá na součin lineárních činitelů (polynomů stupně 1).

Důsledek 8.7. *Buď* $\mathbf{T}$ *těleso a* $f \in T[x]$ *stupně* ≥ 1. *Pak existuje těleso* $\mathbf{S} \geq \mathbf{T}$, *kde se polynom* f *rozkládá na součin polynomů stupně 1.*

Důkaz. Budeme postupovat indukcí podle stupně polynomu f. Je-li f stupně 1, $f = ax - b$, pak má kořen $a^{-1}b$ již v tělese $\mathbf{T}$. V opačném případě uvažujme nadtěleso $\mathbf{U} \geq \mathbf{T}$, kde má polynom f kořen u, a uvažujme polynom $g \in U[x]$ takový že $f = g \cdot (x - u)$. Protože $\deg g < \deg f$, podle indukčního předpokladu existuje nadtěleso $\mathbf{S} \geq \mathbf{U}$, kde se g rozkládá na součin polynomů stupně 1, čili se tam rozkládá i f. □

9. Konečná tělesa a jejich aplikace

9.1. Konečná tělesa a počítačová reprezentace dat.

Důležitou aplikací faktorokruhů je konstrukce konečných těles. Buď p prvočíslo a uvažujme ireducibilní polynom $m \in \mathbb{Z}_p[\alpha]$ stupně k. Faktorokruh $\mathbb{Z}_p[\alpha]/(m)$ je podle Tvrzení 8.5 tělesem, jeho prvky jsou polynomy stupně $< k$ nad $\mathbb{Z}_p$, čili toto těleso má právě p^k prvků. Například,

- čtyřprvkové těleso můžeme zkonstruovat jako $\mathbb{Z}_2[\alpha]/(\alpha^2 + \alpha + 1)$,
- osmiprvkové jako $\mathbb{Z}_2[\alpha]/(\alpha^3 + \alpha + 1)$ nebo $\mathbb{Z}_2[\alpha]/(\alpha^3 + \alpha^2 + 1)$,
- devítiprvkové jako $\mathbb{Z}_3[\alpha]/(\alpha^2 + 1)$ nebo $\mathbb{Z}_3[\alpha]/(\alpha^2 \pm \alpha + 2)$.

Pozor, pro $k > 1$ je p^k-prvkové těleso něco jiného než okruh $\mathbb{Z}_{p^k}$!

V pokročilejším kurzu algebry se dokazuje následující stěžejní věta.

Věta 9.1. *Buď* p *prvočíslo,* $k \in \mathbb{N}$.

(1) *Existuje ireducibilní polynom stupně* k *nad* $\mathbb{Z}_p$, *čili existuje konečné těleso velikosti* p^k,
(2) *každé konečné těleso velikosti* p^k *lze sestrojit jako faktorokruh* $\mathbb{Z}_p[\alpha]/(m)$, *pro nějaký polynom* m *stupně* k,
(3) *na volbě* m *nezáleží, tj. jsou-li* m_1, m_2 *dva ireducibilní polynomy stupně* k *nad tělesem* $\mathbb{Z}_p$, *pak jsou tělesa* $\mathbb{Z}_p[\alpha]/(m_1)$ *a* $\mathbb{Z}_p[\alpha]/(m_2)$ *izomorfní.*

Důkaz uvedených vlastností je složitější, než se teď může zdát. Dokázat samotnou existenci konečného tělesa velikosti p^k není těžké: dostaneme jej jako rozkladové nadtěleso polynomu $x^{p^k} - x$ nad tělesem $\mathbb{Z}_p$ (konstrukcí ve stylu Důsledku 8.7). Avšak k reprezentaci pomocí faktorokruhů je potřeba spousta dalších ingrediencí, jako je teorie tělesových rozšíření konečného stupně, jednoznačnost rozkladových nadtěles, a také znalosti o struktuře cyklických grup.

Konečné těleso velikosti p^k budeme značit $\mathbb{F}_{p^k}$ (používá se také označení $\mathbf{GF}(p^k)$, jako *Galois field*). Vzhledem k tomu, že jsou stejně velká tělesa izomorfní, na konkrétní reprezentaci zpravidla nezáleží.

Konečná tělesa, zejména ta velikosti 2^k, mají zásadní využití v informatice. Jejich pomocí lze reprezentovat počítačová data a provádět s nimi různé operace, anebo v nich datové operace analyzovat. Reprezentaci dat si nyní vysvětlíme.

Základním datovým objektem, se kterým pracují počítače, jsou tzv. *bitvektory*, tedy k-tice nul a jedniček, tzv. *bitů*. Bitvektory délky k lze přirozeně reprezentovat pomocí konečného tělesa $\mathbb{F}_{2^k} = \mathbb{Z}_2[\alpha]/(m)$, kde m je ireducibilní stupně k: vektor $(a_0, \ldots, a_{k-1})$ se reprezentuje polynomem $a_0 + a_1\alpha + \ldots + a_{k-1}\alpha^{k-1}$.

Obrázek 4. Bitvektor a jeho reprezentace polynomem.

Běžné operace na bitvektorech mají často přirozenou tělesovou interpretaci. Například posun bitvektoru vlevo či vpravo odpovídá násobení a dělení prvkem α. Logická spojka XOR po bitech odpovídá tělesovému sčítání, logická spojka AND po bitech odpovídá násobení po koeficientech (pozor, to je něco jiného, než násobení v tělese). Konečná tělesa přinášejí navíc dvě zajímavé operace, které pracují se všemi bity najednou: tělesové násobení a invertování. Tyto operace mají zajímavé vlastnosti, které nacházejí uplatnění v konstrukci šifer (zjednodušeně řečeno, dokážou rozprostřít lokální změnu na všechny pozice).

Příklad. V současnosti nejpoužívanější symetrická šifra AES (*Advanced Encryption Standard*, též známá jako *Rijndael*) pracuje s bitvektory délky 8, které reprezentuje pomocí prvků tělesa

$$\mathbb{F}_{256} = \mathbb{Z}_2[\alpha]/(\alpha^8 + \alpha^4 + \alpha^3 + \alpha + 1).$$

Text je rozdělen na bloky po 128 bitech a každý blok je reprezentován jako matice 4×4 prvků tělesa $\mathbb{F}_{256}$. Šifra opakuje pro každý blok několikrát za sebou čtyři fáze (první a poslední průběh je trochu jiný, ale to teď není důležité). V první se provádí pro každý prvek matice následující operace:

$$u \mapsto u^{-1} \cdot (1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4) + (1 + \alpha + \alpha^5 + \alpha^8) \bmod (\alpha^8 + 1),$$

kde inverz se bere v tělese $\mathbb{F}_{256}$ a zbytek výpočtu probíhá v $\mathbb{Z}_2[\alpha]$. V druhé fázi se rotuje každý řádek o jistý počet pozic. Ve třetí se mixuje každý sloupec tak, že se interpretuje jako polynom z $\mathbb{F}_{256}[x]$ stupně < 4, na který se provede operace

$$f \mapsto f \cdot (\alpha + x + x^2 + (\alpha + 1)x^3) \bmod (x^4 + 1).$$

Ve čtvrté fázi se pak přičítá jistým způsobem vybraná část klíče (po bitech). Smysl prvních tří fází je rozprostřít změnu provedenou přičítáním klíče do celé tabulky (míchání prvků, řádků, sloupců), a to tak, aby se co nejlépe ztratily slabiny opakování klíče. Pro praxi je zásadní, že lze velmi rychle nejen šifrovat, ale také dešifrovat: není těžké odvodit, že operace inverzní k těm výše uvedeným mají podobně jednoduchý algebraický zápis.

Velký význam má také Důsledek 8.3 a jeho zobecnění pro zobrazení více proměnných (viz cvičení v sekci 8.1), které říká, že každou operaci na datech lze interpretovat jako polynomiální zobrazení nad příslušným tělesem. Tento fakt nachází uplatnění například v kryptoanalýze. Mezi další aplikace věty o interpolaci patří samoopravné kódy (sekce 9.3) nebo sdílení tajemství (sekce 9.2).

Další oblastí aplikace konečných těles jsou konečné geometrie (afinní a projektivní prostory nad konečnými tělesy). Konečné geometrie jsou zdrojem zajímavých kombinatorických objektů, jak si ukážeme v sekci 9.4 na příkladu konstrukce navzájem ortogonálních latinských čtverců pomocí afinních zobrazení. Konečné geometrie jsou také zdrojem zajímavých výpočetních problémů, jako je například počítání na eliptických křivkách nad konečnými tělesy, opět s aplikací v návrhu šifer.

9.2. **Sdílení tajemství.**

Motivační úloha je následující: armáda má tajný kód, který umožňuje odpálit jaderné rakety. Zřejmě není dobré, aby jeden šílenec mohl odpálit rakety o své vůli. Ani dva šílenci by neměli mít možnost odpálit rakety. Prezident nařídil, že k odpálení raket je potřeba souhlas aspoň tří šílenců z pětičlenného generálního štábu. Jak to zařídit?

Obecně hovoříme o (k, n)-*schématu sdílení tajemství*, pokud se n účastníků dělí o tajemství, k jehož odhalení je potřeba přítomnost alespoň k z nich (kterýchkoliv). V celém odstavci budeme uvažovat, že sdílíme tajemství t z nějakého tělesa $\mathbf{T}$ (v praxi se sdílí bitvektor délky m, interpretovaný buď jako m tajemství z tělesa $\mathbf{T} = \mathbb{Z}_2$, nebo jedno tajemství z $\mathbf{T} = \mathbb{F}_{2^m}$).

Pro případ $k = n$ lze použít jednoduché schéma založené na *maskování hodnot*. Vlastník tajemství vydá každému účastníku náhodný prvek $a_i \in T$ a zveřejní hodnotu $c = t + \sum a_i$. Pokud má dojít k odhalení, každý účastník sdělí své a_i a společně spočtou $t = c - \sum a_i$. Pokud se sejde účastníků méně, byť jen $n - 1$, o hodnotě t nemohou říci vůbec nic: chybějící prvek může hodnotu součtu změnit na libovolnou jinou hodnotu, s pravděpodobností přesně $\frac{1}{|T|}$ (protože zobrazení $x \mapsto a + x$ je permutace). V praxi se používá těleso $\mathbb{Z}_2$: pravděpodobnost uhodnutí jednoho bitu je $\frac{1}{2}$, čili pro m-bitový klíč je pravděpodobnost $(\frac{1}{2})^m$.

Klasickým řešením obecného (k, n)-schématu je tzv. *Shamirův protokol*. Vlastník tajemství náhodně zvolí polynom $f \in T[x]$ stupně $< k$ takový, že $f(0) = t$ (tj. tajemství je absolutní člen f), vybere n po dvou různých prvků $0 \neq a_1, \ldots, a_n \in T$ (ta mohou být veřejná) a jednotlivým účastníkům rozdá hodnoty $f(a_1), \ldots, f(a_n)$. Pokud se sejde libovolných k účastníků, vezmou své hodnoty, provedou interpolaci ve svých bodech a spočtou (ten jediný) polynom stupně $< k$, který vyhovuje jejich podmínkám; tajemství je jeho absolutní člen. Naopak, pokud se jich sejde méně, byť jen $k-1$, o absolutním členu nezjistí nic: $k-1$ nenulovými body lze proložit polynom s libovolnou hodnotou v bodě 0, a navíc rozložení hodnot v 0 je rovnoměrné. V praxi se pro m-bitový klíč používá těleso s 2^m prvky (musí být $2^m > n$), které zajistí pravděpodobnost náhodného uhodnutí $\frac{1}{2^m}$.

Schéma lze snadno modifikovat pro sofistikovanější úlohy. Například, co kdyby prezident rozhodl, že rakety mohou odpálit buď aspoň tři z pěti šílených generálů, nebo on sám? Snadná pomoc: vyrobíme $(3, 8)$-schéma, každému z generálů dáme po jednom dílu a prezidentu dáme tři. A tak podobně. V reálném životě se schéma používá například pro rozhodnutí komisí, tajemstvím je klíč k elektronickému podpisu.

Cvičení 9.2. *V dvoupatrovém úřadě v Kocourkově sídlí 20 úředníků, v každém patře 10, a ředitel. Úřad smí vydat rozhodnutí s kulatým razítkem, je-li přítomno aspoň 5 úředníků z 1. patra a 3 z 2. patra, nebo aspoň 2 z 1. patra, 8 z 2. patra a ředitel. Navrhněte schéma sdílení klíče k sejfu s kulatým razítkem. (Podobnost s Magistrátem hlavního města Prahy je čistě náhodná.)*

9.3. **Samoopravné kódy.**

Problém zní: jak v proudu dat odhalit a odstranit náhodně vznikající chyby? Typickou situací je přenos informace nespolehlivým kanálem (šum, ztráta dat atd.), ale také dlouhodobé uložení v paměti. Matematický model situace je následující: vysílač pošle slovo délky k nad abecedou A (často $A = \mathbb{Z}_2$, ale používají se i jiná konečná tělesa), kanál náhodně změní některá písmena, a přijímač má odhalit, zda

došlo k chybě, či přímo najít původní zprávu. V základních metodách se (ne zcela realisticky) předpokládá, že v každém slově délky k nastane nejvýše e chyb.

Asi nejjednodušším schematem je test parity: ke zprávě sestávající z k bitů přidáme jeden bit, který bude součtem všech bitů v původní zprávě (modulo 2). Pokud nastane při přenosu právě jedna chyba, kontrolní součet vyjde špatně a víme, že je třeba přenos opakovat. Systém je časově i prostorově efektivní, ale neumožňuje chybu opravit — nevíme, kde nastala. A pokud je chyb sudý počet, test mylně projde.

Druhým velmi jednoduchým schematem je opakování písmen: každé písmeno zopakujeme n-krát a předpokládáme, že v každé po sobě jdoucí n-tici nastane méně než $n/2$ chyb. Zprávu pak snadno zrekonstruujeme hlasováním: kterého znaku je v přijaté n-tici víc, ten byl v původní zprávě. Schema má vcelku slabý předpoklad na spolehlivost kanálu, ale je prostorově náročné: délka zprávy se z-n-násobí. Existuje lepší postup?

Nejprve si shrneme, co hledáme. Každé slovo délky k chceme nahradit kódovým slovem délky $n \geq k$, přičemž tato kódová slova by měla být dostatečně odlišná, aby při záměně omezeného množství znaků bylo možné jednoznačně zrekonstruovat původní zprávu. Nejprve definujeme tzv. *Hammingovu vzdálenost*: pro slova $u, v \in A^n$ je vzdálenost $\delta(u, v)$ rovna počtu pozic, na kterých se tato slova liší. *Samoopravným kódem* typu $(k, n; d)$ pak je libovolné prosté zobrazení $\varphi : A^k \to A^n$ takové, že pro všechna $u, v \in \varphi(A^k)$, $u \neq v$, platí $\delta(u, v) \geq d$. Například, přidání paritního bitu je $(k, k+1; 2)$-kód, zatímco opakování znaku je $(1, n; n)$-kód.

Označme $C = \varphi(A^k)$ množinu *kódových slov*. Pro praktické využití je dále potřeba, aby

- kódovací i dekódovací zobrazení $\varphi : A^k \to C$ a $\varphi^{-1} : C \to A^k$ byly efektivně počítatelné;
- pro každé $u \in A^n$ šlo efektivně najít kódové slovo $v \in C$ takové, že vzdálenost $\delta(u, v)$ je nejmenší možná.

Často se používají tzv. *lineární kódy*: abeceda je reprezentována konečným tělesem a požaduje se, aby zobrazení φ bylo lineární, a tedy aby množina C byla podprostorem.

Důležitým pozorováním je, že kód typu $(k, n; d)$ je schopen opravit $e = \lfloor \frac{d-1}{2} \rfloor$ chyb: při záměně $\leq e$ znaků ve slově u vznikne slovo v takové, že $\delta(u, v) \leq e$. Protože jsou kódová slova vzdálená alespoň $2e + 1$, slovo u je nutně nejbližším kódovým slovem k v.

První zajímavý samoopravný kód objevil Richard Hamming kolem roku 1950 při konstrukci prvních počítačů v Bellových laboratořích. Tzv. *Hammingův* $(4, 7)$*-kód* je schopen opravit jednu chybu v každé přenášené sedmici, za cenu prodloužení zprávy na $\frac{7}{4}$ původní délky. Základní myšlenkou je volit $\varphi : \mathbb{Z}_2^4 \to \mathbb{Z}_2^7$ jako lineární zobrazení, které za původní slovo délky 4 přídá kontrolní sekvenci délky 3 tak, aby různá slova byla od sebe dostatečně vzdálená. Konkrétně, kód je dán maticí

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix},$$

lineární zobrazení předpisem $\varphi(u) = uM$. Všimněte si, že obrazy bázových vektorů (tj. řádky matice) jsou vzdáleny aspoň 3, a není těžké dokázat, že tento odhad platí

pro všechny dvojice kódových slov. Jde tedy o kód typu $(4, 7; 3)$, který opravuje 1 chybu. Ověření, zda je dané slovo kódové, je velmi snadné: vezmeme první čtyři znaky, ty zakódujeme a ověříme, zda vyšlo stejné kódové slovo. Pokud ne, předpokládaje, že nastala právě jedna chyba, máme přesně 7 možností, kde mohla být, a všechny rychle ověříme. Pomocí lineární algebry lze najít i efektivnější postup dekódování, ale to je již mimo rozsah našeho textu.

V roce 1960 byly objeveny tzv. *Reed-Salomonovy kódy* založené na interpolaci polynomů. Jejich poměr prodloužení zprávy vůči počtu opravovaných chyb je v jistém smyslu optimální a patří v praxi mezi nejpoužívanější kódy. Abecedou je konečné těleso $\mathbf{T}$. Zvolíme n pod dvou různých prvků $u_1, \dots, u_n \in T$. Kódovat budeme k-tice prvků T interpretované jako polynomy stupně $< k$. Kódovým slovem pak bude n-tice hodnot v uvedených bodech. Formálně, Reed-Salomonovým (k, n)-kódem je zobrazení

$$\varphi : T^k \to T^n, \quad f = \sum a_i x^i \mapsto (f(u_1), \dots, f(u_n)).$$

Inverzním zobrazením je interpolace v daných bodech. Jaká je vzdálenost kódových slov? Pokud se dva polynomy stupně $< k$ shodují v k hodnotách, musí být totožné (jednoznačnost v Důsledku 8.2). Jinými slovy, různé polynomy f, g mají $< k$ stejných hodnot, čili $> n - k$ různých hodnot, takže jde o kód typu $(k, n; d)$ pro nějaké $d \geq n - k + 1$, opravující alespoň $\lfloor \frac{n-k}{2} \rfloor$ chyb.

V praxi se často používá $\mathbf{T} = \mathbb{F}_{256}$, $n = 255$ a k se volí podle spolehlivosti kanálu (čím menší k, tím více chyb kód opravuje, ale tím horší je poměr délek původních a kódových slov). Například pro $k = 253$ kód opravuje jednu chybu za cenu prodloužení slov cca o 1%, pro $k = 127$ kód opravuje 64 chyb za cenu prodloužení slov na dvojnásobek. Dále se v praxi používá speciální volba bodů $u_i = \alpha^{i-1}$, $i = 1, \dots, 255$, kde α je generátor cyklické grupy $\mathbf{T}^*$ (viz sekce 13.2), aby bylo možné použít na kódování a dekódování rychlou Fourierovu transformaci (standardní algoritmy dosazování a interpolace běží v kvadratickém čase, což je příliš pomalé).

K úspěšnému uvedení do praxe zbývá popsat algoritmus na vyhledání nejbližšího kódového slova. Pro krátké kódy opravující jednu chybu lze zkoušet všechny možnosti (jako u Hammingova kódu, viz cvičení), ale obecně to není snadná úloha a čtenáře odkazujeme do specializované literatury.

9.4. **Vzájemně ortogonální latinské čtverce a návrh experimentů.**

Latinským čtvercem na množině X rozumíme čtvercovou matici $(a_{i,j})_{i,j \in I}$ indexovanou množinou I s prvky $a_{i,j}$ z množiny X splňující následující podmínku: každý prvek množiny X je v každém řádku a každém sloupci právě jednou (z toho ihned plyne $|I| = |X|$). Pokud $I = X$, můžeme se na latinský čtverec dívat jako na binární operaci, kde $u * v = a_{u,v}$.

Příklad.

(1) Libovolné řešení sudoku je latinským čtvercem na množině $X = \{1, \dots, 9\}$.

(2) Následující tabulky jsou latinské čtverce na množině $X = \{0, 1, 2\}$:

0	1	2
1	2	0
2	0	1

0	2	1
2	1	0
1	0	2

0	2	1
1	0	2
2	1	0

Příslušné operace lze zapsat jako $u * v = u + v \bmod 3$, resp. $-u - v \bmod 3$, resp. $u - v \bmod 3$.

(3) Je-li **R** okruh, pak operace + určuje latinský čtverec na množině R daný předpisem $(a+b)_{a,b\in R}$. Podobně, je-li **T** těleso, pak operace $\cdot$ na nenulových prvcích určuje latinský čtverec, $(a \cdot b)_{a,b\in T^*}$. Obecněji, jak uvidíme v sekci 10, multiplikační tabulky grup jsou latinské čtverce.

Dva latinské čtverce $(a_{i,j})_{i,j\in I}$ a $(b_{i,j})_{i,j\in I}$ na množinách X a Y nazveme *vzájemně ortogonální* pokud se každá dvojice z $X\times Y$ vyskytuje na seznamu $((a_{i,j}, b_{i,j}) : i, j \in I)$ právě jednou.

Příklad. Vzájemně ortogonální čtverce 2×2 neexistují, neboť jedinými latinskými čtverci na $X = \{0, 1\}$ jsou

$$\begin{array}{|cc|}\hline 0 & 1\\ 1 & 0\\ \hline\end{array} \qquad \begin{array}{|cc|}\hline 1 & 0\\ 0 & 1\\ \hline\end{array}$$

a ty ortogonální nejsou: dvojice $(0, 1)$ a $(1, 0)$ se opakují dvakrát, naopak dvojice $(0, 0)$ a $(1, 1)$ tam chybí.

Vzájemně ortogonální čtverce 3×3 existují, jsou jimi například první a třetí tabulka ve výše uvedeném příkladu. Naopak první a druhá vzájemně ortogonální nejsou, dvojice $(0, 0)$ se opakuje třikrát.

Úloha. Z balíčku karet vytáhneme figury J,Q,K,A od každé ze čtyř barev. Rozložte karty do čtverce 4×4 tak, aby v každém řádku a každém sloupci bylo po jedné kartě od každé barvy a každé figury.

Řešení. Úloha se vlastně ptá po nalezení dvou vzájemně ortogonálních čtverců na čtyřech prvcích. V prvním případě jde o množinu $X = \{J, Q, K, A\}$, v druhém případě o množinu $Y = \{\diamondsuit, \heartsuit, \clubsuit, \spadesuit\}$. Čtverce musí být latinské, protože každý řádek a sloupec má obsahovat po jednom symbolu. A čtverce musí být vzájemně ortogonální, protože každou kartu (tj. dvojici symbolů z $X \times Y$) musíme použít právě jednou. Řešení není vidět na první pohled, ale není těžké ověřit, že jím je následující konfigurace:

A$\diamondsuit$	K$\heartsuit$	Q$\clubsuit$	J$\spadesuit$
K$\clubsuit$	A$\spadesuit$	J$\diamondsuit$	Q$\heartsuit$
Q$\spadesuit$	J$\clubsuit$	A$\heartsuit$	K$\diamondsuit$
J$\heartsuit$	Q$\diamondsuit$	K$\spadesuit$	A$\clubsuit$

Návodem k jejímu nalezení bude důkaz Tvrzení 9.3. □

Úlohu lze samozřejmě zobecnit na libovolné množství karet/barev, nebo na hledání většího množství navzájem ortogonálních čtverců. Problémem se zabýval již Leonhard Euler v 18. století a traduje se, že jeho zájem byl podnícen následující úlohou, kterou nedokázal vyřešit.

Úloha. Vojenské přehlídky se má zúčastnit 36 vojáků z šesti regimentů, z každého po šesti různých hodnostech. Sestavte vojáky do čtverce 6×6 tak, aby v každém řádku a každém sloupci bylo po jednom zástupci každého regimentu a každé hodnosti.

Vidíme, že jde opět o nalezení dvou vzájemně ortogonálních čtverců, tentokrát na šesti prvcích. Až Gaston Tarry v roce 1901 ukázal, že takové čtverce neexistují.

Euler roku 1782 objevil dva typy konstrukcí, které jsou obsahem následujících dvou tvrzení (ačkoliv sám by je popsal jinak, v jeho době nebyl znám koncept konečných těles).

Tvrzení 9.3. *Buď n mocnina prvočísla, $n \neq 2$. Pak existuje $n-1$ po dvou vzájemně ortogonálních latinských čtverců.*

Důkaz. Uvažujme těleso $\mathbf{T}$ s n prvky a pro každé $0 \neq a \in T$ definujeme matici $(au+v)_{u,v\in T}$. Tyto matice jsou latinskými čtverci: kdyby se v řádku u vyskytovaly na pozici v_1, v_2 dva stejné prvky, tedy kdyby $au + v_1 = au + v_2$, odečtením au dostaneme $v_1 = v_2$, a podobně pro sloupce, kdyby $au_1+v = au_2+v$, pak odečtením a vykrácením $u_1 = u_2$. Protože jde o konečné čtverce, každý prvek se musí vyskytovat právě jednou. Pro různá a, b jsou tyto čtverce vzájemně ortogonální: kdyby se na dvou pozicích (u_1, v_1), (u_2, v_2) vyskytoval ten samý prvek, tedy kdyby

$$au_1 + v_1 = au_2 + v_2 \quad \text{a} \quad bu_1 + v_1 = bu_2 + v_2,$$

pak $a(u_1 - u_2) = v_2 - v_1 = b(u_1 - u_2)$, a protože $a \neq b$, musí být $u_1 = u_2$, a tudíž také $v_1 = v_2$. □

Tvrzení 9.4. *Pokud existují vzájemně ortogonální latinské čtverce velikostí m a n, pak existují také velikosti $m \cdot n$.*

Důkaz. Označme jednu dvojici $(a_{i,j})_{i,j\in I}$, $(b_{i,j})_{i,j\in I}$ na množinách X_1, X_2, a druhou dvojici $(u_{k,l})_{k,l\in J}$, $(v_{k,l})_{k,l\in J}$ na množinách Y_1, Y_2. Definujeme matici $((a_{i,j}, u_{k,l}))_{(i,k),(j,l)\in I\times J}$ na množině $X_1 \times Y_1$ a matici $((b_{i,j}, v_{k,l}))_{(i,k),(j,l)\in I\times J}$ na množině $X_2 \times Y_2$. Vidíme, že jde o matice velikosti $m \cdot n$. Jako cvičení ověřte, že jsou tyto matice latinskými čtverci a že jsou vzájemně ortogonální. □

Důsledek 9.5. *Je-li $n \not\equiv 2 \pmod 4$, pak existují vzájemně ortogonální latinské čtverce velikosti n.*

Důkaz. Pro $n = p_1^{k_1} \cdot \ldots \cdot p_m^{k_m}$, kde $p_i^{k_i}$ jsou libovolné mocniny prvočísla s výjimkou 2^1, lze vzít ortogonální čtverce velikosti $p_i^{k_i}$ zkonstruované v Tvrzení 9.3 a na ty aplikovat Tvrzení 9.4. Jedinou výjimku tak tvoří čísla tvaru dva krát liché číslo, tj. $n \equiv 2 \pmod 4$. □

Pro $n = 2$ jsme si již ukázali, že vzájemně ortogonální čtverce neexistují, pro $n = 6$ jde o zmíněný Tarryho výsledek. V roce 1958 pak byla nalezena konstrukce pro všechna $n \geq 10$, čímž byl problém existence vyřešen. Dodnes však zůstává nejasné, kolik po dvou vzájemně ortogonálních čtverců velikosti n může existovat. Není těžké dokázat, že jich je nejvýše $n-1$, čili Tvrzení 9.3 dává optimální odpověď pro mocniny prvočísla. V případě $n = 10$ se ví, že jich není 9, ale přesný počet není znám. Je známo, že existence $n - 1$ vzájemně ortogonálních latinských čtverců je ekvivalentní existenci projektivní roviny řádu n, což dále ukazuje na složitost problému díky jeho souvislosti z konečnými geometriemi.

A jak je to se zmíněným *návrhem experimentů*? Uvažujme následující zadání: máme k dispozici n odrůd dané plodiny, n druhů hnojiva a n typů pesticidu. Chceme zjistit, která kombinace je nejlepší. Kdybychom chtěli zkoušet všechny kombinace, potřebujeme n^3 experimentů; to je mnoho. Nestačilo by n^2? Podmínka je, že každý objekt (odrůdu, hnojivo, pesticid) musíme použít stejně-krát (tedy n-krát) a dále bychom rádi, aby se každá dvojice objektů použila právě jednou, abychom mohli zkoumat závislost jednotlivých veličin. Odpovědí je libovolný latinský čtverec $(a_{i,j})$ na množině $\{1, \ldots, n\}$: experimentální pole rozdělíme na $n \times n$ políček, do i-tého řádku sejeme i-tou odrůdu, do j-tého sloupce sypeme j-té hnojivo a na pole s indexem (i, j) dáme $a_{i,j}$-tý pesticid.

Nyní uvažujme další položku, třeba n stupňů zavlažování. Uvažujme druhý latinský čtverec $(b_{i,j})$ a budeme zavlažovat pole s indexem (i, j) dávkou $b_{i,j}$. Není dobré brát tento čtverec libovolně: pokud například zvolíme $b_{i,j} = a_{i,j}$, pak bude pro daný pesticid vždy stejná úroveň zavlažování a o závislosti těchto veličin nezjistíme nic. Jednou z dobrých voleb jsou vzájemně ortogonální čtverce, které zajistí, že každá kombinace pesticid–zavlažování se objeví právě jednou.

Tato jednoduchá úloha byla jedním z počátků rozsáhlé *teorie designů*, která se zabývá konstrukcí objektů s nejrůznějšími požadavky na vyváženost, s aplikací v návrhu statistických experimentů.

Grupy

10. Pojem grupy

10.1. **Definice a příklady.**

Hlavní motivací teorie grup je studium nejrůznějších typů symetrií a transformací matematických objektů. Pojem pochází z Galoisovy teorie a původně označoval množinu (skupinu) permutací G uzavřenou na skládání, tj. splňující $\pi \circ \sigma \in G$ pro všechna $\pi, \sigma \in G$. Abstrakcí tohoto pojmu vznikla rozsáhlá větev algebry, zvaná teorie grup. Aplikace nachází všude, kde se vyskytuje pojem symetrie či transformace, především v kombinatorice (konečné grupy) a geometrii (lineární grupy).

Definice. *Grupou* rozumíme čtveřici $\mathbf{G} = (G, *, ', e)$, kde G je množina, na které jsou definovány binární operace $*$, unární operace $'$ a konstanta e splňující pro každé $a, b, c \in G$ následující podmínky:

$$a * (b * c) = (a * b) * c, \qquad a * e = e * a = a, \qquad a * a' = a' * a = e.$$

Grupu nazýváme *abelovskou*, pokud navíc pro všechna $a, b \in G$ platí

$$a * b = b * a.$$

Prvku e se říká *jednotka*, prvku a' *inverzní prvek* k prvku a.

Formálně rozlišujeme mezi množinou G, tzv. *nosnou množinou*, a čtveřicí $\mathbf{G} = (G, *, ', e)$, která navíc obsahuje informaci o algebraické struktuře definované na množině G. V konkrétních příkladech bývá typickou trojicí operací buď $+, -, 0$, pak hovoříme o *aditivním zápise* (a místo $x + (-y)$ píšeme $x - y$), anebo trojice $\cdot, ^{-1}, 1$, čemuž říkáme *multiplikativní zápis*.

Definice. Buď $\mathbf{G} = (G, *, ', e)$ grupa a $H \subseteq G$ podmnožina její nosné množiny taková, že $e \in H$ a pro každé $a, b \in H$ platí

$$a' \in H \quad \text{a} \quad a * b \in H.$$

Říkáme, že H je *uzavřena na grupové operace* a že *tvoří podgrupu* grupy $\mathbf{G}$. Čtveřici $\mathbf{H} = (H, *|_H, '|_H, e)$ pak nazýváme *podgrupou*, přičemž $|_H$ značí restrikci operací na množinu H. Značíme $\mathbf{H} \leq \mathbf{G}$. Podgrupy $\mathbf{G}$ a $\{e\}$ nazýváme *nevlastní*.

V matematice se vyskytují nejrůznější příklady grup, nicméně je možné identifikovat čtyři základní rodiny, které nacházejí asi největší využití: permutační grupy, maticové grupy, grupy geometrických zobrazení a číselné grupy.

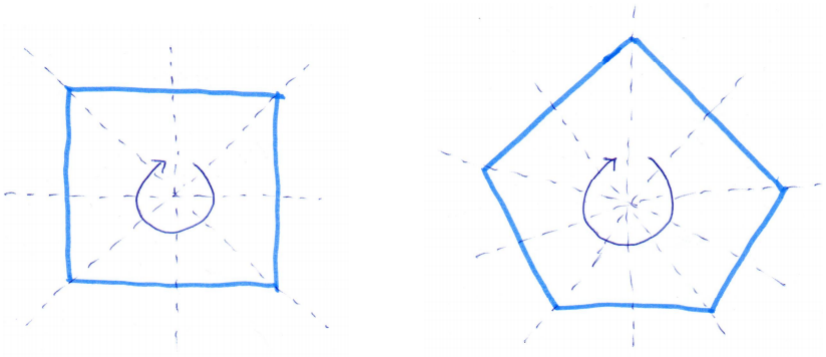
Příklad. *Permutační grupy*[1]. Základním příkladem je *symetrická grupa* sestávající z permutací na dané neprázdné množině X s operacemi $\circ$ skládání permutací, $^{-1}$ invertování permutací a konstantou $id : x \mapsto x$ (identické zobrazení), tj.

$$\mathbf{S}_X = (\{\pi : \ \pi \text{ je permutace na množině } X\}, \circ, ^{-1}, id).$$

Je-li $X = \{1, \ldots, n\}$, pak místo $\mathbf{S}_X$ píšeme $\mathbf{S}_n$. Podgrupy symetrických grup se nazývají *permutační grupy*, např.

- *alternující grupa* $\mathbf{A}_n \leq \mathbf{S}_n$ všech sudých permutací na n prvcích;
- *dihedrální grupa* $\mathbf{D}_{2n} \leq \mathbf{S}_n$ všech permutací, které odpovídají symetriím pravidelného n-úhelníka vztaženým na jeho vrcholy očíslované po směru hodinových ručiček. Tyto permutace odpovídají n rotacím a n reflexím, proto značení $\mathbf{D}_{2n}$.

[1]Typický čtenář by měl znát základní fakta o permutacích z kurzu lineární algebry nebo diskrétní matematiky. Ostatním doporučujeme nahlédnut do sekce 13.3, kde jsou tyto znalosti zopakovány a doplněny.



Obrázek 5. Grupy $\mathbf{D}_8$ a $\mathbf{D}_{10}$.

- nejrůznější grupy symetrií geometrických těles, automorfismů grafů a dalších matematických struktur.

Příklad. Speciálním případem permutačních grup jsou grupy zobrazení na různých typech geometrických prostorů (eukleidovské, afinní, projektivní apod.) zachovávajících jisté vlastnosti (afinní zobrazení, projektivní zobrazení apod.). Příkladem je *eukleidovská grupa* $\mathbf{E}_n$ sestávající ze všech izometrií (tj. zobrazení zachovávajících vzdálenosti) eukleidovského prostoru $\mathbb{R}^n$.

Na grupy symetrií geometrických objektů daných konečně mnoha body (např. na grupy $\mathbf{D}_{2n}$ pro pravidelné n-úhelníky) lze nahlížet dvojím způsobem: jako na podgrupu eukleidovské grupy, sestávající z izometrií zachovávajících daný objekt, nebo jako na permutační grupu na bodech, které tento objekt určují.

Příklad. *Maticové grupy*[2]. Základním příkladem je *obecná lineární grupa* nad tělesem $\mathbf{T}$ sestávající z regulárních matic dané velikosti s operacemi $\cdot$ maticového násobení, $^{-1}$ maticového invertování a jednotkovou maticí jako jednotkou, tj.

$$\mathbf{GL}_n(\mathbf{T}) = (\{A : \ A \text{ je regulární matice } n \times n \text{ nad tělesem } \mathbf{T}\}, \cdot, ^{-1}, I),$$

Podgrupy lineárních grup se nazývají *maticové grupy*, např.

- *speciální lineární grupa* $\mathbf{SL}_n(\mathbf{T})$ všech matic s determinantem 1;
- *ortogonální grupa* $\mathbf{O}_n(\mathbf{T})$ všech ortogonálních matic, tj. takových A, které splňují $AA^T = I$ (nad tělesem $\mathbb{R}$ jde o matice, jejichž řádky, resp. sloupce, jsou ortonormální vektory vzhledem k standardnímu skalárnímu součinu).

Permutační a maticové grupy jsou v jistém smyslu univerzální příklady: každou grupu lze reprezentovat jako permutační grupu a každou konečnou grupu lze reprezentovat jako maticovou grupu. Více se dozvíte v letním semestru.

Pro některé grupy je taková reprezentace přirozená, ale leckdy ne: příkladem je osmiprvková kvaternionová grupa.

Příklad. *Kvaternionová grupa* $\mathbf{Q}_8$ je definovaná na množině $\{\pm 1, \pm i, \pm j, \pm k\}$. Násobení je dáno vzorci

$$i^2 = j^2 = k^2 = -1, \quad ij = k, \quad jk = i, \quad ki = j,$$

a dále pravidly $xy = -(yx)$ a $(-x)y = x(-y) = -(xy)$ pro všechna $x, y \in \{i, j, k\}$.

[2]Typický čtenář by měl znát základní fakta o maticích z kurzu lineární algebry.

Základním zdrojem příkladů abelovských grup jsou grupy odvozené od komutativních okruhů, zejména pak *číselné grupy*, odvozené od číselných oborů.

Příklad. Buď $\mathbf{R}$ okruh. Pak $(R, +, -, 0)$ je abelovská grupa, tzv. *aditivní grupa* okruhu $\mathbf{R}$. Důležité jsou zejména číselné grupy $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, a také grupy $\mathbb{Z}_n$ sestávající z čísel $0, \ldots, n-1$ s operacemi modulo n.

Příklad. Buď $\mathbf{R}$ komutativní okruh s jednotkou, označme R^* množinu všech invertibilních prvků v $\mathbf{R}$. Pak $\mathbf{R}^* = (R^*, \cdot, ^{-1}, 1)$ je abelovská grupa, tzv. *multiplikativní grupa* okruhu $\mathbf{R}$. Skutečně jde o grupu: inverz invertibilního prvku je invertibilní (protože $(a^{-1}) \cdot a = 1$), součin dvou invertibilních prvků a, b je invertibilní (protože $(ab)(b^{-1}a^{-1}) = 1$) a grupové axiomy jsou obsaženy v definici okruhu.

- Je-li $\mathbf{R}$ těleso, pak $\mathbf{R}^* = (R \setminus \{0\}, \cdot, ^{-1}, 1)$.
- Pro polynomiální okruhy platí $\mathbf{R}[x]^* = \mathbf{R}^*$, protože invertibilní jsou právě konstantní polynomy invertibilní v $\mathbf{R}$.
- $\mathbb{Z}^* = (\{1, -1\}, \cdot, ^{-1}, 1)$.
- Prvky grupy $\mathbb{Z}_n^*$ jsou právě všechna čísla $a \in \{1, \ldots, n-1\}$ nesoudělná s n. Soudělná čísla invertibilní nejsou: je-li $d \nmid 1$ společný dělitel a, n, pak $d \mid (ab \bmod n)$ pro libovolné b, takže součin ab nikdy nemůže být 1. Naopak, jsou-li a, n nesoudělná, uvažujme Bézoutovy koeficienty u, v splňující $1 = \mathrm{NSD}(a, n) = ua + vn$. Podíváme-li se na rovnost modulo n, dostaneme $1 \equiv ua \pmod{n}$, a tedy $a^{-1} = u \bmod n$.

Existuje řada dalších geometrických i algebraických konstrukcí abelovských grup, například grupy odvozené od eliptických křivek nebo třídové grupy prvoideálů v číselných tělesech. Některé z těchto konstrukcí mají významné aplikace v kryptografii (sekce 13.3), v praxi se hojně využívá například Diffie-Hellmanův protokol s grupami na eliptických křivkách nad konečnými tělesy.

Důležitou konstrukcí grup je direktní součin.

Definice. *Direktním součinem* grup $\mathbf{G}_i = (G_i, *_i, {}'^i, e_i)$, $i = 1, \ldots, n$, rozumíme grupu

$$\prod_{i=1}^{n} \mathbf{G}_i = \mathbf{G}_1 \times \cdots \times \mathbf{G}_n = (G_1 \times \cdots \times G_n, *, ', e),$$

jejíž operace jsou definovány po složkách, tj.

$$\begin{aligned}(a_1, \ldots, a_n) * (b_1, \ldots, b_n) &= (a_1 *_1 b_1, \ldots, a_n *_n b_n),\\ (a_1, \ldots, a_n)' &= ((a_1)'^1, \ldots, (a_n)'^n),\\ e &= (e_1, \ldots, e_n).\end{aligned}$$

pro všechna $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in G_1 \times \ldots \times G_n$. Je snadné ověřit, že direktní součin splňuje všechny axiomy grup.

V případě, kdy $\mathbf{G}_1 = \ldots = \mathbf{G}_n = \mathbf{G}$, hovoříme o *direktní mocnině* a značíme ji $\mathbf{G}^n$.

Analogicky bychom mohli definovat také direktní součin okruhů, či jakéhokoliv jiného typu algebraických struktur.

Podobně jako v případě komutativních okruhů, definice grupy obsahuje pouze minimální množství podmínek. Následující tvrzení ukazuje několik aritmetických pravidel (mj. krácení, jednoznačnost jednotky a jednoznačnost inverzních prvků), které z definice snadno plynou a v dalším textu je budeme volně používat.

Tvrzení 10.1 (základní vlastnosti grup)**.** *Buď* $\mathbf{G} = (G, *, ', e)$ *grupa a* $a, b, c \in G$. *Pak*

(1) *jestliže* $a * c = b * c$ *nebo* $c * a = c * b$, *pak* $a = b$;
(2) *jestliže* $a * u = a$ *nebo* $u * a = a$ *pro nějaké* $u \in G$, *pak* $u = e$;
(3) *jestliže* $a * u = e$ *nebo* $u * a = e$ *pro nějaké* $u \in G$, *pak* $u = a'$;
(4) $(a')' = a$;
(5) $(a * b)' = b' * a'$.

Důkaz. (1) Je-li $a * c = b * c$, pak také $(a * c) * c' = (b * c) * c'$ a použitím všech tří axiomů dostaneme $(a * c) * c' = a * (c * c') = a * e = a$ a podobně $(b * c) * c' = b$. Tedy $a = b$. Analogicky pro $c * a = c * b$.

(2) Je-li $a * u = a = a * e$, krácením dostáváme $u = e$. Analogicky pro $u * a = a$.

(3) Je-li $a * u = e = a * a'$, krácením dostáváme $u = a'$. Analogicky pro $u * a = e$.

(4) Protože $a' * a = e$, z jednoznačnosti inverzních prvků dostáváme $a = (a')'$.

(5) Protože $(a*b)*(b'*a') = a*(b*b')*a' = a*e*a' = a*a' = e$, z jednoznačnosti inverzních prvků dostáváme $(a * b)' = b' * a'$. □

10.2. **Mocniny a řád prvku.**

Čtenář si snad již zažil, že se grupové operace značí nejrůznějšími způsoby. Nadále se budeme držet multiplikativního zápisu. Nebude-li výslovně uvedeno jinak, uvažujeme-li grupu $\mathbf{G}$, implicitně rozumíme $\mathbf{G} = (G, \cdot, ^{-1}, 1)$. Ze začátku je dobré si u všech výrazů rozmýšlet, jak bychom je přepsali do ostatních značení.

Nyní definujeme *mocniny*. Buď $\mathbf{G}$ grupa, $a \in G$, $n \in \mathbb{Z}$. Označme

$$a^n = \begin{cases} 1 & n = 0 \\ \underbrace{a \cdot a \cdot \ldots \cdot a}_{n} & n > 0 \\ \underbrace{a^{-1} \cdot a^{-1} \cdot \ldots \cdot a^{-1}}_{-n} & n < 0 \end{cases}$$

Tvrzení 10.2 (mocniny)**.** *Buď* $\mathbf{G}$ *grupa,* $a, b \in G$ *a* $k, l \in \mathbb{Z}$. *Pak*

$$a^{k+l} = a^k \cdot a^l, \qquad a^{kl} = (a^k)^l = (a^l)^k$$

a je-li $\mathbf{G}$ *abelovská, pak navíc* $(ab)^k = a^k b^k$.

Důkaz. Pokud $k, l > 0$, ihned vidíme, že počet prvků a ve výrazech na obou stranách každé rovnosti je stejný. V případě záporných exponentů je třeba vzít v úvahu, že a a a^{-1} se navzájem pokrátí. Např. v první rovnosti, pro $k > 0 > l$, $|l| < |k|$, máme na levé straně součin $k + l$ prvků a, zatímco na pravé straně součin k prvků a a $-l$ prvků a^{-1}. Po vykrácení dostaneme rovnost obou výrazů. Ostatní případy se rozeberou podobně. □

V aditivním značení je mocninou výraz $a + \ldots + a$, resp. $(-a) + \ldots + (-a)$; tyto výrazy zkracujeme jako $n \cdot a$. Tvrzení 10.2 se pak přepíše jako

$$(k + l) \cdot a = k \cdot a + l \cdot a, \qquad (kl) \cdot a = k \cdot (l \cdot a), \qquad k \cdot (a + b) = k \cdot a + k \cdot b,$$

poslední rovnost samozřejmě platí pouze pro abelovské grupy. Pokud vám tyto podmínky připomínají definici vektorového prostoru, jste na správně stopě. Teorie abelovských grup je do značné míry teorií „vektorových prostorů nad $\mathbb{Z}$“, neboli $\mathbb{Z}$-modulů, s řadou aplikací v teorii čísel. Tímto směrem se však v úvodním kurzu ubírat nebudeme.

Definice. *Řádem grupy* $\mathbf{G}$ se rozumí počet prvků její nosné množiny, značíme jej $|\mathbf{G}|$ (tj., formálně vzato, $|\mathbf{G}| = |G|$).

Řádem prvku a v grupě $\mathbf{G}$ se rozumí nejmenší $n \in \mathbb{N}$ takové, že $a^n = 1$, pokud takové n existuje, resp. ∞ v opačném případě. Značíme jej $\mathrm{ord}(a)$.

V Tvrzení 11.6 si ukážeme, že řád prvku je roven řádu jisté podgrupy, ale zatím si vystačíme s definicí pomocí mocnin.

Příklad. Pokud mluvíme o řádu jistého prvku, je třeba říci, v které grupě!

- $\mathrm{ord}(2) = 7$ v grupě $\mathbb{Z}_7$, protože $7 \cdot 2 \equiv 0 \pmod 7$, ale $n \cdot 2 \not\equiv 0 \pmod 7$ pro $n = 1, \ldots, 6$;
- $\mathrm{ord}(2) = 3$ v grupě $\mathbb{Z}_7^*$, protože $2^3 \equiv 1 \pmod 7$, ale $2^n \not\equiv 1 \pmod 7$ pro $n = 1, 2$.

Příklad. V nekonečných grupách mohou řády vycházet všelijak:

- v grupě $\mathbb{Q}$ je $\mathrm{ord}(0) = 1$ a $\mathrm{ord}(a) = \infty$ pro všechna $a \neq 0$;
- v grupě $\mathbb{Q}^*$ je $\mathrm{ord}(1) = 1$, $\mathrm{ord}(-1) = 2$ a $\mathrm{ord}(a) = \infty$ pro všechna $a \neq \pm 1$;
- v grupě $\mathbb{C}^*$ existuje prvek libovolného řádu: $\mathrm{ord}(e^{2\pi i/k}) = k$.

Příklad. V konečných grupách řády nevycházejí všelijak:

- v grupě $\mathbb{Z}_6$ je $\mathrm{ord}(0) = 1$, $\mathrm{ord}(1) = 6$, $\mathrm{ord}(2) = 3$, $\mathrm{ord}(3) = 2$, $\mathrm{ord}(4) = 3$ a $\mathrm{ord}(5) = 6$, čili vyskytují se řády $1, 2, 3, 6$;
- v grupě $\mathbf{S}_3$ je $\mathrm{ord}(id) = 1$, $\mathrm{ord}((i\ j)) = 2$, $\mathrm{ord}((i\ j\ k)) = 3$, čili vyskytují se řády $1, 2, 3$.

Všimněte si, že řád každého prvku dělí řád celé grupy. To není náhoda, nýbrž pravidlo, které je speciálním případem Lagrangeovy věty (Věta 11.9), která je náplní příští sekce.

Tvrzení 10.3 (řád permutace)**.** *Řád permutace v grupě* $\mathbf{S}_n$ *je roven nejmenšímu společnému násobku délek jejích cyklů.*

Důkaz. Cyklus délky n má řád n. Jsou-li $C_1, \ldots, C_m$ disjunktní cykly v π, pak $\pi^k = (C_1 \circ \ldots \circ C_m)^k = C_1^k \circ \ldots \circ C_m^k$. Z toho plyne, že $(C_1 \circ \ldots \circ C_m)^k = id$ právě tehdy, když je k násobkem všech délek cyklů. Čili řád je roven nejmenšímu společnému násobku. □

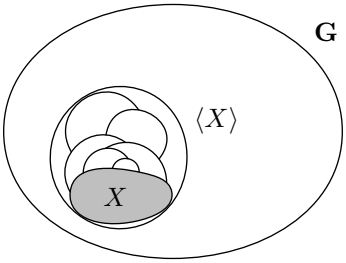
11. Podgrupy

11.1. Generátory.

Lemma 11.1. *Průnik podgrup je podgrupa.*

Důkaz. Buď $\mathbf{G}$ grupa, uvažujme podgrupy $\mathbf{H}_i$, $i \in I$, a označme $H = \bigcap_{i \in I} H_i$. Dokážeme, že je množina H uzavřená na grupové operace. Protože $1 \in H_i$ pro všechna $i \in I$, bude 1 náležet i jejich průniku. Nyní uvažujme $a, b \in H$. Tyto leží v každém H_i a díky uzavřenosti na operace tam leží také prvky a^{-1} a $a \cdot b$. Takže tyto prvky leží i v průniku všech H_i, čili v H. □

Definice. Uvažujme podmnožinu $X \subseteq G$ grupy $\mathbf{G}$. Podgrupou *generovanou množinou* X rozumíme nejmenší podgrupu (vzhledem k inkluzi) grupy $\mathbf{G}$ obsahující podmnožinu X, značíme ji $\langle X \rangle_{\mathbf{G}}$.



Obrázek 6. Ilustrace generování podgrupy $\langle X\rangle_{\mathbf{G}}$.

Taková podgrupa jistě existuje: stačí vzít průnik všech podgrup obsahujících množinu X, tj.

$$\langle X\rangle_{\mathbf{G}} = \bigcap\{H :\ X \subseteq H,\ \mathbf{H} \leq \mathbf{G}\}.$$

Podle předchozího lemmatu jde skutečně o podgrupu, mezi všemi podgrupami obsahujícími množinu X bude jistě nejmenší.

Jak najít podgrupu generovanou danou množinou? Pro konečné grupy lze v principu použít následující postup: začneme s prvky množiny X a postupně přidáváme všemožné součiny a inverzy. Ve chvíli, kdy nejsme schopni získat žádné nové prvky, naše množina je uzavřená na grupové operace a podgrupa je nalezena (viz obrázek). Leckdy je však efektivnější použít následující tvrzení.

Tvrzení 11.2. *Buď* $\mathbf{G}$ *grupa a* $\emptyset \neq X \subseteq G$. *Pak*

$$\langle X\rangle_{\mathbf{G}} = \{a_1^{k_1} \cdot a_2^{k_2} \cdot \ldots \cdot a_n^{k_n} :\ n \in \mathbb{N},\ a_1, \ldots, a_n \in X,\ k_1, \ldots, k_n \in \mathbb{Z}\}.$$

Důkaz. Označme M množinu na pravé straně rovnosti. Je potřeba dokázat, že množina M

(1) tvoří podgrupu,
(2) obsahuje X,
(3) je nejmenší podmnožinou grupy $\mathbf{G}$ splňující tyto podmínky.

(1) Součin dvou prvků z M je jistě v M, jednotka $1 = a^0$ je tam také, a uzavřenost na inverzy plyne ze vztahu $(a_1^{k_1} \cdot \ldots \cdot a_n^{k_n})^{-1} = a_n^{-k_n} \cdot \ldots \cdot a_1^{-k_1} \in M$.

(2) Volbou $n = 1, k_1 = 1$ dostaneme libovolný prvek X.

(3) Uvažujme libovolnou podgrupu $\mathbf{H}$ obsahující X. Tato podgrupa musí obsahovat všechny mocniny a^i, $a \in X$, i jejich libovolné násobky, čili celé M. □

Obecné tvrzení o tvaru podgrup generovaných danou podmnožinou má dva důležité speciální případy.

Důsledek 11.3. *Buď* $\mathbf{G}$ *grupa a* $a \in G$. *Pak* $\langle a\rangle_{\mathbf{G}} = \{a^k :\ k \in \mathbb{Z}\}$.

Důsledek 11.4. *Buď* $\mathbf{G}$ *abelovská grupa a* $u_1, \ldots, u_n \in G$. *Pak*

$$\langle u_1, \ldots, u_n\rangle_{\mathbf{G}} = \{u_1^{k_1} \cdot u_2^{k_2} \cdot \ldots \cdot u_n^{k_n} :\ k_1, \ldots, k_n \in \mathbb{Z}\}.$$

Vidíme, že v abelovských grupách je generování podgrup podobné jako generování vektorových prostorů: v aditivním zápise, tj. pro abelovskou grupu $\mathbf{G} = (G, +, -, 0)$, dostáváme

$$\langle u_1, \ldots, u_n\rangle_{\mathbf{G}} = \{k_1u_1 + k_2u_2 + \ldots + k_nu_n :\ k_1, \ldots, k_n \in \mathbb{Z}\}.$$

(S nezávislostí a jednoznačností zápisu je to složitější, ale tím se v této učebnici zabývat nebudeme.)

Příklad. Důležitým typem úlohy je zjistit, jakou podgrupu generuje daná podmnožina. Například:

- $\langle \frac{3}{4}, \frac{1}{3} \rangle_{\mathbb{Q}} = \{k\frac{3}{4} + l\frac{1}{3} : \ k, l \in \mathbb{Z}\} = \{\frac{k}{12} : \ k \in \mathbb{Z}\} = \langle \frac{1}{12} \rangle_{\mathbb{Q}}$. První a poslední rovnost plynou z Důsledku 11.4. K důkazu prostřední je potřeba si uvědomit, že na jednu stranu $\frac{3}{4}, \frac{1}{3} \in \langle \frac{1}{12} \rangle_{\mathbb{Q}}$, a na druhou stranu $\frac{1}{12} = \frac{3}{4} - 2 \cdot \frac{1}{3}$, a tedy $\frac{1}{12} \in \langle \frac{3}{4}, \frac{1}{3} \rangle_{\mathbb{Q}}$.
- $\langle \frac{3}{4}, \frac{1}{3} \rangle_{\mathbb{Q}^*} = \{(\frac{3}{4})^k \cdot (\frac{1}{3})^l : \ k, l \in \mathbb{Z}\} = \{3^k \cdot 4^l : \ k, l \in \mathbb{Z}\}$.

Příklad. Jiným důležitým typem úlohy je, najít k dané grupě $\mathbf{G}$ co nejmenší množinu generátorů, tj. podmnožinu $X \subseteq G$ takovou, že $\mathbf{G} = \langle X \rangle_{\mathbf{G}}$. Například:

- $\mathbb{Z} = \langle 1 \rangle = \langle -1 \rangle$, $\mathbb{Z}^* = \langle -1 \rangle$, $\mathbb{Q}^* = \langle -1, \text{prvočísla} \rangle$.
- $\mathbb{Z}_n = \langle 1 \rangle$, ale najít malou generující množinu grupy $\mathbb{Z}_n^*$ není obecně snadné. Například, $\mathbb{Z}_7^* = \langle 3 \rangle$, ale $\mathbb{Z}_8^* = \langle 3, 5 \rangle$ a nelze ji nagenerovat jedním prvkem.
- Pro některé grupy neexistuje minimální množina generátorů. Např. $\mathbb{Q} = \langle \frac{1}{n} : n \in \mathbb{N} \rangle$, libovolné konečné množství generátorů lze vypustit, ale minimální podmnožina neexistuje.

Tvrzení 11.5 (generátory permutačních grup)**.**

(1) *Grupa $\mathbf{S}_n$ je generovaná množinou všech transpozic.*
(2) *Grupa $\mathbf{A}_n$ je generovaná množinou všech trojcyklů.*

Důkaz. (1) Danou permutaci nejprve napíšeme jako složení svých cyklů a každý cyklus pak rozložíme podle následujícího vzoru:

$$(a_1\ a_2\ \ldots\ a_k) = (a_1\ a_k) \circ \ldots \circ (a_1\ a_3) \circ (a_1\ a_2).$$

(2) Danou sudou permutaci nejprve napíšeme jako složení sudého počtu transpozic a ty seskupíme do sousedících dvojic. Pokud jsou sousedící transpozice stejné, můžeme je vypustit. Pokud mají společný jeden prvek, pak $(i\ j) \circ (j\ k) = (i\ j\ k)$. A jsou-li disjunktní, pak $(i\ j) \circ (k\ l) = (k\ i\ l) \circ (i\ j\ k)$. Tímto způsobem přepíšeme rozklad na transpozice na složení trojcyklů. □

Uvedené množiny generátorů nejsou optimální. Například grupu $\mathbf{S}_n$ je možné generovat jednou transpozicí a jedním n-cyklem. Více příkladů najdete v cvičeních.

Příklad. Ukážeme, že

$$\mathbf{S}_n = \langle (1\ 2), (1\ 2\ \ldots\ n) \rangle.$$

Díky Tvrzení 11.5 stačí dokázat, že lze nagenerovat všechny transpozice. Nejprve nagenerujeme transpozice $(k\ k+1)$, $k = 1, \ldots, n-1$: induktivně

$$(k+1\ k+2) = (1\ 2\ \ldots\ n) \circ (k\ k+1) \circ (1\ 2\ \ldots\ n)^{-1}.$$

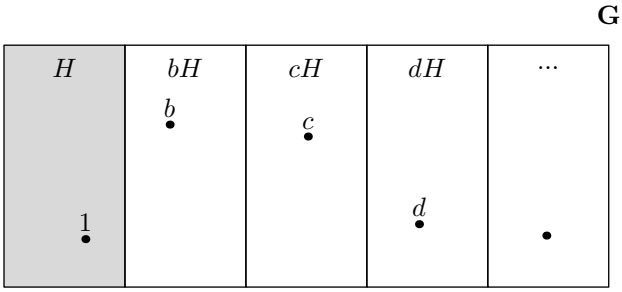
Dále, pro každé k nagenerujeme ostatní transpozice $(k\ k+i)$, $i > 0$: opět induktivně

$$(k\ k+i+1) = (k+i\ k+i+1) \circ (k\ k+i) \circ (k+i\ k+i+1)^{-1}.$$

Na závěr dokážeme, že řád prvku (definovaný pomocí mocnin) je roven řádu podgrupy jím generované.

Tvrzení 11.6 (řád prvku vs. řád podgrupy)**.** *Buď $\mathbf{G}$ grupa a $a \in G$. Pak*

$$\operatorname{ord}(a) = |\langle a \rangle_{\mathbf{G}}|.$$



Obrázek 7. Rozklad grupy $\mathbf{G}$ podle podgrupy $\mathbf{H}$ a jeho transverzála.

Důkaz. Podle Důsledku 11.3 je $\langle a\rangle_{\mathbf{G}} = \{a^k :\ k \in \mathbb{Z}\}$. Všimněte si, že $a^i = a^j$ právě tehdy, když $a^{i-j} = 1$. Je-li $\operatorname{ord}(a) = \infty$, pak žádné $n \neq 0$ s vlastností $a^n = 1$ neexistuje, čili mocniny a^k jsou po dvou různé a podgrupa $\langle a\rangle$ je nekonečná. Je-li $\operatorname{ord}(a) = n < \infty$, pak jsou mocniny $a^0, a^1, \ldots, a^{n-1}$ po dvou různé, ovšem další mocniny nové prvky nepřidají: $a^n = a^0 = 1$, $a^{n+1} = a^n \cdot a^1 = a^1$, $a^{n+2} = a^n \cdot a^2 = a^2$ atd., obecně $a^{qn+r} = (a^n)^q \cdot a^r = a^r$. Tedy $\langle a\rangle_{\mathbf{G}} = \{a^0, a^1, \ldots, a^{n-1}\}$ obsahuje přesně n prvků. □

11.2. **Lagrangeova věta.**

Základní aritmetickou vlastností konečných grup je fakt, že řády podgrup dělí řád celé grupy, tj.

$$\mathbf{H} \leq \mathbf{G} \quad \Rightarrow \quad |\mathbf{H}| \text{ dělí } |\mathbf{G}|.$$

Speciálně, díky Tvrzení 11.6, řád prvku dělí řád celé grupy.

Myšlenka důkazu Lagrangeovy věty není složitá: celou grupu $\mathbf{G}$ rozložíme na několik podmnožin, které jsou po dvou disjunktní a stejně velké jako daná podgrupa $\mathbf{H}$. Počet prvků grupy $\mathbf{G}$ tak bude roven počtu prvků $\mathbf{H}$ krát počet těchto podmnožin. Nesamozřejmou částí důkazu je konstrukce tohoto rozkladu.

Definice. Buď $\mathbf{G}$ grupa a $\mathbf{H}$ její podgrupa:

- množiny $aH = \{ah :\ h \in H\}$, kde $a \in G$, se nazývají *rozkladové třídy* podgrupy $\mathbf{H}$;
- podmnožina $T \subseteq G$ s vlastností $|T \cap aH| = 1$ pro každé $a \in G$ se nazývá *transverzála* rozkladu $\mathbf{G}$ podle $\mathbf{H}$;
- počet rozkladových tříd se nazývá *index* podgrupy $\mathbf{H}$ v grupě $\mathbf{G}$ a značí se

$$[\mathbf{G} : \mathbf{H}] = |\{aH :\ a \in G\}|.$$

Pojmy, které jsme definovali, se někdy používají s přívlastkem *levý*, tj. levé rozkladové třídy, levá transverzála, levý index. *Pravými rozkladovými třídami* pak rozumíme množiny $Ha = \{ha :\ h \in H\}$ a ostatní pojmy se definují analogicky. Levé a pravé varianty mohou být stejné či různé (viz příklady níže), ale jak uvidíme, počet rozkladových tříd, tj. index, vyjde z obou stran stejně.

Příklad. Buď $\mathbf{G} = \mathbb{Z}$ a $\mathbf{H} = \{h \in \mathbb{Z} :\ n \mid h\}$. Rozkladovou třídu určenou prvkem $a \in \mathbb{Z}$ můžeme vyjádřit

$$aH = \{a + h :\ h \in H\} = \{a + nk :\ k \in \mathbb{Z}\} = \{u \in \mathbb{Z} :\ u \equiv a \pmod n\}.$$

Dvě rozkladové třídy aH, bH jsou buď stejné, nebo disjunktní, přičemž $aH = bH$ právě tehdy, když $a \equiv b \pmod n$. Dostáváme tak n různých po dvou disjunktních

rozkladových tříd, $[\mathbf{G} : \mathbf{H}] = n$. Jako transverzálu lze zvolit např. $T = \{0, \dots, n-1\}$, množinu všech možných zbytků po dělení n.

Příklad. Buď $\mathbf{G} = \mathbf{S}_n$ a $\mathbf{H} = \mathbf{A}_n$. Pak $\pi A_n = A_n \pi = A_n$ pro libovolnou π sudou a $\pi A_n = A_n \pi$ sestává ze všech lichých permutací pro libovolnou π lichou. Grupa $\mathbf{S}_n$ se tedy rozkládá na dvě disjunktní rozkladové třídy (levé i pravé jsou stejné), $[\mathbf{S}_n : \mathbf{A}_n] = 2$ a jako transverzálu lze zvolit např. $T = \{id, (1\ 2)\}$.

Levé a pravé rozkladové třídy nemusí být vždy stejné, nejmenším příkladem je následující situace.

Příklad. Buď $\mathbf{G} = \mathbf{S}_3$ a $\mathbf{H} = \{id, (1\ 2)\}$. Snadno spočteme, že levý i pravý rozklad obsahuje tři dvouprvkové třídy, avšak

$$(1\ 3)H = \{(1\ 3), (1\ 2\ 3)\}, \quad \text{ale} \quad H(1\ 3) = \{(1\ 3), (1\ 3\ 2)\}.$$

Lagrangeovu větu dokážeme pomocí dvou základních vlastností rozkladů: za prvé, různé rozkladové třídy jsou disjunktní, a za druhé, všechny rozkladové třídy jsou stejně velké. Analogická tvrzení platí i pro pravé rozkladové třídy.

Lemma 11.7 (disjunkce rozkladových tříd)**.** *Buď* $\mathbf{G}$ *grupa a* $\mathbf{H}$ *její podgrupa. Pro každé* $a, b \in G$ *platí buď* $aH = bH$*, nebo* $aH \cap bH = \emptyset$*.*

Důkaz. Předpokládejme $aH \cap bH \neq \emptyset$, dokážeme, že $aH = bH$. Uvažujme $c \in aH \cap bH$ a napišme $c = ah_1 = bh_2$ pro nějaká $h_1, h_2 \in H$. Pak pro každé $ah \in aH$ platí

$$ah = ch_1^{-1}h = b\underbrace{h_2 h_1^{-1} h}_{\in H} \in bH$$

a podobně pro každé $bh \in bH$ platí

$$bh = ch_2^{-1}h = a\underbrace{h_1 h_2^{-1} h}_{\in H} \in aH.$$

Tedy $aH = bH$. □

Lemma 11.8 (velikost rozkladových tříd)**.** *Buď* $\mathbf{G}$ *grupa a* $\mathbf{H}$ *její podgrupa. Pro každé* $a \in G$ *platí* $|aH| = |H|$*.*

Důkaz. Uvažujme zobrazení $f : G \to G$ definované $f(x) = ax$. Toto zobrazení je prosté: kdyby $ax = f(x) = f(y) = ay$, krácením dostaneme $x = y$. Přitom $f(H) = aH$, tedy $f|_H$ je bijekce mezi H a aH, takže jsou tyto množiny stejně velké. □

Lagrangeovu větu lze formulovat i pro nekonečné grupy, s použitím kardinálních čísel pro označení velikostí množin. Čtenáři, který kardinální čísla neviděl, postačí k porozumění tvrzení vlastnost, že součin velikostí množin je roven velikosti kartézského součinu, tj. $|X| \cdot |Y| = |X \times Y|$. Důkaz věty je pro konečné i nekonečné množiny stejný.

Věta 11.9 (Lagrangeova věta)**.** *Buď* $\mathbf{G}$ *grupa a* $\mathbf{H}$ *její podgrupa. Pak*

$$|\mathbf{G}| = |\mathbf{H}| \cdot [\mathbf{G} : \mathbf{H}].$$

Důkaz. Zvolme nějakou transverzálu T a napišme

$$G = \bigcup_{a \in T} aH.$$

Podle Lemmatu 11.7 jde o disjunktní sjednocení, takže počet prvků lze spočítat jako součet velikostí jednotlivých podmnožin:

$$|\mathbf{G}| = \sum_{a \in T} |aH| = \sum_{a \in T} |H| = |T| \cdot |H| = [\mathbf{G} : \mathbf{H}] \cdot |\mathbf{H}|.$$

V druhé rovnosti jsme použili Lemma 11.8 a ve čtvrté rovnosti jsme použili vztah $|T| = [\mathbf{G} : \mathbf{H}]$, který plyne z Lemmatu 11.7. □

Příklad. Speciálním případem Lagrangeovy věty je *Eulerova věta* (Věta 2.4), jejíž elementární důkaz jsme předvedli v sekci 2.2. Zvolme $\mathbf{G} = \mathbb{Z}_n^*$ a $a \in \mathbb{Z}_n^*$, tedy a je celé číslo nesoudělné s n. Pak $\mathrm{ord}(a)$ dělí $|\mathbb{Z}_n^*| = \varphi(n)$, čili $\varphi(n) = k \cdot \mathrm{ord}(a)$ pro nějaké k. V grupě $\mathbb{Z}_n^*$ tedy platí

$$a^{\varphi(n)} = (a^{\mathrm{ord}(a)})^k = 1^k = 1,$$

čili v jazyce teorie čísel $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Na závěr sekce ukážeme kritérium, podle kterého se snadno pozná, zda jsou dvě rozkladové třídy stejné.

Tvrzení 11.10 (rovnost rozkladových tříd)**.** *Buď* $\mathbf{G}$ *grupa a* $\mathbf{H}$ *její podgrupa. Pro každé* $a, b \in G$ *platí*

(1) $aH = bH$ *právě tehdy, když* $a^{-1}b \in H$*;*
(2) $Ha = Hb$ *právě tehdy, když* $ab^{-1} \in H$*.*

Důkaz. (1) ($\Rightarrow$) Protože $aH = bH$, máme $b \in aH$, a tedy $b = ah$ pro nějaké $h \in H$. Tudíž $a^{-1}b = h \in H$. ($\Leftarrow$) Jestliže $a^{-1}b \in H$, pak pro každé $ah \in aH$ platí

$$ah = bb^{-1}ah = b\underbrace{(a^{-1}b)^{-1}h}_{\in H} \in bH$$

a podobně pro každé $bh \in bH$ platí

$$bh = a\underbrace{(a^{-1}b)h}_{\in H} \in aH.$$

Tedy $aH = bH$. (2) se dokáže analogicky. □

Cvičení 11.11. *Pomocí Tvrzení 11.10 dokažte, že mezi levými a pravými rozkladovými třídami existuje bijekce, daná předpisem* $aH \mapsto Ha^{-1}$*.*

12. Působení grupy na množině

12.1. Abstraktní grupa jako grupa permutací.

V mnoha situacích se hodí interpretovat danou abstraktní grupu jako grupu permutací na jisté množině. Například číselnou grupu $\mathbb{Z}_n$ lze interpretovat jako grupu permutací roviny, kde číslu k odpovídá rotace o úhel $k \cdot 2\pi/n$ podle daného středu. Součet dvou čísel modulo n odpovídá složení příslušných rotací, opačné číslo odpovídá opačné rotaci a nula identické permutaci. Toto pozorování motivuje následující definici.

Definice. *Působením grupy* $\mathbf{G}$ *na množině* X rozumíme libovolné zobrazení $\pi : G \to S_X$ splňující

$$\pi(gh) = \pi(g) \circ \pi(h),\ \pi(g^{-1}) = \pi(g)^{-1} \text{ a } \pi(1) = id$$

pro každé $g, h \in G$. Hodnotu permutace $\pi(g)$ na prvku $x \in X$ budeme značit krátce $g(x)$.

Z definice plyne, že jednotka v $\mathbf{G}$ působí jako identita, g^{-1} působí jako inverzní permutace k $\pi(g)$ a platí vztah $(g \cdot h)(x) = g(h(x))$. Můžeme si představovat, že prvky grupy $\mathbf{G}$ „hýbou“ s prvky množiny X, přičemž jak se prvky v $\mathbf{G}$ násobí, tak se příslušné „pohyby“ skládají.

Příklad. Triviálním případem je přirozené působení permutační grupy $\mathbf{G} \leq \mathbf{S}_X$ na množinu X, kde $\pi(g) = g$.

Příklad. Působení z úvodního odstavce odpovídá následující konfiguraci: $\mathbf{G} = \mathbb{Z}_n$, $X = \mathbb{R}^2$ a $\pi(k)$ je permutace na X daná předpisem

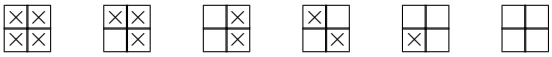
$$\begin{pmatrix} a \\ b \end{pmatrix} \mapsto \begin{pmatrix} \cos(k \cdot 2\pi/n) & -\sin(k \cdot 2\pi/n) \\ \sin(k \cdot 2\pi/n) & \cos(k \cdot 2\pi/n) \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix}.$$

Příklad. Maticové grupy lze interpretovat jako permutace příslušného vektorového prostoru dané příslušným lineárním zobrazením: zde $\mathbf{G} \leq \mathbf{GL}_n(\mathbf{T})$, $X = T^n$ a $\pi(A)$ je permutace množiny T^n, která vektor v zobrazí na součin Av.

Jako motivaci, proč uvažovat abstraktní koncept působení grupy na množině, si ukážeme jednu pěknou aplikaci v kombinatorice. Jak spočítat jistý typ objektů až na dané symetrie? Například, kolika způsoby je možné obarvit stěny krychle dvěma barvami až na otočení, tj. když dvě obarvení považujeme za totožná, pokud jedno z druhého dostaneme rotací krychle? Pro jednoduchost budeme metodu ilustrovat v dvojrozměrném případě.

Úloha. Kolika způsoby je možné obarvit políčka čtvercové tabulky 2×2 dvěma barvami až na otočení? Tj. dvě obarvení považujeme za totožná, pokud jedno z druhého dostaneme otočením tabulky.

Tuto úlohu je snadné řešit prostým výčtem všech možných obarvení, vyjde nám následujících šest:



Ale při větším počtu barev nebo větším počtu políček bychom se nedopočítali. Nejprve si ujasněme, co přesně znamená počítání „až na danou symetrii“. Dva objekty považujeme za totožné, pokud jeden z druhého dostaneme pomocí nějakého povoleného zobrazení. V naší úloze jsou to rotace, které zachovávají daný čtverec, tj. rotace roviny o 0, 90, 180 a 270 stupňů kolem středu čtverce. Uvažujme tedy působení grupy $\mathbf{G}$ sestávající z těchto čtyřech rotací na množině X sestávající ze včech možných obarvení čtverce 2×2 dvěma barvami (čili $|X| = 2^4 = 16$), kde $\pi(g)$ je permutace, která otočí tabulku o příslušný úhel i s daným obarvením.

Nyní zpět k teorii. V celém zbytku sekce budeme uvažovat libovolné působení grupy $\mathbf{G}$ na množinu X. Budeme potřebovat několik užitečných definic a vlastností.

Definice. Zavedeme tzv. *relaci tranzitivity* $\sim$ na množině X: definujeme $x \sim y$, pokud existuje $g \in G$ takové, že $g(x) = y$.

Volně řečeno, $x \sim y$, pokud nějaká permutace přesouvá prvek x na prvek y.

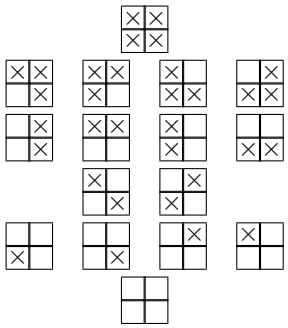
Lemma 12.1. *Relace $\sim$ je ekvivalence na množině X.*

Důkaz. Reflexivita: jednotka působí jako identita, tj. $1(x) = id(x) = x$. Symetrie: inverz prvku působí jako inverzní permutace, tj. $g(x) = y$ implikuje $g^{-1}(y) = x$. Tranzitivita: násobení odpovídá skládání permutací, čili pokud $x \sim y \sim z$, tj. $g(x) = y$ a $h(y) = z$ pro nějaká $g, h \in G$, pak $(h \cdot g)(x) = h(g(x)) = h(y) = z$, a tedy $x \sim z$. □

Bloky ekvivalence $\sim$ nazýváme *orbity*. Orbitu obsahující prvek x budeme značit

$$[x] = \{y \in X : \ x \sim y\} = \{g(x) : \ g \in G\}.$$

Příklad. V motivační úloze jsou v relaci $\sim$ právě ty dvojice obravení, kde lze jedno z druhého dostat otočením. Množina všech obarvení se tedy rozpadne na šest orbit následujícím způsobem:



Řešením motivační úlohy je *počet orbit* v tomto působení.

Definice. Bod $x \in X$ se nazývá *pevným bodem* prvku $g \in G$, pokud $g(x) = x$. Množinu všech pevných bodů prvku $g \in G$ budeme značit

$$X_g = \{x \in X : \ g(x) = x\}$$

a *stabilizátorem prvku* $x \in X$ nazveme množinu

$$G_x = \{g \in G : \ g(x) = x\}.$$

Příklad. Stabilizátorem obou jednobarevných obarvení je celá grupa $\mathbf{G}$. Stabilizátor obarvení  obsahuje pouze identitu. Stabilizátor obarvení obsahuje identitu a rotaci o 180 stupňů.

Lemma 12.2. *Stabilizátor G_x tvoří podgrupu grupy $\mathbf{G}$.*

Důkaz. Jednotka náleží G_x, neboť $1(x) = id(x) = x$. Pokud $g(x) = x$, pak také $g^{-1}(x) = x$. A pokud $g, h \in G_x$, tj. platí $g(x) = h(x) = x$, pak $(gh)(x) = g(h(x)) = g(x) = x$, čili $gh \in G_x$. □

Zásadní význam má následující tvrzení, které dává do souvislosti velikost stabilizátoru a velikost orbity.

Tvrzení 12.3 (velikost orbity vs. index stabilizátoru)**.** *Nechť grupa $\mathbf{G}$ působí na množině X. Pak pro každé $x \in X$ platí*

$$\big|[x]\big| = [\mathbf{G} : \mathbf{G}_x].$$

Důkaz. Index $[\mathbf{G} : \mathbf{G}_x]$ značí počet rozkladových tříd podgrupy $\mathbf{G}_x$, stačí tedy najít bijekci mezi prvky orbity a množinou rozkladových tříd. Uvažujme zobrazení

$$\varphi : \{gG_x : \ g \in G\} \to [x], \qquad gG_x \mapsto g(x).$$

Dokážeme, že to je bijekce. Předně je třeba ověřit, že jsme dobře definovali zobrazení: mohlo by se stát, že tutéž rozkladovou třídu máme označenu dvěma různými způsoby, tj. že $gG_x = hG_x$, a přitom se jí snažíme přiřadit různé hodnoty $g(x) \neq h(x)$. Ovšem podle Tvrzení 11.10 platí

$$gG_x = hG_x \;\Leftrightarrow\; h^{-1}g \in G_x \;\Leftrightarrow\; h^{-1}g(x) = x \;\Leftrightarrow\; g(x) = h(x),$$

a tedy φ je nejen dobře definované, ale také prosté. Navíc pro každý prvek $y \in [x]$ existuje $g \in G$ splňující $g(x) = y$, tedy φ je bijekce. □

Je-li grupa $\mathbf{G}$ konečná, kombinací Tvrzení 12.3 a Lagrangeovy věty dostáváme vztah

$$|\mathbf{G}| = |\mathbf{G}_x| \cdot |[x]|.$$

Speciálně, velikost každé orbity dělí řád grupy $\mathbf{G}$ (všimněte si, že to je splněno v naší motivační úloze).

12.2. **Burnsideova věta a počítání orbit.**

Označme $X/\!\sim$ množinu všech bloků ekvivalence $\sim$ na množině X. V našem kontextu bude $|X/\!\sim|$ značit počet orbit daného působení.

Věta 12.4 (Burnsideova věta)**.** *Nechť konečná grupa* $\mathbf{G}$ *působí na konečnou množinu* X*. Pak*

$$\big|X/\!\sim\big| = \frac{1}{|G|} \cdot \sum_{g \in G} \big|X_g\big|.$$

Vzorec lze interpretovat tak, že *počet orbit je roven průměrnému počtu pevných bodů*, kde průměr počítáme přes všechny prvky grupy $\mathbf{G}$.

Důkaz. Označme

$$M = \big\{(g, x) \in G \times X :\; g(x) = x\big\}.$$

Prvky této množiny můžeme spočítat dvěma způsoby: buď ke každému g spočítáme počet x takových, že $(g, x) \in M$, nebo naopak, ke každému x spočítáme počet g takových, že $(g, x) \in M$. Dostaneme tak následující rovnost:

$$|M| = \sum_{g \in G} |X_g| = \sum_{x \in X} |G_x|.$$

Použitím této rovnosti dopočítáme uvedený vzorec:

$$\frac{1}{|G|} \cdot \sum_{g \in G} |X_g| \;=\; \frac{1}{|G|} \cdot \sum_{x \in X} |G_x| \;\overset{12.3}{=}\; \frac{1}{|G|} \cdot \sum_{x \in X} \frac{|G|}{\big|[x]\big|} \;=\; \sum_{x \in X} \frac{1}{\big|[x]\big|} \;=$$

$$\sum_{O \in (X/\sim)} \sum_{x \in O} \frac{1}{\big|[x]\big|} \;=\; \sum_{O \in (X/\sim)} \sum_{x \in O} \frac{1}{|O|} \;=\; \sum_{O \in (X/\sim)} |O| \cdot \frac{1}{|O|} \;=\; \sum_{O \in (X/\sim)} 1.$$

Výsledek je tedy roven velikosti množiny $X/\!\sim$. □

Příklad. Vraťme se k motivační úloze. Rotace o 0 stupňů (tj. identita) zachovává všechna obarvení, tedy $|X_0| = |X| = 16$. Rotace o 90 stupňů zobrazuje levé dolní políčko na levé horní, levé horní na pravé horní, atd., čili abychom dostali stejné obarvení, musí mít všechna čtyři políčka stejnou barvu. Tedy $|X_{90}| = 2$. Podobně $|X_{270}| = 2$. Rotace o 180 stupňů zaměňuje levé dolní políčko za pravé horní a levé horní za pravé dolní. Tyto dvě dvojice tedy musí být stejnobarevné a to lze provést čtyřmi způsoby. Tedy $|X_{180}| = 4$. Podle Burnsideovy věty je počet obarvení až na otočení $\frac{1}{4} \cdot (16 + 2 + 4 + 2) = 6$.

Metodu ilustrujeme na několika dalších úlohách.

Úloha. (a) Dětská stavebnice obsahuje tři červené, tři zelené a tři modré čtvercové destičky. Kolika způsoby je lze sestavit do velkého čtverce 3×3? Dvě sestavy považujeme za totožné, pokud jednu z druhé dostaneme otočením. (b) Jak se výsledek změní, pokud je možné dílky pevně spojovat? Tedy pokud dvě sestavy považujeme za totožné, dostaneme-li jednu z druhé otočením a převrácením.

Řešení. Místo sestav budeme uvažovat barvení jednotlivých políček čtverce. Čili X bude množina všech obarvení čtverce 3×3 daným počtem barev a $\mathbf{G}$ bude (a) grupa $\mathbb{Z}_4$ interpretovaná jako rotace čtverce, (b) grupa $\mathbf{D}_8$ všech symetrií čtverce. Grupa $\mathbf{G}$ působí na X tak, že příslušná permutace otočí/převrátí čtverec i s jeho obarvením. Řešením úlohy je počet orbit tohoto působení (dvě obarvení jsou v jedné orbitě právě tehdy, když jedno z druhého dostaneme otočením, resp. převrácením).

Napíšeme tabulku, v jejímž prvním sloupci bude seznam prvků grupy $\mathbf{G}$, přičemž zobrazení „podobného typu“ budeme sdružovat (rozumí se, že prvky „podobného typu“ mají stejně velké množiny pevných bodů), v druhém sloupci bude počet prvků daného typu a ve třetím počet pevných bodů těchto prvků. Pevným bodem se rozumí takové obarvení, které po daném otočení/převrácení vypadá stejně.

g	#	$\lvert X_g\rvert$
id	1	1680
rotace o $\pm 90°$	2	0
rotace o $180°$	1	0
osa přes vrcholy	2	36
osa středem hran	2	36

Podle Burnsideovy věty je počet obarvení

(a) $\frac{1}{4} \cdot (1680 + 2 \cdot 0 + 1 \cdot 0) = 420$,
(b) $\frac{1}{8} \cdot (1680 + 2 \cdot 0 + 1 \cdot 0 + 2 \cdot 36 + 2 \cdot 36) = 228$.

□

Úloha. Kolik náhrdelníků lze sestavit (a:Sparta) ze tří červených, tří žlutých a tří modrých kuliček, (b:Bohemians) z šesti zelených a tří bílých kuliček? (Nezáleží na poloze náhrdelníku, je možno jej převracet či otáčet.)

Řešení. Náhrdelník reprezentujeme jako obarvení vrcholů pravidelného devítiúhelníka. Čili X, resp. Y, bude množina všech obarvení vrcholů pravidelného devítiúhelníka danými barvami a $\mathbf{G} = \mathbf{D}_{18}$ bude grupa všech symetrií pravidelného devítiúhelníka, která působí na X, resp. Y, tak, že příslušná permutace otočí/převrátí devítiúhelník i s jeho obarvením. Každé orbitě tohoto působení odpovídá právě jeden náhrdelník (jehož kuličky jsou uspořádány podle toho obarvení). Napíšeme tabulku podobně jako v předchozí úloze.

g	#	$\lvert X_g\rvert$	$\lvert Y_g\rvert$
id	1	1680	84
rotace o ± 1 vrchol	2	0	0
rotace o ± 2 vrcholy	2	0	0
rotace o ± 3 vrcholy	2	6	3
rotace o ± 4 vrcholy	2	0	0
reflexe	9	0	4

Podle Burnsideovy věty je počet náhrdelníků (a) $\frac{1}{18} \cdot (1680 + 2 \cdot 6) = 94$, resp. (b) $\frac{1}{18} \cdot (84 + 2 \cdot 3 + 9 \cdot 4) = 7$. □

Úloha. Kolika způsoby je možné obarvit stěny krychle dvěma barvami? Kolika způsoby lze přiřadit stěnám čísla 1 až 6? A kolik existuje hracích kostek, tj. kolika způsoby lze přiřadit čísla 1 až 6 tak, že součet protilehlých stěn je sedm? Dvě obarvení/přiřazení považujeme za totožná, pokud lze jedno z druhého dostat otočením krychle.

Řešení. Buď X množina všech obarvení stěn krychle dvěma barvami, Y množina všech přiřazení čísel 1 až 6 stěnám a Z množina těch přiřazení z Y, jejichž protilehlé stěny dávají součet sedm. $\mathbf{G}$ bude grupa všech rotací krychle působící na X, Y i Z tak, že příslušná permutace otočí krychli i s jejím obarvením/přiřazením. Napíšeme tabulku podobně jako v předchozí úloze.

g	#	$\lvert X_g\rvert$	$\lvert Y_g\rvert$	$\lvert Z_g\rvert$
identita	1	2^6	6!	48
osa přes středy protilehlých stěn, $\pm 90°$	6	2^3	0	0
osa přes středy protilehlých stěn, $+180°$	3	2^4	0	0
osa přes středy protilehlých hran, $+180°$	6	2^3	0	0
osa přes protilehlé vrcholy, $\pm 120°$	8	2^2	0	0

Tedy počty orbit jsou

- $\lvert X/\!\sim\rvert = \frac{1}{24} \cdot (2^6 + 3 \cdot 2^4 + 12 \cdot 2^3 + 8 \cdot 2^2) = 10$,
- $\lvert Y/\!\sim\rvert = \frac{1}{24} \cdot 6! = 30$,
- $\lvert Z/\!\sim\rvert = \frac{1}{24} \cdot 48 = 2$.

Jak známo, hrací kostky jsou dvě, pravotočivá a levotočivá, podle pořadí stěn 1, 2, 3 při pohledu na roh kostky. který tato čísla sdílí. □

Burnsideovu větu lze použít v řadě dalších aplikací, např. pokud chceme zjistit *počet nějakých struktur dané velikosti až na izomorfismus*. Metodu ilustrujeme na grafech se čtyřmi vrcholy.

Příklad. Buď X množina všech grafů s vrcholy $1, 2, 3, 4$. Dva grafy jsou izomorfní, pokud existuje permutace z $\mathbf{S}_4$, která převádí hrany na hrany a mezery na mezery. Uvažujme tedy působení grupy $\mathbf{S}_4$ na X tak, že daná permutace přehází vrcholy i s hranami, které do nich vedou. Orbity tohoto působení budou obsahovat právě všechny navzájem izomorfní grafy, počet neizomorfních grafů je tedy roven počtu orbit. Řešením je tabulka

g	#	$\lvert X_g\rvert$
id	1	2^6
(..)	6	2^4
(..)(..)	3	2^4
(...)	8	2^2
(....)	6	2^2

Vidíme, že čtyřprvkových grafů je 11.

Na závěr ukážeme jednu zajímavost s elegantním důkazem. Permutační grupa se nazývá *tranzitivní*, má-li jen jednu orbitu (ve svém přirozeném působení). Např. grupy $\mathbf{S}_n$, $\mathbf{A}_n$, $\mathbf{D}_{2n}$ jsou tranzitivní, grupa $\langle (1\ 2)(3\ 4) \rangle_{\mathbf{S}_4}$ není.

Věta 12.5 (Jordanova věta)**.** *Každá alespoň dvouprvková konečná tranzitivní grupa obsahuje alespoň jednu permutaci bez pevného bodu.*

Důkaz. Podle Burnsideovy věty je počet orbit roven průměrnému počtu pevných bodů. Tranzitivita říká, že počet orbit je 1. Přitom identita má alespoň dva pevné body, tedy *nadprůměrné* množství, takže musí existovat permutace, která má *podprůměrné* množství pevných bodů. Protože je počet pevných bodů nezáporné celé číslo, jediná podprůměrná hodnota je 0. Tedy existuje permutace bez pevného bodu. □

13. Cyklické grupy

13.1. Podgrupy, generátory, řády prvků.

Grupa $\mathbf{G}$ se nazývá *cyklická*, pokud je generovaná jedním prvkem, tj.

$$\mathbf{G} = \langle a \rangle_{\mathbf{G}}$$

pro nějaké $a \in G$. Její prvky lze díky Důsledku 11.3 vyjádřit jako mocniny generátoru,

$$G = \{a^k : \ k \in \mathbb{Z}\},$$

z čehož je vidět, že je to nutně grupa abelovská. Z Tvrzení 11.6 plyne, že je-li řád a nekonečný, pak jsou tyto mocniny po dvou různé, a je-li $\mathrm{ord}(a) = n$ konečný, pak $G = \{a^0, a^1, \ldots, a^{n-1}\}$. Odsud pochází název pro cyklické grupy: při násobení daným prvkem a cyklicky procházíme přes všechny prvky grupy $\mathbf{G}$.

Příklady.

- Grupy $\mathbb{Z}$ a $\mathbb{Z}_n$, $n \in \mathbb{N}$, jsou cyklické, generované prvkem 1.
- Grupy $\mathbb{C}_n \leq \mathbb{C}^*$ sestávající ze všech komplexních kořenů polynomu $x^n - 1$ jsou cyklické, $\mathbb{C}_n = \langle e^{2\pi i/n} \rangle$.
- V této sekci si dokážeme, že grupy $\mathbb{Z}_p^*$ jsou cyklické pro každé prvočíslo p (Věta 13.7). Například $\mathbb{Z}_5^* = \langle 2 \rangle$, $\mathbb{Z}_7^* = \langle 3 \rangle$, $\mathbb{Z}_{11}^* = \langle 2 \rangle$.
- Některé grupy $\mathbb{Z}_n^*$, n složené, jsou cyklické, např. $\mathbb{Z}_6^* = \{1, 5\} = \langle 5 \rangle$, ale některé ne, např. grupa $\mathbb{Z}_8^*$ cyklická není.
- Každá grupa $\mathbf{G}$ prvočíselného řádu je cyklická. Uvažujme podgrupu $\langle a \rangle$, $a \neq 1$. Podle Lagrangeovy věty je $|\langle a \rangle|$ dělí $|\mathbf{G}|$, přitom $|\langle a \rangle| > 1$, tedy $|\langle a \rangle| = |\mathbf{G}|$ a prvek a tuto grupu generuje.

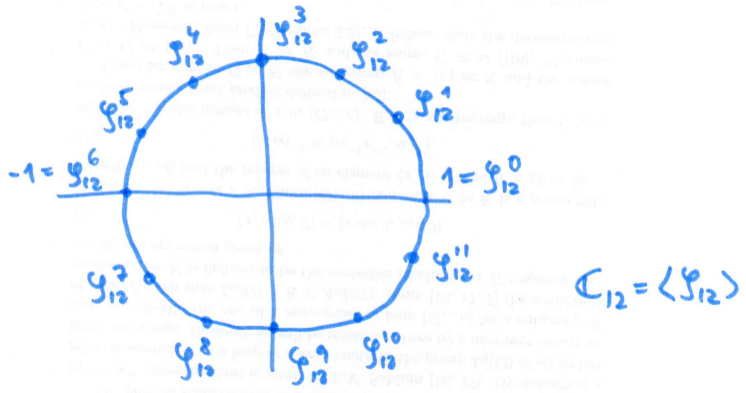
Nejprve se podíváme, jak vypadají podgrupy cyklických grup.

Tvrzení 13.1. *Každá podgrupa cyklické grupy je cyklická.*

Důkaz. Buď $\mathbf{H}$ podgrupa cyklické grupy $\mathbf{G} = \langle a \rangle$. Je-li $H = \{1\}$, pak $\mathbf{H} = \langle 1 \rangle$. V opačném případě označme k nejmenší kladné číslo takové, že $a^k \in H$ (takové jistě existuje: je-li $1 \neq b \in H$, pak $b = a^l$ pro nějaké l a buď b nebo b^{-1} má exponent kladný). Dokážeme, že $\mathbf{H} = \langle a^k \rangle$. Inkluze $\langle a^k \rangle \subseteq H$ je zřejmá. Pro spor tedy předpokládejme, že existuje nějaký prvek $a^n \in H \setminus \langle a^k \rangle$. Nutně $k \nmid n$, jinak bychom měli $a^n = (a^k)^{n/k} \in \langle a^k \rangle$. Napišme $n = kq + r$, kde $0 < r < k$. Pak

$$a^r = a^{n-kq} = a^n \cdot (a^k)^{-q} \in H,$$

protože a^n i a^k leží v H, což je spor s volbou k jako nejmenšího kladného čísla s vlastností $a^k \in H$. □



Obrázek 8. Ilustrace faktu, proč se cyklické grupy nazývají cyklické.

Příklad. Grupa $\mathbb{Z}$ je cyklická, čili její podgrupy jsou cyklické, tedy tvaru

$$\mathbf{H} = \langle k \rangle = k\mathbb{Z} = \{a \in \mathbb{Z} : \ k \mid a\}.$$

Přitom $k\mathbb{Z} = l\mathbb{Z}$ právě tehdy, když $k = \pm l$. Podgrupy jsou tedy ve vzájemně jednoznačné korepondenci s nezápornými čísly a $k\mathbb{Z} \subseteq l\mathbb{Z}$ právě tehdy, když $l \mid k$. Čili podgrupy jsou uspořádány vzhledem k inkluzi opačně než množina $\mathbb{N} \cup \{0\}$ dělitelností.

Pro konečné cyklické grupy je situace složitější, mnoho různých prvků může generovat stejné podgrupy.

Lemma 13.2 (podgrupy cyklických grup)**.** *Buď $\mathbf{G} = \langle a \rangle$ cyklická grupa. Pak*

(1) $\langle a^k, a^l \rangle = \langle a^{\mathrm{NSD}(k,l)} \rangle$,
(2) *je-li $|\mathbf{G}| = n$, pak $\langle a^k \rangle = \langle a^{\mathrm{NSD}(k,n)} \rangle$.*

Důkaz. (1) Protože $\mathrm{NSD}(k,l)$ dělí k i l, platí $a^k, a^l \in \langle a^{\mathrm{NSD}(k,l)} \rangle$, čímž máme prokázánu inkluzi $\subseteq$. Naopak, podle Bézoutovy rovnosti je $\mathrm{NSD}(k,l) = uk + vl$ pro nějaká $u, v \in \mathbb{Z}$, a tedy

$$a^{\mathrm{NSD}(k,l)} = a^{uk+vl} = (a^k)^u \cdot (a^l)^v \in \langle a^k, a^l \rangle,$$

čímž máme prokázánu inkluzi $\supseteq$.

(2) Dosadíme $l = n$: pak $\langle a^{\mathrm{NSD}(k,n)} \rangle = \langle a^k, a^n \rangle = \langle a^k \rangle$, protože $a^n = 1$. □

Tvrzení 13.3 (generátory cyklických grup)**.** *Buď $\mathbf{G} = \langle a \rangle$ cyklická grupa.*

(1) *Pokud je $\mathbf{G}$ nekonečná, generátorem jsou pouze prvky a, a^{-1}.*
(2) *Pokud je $\mathbf{G}$ konečná řádu n, generátorem jsou právě prvky a^k, kde $k \in \{1, \ldots, n-1\}$ je nesoudělné s n.*

Důkaz. (1) Oba prvky a, a^{-1} grupu $\mathbf{G}$ generují, protože $\{a^k : \ k \in \mathbb{Z}\} = \{a^{-k} : k \in \mathbb{Z}\}$. Žádný jiný generátor grupa $\mathbf{G}$ nemá: kdyby $\mathbf{G} = \langle a^n \rangle$ pro nějaké n, pak by existovalo $m \in \mathbb{Z}$ takové, že $a = (a^n)^m$, a dostali bychom $1 = (a^n)^m \cdot a^{-1} = a^{mn-1}$; řád a je ovšem nekonečný, a tedy $mn = 1$, čili $n = \pm 1$.

(2) Podle Lemmatu 13.2 je $\langle a^k\rangle = \langle a^{\mathrm{NSD}(k,n)}\rangle$. Pokud $\mathrm{NSD}(k,n) = 1$, pak $\langle a^k\rangle = \langle a\rangle = \mathbf{G}$. Pokud $\mathrm{NSD}(k,n) = d \neq 1$, pak $\langle a^k\rangle = \langle a^d\rangle = \{a^d, a^{2d}, \ldots, a^{\frac{n}{d}d}\}$ je vlastní podgrupa. □

Příklad (podgrupy grupy $\mathbb{Z}_n$)**.** Grupa $\mathbb{Z}_n$ je cyklická, čili její podgrupy jsou cyklické, tedy tvaru

$$\langle k\rangle = k\mathbb{Z}_n = \{ku \bmod n :\ u = 0, \ldots, n-1\},$$

pro nějaké $k \in \{0, \ldots, n-1\}$. Z Lemmatu 13.2(2) s volbou $a = 1$ plyne, že $k\mathbb{Z}_n = \mathrm{NSD}(k,n)\mathbb{Z}_n$, tedy $k\mathbb{Z}_n = l\mathbb{Z}_n$ právě tehdy, když $\mathrm{NSD}(k,n) = \mathrm{NSD}(l,n)$. Podgrupy jsou tedy ve vzájemně jednoznačné korepondenci s děliteli čísla n. Pro $k, l \mid n$ pak platí, že $k\mathbb{Z}_n \subseteq l\mathbb{Z}_n$ právě tehdy, když $l \mid k$. Čili podgrupy jsou uspořádány vzhledem k inkluzi opačně než množina všech dělitelů čísla n dělitelností. Podle Tvrzení 13.3 je $\mathbb{Z}_n = \langle k\rangle$ právě tehdy, když jsou k, n nesoudělná.

Příklad (podgrupy grupy $\mathbb{Z}_p^*$)**.** Grupa $\mathbb{Z}_{11}^* = \langle 2\rangle$ je cyklická řádu 10, čili její podgrupy jsou cyklické, tedy tvaru

$$\langle 2^k\rangle = \{2^{uk} \bmod 11 :\ u = 0, \ldots, 10\},$$

pro nějaké $k \in \{0, \ldots, 9\}$. Z Lemmatu 13.2(2) plyne, že $\langle 2^k\rangle = \langle 2^l\rangle$ právě tehdy, když $\mathrm{NSD}(k,n) = \mathrm{NSD}(l,n)$. Podgrupy jsou tedy ve vzájemně jednoznačné korepondenci s děliteli čísla 10, máme tedy čtyři podgrupy,

$$\langle 2^1\rangle = \mathbb{Z}_{11}^*,\ \langle 2^2\rangle = \{1,4,5,9,3\},\ \langle 2^5\rangle = \{1,10\},\ \langle 2^{10}\rangle = \{1\}.$$

Generátory jsou prvky 2^k takové, že k je nesoudělné s 10, tedy prvky $2^1 = 2$, $2^3 = 8$, $2^6 = 7$ a $2^9 = 6$. Všimněte si, že to jsou právě čísla, která nepatří do žádné z vlastních podgrup vypsaných výše.

Úloha se přímočaře zobecní na libovolnou grupu $\mathbb{Z}_p^*$, p prvočíslo, o které si ukážeme, je vždy cyklická, byť není zřejmé, který prvek a je generátorem. Podgrupy pak budou právě $\langle a^k\rangle$, kde $k \mid p-1$, generátory budou právě prvky a^k, kde $\mathrm{NSD}(k, p-1) = 1$.

Z Tvrzení 13.3 plyne, že cyklická grupa řádu n má právě $\varphi(n)$ generátorů, kde φ značí Eulerovu funkci. Tohoto faktu využijeme k řešení obecnější úlohy: spočítáme počet prvků každého řádu. V nekonečných cyklických grupách mají všechny prvky kromě jednotky řád nekonečný. V konečných grupách dává Lagrangeova věta omezení na přípustné řády. Ukážeme si, že v cyklických grupách prvky všech přípustných řádů existují a jejich počet je dán Eulerovou funkcí.

Tvrzení 13.4 (řády prvků cyklických grup)**.** *Cyklická grupa konečného řádu n obsahuje právě $\varphi(d)$ prvků řádu d pro každé $d \mid n$.*

Důkaz. Buď $\mathbf{G}$ cyklická grupa konečného řádu n. Každý prvek řádu $d \mid n$ je generátorem nějaké cyklické podgrupy řádu d. Taková podgrupa však v $\mathbf{G}$ existuje pouze jedna: podle Lemmatu 13.2 jsou všechny podgrupy v $\mathbf{G}$ tvaru $\langle a^k\rangle$, $k \mid n$. Přitom $|\langle a^k\rangle| = \frac{n}{k}$, čili $\langle a^{\frac{n}{d}}\rangle$ je jediná podgrupa řádu d. Ta má podle Tvrzení 13.3 právě $\varphi(d)$ generátorů. □

Tvrzení o počtu prvků daného řádu lze použít k důkazu následující kombinatorické identity.

Tvrzení 13.5. *Pro každé $n \in \mathbb{N}$ platí $\sum_{d\mid n} \varphi(d) = n$.*

Důkaz. Budeme počítat počet prvků grupy $\mathbb{Z}_n$ dvěma způsoby. Jeden způsob je triviální: grupa obsahuje čísla $0, \ldots, n-1$, tedy $|\mathbb{Z}_n| = n$. Podruhé spočítáme prvky podle řádů: podle Lagrangeovy věty jsou přípustné řády $d \mid n$, tedy $|\mathbb{Z}_n| = \sum_{d|n} u_d$, kde u_d značí počet prvků řádu d. Tvrzení 13.4 říká, že $u_d = \varphi(d)$. □

13.2. **Multiplikativní grupy konečných těles jsou cyklické.**

Tvrzení uvedené v názvu podsekce má dalekosáhlé důsledky v teorii konečných těles. K jeho důkazu použijeme následující kritérium cykličnosti.

Lemma 13.6. *Buď* $\mathbf{G}$ *konečná grupa a předpokládejme, že pro každé* k *grupa* $\mathbf{G}$ *obsahuje nejvýše* k *prvků a splňujících* $a^k = 1$. *Pak je grupa* $\mathbf{G}$ *cyklická.*

Důkaz. Označme $n = |\mathbf{G}|$ a u_k počet prvků řádu k v grupě $\mathbf{G}$. Podle Lagrangeovy věty je $u_k = 0$ pro všechna $k \nmid n$, a tedy $n = \sum_{d|n} u_d$ (počítáme prvky $\mathbf{G}$ podle jejich řádu jako v Tvrzení 13.5).

Uvažujme nějaký prvek a řádu k v $\mathbf{G}$. Podgrupa $\langle a \rangle$ je cyklická řádu k a všechny prvky $b \in \langle a \rangle$ splňují $b^k = 1$. Podle předpokladu v $\mathbf{G}$ žádné jiné prvky s touto vlastností nejsou, takže $\langle a \rangle$ je jediná cyklická podgrupa řádu k v $\mathbf{G}$. Podle Tvrzení 13.3 má $\varphi(k)$ generátorů, a tedy $u_k = \varphi(k)$.

Čili pro každé $d \mid n$ platí $u_d = 0$ nebo $u_d = \varphi(d)$. Dokážeme, že vždy nastane druhá možnost: podle Tvrzení 13.5 je $\sum_{d|n} \varphi(d) = n = \sum_{d|n} u_d$, takže $u_d = \varphi(d)$ pro všechna $d \mid n$. Speciálně $u_n \neq 0$, a tedy v $\mathbf{G}$ existuje prvek řádu n, neboli generátor. □

Věta 13.7. *Buď* $\mathbf{T}$ *těleso a* $\mathbf{G}$ *konečná podgrupa grupy* $\mathbf{T}^*$. *Pak* $\mathbf{G}$ *je cyklická.*

Důkaz. Podle Věty 4.4 má polynom $x^k - 1$ nejvýše k kořenů v tělese $\mathbf{T}$. Tedy grupa $\mathbf{G} \leq \mathbf{T}^*$ může obsahovat nejvýše k prvků a splňujících $a^k = 1$ a můžeme aplikovat předchozí kritérium. □

Speciálně, multiplikativní grupy konečných těles jsou cyklické. Jejich generátorům se říká *primitivní prvky*. Pozor, prvek α v tělese $\mathbb{Z}_p[\alpha]/(m)$ být primitivní může, ale nemusí: pozitivním příkladem $\mathbb{F}_4 = \mathbb{Z}_2[\alpha]/(\alpha^2 + \alpha + 1)$, negativním příkladem je $\mathbb{F}_9 = \mathbb{Z}_3[\alpha]/(\alpha^2 + 1)$, kde α generuje pouze čtyřprvkovou podgrupu grupy $\mathbb{F}_9^*$. Primitivní prvky se používají například v algoritmu rychlé Fourierovy transformace, která umí vyhodnocovat a interpolovat polynomy v bodech $1, \alpha, \alpha^2, ..., \alpha^{n-1}$ v čase $O(n \log n)$ (zatímco pro n náhodně zvolených bodů bychom standardními metodami potřebovali kvadratický čas).

Není bez zajímavosti, že pro grupy $\mathbb{Z}_p^*$ lze znění Věty 13.7 interpretovat čistě v jazyce elementární teorie čísel: pro každé prvočíslo p existuje číslo a (generátor té grupy) takové, že každé $b \in \{1, \ldots, p-1\}$ lze vyjádřit právě jedním způsobem jako $b = a^k \bmod p$ pro nějaké $k \in \{0, \ldots, p-2\}$.

13.3. **Diskrétní logaritmus a kryptografie.**

Buď $\mathbf{G} = \langle a \rangle$ cyklická grupa řádu n. Podívejme se zobrazení

$$exp : \mathbb{Z}_n \to \mathbf{G}, \quad k \mapsto a^k.$$

Zobrazení se nazývá *diskrétní exponenciála* a z výše uvedených úvah plyne, že je bijektivní. Inverznímu zobrazení se říká *diskrétní logaritmus*. Jinými slovy, diskrétní logaritmus prvku $b \in G$ přiřadí to jediné $k \in \{0, \ldots, n-1\}$, pro které $b = a^k$; budeme jej značit $k = \log_a b$. (Pro nekonečné grupy se používá analogická terminologie, ale z výpočetního hlediska nejsou tak zajímavé.)

Počítat diskrétní exponenciálu je zpravidla snadné. Přesněji řečeno, kdykoliv lze v dané grupě efektivně násobit, pak lze také efektivně mocnit. Určitě ne tak, že bychom počítali a^k jako k součinů. Stačí jich méně než $2\lfloor \log_2 k \rfloor$: napíšeme si k ve dvojkové soutavě, $k = \sum_{i=0}^{\lfloor \log_2 k \rfloor} u_i 2^i$, kde $u_i \in \{0,1\}$, a vyjádříme mocninu jako

$$a^k = a^{\sum_{i=0}^{\lfloor \log_2 k \rfloor} u_i 2^i} = \prod_{i:\, u_i = 1} a^{2^i}.$$

Přitom prvků tvaru a^{2^i} se ve vzorci vyskytuje nejvýše $\lfloor \log_2 k \rfloor + 1$ a spočteme je postupným mocněním

$$a,\ a^2,\ (a^2)^2,\ \ldots,\ a^{2^i} = (a^{2^{i-1}})^2,\ \ldots,$$

čili pomocí $\lfloor \log_2 k \rfloor$ součinů. Vidíme, že spočítat libovolnou mocninu v n-prvkové grupě vyžaduje méně než $2 \log_2 n$ násobení.

Empirická zkušenost ukazuje, že pro spoustu grup je výpočet diskrétního logaritmu obtížný. Hledání logaritmu hrubou silou, procházením všech n možných exponentů, je exponenciálně pomalejší než výpočet mocniny. Pro některé grupy je výpočet snadný (viz příklad níže), ale například pro grupy $\mathbb{Z}_p^*$, p prvočíslo, nebo grupy odvozené z eliptických křivek nad konečnými tělesy, není v současnosti znám výrazně lepší postup než hrubá síla.

Příklad. Uvažujme cyklickou grupu $\mathbb{Z}_n = \langle a \rangle$. Logaritmus $\log_a b$ je roven tomu (jedinému) $k \in \{0, \ldots, n-1\}$, pro které

$$ka \equiv b \pmod{n}.$$

Takové k najdeme snadno Eukleidovým algoritmem: podle Tvrzení 13.3 je generátor nesoudělný s n, spočteme Bézoutovy koeficienty $1 = \mathrm{NSD}(a,n) = ua + vn$ a vidíme, že $b = uab + vnb \equiv uba \pmod{n}$, čili $\log_a b = ub \bmod n$.

Nadále uvažujme libovolnou cyklickou grupu $\mathbf{G}$, pro kterou je diskrétní exponenciála výpočetně zvladatelná, ale logaritmus nikoliv (používají se např. grupy $\mathbb{Z}_p^*$ pro prvočísla $p \geq 2^{1000}$). Ukážeme si dva kryptografické algoritmy založené na diskrétním logaritmu: *Diffie-Hellmanův protokol* pro výměnu klíče (jde o nejpoužívanější algoritmus svého druhu) a *El Gamalův algoritmus* pro kryptografii s veřejným klíčem (v praxi se používá, i když algoritmus RSA ze sekce 2.2, který řeší stejnou úlohu, je výrazně populárnější).

Diffie-Hellmanův protokol. Alice a Bob se potřebují dohodnout na nějakém společném hesle (odborně *klíči*), přičemž k dispozici mají pouze veřejný kanál (např. odposlouchávaný telefon). Jak to provést?

Nejprve se Alice a Bob dohodnou na nějaké cyklické grupě a generátoru, $\mathbf{G} = \langle a \rangle$. Dále si Alice zvolí číslo m a Bob číslo n z intervalu $2, \ldots, |G| - 1$, přičemž každý bude svoje číslo držet v tajnosti. Pak provedou následující úkony: Alice spočte $u = a^m$ a pošle u Bobovi, Bob spočte $v = a^n$ a pošle v Alici. Poté Alice spočte $v^m = (a^n)^m = a^{mn}$ a Bob spočte $u^n = (a^m)^n = a^{mn}$. Oba získali stejný prvek a^{mn} a ten prohlásí za společný klíč.

Kdyby nepřítel poslouchal jejich komunikaci, co zjistí? Bude znát grupu $\mathbf{G}$, generátor a a hodnoty $u = a^m$ a $v = a^n$. Chtěl by spočítat prvek a^{mn}. Této úloze se říká *Diffie-Hellmanův problém*. Očividným řešením je provést diskrétní logaritmus, získat čísla m, n, vynásobit je a dopočítat a^{mn}. Toto řešení však není výpočetně zvladatelné a žádný efektivní postup není v současné době znám.

El Gamalův protokol. Příjemce zvolí cyklickou grupu $\mathbf{G} = \langle a \rangle$, náhodné číslo k z intervalu $2, \ldots, |G| - 1$ a spočte $b = a^k$. *Veřejným klíčem* bude $\mathbf{G}, a, b$, jeho *soukromým klíčem* bude k.

Odesílatel zprávy zvolí náhodné číslo l z intervalu $2, \ldots, |G|-1$, které po odeslání zprávy zničí, a zprávu $x \in G$ zašifruje jako dvojici

$$y = (a^l, x \cdot b^l).$$

Příjemce obdrží dvojici $y = (u, v)$ a dešifruje ji pomocí k takto:

$$v \cdot u^{-k} = x \cdot b^l \cdot (a^l)^{-k} = x \cdot (a^k)^l \cdot (a^l)^{-k} = x.$$

Kdybychom uměli rychle počítat diskrétní logaritmus, okamžitě získáme soukromý klíč. Jsou známy i jiné způsoby útoku na El Gamalův algoritmus, díky nimž například grupy $\mathbb{Z}_p^*$ nejsou považovány za bezpečné. Obecný postup však znám není a El Gamal je považován za bezpečný například na dostatečně velkých grupách odvozených z eliptických křivek.

Jedním ze základních konceptů v kryptografii je pojem *jednosměrné funkce*. Velmi zjednodušeně řečeno, je to bijektivní zobrazení f takové, že hodnoty $f(x)$ se dají počítat rychle, ale není znám postup, kterým by bylo možné získat nějakou statisticky významnou informaci o hodnotách inverzního zobrazení $f^{-1}(y)$. Příkladem je

- diskrétní exponenciála v některých grupách, např. zobrazení $\mathbb{Z}_{p-1} \to \mathbb{Z}_p^*$, $k \mapsto a^k \bmod p$ pro dostatečně velká prvočísla p;
- zobrazení $\mathbb{Z}_N \to \mathbb{Z}_N$, $a \mapsto a^k \bmod N$ pro vhodná k, N, např. je-li N součinem dvou dostatečně velkých velkých prvočísel (na tomto příkladu je založena šifra RSA, viz sekce 2.2).

K čemu může být dobrá funkce, kterou lze zprávu důkladně zašifrovat, ale za žádných okolností dešifrovat? Alice a Bob si chtějí na dálku zahrát hru „*panna nebo orel*“. Alice bude házet mincí, Bob hádat. Jak to ale udělat, aby Alice Boba nepodvedla, když se Bob nemůže na minci podívat? Zvolme nějakou jednosměrnou funkci f na množině $\{1, \ldots, n\}$. Pokud Alice hodí orla, zvolí náhodné liché číslo x, v opačném případě zvolí sudé číslo. Bobovi pošle hodnotu $f(x)$. Protože je f jednosměrná, Bob neumí spočítat, co padlo, zvolí tedy odpověď náhodně a sdělí ji Alici. Nyní Alice zveřejní číslo x a Bob ihned vidí, zda vyhrál. Pro kontrolu si spočte $f(x)$ a porovná ho s hodnotou, kterou dostal na začátku. Pokud se hodnoty neshodují, Alice podvádí. Může Alice Boba podvést tak, aby na to nepřišel? Dejme tomu, že padl orel a to samé si tipnul Bob. Aby Alice Boba podvedla, musela by Bobovi ukázat sudé x' takové, že $f(x') = f(x)$. Jenže takové x' neexistuje, když je f bijekce.

Pro kryptografii jsou zvlášť cenné jednosměrné funkce, ke kterým existují tzv. *zadní vrátka*, dodatečná informace, pomocí které lze inverzní zobrazení počítat efektivně. Příkladem je zobrazení z šifry RSA: počítat odmocniny modulo N je obecně obtížné, ale známe-li prvočíselný rozklad čísla N, je to snadné.

V poslední době je populárním tématem tzv. *homomorfní šifrování*. Cílem je najít jednosměrné funkce (se zadními vrátky), které by byly homomorfismem vůči nějaké zajímavé operaci. Motivace vychází z praxe vzdáleného počítání (*cloud computing*): rádi bychom, aby pro nás někdo spočítal časově náročné úlohy na našich datech, ale zároveň bychom nechtěli tato data prozradit. Přeloženo do matematického jazyka, máme nějakou operaci $*$ na datech (typickým příkladem z praxe je

třeba součin velkých matic) a chceme jednosměrnou funkci $X \to X$ takovou, že když pošleme poskytovateli služeb hodnoty $f(x), f(y)$, on spočte výsledek $f(x) * f(y)$ a my jej dešifrujeme zobrazením f^{-1}, dostaneme správný výsledek $x * y$. To jest, chceme, aby platilo $f(x) * f(y) = f(x * y)$, jinými slovy, aby f byl homomorfismus vzhledem k operaci $*$ (která je často grupová, ale jsou i obecnější koncepty). O homomorfismech se dozvíte více v letním semestru.

Cvičení 13.8. *Vysvětlete nějaké geograficky vzdálené osobě (humanitně založené sestře, spolužákovi, který nemá rád algebru, rodičům, zvídavému dědečkovi, ...), jak hrát kámen-nůžky-papír v době karantény, a zahrajte si.*

Dodatek o permutacích

Základní vlastnosti permutací.

V tomto odstavci shrneme poznatky, které by měl typický čtenář mít ze základního kurzu lineární algebry nebo diskrétní matematiky, a doplníme je o pojem konjugace.

Permutací na množině X rozumíme bijekci (vzájemně jednoznačné zobrazení) $X \to X$. Pro permutace π, σ na X definujeme operace $\circ$, $^{-1}$, id předpisy

- $\pi \circ \sigma : x \mapsto \pi(\sigma(x))$,
- $\pi^{-1} : x \mapsto$ (ten jediný) prvek y splňující $\pi(y) = x$,
- $id : x \mapsto x$.

Označíme-li S_X množinu všech permutací na množině X, pak $\mathbf{S}_X = (S_X, \circ, ^{-1}, id)$ je tzv. *symetrická grupa* na X. Podgrupám této grupy se říká *permutační grupy*. Je-li $X = \{1, \dots, n\}$, značíme $\mathbf{S}_X = \mathbf{S}_n$.

Cyklus v permutaci π je posloupnost $a_1, \dots, a_k$ navzájem různých prvků množiny X splňující $\pi(a_1) = a_2$, $\pi(a_2) = a_3$, ..., $\pi(a_k) = a_1$. *Rozkladem na cykly* se rozumí zápis

$$(a_{11}\ a_{12}\ \dots\ a_{1k_1})(a_{21}\ a_{22}\ \dots\ a_{2k_2}) \cdots (a_{m1}\ a_{m2}\ \dots\ a_{mk_m}),$$

kde $a_{i1}, a_{i2}, \dots, a_{ik_i}$, $i = 1, \dots, m$, jsou pod dvou různé prvky. Cykly délky 1 se ze zápisu zpravidla vynechávají. (Je-li X konečná množina, rozklad na cykly jistě existuje; pro nekonečné množiny bychom museli povolit „nekonečné cykly“.)

Transpozicí rozumíme permutaci tvaru $(x\ y)$. Permutace na konečné množině se nazývá *sudá*, pokud se skládá ze sudého počtu transpozic, *lichá* v opačném případě (máme-li dva různé rozklady jedné permutace, mohou mít různé délky, ale, jak lze snadno nahlédnout, stejnou paritu). Definujeme *znaménko permutace*: $\operatorname{sgn} \pi = 1$, je-li π sudá, a $\operatorname{sgn} \pi = -1$, je-li π lichá. Z definice snadno plyne, že

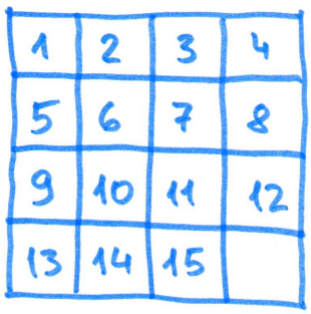
$$\operatorname{sgn}(\pi \circ \sigma) = \operatorname{sgn} \pi \cdot \operatorname{sgn} \sigma \quad \text{a} \quad \operatorname{sgn} \pi^{-1} = \operatorname{sgn} \pi.$$

Z rozkladu $(a_1\ a_2\ \dots\ a_k) = (a_1\ a_k) \circ \dots \circ (a_1\ a_3) \circ (a_1\ a_2)$ vidíme, že cykly sudé délky jsou liché a naopak, a že

$$\operatorname{sgn} \pi = (-1)^{n - \text{počet cyklů v } \pi} = (-1)^{\text{počet cyklů v } \pi \text{ sudé délky}}.$$

Díky uvedeným vztahům tvoří sudé permutace podgrupu v $\mathbf{S}_n$, tzv. *alternující grupu* $\mathbf{A}_n$.

Definice. Permutaci $\rho \circ \pi \circ \rho^{-1}$ nazýváme *permutací konjugovanou* s permutací π podle permutace ρ.



Obrázek 9. Loydova patnáctka a číslování polí

Konjugace má velmi přirozenou interpretaci: pro

$$\pi = (a_{11}\ a_{12}\ \dots\ a_{1k_1}) \cdots (a_{m1}\ a_{m2} \dots\ a_{mk_m})$$

dostáváme

$$\rho \circ \pi \circ \rho^{-1} = (\rho(a_{11})\ \rho(a_{12})\ \dots\ \rho(a_{1k_1})) \cdots (\rho(a_{m1})\ \rho(a_{m2})\ \dots\ \rho(a_{mk_m})),$$

neboť pro každé i, j platí

$$(\rho \circ \pi \circ \rho^{-1})(\rho(a_{ij})) = \rho(\pi(a_{ij})) = \rho(a_{i(j\oplus 1)}),$$

kde $j \oplus 1 = j + 1$ pro $j < k_j$ a $k_j \oplus 1 = 1$. Konjugace podle ρ tedy funguje jako „kopírování“ zápisu podle pravidel daných permutací ρ, každý prvek a v zápise permutace π se přepíše na $\rho(a)$, přičemž struktura cyklů zůstane zachována.

Tvrzení 13.9 (konjugace pro permutace)**.** *Permutace π, σ jsou konjugované v grupě $\mathbf{S}_n$ právě tehdy, když mají stejný počet cyklů každé délky (říká se* stejný typ*).*

Důkaz. ($\Rightarrow$) Plyne bezprostředně z výše uvedeného výpočtu.

($\Leftarrow$) Jsou-li

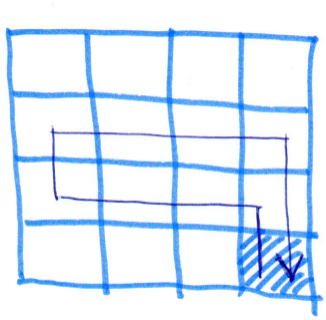
$$\pi = (a_{11}\ a_{12}\ \dots\ a_{1k_1})(a_{21}\ a_{22}\ \dots\ a_{2k_2}) \cdots (a_{m1}\ a_{m2} \dots\ a_{mk_m}),$$
$$\sigma = (b_{11}\ b_{12}\ \dots\ b_{1k_1})(b_{21}\ b_{22}\ \dots\ b_{2k_2}) \cdots (b_{m1}\ b_{m2} \dots\ b_{mk_m}),$$

dvě permutace stejného typu, definujeme $\rho(a_{ij}) = b_{ij}$ a výše uvedeným výpočtem dostaneme $\sigma = \rho \circ \pi \circ \rho^{-1}$. □

Příklad. Permutace (1 2 3) a (2 3 4) jsou konjugované v grupě $\mathbf{S}_4$, protože obě mají jeden cyklus délky 1 a jeden cyklus délky 3. Tyto permutace ovšem nejsou konjugované v grupě $\mathbf{A}_4$: jak plyne z důkazu Tvrzení 13.9, jediné permutace ρ splňující $(2\ 3\ 4) = \rho \circ (1\ 2\ 3) \circ \rho^{-1}$ jsou (1 4), (1 2 3 4) a (1 3 2 4). Žádná z nich ovšem není sudá.

Loydova patnáctka a generátory alternující grupy.

Loydova patnáctka je známý hlavolam, ve kterém se po hrací ploše ve formě čtvercového pole 4×4 posouvá patnáct čtvercových kostiček s čísly $1, ..., 15$. V jednom kroku je možné posunout na prázdné pole jednu ze sousedních kostiček. Cílem je stav, kde jsou kostičky seřazeny vzestupně po řádcích zleva doprava, shora dolů, přičemž prázdné políčko je vpravo dole (viz obrázek). Otázka zní: pro které počáteční stavy lze kostičky přesunout do cílového stavu?



Obrázek 10. Loydova patnáctka: průchod ze základního stavu

Matematicky lze hlavolam popsat následujícím způsobem. Místo prázdného pole budeme uvažovat kostičku s číslem 16. Označme pole hrací plochy jako v cílovém stavu (pole vpravo dole bude mít číslo 16). *Stav hry* lze popsat jako permutaci $\pi \in S_{16}$, kde na poli číslo i je kostička s číslem $\pi(i)$. *Cílový stav* je popsán identickou permutací. V jednom *kroku* je možné prohodit kostičku číslo 16 se sousední kostičkou, čili ze stavu daného permutací π se dostaneme do stavu daného permutací $\pi \circ (i\ j)$, kde i, j jsou sousední pole a $\pi(i) = 16$. Všimněte si, že v jednom kroku se vždy změní znaménko stavové permutace.

Nejprve ukážeme, které stavy nemají řešení. Popíšeme jistou vlastnost stavu, tzv. *invariant*, kterou zachovává každý krok hry. Stavy, které mají jinou hodnotu invariantu než cílový stav, nemohou být řešitelné. Označme $\mathrm{ny}(\pi)$ tzv. *newyorskou vzdálenost* prázdného pole od pravého dolního pole ve stavu daném permutací π (tj. nejmenší počet tahů, který je potřeba na přesunutí prázdného pole na pozici 16). Označme

$$I(\pi) = \mathrm{sgn}(\pi) \cdot (-1)^{\mathrm{ny}(\pi)}$$

součin znaménka permutace a (multiplikativní) parity newyorské vzdálenosti prázdného pole. Vidíme, že $I(\pi)$ se v žádném kroku nemění: jeden krok změní znaménko permutace, ale také paritu newyorské vzdálenosti, čili I je invariantem. Vzhledem k tomu, že $I(id) = 1$, stavy dané permutací π s $I(\pi) = -1$ určitě řešitelné nejsou.

Těžší je dokázat, že všechny stavy s $I(\pi) = 1$ jsou řešitelné. Stav nazveme *základní*, pokud je prázdné pole vpravo dole, tj. pro jeho stavovou permutaci platí $\pi(16) = 16$. Bez újmy na obecnosti se lze soustředit na řešitelnost základních stavů: libovolný jiný stav, daný permutací σ, lze několika kroky převést na základní stav daný permutací σ', přičemž $I(\sigma) = I(\sigma') = \mathrm{sgn}(\sigma')$. Otázka tedy zní, zda jsou všechny základní stavy dané sudou permutací řešitelné.

Všimněte si, že ze základního stavu daného permutací π se lze dostat několika kroky do základních stavů daných permutacemi

- $\pi \circ (9\ 10\ 11\ 12\ 15\ 14\ 13)$ — prázdným polem objedeme dolní obdélník 2×4 (po směru hodinových ručiček).
- $\pi \circ (5\ 6\ 7\ 8\ 11\ 10\ 9)$ — prázdné pole posuneme o 1 nahoru, objedeme s ním prostřední obdélník 2×4 a vrátíme jej dolů (viz obrázek).
- $\pi \circ (1\ 2\ 3\ 4\ 7\ 6\ 5)$ — prázdné pole posuneme o 2 nahoru, objedeme s ním horní obdélník 2×4 a vrátíme jej dolů.

Které základní stavy lze převést na cílový, daný permutací id? Nebo raději obráceně, na které stavy se lze dostat z identity? Jistě na ty, které jsou dané permutacemi, které poskládáme z tří výše uvedených permutací. Problém řešitelnosti Loydovy patnáctky se tedy redukuje na úlohu, zda

$$\mathbf{A}_{15} = \langle (1\ 2\ 3\ 4\ 7\ 6\ 5), (5\ 6\ 7\ 8\ 11\ 10\ 9), (9\ 10\ 11\ 12\ 15\ 14\ 13) \rangle$$

(vzhledem k tomu, že 16 je pevná, můžeme permutace určující základní stavy považovat za prvky $\mathbf{A}_{15}$). Tuto úlohu necháváme jako cvičení ve stylu sekce 11.1.