Kapitola 1

Množiny

1.1 Základní množinové pojmy

Pojem množiny nedefinujeme, pouze připomínáme, že množina je „souhrn, nebo soubor“ navzájem rozlišitelných objektů, kterým říkáme prvky. Pro známé množiny používáme běžné značení — $\mathbb{N}$ pro množinu přirozených čísel , $\mathbb{Z}$ pro množinu celých čísel, $\mathbb{Q}$ pro množinu racionálních čísel a $\mathbb{R}$ pro množinu reálných čísel. Poznamenejme, že do množiny přirozených čísel počítáme i 0.

Nejprve shrneme pojmy a fakta, které znáte ze střední školy.

1.1.1 Stejné množiny. To, kdy jsou dvě množiny stejné, zavádí tzv. princip rovnosti.

Princip rovnosti. Dvě množiny S a T jsou si rovny (píšeme $S = T$), jestliže každý prvek množiny S je prvkem množiny T a naopak každý prvek T je také prvkem S. □

1.1.2 Zadání množiny. Množinu můžeme zadat buď výčtem, tj. vypíšeme všechny její prvky, nebo naznačíme, které prvky obsahuje. Příkladem je např. množina sudých přirozených čísel $S = \{0, 2, 4, 6, 8, \ldots\}$. Množinu zadáváme také vlastností, která charakterizuje její prvky. Je-li $p(x)$ vlastnost, kterou prvek má nebo nemá, pak množinu C všech prvků x s vlastností $p(x)$ a žádných jiných zapisujeme

$$C = \{x \mid p(x)\}.$$

Proto množinu všech sudých přirozených čísel můžeme zadat jako $\{m \mid m = 2k, k \in \mathbb{N}\}$. Množina všech lichých přirozených čísel je množina $\{m \mid m = 2k + 1, k \in \mathbb{N}\}$.

1.1.3 Podmnožiny. Mějme dvě množiny S a T. Jestliže každý prvek množiny S je také prvkem množiny T, říkáme, že S je *podmnožina* T a píšeme $S \subseteq T$.

Jestliže platí $S \subseteq T$ a S a T jsou různé množiny, říkáme též, že S je *vlastní podmnožina* T.

Platí: $S = T$ právě tehdy, když $S \subseteq T$ a současně $T \subseteq S$. □

Jedná se o jednoduchý důsledek principu rovnosti a definice podmnožiny.

1.1.4 Prázdná množina. Mezi množinami hraje významnou roli tzv. prázdná množina. Definujeme: *Prázdná množina* je množina, která nemá žádný prvek; obvykle se značí $\emptyset$.

Platí: $\emptyset \subseteq A$ pro každou množinu A.

Kdyby totiž $\emptyset \not\subseteq A$ pro nějakou množinu A, musel by existovat prvek $x \in \emptyset$, který by neležel v A. Ale $\emptyset$ žádný prvek nemá. Proto platí $\emptyset \subseteq A$ pro každou množinu A.

1.1.5 Operace s množinami. Připomeňme známé operace s množinami.

Mějme dvě množiny A a B. Jejich *sjednocením* je množina

$$A \cup B = \{x \mid x \in A \text{ nebo } x \in B\}.$$

Průnikem těchto dvou množin je množina

$$A \cap B = \{x \,|\, x \in A \text{ a } x \in B\}.$$

Rozdílem množin A a B (v tomto pořadí) je množina

$$A \setminus B = \{x \,|\, x \in A \text{ a } x \notin B\}.$$

Je-li $A \subseteq U$, potom *doplňkem množiny A v množině U* je množina $U \setminus A$. Dolněk množiny A značíme $\overline{A}$.

Je-li $A \cap B = \emptyset$, říkáme, že množiny A a B jsou *disjunktní*.

1.1.6 Kartézský součin. *Kartézský součin* množin A, B (značíme $A \times B$) je definován

$$A \times B = \{(a, b) \,|\, a \in A, b \in B\}.$$

Jestliže se jedná o kartézský součin stejných množin, mluvíme o *kartézských mocninách* množiny A a píšeme A^2 místo $A \times A$. Obecně definujeme pro $n \in \mathbb{N}$, $n > 0$

$$A^n = \{(a_1, a_2, \ldots, a_n) \,|\, a_i \in A\}.$$

1.1.7 Potenční množina. Uvedeme ještě jednu operaci s množinami. Pro danou množinu A můžeme vytvořit množinu obsahující všechny podmnožiny $X \subseteq A$.

Definice. Je dána množina A. Pak množina $P(A)$ definovaná

$$P(A) = \{B \,|\, B \subseteq A\}.$$

se nazývá *potenční množina*. □

Uvědomte si, že potenční množina je vždy neprázdná; ano, obsahuje vždy prázdnou množinu.

1.1.8 Charakteristická funkce podmnožiny. Pro práci s podmnožinami dané množiny se velmi hodí následující pojem charakteristické funkce dané podmnožiny.

Definice. Je dána množina U a její podmnožina A. Pak *charakteristická funkce* χ_A podmnožiny $A \subseteq U$ je zobrazení $\chi_A \colon U \to \{0, 1\}$ definované

$$\chi_A(x) = \begin{cases} 1, & x \in A; \\ 0, & x \notin A. \end{cases}$$

□

Zhruba řečeno hodnota $\chi_A(x)$ je rovna odpovědi na otázku, zda prvek x leží v podmnožině A s tím, že 1 znamená ano, 0 znamená ne.

Poznámky.

- Pro různé podmnožiny A, B je $\chi_A \neq \chi_B$. Ano, je-li x prvek, který leží v jedné množině a neleží ve druhé, pak $\chi_A(x) \neq \chi_B(x)$.
- Ke každé funkci $\chi \colon U \to \{0, 1\}$ existuje $C \subseteq U$, taková, že $\chi = \chi_C$. Je proto jedno, zda pracujeme s podmnožinami dané množiny U nebo s charakteristickými funkcemi podmnožin.
- Platí χ_U je konstantní 1, $\chi_\emptyset$ konstantní 0.
- Známe-li charakteristické funkce podmnožin A, B, odvodíme z nich jednoduše i charakteristické funkce podmnožin $A \cup B$, $A \cap B$ a $\overline{A}$. Ano, $\chi_{A \cup B}(x) = \max(\chi_A(x), \chi_B(x))$, $\chi_{A \cap B}(x) = \min(\chi_A(x), \chi_B(x))$ a $\chi_{\overline{A}}(x) = 1 - \chi_A$.

1.2 Russellův paradox

Pro zvídavé studenty.

1.2.1 V odstavci 1.1.2 jsme uvedli možnost zadat množinu vlastností, tj. utvořit množinu všech prvků s danou vlastností. Formálně se jedná o princip abstrakce:

1.2.2 Princip abstrakce. Je dána vlastnost $\mathcal{K}$, kterou prvky mají nebo nemají. Pak existuje množina

$$C = \{x \mid x \text{ má vlastnost } \mathcal{K}\}.$$

□

Tento princip v sobě ale skrývá spor, jak ukazuje následující paradox tzv. *Russellův paradox*.

1.2.3 Russellův paradox. Uvažujme vlastnost „nebýt prvkem sebe sama“. Tuto vlastnost má řada množin: Např. množina $A = \{2\}$ není prvkem sebe sama, protože množina A má jediný prvek a to 2. Jistě byste našli řadu jiných množin, které nejsou prvkem sebe sama. Utvořme tedy tuto množinu:

$$R = \{x \mid x \notin x\}.$$

Podle principu abstrakce se jedná o dobře utvořenou množinu (nebýt prvkem sebe sama jako množiny je vlastnost $\mathcal{K}$). Nyní se můžeme ptát, zda množina R je prvkem sebe sama nebo ne. Jsou pouze dvě možnosti:

- $R \in R$; ale v tomto případě R musí splňovat vlastnost $\mathcal{K}$, a proto $R \notin R$. Tedy má současně platit $R \in R$ a $R \notin R$ a to nastat nemůže.
- $R \notin R$; v tomto případě R nesplňuje vlastnost $\mathcal{K}$, a proto není pravda, že $R \notin R$, Opět má současně platit $R \notin R$ a neplatit $R \notin R$, takže ani tato situace nastat nemůže.

Chyba spočívá v tom, že jsme předpokládali, že R je množina.

Na základě předchozí úvahy je princip abstrakce v axiomatické teorii množin nahrazen principem vydělení.

1.2.4 Princip vydělení. Mějme vlastnost $\mathcal{K}$, kterou každý prvek má nebo nemá. Pak pro každou množinu U existuje množina skládající se právě ze všech prvků množiny U, které mají vlastnost $\mathcal{K}$. Tuto množinu zapisujeme $\{x \mid x \in U \text{ a } x \text{ má vlastnost } \mathcal{K}\}$, nebo krátce $\{x \in U \mid \mathcal{K}\}$. □

Poznámka. Uvědomte si, že existence množiny vytvořené podle principu vydělení

$$R = \{x \mid x \in U, x \notin x\}$$

již nevede ke sporu. Na otázku, zda R je prvkem sebe sama, můžeme dát tuto odpověď:

- $R \notin R$; to znamená, že není pravda tvrzení $R \in U$ a současně $R \notin R$. Tedy buď není pravda $R \in U$ nebo není pravda $R \notin R$ nebo oboje. Protože $R \notin R$ platí, dostáváme, že $R \notin U$. Tato situace nastat může.

1.3 Mohutnost množin

Než uvedeme, co znamená, že dvě množiny mají stejnou mohutnost, připomeneme ještě pojem bijektivního zobrazení, též zvané vzájemně jednoznačného zobrazení.

1.3.1 Vzájemně jednoznačné zobrazení. Zobrazení f množiny A do množiny B je *bijektivní*, též *vzájemně jednoznačné*, jestliže je prosté a na.

Prosté zobrazení, též *injektivní zobrazení*, je takové, které dvěma různým prvkům x, y množiny A přiřazuje různé prvky $f(x)$, $f(y)$ množiny B.

Zobrazení je na B, též *surjektivní zobrazení*, jestliže pro každý prvek $y \in B$ existuje prvek $x \in A$ takový, že $f(x) = y$.

1.3.2 Mohutnost množin. Pojem mohutnost množin zobecňuje to, čemu v případě konečných množin říkáme počet prvků.

Definice. Řekneme, že dvě množiny A, B mají *stejnou mohutnost*, jestliže existuje vzájemně jednoznačné zobrazení množiny A na množinu B. Tento fakt značíme $|A| = |B|$. □

Poznámka. Poznamenejme, že existuje-li vzájemně jednoznačné zobrazení f množiny A na množinu B, pak také existuje vzájemně jednoznačné zobrazení množiny B na množinu A; ano, stačí vzít inverzní zobrazení f^{-1} k zobrazení f. Je proto v pořádku mluvit o množinách stejné mohutnosti.

1.3.3 Příklad. Množina všech sudých čísel $S = \{0, 2, 4, \ldots\} = \{2n \,|\, n \in \mathbb{N}\}$ a množina všech lichých čísel $L = \{1, 3, 5, \ldots\} = \{2n + 1 \,|\, n \in \mathbb{N}\}$ mají stejnou mohutnost.

Zdůvodnění. Definujme zobrazení $f\colon S \to L$ předpisem:

$$f(2n) = 2n + 1 \text{ pro všechna } n \in \mathbb{N}.$$

Toto zobrazení je prosté, protože z rovnosti $f(2n) = f(2m)$ pro dvě přirozená čísla n, m vyplývá $2n + 1 = 2m + 1$ a tedy $n = m$.

Z popisu množiny L také vidíme, že zobrazení f je na: Ano, každé liché přirozené číslo je tvaru $2m + 1$ a je tedy obrazem sudého čísla $2m$.

1.3.4 Spočetné a nespočetné množiny.

Definice. Řekneme, že množina A je *spočetná*, má-li stejnou mohutnost jako množina všech přirozených čísel $\mathbb{N}$. Jestliže množina A je nekonečná a není spočetná, řekneme, že je *nespočetná*.

Pro množinu, která je buď spočetná nebo konečná, se často používá termín *nejvýše spočetná množina*. □

1.3.5 Uvedeme jednoduchý test, kterým se dá rozhodnout, zda je daná množina spočetná.

Tvrzení. Množina A je spočetná právě tehdy, když ji lze uspořádat do prosté nekonečné posloupnosti (tj. do posloupnosti, ve které se neopakují prvky). □

Zdůvodnění. Jestliže existuje bijekce $f\colon \mathbb{N} \to A$, pak posloupnost

$$f(0), f(1), \ldots, f(n), \ldots$$

je prostá nekonečná posloupnost, do které jsme vypsali všechny prvky množiny A.

Umíme-li množinu A vypsat do prosté nekonečné posloupnosti

$$a_0, a_1, \ldots, a_n, \ldots$$

pak zobrazení f definované $f(i) = a_i$ je bijekce množiny $\mathbb{N}$ na A.

1.3.6 Příklad. Množina všech celých čísel je spočetná.

Zdůvodnění. Využijeme předchozí tvrzení.

Množinu celých čísel uspořádáme do nekonečné prosté posloupnosti tak, že střídavě vypisujeme kladná a záporná čísla; tj.:

$$0, 1, -1, 2, -2, 3, -3, 4, -4, \ldots, n, -n, \ldots$$

Přesněji, $0 = a_0$, $1 = a_1$, $-1 = a_2$, $2 = a_3$, $-2 = a_4$, atd. Obecně: celé kladné číslo n je prvek a_{2n-1} a číslo $-n$ je prvek a_{2n}.

1.3.7 Podobnou úvahou jakou jsme udělali v předchozím příkladě dokážeme i následující tvrzení.

Tvrzení. Sjednocení dvou nejvýše spočetných množin je nejvýše spočetná množina. Sjednocení dvou množin, z nichž jedna je spočetná a druhá je nejvýše spočetná, je spočetná množina. □

1.3.8 Uvedeme ještě jedno jednoduché tvrzení.

Tvrzení. Nekonečná podmnožina spočetné množiny je opět spočetná množina. □

Tedy např. množina všech kladných přirozených čísel je spočetná, ano je to nekonečná podmnožina množiny přirozených čísel. Podobně se ukáže, že množina všech sudých čísel větších než milión je také spočetná.

1.3.9 Mohutnost kartézského součinu. Trochu obtížněji se ukazuje, že i kartézský součin dvou spočetných množin je spočetná množina.

Tvrzení. Kartézský součin dvou nejvýše spočetných množin je nejvýše spočetná množina. □

Myšlenka důkazu. Ukážeme, že množinu $C = A \times B$, kde A a B jsou spočetné množiny ($A = \{a_0, a_1, a_2, \ldots\}$ a $B = \{b_0, b_1, b_2, \ldots\}$), lze uspořádat do prosté posloupnosti podle následujícího schematu:

$$\begin{array}{ccccc} (a_0,b_0) & & (a_0,b_1) & & (a_0,b_2) \quad \ldots \\ & \swarrow & & \swarrow & \\ (a_1,b_0) & & (a_1,b_1) & & (a_1,b_2) \quad \ldots \\ & \swarrow & & \swarrow & \\ (a_2,b_0) & & (a_2,b_1) & & (a_2,b_2) \quad \ldots \\ & \swarrow & & \swarrow & \\ \vdots & & \vdots & & \vdots \end{array}$$

Na schematu je naznačeno, jak množinu C uspořádat. Máme

$$C = \{(a_0,b_0), (a_0,b_1), (a_1,b_0), (a_0,b_2), (a_1,b_1), (a_2,b_0), (a_0,b_3), \ldots\}.$$

Pro zvídavé studenty: dvojice (a_i, b_j) bude v posloupnosti na místě $k = i + \frac{(i+j)(i+j+1)}{2}$.

Platí tedy, že kartézský součin dvou spočetných množin je spočetná množina. Nyní již není těžké nahlédnout, že je-li jedna z množin A, B konečná neprázdná a druhá spočetná, pak kartézský součin bude (nekonečná) množina, která je spočetná. Kartézský součin dvou konečných množin je opět konečná množina.

Ukázali jsme, že tvrzení platí.

1.3.10 Příklad. Množina $\mathbb{Q}$ všech racionálních čísel je spočetná.

Zdůvodnění. Každé racionální číslo lze reprezentovat jako zlomek $\frac{p}{q}$, kde q je nenulové přirozené číslo a p je celé číslo. Tedy zlomky můžeme chápat jako uspořádané dvojice (p, q), kde p je čitatel a q jmenovatel racionálního čísla $\frac{p}{q}$. Navíc množina celých čísel je spočetná, stejně jako množina všech nenulových přirozených čísel. Proto množina M všech dvojic (p, q) je spočetná. Množina racionálních čísel $\mathbb{Q}$ je nyní nekonečná podmnožina množiny M, která obsahuje pouze ty dvojice (p, q), kde p a q jsou nesoudělné. Proto je množina $\mathbb{Q}$ spočetná.

1.3.11 Tvrzení. Sjednocení spočetně mnoha nejvýše spočetných množin je nejvýše spočetná množina.

Jinak řečeno: Jsou-li $A_0, A_1, \ldots A_n, \ldots$ nejvýše spočetné množiny, pak jejich sjednocení $A_0 \cup A_1 \cup \ldots = \bigcup_{i\in\mathbb{N}} A_i$ je nejvýše spočetná množina. □

V případě konečných množin $A_0, A_1, \ldots A_n, \ldots$ můžeme použít obdobné schema jako pro kartézský součin. Pro spočetné množiny A_i potřebujeme ještě tzv. axiom výběru, ale to přesahuje rámec naší přednášky.

1.3.12 Příklad. Mějme konečnou neprázdnou množinu A. Množina $A^\star$ všech konečných posloupností prvků z A je spočetná.

Zdůvodnění. Množinu všech konečných posloupností rozdělíme do množin A_i, $i \in \mathbb{N}$, tak, že množina A_i obsahuje přesně všechny konečné posloupnosti délky i. Pak platí:

$$A^\star = \bigcup_{i\in\mathbb{N}} A_i.$$

Přitom každá z množin A_i je konečná. Je tedy $A^\star$ spočetné sjednocení neprázdných konečných množin A_i splňujících

$$A_i \cap A_j = \emptyset \quad \text{pro různá} \quad i, j \in \mathbb{N}.$$

Proto je $A^\star$ spočetná množina.

1.3.13 Nespočetné množiny. Z výše uvedených příkladů by se mohlo zdát, že každá nekonečná množina je spočetná. Není tomu tak. V dalším ukážeme, že množina všech **nekonečných** posloupností 0 a 1 není spočetná. Tohoto faktu dále využijeme k tomu, abychom ukázali, že všech podmnožin množiny $\mathbb{N}$ je nespočetně. To znamená, že množina všech podmnožin spočetné množiny $\mathbb{N}$ má více prvků než sama $\mathbb{N}$.

1.3.14 Cantorova diagonální metoda ukazuje, že množina všech nekonečných posloupností 0 a 1 není spočetná. □

Předpokládejme (pro spor), že množina všech nekonečných posloupností nul a jedniček je spočetná, tudíž že ji můžeme uspořádat do prosté nekonečné posloupnosti. Znázorněme si nějaké takové uspořádání do posloupnosti schematicky — v prvním řádku máme posloupnost s_0, v druhém řádku posloupnost s_1, ve třetím řádku posloupnost s_2, atd.

$$\begin{array}{lcccccccc}
s_0 & = & \boxed{s_0(0),} & s_0(1), & s_0(2), & s_0(3), & s_0(4), & s_0(5), & \dots \\
s_1 & = & s_1(0), & \boxed{s_1(1),} & s_1(2), & s_1(3), & s_1(4), & s_1(5), & \dots \\
s_2 & = & s_2(0), & s_2(1), & \boxed{s_2(2),} & s_2(3), & s_2(4), & s_2(5), & \dots \\
s_3 & = & s_3(0), & s_3(1), & s_3(2), & \boxed{s_3(3),} & s_3(4), & s_3(5), & \dots \\
s_4 & = & s_4(0), & s_4(1), & s_4(2), & s_4(3), & \boxed{s_4(4),} & s_4(5), & \dots \\
s_5 & = & s_5(0), & s_5(1), & s_5(2), & s_5(3), & s_5(4), & \boxed{s_5(5),} & \dots \\
\vdots & & & & & & & &
\end{array}$$

Vytvoříme novou posloupnost $\overline{s}$ nul a jedniček a ukážeme o ní, že nebyla vypsána; to bude ve sporu s tím, že jsme vypsali **všechny** posloupnosti nul a jedniček.

Definujeme posloupnost $\overline{s}$:

- Jestliže v prvním rámečku schematu bylo číslo 1, začíná $\overline{s}$ číslem 0,
- jestliže v prvním rámečku schematu bylo číslo 0, začíná posloupnost $\overline{s}$ číslem 1.

Jinými slovy, posloupnost $\overline{s}$ má na nultém místě to druhé z čísel 0 a 1, než které má posloupnost s_0, což můžeme zapsat: $\overline{s}(0) = 1 - s_0(0)$. Dále postupujeme obdobně: Jestliže v druhém rámečku má posloupnost s_1 číslo 0, položíme $\overline{s}(1) = 1$; je-li $s_1(1) = 1$, položíme $\overline{s}(1) = 0$, tj. $\overline{s}(1) = 1 - s_1(1)$. Dále $\overline{s}(2)$ bude číslo $1 - s_2(2)$, atd.

Formálně zápis vypadá takto: $\overline{s} = \{\overline{s}(0), \overline{s}(1), \overline{s}(2), \dots, \overline{s}(n), \dots\}$, kde $\overline{s}(n) = 1 - s_n(n)$ pro všechna $n \in \mathbb{N}$.

Posloupnost $\overline{s}$ mezi posloupnostmi $s_0, s_1, s_2, \dots, s_n, \dots$ není, neboť od posloupnosti s_0 se liší na nultém místě, od posloupnosti s_1 se liší na prvním místě, od posloupnosti s_2 se liší na druhém místě, ... od n-té posloupnosti s_n se liší na n-tém místě. Dospěli jsme ke sporu s předpokladem, že jsme na začátku vypsali všechny posloupnosti. Proto množina všech nekonečných posloupností nul a jedniček není spočetná.

Věta. Množina $P(\mathbb{N})$ je nespočetná. □

Důkaz. Víme, že každé podmnožině $X \subseteq \mathbb{N}$ odpovídá její charakteristická funkce $\chi_X \colon \mathbb{N} \to \{0, 1\}$. A toto zobrazení je vlastně nekonečná posloupnost 0 a 1. Proto je $P(\mathbb{N})$ je nespočetná.

1.3.15 Poznámka. Obdobně jako jsme ukázali, že množina všech podmnožin množiny přirozených čísel je nespočetná, je možné ukázat, že množina všech reálných čísel v otevřeném intervalu $(0, 1)$ je také nespočetná.

1.3.16 Pro zvídavé studenty uvádíme ještě jednu větu, která ukazuje, že nespočetné množiny mohou mít různou mohutnost. Tím, že vytvoříme množinu všech podmnožin $P(\mathbb{N})$ dostaneme „větší nekonečno“, atd.

Věta. Pro žádnou množinu S neexistuje zobrazení $f\colon S \to P(S)$, které je na $P(S)$. □

Zdůvodnění. Předpokládejme, že f je zobrazení množiny S do $P(S)$. Definujme

$$C = \{x \in S \mid x \notin f(x)\}.$$

(Uvědomte si, že $f(x) \in P(S)$, tj. $f(x) \subseteq S$.)

Tím jsme definovali podmnožinu C množiny S. Ukážeme, že C není f-obrazem žádného prvku $y \in S$ (tj. $C \neq f(y)$ pro žádné $y \in S$). Tím bude ukázáno, že zobrazení f není na.

Vybereme libovolný prvek $a \in S$. Pak platí buď $a \in C$ nebo $a \notin C$.

Jestliže $a \in C$, pak podle definice množiny C máme $a \notin f(a)$. Tudíž C a $f(a)$ musí být různé množiny, protože C obsahuje prvek a a ten není prvkem $f(a)$.

Jestliže $a \notin C$, pak podle definice množiny C platí $a \in f(a)$. Tudíž i v tomto případě $C \neq f(a)$, protože prvek a leží v $f(a)$, ale neleží v C).

Proto množina C není rovna obrazu $f(a)$ pro žádné $a \in A$. Ukázali jsme, že zobrazení f není na množinu $P(S)$.

Protože f bylo libovolné zobrazení S do $P(S)$, ukázali jsme, že neexistuje zobrazení množiny S na $P(S)$.

1.3.17 Poznámka. Ukázali jsme, že pro žádnou množinu S neexistuje zobrazení množiny S *na* množinu všech jejích podmnožin $P(S)$. To intuitivně říká, že S „má méně prvků“ než $P(S)$. Povšimněte si přitom podobnosti předchozí úvahy s Russellovým paradoxem. (Cantor ovšem takto uvažoval více než 10 let před Russellem!)

Kapitola 2

Relace

2.1 Binární relace

Jeden ze základních matematických pojmů je pojem relace, speciálně binární relace. Zhruba řečeno binární relace je „vztah“ mezi dvěma prvky, který mezi nimi buď „je“ nebo „není“. Později se seznámíte i s unárními relacemi, ternárními relacemi a dalšími, které využijete např. v relačních databázích.

2.1.1 Definice. *Relace* (přesněji *binární relace*) *z množiny A do množiny B* je libovolná množina uspořádaných dvojic $R \subseteq A \times B$. Jestliže $A = B$, mluvíme o *relaci na množině A*. □

Příklady.

1. Být dědečkem. Jedná se o relaci R na množině A všech lidí. Dvojice (a, b) patří do R právě tehdy, když osoba a je dědečkem osoby b.

2. Být stejně dlouhé. Jedná se o relaci na množině všech objektů (tady se musíte rozhodnout, které objekty chcete uvažovat); pro dva objekty a, b platí $(a, b) \in R$ právě tehdy, když oba objekty jsou stejně dlouhé.

3. Být podmnožinou. Jedná o relaci R na množině všech podmnožin množiny U. Pro dvě množiny X, Y, $X \subseteq U$, $Y \subseteq U$ platí: (X, Y) je v relaci R právě tehdy, když množina X je podmnožinou množiny Y.

4. Být menší nebo rovno. Jedná se např. o relaci R na množině všech přirozených čísel $\mathbb{N}$, kde $(m, n) \in R$ právě tehdy, když $m \leq n$.

5. Být studentem studijní skupiny. Jedná se o relaci R z množiny A všech studentů prvého ročníku FEL do množiny B všech studijních skupin. Dvojice (a, K), kde a je student a K je studijní skupina, patří do relace R právě tehdy, když je student a zapsán do skupiny K.

6. Funkce sinus. Jedná o relaci R na množině reálných čísel $\mathbb{R}$ definovanou: $(x, y) \in R$ právě tehdy, když $y = \sin x$.

Konvence. Zápis $(a, b) \in R$ je často nepříliš šťastný; nikoho by nenapadlo psát $(X, Y) \in \subseteq$, či dokonce $(2, 3) \in \leq$ (místo toho píše $X \subseteq Y$ a $2 \leq 3$). Proto i my v dalším textu budeme místo zápisu $(a, b) \in R$ používat zápis $a \, R \, b$.

2.1.2 Poznámka. Každé zobrazení $f : A \to B$ je relace (nebo přesněji definuje relaci); a to relace f z A do B, kde $x \, f \, y$ právě tehdy, když $y = f(x)$.

Naproti tomu ne každá relace z A do B je zobrazením z množiny A do množiny B; k tomu, aby relace R byla zobrazením je třeba (a stačí), aby pro každé $a \in A$ existovalo nejvýše jedno $b \in B$ takové, že $a \, R \, b$.

Pro relace přejímáme termíny, které jsou běžné, když mluvíme o zobrazeních. *Definičním oborem* relace R je množina všech $a \in A$, pro něž existuje $b \in B$ takové, že $a\ R\ b$; *oborem hodnot* relace R je množina všech $b \in B$, pro něž existuje $a \in A$ takové, že $a\ R\ b$.

2.1.3 Matice relace. Jestliže obě množiny A a B jsou konečné, pak relaci $R \subseteq A \times B$ můžeme reprezentovat také maticí:

Označme $A = \{a_1, a_2, \ldots, a_n\}$ a $B = \{b_1, b_2, \ldots, b_k\}$. Položme $M_R = (m_R(i,j))$ s n řádky a k sloupci, kde

$$m_R(i,j) = 1 \text{ pro } (a_i, b_j) \in R \quad \text{a} \quad m_R(i,j) = 0 \text{ pro } (a_i, b_j) \notin R.$$

Uvědomte si, že se vlastně jedná o charakteristickou funkci relace R jakožto podmnožiny množiny všech uspořádaných dvojic $A \times B$, pouze zapsané do „obdélníkového schematu“, tj. do matice.

2.1.4 Podrelace.

Definice. Řekneme, že relace R je *podrelací* relace S, jestliže $R \subseteq S$; tj. kdykoli platí $a\ R\ b$, pak platí i $a\ S\ b$. □

Např. „býti menší než“ je podrelací relace „býti menší nebo rovno“.

2.2 Operace s relacemi

Z relací můžeme vytvářet další relace; jsou to jednak množinové operace (vždyť relace jsou vlastně množiny uspořádaných dvojic), jednak operace, které jsou pro množiny specifické — což je vytvoření inverzní relace a skládání relací.

2.2.1 Množinové operace s relacemi.

Definice. Mějme dvě relace R a S z množiny A do množiny B. Pak

- *průnikem* relací R a S je relace $R \cap S$;
- *sjednocením* těchto relací je relace $R \cup S$;
- *doplňkem* relace R je relace $\overline{R} = (A \times B) \setminus R$.

□

Např. označíme-li T relaci rovnosti na množině reálných čísel $\mathbb{R}$ a S relaci býti ostře menší také na množině $\mathbb{R}$, pak $T \cap S = \emptyset$ a $T \cup S$ je relace býti menší nebo roven. Doplňkem relace T je nerovnost, tj. relace $\overline{T} = \{(a,b) \mid a, b \in \mathbb{R}, a \neq b\}$.

2.2.2 Inversní relace.

Definice. Mějme relaci R z množiny A do množiny B. Pak *inversní relací* k relaci R je relace R^{-1} z množiny B do množiny A definovaná takto:

$$x\ R^{-1}\ y \quad \text{právě tehdy, když} \quad y\ R\ x.$$

□

Poznámky.

1. Uvědomte si, že inversní relace k relaci z příkladu 6 na straně 8 existuje: Je to relace R^{-1}, kde $x \in \mathbb{R}$, $y \in \mathbb{R}$ a platí $x\ R^{-1}\ y$ právě tehdy, když $y\ R\ x$, tj. právě tehdy, když $x = \sin y$.

   Přesto z matematické analýzy víte, že inversní funkce k funkci $y = \sin x$ ($y = \arcsin(x)$) má za definiční obor interval $\langle -\pi/2, \pi/2 \rangle$. Je to proto, že inversní relace sice existuje, ale není již zobrazením; není totiž pravda, že pro každé $-1 \leq x \leq 1$ existuje právě jedno $y \in \mathbb{R}$ tak, že $x = \sin y$; ano, např. $\sin 0 = \sin \pi$.

2. Jsou-li množiny A, B konečné a representujeme-li relaci R maticí M_R, pak matice inversní relace R^{-1} je matice transponovaná k matici M_R. Tj.

$$m_{R^{-1}}(j,i) = m_R(i,j),\ i = 1, \ldots, n,\ j = 1, \ldots, k.$$

2.2.3 Skládání relací.

Definice. Mějme relaci R z množiny A do množiny B a S relaci z množiny B do množiny C. Pak *složená relace* $R \circ S$ je relace z množiny A do množiny C definovaná předpisem:

$$a \; R \circ S \; c \quad \text{právě tehdy, když existuje } b \in B \text{ takové, že} \quad a \; R \; b \;\text{ a }\; b \; S \; c.$$

□

Poznámka. Je-li M_R matice relace R a M_S matice relace S, pak matice relace $R \circ S$ je rovna součinu

$$M_{R \circ S} = M_R \cdot M_S,$$

s tím, že „počítáme“ $1 + 1 = 1$. □

2.2.4 Tvrzení. Skládání relací je asociativní. Přesněji, je-li R relace z množiny A do množiny B, relace S z množiny B do množiny C a relace T z množiny C do množiny D, pak platí

$$R \circ (S \circ T) = (R \circ S) \circ T.$$

Zdůvodnění. K důkazu tvrzení si stačí uvědomit, že $a \; R \circ (S \circ T) \; d$ platí právě tehdy, když existují prvky $b \in B$ a $c \in C$ takové, že $a \; R \; b$, $b \; S \; c$ a $c \; T \; d$, a to je právě tehdy, když $a \; (R \circ S) \circ T \; d$.

2.2.5 Poznámka. Skládání relací **není** komutativní, tj. existují relace R a S na množině A, pro které neplatí $R \circ S = S \circ R$. To není překvapující, skládání funkcí také není komutativní.

Uvedeme jeden „nematematický“ příklad: Uvažujme množinu A všech lidí v České republice a dvě relace R, S definované na A:

$$\begin{aligned} a \; R \; b &\quad \text{právě tehdy, když } a \text{ je sourozenec } b \text{ a } a \neq b \\ c \; S \; d &\quad \text{právě tehdy, když } c \text{ je dítětem } d. \end{aligned}$$

Ukážeme, že $R \circ S \neq S \circ R$.

Abychom ukázali, že $R \circ S \neq S \circ R$, stačí najít dva lidi a, b tak, že platí $a \; R \circ S \; b$ a neplatí $a \; S \circ R \; b$. Uvažujme dvojici synovec a a strýc b. Platí $a \; S \circ R \; b$, protože jeden z rodičů a je sourozencem strýce b. Neplatí ale, že $a \; R \circ S \; b$ protože to by znamenalo, že některý ze sourozenců a by byl rodičem strýce b.

2.3 Relace na množině

Jak už jsme viděli i ve výčtu příkladů binárních relací, řada z relací je relacemi na množině. V dalším se zaměříme právě na ně. (Připomeňme, že relace $R \subseteq A \times B$ se nazývá relace na množině A vždy, když $A = B$.)

Nejprve uvedeme čtyři vlastnosti, které některé relace na množině mají (a některé nemají). Ty nám později umožní zavést dva speciální typy relací na množině — relaci ekvivalence a částečné uspořádání.

2.3.1 Reflexivita, symetrie, antisymetrie a tranzitivita.

Definice. Řekneme, že relace R na množině A je

1. *reflexivní*, jestliže pro všechna $a \in A$ platí $a \; R \; a$;
2. *symetrická*, jestliže pro všechna $a, b \in A$ platí: je-li $a \; R \; b$, pak také $b \; R \; a$;
3. *antisymetrická*, jestliže pro všechna $a, b \in A$ platí: je-li $a \; R \; b$ a $b \; R \; a$, pak nutně $a = b$;
4. *tranzitivní*, jestliže pro všechna $a, b, c \in A$ platí: je-li $a \; R \; b$ a $b \; R \; c$, pak nutně $a \; R \; c$.

□

Příklady.

1. Uvažujme relaci nerovnosti R na množině přirozených čísel $\mathbb{N}$ (tj. $n\ R\ m$ právě tehdy, když n a m jsou různá přirozená čísla).

Tato relace není reflexivní, protože pro žádné $n \in \mathbb{N}$ neplatí $n \neq n$.

Tato relace je symetrická: Je-li $n \neq m$, pak také $m \neq n$.

Tato relace R není antisymetrická, protože např. $2 \neq 3$, $3 \neq 2$ a 2 a 3 jsou různá čísla (tj. $2\ R\ 3$ a $3\ R\ 2$ a přesto $2 \neq 3$).

Tato relace také není tranzitivní, protože např. $2 \neq 3$ a $3 \neq 2$ a přesto $2 = 2$ (tj. $2\ R\ 3$ a $3\ R\ 2$ a přesto není $2\ R\ 2$).

2. Relace menší nebo rovno $\leq$ na množině $\mathbb{R}$ je reflexivní, neboť $a \leq a$ pro všechna reálná a. Není symetrická, protože např. $2 \leq 3$, ale $3 \not\leq 2$. Je antisymetrická, neboť jakmile pro dvě reálná čísla a, b platí $a \leq b$ a $b \leq a$, pak jsou nutně stejná, tj. $a = b$. Relace je také tranzitivní, neboť je-li $a \leq b$ a $b \leq c$, pak i $a \leq c$.

2.3.2 Relace ekvivalence. Jeden z nedůležitějších typů relací na množině je relace ekvivalence. Zhruba řečeno, ekvivalence na množině A je „zobecněná rovnost“. Tím, že dva prvky množiny A jsou ekvivalentní (jsou v relaci), vlastně říkáme, že jsou pro nás v jistém smyslu „stejné“. Jako příklad uveďme podobnost trojúhelníků v rovině: dva trojúhelníky jsou podobné (v jistém smyslu „stejné“) mají-li stejné úhly u odpovídajících stran. Další příklady uvedeme dále.

Definice. Relace R na množině A se nazývá *ekvivalence*, jestliže je reflexivní, symetrická a tranzitivní. □

2.3.3 Příklad. Relace R na množině všech celých čísel $\mathbb{Z}$ definovaná předpisem:

$$m\ R\ n \quad \text{právě tehdy, když} \quad m - n \text{ je sudé}, \ (m, n \in \mathbb{Z}),$$

je ekvivalence.

Zdůvodnění: Relace R je reflexivní, protože pro každé $m \in \mathbb{Z}$ je $m - m = 0$ a nula je sudé číslo. Tedy $m\ R\ m$.

Relace R je symetrická protože, je-li $m\ R\ n$, tj. $m - n = 2k$ pro nějaké celé číslo k, je i $n - m$ sudé ($n - m = -2k$) a proto $n\ R\ m$.

Navíc R je tranzitivní: Máme-li tři čísla $m, n, p \in \mathbb{Z}$ taková, že $m\ R\ n$ a $n\ R\ p$, tj. $m - n = 2k$ a $n - p = 2l$ pro nějaká celá čísla k a l, potom $m - p = (m - n) + (n - p) = 2k + 2l = 2(k + l)$. Odtud plyne $m\ R\ p$.

2.3.4 Příklad. Uvažujme realaci S na množině $A = \{(p, q) \mid q \neq 0,\ p, q \in \mathbb{Z}\}$ definovánou předpisem:

$$(p, q)\ S\ (m, n) \quad \text{právě tehdy, když} \quad pn = qm$$

Pak S je ekvivalence na množině A.

Zdůvodnění: S je reflexivní: Pro všechny $(p, q) \in A$ máme $(p, q)\ S\ (p, q)$, protože $pq = qp$.

S je symetrická: je-li $(p, q)\ S\ (m, n)$, tj. $pn = qm$, je i $mq = np$. To znamená, že $(m, n)\ S\ (p, q)$.

S je tranzitivní: Předpokládejme $(p, q)\ S\ (m, n)$ a $(m, n)\ S\ (r, s)$, tj. $pn = qm$ a $ms = nr$. K tomu, abychom ukázali, že S je tranzitivní relace, potřebujeme ověřit, že $(p, q)\ S\ (r, s)$, tj., že $ps = qr$. Protože $pn = qm$ a n je nenulové, máme $p = \frac{qm}{n}$. Odtud $ps = \frac{qm}{n}\, s = \frac{q}{n}\, ms = \frac{q}{n}\, nr = qr$. (Užili jsme rovnost $ms = nr$.) Tedy S je tranzitivní relace.

2.3.5 Rozklad množiny. Protože je relace ekvivalence „zobecněná rovnost“, měli bychom mít možnost mluvit o „stejných prvcích“; tj. měli bychom mít možnost rozdělit všechny prvky množiny na skupiny „pro nás se nelišících“. Rozdělení prvků dané množiny zachycuje pojem *rozklad množiny*.

Definice. Mějme neprázdnou množinu A. Množina $\mathcal{S}$ neprázdných podmnožin množiny A se nazývá *rozklad množiny* A, jestliže jsou splněny následující podmínky:

1. Každý prvek $a \in A$ leží v některé podmnožině z $\mathcal{S}$, tj. $\bigcup \mathcal{S} = A$.
2. Prvky množiny $\mathcal{S}$ jsou po dvou disjunktní; tj. pro všechna $X, Y \in \mathcal{S}$ platí: jestliže $X \cap Y \neq \emptyset$, pak $X = Y$.

□

2.3.6 Třídy ekvivalence. Nyní upřesníme pojem „stejných“ prvků vzhledem k dané relaci ekvivalence. „Stejné prvky“ jako prvek $a \in A$ budou tvořit třídu ekvivalence obsahující prvek a. Přesněji

Definice. Je dána relace ekvivalence R na množině A. *Třídou ekvivalence* R odpovídající prvku $a \in A$ nazýváme množinu

$$R[a] = \{b \in A \mid a \; R \; b\}.$$

Množinu všech tříd dané ekvivalence, tj. množinu $\{R[a] \mid a \in A\}$ často nazýváme *faktorovou množinou podle ekvivalence* R a značíme A/R. □

2.3.7 Příklady. Uvažujme relaci ekvivalence R z příkladu 2.3.3. Tato relace má dvě třídy ekvivalence, a to $R[0]$, což je množina všech sudých čísel, a $R[1]$, což je množina všech lichých čísel.

Dříve než najdeme třídu ekvivalence S z příkladu 2.3.4 k obecné dvojici $(p, q) \in A$, zaměříme se na třídu ekvivalence obsahující dvojici $(1, 2)$. Tato třída obsahuje všechny dvojice (s, t) pro něž $1 \cdot t = s \cdot 2$, tj. pro něž $t = 2s$. Kdybychom si dvojice představili jako zlomky, byly by to všechny zlomky, které po zkrácení dávají racionální číslo $\frac{1}{2}$.

Pro obecnou dvojici $(p, q) \in A$ platí, že $(s, t) \; S \; (p, q)$ právě tehdy, když $\frac{s}{t} = \frac{p}{q}$. Tedy třídy ekvivalence odpovídají jednotlivým racionálním číslům.

2.3.8 Tvrzení. Nechť R je ekvivalence na množině A. Množina tříd ekvivalence $\{R[a] \mid a \in A\}$ tvoří rozklad množiny A, tj. má tyto vlastnosti:

1. Každá množina $R[a]$ je neprázdná a platí $\bigcup\{R[a] \mid a \in A\} = A$.
2. Třídy ekvivalence $R[a]$ jsou po dvou disjunktní, tj. jestliže $R[a] \cap R[b] \neq \emptyset$, pak $R[a] = R[b]$.

□

Zdůvodnění. Vlastnost 1 vyplývá z faktu, že relace R je reflexivní; ano, $a \in R[a]$ a proto nejen je množina $R[a]$ vždy neprázdná, ale navíc platí $\bigcup\{R[a] \mid a \in A\} = A$.

Abychom dokázali vlastnost 2, předpokládejme, že $R[a] \cap R[b] \neq \emptyset$ pro nějaké $a, b \in A$. To znamená, že existuje $c \in R[a] \cap R[b]$. Proto $a \; R \; c$ a $b \; R \; c$. Ze symetrie relace R dostáváme $a \; R \; c$, $c \; R \; b$; tudíž z tranzitivity relace R máme $a \; R \; b$. Nyní použijeme-li opět symetrii relace R dostáváme i $b \; R \; a$.

Vezměme nyní některý prvek $d \in R[a]$. Pak platí $b \; R \; a$ a $a \; R \; d$, proto $b \; R \; d$ (použili jsme tranzitivitu). Proto $R[a] \subseteq R[b]$. Analogicky se ukáže, že $R[b] \subseteq R[a]$ a proto platí $R[a] = R[b]$.

2.3.9 Už víme, že každé relaci ekvivalence můžeme utvořit rozklad „na množiny stejných prvků“. Ukážeme, že z rozkladu množiny A se dá sestrojit ekvivalence, která má původní rozklad jako rozklad na třídy ekvivalence.

Tvrzení. Nechť $\mathcal{S}$ je rozklad množiny A. Pak relace $R_{\mathcal{S}}$ definovaná:

$$a \; R_{\mathcal{S}} \; b \quad \text{právě tehdy, když} \quad \text{existuje } X \in \mathcal{S} \text{ takové, že } a, b \in X$$

je ekvivalence na množině A. □

Zdůvodnění. Protože $\mathcal{S}$ je rozklad, každý prvek $a \in A$ leží v některé množině rozkladu, tudíž relace $R_{\mathcal{S}}$ je reflexivní. Navíc, být prvkem stejné množiny je symetrická vlastnost, proto je relace $R_{\mathcal{S}}$ i symetrická.

Ukážeme, že $R_{\mathcal{S}}$ je tranzitivní: Předpokládejme, že $a \; R_{\mathcal{S}} \; b$ a $b \; R_{\mathcal{S}} \; c$ pro nějaké $a, b, c \in A$. To znamená, že existují $S_1, S_2 \in \mathcal{S}$ takové, že $a, b \in S_1$ a $b, c \in S_2$. Máme tedy $b \in S_1 \cap S_2$ a proto z vlastnosti 2 rozkladu víme, že $S_1 = S_2$. Odtud $a, c \in S_1$ a $a \; R_{\mathcal{S}} \; c$.

2.3.10 Poznámka. Je dobré si uvědomit, že vyjdeme-li od ekvivalence R, utvoříme k ní rozklad na její třídy ekvivalence $\mathcal{S} = \{R[a] \mid a \in A\}$, načež k tomuto rozkladu utvoříme (podle předchozího tvrzení) opět ekvivalenci $R_{\mathcal{S}}$, pak dostaneme původní ekvivalenci R.

Podobně, kdybychom vyšli od rozkladu, utvořili k němu ekvivalenci a k této ekvivalenci rozklad na její třídy, dostali bychom původní rozklad.

Je proto jedno, zda se na relaci ekvivalence díváme jako na množinu uspořádaných dvojic nebo s ní pracujeme jako s rozkladem na množiny těch prvků, které jsou pro nás „stejné“ — ekvivalentní. Jedná se tedy o dva různé pohledy na tutéž matematickou realitu.

2.4 Částečné uspořádání

Je ještě jeden typ binárních relací, které v aplikacích hrají důležitou roli. Jsou to tzv. částečná uspořádání a částečně uspořádané množiny (v angličtině poset).

Definice. Mějme relaci R na množině A. Relaci R nazveme *částečné uspořádání* na A, jestliže je reflexivní, antisymetrická a tranzitivní.

Částečně uspořádaná množina (v angličtině *poset*) je dvojice (A, R), kde A je množina a R je částečné uspořádání na množině A. □

Příklady.

1. Relace $\leq$ na množině reálných čísel $\mathbb{R}$ je částečné uspořádání.
2. Relace „ býti podmnožinou“ na množině $P(U) = \{X \mid X \subseteq U\}$ je částečné uspořádání.
3. Relace „býti dělitelem“ na množině přirozených čísel $\mathbb{N}$ je částečné uspořádání.

V dalším textu značíme částečné uspořádání místo R symbolem $\sqsubseteq$. To proto, abychom zdůraznili, že relace není libovolná, ale je částečným uspořádáním.

2.4.1 Největší, maximální, nejmenší a minimální prvek. Máme-li na množině A definované částečné uspořádání $\sqsubseteq$, pak v množině A může a nemusí existovat největší či nejmenší prvek; tj. prvek, který je „větší“ či „menší“ než všechny ostatní prvky. Kromě těchto mohou existovat maximální či minimální prvky, které „nejsou menší než žádný jiný prvek“ či „nejsou větší než žádný jiný prvek“. Přesněji:

Definice. Mějme částečně uspořádanou množinu $(A, \sqsubseteq)$. Řekneme, že prvek $a \in A$ je *největší* (resp. *nejmenší*) prvek, jestliže pro každé $b \in A$ platí $b \sqsubseteq a$ (resp. $a \sqsubseteq b$).

Řekneme, že prvek $a \in A$ je *maximální* (resp. *minimální*) prvek, jestliže neexistuje prvek $b \in A$, $b \neq a$, takový, že $a \sqsubseteq b$ (resp. $b \sqsubseteq a$). □

Poznámka. Jestliže existuje největší (resp. nejmenší) prvek, pak je i maximální (resp. minimální) a další maximální (resp. minimální) prvek neexistuje. Jestliže má uspořádaná množina víc než jeden maximální (resp. minimální) prvek, tak nemá největší (resp. nejmenší) prvek. Jsou ovšem i příklady částečně uspořádaných množin, které nemají ani maximální, ani minimální prvky.

Kapitola 3

Celá čísla

3.1 Celá čísla a jejich vlastnosti

Znalost a pochopení celých čísel a jejich vlastností je základ mnoha aplikací; za všechny uveďme kódování a šifrování. A právě znalosti potřebné k těmto aplikacím probereme v této kapitole.

Nejprve připomeneme známá fakta ze střední školy o dělení celých čísel. Jedná se o dělení se zbytkem (tj. o celočíselné dělení), pojem společného dělitele a největšího společného dělitele, pojmy prvočísla, čísla složená a čísla nesoudělná. Znalosti rozšíříme o Euklidův algoritmus pro nalezení největšího společného dělitele a jeho důsledky, speciálně důsledky pro řešení diofantických rovnic — lineárních rovnic, kde hledáme pouze celočíselná řešení.

3.1.1 Věta o dělení. Pro každá dvě celá čísla a, b, $b > 0$, existují celá čísla q, r taková, že

$$a = q\,b + r, \quad 0 \le r < b.$$

Tato čísla jsou určena jednoznačně. □

3.1.2 Poznámky: 1. Číslo q z předchozí věty se nazývá *částečný podíl* a číslo r *zbytek* při dělení čísla a číslem b.

2. S větou o dělení jste se setkali hlavně v případě přirozených čísel. Pro kladná čísla a, b není třeba dokazovat existenci částečného podílu a zbytku; jedná se o použití písemného dělení.

My jsme formulovali větu o dělení i pro záporná čísla a. Ukážeme, že i v tomto případě q a r existují. Jestliže je a záporné, nejprve vydělíme $|a|$ číslem b, tj. máme $|a| = q'b + r'$ pro $0 \le r' < b$, $q' \ge 0$.

Protože $a < 0$, tak $a = -q'b - r'$. Jestliže $r' = 0$, položíme $r = 0$ a $q = -q'$.

Jestliže $a < 0$ a $r' > 0$. V takovém případě platí $a = -(q' + 1)b + b - r'$ a $b - r' > 0$. Proto položíme $q = -(q' + 1)$ a $r = b - r'$.

Ukážeme si postup na příkladu. Při dělení čísla $a = -7$, $b = 4$, vydělíme číslo 7 číslem 4 a dostáváme $7 = 1 \cdot 4 + 3$, tedy $-7 = (-1) \cdot 4 - 3 = (-2) \cdot 4 + 4 - 3$. Zbytek při dělení je $r = 4 - 3 = 1$ a částečný podíl je -2. Ano, $-7 = (-2) \cdot 4 + 1$ a $1 < 4$.

3.1.3 Důkaz jednoznačnosti ve větě o dělení. Ještě dokážeme jednoznačnost částečného podílu a zbytku. Předpokládejme, že by čísla q a r z 3.1.1 nebyla jednoznačná; tj. pro některá čísla a, b by existovala čísla q_1, r_1 a q_2, r_2, kde $0 \le r_1, r_2 < b$ takové, že

$$a = q_1\,b + r_1, \quad \text{a} \quad a = q_2\,b + r_2.$$

Pak

$$q_1\,b + r_1 = q_2\,b + r_2, \quad \text{tj.} \quad (q_1 - q_2)\,b = r_2 - r_1.$$

Protože $|r_2 - r_1| < b$ se má rovnat násobku čísla b, musí $q_1 - q_2 = 0$ (v opačném případě by $|q_1 - q_2|\,b \ge b$). To ale znamená, že $q_1 = q_2$ a $r_1 = r_2$. Ukázali jsme, že částečný podíl i zbytek jsou jediné.

3.1.4 Další známé pojmy.

Relace dělitelnosti. Celé číslo b *dělí* celé číslo a jestliže $a = k \cdot b$ pro nějaké celé číslo k. Tento fakt značíme $b \mid a$.

Jestliže $b \mid a$ říkáme také, že a je *násobkem* čísla b. □

Prvočíslo, složené číslo. Přirozené číslo p se nazývá *prvočíslo*, jestliže $p > 1$ a p je dělitelné pouze 1 a p. Přirozené číslo $n > 1$ se nazývá *složené*, jestliže není prvočíslo; jinými slovy má dělitele r takového, že $1 < r < n$. □

Všimněte si, že

1. 0 dělí sebe sama. Ano, například platí $0 = 1 \cdot 0$ a 1 je celé číslo.
2. Je-li $b > 0$, pak b dělí a právě tehdy, když je zbytek při dělení čísla a číslem b roven 0.
3. Číslo 1 není ani složené, ani prvočíslo.

3.1.5 Společný dělitel, největší společný dělitel, nesoudělná čísla.

Definice.

1. Číslo c je *společný dělitel* čísel a, b, jestliže $c \mid a$ a $c \mid b$.
2. Číslo d je *největší společný dělitel* a, b, jestliže
   - $d \geq 0$,
   - d je společný dělitel a, b, tj. $d \mid a$ a $d \mid b$,
   - a kdykoli je c společný dělitel a, b, pak $c \mid d$.

   Největší společný dělitel čísel a, b značíme $\gcd(a, b)$.
3. Čísla a, b se nazývají *nesoudělná*, jestliže $\gcd(a, b) = 1$.

□

3.1.6 Poznámky.

1. Pro každé celé nezáporné číslo a je $a = \gcd(a, 0)$.
2. Jestliže kladné číslo a dělí b, pak $\gcd(a, b) = a$.
3. Pro každá celá čísla a, b je $\gcd(a, b) = \gcd(-a, b) = \gcd(a, -b) = \gcd(-a, -b)$.

3.1.7 Euklidův algoritmus. Na střední škole jste největší společný dělitel dvou čísel a, b získávali z rozkladu čísel a, b na součin prvočísel. Pro velká čísla je to ale velmi zdlouhavá metoda, protože problém nalezení prvočíselného rozkladu velkého čísla je velmi obtížný. Euklidův algoritmus dává rychlou možnost, jak $\gcd(a, b)$ najít a to bez znalosti prvočíselného rozkladu.

Euklidův algoritmus.

Vstup: přirozená čísla a, b, $b \neq 0$.
Výstup: $d = \gcd(a, b)$.

1. (Inicializace)
   $z := a$, $t := b$;
2. (Výpočet částečného podílu a zbytku.)
   `repeat`
   `do` $z = q \cdot t + r$;
   $z := t$, $t := r$.
   `until` $t = 0$.
3. (Největší společný dělitel)
   $d := z$.

3.1.8 Správnost Euklidova algoritmu vyplývá z následujícího tvrzení.

Tvrzení. Dvojice čísel z, t a dvojice čísel t, r z Euklidova algoritmu mají stejné společné dělitele (tedy i největšího společného dělitele).

Zdůvodnění předchozího tvrzení je jednoduché; stačí si uvědomit, že $r = z - q\,t$ (q je částečný podíl, r zbytek při dělení čísla z). Proto každý společný dělitel čísel z, t je také dělitelem r, tj. společným dělitelem t, r.

Naopak každý společný dělitel čísel t, r je také dělitelem čísla z, protože $z = q\,t + r$.

3.1.9 Bezoutova věta. Následující věta má velký význam při řešení lineárních rovnic, kdy hledáme pouze celočíselná řešení. K důkazu této věty využijeme Euklidův algoritmus, který trochu „rozšíříme“.

Bezoutova věta. Jsou dána přirozená čísla a, b, označme $d = \gcd(a, b)$. Pak existují celá čísla x, y tak, že

$$ax + by = d.$$

□

Důkaz Bezoutovy věty vyplývá z rozšířeného Euklidova algoritmu, který pro každá dvě přirozená čísla a, b zkonstruuje celá čísla x, y z Bezoutovy věty.

3.1.10 Rozšířený Euklidův algoritmus.

Vstup: přirozená čísla a, b, $b \neq 0$.
Výstup: $d = \gcd(a, b)$ a celá čísla x, y taková, že $ax + by = d$.

1. (Inicializace)
   $z := a$, $x_z := 1$, $y_z := 0$, $t := b$, $x_t := 0$, $y_t := 1$;
2. (Výpočet částečného podílu a zbytku)
   `repeat`
   `do` $z = q\,t + r$, $x_r := x_z - q\,x_t$, $y_r := y_z - q\,y_t$;
   $z := t$, $x_z := x_t$, $y_z := y_t$
   $t := r$, $x_t := x_r$, $y_t := y_r$.
   `until` $t = 0$
3. (Největší společný dělitel a čísla x a y)
   $d := z$, $x := x_z$, $y := y_z$.

Zdůvodnění správnosti rozšířeného Euklidova algoritmu je podobné jako v 3.1.7. Platí totiž následující:

Předpokládejme, že platí $z = a\,x_z + b\,y_z$ a $t = a\,x_t + b\,y_t$. Pak

$$r = z - q\,t = a\,x_z + b\,y_z - q\,(a\,x_t + b\,y_t) = a\,(x_z - q\,x_t) + b\,(y_z - q\,y_t).$$

Z předcházející rovnosti je jasné, že čísla x_r a y_r jsou v rozšířeném Euklidově algoritmu určena správně.

3.1.11 Důsledky. 1) Jsou dána dvě nesoudělná čísla a, b. Jestliže číslo a dělí součin $b \cdot c$, pak a dělí číslo c.

2) Jestliže prvočíslo p dělí součin $a \cdot b$, pak p dělí aspoň jedno z čísel a, b. □

Zdůvodnění: Nejprve ukážeme první část tvrzení. Předpokládejme, že čísla a a b jsou nesoudělná. Pak podle Bezoutovy věty existují celá čísla x, y taková, že

$$1 = a\,x + b\,y.$$

Vynásobme tuto rovnici číslem c a dostaneme

$$c = (ac)\,x + (bc)\,y.$$

Číslo a dělí součin ac, dále víme (z předpokladu), že a dělí i součin bc. Proto číslo a dělí i číslo c.

Nyní druhá část tvrzení vyplývá z faktu, že jestliže číslo a je soudělné s prvočíslem p, pak p dělí a.

3.2 Něco o prvočíslech

Pro zvídavé studenty uvádíme ještě několik faktů týkajících se prvočísel. První věta se týká existence rozkladu přirozeného čísla většího než 1 na součin vhodných prvočísel, tzv. prvočíselný rozklad čísla.

3.2.1 Věta. Pro každé přirozené číslo $n > 1$ existují prvočísla $p_1, \ldots, p_k$ a kladná celá čísla $r_1, \ldots, r_k$ tak, že

$$n = p_1^{r_1}\, p_2^{r_2}\, \ldots,\, p_k^{r_k}. \tag{3.1}$$

Je-li navíc $p_1 < p_2 < \ldots < p_k$, pak je rozklad 3.1 jednoznačný. □

Myšlenka důkazu. Existence rozkladu se dá dokázat (silnou) matematickou indukcí. Základní krok je jednoduchý: číslo 2 je prvočíslo, tudíž k z věty je 1 a $p_1 = 2$.

Předpokládejme, že se všechna čísla $1 < i < n$, $n \geq 3$, dají napsat jako součin mocnin prvočísel. Uvažujme číslo n; nyní buď je n prvočíslo a tudíž $k = 1$ a $p_1 = n$, nebo je n složené číslo. V druhém případě platí $n = r\,s$, kde $1 < r, s < n$. Můžeme proto na r i s použít indukční předpoklad, tj. jak r, tak s mají rozklad na součin mocnin prvočísel. Nyní je jednoduché získat z těchto rozkladů i prvočíselný rozklad $n = r\,s$.

Jednoznačnost rozkladu se dokáže pomocí druhého tvrzení v důsledcích 3.1.11.

3.2.2 Uvedeme ještě dvě věty týkající se prvočísel; první z nich říká, že množina všech prvočísel je nekonečná, tudíž musí být spočetná. Ano je to nekonečná podmnožina množiny přirozených čísel. Tato věta má jednoduchý důkaz, který uvedeme.

Naproti tomu druhou větu není jednoduché dokázat a my zde důkaz neuvádíme.

Věta. (Euklid) Prvočísel je nekonečně mnoho. □

Zdůvodnění: Předpokládejme, že existuje jen konečně mnoho prvočísel, řekněme $p_1, p_2, \ldots, p_k$ jsou všechna prvočísla. Uvažujme číslo $N = p_1 p_2 \ldots p_k + 1$. Podle předchozí věty se dá toto číslo napsat jako součin mocnin prvočísel. Ale to není možné, protože N není dělitelné žádným prvočíslem $p_1, \ldots, p_k$. Dostali jsme spor s předpokladem, že $p_1, p_2, \ldots, p_k$ byla všechna prvočísela. Není proto pravda, že by prvočísel bylo konečně mnoho.

Věta. Pro každá dvě nesoudělná čísla a, b, $a \geq 2$, je mezi čísly $an + b$, kde n je přirozené číslo, nekonečně mnoho prvočísel. □

3.3 Diofantické rovnice.

Rozšířený Euklidův algoritmus a Bezoutovu větu použijeme k řešení lineární diofantické rovnice o dvou proměnných. To, že o rovnici s koeficienty z množiny celých čísel prohlásíme, že je diofantická, znamená, že nás nezajímají všechna reálná řešení, ale jenom celočíselná řešení. Budeme se zabývat následující úlohou:

3.3.1 Pro daná celá čísla a, b, c najděte všechna **celá čísla** x, y, která jsou řešením rovnice

$$ax + by = c. \tag{3.2}$$

Je zřejmé, že ne každá rovnice 3.2 má celočíselné řešení. Následující tvrzení charakterizuje rovnice, které celočíselné řešení mají.

Tvrzení. Rovnice 3.2 má aspoň jedno řešení v oboru celých čísel právě tehdy, když číslo c je násobkem největšího společného dělitele $\gcd(a, b)$. □

Zdůvodnění: Označme $d = \gcd(a, b)$. Jestliže číslo c je násobkem d, řekněme: $c = k\,d$, pak stačí najít čísla x', y' z Bezoutovy věty, tj.

$$d = a\,x' + b\,y' \quad \text{a dostáváme} \quad c = kd = a\,(kx') + b\,(ky').$$

Nyní $x := kx'$ a $y := ky'$ je jedno z celočíselných řešení rovnice 3.2. Předpokládejme, že opravdu existují celá čísla x, y taková, že

$$c = a\,x + b\,y.$$

Pak každý společný dělitel čísel a, b také dělí číslo c; mezi jinými i největší společný dělitel $\gcd(a, b)$ dělí číslo c.

3.3.2 Postup řešení diofantických rovnic (3.2). Lineární diofantické rovnice řešíme následujícím způsobem:

1. Pomocí rozšířeného Euklidova algoritmu najdeme jedno řešení x_0, y_0 nehomogenní rovnice 3.2 nebo zjistíme, že rovnice nemá řešení.

2. Řešíme přidruženou homogenní rovnici $a\,x + b\,y = 0$ a to takto: Rovnici nejprve zkrátíme číslem $\gcd(a, b)$; obecné řešení homogenní rovnice $a_1\,x + b_1\,y = 0$, kde a_1 a b_1 jsou nesoudělné, je $x = b_1 k$, $y = -a_1 k$ pro libovolné $k \in \mathbb{Z}$.

3. Obecné řešení rovnice 3.2 je

$$x = x_0 + b_1\,k, \quad y = y_0 - a_1\,k, \quad k \in \mathbb{Z}.$$

Předchozí postup je zdůvodněn dvěma následujícími tvrzeními. První tvrzení je zřejmé a nevyžaduje žádné speciální zdůvodnění, protože vyplývá z důsledku 3.1.11.

3.3.3 Tvrzení. Rovnice $ax + by = 0$ má vždy nekonečně mnoho řešení. Jsou-li a, b nesoudělná, pak každé řešení této rovnice je tvaru $x = kb$, $y = -ka$ pro nějaké celé číslo k. □

3.3.4 Tvrzení. Jestliže číslo c je násobkem největšího společného dělitele čísel a, b, pak každé celočíselné řešení rovnice 3.2 je tvaru

$$x = x_0 + kb_1, \quad y = y_0 - ka_1,$$

kde x_0, y_0 je jedno celočíselné řešení rovnice 3.2, $a_1 = \frac{a}{\gcd(a,b)}$, $b_1 = \frac{b}{\gcd(a,b)}$ a k je libovolné celé číslo.

Zdůvodnění je obdobné tomu, které máte v lineární algebře:

Jsou-li x_1, y_1 a x_2, y_2 dvě řešení nehomogenní rovnice, pak jejich rozdíl $x_1 - x_2, y_1 - y_2$ je řešením přidružené homogenní rovnice.

Dále, je-li x_1, y_1 řešení nehomogenní rovnice a x_0, y_0 řešení přidružené homogenní rovnice, pak jejich součet $x_1 + x_0, y_1 + y_0$ je řešením nehomogenní rovnice.

Na závěr si stačí uvědomit, že je-li x_1, y_1 jedno řešení nehomogenní rovnice, pak libovolné řešení x, y nehomogenní rovnice je rovno $x = x_1 + (x - x_1)$, $y = y_1 + (y - y_1)$.

3.4 Kongruence modulo n

Poznatky o ekvivalencích i o celých číslech využijeme ke konstrukci tzv. kongruence modulo n. Třídy této ekvivalence nám pak umožní zavést nové „počítání“, jiné než s reálnými čísly, a přesto některá nová počítání budou mít všechny vlastnosti, které platí pro reálná čísla.

3.4.1 Definice. Je dáno přirozené číslo $n > 1$. Řekneme, že celá čísla a, b jsou *kongruentní modulo* n, píšeme $a \equiv b \pmod n$, jestliže jejich rozdíl $a - b$ je dělitelný číslem n. □

3.4.2 Jiná charakterizace kongruence modulo n. Někdy je výhodnější používat i jinou charakterizaci dvojic čísel, které jsou v relaci modulo n. Z následujících dvou je užitečná hlavně druhá charakteristika.

Tvrzení. Je dáno přirozené číslo $n > 1$. Pak

1. $a \equiv b \pmod n$ právě tehdy, když $a = b + kn$ pro nějaké celé číslo k;
2. $a \equiv b \pmod n$ právě tehdy, když čísla a i b mají stejné zbytky při dělení číslem n.

□

Zdůvodnění. 1. Je zřejmé, že $a - b$ je dělitelné číslem n, tj. $a - b = kn$ pro nějaké $k \in \mathbb{Z}$, právě tehdy, když $a = b + kn$.

2. Z věty o dělení víme, že $a = q_1\, n + z_1$ a $b = q_2\, n + z_2$, kde $0 \le z_1, z_2 < n$. Jsou-li zbytky při dělení stejné, tj. jestliže se $z_1 = z_2$, pak $a - b = (q_1 - q_2)\, n$ a platí $a \equiv b \pmod n$.

Jsou-li zbytky různé, pak $a - b = (q_1 - q_2)\, n + (z_1 - z_2)$. Přitom $0 < |z_1 - z_2| < n$ a tudíž $a - b$ není dělitelné n, tj. neplatí $a \equiv b \pmod n$.

3.4.3 Relace modulo n je ekvivalence.

Věta. Relace modulo n na množině celých čísel je relace ekvivalence. Přesněji: pro každá celá čísla a, b a c platí

1. $a \equiv a \pmod n$;
2. jestliže $a \equiv b \pmod n$, pak i $b \equiv a \pmod n$;
3. jestliže $a \equiv b \pmod n$ a $b \equiv c \pmod n$, pak $a \equiv c \pmod n$.

□

Zdůvodnění. Fakt, že relace modulo n je reflexivní a symetrická, je zřejmý. K tomu, abychom ověřili tranzitivitu relace, použijeme ekvivalentní podmínku 2 z tvrzení 3.4.2. Jestliže čísla a, b mají stejný zbytek při dělení n, a také čísla b, c mají stejný zbytek při dělení n, pak totéž platí i pro čísla a, c.

3.4.4 Vlastnosti ekvivalence modulo n. Ekvivalence modulo n má ještě další vlastnosti. Ukážeme si ty nejdůležitější v následujícím tvrzení a jeho důsledcích.

Tvrzení. Jestliže pro celá čísla a, b, c a d platí $a \equiv b \pmod n$ a $c \equiv d \pmod n$, pak

$$(a + c) \equiv (b + d) \pmod n \quad \text{a také} \quad (a \cdot c) \equiv (b \cdot d) \pmod n.$$

Zdůvodnění. Předpokládejme, že $a \equiv b \pmod n$ a $c \equiv d \pmod n$. Pak $a = b + kn$ a $c = d + ln$ pro nějaká $k, l \in \mathbb{Z}$. Proto $a + c = b + d + (k + l)n$ a $a \cdot c = (b + kn)(d + ln) = bd + (bl + dk + kln)n$.

Důsledky. Jestliže pro celá čísla a, b platí $a \equiv b \pmod n$, pak platí

1. $ra \equiv rb \pmod n$ pro každé celé číslo r.
2. $a^k \equiv b^k \pmod n$ pro každé přirozené číslo k.
3. Jestliže $a_i \equiv b_i \pmod n$ pro každé $i = 0, \ldots, k$, a jestliže $r_0, \ldots, r_k$ jsou libovolná celá čísla, pak
$$(r_0\, a_0 + \ldots + r_k\, a_k) \equiv (r_0\, b_0 + \ldots + r_k\, b_k) \pmod n.$$

□

Zdůvodnění. 1. K důkazu stačí použít předcházející tvrzení 3.4.4 na dvojici $r \equiv r \pmod n$ a $a \equiv b \pmod n$.

2. Pro $k = 0$ je $a^0 = 1 = b^0$, tedy vlastnost platí; pro $k = 1$ je daná vlastnost předpoklad.

Z předcházejícího tvrzení víme, že $a^2 \equiv b^2 \pmod n$ (použili jsme $a \equiv b \pmod n$ a $a \equiv b \pmod n$). Nyní z $a \equiv b \pmod n$ a $a^2 \equiv b^2 \pmod n$ dostáváme $a^3 \equiv b^3 \pmod n$, obdobně $a^4 \equiv b^4 \pmod n$, atd. (Formální důkaz se vede indukcí a to podle n.)

3. Zde se jedná i opakované použití bodů 1 a 2. Formálně přesně by se tvrzení dokazovalo indukcí podle k.

3.4.5 Aplikace. Výše uvedené důsledky můžeme využít k tomu, abychom dokázali pravdivost testů dělitelnosti 3, 4, 5, 7, 9, 11 a 13. Všechny testy mají stejný základ: jsou založeny na tom, že pro tato čísla umíme najít zbytky při dělení mocnin 10 těmito čísly. Na přednášce testy odvodíme.

3.4.6 Z tvrzení v 3.4.4 víme, že relaci modulo n můžeme „násobit“. S krácením je situace ale složitější, viz následující tvrzení.

Tvrzení. Jestliže $ra \equiv rb \pmod n$ pro nějaká celá čísla r, a, b a přirozené číslo n, pak

$$a \equiv b \left(\mathrm{mod}\, \frac{n}{\gcd(n,r)}\right).$$

□

Zdůvodnění. Víme, že $ra - rb = kn$ pro nějaké celé číslo k. Odtud $r(a-b) = kn$. Označme $d = \gcd(r,n)$. Dostáváme $r = sd$, $n = md$, navíc jsou čísla s a m nesoudělná. Dosadíme do předchozí rovnosti a máme

$$sd(a-b) = kmd \quad \text{a po krácení} \quad s(a-b) = km.$$

Protože jsou čísla s a m nesoudělná a s dělí součin $k\,m$, musí s dělit k. Odtud $s(a-b) = slm$ (pro nějaké $l \in \mathbb{Z}$) a $a - b = lm$. Ukázali jsme, že $a \equiv b \pmod m$, tj. $a \equiv b \pmod{\frac{n}{\gcd(n,r)}}$.

3.4.7 Další otázka, kterou si položíme, je, zda je možné „řešit“ některé rovnice modulo n. Jde-li pouze o sčítání, je situace jednodušší, chceme-li řešit rovnice s násobením, potřebujeme řešit diofantické rovnice.

1. Jsou dána celá čísla a, b a přirozené číslo $n > 1$. Máme najít všechna celá čísla x, pro která platí

$$(a + x) \equiv b \pmod n. \tag{3.3}$$

Tato úloha má vždy jediné řešení a to $x \equiv (b - a) \pmod n$.

2. Jsou dána celá čísla a, b a přirozené číslo $n > 1$. Máme najít všechna celá čísla x, pro která platí

$$a\,x \equiv b \pmod n. \tag{3.4}$$

Tato úloha nemusí mít vždy řešení. Např. neexistuje celé číslo x, pro které $2x \equiv 3 \pmod 4$. Následující tvrzení říká za jakých podmínek řešení existuje.

3.4.8 Tvrzení. Rovnice 3.4 má řešení právě tehdy, když číslo b je násobkem největšího společného dělitele $\gcd(a,n)$. V takovém případě najdeme všechna čísla x jako řešení diofantické rovnice

$$ax + ny = b. \tag{3.5}$$

□

Zdůvodnění. Ano, $ax \equiv b \pmod n$ znamená $ax - b = kn$ pro nějaké celé číslo k. Nejprve přepíšeme tuto rovnici na $ax - kn = b$ a pak provedeme substituci $y = -k$. Dostali jsme požadovanou diofantickou rovnici 3.5.

3.4.9 Uvedeme ještě jedno tvrzení týkající se relací modulo n.

Tvrzení. Předpokládejme, že pro dvě celá čísla a, b platí $a \equiv b \pmod n$ a $a \equiv b \pmod m$, $m, n > 1$. Pak pro nesoudělná čísla m, n platí i $a \equiv b \pmod{nm}$.

Zdůvodnění. Víme, že $a - b = kn$ a $a - b = lm$ pro nějaká čísla $k, l \in \mathbb{Z}$. Odtud dostáváme $kn = lm$ pro nějaká $k, l \in \mathbb{Z}$. Čísla n a m jsou nesoudělná a n dělí součin lm. Proto n dělí l a $a - b = r\,(nm)$. Ukázali jsem $a \equiv b \pmod{nm}$.

3.4.10 Eulerova funkce. V dalším textu využijeme Eulerovu funkci, která pro dané přirozené číslo $n > 1$ udává počet kladných čísel nesoudělných s n a menších než n.

Definice. Je dáno přirozené číslo $n > 1$. Pak hodnota Eulerovy funkce $\phi(n)$ je rovna počtu všech přirozených čísel i, $0 \leq i < n$, která jsou nesoudělná s číslem n. Tj.

$$\phi(n) = |\{i \,|\, 0 < i < n, \gcd(i, n) = 1\}|.$$

□

Tedy např. $\phi(6) = 2$, protože mezi 0 a 6 jsou pouze dvě čísla nesoudělná s číslem 6 a to 1 a 5.

3.4.11 Vlastnosti Eulerovy funkce. Následující tvrzení nám pomáhá určit hodnotu Eulerovy funkce, ovšem pouze za předpokladu, že známe prvočíselný rozklad čísla.

Tvrzení.

1. Je-li p prvočíslo, pak $\phi(p) = p - 1$.
2. Je-li $n = p^k$ pro nějaké prvočíslo p, pak $\phi(n) = p^k - p^{k-1}$.
3. Jestliže n a m jsou nesoudělná přirozená čísla, pak $\phi(nm) = \phi(n) \cdot \phi(m)$.

□

Zdůvodnění. 1 a 2. Je-li p prvočíslo, pak všechna čísla $1, 2, \ldots, p-1$ jsou nesoudělná s p. Soudělná čísla mezi 0 a $p^k - 1$ jsou všechna čísla dělitelná p. Vydělíme-li každé z nich číslem p, dostaneme všechna čísla mezi 0 a p^{k-1}. Proto $\phi(p^k) = p^k - p^{k-1}$.

Zdůvodnění vlastnosti 3 je obtížnější. Nejjednodušší je použít k důkazu čínskou větu o zbytcích, o které budeme mluvit později.

Dříve než ukážeme šifrovací systém RSA, uvedeme tzv. malou Fermatovu větu, která je základem tohoto šifrovacího systému.

3.4.12 Malá Fermatova věta. Přestože tato věta má i elementární důkaz, neuvádíme ho; věta je jednoduchým důsledkem vlastností konečných grup a dokážeme ji až budeme probírat vlastnosti konečných grup.

Věta. Jestliže číslo a je nesoudělné s prvočíslem p, pak

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

□

Poznámka. Z malé Fermatovy věty vyplývá toto: Je-li p prvočíslo, pak pro každé celé číslo a platí

$$a^p \equiv a \pmod{p}$$

Je-li totiž a soudělné s prvočíslem p, pak je a dělitelné p a tudíž $a \equiv 0 \pmod{p}$. Proto $a^p \equiv 0 \equiv a \pmod{p}$. Pro nesoudělná čísla a stačí obě strany ekvivalence vynásobit číslem a.

Malá Fermatova věta je základem šifrovacího systému RSA, jednoho z tzv. šifrovacích systémů s veřejným klíčem.

3.4.13 Aplikace — RSA. Alice chce od Boba (či někoho jiného) dostávat tajné zprávy. Alice si zvolí dvě velká prvočísla p a q (za velká se zde považují čísla aspoň patnáctimístná) a spočítá jejich součin $N = pq$.

Dále **Alice** zvolí číslo e_A, které je nesoudělné s $\phi(N) = (p-1)(q-1)$. K číslu e_A najde číslo d_A tak, aby

$$e_A \cdot d_A \equiv 1 \pmod{\phi(N)}.$$

Takové číslo existuje, viz. 3.4.8. Číslo e_A se nazývá *šifrovací index* a číslo d_A se nazývá *dešifrovací index*.

Alice zveřejní: N a e_A; **Alice utají**: d_A, tedy i $\phi(N)$ a proto p, q.

Bob chce poslat Alici číslo x, $x < N$. Bob spočítá

$$x^{e_A} \equiv y \pmod N,\ 0 \le y < N,$$

a pošle Alici číslo y. Alice přijme číslo y a spočítá

$$y^{d_A} \equiv z \pmod N,\ 0 \le z < N.$$

Věta.

$$x = z.$$

□

Myšlenka důkazu správnosti RSA. Víme, že

$$z \equiv y^{d_A} \pmod N, \quad y \equiv x^{e_A} \pmod N.$$

Proto

$$z \equiv x^{e_A \cdot d_A} \pmod N.$$

Navíc $e_A \cdot d_A = 1 + k\,(p-1)\,(q-1)$ pro nějaké celé číslo $k \in \mathbb{Z}$.

Nejprve ukážeme, že $z \equiv x \pmod p$. K tomu stačí si uvědomit, že pro každé přirozené číslo $s > 0$ máme

$$x^{1+s(p-1)} \equiv x^{1+p-1+(s-1)(p-1)} \equiv x^p\, x^{(s-1)(p-1)} \equiv x\, x^{(s-1)(p-1)} \equiv x^{1+(s-1)(p-1)} \pmod p.$$

(Využili jsme poznámku za malou Fermatovou větou, speciálně $x^p \equiv x \pmod p$.) Nyní

$$x^{e_A d_A} = x^{1+(k(q-1))(p-1)} = x^{1+s(p-1)} = \ldots = x^{1+(p-1)} \equiv x \pmod p.$$

Obdobně se dá ukázat, že $z \equiv x \pmod q$.

Protože p a q jsou nesoudělná (jsou to různá prvočísla), dostáváme s využitím tvrzení 3.4.9

$$z \equiv x \pmod N, \quad \text{kde} \quad N = p \cdot q.$$

Nyní si stačí uvědomit, že $0 \le x, z < N$, a proto

$$x = z.$$

3.5 Čínská věta o zbytcích

V minulém odstavci jsme řešili rovnice, které obsahovaly kongruenci modulo n. Nyní nás čeká trochu jiný úkol: Najít celé číslo takové, že má předepsané zbytky při dělení různými přirozenými čísly většími než 1. Čínská věta o zbytcích ukazuje případ, kdy takové číslo **vždy** existuje.

3.5.1 Čínská věta o zbytcích. Jsou dána přirozená čísla $m_1, m_2, \ldots, m_k$ po dvou nesoudělná. Pak pro libovolná celá čísla $a_1, a_2, \ldots, a_k$ existuje celé číslo x takové, že

$$\begin{aligned} x &\equiv a_1 \pmod{m_1} \\ x &\equiv a_2 \pmod{m_2} \\ &\ldots \\ x &\equiv a_k \pmod{m_k} \end{aligned} \tag{3.6}$$

Navíc, jestliže nějaké celé číslo y také splňuje 3.6, pak nutně $x \equiv y \pmod N$, kde N je součin všech čísel m_i, tj. $N = m_1 m_2 \ldots m_k$. □

Myšlenka důkazu — nalezení x splňující 3.6. K důkazu stačí, abychom našli celá čísla q_1, q_2, …, q_k taková, že pro každé $i = 1, 2, \ldots, k$ platí

$$q_i \equiv 1 \pmod{m_i} \ \text{ a } \ q_i \equiv 0 \pmod{m_j} \ \text{ pro } j \ne i. \tag{3.7}$$

Pak námi hledané číslo x bude mít tvar:

$$x = a_1\, q_1 + a_2\, q_2 + \ldots + a_k\, q_k.$$

Čísla q_1, q_2, …, q_k nalezneme např. takto: Utvoříme číslo M_i, které je součinem všech m_j, $j \neq i$. Protože jsou čísla m_1, m_2, …, m_k po dvou nesoudělná, jsou nesoudělná i čísla m_i a M_i. Proto (podle Bezoutovy věty) existují celá čísla t_1 a t_2 taková, že

$$m_i t_1 + M_i t_2 = 1.$$

Nyní číslo $q_i = M_i t_2$ má požadované vlastnosti. Ano, $M_i t_2$ je dělitelné všemi m_j pro $j \neq i$ a $M_i t_2 \equiv 1 \,(\mathrm{mod}\, m_i)$.

3.5.2 Poznámka. Číslo x z 3.6 může pro některé $a_1, \ldots, a_k$ existovat i v případě, že čísla $m_1, \ldots, m_k$ nejsou po dvou nesoudělná. V takovém případě ale x neexistuje **pro každé** $a_1, \ldots, a_k$.

3.6 Zbytkové třídy

Nyní již můžeme zavést „nové počítání“.

3.6.1 Zbytkové třídy modulo n Víme, že relace modulo n je relace ekvivalence na množině celých čísel. Novými „čísly“ v novém počítání budou třídy ekvivalence modulo n.

Definice. *Zbytková třída modulo n obsahující číslo $i \in \mathbb{Z}$* je

$$[i]_n = \{j \mid j = i + kn \text{ pro nějaké } k \in \mathbb{Z}\}. \tag{3.8}$$

Množinu všech různých zbytkových tříd značíme $\mathbb{Z}_n$, je tedy

$$\mathbb{Z}_n = \{[0]_n, [1]_n, \ldots, [n-1]_n\}.$$

□

Poznamenejme, že množiny $[i]_n$ se nazývají zbytkové třídy proto, že do $[i]_n$ patří všechna celá čísla, která mají stejný zbytek při dělení číslem n jako číslo i. Také si uvědomte, že množina $\mathbb{Z}_n$ je vlastně faktorová množina celých čísel podle ekvivalence $\mathrm{mod}\, n$.

3.6.2 Počítání se zbytkovými třídami. Z tvrzení 3.4.4 víme, že ekvivalence modulo n „zachovává“ operace sčítání a násobení. Vlastnosti, které jsme uvedli, se dají přeformulovat:

Jestliže vybereme libovolné číslo a z $[i]_n$ a libovolné číslo b z $[j]_n$, pak číslo $a+b$ leží v $[i+j]_n$ a číslo $a \cdot b$ leží v $[i \cdot j]_n$. To nám umožňuje definovat na množině zbytkových tříd $\mathbb{Z}_n$ operace sčítání $\oplus$ a násobení $\odot$ takto:

$$[i]_n \oplus [j]_n = [i+j]_n, \qquad [i]_n \odot [j]_n = [i \cdot j]_n. \tag{3.9}$$

3.6.3 Vlastnosti operací sčítání $\oplus$ a násobení $\odot$. Následující vlastnosti jednoduše vyplývají z definice obou operací se zbytkovými třídami

Tvrzení.

1. Sčítání $\oplus$ je asociativní, tj. pro každá tři celá čísla i, j, k platí:

$$([i]_n \oplus [j]_n) \oplus [k]_n = [i]_n \oplus ([j]_n \oplus [k]_n).$$

2. Sčítání $\oplus$ je komutativní, tj. pro každá dvě celá čísla i, j platí:

$$[i]_n \oplus [j]_n = [j]_n \oplus [i]_n.$$

3. Třída $[0]_n$ hraje roli “nuly”, tj. pro každé celé číslo i platí:

$$[0]_n \oplus [i]_n = [i]_n.$$

4. Můžeme odčítat, přesněji pro každou třídu $[i]_n$ existuje třída $-[i]_n$ taková, že
$$[i]_n \oplus (-[i]_n) = [0]_n.$$

□

Uvědomte si, že třída $-[i]_n$ je třída $[-i]_n = [n-i]_n$. Ano, $[i]_n \oplus [n-i]_n = [n]_n = [0]_n$.

Tvrzení.

1. Násobení $\odot$ je asociativní, tj. pro každá tři celá čísla i, j, k platí:
$$([i]_n \odot [j]_n) \odot [k]_n = [i]_n \odot ([j]_n \odot [k]_n).$$
2. Násobení $\odot$ je komutativní, tj. pro každá dvě celá čísla i, j platí:
$$[i]_n \odot [j]_n = [j]_n \odot [i]_n.$$
3. Třída $[1]_n$ hraje roli "jedničky", tj. pro každé celé číslo i platí:
$$[1]_n \odot [i]_n = [i]_n.$$

□

Navíc mezi sčítáním $\oplus$ a násobením $\odot$ platí distributivní zákon.

Tvrzení. Pro každá tři celá čísla i, j, k platí
$$[i]_n \odot ([j]_n \oplus [k]_n) = [i]_n \odot [j]_n \oplus [i]_n \odot [k]_n.$$

□

3.6.4 Poznámka. Mezi vlastnostmi operace $\odot$ nebylo „krácení", tj. neuvedli jsme za jakých podmínek pro dané celé číslo i existuje celé číslo j takové, že $[i]_n \odot [j]_n = [1]_n$. To je proto, že ne každá rovnice tvaru $[i]_n \odot [x]_n = [1]_n$ má řešení.

Tvrzení. Pro zbytkové třídy $[i]_n$, $[j]_n$, $0 \le i, j < n$, existuje zbytková třída $[x]_n$ splňující rovnici
$$[i]_n \odot [x]_n = [j]_n \tag{3.10}$$
právě tehdy, když j je násobek největšího společného dělitele $d = \gcd(i, n)$. V takovém případě má rovnice 3.10 d různých řešení.

Speciálně: Ke zbytkové třídě $[i]_n$ existuje zbytková třída $[x]_n$ taková, že
$$[i]_n \odot [x]_n = [1]_n \tag{3.11}$$
právě tehdy, když i a n jsou nesoudělná čísla a tato třída je jediná. □

Zdůvodnění. Rovnici 3.10 můžeme přeformulovat na $[ix]_n = [j]_n$ a tedy $ix = j + kn, k \in \mathbb{Z}$. Nyní je zřejmé, že se jedná o diofantickou rovnici $ix + ny = j$. Proto třída $[x]_n$ existuje právě tehdy, když j je násobek $\gcd(i, n)$. Z vlastností řešení diofantických rovnic vyplývá, že takových různých řešení (různých tříd) je $d = \gcd(i, n)$.

3.6.5 Poznámka. Je-li p prvočíslo, pak množina $\mathbb{Z}_p$ splňuje všechny vlastnosti, na které jsme zvyklí z počítání s reálnými čísly, tj. vlastnosti, které má sčítání a násobení reálných čísel. Jestliže n není prvočíslo, pak už je situace složitější. Na příklad, jestliže $n = rs$, $0 < r < n$ a $0 < s < n$, pak $[r]_n \odot [s]_n = [0]_n$ a přitom třídy $[r]_n$ a $[s]_n$ jsou obě nenulové. (Znamená to tedy, že takovým číslem $[r]_n$ nebo $[s]_n$ nemůžeme krátit.)

3.6.6 Konvence. V dalším textu zjednodušíme značení zbytkových tříd: budeme psát $\mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ místo zdlouhavého $\mathbb{Z}_n = \{[0]_n, [1]_n, \ldots, [n-1]_n\}$; a dále budeme psát $+$ a $\cdot$ místo $\oplus$ a $\odot$. Přitom budeme mít na paměti, že mluvíme-li o sčítání a násobení v $\mathbb{Z}_n$ jedná se vždy o operace se zbytkovými třídami a ne operace s celými čísly.

Uvědomte si, že při tomto značení operací v $\mathbb{Z}_n$ platí: pro každé $i, j \in \mathbb{Z}_n$

$i + j = k$, kde k je zbytek při dělení součtu čísel i a j číslem n;

$i \cdot j = l$, kde l je zbytek při dělení součinu čísel i a j číslem n.

Kapitola 4

Binární operace

4.1 Grupoidy, pologrupy, monoidy

V minulé kapitole jsme zavedli sčítání a násobení se zbytkovými třídami modulo n. Nyní se zaměříme na společné vlastnosti obecných „počítání“.

4.1.1 Grupoid.

Definice. *Binární operace na množině S* je zobrazení množiny všech uspořádaných dvojic $S \times S$ do množiny S. (To znamená, že binární operace musí **každým** prvkům a, b přiřadit výsledek.)

Dvojice $(S, \circ)$, kde S je množina a $\circ$ je binární operace na S, se nazývá *grupoid*. □

Binární operace značíme $\cdot$, nebo $+$, $\circ$, $*$ atd. (Prvkům x, y binární operace $\circ$ přiřazuje prvek $x \circ y$.)

Příklady grupoidů.

1. Sčítání na množině všech reálných čísel.
2. Sčítání na množině všech celých čísel.
3. Sčítání na množině všech kladných přirozených čísel.
4. Násobení na množině všech reálných čísel.
5. Násobení na množině všech celých čísel.
6. Násobení matic na množině všech reálných čtvercových matic.
7. Sčítání na množině $\mathbb{Z}_n$.
8. Násobení na množině $\mathbb{Z}_n$.
9. Odčítání na množině celých čísel.

Odčítání **není** binární operace na množině přirozených čísel, protože např. výsledek $3 - 4$ není přirozené číslo.

4.1.2 Neutrální prvek. Grupoid může obsahovat prvek, kterým se chová „neutrálně“, když s ním provádíme operaci. Takovému prvku budeme říkat neutrální a ukážeme, že v grupoidu může existovat nejvýše jeden.

Definice. Máme dán grupoid $(S, \circ)$. Prvek $e \in S$ se nazývá *neutrálním* (též *jednotkovým*, nebo *nulovým*), jestliže

$$e \circ x = x = x \circ e \quad \text{pro každé } x \in S. \tag{4.1}$$

□

Jestliže se operace značí $+$ a říkáme jí sčítání, mluvíme o neutrálním prvku. Značíme-li operaci podobně jako násobení, najdete v literatuře většinou termín jednotkový prvek. My o jednotkovém prvku budeme mluvit v případě, že na jedné množině budeme mít definovány dvě operace – sčítání

(s neutrálním či nulovým prvkem) a násobení (s jednotkovým prvkem). To proto, abychom oba prvky odlišili.

Příklady neutrálních prvků.

1. Sčítání na množině všech reálných čísel stejně jako na množině celých čísel má 0 jako neutrální prvek.
2. Násobení na množině všech reálných čísel stejně jako na množině všech celých čísel má 1 jako neutrální prvek.
3. Násobení matic na množině všech reálných čtvercových matic má jednotkovou matici jako svůj neutrální prvek.
4. Sčítání na množině $\mathbb{Z}_n$ má třídu $[0]_n$ jako neutrální prvek.
5. Násobení na množině $\mathbb{Z}_n$ má třídu $[1]_n$ jako neutrální prvvek.
6. Skládání zobrazení na množině všech zobrazení množiny X do sebe má identické zobrazení jako neutrální prvek.

Sčítání na množině všech kladných přirozených čísel **nemá** neutrální prvek. Neexistuje totiž kladné přirozené číslo e takové, že $n + e = n = e + n$.

4.1.3 Tvrzení. Jestliže v grupoidu $(S, \circ)$ existují prvky e a f takové, že pro každé $x \in S$ platí $e \circ x = x$ a $x \circ f = x$, pak $e = f$ a e je neutrální prvek.

Zdůvodnění. Stačí se podívat na $e \circ f$. Z vlastnosti prvku e je $e \circ f = f$, z vlastnosti prvku f je $e \circ f = e$. Tedy $e = f$.

Poznamenejme, že předchozí tvrzení také říká, že jestliže v grupoidu existuje neutrální prvek, pak je jediný.

4.1.4 Pologrupy. Některé binární operace mohou splňovat vlastnosti, které umožňují jednodušší „počítání“. Základní výhodou při počítání je „nemuset závorkovat“.

Definice. Máme dánu binární operaci $\circ$ na množině S. Jestliže pro každé $x, y, z \in S$ platí

$$x \circ (y \circ z) = (x \circ y) \circ z \tag{4.2}$$

nazývá se množina S spolu s binární operací $\circ$ *pologrupa*. □

Vlastnosti 4.2 říkáme asociativní zákon. Pologrupa je tedy každý grupoid, který splňuje asociativní zákon.

Příklady pologrup.

1. Sčítání na množině všech reálných čísel.
2. Sčítání na množině všech celých čísel.
3. Sčítání na množině všech kladných přirozených čísel.
4. Násobení na množině všech reálných čísel.
5. Násobení na množině všech celých čísel.
6. Násobení matic na množině všech reálných čtvercových matic.
7. Násobení na množině $\mathbb{Z}_n$.
8. Skládání zobrazení na množině všech zobrazení množiny X do sebe.
9. Zřetězení na množině všech binárních slov.

Příkladem grupoidu, který **není** pologrupou, je např. odčítání na množině všech celých čísel. Ano, např. $(3 - 4) - 1 = -2$, kdežto $3 - (4 - 1) = 0$.

4.1.5 Monoidy. Pro nás má neutrální prvek význam hlavně v grupoidech, které splňují asociativní zákon; tedy v pologrupách. Pologrupám s neutrálním prvkem říkáme monoidy.

Definice. *Monoid* je pologrupa, která obsahuje neutrální prvek. □

V dalším textu napíšeme-li $(S, \circ, e)$ je monoid, je e neutrální prvek pologrupy $(S, \circ)$.

Zvláštní příklady monoidů neuvádíme, student příklady najde mezi příklady pologrup sám.

4.1.6 Invertibilní prvek. Viděli jsme v minulých kapitolách, že pro řešení rovnic je potřeba mít možnost „krátit“. Prvky, kterými lze krátit se nazývají invertibilní prvky. Formálně:

Definice. Máme dán monoid $(S, \circ, e)$. Řekneme, že prvek $x \in S$ je *invertibilní*, jestliže existuje prvek $y \in S$ takový, že

$$x \circ y = e = y \circ x. \tag{4.3}$$

Prvek y, pro který platí 4.3, se nazýváme *inverzní* prvek k prvku x a značíme ho x^{-1}. □

Fakt, že prvek y, pro který platí 4.3, je jediný, musíme teprve dokázat. To nám říká následující tvrzení.

4.1.7 Tvrzení. Jestliže v monoidu $(S, \circ, e)$ existují prvky x, y_1, y_2 takové, že

$$y_1 \circ x = e \quad \text{a} \quad x \circ y_2 = e,$$

pak $y_1 = y_2$ a jedná se o inverzní prvek k prvku x. □

Zdůvodnění. Uvažujme prvek $y_1 \circ x \circ y_2$. Platí

$$y_2 = e \circ y_2 = (y_1 \circ x) \circ y_2 = y_1 \circ (x \circ y_2) = y_1 \circ e = y_1.$$

4.1.8 Poznámka. Někdy se místo termínu inverzní prvek používá termín *opačný* prvek. Je to tehdy, jestliže se operace značí + a říkáme jí sčítání, v tomto případě opačný prvek značíme nejčastěji $-x$. Značíme-li operaci podobně jako násobení, mluvíme o inverzním prvku a značíme ho x^{-1}.

Pro inverzní prvky platí následující tvrzení.

4.1.9 Tvrzení. Je dán monoid $(S, \circ, e)$. Pak

1. e je invertibilní prvek a $e^{-1} = e$.
2. Je-li x invertibilní prvek, pak x^{-1} je také invertibilní prvek a platí $(x^{-1})^{-1} = x$.
3. Jestliže x a y jsou invertibilní prvky, pak také $x \circ y$ je invertibilní prvek a platí $(x \circ y)^{-1} = y^{-1} \circ x^{-1}$.

□

Zdůvodnění všech tří tvrzení je jednoduché. Ukážeme pouze tvrzení 3. Platí

$$(x \circ y) \circ (y^{-1} \circ x^{-1}) = x \circ (y \circ y^{-1}) \circ x^{-1} = x \circ x^{-1} = e,$$

a

$$(y^{-1} \circ x^{-1}) \circ (x \circ y) = y^{-1} \circ (x^{-1} \circ x) \circ y = y^{-1} \circ y = e.$$

Tedy $(x \circ y)^{-1} = y^{-1} \circ x^{-1}$.

4.1.10 Krácení. Ukážeme, že v monoidu každým invertibilním prvkem můžeme krátit.

Tvrzení. Je dán monoid $(S, \circ, e)$. Jestliže prvek $a \in S$ je invertibilní a navíc $a \circ b = a \circ c$ nebo $b \circ a = c \circ a$, pak $b = c$. □

Zdůvodnění. Je-li prvek $a \in S$ invertibilní, existuje k němu inverzní prvek a^{-1}. Rovnost $a \circ b = a \circ c$ vynásobíme prvkem a^{-1} zleva a dostaneme

$$a^{-1} \circ (a \circ b) = a^{-1} \circ (a \circ c) \quad \text{a tedy} \quad (a^{-1} \circ a) \circ b = (a^{-1} \circ a) \circ c.$$

Protože $a^{-1} \circ a = e$, dostáváme $b = c$.

Analogicky postupujeme v případě $b \circ a = c \circ a$, pouze násobíme rovnici inverzním prvkem zprava.

4.1.11 Grupy. Ukázali jsme si, že invertibilní prvky jsou ty prvky, kterými můžeme krátit. Monoidy, ve kterých jsou všechny prvky invertibilní, hrají důležitou roli a nazývají se grupy.

Definice. Monoid $(S, \circ, e)$, ve kterém je každý prvek invertibilní, se nazývá *grupa*. □

Příklady grup.

1. Reálná čísla spolu se sčítáním a 0, $(\mathbb{R}, +, 0)$, tvoří grupu. Ano, ke každému reálnému číslu x existuje prvek $-x$ pro nějž $x + (-x) = 0 = (-x) + x$.
2. Celá čísla spolu se sčítáním a 0, $(\mathbb{Z}, +, 0)$, tvoří grupu. Ano, ke každému celému číslu x existuje číslo $-x$ pro nějž $x + (-x) = 0 = (-x) + x$.
3. Kladná reálná čísla spolu s násobením a 1 $(\mathbb{R}^+, \cdot, 1)$ tvoří grupu. Ano, ke každému kladnému číslu x existuje kladné číslo $\frac{1}{x}$, pro nějž $x \cdot \frac{1}{x} = 1 = \frac{1}{x} \cdot x$.
4. Množina $\mathbb{Z}_n$ spolu se sčítáním a $[0]_n$ tvoří grupu. Ano, k prvku $[i]_n$ existuje prvek $[n-i]_n$, pro který $[i]_n + [n-i]_n = [0]_n = [n-i]_n + [i]_n$.
5. Množina všech permutací množiny $\{1, 2, \ldots, n\}$ spolu se skládáním a identickou permutací tvoří grupu. Ano, ke každé permutaci f existuje inverzní permutace f^{-1} a ta je inverzním prvkem k f.

Příklady monoidů, které nejsou grupami.

1. Celá čísla spolu s násobením. Jedná se sice o monoid, kde neutrální prvek je 1, ale např. číslo 2 nemá inverzní prvek.
2. Množina $\mathbb{Z}_n$ spolu se násobením. Jedná se sice o monoid, kde $[1]_n$ je neutrální prvek, ale k prvku $[0]_n$ neexistuje inverzní prvek.
3. Množina všech zobrazení množiny $\{1, 2, \ldots, n\}$ pro $n > 1$ spolu se skládáním zobrazení. Jedná se sice o monoid, kde e je identické zobrazení, ale zobrazení, které není prosté, nemá inverzní prvek.

4.1.12 Následující tvrzení říká, že v každé grupě je možné řešit všechny rovnice typu $a \circ x = b$ a $y \circ a = b$ a řešení jsou jediná. V obecném monoidu by obdobné tvrzení platilo pouze v případě, že a je invertibilní prvek.

Tvrzení. Je dána grupa $(S, \circ)$ s neutrálním prvkem e. Pak pro každé dva prvky $a, b \in S$ existují prvky $x, y \in S$ tak, že

$$a \circ x = b, \qquad y \circ a = b.$$

Tyto prvky jsou určeny jednoznačně. □

Zdůvodnění. Nejprve dokážeme, že takové x existuje. Protože pracujeme v grupě a $a \in S$, existuje inverzní prvek a^{-1}. Vynásobíme rovnici $a \circ x = b$ prvkem a^{-1} zleva, dostáváme

$$x = (a^{-1} \circ a) \circ x = a^{-1} \circ (a \circ x) = a^{-1} \circ b.$$

Obdobně dostáváme $y = b \circ a^{-1}$ z druhé rovnice vynásobením prvkem a^{-1} tentokrát zprava.

Nyní ukážeme, že takový prvek x je jediný. Předpokládejme, že $a \circ x_1 = b$ a $a \circ x_2 = b$. Pak $a \circ x_1 = a \circ x_2$. Jsme v grupě, proto můžeme prvek a zkrátit zleva; přesněji rovnici vynásobit prvkem a^{-1} zleva, a dostáváme $x_1 = x_2$. Obdobně se ukáže, že i řešení rovnice $y \circ a = b$ je jediné.

4.1.13 Z předchozího odstavce víme, že v grupě má každá rovnice jednoznačné řešení. Následující věta ukazuje, že existence řešení každé rovnice už zaručuje, že pologrupa je grupa. Přesněji:

Věta. Pologrupa $(S, \circ)$ je grupa právě tehdy, když každá rovnice tvaru $a \circ x = b$ a každá rovnice tvaru $y \circ a = b$ má řešení.

Přesněji: Pologrupa $(S, \circ)$ je grupa právě tehdy, když pro každé dva prvky $a, b \in S$ existují prvky $x, y \in S$ takové, že $a \circ x = b$ a $y \circ a = b$. □

Zdůvodnění. a) Jestliže je pologrupa $(S, \circ)$ grupa, pak podle tvrzení 4.1.12 víme, že každá rovnice řešení má a to jediné.

b) Ukážeme nejprve, že existence všech řešení rovnic zajišťuje, že v $(S, \circ)$ existuje neutrální prvek. Vyberme libovolně prvek $a \in S$. Protože rovnice $a \circ x = a$ a $y \circ a = a$ mají řešení, existují prvky $e_a, f_a \in S$ takové, že $a \circ e_a = a$ a $f_a \circ a = a$. Ukážeme, že $b \circ e_a = b$ a $f_a \circ b = b$ pro každé $b \in S$. Tím bude dokázáno, že $e_a = f_a$ je neutrální prvek pologrupy $(S, \circ)$ (viz 4.1.3).

Vezměme libovolný prvek $b \in S$. Pak

$$b \circ e_a = (y \circ a) \circ e_a = y \circ (a \circ e_a) = y \circ a = b.$$

První a poslední rovnost jsme dostali z faktu, že $b = a \circ y$ pro vhodné $y \in S$. Obdobně se dokáže, že $b = f_a \circ b$. Proto $e_a = f_a$ je neutrální prvek; označme ho pro jednoduchost e.

Nyní už je lehké ukázat, že každý prvek $a \in S$ je invertibilní. Víme, že existují prvky $x, y \in S$, pro které $a \circ x = e$ a $y \circ a = e$. Tedy podle tvrzení 4.1.7 platí $x = y$, a x je inverzní prvek v prvku a.

Proto je $(S, \circ)$ grupa.

4.1.14 Podpologrupy. Známe příklady pologrup, monoidů, grup, které jsou podmnožinami jiných; stačí uvést příklad celých čísel se sčítáním a množinu přirozených čísel, nebo množinu všech kladných přirozených čísel. Student sám najde řadu dalších příkladů. V tomto odstavci a dalších se zabýváme společnými vlastnostmi, které takové příklady mají.

Definice. Máme dánu pologrupu $(S, \circ)$. Podmnožina $T \subseteq S$ spolu s operací $\circ$ se nazývá *podpologrupa* pologrupy $(S, \circ)$ jestliže pro každé dva prvky $x, y \in T$ je $x \circ y \in T$. (V tomto případě množina $(T, \circ)$ je také pologrupa.) □

Příklady podpologrup.

1. Množina přirozených čísel $\mathbb{N}$ spolu se sčítáním je podpologrupou $(\mathbb{Z}, +)$.
2. Množina všech regulárních čtvercových matic spolu s násobením matic je podpologrupou $(M_n, \cdot)$, kde M_n je množina všech čtvercových matic.
3. Množina všech kladných reálných čísel spolu s násobením je podpologrupou $(\mathbb{R}, \cdot)$.

Množina všech regulárních čtvercových matic spolu se sčítáním matic **není** podpologrupou $(M_n, +)$, protože součtem dvou regulárních matic můžeme dostat matici, která je singulární. Ano, např. $E + (-E) = O$, (zde E je jednotková matice).

4.1.15 Podmonoid.

Definice. Máme dán monoid $(S, \circ, e)$. Podpologrupa, která obsahuje neutrální prvek e, se nazývá *podmonoid* monoidu $(S, \circ, e)$. □

Poznámka. Uvědomte si, že k tomu, aby podpologrupa $(T, \circ)$ byla podmonoidem nestačí, aby v ní existoval neutrální prvek; neutrální prvek $(T, \circ)$ musí být stejný jako byl ve „velkém“ monoidu $(S, \circ, e)$, tj. musí být e.

Příklady podmonoidů.

1. Množina přirozených čísel $\mathbb{N}$ spolu se sčítáním tvoří podmonoid $(\mathbb{Z}, +)$, protože $0 \in \mathbb{N}$.
2. Množina všech regulárních čtvercových matic spolu s násobením matic tvoří podmonoid $(M_n, \cdot)$, protože jednotková matice je regulární.
3. Označme T_X množinu všech zobrazení množiny X do sebe. Pak $(T_X, \circ)$, kde $\circ$ je skládání zobrazení, je monoid (identické zobrazení je jeho neutrální prvek). Množina všech permutací na množině X je podmonoid $(T_X, \circ)$. Ano, identické zobrazení je také permutace.

4.1.16 Komutativní pologrupy, monoidy, grupy. V mnoha příkladech, které jsme uvedli, binární operace měly ještě jednu vlastnost — nezáleželo na pořadí, ve kterém jsme operaci prováděli. Takové operace umožňují „jednodušší“ počítání a my se jim věnujeme.

Definice. Pologrupa $(S, \circ)$ (monoid, grupa) se nazývá *komutativní pologrupa* (*komutativní monoid*, *komutativní grupa*), platí-li navíc *komutativní zákon*, tj. pro každé dva prvky $x, y \in S$

$$x \circ y = y \circ x.$$

□

4.1.17 Grupa invertibilních prvků. V každém monoidu existuje alespoň jeden invertibilní prvek; ano, neutrální prvek je vždy invertibilní. Následující tvrzení ukazuje, že množina všech invertibilních prvků monoidu je podpologrupa, která je grupou. To nám umožní definovat grupu invertibilních prvků libovolného monoidu.

Tvrzení. Je dán monoid $(S, \circ, e)$. Označme T množinu všech invertibilních prvků monoidu $(S, \circ, e)$. Pak $(T, \circ, e)$ je podmonoid monoidu $(S, \circ, e)$, který je grupa. □

Zdůvodnění. Podle tvrzení 4.1.9 platí $e^{-1} = e$ a $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$, kdykoli a^{-1} a b^{-1} existují. Tedy množina T obsahuje e a je uzavřena na operaci $\circ$. Jedná se proto o podmonoid monoidu $(S, \circ, e)$. Fakt, že $(T, \circ, e)$ je grupa vyplývá z faktu, že $(a^{-1})^{-1} = a$, jinými slovy, je-li a invertibilní prvek, je invertibilní i prvek a^{-1}.

Definice. Grupu $(T, \circ, e)$ nazýváme *grupa invertibilních prvků* monoidu $(S, \circ, e)$. □

4.1.18 Grupa invertibilních prvků $(\mathbb{Z}_n^\star, \cdot)$.

Víme, že $(\mathbb{Z}_n, \cdot)$ je monoid s neutrálním prvkem 1. Množina všech invertibilních prvků tohoto monoidu je

$$\mathbb{Z}_n^\star = \{i \,|\, 0 \leq i < n, \;\; i \text{ a } n \text{ jsou nesoudělná}\}.$$

Je tedy $(\mathbb{Z}_n^\star, \cdot, 1)$ grupa o $\phi(n)$ prvcích, kde $\phi(n)$ je hodnota Eulerovy funkce pro číslo n.

4.1.19 Věta. Je dána konečná komutativní grupa $(G, \circ, e)$, kde e je neutrální prvek. Označme n počet prvků množiny G. Pak pro každý prvek $g \in G$ platí

$$g^n = e.$$

□

Zdůvodnění. Označme $G = \{a_1, a_2, \ldots, a_n\}$ a utvořme množinu $g \circ G = \{g \circ a_1, g \circ a_2, \ldots, g \circ a_n\}$. Protože $g \circ a_i \neq g \circ a_j$ pro $i \neq j$ (ano, v grupě můžeme prvkem g krátit), jsou množiny G a $g \circ G$ stejné.

Navíc pracujeme v komutativní grupě, proto

$$a_1 \circ a_2 \circ \ldots \circ a_n = (g \circ a_1) \circ (g \circ a_2) \circ \ldots \circ (g \circ a_n).$$

a dále

$$(a_1 \circ a_2 \circ \ldots \circ a_n) = g^n \circ (a_1 \circ a_2 \circ \ldots \circ a_n).$$

Opět užijeme fakt, že pracujeme v grupě a poslední rovnici zkrátíme prvkem $a_1 \circ a_2 \circ \ldots \circ a_n$. Dostaneme $g^n = e$.

4.1.20 Euler-Fermatova věta. Jak jsme slíbili v minulé kapitole využijeme 4.1.19 k důkazu malé Fermatovy věty. Nejprve ale uvedeme obecnější větu, které se říká Euler-Fermatova věta.

Věta (Euler-Fermat). Je dáno přirozené číslo $n > 1$. Pak pro každé celé číslo a nesoudělné s n platí

$$a^{\phi(n)} \equiv 1 \,(\mathrm{mod}\, n).$$

□

Zdůvodnění. Grupa invertibilních prvků monoidu $(\mathbb{Z}_n, \cdot, 1)$ má $\phi(n)$ prvků. Proto z předchozí věty víme, že pro každé $b \in \mathbb{Z}_n^\star$

$$b^{\phi(n)} = 1.$$

To ale znamená, že

$$b^{\phi(n)} \equiv 1 \,(\mathrm{mod}\, n).$$

Protože pro každé číslo a nesoudělné s n existuje $b \in \mathbb{Z}_n^\star$ takové, že $a \equiv b \,(\mathrm{mod}\, n)$ (ano, stačí vzít zbytek při dělení a číslem n), platí také

$$a^{\phi(n)} \equiv 1 \,(\mathrm{mod}\, n).$$

Poznámka. Jestliže číslo n je prvočíslo, pak $\phi(n) = n-1$, a proto je malá Fermatova věta zvláštní případ Euler-Fermatovy věty.

4.2 Podgrupy

Podgrupy grup, speciálně konečných grup, jsou velmi důležité nejen v aplikacích. Věnujeme jim proto celý oddíl.

4.2.1 Definice. Je dána grupa $\mathcal{G} = (G, \circ, e)$. Množina $H \subseteq G$ *tvoří podgrupu* grupy $\mathcal{G}$, jestliže platí

1. pro každé $x, y \in H$ je také $x \circ y \in H$;
2. $e \in H$;
3. pro každé $x \in H$ je také $x^{-1} \in H$.

□

V takovém případě je $(H, \circ, e)$ také grupou.

Poznámka. Často se říká, že množina $H \subseteq G$ *je podgrupa* grupy $\mathcal{G}$, místo přesného H *tvoří podgrupu* grupy $\mathcal{G}$. Je to v případě, kdy je z kontextu jasné o jaké operace se jedná. I my tuto konvenci budeme občas používat.

4.2.2 Věta. Je dána konečná grupa $(G, \circ, e)$ a její podgrupa H. Pak počet prvků H dělí počet prvků G. □

Myšlenka důkazu. Označme $n = |G|$ a $k = |H|$. Pro každé $g \in G$ vytvoříme množinu $g \circ H = \{g \circ x \mid x \in H\}$. Pro různá $g_1, g_2 \in G$ jsou množiny $g_1 \circ H$ a $g_2 \circ H$ buď stejné nebo disjunktní. Protože $g \in g \circ H$ a tudíž každý prvek G leží v některé množině $g \circ H$, tvoří $\{g \circ H \mid g \in G\}$ rozklad množiny G.

Všechny množiny $g \circ H$ mají stejný počet prvků, navíc protože $e \circ H = H$, je tento počet roven k. Je tedy množina G o n prvcích rozdělena na několik množin o k prvcích, proto k dělí n. (Uvědomte si, že různých množin $g \circ H$ je přesně n/k.)

Poznámka. Počet prvků konečné grupy se také nazývá *řád grupy*. Pak předchozí věta může znít: Řád libovolné podgrupy konečné grupy dělí řád grupy.

4.2.3 Řád prvku konečné grupy. Mějme konečnou grupu $(G, \circ, e)$ a zvolme její prvek $a \in G$. Označme a^i součin i-krát prvku a sama se sebou, přesněji

$$a^0 = e,\ a^1 = a,\ a^i = a^{i-1} \circ a.$$

Utvoříme množinu

$$\{a, a^2, a^3, \ldots, a^k, \ldots\}.$$

Protože G je konečná množina, musí existovat i, j takové, že $i \neq j$ a $a^i = a^j$. Označme jako i ten menší exponent, tj. $i < j$. Protože pracujeme v grupě, můžeme prvkem a krátit (tj. násobit obě strany rovnice prvkem a^{-1}). Proto $a^i = a^j$ implikuje $a^{i-1} = a^{j-1}$, atd. až $e = a^{j-i}$.

Definice. Označme $r(a)$ to nejmenší kladné přirozené číslo, pro které $a^{r(a)} = e$. Číslo $r(a)$ se nazývá *řád prvku* a. □

4.2.4 Podgrupa generovaná prvkem.

Definice. Mějme konečnou grupu $(G, \circ, e)$ a její prvek $a \in G$. Množina

$$\{a, a^2, \ldots, a^{r(a)} = e\},$$

kde $r(a)$ je řád prvku a, tvoří podgrupu, která se nazývá *podgrupa generovaná prvkem a* a značíme ji $\langle a \rangle$. □

Fakt, že $\langle a \rangle$ je grupou je zřejmé např. z toho, že 1) $a^i \circ a^j = a^m$, kde m je zbytek při dělení čísla $i + j$ číslem $r(a)$, a 2) $(a^i)^{-1} = a^{r(a)-i}$.

Uvědomte si, řád prvku a je vlastně řád podgrupy $\langle a \rangle$.

4.2.5 Řád prvku a můžeme charakterizovat i jinak, jak ukazuje následující tvrzení. Novou charakterizaci využijeme později.

Tvrzení. Číslo $r > 0$ je řád prvku a v konečné grupě $(G, \cdot, e)$ právě tehdy, když platí následující dvě podmínky:

1) Platí $a^r = e$.
2) Kdykoli $a^s = e$, tak číslo r dělí s.

□

Zdůvodnění. a) Předpokládejme, že číslo r splňuje obě podmínky předchozího tvrzení. Pak zřejmě je r nejmenší číslo takové, že $a^r = e$ a proto je to řád prvku a.

b) Předpokládejme, že r je řád prvku a. Pak zřejmě platí podmínka 1).

Vezměme libovolné číslo s takové, že $a^s = e$. Vydělíme číslo s číslem r, tj. spočítáme $s = qr + z$, kde $0 \leq z < r$. Pak

$$e = a^s = a^{qr+z} = (a^r)^q \cdot a^z = e^q \cdot a^z = a^z.$$

Protože z je menší než r a r je nejmenší kladné číslo pro nějž $a^i = e$, musí být $z = 0$. Tj. r dělí s.

Důsledek. Mějme konečnou grupu $(G, \circ, e)$ o n prvcích. Pak řád každého prvku $a \in G$ dělí řád grupy $(G, \circ, e)$. □

Toto tvrzení je zřejmým důsledkem 4.2.2. Ano, $\langle a \rangle$ je podgrupa grupy $(G, \cdot, e)$ a počet jejích prvků je $r(a)$.

4.2.6 Věta. Mějme konečnou grupu $(G, \circ, e)$ o n prvcích. Pak pro každý její prvek $a \in G$ platí

$$a^n = e.$$

Zdůvodnění. Protože řád libovolného prvku $a \in G$ dělí n, je $a^n = a^{k\, r(a)} = (a^{r(a)})^k = e^k = e$. □

Uvědomte si, že věta 4.1.19 tvrdila to samé pro komutativní konečné grupy. Nyní víme, že tvrzení platí i pro nekomutativní konečné grupy.

4.2.7 Generátor grupy, cyklické grupy. Grupy, které mají nejjednodušší strukturu, jsou ty, jejichž každý prvek je mocninou některého prvku g grupy. V těchto grupách „násobit“ prvky znamená vlastně „sčítat“ exponenty. Takovým grupám říkáme cyklické a prvku g říkáme generátor.

Definice. Je dána konečná grupa $\mathcal{G} = (G, \circ, e)$. Jestliže pro prvek $a \in G$ platí, že $\langle a \rangle = G$, pak se prvek a nazývá *generátor grupy* $\mathcal{G}$. Každá grupa, která má generátor, se nazývá *cyklická grupa.* □

Konečná grupa $\mathcal{G} = (G, \circ, e)$ řádu n (tj. o n prvcích) je cyklická právě tehdy, když v ní existuje prvek řádu n.

Poznamenejme, že o cyklické grupě mluvíme i v případě, že se jedná o nekonečnou grupu (příkladem takové grupy je např. množina celých čísel se sčítání), ale těmi se nebudeme zabývat.

Příklady.

1. Pro každé přirozené číslo $n > 1$ je grupa $(\mathbb{Z}_n, +, 0)$ cyklická grupa s generátorem 1.
2. Pro každé prvočíslo p je grupa $(\mathbb{Z}_p^\star, \cdot, 1)$ cyklická grupa. Najít její generátor není vždy jednoduché.
3. Grupa $(\mathbb{Z}_8^\star, \cdot, 1)$ není cyklická; skládá se ze čtyř prvků: $\mathbb{Z}_8^\star = \{1, 3, 5, 7\}$ a $3^2 = 1$, $5^2 = 1$ a $7^2 = 1$.

4.2.8 Následující věta nám umožňuje spočítat řády všech prvků v cyklické grupě. S jejím využitím budeme např. schopni říci kolik generátorů cyklická grupa o n prvcích má.

Věta. Mějme konečnou grupu $\mathcal{G} = (G, \circ, e)$ a prvek $a \in G$ řádu $r(a)$. Pak prvek a^i má řád:

$$r(a^i) = \frac{r(a)}{\gcd(r(a), i)}.$$

□

Zdůvodnění. Dokážeme, že číslo $\frac{r(a)}{\gcd(r(a),i)}$ splňuje podmínky věty 4.2.5.

Označme $r := r(a)$, $d := \gcd(i, r)$. Pak $i = d\,i'$ a $r = d\,r'$, kde $\gcd(i', r') = 1$. Číslo $\frac{r(a)}{\gcd(r(a),i)}$ je při tomto značení rovno r'.

Platí:

$$(a^i)^{r'} = a^{ir'} = a^{i'dr'} = (a^{dr'})^{i'} = (a^r)^{i'} = e.$$

Navíc, jestliže $(a^i)^s = e$, pak $a^{is} = e$. Protože r je řád prvku a, znamená to, že číslo r dělí součin is. Máme proto

$$is = kr, \quad \text{tj.} \quad i'ds = kr'd \quad \text{a} \quad i's = kr'.$$

Čísla i' a r' jsou nesoudělná a číslo r' dělí součin $i'\,s$, proto r' dělí s.

Pozorování. Předchozí věta nám dává návod, jak spočítat řády všech prvků b z podgrupy $\langle a \rangle$. O podgrupě $\langle a \rangle$ víme, že je cyklická a že a je její generátor. Můžeme proto na každý prvek $b \in \langle a \rangle$ použít tvrzení 4.2.5. Speciálně, známe-li generátor cyklické grupy, můžeme spočítat řády všech prvků této grupy.

4.2.9 Důsledek. Mějme konečnou cyklickou grupu $\mathcal{G} = (G, \circ, e)$ o n prvcích. Pak $\mathcal{G}$ má $\phi(n)$ generátorů. □

Zdůvodnění. Označme a některý generátor grupy $\mathcal{G}$. Pak a^i je také generátor $\mathcal{G}$ právě tehdy, když je i nesoudělné s n.

4.2.10 Tvrzení. Mějme konečnou cyklickou grupu $\mathcal{G} = (G, \circ, e)$ o n prvcích. Pak pro každé číslo d, které dělí n, existuje podgrupa $\mathcal{G}$ o d prvcích. □

Zdůvodnění. Označme a některý generátor grupy $\mathcal{G}$. Pak hledaná podgrupa je podgrupa generovaná prvkem a^k, kde $k = \frac{n}{d}$; tj.

$$\langle a^k \rangle = \{a^k, a^{2k}, \ldots, a^{dk} = e\}.$$

4.2.11 Důsledek. Konečná cyklická grupa má všechny podgrupy cyklické, a to výše uvedeného tvaru. □

Zdůvodnění. Mějme cyklickou grupu s generátorem a. Vezměme dva prvky této grupy, řekněme a^i a a^j. Podgrupa, která tyto prvky obsahuje, obsahuje též všechny prvky tvary a^{ix+jy}, kde x a y jsou libovolná celá čísla. Z Bezoutovy věty víme, že rovnice $ix + jy = k$ má celočíselné řešení právě tehdy, když největší společný dělitel čísel i a j dělí číslo k. Odtud plyne, že nejmenší podgrupa obsahující prvky a^i a a^j je $\langle a^d \rangle$, kde d je největší společný dělitel čísel i a j.

Kapitola 5

Struktury se dvěma binárními operacemi

V minulých kapitolách jsme studovali množiny s jednou binární operací. Přitom např. na množině všech reálných čísel, na množině čtvercových matic, i na množině zbytkových tříd modulo n existují dvě různé operace, navíc svázáné distributivním zákonem. To přináší další možnosti využití těchto operací nad rámec jednotlivých operací.

V této kapitole nejprve zkonstruujeme nové příklady množin se dvěma operacemi; na to využijeme polynomy nad $\mathbb{Z}_p$, kde p je prvočíslo. K tomu zadefinujeme kongruence modulo polynom na množině všech polynomů stupně menšího než dané k. Jedná se o konstrukci obdobnou konstrukci $\mathbb{Z}_n$. Těchto příkladů využijeme k zavedení nových typů struktur – okruhů a těles – jejichž příklady budou i známá reálná čísla.

V další části kapitoly zavedeme svazy a Booleovy algebry. Jsou to také příklady množin s dvěma binárními operacemi, ale v tomto případě se jedá o zobecnění příkladu $(P(U), \cup, \cap)$ (pro nějakou množinu U).

Než přistoupíme ke studiu polynomů připomeňme, že pro všechna přirozená čísla n, $n > 1$, je $(\mathbb{Z}_n, +, 0)$ komutativní grupa; $(\mathbb{Z}_n, \cdot, 1)$ je komutativní monoid a $(\mathbb{Z}_n^\star, \cdot, 1)$ je jeho grupa invertibilních prvků. Jestliže p je prvočíslo, pak invertibilní je každý nenulový prvek $\mathbb{Z}_p$ (tj. $\mathbb{Z}_p^\star = \mathbb{Z}_p \setminus \{0\}$).

Navíc, mezi sčítáním a násobením platí distributivní zákon, tj.

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{pro všechny} \quad a, b, c \in \mathbb{Z}_p. \tag{5.1}$$

Uvědomme si, že pro $\mathbb{Z}_p$ se jedná o všechny vlastnosti, které známe z počítání s reálnými čísly.

Pro jednoduchost budeme místo dlouhého $(\mathbb{Z}_n, +, \cdot, 0, 1)$ psát pouze $\mathbb{Z}_n$ všude tam, kde bude jasné, že na množině zbytkových tříd $\mathbb{Z}_n$ pracujeme se dvěma binárními operacemi a to sčítáním a násobením.

5.1 Polynomy nad $\mathbb{Z}_p$

Polynomy studujeme pouze nad $\mathbb{Z}_p$, kde p je prvočíslo. Poznamenejme, že polynomy můžeme vytvářet také nad $\mathbb{Z}_n$ pro n složené. Ovšem v tomto případě ne vždy se dá dělit. Ano, např. v $\mathbb{Z}_{15}$ nelze dělit číslem 3.

5.1.1 Polynomy nad $\mathbb{Z}_p$.

Definice. *Polynomem* nad $\mathbb{Z}_p$ rozumíme každý výraz

$$a_0 + a_1\, x + a_2\, x^2 + \ldots + a_n\, x^n, \tag{5.2}$$

kde všechny koeficienty a_i jsou ze $\mathbb{Z}_p$. Symbol x nazýváme *proměnná*. Jsou-li všechny koeficienty a_i nulové, mluvíme o *nulovém* polynomu. Jestliže alespoň jeden z koeficientů a_i je nenulový, mluvíme o *nenulovém* polynomu. Množinu všech polynomů nad $\mathbb{Z}_p$ značíme $\mathbb{Z}_p[x]$. □

5.1.2 Stupeň polynomu. Podobně jako pro reálné polynomy definujeme stupeň polynomu.

Definice. *Stupeň* nulového polynomu je roven -1, *stupeň* nenulového polynomu je roven největšímu n takovému, že a_n je nenulové. Stupeň polynomu $f(x)$ značíme $\mathrm{st}(f)$. □

5.1.3 Rovnost polynomů.

Definice. Dva polynomy jsou si **rovny**, rovnají-li se všechny jejich odpovídající koeficienty, tj. mají-li stejný stupeň a koeficienty u stejných mocnin jsou stejné. □

5.1.4 Funkce odpovídající polynomu. Každému polynomu $f(x)$ ze $\mathbb{Z}_p[x]$ můžeme přiřadit zobrazení $\mathbb{Z}_p$ do sebe a to tímto způsobem: Je-li $f(x) = a_0 + a_1\,x + a_2\,x^2 + \ldots + a_n\,x^n$, pak jemu odpovídající zobrazení ze $\mathbb{Z}_p$ do $\mathbb{Z}_p$ je definováno:

$$\text{pro každé } b \in \mathbb{Z}_p \quad b \mapsto a_0 + a_1\,b + a_2\,b^2 + \ldots + a_n\,b^n.$$

(Všechny operace sčítání a násobení jsou operace v $\mathbb{Z}_p$.)

Např. polynomu $f(x) = 1+x^2$ nad $\mathbb{Z}_2$ odpovídá zobrazení $\mathbb{Z}_2$ do $\mathbb{Z}_2$, kde $0 \mapsto 1$, $1 \mapsto 1+1^2 = 0$. Uvědomte si, že se jedná o stejné zobrazení jako to zobrazení, které odpovídá polynomu $g(x) = 1+x$. Není tedy pravda, že různým polynomům odpovídají vždy různá zobrazení. To není překvapivé – všech polynomů je spočetně mnoho, kdežto zobrazení za $\mathbb{Z}_p$ do sebe je jen konečně mnoho.

5.1.5 Poznámka. Poznamenejme, že polynomy nad reálnými nebo komplexními čísly se často definují jako funkce (reálné nebo komplexní) dané výrazem (5.2) pro reálné, nebo komplexní koeficienty a_i. To je možné proto, že v případě reálných polynomů platí, že dva polynomy jsou si rovny jakožto funkce právě tehdy, když jsou si rovny jako formální výrazy.

5.1.6 Sčítání a odčítání polynomů. Analogicky jako pro reálné (komplexní) polynomy, i pro polynomy nad $\mathbb{Z}_p$ definujeme operaci sčítání, a také odčítání dvou polynomů.

Máme dány polynomy nad $\mathbb{Z}_p$: $f(x) = a_0 + a_1\,x + a_2\,x^2 + \ldots + a_n\,x^n$ a $g(x) = b_0 + b_1\,x + b_2\,x^2 + \ldots + b_m\,x^m$ stupně n, resp. m, $n \geq m$, pak jejich *součtem* je polynom

$$(a_0 + b_0) + (a_1 + b_1)\,x + \ldots + (a_m + b_m)\,x^m + a_{m+1}\,x^{m+1} + \ldots + a_n\,x^n,$$

rozdílem $f(x)$ mínus $g(x)$ je polynom

$$(a_0 - b_0) + (a_1 - b_1)\,x + \ldots + (a_m - b_m)\,x^m + a_{m+1}\,x^{m+1} + \ldots + a_n\,x^n.$$

□

Pro stupeň součtu nebo rozdílu dvou polynomů platí

$$\mathrm{st}(f \pm g) \leq \max(\mathrm{st}(f), \mathrm{st}(g)).$$

5.1.7 Násobení polynomů. Polynomy násobíme opět obdobně jako v reálné polynomy.

Součinem polynomů $f(x)$ a $g(x)$ je polynom $h(x) = c_0 + c_1\,x + c_2\,x^2 + \ldots c_{n+m}\,x^{n+m}$, kde koeficienty $c_0, c_1, \ldots, c_{n+m}$ dostaneme tak, že polynomy „vynásobíme jako mnohočleny“, tedy např. $c_0 = a_0\,b_0$, $c_1 = a_0\,b_1 + a_1\,b_0$, …, $c_{n+m} = a_n\,b_m$. □

Platí $\mathrm{st}(f \cdot g) = \mathrm{st}(f) + \mathrm{st}(g)$ kdykoli jsou oba polynomy nenulové, v opačném případě je $\mathrm{st}(f \cdot g) = -1$.

5.1.8 Věta o dělení polynomů. Pro dělení polynomů platí následující věta (opět můžete srovnat s větou o dělení reálných polynomů).

Věta. Mějme dva polynomy $f(x)$ a $g(x)$ ze $\mathbb{Z}_p[x]$, kde $g(x)$ je nenulový. Pak existují polynomy $r(x)$ a $z(x)$ v $\mathbb{Z}_p[x]$ takové, že

$$f(x) = r(x)\,g(x) + z(x) \quad \text{a} \quad \mathrm{st}(z) < \mathrm{st}(g).$$

Tyto polynomy jsou určeny jednoznačně. □

Polynomu $r(x)$ z věty o dělení se říká *částečný podíl*, polynom $z(x)$ je *zbytek při dělení* polynomu $f(x)$ polynomem $g(x)$.

5.1.9 Zdůvodnění věty o dělení. Máme-li dokázat větu o dělení, měli bychom nejprve ukázat existenci částečného podílu a zbytku a pak jejich jednoznačnost. Existence vyplývá z algoritmu dělení:

Existence — algoritmus dělení: Polynomy $r(x)$ a $z(x)$ ze znění věty najdeme stejným způsobem jako v při dělení reálných polynomů, pouze místo „dělení prvkem“ používáme „násobení inverzním prvkem“. Ukažme si postup podrobněji:

Máme $f(x) = a_0 + a_1\,x + \ldots a_n\,x^n$, $g(x) = b_0 + b_1\,x + \ldots b_m\,x^m$, $\mathrm{st}(f) = n$, $\mathrm{st}(g) = m$. Jestliže $n < m$, pak platí

$$f(x) = 0 \cdot g(x) + f(x) \quad \text{a} \quad \mathrm{st}(f) < \mathrm{st}(g).$$

Tedy polynomy $r(x) = 0$ a $z(x) = f(x)$ jsou hledané polynomy.

Předpokládejme, že $n \geq m$. Vezmeme členy polynomů $f(x)$ a $g(x)$ s největší mocninou, tj. $a_n\,x^n$ a $b_m\,x^m$. Vydělíme $a_n\,x^n$ členem $b_m\,x^m$; podíl je roven $a_n\,b_m^{-1}\,x^{n-m}$. Polynom $a_n\,b_m^{-1}\,x^{n-m} \cdot g(x)$ je polynom stupně n a má koeficient u nejvyšší mocniny a_n. Položíme

$$h(x) = f(x) - a_n\,b_m^{-1}\,x^{n-m} \cdot g(x).$$

Polynom $h(x)$ má stupeň ostře menší než n.

Nyní buď $\mathrm{st}(h) < \mathrm{st}(g)$ a jsme hotovi; částečný podíl je $r(x) = a_n\,b_m^{-1}\,x^{n-m}$, zbytek při dělení je $z(x) = h(x)$.

Nebo $\mathrm{st}(h) \geq \mathrm{st}(g)$ a částečný podíl při dělení polynomu $f(x)$ polynomem $g(x)$ je roven součtu $a_n\,b_m^{-1}\,x^{n-m}$ a částečného podílu při dělení polynomu $h(x)$ polynomem $g(x)$. Zbytek při dělení $f(x)$ polynomem $g(x)$ je stejný jako zbytek při dělení $h(x)$ polynomem $g(x)$.

Jednoznačnost: Kdyby pro nějaké polynomy $f(x)$ a $g(x)$ existovaly polynomy $r_1(x)$, $r_2(x)$ a $z_1(x)$ a $z_2(x)$ takové, že

$$f(x) = r_i(x)\,g(x) + z_i(x), \;\; 0 \leq \mathrm{st}(z_i) < \mathrm{st}(g), \;\; i = 1, 2,$$

pak by

$$(r_1(x) - r_2(x))g(x) = z_1(x) - z_2(x).$$

Protože stupeň polynomu $(r_1(x) - r_2(x))g(x))$ je buď -1 nebo alespoň $\mathrm{st}(g)$, a stupeň polynomu $z_1(x) - z_2(x)$ je ostře menší než $\mathrm{st}(g)$, musí $\mathrm{st}((r_1(x) - r_2(x))g(x))$ být roven -1, tedy musí to být nulový polynom. Proto $z_1(x) - z_2(x)$ je také nulový polynom a platí

$$r_1(x) = r_2(x), \quad z_1(x) = z_2(x).$$

5.1.10 Dělitelnost polynomů.

Definice. Jestliže zbytek při dělení polynomu $f(x)$ polynomem $g(x)$ je nulový polynom, říkáme, že polynom $g(x)$ *dělí* polynom $f(x)$, nebo že polynom $f(x)$ *je dělitelný* polynomem $g(x)$, nebo také, že polynom $g(x)$ *je dělitelem* polynomu $f(x)$. □

5.1.11 Kořen polynomu. Obdobně jako pro polynomy nad reálnými čísly, i v případě polynomů nad $\mathbb{Z}_p$ hraje důležitou roli pojem kořene.

Definice. Prvek $a \in \mathbb{Z}_p$ se nazývá *kořen* polynomu $f(x) \in \mathbb{Z}_p[x]$, jestliže platí $f(a) = 0$. □

Tvrzení. Prvek $a \in \mathbb{Z}_p$ je kořenem polynomu $f(x) \in \mathbb{Z}_p[x]$ právě tehdy, když polynom $(x-a)$ dělí polynom $f(x)$. □

Zdůvodnění je jednoduché. Z věty o dělení víme, že

$$f(x) = h(x)\,(x-a) + z(x), \quad \mathrm{st}(z) < \mathrm{st}(x-a) = 1. \tag{5.3}$$

Proto je $z(x) = z_0$ pro $z_0 \in \mathbb{Z}_p$. Navíc, dosadíme-li $x = a$, dostaneme

$$f(a) = h(x) \cdot 0 + z_0, \quad \text{tedy} \quad z_0 = f(a).$$

Odtud vidíme, že $f(a) = 0$ právě tehdy, když zbytek $z(x)$ je nulový polynom.

5.1.12 Ireducibilní polynomy. Ireducibilní polynomy v $\mathbb{Z}_p[x]$ mají obdobnou roli jako prvočísla v $\mathbb{Z}$.

Definice. Polynom $f(x)$ ze $\mathbb{Z}_p[x]$ se nazývá *ireducibilní* (nad $\mathbb{Z}_p$), jestliže se polynom $f(x)$ nedá napsat jako součin dvou polynomů menšího stupně než je stupeň $f(x)$. □

Jinými slovy, jestliže z rovnosti $f(x) = g(x) \cdot h(x)$ plyne buď $\mathrm{st}(g) = \mathrm{st}(f)$ (a $h(x)$ je polynom stupně 0, tj. konstanta) nebo $\mathrm{st}(h) = \mathrm{st}(f)$ (a $g(x)$ je polynom stupně 0, tj. konstanta).

Příklady.

1. Polynom $f(x) = x^2 + 1$ je ireducibilní nad $\mathbb{Z}_3$, protože se nedá napsat jako součin dvou lineárních polynomů; to by totiž musel mít kořen.
2. Polynom $f(x) = x^2 + 1$ není ireducibilní nad $\mathbb{Z}_2$. Stačí si uvědomit, že v $\mathbb{Z}_2$ platí $(x^2+1) = (x+1)\,(x+1)$.
3. Polynom $g(x) = x^2 + x + 1$ je ireducibilní nad $\mathbb{Z}_2$, nemá totiž kořen.

5.1.13 Pozorování. Polynom stupně 2 nebo 3 je ireducibilní právě tehdy, když nemá kořen.

Pro polynomy stupně většího než 3 už takové tvrzení neplatí; součin dvou ireducibilních polynomů stupně 2 je polynom stupně 4, který není ireducibilní a přesto nemá kořen. Samozřejmě, polynom, který kořen má, nikdy není ireducibilní (viz 5.1.11).

5.1.14 Víme, že reálné ireducibilní polynomy jsou pouze polynomy stupně jedna nebo kvadratické polynomy se záporným diskriminantem. Pro polynomy nad $\mathbb{Z}_p$ takový fakt neplatí. Dokonce platí následující věta, kterou ale nedokazujeme. Její důkaz není jednoduchý a je nad rámec tohoto kurzu.

Tvrzení. Nad $\mathbb{Z}_p$ existují ireducibilní polynomy libovolného stupně. □

5.1.15 Největší společný dělitel dvou polynomů. Stejně jako pro celá čísla, můžeme i pro polynomy definovat pojem společný dělitel a největší společný dělitel.

Definice. Mějme dva polynomy $f(x)$ a $g(x)$ ze $\mathbb{Z}_p[x]$. Polynom $h(x)$ ze $\mathbb{Z}_p[x]$ se nazývá *největší společný dělitel* polynomů $f(x)$ a $g(x)$, jestliže splňuje

1. $h(x)$ dělí oba polynomy $f(x)$ a $g(x)$;
2. kdykoli nějaký polynom $k(x)$ dělí oba polynomy $f(x)$ a $g(x)$, pak $k(x)$ dělí i $h(x)$.

□

Poznamenejme, že jsme také mohli definovat největší společný dělitel dvou polynomů jako společný dělitel, který má největší stupeň mezi všemi společnými děliteli.

Poznámka. Uvědomme si, že pro $\mathbb{Z}_p[x]$, kde p je liché prvočíslo, má vlastnosti největšího společného dělitele dvou polynomů víc než jeden polynom. Ano, jestliže je polynom $h(x)$ největší

společný dělitel polynomů $f(x)$ a $g(x)$, pak také polynom $b \cdot h(x)$, kde b je nenulový prvek $\mathbb{Z}_p$, je největší společný dělitel polynomů $f(x)$ a $g(x)$.

Mezi všemi největšími společnými děliteli polynomů f a g je přesně jeden, který má u nejvyšší mocniny koeficient roven 1 (takovým polynomům se říká *monické*). V některé literatuře se za největší společný dělitel dvou polynomů považuje právě monický největší společný dělitel. V tomto případě je pak určen jednoznačně.

Definice. Dva polynomy nazýváme *nesoudělné*, jestliže $h(x) = 1$ je jejich největší společný dělitel. □

5.1.16 Euklidův algoritmus.

Protože platí věta o dělení, analogicky jako u celých čísel, můžeme hledat největší společný dělitel dvou polynomů Euklidovým algoritmem. Ani Euklidův algoritmus, ani z něj vyplývající Bezoutovu větu pro polynomy nedokazujeme. Důkazy jsou téměř stejné jako v případě celých čísel; jediný rozdíl je, že místo čísla píšeme polynom.

Vstup: Dva polynomy $a(x)$ a $b(x)$ ze $\mathbb{Z}_p[x]$.
Výstup: polynom $h(x)$, který je jeden z největších společných dělitelů $d(x)$ polynomů $a(x)$ a $b(x)$.

1. (Inicializace)
   $t(x) := a(x)$, $r(x) := b(x)$;
2. (Výpočet částečného podílu a zbytku.)
   `repeat`
   `do` $t(x) = q(x) \cdot r(x) + z(x)$;
   $t(x) := r(x)$, $r(x) := z(x)$
   `until` $z(x) = 0$.
3. (Největší společný dělitel)
   $d(x) := r(x)$.

Bezoutova věta. Jsou-li $a(x)$ a $b(x)$ dva polynomy ze $\mathbb{Z}_p[x]$ a $h(x)$ je některý z jejich největších společných dělitelů, pak existují polynomy $f(x)$ a $g(x)$ ze $\mathbb{Z}_p[x]$ takové, že

$$h(x) = f(x)\, a(x) + g(x)\, b(x).$$

□

5.1.17 Řešení polynomiálních rovnic. Jako v případě celých čísel, můžeme (a potřebujeme) řešit tzv. polynomiální rovnice. V našem případě to jsou lineární rovnice, kde koeficienty i proměnné jsou polynomy nad $\mathbb{Z}_p$.

Tvrzení. Jsou dány polynomy $a(x)$, $b(x)$ a $c(x)$ nad $\mathbb{Z}_p$, kde p je prvočíslo. Pak rovnice

$$y(x)\, a(x) + z(x)\, b(x) = c(x) \tag{5.4}$$

má řešení, tj. existují polynomy $y(x)$ a $z(x)$ které splňují rovnici (5.4), právě tehdy, když největší společný dělitel $d(x)$ polynomů $a(x)$ a $b(x)$ dělí pravou stranu, tj. polynom $c(x)$. □

Postup řešení. Jedno řešení rovnice (5.4) najdeme rozšířeným Euklidovým algoritmem, označme ho y_0, z_0. Obecné řešení rovnice (5.4) je pak součtem jednoho řešení (nehomogenní) rovnice (5.4) a obecného řešení přidružené homogenní rovnice

$$a(x)\, y(x) + b(x)\, z(x) = 0. \tag{5.5}$$

Označme $a_1(x)$ ten polynom, pro který $a(x) = d(x)\, a_1(x)$ a $b_1(x)$ ten polynom, pro který $b(x) = d(x)\, b_1(x)$. (Jinými slovy, $a_1(x)$ je polynom, který dostaneme vydělením polynomu $a(x)$ největším společným dělitelem $d(x)$, obdobně $b_1(x)$.) Řešení homogenní rovnice (5.5) je tvaru:

$$y_1(x) = t(x)\, b_1(x), \quad \text{a} \quad z_1(x) = -t(x)\, a_1(x),$$

kde $t(x)$ je libovolný polynom nad $\mathbb{Z}_p$.

Nyní obecné řešení 5.4 je

$$y(x) = y_0(x) + t(x)\, b_1(x) \;\text{ a }\; z(x) = z_0(x) - t(x)\, a_1(x), \quad \text{kde } t(x) \in \mathbb{Z}_p[x].$$

5.1.18 Kongruence modulo polynom. Analogií kongruence modulo n na množině celých čísel je kongruence modulo polynom $q(x)$ na množině všech polynomů $\mathbb{Z}_p[x]$.

Definice. Je dán polynom $q(x) \in \mathbb{Z}_p[x]$. Na množině $\mathbb{Z}_p[x]$ definujeme relaci *modulo* $q(x)$ takto:

$$a(x) \equiv b(x) \,(\mathrm{mod}\, q(x)) \text{ právě, když polynom } (a(x) - b(x)) \text{ je dělitelný } q(x).$$

□

Relace kongruence modulo polynom mají obdobné vlastnosti jako má kongruence modulo n. Uvádíme je v následujících dvou odstavcích. Důkazy jsou stejné jako pro modulo n a proto je neuvádíme.

5.1.19 Tvrzení. Pro každé dva polynomy $a(x)$, $b(x)$ ze $\mathbb{Z}_p[x]$ platí:

$$a(x) \equiv b(x) \,(\mathrm{mod}\, q(x))$$

právě tehdy, když platí jedna z následujících podmínek:

1. $a(x) = b(x) + t(x)\, q(x)$ pro vhodný polynom $t(x) \in \mathbb{Z}_p[x]$;
2. $a(x)$ i $b(x)$ mají stejný zbytek při dělení polynomem $q(x)$.

□

5.1.20 Tvrzení. Relace modulo $q(x)$ je relace ekvivalence na množině $\mathbb{Z}_p[x]$, jinými slovy tato relace je reflexivní, symetrická a tranzitivní.

Navíc, jestliže $a(x) \equiv b(x) \,(\mathrm{mod}\, q(x))$ a $c(x) \equiv d(x) \,(\mathrm{mod}\, q(x))$, pak také

$$(a(x) + c(x)) \equiv (b(x) + d(x) \,(\mathrm{mod}\, q(x))$$

$$a(x) \cdot c(x) \equiv b(x) \cdot d(x) \,(\mathrm{mod}\, q(x)).$$

□

5.1.21 Označme k stupeň polynomu $q(x)$. Předchozí tvrzení umožňuje definovat množinu tříd ekvivalence modulo $q(x)$, budeme ji označovat $\mathbb{Z}_p[x]/q(x)$. Protože každá třída obsahuje přesně jeden polynom stupně menšího než je k, lze psát

$$\mathbb{Z}_p[x]/q(x) = \{[a(x)] \,|\, \mathrm{st}(a) < \mathrm{st}(q)\}.$$

Navíc, na množině $\mathbb{Z}_p[x]/q(x)$ definujeme dvě operace a to sčítání a násobení takto:

$$[a(x)] + [c(x)] = [a(x) + c(x)]$$

$$[a(x)] \cdot [c(x)] = [a(x) \cdot c(x)].$$

Není těžké ukázat, analogicky jako pro zbytkové třídy modulo n, že $(\mathbb{Z}_p[x]/q(x), +)$ je komutativní grupa s neutrálním prvkem $[0]$, $(\mathbb{Z}_p[x]/q(x), \cdot)$ je komutativní monoid s neutrálním prvkem $[1]$, a že mezi oběma operacemi platí distributivní zákony.

5.2 Okruhy a tělesa

V tomto oddíle ukážeme společné vlastnosti množin se dvěma binárními operacemi, jejichž příkladem mohou být reálná čísla, $(\mathbb{Z}_n, +, \cdot)$, nebo $(\mathbb{Z}_p[x]/q(x), +, \cdot)$.

5.2.1 Okruh s jednotkou.

Definice. Množinu M spolu se dvěma binárními operacemi + a $\cdot$ nazýváme *okruh s jednotkou*, jestliže $(M, +, 0)$ je komutativní grupa, $(M, \cdot, 1)$ je monoid a platí distributivní zákony, tj. pro každé $a, b, c \in M$

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{a} \quad (b + c) \cdot a = b \cdot a + c \cdot a.$$

Jestliže navíc je násobení komutativní, pak okruh nazýváme *komutativní okruh s jednotkou.* □

Příklady.

a) $(\mathbb{R}, +, \cdot)$. kde $\mathbb{R}$ je množina všech reálných čísel spolu se sčítáním a násobením tvoří komutativní okruh s jednotkou.
b) $(M_n, +, \cdot)$, kde M_n je množina všech reálných čtvercových matic řádu n, + je sčítání matic a $\cdot$ je násobení matic, tvoří okruh s jednotkou, ale ne komutativní okruh (násobení matic není komutativní).
c) $(\mathbb{Z}_n, +, \cdot)$ je komutativní okruh s jednotkou, kde jednotkový prvek je 1. Tento okruh má n prvků.
d) $(\mathbb{Z}_p[x]/q(x), +, \cdot)$ je komutativní okruh s jednotkou, kde jednotkový prvek je třída $[e(x)]$, kde platí $e(x) = 1$ je polynom stupně 0. Tento okruh má p^k prvků, kde $k = \text{st}(q)$.

5.2.2 Dělitele nuly. V okruhu s jednotkou se může stát, že součin dvou nenulových prvků je roven nulovému prvku. To znáte z lineární algebry; existují nenulové čtvercové matice tak, že jejich součin je nulová matice.

Definice. Nenulovým prvkům a, b okruhu říkáme *dělitele nuly*, jestliže $a \cdot b = 0$. □

Příklady.

a) Okruh $(M_2, +, \cdot)$ má dělitele 0, jsou to např. matice

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

b) V okruhu $(\mathbb{Z}_6, +, \cdot)$ jsou např. 2 a 3 dělitele nuly.
c) Jestliže polynom $q(x)$ není ireducibilní nad $\mathbb{Z}_p$, tj. $q(x) = m(x) \cdot n(x)$, kde $\text{st}(m) > 0$, $\text{st}(n) > 0$, pak třídy $[m(x)]$ a $[n(x)]$ jsou dělitele nuly v $(\mathbb{Z}_p[x]/q(x), +, \cdot)$.

5.2.3 Těleso.

Definice. Komutativní okruh s jednotkou se nazývá *těleso*, jestliže každý jeho nenulový prvek je invertibilní (vzhledem k násobení) a $0 \neq 1$. □

Poznamenejme, že podmínka $0 \neq 1$ znamená, že těleso musí mít alespoň dva prvky. Ty také stačí $(\mathbb{Z}_2, +, \cdot)$ je těleso o přesně dvou prvcích — 0 a 1.

Příklady.

a) $(\mathbb{R}, +, \cdot)$, kde $\mathbb{R}$ je množina všech reálných čísel.
b) $(\mathbb{Z}_p, +, \cdot)$, kde p je prvočíslo.
c) $(\mathbb{Z}_p[x]/q(x), +, \cdot)$, kde p je prvočíslo a $q(x)$ je ireducibilní polynom v $\mathbb{Z}_p[x]$.
d) $(\mathbb{C}, +, \cdot)$, kde $\mathbb{C}$ je množina všech komplexních čísel.

5.2.4 Tvrzení. V tělese neexistují dělitele nuly. □

Zdůvodněni. Ano, kdyby v tělese platilo $a \cdot b = 0$ a kdyby a byl nenulový prvek, pak by existoval prvek a^{-1} inverzní k a. Proto by platilo

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

5.2.5 Poznámka. V předmětu zabývající se lineární algebrou jste vyšetřovali vektorové prostory, řešili soustavy rovnic, pracovali s maticemi; vše nad reálnými čísly. To ale není nutné — pojmy lineární algebry je možné studovat i nad jakým tělesem, třeba tělesem $\mathbb{Z}_2$. To se v aplikacích také často využívá.

5.3 Konečná (Galoisova) tělesa

Ukázali jsme, jak je možné zkonstruovat konečná tělesa pomocí polynomů nad $\mathbb{Z}_p$. Nyní ukážeme několik vlastností, které splňuje každé konečné těleso.

5.3.1 Máme dáno konečné těleso $(F, +, \cdot, 0, 1)$. Připomeňme, že $(F, +, 0)$ je komutativní grupa, $(F, \cdot, 1)$ je komutativní monoid, platí distributivní zákony a $0 \neq 1$.

Vezměme prvek $1 \in F$ a utvořme prvky

$$1,\ 1+1,\ 1+1+1,\ \ldots,\ \underbrace{1+1+\ldots+1}_{i-\text{krát}},\ \ldots$$

Protože F je konečná množina, musí existovat i, j, $i < j$, tak že

$$\underbrace{1+1+\ldots+1}_{i-\text{krát}} = \underbrace{1+1+\ldots+1}_{j-\text{krát}}.$$

Pak

$$0 = \underbrace{1+1+\ldots+1}_{(j-i)-\text{krát}}$$

5.3.2 Tvrzení. Označme n nejmenší kladné přirozené číslo, pro které

$$0 = \underbrace{1+1+\ldots+1}_{n-\text{krát}}.$$

Pak n je prvočíslo.

5.3.3 Charakteristika tělesa. Prvočíslo p z tvrzení 5.3.2 se nazývá *charakteristika* tělesa $(F, +, \cdot, 0, 1)$.

Jestliže konečné těleso je charakteristiky p, pak obsahuje $\mathbb{Z}_p$ jako podtěleso.

5.3.4 Poznámka. Uvědomte si, že charakteristika p je vlastně řád prvku 1 v grupě $(F, +, 0)$.

5.3.5 Tvrzení. Je dáné konečné těleso $(F, +, \cdot, 0, 1)$ charakteristiky p. Pak pro každé $a, b \in F$ platí

$$(a+b)^p = a^p + b^p.$$

5.3.6 Poznámka. Konečná tělesa studovat a popsal francouzský matematik Évariste Galois a proto se konečným tělesům také říká Galoisova tělesa.

É. Gaolois dokázal následující větu:

Věta. Pro každé prvočíslo p a každé přirozené číslo $k \geq 1$ existuje konečné těleso o p^k prvcích. Toto těleso je jediné (až na pojmenování jeho prvků).

Těleso o p^k se často označuje jako $\mathrm{GF}(p^k)$ a mluví se o něj jako o Galoisově tělese s p^k prvky. Uvědomte si, že předchozí věta také říká, že kdybychom znali ireducibilní polynom nad $\mathbb{Z}_p$ libovolného stupně $k \geq 1$, mohli bychom zkonstruovat libovolné konečné těleso.

5.4 Svazy a Booleovy algebry

Už jsme si ukázali několik příkladů množin, na kterých jsme měli dány dvě operace. Jednalo se o příklady okruhů a těles. V aplikacích se ale setkáte i se strukturami se dvěma binárními operacemi, které mají jiné vlastnosti — jsou to svazy a Booleovy algebry. Tyto struktury kromě binárních operací „obsahují“ i z operací odvozené částečné uspořádání.

Při studiu svazů mějte na paměti základní příklad — je to množina všech podmnožin nějaké pevně dané množiny U s operacemi průnik a sjednocení.

5.4.1 Nejprve připomene vlastnosti operací průnik a sjednocení množin.

Příklad — podmnožiny s průnikem a sjednocením. Uvažujme množinu $\mathcal{P}(U)$ a na ní operace $\cap$ a $\cup$. Pak pro každé tři množiny $A, B, C \subseteq U$ platí

1. $A \cap (B \cap C) = (A \cap B) \cap C$, $A \cup (B \cup C) = (A \cup B) \cup C$ (asociativní zákon).
2. $A \cap B = B \cap A$, $A \cup B = B \cup A$ (komutativní zákon).
3. $A \cap A = A$, $A \cup A = A$ (zákon idempotence).
4. $A \cap (B \cup A) = A$, $A \cup (B \cap A) = A$ (zákon pohlcení).

5.4.2 Svaz. Čtyři vlastnosti průniku a sjednocení, které jsem uvedli v minulém odstavci, jsou vlastnosti, které budeme vyžadovat pro libovolný svaz.

Definice. *Svaz* je trojice $(M, \wedge, \vee)$, kde M je neprázdná množina, $\wedge$ a $\vee$ jsou dvě binární operace splňující asociativní zákon, komutativní zákon, zákon idempotence a zákon pohlceni (jak byly definovány v 5.4.1. Operace $\wedge$ se nazývá *průsek* operace a $\vee$ *spojení*. □

5.4.3 Uspořádání na svazu. Jestliže $(M, \wedge, \vee)$ je svaz, pak pomocí operací $\wedge$ nebo $\vee$ lze definovat relaci částečného uspořádání na množině M. Připomeňme, že částečné uspořádání je každá relace, která je reflexivní, antisymetrická a tranzitivní.

Definice. Je dán svaz $(M, \vee, \wedge)$. Na množině M definujeme relaci $\sqsubseteq$ předpisem:

$$a \sqsubseteq b \text{ právě tehdy, když } a \wedge b = a$$

□

Dá se dokázat, že $a \wedge b = a$ právě tehdy, když $a \vee b = b$. Fakt, že relace $\sqsubseteq$ je skutečně částečné uspořádání, ukazuje následující tvrzení.

Tvrzení. Relace $\sqsubseteq$ na množině prvků nějakého svazu je reflexivní, antisymetrická a tranzitivní. Jedná se proto o relaci částečného uspořádání. □

Zdůvodnění. Reflexivita: Protože pro každý prvek $a \in M$ je $a \wedge a = a$, platí $a \sqsubseteq a$. Ukázali jsme, že relace $\sqsubseteq$ je reflexivní.

Antisymetrie: Předpokládejme, že pro $a, b \in M$ platí $a \sqsubseteq b$ a $b \sqsubseteq a$. To znamená, že $a \wedge b = a$ a $b \wedge a = b$. Protože $a \wedge b = b \wedge a$, dostáváme $a = b$ a relace je antisymetrická.

Tranzitivita: Předpokládejme, že pro $a, b, c \in M$ platí $a \sqsubseteq b$ a $b \sqsubseteq c$. Tedy $a \wedge b = a$ a $b \wedge c = b$. Potom

$$a \wedge c = (a \wedge b) \wedge c = a \wedge (b \wedge c) = a \wedge b = a.$$

Proto $a \sqsubseteq c$ a relace je tranzitivní.

5.4.4 Poznámka. Operacím $\wedge$ a $\vee$ ze svazu se také říká *infimum* a *supremum*. Je to proto, že pro relaci $\sqsubseteq$ je prvek $a \wedge b$ největší dolní mezí a prvek $a \vee b$ nejmenší dolní mezí množiny $\{a, b\}$.

5.4.5 Nejmenší prvek 0 a největší prvek 1 ve svazu.

Definice. Je dán svaz $(M, \wedge, \vee)$. *Nejmenší prvek*, značíme ho většinou **0**, je ten prvek $y \in M$, pro který platí

$$y \wedge a = y, \quad y \vee a = a \quad \text{pro každé } a \in M.$$

Největší prvek, značíme ho většinou **1**, je ten prvek $x \in M$, pro který platí

$$x \wedge a = a, \quad x \vee a = x \quad \text{pro každé } a \in M.$$

□

Uvědomte si, že jména nejmenší a největší prvek jsou „správně“, protože pro nejmenší prvek **0** platí $0 \sqsubseteq a$ pro každé $a \in M$ a pro největší prvek **1** platí $a \sqsubseteq \mathbf{1}$ pro každé $a \in M$.

5.4.6 Distributivní svaz. Náš základní příklad svazu, tj. $(P(U), \cap, \cup)$ (viz 5.4.1), splňuje ještě další vlastnosti. Už z první přednášky víme, že průnik a sjednocení splňují distributivní zákony. Distributivní zákon ovšem ze základních vlastností operací ve svazu nevyplývá. (Na přednášce si ukážeme dva příklady malých svazů, které distributivní zákon nesplňují.) Na druhou stranu svazy, které distributivní zákony splňují, hrají v aplikacích důležitou roli.

Definice. Svaz $(M, \wedge, \vee)$ se nazývá *distributivní*, jestliže v něm platí distributivní zákony, tj. pro každé tři prvky $a, b, c \in M$ platí

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c), \quad a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c).$$

□

5.4.7 V každém distributivním svazu platí následující tvrzení:

Tvrzení. Uvažujme distributivní svaz $(M, \wedge, \vee)$. Jestliže pro prvky $a, b, c \in M$ platí

$$a \wedge b = a \wedge c \quad \text{a} \quad a \vee b = a \vee c,$$

pak prvky b a c jsou stejné. □

Zdůvodnění. V libovolném distributivním svazu platí:

$$b = b \wedge (a \vee b) = b \wedge (a \vee c) = (b \wedge a) \vee (b \wedge c) = (a \wedge c) \vee (b \wedge c) = (a \vee b) \wedge c = (a \vee c) \wedge c = c.$$

Uvědomte si, že jsme dvakrát použili zákon pohlcení (na začátku a na konci výpočtu), dvakrát distributivní zákon a dvakrát komutativní zákon. Zbytek byly předpoklady tvrzení.

5.4.8 Doplněk ve svazu. Uvedeme ještě jeden pojem, který množinový svaz má, ale obecný svaz mít nemusí. Jedná se o doplněk prvku. V příkladě $(P(U), \cap, \cup)$ je pro $X \in P(U)$ roven $U \setminus X = \overline{X}$.

Definice. Je dán svaz $(M, \wedge, \vee)$ s $\mathbf{0}$ a $\mathbf{1}$. Řekneme, že prvek b je *doplňkem prvku* a, jestliže platí

$$a \wedge b = \mathbf{0}, \quad \text{a} \quad a \vee b = \mathbf{1}. \tag{5.6}$$

□

V obecném svazu může existovat víc prvků, které mají vlastnost doplňku prvku a. Jestliže svaz je distributivní, pak doplněk existuje nejvýše jeden.

Tvrzení. Je-li $(M, \wedge, \vee)$ distributivní a k prvku a existuje doplněk, pak tento doplněk je jediný a značíme ho $\overline{a}$.

Toto tvrzení vyplývá z 5.4.7.

5.4.9 Booleova algebra. V řadě aplikací budete pracovat se zvláštní třídou distributivních svazů, s tzv. Booleovou algebrou.

Definice. *Booleova algebra* je distributivní svaz s $\mathbf{0}$ a $\mathbf{1}$, kde každý prvek má doplněk. □

Tvrzení. Je-li $(B, \wedge, \vee, \mathbf{0}, \mathbf{1},)$ Booleova algebra, pak pro každé tři prvky $a, b, c \in B$ platí

a) $\overline{\mathbf{0}} = \mathbf{1}$, $\overline{\mathbf{1}} = \mathbf{0}$.
b) $\overline{a \wedge b} = \overline{a} \vee \overline{b}$, $\overline{a \vee b} = \overline{a} \wedge \overline{b}$.
c) $\overline{\overline{a}} = a$.

Zdůvodnění. Vlastnost a) je zřejmá; stačí si uvědomit, že

$$\mathbf{0} \wedge \mathbf{1} = \mathbf{0} \quad \text{a} \quad \mathbf{0} \vee \mathbf{1} = \mathbf{1}.$$

Také vlastnost c) není těžké nahlédnout. Víme, že

$$a \wedge \overline{a} = \mathbf{0} \quad \text{a} \quad a \vee \overline{a} = \mathbf{1}.$$

Podíváme-li se na tyto dva vztahy „očima“ prvku $\overline{a}$, vidíme, že a je doplněk $\overline{a}$.

Abychom ukázali vlastnost b), spočítáme (využíváme distributivní zákon)

$$(a \wedge b) \wedge (\overline{a} \vee \overline{b}) = (a \wedge b \wedge \overline{a}) \vee (a \wedge b \wedge \overline{b}) = \mathbf{0} \vee \mathbf{0} = \mathbf{0}.$$

Obdobně se ukáže i $(a \wedge b) \vee (\overline{a} \vee \overline{b}) = \mathbf{1}$.

Proto $\overline{a \wedge b} = \overline{a} \vee \overline{b}$.

Druhý vztah se dokáže analogicky.

Poznámka. Z teorie množin víme, že $(P(U), \cap, \cup)$ spolu s nejmenším prvkem $\mathbf{0} = \emptyset$, největším prvkem $\mathbf{1} = U$ a doplňkem $\overline{X}$ tvoří Booleovu algebru. Často se ale budete setkávat i s jinými příklady Booleovy algebry.

5.4.10 Nejmenší Booleova algebra. Nejmenší Booleova algebra má dva prvky, tj. $B = \{0, 1\}$, a platí $i \wedge j = \min\{i, j\}$, $i \vee j = \max\{i, j\}$, $\mathbf{0} = 0$, $\mathbf{1} = 1$ a $\overline{0} = 1$, $\overline{1} = 0$. □

Uvědomte si, že vlastně jedná o charakteristické funkce podmnožin jednoprvkové množiny.

5.4.11 Konečné Booleovy algebry. Všechny konečné Booleovy algebry (až na přejmenování prvků) jsou tohoto tvaru.

Pro každé $n \geq 1$ označme B_n množinu všech n-tic 0 a 1, tj.

$$B_n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in \{0, 1\}\}.$$

Definujeme

a) $(a_1, a_2, \ldots, a_n) \wedge (b_1, b_2, \ldots, b_n) = (a_1 \wedge b_1, a_2 \wedge b_2, \ldots, a_n \wedge b_n)$,
b) $(a_1, a_2, \ldots, a_n) \vee (b_1, b_2, \ldots, b_n) = (a_1 \vee b_1, a_2 \vee b_2, \ldots, a_n \vee b_n)$,
c) $\mathbf{0} = (0, 0, \ldots, 0)$, $\mathbf{1} = (1, 1, \ldots, 1)$,
d) $\overline{(a_1, a_2, \ldots, a_n)} = (\overline{a_1}, \overline{a_2}, \ldots, \overline{a_n})$.

Pak B_n spolu s operacemi tvoří Booleovu algebru o 2^n prvcích. □

Uvědomte si, že vlastně jedná o charakteristické funkce podmnožin n-prvkové množiny.

Všimněte si, že nejmenší Booleova algebra je vlastně B_1. Na B_n, pro $n > 1$, se můžeme dívat jako na kartézský součin n kopií nejmenší Booleovy algebry B.

5.5 Booleovy funkce

V řadě aplikací se v nejmenší Booleově algebře B operace značí jinak; $\wedge$ se označuje $\cdot$ a $\vee$ se označuje $+$. Je třeba však mít na paměti, že pro operace $+$ platí

$$a + b = \max\{a, b\}, \text{ tj. platí } 1 + 1 = 1, \quad a \cdot b = \min\{a, b\}.$$

Při tomto značení platí následující rovnosti (je to obdoba rovností, které platí v Booleově algebře viz výše).

5.5.1 Tvrzení. Pro všechna $x, y, z \in \{0, 1\}$ platí:

1. $x \cdot x = x$, $x + x = x$;
2. $x \cdot y = y \cdot x$, $x + y = y + x$;
3. $x \cdot (y \cdot z) = (x \cdot y) \cdot z$, $x + (y + z) = (x + y) + z$;
4. $x \cdot (y + x) = x$, $x + (y \cdot x) = x$;
5. $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$, $x + (y \cdot z) = (x + y) \cdot (x + z)$;
6. $\overline{\overline{x}} = x$;
7. $\overline{x + y} = \overline{x} \cdot \overline{y}$, $\overline{x \cdot y} = \overline{x} + \overline{y}$;
8. $x + \overline{x} = 1$, $x \cdot \overline{x} = 0$;
9. $x \cdot 0 = 0$, $x \cdot 1 = x$;
10. $x + 1 = 1$, $x + 0 = x$.

□

5.5.2 Booleovy funkce.

Definice. Libovolné zobrazení $f\colon\{0,1\}^n \to \{0,1\}$ se nazývá *Booleova funkce arity n.* □

Velmi často je potřeba takovou funkci f napsat pomocí operací Booleovy algebry. Např. $f(x,y) = (\overline{x}\cdot y)+y$ má hodnoty $f(0,0) = (1\cdot 0)+0 = 0$, $f(0,1) = (1\cdot 1)+1 = 1$, $f(1,0) = (0\cdot 0)+0 = 0$ a $f(1,1) = (0\cdot 1)+1 = 1$.

5.5.3 Disjunktivní normální forma — DNF.

Definice. Řekneme, že Booleova funkce je v *disjunktivní normální formě* (*disjunktivním normálním tvaru*), zkráceně *DNF*, jestliže $f(x_1,\dots,x_n)$ je rovna součtu konečně mnoha součinů proměnných x_i nebo doplňků proměnných $\overline{x_i}$. □

Poznamenejme, že proměnné nebo doplňku proměnné se také říká *literál*.

Příklad. Příkladem Booleovy funkce napsané v disjunktivní normální formě je např. tato funkce

$$f(x,y,z) = (x\cdot\overline{y}\cdot z)+(y\cdot z) = (x\cdot z)+(y\cdot z).$$

5.5.4 Konjunktivní normální forma — CNF.

Definice. Řekneme, že Booleova funkce je v *konjunktivní normální formě* (*konjunktivním normálním tvaru*), zkráceně *CNF*, jestliže $f(x_1,\dots,x_n)$ je rovna součinu konečně mnoha součtů proměnných x_i nebo doplňků proměnných $\overline{x_i}$. □

Příklad. Příkladem Booleovy funkce napsané v konjunktivní normální formě je např. tato funkce

$$f(x,y,z) = z\cdot(x+y+\overline{z}) = z\cdot(x+y).$$

Poznamenejme, že v obou příkladech se jedná o tutéž Booleovu funkci.

5.5.5 K tomu, abychom zapsali danou Booleovu funkci 2 až 4 proměnných buď v disjunktivní nebo konjunktivní formě se hodí Karnaughovy mapy. Na přednášce si ukážeme Karnaughovy mapy pro funkce 3 a 4 proměnných.

5.6 Isomorfismy, homomorfismy.

Pro zvídavé studenty.

5.6.1 Příklad — logaritmus jako isomorfismus grup. Uvažujme grupy $(\mathbb{R}^+,\cdot,1)$ a $(\mathbb{R},+,0)$ (tj. kladná reálná čísla s násobením a reálná čísla se sčítáním). Dále uvažujme zobrazení $f\colon\mathbb{R}^+\to\mathbb{R}$ takové, že $f(x) = \log_{10} x$. Pak $f(x)$ je bijekce množiny $\mathbb{R}^+$ na množinu $\mathbb{R}$, která navíc splňuje

1) $\log_{10}(x\cdot y) = \log_{10}(x)+\log_{10}(y)$,
2) $\log_{10} 1 = 0$,
3) $\log_{10}(x^{-1}) = -\log_{10}(x)$.

Funkce $\log_{10}$ umožnila násobit velká reálná čísla i v době, kdy ještě nebyly známy počítače ani kalkulačky. Když lidé chtěli násobit čísla a,b, nejprve v tabulkách našli $\log_{10}(a)$ a $\log_{10}(b)$. Tyto hodnoty sečetli a dostali $d = \log_{10}(a\cdot b)$. Na závěr číslo d odlogaritmovali — tj. našli vzor v $\log_{10}$ (opět v tabulkách).

5.6.2 Homomorfismus, isomorfismus grup. Máme dány dvě grupy $(G,\circ,e_1)$ a $(H,\star,e_2)$. Zobrazení $f\colon G\to H$ se nazývá *homomorfismus*, jestliže splňuje následující tři podmínky: Pro každé $x,y\in G$ platí

1) $f(x\circ y) = f(x)\star f(y)$,
2) $f(e_1) = e_2$,
3) $f(x^{-1}) = (f(x))^{-1}$.

Jestliže navíc je zobrazení f prosté a na, nazývá se f *isomorfismus.* □

5.6.3 Homomorfismus okruhů s jednotkou. Jsou dány dva okruhy $(M,+,\cdot)$ a $(N,+,\cdot)$. Zobrazení $f\colon M \to N$ se nazývá *homomorfismus* okruhů, jestliže je to homomorfismus jak grup $(M,+)$ do $(N,+)$, tak homomorfismus monoidů $(M,\cdot)$ a $(N,\cdot)$ (tj. platí podmínky 1) a 2)). □

5.6.4 Příklad — využití Čínské věty o zbytcích. Máme nesoudělná čísla $m,n > 1$. Definujme operace $+$ a $\cdot$ na množině $\mathbb{Z}_m \times \mathbb{Z}_n$ takto:

$$(i,j)+(k,l) = (i+k, j+l) \quad \text{a} \quad (i,j)\cdot(k,l) = (i\cdot k.j\cdot l).$$

kde v první složce sčítáme a násobíme v $\mathbb{Z}_m$, ve druhé složce v $\mathbb{Z}_n$.

Pak $(\mathbb{Z}_m \times Z_n, +, \cdot)$ tvoří komutativní okruh s jednotkou, kde $(0,0)$ je neutrální prvek vzhledem ke sčítání a $(1,1)$ je neutrální prvek vzhledem k násobení.

Čínská věta o zbytcích vlastně říká, že zobrazení $f\colon \mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n$, definované

$$f(a) = (a_m, a_n) \quad \text{kde} \quad a_m \equiv a \,(\mathrm{mod}\, m), a_n \equiv a \,(\mathrm{mod}\, n)$$

je isomorfismus okruhů. Navíc isomorfismus okruhů je také isomorfismus grup invertibilních prvků $(\mathbb{Z}^\star_{mn}, \cdot)$ a $(\mathbb{Z}^\star_m \times \mathbb{Z}^\star_n, \cdot)$. Proto

$$\phi(mn) = |\mathbb{Z}^\star_{mn}| = |\mathbb{Z}^\star_m| \cdot |\mathbb{Z}^\star_n| = \phi(m)\cdot\phi(n).$$

Ukázali jsem třetí vlastnost Eulerovy funkce.

Kapitola 6

Grafy

Teorie grafů je matematická disciplína, která má velmi mnoho aplikací a to nejen v technice. My se v této kapitole seznámíme pouze se základními pojmy a ukážeme si několik aplikací. Studentovi, který by se o grafech chtěl dozvědět více, doporučujeme knížku Kapitoly z diskrétní matematiky autorů J. Matoušek, J. Nešetřil, nebo knížku Grafy a jejich aplikace od J. Demla.

6.1 Orientované a neorientované grafy

6.1.1 Orientovaný graf. Zhruba řečeno, orientovaný graf se skládá z prvků, které nazýváme vrcholy, a orientovaných spojnic, které vedou z jednoho vrcholu do druhého vrcholu a nazýváme je (orientované) hrany. Protože z vrcholu u do vrcholu v mohou vést dvě a více hran, musíme v obecném případě hrany rozlišovat jmény. Přesněji:

Definice. *Orientovaný graf* G je trojice (V, E, ε), kde V je neprázdná konečná množina *vrcholů* (též zvaných *uzlů*), E je konečná množina jmen *hran* (též zvaných *orientovaných hran*) a ε je přiřazení, které každé hraně $e \in E$ přiřazuje uspořádanou dvojici vrcholů a nazývá se *vztah incidence*. □

Další pojmy spojené s orientovanými grafy. Jestliže $\varepsilon(e) = (u, v)$ pro $u, v \in V$, říkáme, že vrchol u je *počáteční vrchol* hrany e a vrchol v je *koncový vrchol* hrany e; značíme $PV(e) = u$ a $KV(e) = v$. O vrcholech u, v říkáme, že jsou *krajní vrcholy* hrany e, též že jsou *incidentní* s hranou e.

Jestliže počáteční a koncový vrchol jsou stejné, říkáme, že hrana e je *orientovaná smyčka*. □

6.1.2 Neorientovaný graf. Zhruba řečeno, neorientované grafy jsou grafy, kde každá hrana může být použita „oběma směry“. Přesněji:

Definice. *Neorientovaný graf* je trojice $G = (V, E, \varepsilon)$, kde V je neprázdná konečná množina *vrcholů* (též zvaných *uzlů*), E je konečná množina jmen *hran* a ε je přiřazení, které každé hraně $e \in E$ přiřazuje množinu $\{u, v\}$ (kde $u, v \in V$ jsou vrcholy) a nazývá se *vztah incidence*. □

Další pojmy spojené s neorientovanými grafy. Jestliže $\varepsilon(e) = \{u, v\}$ pro $u, v \in V$, říkáme, že u, v jsou *krajní vrcholy* hrany e, též že jsou *incidentní* s hranou e.

Je-li $u = v$, říkáme že e je *(neorientovaná) smyčka*. □

6.1.3 Paralelní hrany.

Definice. Jestliže v orientovaném nebo neorientovaném grafu existují dvě různé hrany e_1, e_2, pro které $\varepsilon(e_1) = \varepsilon(e_2)$, říkáme, že hrany e_1, e_2 jsou *paralelní*. □

Uvědomte si, že pro orientované grafy to znamená, že počáteční vrcholy i koncové vrcholy hran e_1, e_2 jsou stejné, zatímco pro neorientované grafy to pouze znamená, že krajní vrcholy hran e_1, e_2 jsou stejné.

6.1.4 Prostý graf.

Definice. Graf (orientovaný nebo neorientovaný) se nazývá *prostý graf*, jestliže nemá paralelní hrany. □

Poznámky. V některé literatuře se termínem *graf* rozumí prostý graf a grafům, ať již orientovaným nebo neorientovaným, které mají paralelní hrany, se říká *multigrafy*. Vzhledem k tomu, že v řadě aplikací hrají paralelní hrany podstatnou roli, my tuto terminologii nepoužíváme.

V případě prostých grafů nepotřebujeme speciální jména hran a můžeme v případě orientovaných grafů hranu e pojmenovat (u, v), jestliže $PV(e) = u$ a $KV(e) = v$; v případě neorientovaných grafů $\{u, v\}$, jestliže $\varepsilon(e) = \{u, v\}$. Proto prosté grafy označujeme pouze jako dvojici (V, E).

V dalším textu (není-li řečeno jinak) $V(G)$ označuje množinu vrcholů grafu G a $E(G)$ množinu hran grafu G.

6.1.5 Stupně vrcholů. Zhruba řečeno stupeň vrcholu v je počet hran, které začínají nebo končí ve v. Pro orientované grafy ještě rozlišujeme vstupní stupeň — to je počet hran, které do vrcholu vstupují, a výstupní stupeň — to je počet hran, které z vrcholu vycházejí. Přesněji:

Definice. Je dán orientovaný graf $G = (V, E, \varepsilon)$. *Vstupní stupeň* vrcholu v, značíme jej $d^-(v)$, je roven počtu hran, pro které je v koncovým vrcholem (které do vrcholu v vstupují), tj.

$$d^-(v) = |\{e \in E; KV(e) = v\}|.$$

Výstupní stupeň vrcholu v, značíme jej $d^+(v)$, je roven počtu hran, pro které je v počátečním vrcholem (které z vrcholu v vystupují), tj.

$$d^+(v) = |\{e \in E; PV(e) = v\}|.$$

Stupeň vrcholu v, značíme jej $d(v)$, je roven počtu hran, které jsou incidentní s vrcholem v, tj.

$$d(v) = d^-(v) + d^+(v).$$

Je dán neorientovaný graf $G = (V, E, \varepsilon)$. *Stupeň* vrcholu v, značíme jej $d(v)$, je roven počtu hran, které jsou incidentní s vrcholem v, kde smyčka je počítána dvakrát. □

Všimněte si, že v definici stupně vrcholu je smyčka započítána dvakrát i v případě orientovaných grafů. Proto, „zapomeneme-li v orientovaném grafu na orientaci“ stupně vrcholů zůstanou stejné.

6.1.6 Tvrzení. Pro každý graf G (orientovaný nebo neorientovaný) platí

$$\sum_{v \in V} d(v) = 2\,|E|,$$

kde $|E|$ značí počet hran grafu G. □

Zdůvodnění. Každá hrana grafu, která není smyčka, zvýší stupeň dvou svých krajních vrcholů. Z definice pak smyčka také zvyšuje stupeň vrcholu o 2.

Důsledek. Každý graf má sudý počet vrcholů lichého stupně. □

Zdůvodnění. Je-li součet stupňů sudé číslo, musí mezi stupni být sudý počet lichých čísel.

6.1.7 Zadávání grafu. Orientovaný i neorientovaný graf můžeme zadat seznamem jeho vrcholů, a pro každý vrchol seznamem hran, které v něm začínají a/nebo v něm končí. Graf též můžeme zadat maticí sousednosti nebo maticí incidence.

6.1.8 Matice sousednosti. Je dán graf $G = (V, E, \varepsilon)$, kde jsme očíslovali vrcholy, tj. $V = \{v_1, v_2, \ldots, v_n\}$. Čtvercová matice $\mathbf{M} = (m(i,j))$ řádu n se nazývá *matice sousednosti* grafu G, splňuje-li:

- Pro orientovaný graf je $m(i,j)$ roven počtu hran, pro něž je v_i počáteční vrchol a v_j koncový vrchol.
- Pro neorientovaný graf je $m(i,j)$ roven počtu hran s krajními vrcholy v_i a v_j.

□

Poznamenejme, že pro neorientovaný graf je matice sousednosti symetrická.

6.1.9 Matice incidence. Je dán graf $G = (V, E, \varepsilon)$ bez smyček. Očíslujme vrcholy $V = \{v_1, v_2, \ldots, v_n\}$ a hrany $E = \{e_1, e_2, \ldots, e_m\}$. Matice $\mathbf{B} = (b(i,j))$ typu (n, m) se nazývá *matice incidence* grafu G, splňuje-li:

- Pro orientovaný graf je

$$b(i,j) = \begin{cases} 1, & \text{jestliže } v_i \text{ je počáteční vrchol hrany } e_j, \\ -1, & \text{jestliže } v_i \text{ je koncový vrchol hrany } e_j, \\ 0, & \text{v ostatních případech.} \end{cases}$$

- Pro neorientovaný graf je

$$b(i,j) = \begin{cases} 1, & \text{jestliže } v_i \text{ je krajní vrchol hrany } e_j, \\ 0, & \text{v ostatních případech.} \end{cases}$$

□

Matice incidence orientovaného grafu má v každém sloupci jednu 1 a jednu -1, pro neorientované grafy má v každém sloupci dvě 1.

6.1.10 Porovnávání grafů.

Definice. Řekneme, že dva grafy $G = (V, E, \varepsilon)$ a $G' = (V', E', \varepsilon')$ jsou si *rovny*, jestliže $V = V'$, $E = E'$ a $\varepsilon = \varepsilon'$.

Řekneme, že dva grafy $G = (V, E, \varepsilon)$ a $G' = (V', E', \varepsilon')$ jsou *isomorfní*, jestliže existují bijekce $f\colon V \to V'$ a $g\colon E \to E'$ takové, že pro orientované grafy

$$\varepsilon(e) = (u, v) \quad \text{právě tehdy, když} \quad \varepsilon'(g(e)) = (f(u), f(v))$$

a pro neorientované grafy

$$\varepsilon(e) = \{u, v\} \quad \text{právě tehdy, když} \quad \varepsilon'(g(e)) = \{f(u), f(v)\}.$$

□

6.1.11 Sled v orientovaném grafu. V mnoha aplikacích teorie grafů potřebujeme „chodit“ v grafu. Zavedeme tři pojmy — sled, tah a cesta. Sled je ten nejobecnější pojem, v něm pouze požadujeme, abychom vždy procházeli hranu z vrcholu této hrany. Tah bude speciální případ sledu, kde každou hranou smíme projít nejvýše jednou. Nejvíc omezující je pojem cesta — zde smíme vejít dvakrát do téhož vrcholu pouze v případě, že je to poslední vrchol a vešli jsme do prvního vrcholu. Přesněji:

Definice. Je dán orientovaný graf $G = (V, E, \varepsilon)$. *Orientovaný sled* v G je posloupnost vrcholů a hran

$$v_1, e_1, v_2, e_2, \ldots, v_{k-1}, e_{k-1}, v_k$$

taková, že pro každé $i = 1, 2, \ldots, k-1$ platí $v_i = PV(e_i)$ a $v_{i+1} = KV(e_i)$.

Neorientovaný sled v G je posloupnost vrcholů a hran

$$v_1, e_1, v_2, e_2, \ldots, v_{k-1}, e_{k-1}, v_k$$

taková, že pro každé $i = 1, 2, \ldots, k-1$ platí že vrcholy v_i a v_{i+1} jsou krajní vrcholy hrany e_i. □

Poznámka: Neorientovaný sled se od orientovaného sledu liší tím, že můžeme „jít i proti směru“ hrany.

6.1.12 Sled v neorientovaném grafu.

Definice. Je dán neorientovaný graf. Pak *neorientovaný sled* je posloupnost vrcholů a hran

$$v_1, e_1, v_2, e_2, \ldots, v_{k-1}, e_{k-1}, v_k$$

taková, že hrana e_i je incidentní s vrcholy v_i a v_{i+1} pro všechny $i = 1, 2, \ldots, k-1$. □

Triviální sled je sled, který obsahuje jediný vrchol a žádnou hranu. Považujeme ho jak za orientovaný, tak za neorientovaný.

6.1.13 Uzavřené sledy.

Máme dán orientovaný nebo neorientovaný sled $v_1, e_1, \ldots, e_{k-1}, v_k$. Říkáme, že vrchol v_1 je *počátečním vrcholem sledu* a v_k je *koncovým vrcholem sledu*. Též říkáme, že sled *vede z vrcholu* v_1 *do vrcholu* v_k.

Definice. Orientovaný (neorientovaný) sled se nazývá *uzavřený*, jestliže $v_1 = v_k$ a navíc $k > 1$. V opačném případě mluvíme o *otevřeném* sledu. □

Tedy triviální sled nepovažujeme za uzavřený.

6.1.14 Tah a cesta.

Ukazuje se výhodné definovat (kromě základního sledu) ještě speciální typy sledů — tahy a cesty.

Definice. Orientovaný (neorientovaný) sled nazýváme orientovaným (neorientovaným) *tahem*, jestliže se v něm neopakují hrany.

Orientovaný (neorientovaný) tah je *cestou*, jestliže se v něm neopakují vrcholy s tou výjimkou, že může být uzavřený, tj. může být $v_1 = v_k$. Uzavřená orientovaná cesta se nazývá *cyklus*, uzavřená neorientovaná cesta se nazývá *kružnice*. □

Poznámka. Každá cesta je zároveň tahem i sledem, naopak to ale neplatí. Také každý cyklus je zároveň kružnicí, ale ne každá kružnice je cyklem.

Poznamenejme, že triviální sled je též tahem i cestou **není** však ani kružnicí ani cyklem.

6.1.15 Dostupnost.

Definice. Máme dán graf $G = (V, E, \varepsilon)$. Řekneme, že vrchol v je *orientovaně (neorientovaně) dostupný* z vrcholu w, jestliže existuje orientovaná (neorientovaná) cesta z w do v. □

Poznámka. V definici dostupnosti jsme mohli požadovat existenci sledu (místo cesty) a dostali bychom stejný pojem. Když totiž existuje orientovaný (neorientovaný) sled z vrcholu w do vrcholu v, pak také existuje orientovaná (neorientovaná) cesta z vrcholu w do vrcholu v. Rozmyslete a zdůvodněte si tento fakt.

6.1.16 Souvislý graf.

Definice. Řekneme, že (orientovaný nebo neorientovaný) graf je *souvislý*, jestliže pro každé dva vrcholy u, v grafu existuje neorientovaná cesta z u do v. □

Poznámka. Vždy existuje cesta z vrcholu u do sebe – totiž triviální cesta. Také platí, že neorientovaná cesta z vrcholu u do vrcholu v je také neorientovanou cestou z v do u (stačí cestu „číst pozpátku“).

6.2 Stromy

Pojem teorie grafů, který se v aplikacích velmi často vyskytuje, je pojem stromu. Setkáte se s ním např. v datových strukturách, ale i leckde jinde.

6.2.1 Definice. Orientovaný nebo neorientovaný graf se nazývá *strom*, je-li souvislý a neobsahuje-li kružnici. □

Tvrzení. V každém stromu s alespoň dvěma vrcholy existuje vrchol stupně 1. □

Zdůvodnění. Ukážeme, že kdyby v nějakém grafu měl **každý** vrchol stupeň alespoň 2, pak by graf obsahoval kružnici.

Předpokládejme, že graf G s $n > 1$ vrcholy má stupně všech vrcholů větší nebo rovno 2. Začneme konstruovat cestu takto: Začneme z libovolného vrcholu, označme jej v_1. Protože v_1 má stupeň aspoň 2, vede z něj hrana, označíme ji e_1. Druhý krajní vrchol hrany e_1 nemůže být v_1 (byla by to smyčka a smyčka je zvláštní případ kružnice). Označme tento nový vrchol v_2. Vrchol v_2 má stupeň aspoň 2, musí být krajním vrcholem nějaké další hrany, řekněme e_2. Hrana e_2 má druhý krajní vrchol v_3; a ten nemůže být ani v_1, ani v_2 (v obou případech by graf obsahoval kružnici). Takto postupujeme až dostaneme cestu o $n-1$ hranách, která obsahuje všechny vrcholy grafu G. Její koncový vrchol má ovšem také stupeň aspoň dvě, musí z něj proto vycházet ještě jiná hrana než e_{n-1}. A protože už nemáme další nepoužitý vrchol, musí uzavřít kružnici. A to je spor s faktem, že G nemá kružnice.

6.2.2 Věta. Každý strom o n vrcholech má $n-1$ hran. □

Zdůvodnění: Větu je možné dokázat indukcí podle počtu vrcholů n. Tvrzení zřejmě platí pro stromy o jednom nebo dvou vrcholech.

Předpokládejme, že tvrzení věty platí pro všechny stromy s n vrcholy, $n > 2$. Vezměme libovolný strom G s $n+1$ vrcholy. Označme G' strom, který dostaneme tak, že z G odstraníme vrchol stupně 1 (podle předchozího tvrzení takový vrchol existuje). Graf G' je opět strom a má n vrcholů, tedy obsahuje přesně $n-1$ hran (indukční předpoklad). Proto G má $n-1+1=n$ hran.

6.2.3 Důsledek. V každém stromu s alespoň dvěma vrcholy existují (alespoň) dva vrcholy stupně 1. □

Zdůvodnění: Má-li souvislý graf n vrcholů ($n > 1$) a $n-1$ hran, nemůže mít jen jeden vrchol stupně 1. Ano, pak by platilo

$$\sum_{v \in V} d(v) \geq 1 + 2(n-1) > 2(n-1) = 2|E|,$$

a to není možné.

Jiný způsob zdůvodnění: Není těžké nahlédnout, že vezmeme-li v grafu, který nemá kružnice, cestu o největším počtu hran, pak oba koncové vrcholy této cesty musí mít stupeň 1.

6.2.4 Poznámka. Mějme souvislý graf G. Přidáme-li k němu hranu (aniž bychom zvětšili množinu vrcholů), zůstane graf souvislý.

Mějme graf G bez kružnic. Odebereme-li z grafu G hranu, vzniklý graf opět nebude obsahovat kružnici.

Strom je graf, který má nejmenší počet hran, aby mohl být souvislý, a současně největší počet hran, aby v něm neexistovala kružnice. Poznamenejme, že přidáním hrany zde rozumíme přidání hrany mezi již existující vrcholy; další vrcholy nepřidáváme.

6.2.5 Tvrzení. Je dán graf G, pak následující je ekvivalentní.

1. G je strom.
2. G nemá kružnice a přidáme-li ke grafu libovolnou hranu uzavřeme přesně jednu kružnici.
3. G je souvislý a odebráním libovolné hrany přestane být souvislý.

□

6.2.6 Podgrafy. Zhruba řečeno, podgraf grafu G dostaneme tak, že z grafu G vynecháme některé (nebo žádné) vrcholy a některé (nebo žádné) hrany a to tak, že necháme-li v podgrafu hranu e, pak tam necháme i oba jeho krajní vrcholy.

Mezi podgrafy nás zajímají hlavně dva speciální typy: Faktor je podgraf, kde jsme ponechali všechny vrcholy. Podgraf indukovaný množinou vrcholů A je podgraf s množinou vrcholů A, kde jsme ponechali všechny hrany grafu G, které jsme mohli ponechat.

Definice. Je dán graf $G = (V, E, \varepsilon)$. *Podgraf* grafu G je trojice $G' = (V', E', \varepsilon')$, kde $V' \subseteq V$, $E' \subseteq E$ a ε' je restrikce ε na množině E', taková, že trojice $G' = (V', E', \varepsilon')$ je také graf.

Podgraf $G' = (V', E', \varepsilon')$ se nazývá *faktor* grafu G, jestliže $V' = V$.

Podgraf $G' = (A, E', \varepsilon')$ se nazývá *podgraf indukovaný množinou* A, $A \subseteq V$, jestliže každá hrana grafu G, která má oba krajní vrcholy v množině A, leží v E'. Podgraf indukovaný množinou A se též nazývá *úplný podgraf na množině* A. □

6.2.7 Komponenty souvislosti. Víme, že ne každý graf je souvislý. Na druhé straně vrcholy každého grafu se dají rozdělit na části, na kterých už graf souvislý je; těm maximálním množinám vrcholů říkáme komponenty souvislosti. Přesněji:

Definice. Máme dán (orientovaný nebo neorientovaný) graf G. *Komponenta souvislosti* (někdy též *komponenta slabé souvislosti*) je maximální množina vrcholů A taková, že indukovaný podgraf určený A je souvislý. □

Maximální množinou zde rozumíme takovou množinu A, pro kterou platí, že přidáme-li k množině A libovolný vrchol, podgraf indukovaný touto větší množinou už souvislý nebude.

Uvědomte si, že komponenty souvislosti jsou vlastně třídy ekvivalence (neorientované) dostupnosti na množině vrcholů V. Ano, relace dostupnosti je reflexivní, symetrická a tranzitivní, proto se jedná o ekvivalenci. Třídy této ekvivalence jsou pak maximální množiny vrcholů, na kterých je graf G souvislý.

Poznámka. Graf je souvislý právě tehdy, když má jedinou komponentu souvislosti.

6.3 Minimální kostra

Z minulé kapitoly víme, že stromy jsou souvislé grafy s nejmenším počtem hran. Vlastnosti stromů se využívá v řadě aplikací, kde úkolem je spojit n míst s co nejmenšími náklady. V takovém případě vytvoříme graf, jehož vrcholy jsou místa, která potřebujeme spojit, a hrany jsou spojení mezi místy. Navíc hrany ještě ohodnotíme tak, že ohodnocení nám udává, jak drahé (na výstavbu, údržbu, atd.) je přímé spojení. Pracujeme proto často s tzv. ohodnocenými grafy, tj. s grafy, kde pro každou hranu e je dáno číslo $c(e)$ – její ohodnocení.

6.3.1 Kostra grafu. Definice. Je dán graf G. Faktor grafu G, který je stromem, se nazývá *kostra* grafu G. □

6.3.2 Tvrzení. Graf G má kostru právě tehdy, když je souvislý. □

Zdůvodnění. Jestliže má graf kostru, musí být souvislý; ano, kostra je jeho souvislý podgraf.

Předpokládejme, že graf G je souvislý. Pak buď nemá kružnici, pak je to strom a je svou vlastní kostrou.

Předpokládejme proto, že kružnici má. Odstraníme-li z grafu jednu hranu e kružnice, dostaneme opět souvislý graf G_1. Jestliže G_1 nemá kružnici, je to kostra grafu G. V opačném případě opět vynecháme jednu hranu z kružnice grafu G_1. Tímto způsobem jednou (po odebrání $|E(G)| - |V(G)| + 1$ hran) dostaneme souvislý graf, který už nemá kružnice. A to je kostra grafu G.

6.3.3 Minimální kostra. Je dán souvislý graf G spolu s ohodnocením hran c, tj. pro každou hranu $e \in E(G)$ je dáno číslo $c(e)$ (číslo $c(e)$ nazýváme *cenou hrany* e).

Definice. *Minimální kostra* grafu $G = (V, E)$ je taková kostra grafu $K = (V, L)$, že $\sum_{e \in L} c(e)$ je nejmenší (mezi všemi kostrami grafu G). □

Číslu $\sum_{e \in L} c(e)$ říkáme *cena kostry* L.

6.3.4 Tvrzení. V každém souvislém ohodnoceném grafu existuje minimální kostra. Nemusí však být jediná. □

Zdůvodnění. Všech koster souvislého grafu je konečný počet, proto některá musí mít nejmenší cenu.

Je-li např. ohodnocení hran stejné pro každou hranu, pak každá kostra je minimální kostra. Ano, cena každé kostry je $(|V| - 1) \cdot c(e)$.

6.3.5 Algoritmus pro nalezení minimální kostry. Existuje řada algoritmů, které pro daný souvislý ohodnocený graf najdou jeho minimální kostru. My uvedeme jeden z nich; jedná se o ukázku tzv. hladového algoritmu, tj. algoritmu, kde se v každém okamžiku rozhodujeme tím pro nás nejvýhodnějším způsobem. Uvědomte si, že ve většině problémů (na rozdíl od problému nalezení minimální kostry), hladový postup nevede k nejlepšímu řešení.

Kruskalův algoritmus.

Vstup: Souvislý graf $G = (V, E)$ a ohodnocení hran c.
Výstup: Množina hran minimální kostry T.

1. Setřídíme hrany podle ceny do neklesající posloupnosti, tj.

$$c(e_1) \leq c(e_2) \leq \ldots \leq c(e_m)$$

   Položíme $L = \emptyset$, $\mathcal{S} = \{\{v\}; v \in V\}$.

2. Probíráme hrany v daném pořadí. Hranu e_i přidáme do L, jestliže má oba krajní vrcholy v různých množinách $S, S' \in \mathcal{S}$. V systému $\mathcal{S}$ množiny S a S' nahradíme jejich sjednocením. V opačném případě hranu přeskočíme.

3. Algoritmus končí, jestliže jsme přidali $n - 1$ hran (tj. $\mathcal{S}$ se skládá z jediné množiny).

□

Poznámky.

1) Jestliže na vstupu Kruslalova algoritmu je nesouvislý graf, algoritmus projde všechny hrany aniž by do T zařadil $n - 1$ hran. Nemusíme tedy napřed testovat, zda daný graf je souvislý.

2) Má-li několik hran stejnou cenu, v kroku 1 se mohou tyto hrany vyskytovat v libovolném pořadí. Na tomto pořadí pak záleží, kterou z minimálních koster dostaneme.

6.4 Kořenové stromy

V této sekci se zabýváme grafy s kořenem a speciálně s orientovanými grafy, které jsou stromy a mají kořen. Takové grafy jsou základem řady datových struktur.

6.4.1 Kořen.

Definice. Je dán orientovaný graf $G = (V, E, \varepsilon)$. Řekneme, že vrchol $r \in V$ je *kořen* grafu G, jestliže pro každý vrchol $v \in V$ existuje orientovaná cesta z r do v. □

Jinými slovy, vrchol r je kořen grafu G právě tehdy, když každý vrchol grafu G je orientovaně dostupný z vrcholu r.

6.4.2 Poznámka. Uvědomte si, že pro vrchol r existuje orientovaná cesta z r do sebe – totiž triviální cesta.

Každý orientovaný graf, který má kořen, je souvislý. Naopak to neplatí; existují souvislé orientované grafy, které nemají kořen.

Orientovaný graf může mít i několik kořenů; např. v cyklu je každý vrchol kořenem.

6.4.3 Kořenový strom.

Definice. Orientovaný graf, který má kořen a je strom, se nazývá *kořenový strom*. □

Protože každý graf který má kořen je souvislý, mohli jsme kořenový strom definovat jako graf, který má kořen a nemá kružnice.

6.4.4 Tvrzení. Je-li G kořenový strom, pak má pouze jeden kořen. □

Zdůvodnění. Kdyby nějaký strom měl dva kořeny, řekněme r_1 a r_2, pak by existovala orientovaná cesta z r_1 do r_2, a také orientovaná cesta z r_2 do r_1. Spojením těchto dvou cest bychom dostali uzavřený orientovaný sled, a ten vždy obsahuje cyklus, tedy i kružnici, a to strom obsahovat nemůže.

Poznámka. Je dán kořenový strom s kořenem r. Pak do každého vrcholu v vede přesně jedna orientovaná cesta z r. Ano, existence jedné cesty je zajištěna z definice kořenového stromu; a kdyby existovaly cesty dvě, pak bychom v G měli uzavřený (neorientovaný) sled, tj. i kružnici. (Uvědomte si, že každý uzavřený sled obsahuje kružnici.) A to strom nemůže.

6.4.5 Následník, předchůdce a list. Definice. Je dán kořenový strom $G = (V, E)$. Jestliže (u, v) je hrana grafu G, pak říkáme, že vrchol u je *předchůdce* vrcholu v a vrchol v je *následník* vrcholu u. Vrchol, který nemá následníka, se nazývá *list*. □

6.4.6 Hladiny a výška kořenového stromu. Definice. Je dán kořenový strom $G = (V, E)$ s kořenem r. Řekneme, že vrchol v *leží v hladině* k, jestliže orientovaná cesta z r do v má přesně k hran.

Výška kořenového stromu je největší k takové, že k-tá hladina je neprázdná. □

Víme, že pro každý vrchol v v kořenovém stromě existuje právě jedna orientovaná cesta z kořene r do vrcholu v. Proto jsou hladiny kořenového stromu korektně definované.

Výšku kořenového stromu jsme také mohli definovat jako počet hran v nejdelší orientované cestě (ta musí vést z kořene do některého z listů).

6.4.7 Podstrom určený vrcholem. Definice. Je dán kořenový strom G. *Podstrom určený vrcholem* v je podgraf G indukovaný množinou všech vrcholů, které jsou orientovaně dostupné z vrcholu v. □

Poznámka. Uvědomte si, že podstrom určený vrcholem v je sám kořenovým stromem a jeho kořen je v.

6.4.8 Binární kořenové stromy. Definice. Kořenový strom se nazývá *binární kořenový strom*, jestliže každý vrchol má nejvýše dva následníky. □

V binárním kořenovém stromě mluvíme o *pravém* a *levém následníku* vrcholu. *Levý podstrom*, resp. *pravý podstrom* vrcholu v je podstrom určený levým, resp. pravým následníkem vrcholu v.

6.4.9 Halda. Jednou z četných aplikací kořenových stromů je datová struktura zvaná *halda*. Je např. základem algoritmu Heapsort pro řazení.

Halda je stromová datová struktura s touto vlastností: Je-li vrchol v orientovaně dostupný z vrcholu x, pak číselná hodnota ve vrcholu v je větší nebo rovna číselné hodnotě ve vrcholu x. Navíc se jedná o úplný binární strom; tj. každý vrchol s výjimkou listů a vrcholů v předposlední hladině

má vždy dva následníky, a jestliže v předposlední hladině má vrchol pouze jeden následník, pak je to levý následník a všechny vrcholy „vpravo“ od něho jsou již listy.

Z definice haldy je zřejmé, že nejmenší hodnotu má kořen haldy, proto je nalezení minima velmi jednoduché. Musí se však vyřešit dvě úlohy — a to odstranění kořene a vložení prvku do haldy. Obě tyto operace je možné provést v čase úměrném $\log_2 n$, kde n je počet prvků, které má halda.

Velkou výhodou haldy je to, že ji v počítači nemusíme držet jako stromovou strukturu, ale můžeme se v haldě „pohybovat“ pomocí násobení dvěma a celočíselného dělení dvěma.

6.5 Acyklické grafy

Stromy jsou (orientované nebo neorientované) grafy, které nemají kružnice. Acyklické grafy jsou orientované grafy, které nemají cykly (kružnice ale mít mohou). Acyklické grafy již nemají tak „hezké“ vlastnosti jako stromy — nemůžeme určit počet hran, které mohou mít, na základě počtu vrcholů. Přesto o nich platí řada zajímavých tvrzení.

6.5.1 Definice. Orientovaný graf se nazývá *acyklický*, jestliže neobsahuje žádný cyklus. □

Poznamenejme, že acyklický graf může obsahovat kružnice; nakreslete si příklad grafu, který obsahuje kružnici, ale je acyklický. Ukážeme, že acyklické grafy se dají definovat i jiným způsobem — jsou to grafy, které mají tzv. topologické očíslování vrcholů a/nebo topologické očíslování hran.

6.5.2 Topologické očíslování vrcholů.

Definice. Je dán orientovaný graf $G = (V, E, \varepsilon)$ s n vrcholy. Očíslování vrcholů

$$v_1, v_2, \ldots, v_n$$

se nazývá *topologické očíslování*, jestliže pro každou hranu e s počátečním vrcholem v_i a koncovým vrcholem v_j platí $i < j$. □

Jinými slovy, hrany musí vést vždy z vrcholu s menším indexem do vrcholu s větším indexem.

6.5.3 Topologické očíslování hran.

Definice. Je dán orientovaný graf $G = (V, E, \varepsilon)$ s m hranami. Očíslování hran

$$e_1, e_2, \ldots, e_m$$

se nazývá *topologické očíslování*, jestliže pro každé dvě hrany e_i, e_j pro které koncový vrchol hrany e_i je počátečním vrcholem hrany e_j platí $i < j$. □

Jinými slovy, kdykoli hrana e' navazuje na hranu e, musí být v posloupnosti hrana e vypsána dříve než hrana e'.

6.5.4 Poznámka. Definici topologického očíslování hran jsme mohli formulovat i následujícím způsobem:

Pro každý vrchol v platí: vypisujeme-li do posloupnosti libovolnou hranu s počátečním vrcholem v, musely již být vypsány všechny hrany, které ve vrcholu v končí.

6.5.5 Tvrzení. V každém acyklickém grafu existuje vrchol, který má vstupní stupeň roven 0. □

Myšlenka zdůvodnění je obdobná jako v důkazu tvrzení 6.2.1. Kdyby totiž každý vrchol nějakého grafu měl vstupní stupeň alespoň 2, pak bychom (podobnou úvahou jako v důkazu faktu, že každý strom o $n > 1$ vrcholech má aspoň jeden vrchol stupně 1) dostali cyklus.

6.5.6 Věta. Pro orientovaný graf G jsou následující podmínky ekvivalentní:

1. G je acyklický;
2. G má topologické očíslování vrcholů;
3. G má topologické očíslování hran.

□

Nástin zdůvodnění. Není těžké ukázat, že má-li graf topologické očíslování vrcholů, má i topologické očíslování hran – stačí hrany vypisovat takto: Nejprve vypíšeme hrany začínající v prvním vrcholu (topologického očíslování), pak v druhém vrcholu, atd. (Uvědomte si, že z posledního vrcholu topologického očíslování vrcholů už žádná hrana nevede.)

Obdobně se ukáže i opačná implikace; totiž, že graf, který má topologické očíslování hran, má i topologické očíslování vrcholů. Vypisujeme počáteční vrcholy hran v topologickém očíslování hran a to vždy první výskyt daného vrcholu. Na konci nám zůstanou některé vrcholy (aspoň jeden) a ty pak vypíšeme na konec v libovolném pořadí.

Také je zřejmé, že cyklus není možné topologicky uspořádat. Tudíž, má-li graf cyklus, nemá topologické očíslování (ani vrcholů, ani hran). Fakt, že acyklický graf má topologické očíslování vrcholů ukážeme následujícím algoritmem.

6.5.7 Postup na nalezení topologického očíslování vrcholů.

Vstup: Acyklický graf $G = (V, E)$ s n vrcholy.

Výstup: Topologické očíslování vrcholů $v_0, v_1, \ldots, v_n$.

1) Pro každý vrchol v spočítáme vstupní stupeň $d^-(v)$.
2) Do množiny M dáme všechny vrcholy se vstupním stupněm 0;
   Položíme $i := 1$.
3) Dokud $M \neq \emptyset$ provedeme

   3a) Vybereme vrchol v z množiny M a odstraníme ho z M.
   Položíme $v_i := v$, $i := i + 1$.

   3b) Pro každou hranu e s $PV(e) = v$ provedeme

$$d^-(KV(e)) := d^-(KV(e)) - 1$$

   a v případě, že $d^-(KV(e)) = 0$, přidáme vrchol $KV(e)$ do množiny M.

6.5.8 Věta. Postup 6.5.7 skončí po konečně mnoha krocích a po skončení je posloupnost $v_1, \ldots, v_n$ topologické očíslování vrcholů grafu G. □

Nástin zdůvodnění. Algoritmus určitě skončí, protože vyjmeme-li vrchol z množiny M, již ho nikdy do množiny M nevrátíme. (Také každou hranou se nezabýváme víc než jedenkrát).

Fakt, že jsme našli topologické očíslování vrcholů vyplývá z toho, že jsme vždy vypisovali vrchol, který (ve zbývajícím) podgrafu měl vstupní stupeň 0.

6.5.9 Poznámka. Kdybychom chtěli získat topologické očíslování hran, stačilo by vypisovat hrany do posloupnosti v pořadí, jak je zpracováváme v kroku 3b).

6.6 Silná souvislost

6.6.1 Silně souvislé grafy. Definice. Řekneme, že orientovaný graf G je *silně souvislý*, jestliže pro každou dvojici vrcholů u, v existuje orientovaná cesta z vrcholu u do vrcholu v a orientovaná cesta z vrcholu v do vrcholu u. □

6.6.2 Poznámka. V definici silně souvislého grafu jsme mohli požadovat pouze existenci orientované cesty z vrcholu u do vrcholu v. Je to proto, že existenci takové cesty vyžadujeme pro všechny dvojice vrcholů, tedy i pro dvojici v, u.

6.6.3 Tvrzení. Souvislý graf je silně souvislý právě tehdy, když každá hrana leží v nějakém cyklu. □

Zdůvodnění. Předpokládejme, že graf G je silně souvislý a vyberme jeho libovolnou hranu e s $PV(e) = x$, $KV(e) = y$. Protože je graf silně souvislý, existuje orientovaná cesta z vrcholu y do vrcholu x. Přidáme-li k této cestě hranu e, dostaneme cyklus, který e obsahuje.

Předpokládejme, že graf G je souvislý a každá jeho hrana je obsažena v některém cyklu. Vyberme libovolně dva vrcholy u, v v grafu. Protože je graf souvislý, existuje neorientovaná cesta z vrcholu u do vrcholu v. Každou hranu v této cestě, kterou jde cesta „proti směru hrany“, nahraďme částí cyklu, který tuto hranu obsahuje. Tím dostaneme orientovaný sled z u do v, a ten jistě obsahuje orientovanou cestu. Ukázali jsme, že G je silně souvislý.

6.6.4 Silně souvislé komponenty. Definice. Je dán orientovaný graf G. Množina vrcholů B se nazývá *silně souvislá komponenta*, též *komponenta silné souvislosti*, jestliže je maximální podmnožina vrcholů taková, podgraf indukovaný B je silně souvislý. □

Poznamenejme, že i zde termín „maximální“ znamená „nedá se přidat vrchol při zachování silné souvislosti“. Jinými slovy, pro každý vrchol $v \notin B$ graf indukovaný množinou $B \cup \{v\}$ není silně souvislý.

6.6.5 Poznámka. Každý vrchol orientovaného grafu leží přesně v jedné silně souvislé komponentě. O hranách už takové tvrzení neplatí – mohou existovat hrany, které vedou mezi silně souvislými komponentami.

6.6.6 Hledání silně souvislých komponent. Silně souvislé komponenty orientovaného grafu je možné hledat algoritmem, který je modifikací algoritmu prohledávání do hloubky. Modifikace spočívá v tom, že kromě pořadových čísel přiřazujeme vrcholům též číslo, které uvádí jak hluboko do prohledávacího zásobníku vede z daného vrcholu orientovaná cesta z dosud probraných hran.

6.6.7 Kondenzace grafu. Definice. Je dán orientovaný graf $G = (V, E)$. *Kondenzace grafu* G je graf $\hat{G} = (\hat{V}, \bar{E})$, kde $\bar{V}$ je množina všech silně souvislých komponent grafu G a hrana vede z komponenty K_1 do komponenty K_2 právě tehdy, když $K_1 \neq K_2$ a existují vrcholy $u \in K_1$, $v \in K_2$ takové, že (u, v) je hrana grafu G. □

6.6.8 Poznámka. Kondenzace grafu už je vždy acyklický graf. Kdyby totiž nebyl, pak byly špatně spočítané silně souvislé komponenty.

6.7 Eulerovy grafy

6.7.1 Eulerovské tahy. Připomeňme, že tah je sled, ve kterém se neopakují hrany. Jinými slovy, tah obsahuje hrany grafu vždy nejvýše jedenkrát.

Definice. Tah v grafu se nazývá *eulerovský*, jestliže prochází každou hranou. □

Jinými slovy, tah se nazývá eulerovský, jestliže obsahuje každou hranu přesně jedenkrát. Eulerovské tahy se dělí na uzavřené a otevřené, orientované a neorientované.

6.7.2 Eulerův graf.

Definice. Orientovaný (neorientovaný) graf G se nazývá *eulerovský graf*, jestliže v něm existuje uzavřený orientovaný (neorientovaný) eulerovský tah. □

6.7.3 Aplikace. Eulerovské tahy mají řadu aplikací.

- **Kreslení s co nejmenším počtem tahů.** Je dán neorientovaný souvislý graf. Úkolem je najít co nejmenší počet hranově disjunktních tahů tak, aby všechny hrany grafu byly obsaženy v některém z nich (to znamená, že každá hrana bude obsažena v přesně jednom tahu).

  Je zřejmé, že existuje-li v grafu eulerovský tah, pak je tento tah hledaným řešením. Řešení tohoto problému se dá využít např. při kreslení pomocí počítače (chceme co nejméně „přejezdů“).

- **Úloha čínského pošťáka.** Pošťák musí při své obchůzce projít všechny ulice. Jak to má udělat, aby ušel co nejméně kilometrů?

  Vytvoříme neorientovaný graf takto: vrcholy jsou křižovatky a hrany ulice mezi nimi, kterými musí pošťák projít. Hrany jsou ohodnoceny počtem kilometrů, které daná ulice má, nebo časem, který je potřeba na projití ulice.

  Jestliže v grafu existuje eulerovský tah, pak je to nejkratší možné řešení — pošťák projde každou ulicí přesně jednou. Jestliže eulerovský tah neexistuje, musí pošťák projít některou ulicí dvakrát. Tedy hledáme „zdvojení“ některých hran tak, aby vzniklý graf byl eulerovský a aby součet přidaných hran byl nejmenší možný.

- **De Bruijnova posloupnost**. Je dáno přirozené číslo $k > 1$. Úkolem je najít co nejdelší cyklickou posloupnost 0 a 1 tak, aby žádné dvě po sobě následující k-tice této posloupnosti nebyly stejné. Úloha se dá řešit nalezením uzavřeného orientovaného eulerovského tahu ve speciálním orientovaném grafu.

6.7.4 Tvrzení. V souvislém orientovaném grafu existuje uzavřený orientovaný eulerovský tah právě tehdy, když pro každý vrchol v grafu platí

$$d^-(v) = d^+(v).$$

(Tj. v každém vrcholu končí stejný počet hran jako v něm začíná.)

V souvislém grafu existuje uzavřený neorientovaný eulerovský tah právě tehdy, když každý vrchol má sudý stupeň. □

6.7.5 Postup hledání uzavřeného orientovaného eulerovského tahu. Vybereme libovolný vrchol v grafu. Protože graf je souvislý, v každém vrcholu začíná i končí alespoň jedna hrana.

Z vrcholu v vytváříme náhodně orientovaný tah; tj. procházíme hrany tak, abychom žádnou hranou neprošli dvakrát. Takto pokračujeme, dokud je to možné, tj. dokud se nevrátíme do výchozího vrcholu v a ve vrcholu v již nezačíná žádná dosud nepoužitá hrana.

Tím jsme dostali uzavřený tah C. Jestliže C obsahuje všechny hrany, je to hledaný uzavřený eulerovský tah.

Neobsahuje-li C všechny hrany, pak na C existuje vrchol w takový, že v něm začíná ještě nepoužitá hrana. (To vyplývá ze souvislosti grafu.) Tah C ve vrcholu w rozpojíme a vložíme do něj náhodně zkonstruovaný uzavřený tah z dosud nepoužitých hran, který začíná (a končí) ve vrcholu w. Dostaneme nový tah C'.

Jestliže C' obsahuje všechny hrany, je to hledaný eulerovský tah. V opačném případě opět najdeme na C' vrchol, ve kterém začíná dosud nepoužitá hrana a postup opakujeme dokud nedostaneme tah obsahující všechny hrany.

6.7.6 Tvrzení. V souvislém orientovaném grafu existuje otevřený orientovaný eulerovský tah právě tehdy, když existují vrcholy u_1, u_2 takové, že

$$d^-(u_1) = d^+(u_1) + 1, \;\; d^-(u_2) = d^+(u_2) - 1,$$

a pro každý jiný vrchol v grafu platí $d^-(v) = d^+(v)$.

V souvislém grafu existuje otevřený neorientovaný eulerovský tah právě tehdy, když v grafu existují přesně dva vrcholy lichého stupně.

6.7.7 Tvrzení. Je dán souvislý neorientovaný graf G s $2k$ vrcholy lichého stupně. Pak existuje k hranově disjunktních otevřených tahů takových, že každá hrana grafu G leží v právě jednom z těchto tahů. □

Zdůvodnění. Ke grafu G přidáme k hran a to tak, že každá nově přidaná hrana spojuje vždy dva vrcholy lichého stupně. Tím dostaneme eulerovský graf G' (ano, každý vrchol má již sudý stupeň). V grafu G' najdeme eulerovský uzavřený tah. Jestliže z něj odstraníme všechny přidané vrcholy, rozpadne se na k hranově disjunktních tahů. Tyto tahy splňují podmínky tvrzení.

6.8 Hamiltonovské grafy

Připomeňme, že cesta je tah, ve kterém se neopakují vrcholy (s výjimkou uzavřené cesty, kdy se první vrchol rovná poslednímu).

6.8.1 Hamiltonovské cesty, kružnice, cykly.

Definice. Je dán graf G. Otevřená cesta se nazývá *hamiltonovská cesta*, obsahuje-li všechny vrcholy (a tudíž všechny vrcholy přesně jedenkrát). Obdobně *hamiltonovská kružnice* je kružnice, která obsahuje každý vrchol grafu; *hamiltonovský cyklus* je cyklus, který obsahuje každý vrchol grafu. □

6.8.2 Hamiltonovské grafy.

Definice. Orientovaný (neorientovaný) graf G se nazývá *hamiltonovský*, jestliže obsahuje hamiltonovský cyklus (hamiltonovskou kružnici). □

6.8.3 Úlohy spojené s hamiltonovskými cestami dělíme na existenční, vyhodnocovací a optimalizační. V existenční úloze jde o to zjistit, zda v daném grafu existuje hamiltonovská cesta, kružnice nebo cyklus. Ve vyhodnocovací úloze jde o to najít (aspoň) jednu hamiltonovskou cestu, kružnici nebo cyklus v případě, že existují. V optimalizačních úlohách máme hrany grafu navíc ohodnoceny délkami a požaduje se nalezení hamiltonovské cesty, kružnice nebo cyklu s co nejmenším součtem délek jednotlivých hran cesty, kružnice nebo cyklu.

Na rozdíl od hledání eulerovských tahů, je hledání hamiltonovských cest, hamiltonovských kružnic a hamiltonovských cyklů velmi obtížná úloha. Přesněji, není znám algoritmus, který by pro obecný graf zjistil, zda v daném grafu existuje hamiltonovská cesta, kružnice nebo cyklus, a přitom provedl počet kroků, který závisí na počtu vrcholů a hran daného grafu jen polynomiálně. Přesto, nebo právě proto, jsou úlohy tohoto typu v praxi rozšířené.

6.8.4 Aplikace. Uvedeme některé z aplikací hamiltonovských cest.

- **Problém obchodního cestujícího.** Jde o problém nalezení nejkratší hamiltonovské kružnice v úplném neorientovaném ohodnoceném grafu.
- **Dopravní úlohy.** Jedná se o optimalizaci pohybu nějakého dopravního prostředku; např. při rozvozu zboží, vybírání schránek apod. Často se jedná o nalezení otevřené hamiltonovské cesty. Např. při rozvozu pracovníků na roztroušená pracoviště můžeme požadovat, aby se po skončení směny dopravní prostředek nevracel prázdný, ale aby svezl pracovníky (v opačném pořadí). Hledáme proto nejkratší hamiltonovskou cestu z daného vrcholu, koncový vrchol cesty obvykle není určen.
- **Plánování procesů.** Máme nějaké výrobní zařízení na kterém se provádějí procesy $p_1, p_2, \ldots, p_n$. Přitom pro některé dvojice procesů p_i, p_j platí, že má-li po skončení procesu p_i následovat proces p_j je třeba zařízení vyčistit, přestavět atd., tedy musíme zaplatit jistou cenu, aby po skončení procesu p_i mohl následovat proces p_j. Úkolem je najít takové pořadí procesů $p_1, p_2, \ldots, p_n$ aby cena byla nulová. V případě, že takové pořadí neexistuje, můžeme žádat pořadí procesů tak, aby cena byla nejmenší. Tento druhý případ vede na problém obchodního cestujícího.

6.8.5 Existují jednoduché nutné podmínky proto, aby v daném grafu existovala hamiltonovská cesta, hamiltonovská kružnice či hamiltonovský cyklus. Uvedeme několik takových tvrzení.

- Existuje-li v grafu hamiltonovská cesta, musí být graf souvislý.
- Existuje-li v grafu hamiltonovská kružnice, musí mít každý vrchol stupeň alespoň 2.
- Existuje-li v grafu G hamiltonovský cyklus, musí být graf silně souvislý.

Netriviální nutná a postačující podmínka pro zjištění, zda daný graf obsahuje hamiltonovskou cestu, kružnici nebo cyklus, není známa.

Kapitola 7

Kombinatorika

V této přednášce se zaměříme na počítání různých objektů a věcí. Na první pohled by se mohlo zdát, že je to velmi jednoduché — prostě objekty „spočítáme“. Jednoduchost je ale jen zdání. Ukážeme si několik postupů, které nám při počítání mohou pomoci; ovšem při řešení podobných úloh je vždy důležité umět se v situaci orientovat a umět se na problém podívat nejlépe z několika různých úhlů.

V závěrečné části se zaměříme na porovnávání rychlosti asymptotického růstu jednotlivých funkcí.

7.1 Základní principy

7.1.1 Sčítací princip. Máme zjistit počet nějakých objektů tvořící množinu X. Jestliže se nám podaří objekty rozdělit do, řekněme k, disjunktních skupin X_1, X_2, …, X_k, pak počet prvků množiny X je roven součtu počtu prvků jednotlivých množin X_i. Přesněji

$$|X| = |X_1| + |X_2| + \ldots + |X_k| = \sum_{i=1}^{k} |X_i|.$$

□

7.1.2 Násobící princip. Jestliže množina X obsahuje k-tice prvků množiny A, kde jednotlivé složky jsou na sobě nezávislé, pak počet prvků množiny X je roven $|A|^k$.

Přesněji: Jestliže $X = A_1 \times A_2 \times \ldots \times A_k$, kde všechny množiny A_i jsou konečné, pak

$$|X| = |A|_1 \times |A|_2 \times \ldots \times |A|_k = \prod_{i=1}^{k} |A_i|.$$

7.2 Permutace, variace a kombinace

Připomeňme n-faktoriál, se kterým jste se setkali již na střední škole. Pro $n \geq 1$ je *n-faktoriál* číslo

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1.$$

Pro $n = 0$ je $n! = 1$.

7.2.1 Permutace. Je dána konečná množina $\{1, 2, \ldots, n\}$. *Permutace* je každé prosté zobrazení množiny $\{1, 2, \ldots, n\}$ na sebe (tj. každá bijekce). □

Existuje $n!$ různých permutací množiny $\{1, 2, \ldots, n\}$.

Příklad. V obchodě mají 6 různých druhů čokolád. Kolika různými způsoby je prodavač může vyložit na pult do řady? (Předpokládáme, že na pořadí jednotlivých druhů čokolád záleží.)

Řešení: Vezmeme 6 čokolád od každého druhu jednu. Pak každá permutace této množiny tvoří jednu možnost vystavení čokolád. Existuje tedy

$$6! = 6 \cdot 5 \cdot \ldots \cdot 2 \cdot 1 = 720$$

různých způsobů vystavení čokolád.

7.2.2 Tvrzení. Máme n různých prvků, z toho je n_1 prvků typu 1, n_2 typu 2, až n_k typu k. (Mezi prvky stejného typu nerozlišujeme.)

Pak existuje

$$\frac{n\,!}{n_1\,!n_2\,!\ldots n_k\,!}$$

různých permutací těchto prvků. □

Důkaz: Tvrzení dokážeme matematickou indukcí.

Základní krok. Pro $k = 1$ máme jen jednu permutaci — všechny prvky jsou typu 1. Platí $\frac{n\,!}{n\,!} = 1$.

Indukční krok. Předpokládejme, že vzorec platí pro k různých typů. Uvažujme situaci, kdy máme $k+1$ typů. Nejprve vybereme umístění všech prvků $k+1$-vního typu. To je možné $\binom{n}{n_{k+1}}$ způsoby. Nyní stačí uspořádat $n - n_{k+1}$ prvků mezí typu 1, 2, až k. Podle indukčního předpokladu je to možné

$$\frac{(n - n_{k+1})\,!}{n_1\,!n_2\,!\ldots n_k\,!}.$$

Proto je hledaný počet roven

$$\binom{n}{n_{k+1}} \cdot \frac{(n - n_{k+1})\,!}{n_1\,!n_2\,!\ldots n_k\,!} = \frac{n\,!}{n_1\,!n_2\,!\ldots n_k\,!n_{k+1}\,!}.$$

7.2.3 Variace. Je dána konečná množina $\{1, 2, \ldots, n\}$. Vybíráme $k < n$ různých čísel z množiny $\{1, 2, \ldots, n\}$, kde záleží na pořadí vybíraných čísel. Jeden takový výběr se nazývá *variace* z $\{1, 2, \ldots, n\}$. Existuje

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1) = \frac{n!}{(n-k)!}.$$

různých variací.

Jestliže se prvky ve výběru mohou opakovat (tj. vybíráme k prvků ne nutně různých), pak počet různých výběrů je roven

$$n^k.$$

□

Příklad. Tři kamarádi, Aleš, Pavel a Martin, přijdou do obchodu, kde se prodávají čokolády a kde mají 6 různých druhů čokolády. Každý z kamarádů si koupí jednu čokoládu. Kolika způsoby je to možné, jestliže (a) každý si koupí jiný druh, (b) mohou si koupit i stejné druhy čokolády?

Řešení: (a) počet různých výběrů je $\frac{6!}{3!} = 120$, protože vybíráme trojice z šesti čokolád a záleží nám na pořadí.

(b) Tady se jedná opět o výběr trojic z šesti prvků, ale prvky se mohou opakovat. Proto je počet roven $6 \cdot 6 \cdot 6 = 216$.

7.2.4 Kombinace. Je dána konečná množina $A = \{1, 2, \ldots, n\}$. Vybíráme $k < n$ prvků množiny A, kde nám nezáleží na pořadí jednotlivých prvků. Jestliže vybíráme různé prvky, vybíráme vlastně k prvkové podmnožiny množiny A. Jednotlivému výběru říkáme *kombinace* k prvků z n prvkové množiny. Počet různých kombinací v případě že se prvky nemohou opakovat je

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Jestliže vybíráme k prvků A bez ohledu na pořadí a prvky se mohou opakovat, pak jejich počet je

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}.$$

7.2.5 Kombinační čísla. Nechť $k \leq n$ jsou dvě přirozená čísla. Pak číslo

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

se nazývá *kombinační číslo*, též *binomický koeficient.* □

7.2.6 Tvrzení.

1) Pro všechna $n \in \mathbb{N}$ platí $\binom{n}{0} = 1$
2) Pro všechna $n \in \mathbb{N}$ platí $\binom{n}{1} = n$.
3) Pro všechna $k \leq n$, $k, n \in \mathbb{N}$, platí

$$\binom{n}{k} = \binom{n}{n-k}.$$

4) Pro všechna $k \leq n$, $k, n \in \mathbb{N}$, platí

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}.$$

□

7.2.7 Binomická věta. Připomeňme znění binomické věty.

Věta. Je dáno přirozené číslo n. Pak pro všechna reálná čísla x, y platí

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k}\, y^k.$$

□

Použití binomické věty. Spočítejme $\sum_{i=0}^{n} \binom{n}{i}$. Z binomické věty víme, že

$$2^n = (1+1)^n = \sum_{k=0}^{n} \binom{n}{k} 1^{n-k}\, 1^k = \sum_{k=0}^{n} \binom{n}{k}.$$

Uvědomte si, že předcházející úvaha je vlastně jiný důkaz faktu, že každá n prvková množina má 2^n různých podmnožin. Ano, podmnožiny jsme rozdělili do skupin podle počtu prvků. Tyto množiny jsou disjunktní a fakt vyplývá ze sčítacího principu.

7.2.8 Princip inkluze a exkluze. Pro každé tři množiny A, B, C platí

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

□

7.2.9 Počítání zobrazení

Tvrzení. Jsou dány dvě konečné množiny A a B, kde $|A| = n$, $|B| = k$.

1. Existuje k^n různých zobrazení A do B.
2. Je-li $k \geq n$, pak existuje $k(k-1)\dots(k-n+1)$ různých prostých zobrazení množina A do množiny B.

□

7.2.10 Dirichletův princip, též holubníkový princip. Nechť A a B jsou dvě konečné množiny, $|A| = n$ a $|B| = k$. Jestliže $n > k$, pak neexistuje prosté zobrazení množiny A do množiny B. □

7.3 Asymptotický růst funkcí

V této části přednášky se zaměříme na „měření“ růstu funkcí. Protože pojmy a fakta zde uvedená využijete nejvíce při porovnávání efektivity algoritmů, půjde nám o asymptotické porovnávání růstu. To znamená, že nás nebudou zajímat hodnoty pro „malé“ hodnoty. Navíc, stejný algoritmus může být na dvou různých počítačích různě, budeme „růst“ až na přenásobení konstantou.

K porovnání funkcí použijeme $\mathcal{O}$, Ω a Θ.

7.3.1 Symbol $\mathcal{O}$.

Definice. Je dána nezáporná funkce $g(n)$, $g(n)\colon \mathbb{N} \to \mathbb{R}$. Řekneme, že nezáporná funkce $f(n)$ *je* $\mathcal{O}(g(n))$, jestliže existuje kladná konstanta c a přirozené číslo n_0 tak, že

$$f(n) \leq c\, g(n) \quad \text{pro všechny } n \geq n_0.$$

□

(Slovní spojení „funkce $f(n)$ je $\mathcal{O}(g(n))$“ chápejte jako „vlastnost“ funkce f.)

$\mathcal{O}(g(n))$ si můžeme představit jako třídu všech nezáporných funkcí $f(n)$:

$$\mathcal{O}(g(n)) = \{f(n) \mid \exists c > 0, n_0 \text{ tak, že } f(n) \leq c\, g(n) \ \ \forall n \geq n_0\}.$$

7.3.2 Symbol Ω.

Definice. Je dána nezáporná funkce $g(n)$, $g(n)\colon \mathbb{N} \to \mathbb{R}$. Řekneme, že nezáporná funkce $f(n)$ *je* $\Omega(g(n))$, jestliže existuje kladná konstanta c a přirozené číslo n_0 tak, že

$$f(n) \geq c\, g(n) \quad \text{pro všechny } n \geq n_0.$$

□

$\Omega(g(n))$ můžeme též chápat jako třídu všech nezáporných funkcí $f(n)$:

$$\Omega(g(n)) = \{f(n) \mid \exists c > 0, n_0 \text{ tak, že } f(n) \geq c\, g(n) \ \ \forall\, n \geq n_0\}.$$

Poznámka. Fakt, že funkce $f(n)$ je $\Omega(g(n))$ je ekvivalentní faktu, že funkce $g(n)$ je $\mathcal{O}(f(n))$.

7.3.3 Symbol Θ.

Definice. Je dána nezáporná funkce $g(n)$, $g(n)\colon \mathbb{N} \to \mathbb{R}$. Řekneme, že nezáporná funkce $f(n)$ *je* $\Theta(g(n))$, jestliže existují kladné konstanty c_1, c_2 a přirozené číslo n_0 tak, že

$$c_1\, g(n) \leq f(n) \leq c_2\, g(n) \quad \text{pro všechny } n \geq n_0.$$

□

$\Theta(g(n))$ můžeme též chápat jako třídu všech nezáporných funkcí $f(n)$:

$$\Theta(g(n)) = \{f(n) \mid \exists c_1, c_2 > 0, n_0 \text{ tak, že } c_1\, g(n) \leq f(n) \leq c_2\, g(n) \ \ \forall\, n \geq n_0\}.$$

Poznámka. Platí $f(n)$ je $\Theta(g(n))$ právě tehdy, když $f(n)$ je zároveň $\mathcal{O}(g(n))$ a $\Omega(g(n))$.

7.3.4 Značení. Protože symboly $\mathcal{O}, \Omega, \Theta$ představují množiny funkcí, budeme v dalším textu psát např. $f(n) \in \mathcal{O}(g(n))$. Je ovšem pravda, že v literatuře najdete i zápis $f(n) = \mathcal{O}(g(n))$. Při tomto zápisu je třeba mít na paměti, že znak rovnosti v zápise $f(n) = \mathcal{O}(g(n))$ nemá stejné vlastnosti jako klasická rovnost. Obdobně pro ostatní symboly.

7.3.5 Tvrzení. $f(n) \in \Theta(g(n))$ právě tehdy, když $g(n) \in \Theta(f(n))$. □

Zdůvodnění. Víme, že $f(n) \in \Theta(g(n))$ znamená existenci konstant $c_1, c_2 > 0$ a $n_0 \in \mathbb{N}$ takové, že

$$c_1\, g(n) \leq f(n) \leq c_2\, g(n) \quad \text{pro všechny } n \geq n_0.$$

Protože c_1, c_2 jsou kladné konstanty, z nerovnosti $c_1\, g(n) \leq f(n)$ vyplývá $g(n) \leq \frac{1}{c_1}\, f(n)$, a z nerovnosti $f(n) \leq c_2\, g(n)$ vyplývá $\frac{1}{c_2}\, f(n) \leq g(n)$. Položíme $d_1 := \frac{1}{c_2}$ a $d_2 := \frac{1}{c_1}$ a platí

$$d_1\, f(n) \leq g(n) \leq d_2\, f(n) \quad \text{pro všechny } n \geq n_0.$$

Proto $f(n) \in \Theta(g(n))$ implikuje $g(n) \in \Theta(f(n))$.

Druhá implikace se ukáže stejným způsobem.

7.3.6 Příklady.

1. Pro každé $a > 1$ a $b > 1$ platí
$$\log_a(n) \in \Theta(\log_b(n)).$$
2. V celém textu značíme logaritmus o základu 2 symbolem lg, tj. $\lg(n) = \log_2(n)$. Platí
$$\lg n! \in \Theta(n \lg n).$$

Druhá část tvrzení vyplývá např. z následující věty.

7.3.7 Věta (Gauss). Pro každé $n \geq 1$ platí

$$n^{\frac{n}{2}} \leq n! \leq \left(\frac{n+1}{2}\right)^n.$$

□

Zdůvodněni: Využijeme fakt, že pro každá dvě kladná čísla a, b platí $\frac{a+b}{2} \geq \sqrt{ab}$.

Přepíšeme $(n!)^2$ takto

$$(n!)^2 = n\,(n-1)\,\ldots 2\,1\,1\,2\,\ldots\,(n-1)\,n = \prod_{i=1}^{n}(n-i+1)i.$$

Odtud

$$n! = \prod_{i=1}^{n}\sqrt{(n-i+1)i} \leq \prod_{i=1}^{n}\frac{n+1}{2} = \left(\frac{n+1}{2}\right)^n,$$

protože pro každé i platí $\sqrt{(n-i+1)i} \leq \frac{n-i+1+i}{2}$. Tím jsme dostali horní odhad.

Na druhé straně pro každé i platí $n \leq (n-i+1)i$ a proto je $n^n \leq (n!)^2$. Odmocněním dostaneme dolní odhad, totiž $n^{\frac{n}{2}} \leq n!$.