

Úvod, návod ke knize

Co je matematika? Rozhodně ne vzorečky k učení nazepamět—to jsou počty. Matematika je jazyk, který nám dovoluje vyjádřit, že mezi určitými objekty existují rozličné vztahy, popřípadě že tyto objekty spolu různě interagují a tyto interakce splňují rozličná pravidla. Výhodou tohoto jazyka je jeho přesnost, protože v matematice je přesně definováno (určeno), co jednotlivé pojmy a značky znamenají. Další výhodou je jeho stručnost.

Je například možné říci: „Uvažujme kvantitu takovou, že když udělám čtvercové pole o straně rovné této kvantitě, pak bude mít toto pole výměru 4 jednotky“. Matematika nám umožňuje namísto toho říct „uvažujme číslo $\sqrt{4}$.“ Mimochodem, není to umělý příklad. Ta dlouhá věta ukazuje, jak lidé s matematikou začínali, první spisky o výpočtech a algoritmech (starověký Egypt, Čína, Indie,...) takto fungovaly.

Kromě stručnosti a přesnosti má matematický jazyk ještě další podstatnou výhodu: Umožňuje relativně efektivně hledat způsoby, jak ze zachycených vztahů, interakcí a pravidel získat co nejvíce informací. Například víme, že ona kvantita musí být 2. Matematika je tedy také umění řešit problémy nastolené matematickým jazykem.

Umět matematiku znamená umět kolem sebe nacházet vztahy a pravidelnosti a vyjadřovat je matematickým jazykem, struktury vzniklé matematickým popisem pak prozkoumat a získat nové informace. A naopak, je to i schopnost číst matematický jazyk, porozumět mu a překládat z něj do lidských, intuitivních pojmu, které je pak možno vztáhnout na svět kolem nás. Je to také schopnost umět odůvodnit, proč jsou metody použité k řešení problémů správné, protože základním prvkem matematiky je její spolehlivost. Ta je dána už tím, že je přesně znám význam pojmu, a hlavně tím, že matematici své závěry dokazují. To někdy vede k dojmu, že matematika je suchý kolotoč definic, vět a důkazů. Asi je trochu i oprávněný, ale je to cena, kterou platíme za spolehlivost výsledků.

Nevýhodou matematického jazyka je, že mu normální člověk nerozumí. Pokud jej někdo potřebuje (vědci, inženýři), tak se jej musí (jako každý jiný jazyk) naučit, a to v zásadě klasickým způsobem. I matematický jazyk má svá slovíčka (pojmy, značení) a gramatiku (například pravidla logiky), jeho obtížnost je ale v tom, že se podstatně liší od ostatních lidských jazyků a navíc je při jeho používání nutno silně používat mozek, a to ještě jedním specifickým způsobem (v zásadě tím, co měří IQ testy), což pro hodně lidí představuje problém a pro některé problém nepřekonatelný. Jedním z cílů tohoto textu je tento jazyk čtenáře alespoň trochu naučit. Bude to od něj samozřejmě vyžadovat práci, protože nejlépe se člověk jazyk naučí tak, že jakmile jej trochu pochytí, začne jím mluvit.

Kniha se tedy snaží o dvojí, na jedné straně má být relativně přístupným úvodem do oblasti zvané diskrétní matematika, zároveň na těchto tématech zkouší čtenáře uvést do světa logiky, matematického uvažování a kritického přístupu k vlastnímu myšlení. Pokud uspěje a čtenář si v tomto směru něco odnese, tak mu to pomůže asi i víc než konkrétní materiál, který se zde naučí. Umění uspořádat myšlenky, najít v nich řád, hledat slabá místa v argumentech, posoudit vzájemnou provázanost různých faktů a také umět vyjádřit své nápady strukturovanou formou jsou schopnosti, které se jistě budou hodit nejen při dělání matematiky.

Co je to za knihu

Tato kniha vznikala jako malé skriptum pro kurs diskrétní matematiky, ale trochu se vymkla z ruky. Její původní účel ji nicméně ovlivnil v několika zásadních věcech.

Jedna je obsahová. Diskrétní matematika je rozsáhlá oblast a většina autorů základních učebnic nestíhá probrat pořádně všechno, někde zajdou hlouběji, jinde jen nastíní základy, u této knihy je volba látky ovlivněna obsahem kursu diskrétní matematiky pro studenty computer science. Je to kurs úvodní a částečně přehledový, kniha tedy pokrývá hodně témat, ale málokdy jde hlouběji.

Významným faktorem je, že tento kurs byl prvním matematickým předmětem, který studenti ve svém programu mají, a měl je seznámit s matematickým způsobem myšlení, práce a vyjadřování. Autor chtěl studentům svými přednáškami i knihou představit matematiku jako vědu, ukázat jim, jak funguje, a naučit je rozumět matematickému jazyku tak, aby byli schopni matematiku číst a také do jisté míry psát (důkazy).

Dnes je k dispozici (zejména v angličtině) široké spektrum učebnic diskrétní matematiky. Mnohé se soustředí spíše na praktický pohled a neobtěžují příliš studenty abstraktními matematickými úvahami. Na opačné straně jsou knihy, pro které je diskrétní matematika jen pozadím, na kterém se snaží učit základy matematiky a logiky a zejména umění důkazů, což samozřejmě výrazně ovlivní řazení materiálu. Tato kniha si hledá místo někde uprostřed, je to pokus o kompromis. Pokud čtenáře zajímá spíš jen diskrétní matematika, může ignorovat teoretičtější pasáže (uspořádání knihy mu v tom pomůže) a dostane se mu v zásadě klasického výkladu diskrétní matematiky (autor by si rád myslel, že dost dobré provedený).

Na druhou stranu student, který se hodlá matematice věnovat hlouběji a zatím se s ní moc nepoznal, zde najde dostatek podnětů k přemýšlení. Aby kniha studentovi co nejvíce pomohla, je hlavně ze začátku výrazně rozvláčnější a detailnější, než bývá zvykem. „Výchovnému“ charakteru kursu odpovídá i to, že se zde najde spousta důkazů,

a to i u věcí, které se typicky nechávají na čtenářích nebo ignorují, a kniha se snaží postupně ukazovat, jak z probrané látky vznikají matematické struktury. Text je tedy hlubší a úplnější (z matematického pohledu) než standardní knihy. Je tu také spousta odboček směřujících čtenáře dál, dobrému studentovi by měly pomoci získat lepší představu o tom, co je vlastně matematika. Pokud autor ve svém úsilí o student-friendly knihu uspěl, pak by kniha mohla sloužit jako vstupní brána k matematice nejen studentům computer science, ale i studentům jiných oborů, kteří se budou muset s matematikou skamarádit a začínají v bodě nula.

Zvýšením prostoru pro důkazy a odbočky se už podle zákona zachování čehokoliv nedostalo na životopisné žblebty o slavných matematicích, historické anekdoty a podobně, ale to student najde v bohatém množství v libovolné novější matematické knize.

Návod ke čtení

Asi není nejlepší nápad skriptum otevřít, začít stranou 1, přejít na stranu 2 atd až po poslední. Hlavně student, který se s matematikou zatím moc nepotkal, udělá lépe, pokud ze začátku hutnější pasáže (zejména týkající se důkazů) přeskočí, jen občas nakoukne do těch snažších. Teprve když začne mít pocit, že už do toho trochu vidí, je vhodné se vrátit zpět na začátek a znova si přečíst to, co třeba na první čtení vypadalo nesrozumitelně. To se týká i rad, zejména těch obecnějších. Člověk obecné povídání „jak na to“ ocení mnohem lépe poté, co si to nejprve sám několikrát vyzkouší. Speciální rada ohledně první kapitoly: Rozhodně nedoporučuji přeskakovat její první část o logice, ale k části o důkazech by se začátečníci měli dostat raději až později.

Čtenář si dokonce v zásadě bez větších problémů může jen vyzobat některé kapitoly, v tom mu pomůže sekce Struktura kapitol níže.

Kniha se snaží vycházet vstříc čtenářům začínajícím, středním i pokročilým. Ti prvně jmenovaní budou asi při prvním čtení chtít některé pokročilejší pasáže přeskočit a soustředit se na pochopení základů. Proto jsou věci, které nepovažuji za nějak důležité, odsazeny zleva jako tato část. Odsazeny jsou i důkazy, ty ale pro některé čtenáře důležité budou. Některé pasáže jsou rovnou označeny jako bonusové. Naopak čtenáře pokročilého, kterého základy nudí, mohou právě ty odsazené/bonusové části probudit a nasměrovat dál.

! Je možné, že čtenář opravdu začínající toho bude chtít přeskočit víc. Pak mu může pomoci vykříčník nalevo, kterým značím ty pasáže, které mi přijdou důležité a raději by se přeskakovat neměly. Tvoří jakési jádro látky, na které je možné později nabalit zbytek.

S Touto značkou jsou pro změnu označeny části, které lze brát jako rady do života pro studenty. Jedná se zejména o algoritmy či doporučené postupy pro řešení hlavních typů problémů. Pro lepší orientaci jsem do obsahu přidal seznam těchto poradních koutků.

Text doprovází cvičení, která jsem se pokusil klasifikovat. „Rutinní“ přímo procvičují probraný materiál. Naučíte se řešit rovnice, rutinní cvičení chce řešit rovnici. Náročnější cvičení vyžadující trochu tvůrčí přístup jsem značil „dobrá“. U některých cvičení odhaduji, že jsou „zkoušková“, to jest ani příliš lehká, ani příliš těžká, ani příliš triková. Z hlediska obsahového pak jako „poučná“ značím cvičení, která nějakým způsobem doplňují vyloženou látku. Tyto vlastnosti je samozřejmě možné kombinovat.

Podobně klasifikuji důkazy. „Rutinní“ jsou většinou velice lehké a v typické knize by byly vynechány, protože pokročilejšího čtenáře neskonale nudí, ale já jsem je tu dal jednak proto, abych sám sebe přesvědčil, že opravdu rutinní jsou, a druhak proto, aby začínající student viděl, co všechno takový důkaz obnáší. Doporučuji je brát jako další cvičení. Pokud student zkusí „rutinní“ důkaz nejprve vytvořit sám a pak to porovná s tím, co jsem napsal já, mnoho se naučí. Očekává se, že lepší student bude na konci kursu umět takovéto rutinní důkazy sám tvorit levou zadní (leváci pravou zadní). Mimochodem, spousta důkazů se dá vést více směry, takže pokud váš důkaz nebude úplně stejný jako můj, tak to ještě nutně neznamená, že je špatně (ale u začátečníka většinou ano).

Pak jsou důkazy „poučné“, které podle mého soudu studentovi dobře ukážou, jak matematika funguje. Některé důkazy jsou myslím „drsné“, u těch rozhodně neočekávám, že by je student tvořil, a možná bude mít i problém je pochopit, ale výborný student by měl IMHO přinejmenším vnímat, co se tam děje. U některých důkazů jsem si nemohl pomoci a klasifikoval jsem je „z povinnosti“. Většinou jde o důkazy důležitých věcí, které by tedy měly být uvedeny, ale obvykle jsou dlouhé a přitom nečekám, že by studentovi něco daly, ani mě je nebaivilo psát. Ale přemohl jsem se.

U klasifikací cvičení i důkazů jde samozřejmě o můj subjektivní názor, pro studenta bude zásadní třeba i to, z jakého důvodu knihu čte. Pokud ani není jeho cílem se naučit matematiku psát, pak může všechny důkazy (možná kromě pár poučných) i mnohá cvičení vynechat. Rovněž důležitost jednotlivých pasáží a to, zda je cvičení zkouškové či ne, záleží také na úrovni kurzu.

Text je strukturován klasickým způsobem (definice, věty atd.), tvrzení jsou vymezena rámečkem. Pro přehlednost vyznačujeme i konce stylistických celků, konce důkazů jsou značeny tradičním čtverečkem vpravo, konce příkladů, poznámek a algoritmů trojúhelníčkem. Tvrzení jsou číslována arabskými číslicemi v každé kapitole zvlášť, značení ukazuje kapitolu a za tečkou číslo tvrzení (věta 6d.13). Podobně jsou číslovány příklady, ale mají své vlastní

číslování a používají na to písmena (příklad 6d.k). Abychom čtenáři ulehčili hledání konkrétního tvrzení/příkladu, značíme na spodním okraji poslední tvrzení a příklad, které se na té stránce objevily.

Abych čtenářům pomohl do dalších let, kdy budou nejspíše studovat z anglických zdrojů, uvádím v textu i anglické ekvivalenty termínů, občas celé věty. Poprvé řečeno mě lákalo to napsat celé englicky, protože každý, kdo se kolem computer science motá, musí tento jazyk umět. Nakonec jsem se ale rozhodl být hodný na začátečníky. Pokud půjdete v oboru dál, anglických knih si ještě užijete.

Na závěr pár rad. Lidé se liší v tom, jakou preferenci dávají paměti a jakou porozumění. Zejména u snažší látky býva častou volbou se jen nadrtit hromádku vzorových příkladů a doufat, že u zkoušky se natrefí na podobný. Tato strategie bývá ve škole bohužel docela účinná, ale v praxi často selhává, protože studenta nevyzbrojí na situace, které vybočují z naučených schémát.

Mnohem perspektivnější je látku pochopit. Student přemýšlivý nejprve stráví delší dobu rozjímaním nad podstatou věci a souvislostmi, načež si udělá pár příkladů, aby se ujistil, že to má opravdu v paži. Výhodou tohoto přístupu je, že už nevyžaduje tolik pamatování a navíc je získaná znalost vysoce flexibilní a neztrácí se v čase tak rychle, jako našprtané rutinní postupy. Nevýhoda je, že to vyžaduje přemýšlení a to bolí. Většina studentů oba přístupy kombinuje a poměr si nastavuje dle vlastního vkusu, nicméně pro studenta s ambicemi, který by rád v oboru působil tvůrčím způsobem, je jednoznačnou volbou cesta přes pochopení.

Jak se k takovému pochopení dospěje? Kromě rozjímaní nad textem je důležité správně pracovat s příklady a cvičeními. Klasická situace: Student se podívá na zadání, nevidí jak na to, koukne do řešení, prohlásí „A jo, tak takhle to je“ a jede dál. Výsledek: nenaučí se nic. Dvě věci jsou třeba dělat jinak.

1. Je třeba příklad opravdu zkousit vyřešit, věnovat mu čas, napnout mozek. Teprve když se to nezlomí ani po větším úsilí, je čas kouknout na řešení. Proč? Protože nejvíce si pamatujeme věci, které máme svázány s emocemi, například vztekem, že něco nejde. Pokud bez emocí konstatujeme „to nevidím“, tak máme malou šanci, že něco z této epizody uvízne v paměti.

2. Když se podíváme na řešení (ať už cvičení nebo ukázkového příkladu v textu), tak nestací jen konstatovat, že vidíme, co se tam děje. Klíčové je umět si zodpovědět na otázku, proč se to tam děje. Proč autor řešil tento příklad zrovna tímto způsobem? Jak poznal, že nemá zkousit něco jiného? A co by se stalo, kdyby něco jiného zkousil? Teprve až čtenář najde odpovědi na tyto otázky, tak má jistotu, že až tento (či podobný) příklad zase potká, tak bude vědět, co dělat. Pokud si na tyto otázky odpovědět neumí, tak by se měl vrátit k textu, protože ty odpovědi tam někde jsou.

Struktura kapitol

Z obsahové stránky je diskrétní matematika zajímavá tím, že mezi jednotlivými tématy je relativně slabá vazba (na rozdíl třeba od analýzy). Kapitoly je možné do značné míry studovat samostatně či permutovat, aniž by vznikly větší problémy. Konec konců, tato kniha prošla během přípravy třemi značně rozdílnými pořadími kapitol a ani jedno z nich by nebylo špatně, rozhodovaly maličkosti.

Čtenáři (či učiteli předmětu diskrétní matematika) se tak nabízí značná svoboda, co a v jakém pořadí číst. Omezení je velice málo:

- Silně doporučuji začít kapitolou 1a, dobré je pak alespoň proletět množiny a zobrazení.
- Kapitolou o kombinatorice se dá klidně i začít a může sloužit jako motivace pro později zařazenou kapitolu o rekurentních rovnicích, která také může být samostatně hned na začátku, používá sice indukci, ale jen zlehka, mnohý čtenář ji již zná v dostatečné míře.
- Je jasné, že obě kapitoly o relacích patří za sebe, ale neodvolávají se příliš na ostatní materiál a klidně mohou přijít později. Podobně kapitola o indukci může klidně přijít později nebo naopak dříve. Jediná silná vazba je k relacím (princip dobrého uspořádání a princip indukce). Je tedy třeba umístit povídání o axiomech do té kapitoly, která přijde dříve (indukce či relace).

Pokud se u studentů očekává již nějaká zkušenosť s matematikou, je dobrý nápad dát kapitolu o indukci na začátek, protože nic nevyžaduje a naopak prakticky všechny kapitoly indukci používají, je to tedy z matematického pohledu takto čistější (tak tomu bylo i v původní verzi tohoto textu). Pro začínající studenty ale bývá lepší je nejprve do matematiky uvést na něčem snažším, proto je v této knize nejprve dvojkapitola o relacích.

• Kapitola o dělitelnosti a počítání modulo se částečně odvolává na předchozí výsledky o relacích, ale je možné ji bez větší újmy studovat i samostatně.

• Kapitola o binárních operacích je v zásadě nezávislá na zbytku knihy (používá relace a prostory modulo jako příklady, ale to se dá zvládnout). Vzhledem k silné abstraktnosti materiálu je dobré ji řadit později, až si student trochu na matematiku zvykne. Je to zase otázka vyspělosti čtenáře, pokud ten již abstraktnější matematiku potkal, tak může být lepší zařadit kapitolu o počítání modulo až za binární operace.

- Grafy jsou jedna z věcí, na které v původním kursu nedošlo, zde byla zařazena čistě přehledová kapitola jen pro úplnost. Čtenář doporučujeme nějakou dobrou knihu.

Literatura

V knize chybí tradiční bohaté odkazy na literaturu. Většina látky je totiž klasická, nové je její podání. Při svém studiu i přípravě tohoto kurzu jsem samozřejmě čerpal z mnoha zdrojů, zmíním ty hlavní. Hodně pomohli mí učitelé matematiky, počínaje základní školou a konče MFF UK, své zápisky z tehdejších přednášek mám dodnes schovány a několikrát jsem do nich nakoukl i při přípravě této knihy. Inspirací mi bylo pojetí diskrétní matematiky mých kolegů, prof. Demlové a doc. Velebila, hodně se mi také líbila níže zmíněná kniha od Rosena.

Pokud by se student chtěl podívat i do jiné knihy, pak je jich k dispozici mnoho, stačí vyhledat „discrete mathematics“ na Webu. Většina má podobný obsah (ale prakticky žádná se nekryje s touto, něco chybí a něco je navíc), i zpracování bývá v modernějších knihách podobné a je otázkou osobního vkusu, která více vyhoví. Pokud by čtenáře zajímal můj názor, zde je pár doporučení:

- Rosen, K.H.: *Discrete Mathematics And Its Applications*, 6ed, McGraw-Hill (2007).

Tohle je první liga. Je to kniha „nového“ typu, se spoustou historických poznámek, životopisů matematiků, pěkných příkladů s aplikacemi, počítačovými cvičeními a podobně. Obsah: Logika, množiny a zobrazení, kombinatorika, relace, dělitelnost a počítání modulo, rekurentní rovnice, také teorie grafů a dokonce algoritmizace. Oproti tomuto textu chybí binární operace, kniha nejde tak hluboko a nesnaží se také organizovat materiál do matematických struktur (pro některé čtenáře to může být výhodou). Navíc probírá diskrétní matematiku a algoritmizaci paralelně a ony se pěkně doplňují, velice dobře se čte a má málo chyb, zato spoustu cvičení. Je to také pěkná bichle (1000 stran). Doporučil bych ji jako zajímavou knihu ke čtení zároveň s tímto textem, protože namísto pohledu teoretičtějšího nabízí spíš pohled praktičtější. Moc se mi líbila.

- Velebil, J.: *Diskrétní matematika a logika*, online (pdf soubor).

První polovina skripta nabízí zajímavý pohled na indukci, relace a počítání modulo. Odtud se pak odrazí k obecnějším algebraickým strukturám pro pokročilé. Skriptum je méně upovídané, nabízí výklad z pohledu matematiky a také dost zajímavých příkladů, náročnějšího čtenáře rozhodně nezklame.

- Velebil, J.: *Lecture notes for Mathematics 5(d)*, online (pdf soubor).

Rovněž začíná indukcí a počítáním modulo, i zde se pak rozjede do pokročilých luhů a hájů, v tomto případě ještě dále. Procvičí angličtinu.

- Matoušek, J. a Nešetřil, J.: *Kapitoly z diskrétní matematiky*, Karolinum Praha (2000).

V prvních kapitolách se podrobně proberou množiny a zobrazení, dále relace, dělají to matematicky, ale se spoustou příkladů a povídání. Dobře se to čte. V druhé polovině knihy autoři utečou hlavně ke grafům a dalším tématům mnohem pokročilejším než tento kurz. Jako dobré cvičení bych doporučil anglickou verzi *Invitation to Discrete Mathematics*, Oxford UP (2008).

- Demlová M., Pondělíček B.: *Matematická logika*, ČVUT Praha (1997).

Doplní studentovi logiku, velice pěkné skriptum.

- Demel J.: *Grafy a jejich aplikace*, Academia (2002).

Díky této knize zde stačil jen stručný úvod o grafech, zbytek je tam.

Upozornění: Snažil jsem se psát česky, pokud se tedy někdy od pravidel českého pravopisu odchyluji (a není to překlep), pak je to záměr, protože si naivně myslím, že to tak je hezčí.

Poděkování: Rád bych poděkoval všem, od kterých jsem se tuto látku naučil a kteří mě různým způsobem podporovali při psaní tohoto textu (třeba děti mě nechaly občas i vyspat). Special thanks jdou panu Knuthovi, jehož sázecí systém *TeX* byl naprostou revolucí v přípravě textů, používám jeho mutaci *AMS-TEX*. Obrázky jsem přímo ve zdrojovém kódu tvořil pomocí PGF/TikZ, který jsem se kvůli této knize naučil a bylo to šťastné setkání.

Děkuji také studentům za odchycení chyb a překlepů, zejména panu Michalu Souchovi, který byl obzvláště pečlivým čtenářem.

Příjemné čtení přeje autor.

Značení

Varování: U čísel používám zásadně desetinnou tečku, protože je tomu tak na kalkulačkách, počítačích i v anglických knihách, lidi od computer science jsou na to zvyklí a já taky. Jako milovníkovi literatury se mi neporušují pravidla snadno, ale jsem přesvědčen, že odborná literatura by měla čtenáři pochopení usnadňovat, ne komplikovat.

\mathbb{N}	přirozená čísla $1, 2, 3, 4, \dots$
\mathbb{Z}	celá čísla $0, 1, -1, 2, -2, \dots$
A, B, C, \dots	množiny
a, b, c, \dots, x, y, z	prvky množin, čísla
$A \cap B, A \cup B$	průnik, sjednocení množin
R, S, T, \dots	zobrazení, relace
$S \circ R$	skládání dvou relací/zobrazení v pořadí R , pak S
$T: A \mapsto B$	zobrazení z A do B
$ A = B $	množiny A, B mají stejnou mohutnost
aRb	dvojice (a, b) je v relaci R
$a \preceq b$	dvojice (a, b) je v relaci \preceq , což je částečné uspořádání
$a b$	a dělí b
$r = a \bmod d$	r je zbytek po dělení a číslem d
$a \equiv b \pmod{n}$	čísla a, b jsou kongruentní (mají stejný zbytek) modulo n ,
$a = b \pmod{n}$	viz $a \equiv b \pmod{n}$
$a \circ b$	binární operace \circ aplikovaná na prvky a, b
$\{a_k\}_{k=1}^n, \{a_k\}_{k=1}^\infty$	posloupnost čísel a_1, a_2, \dots konečná či nekonečná
$\{a_k\}$	posloupnost čísel a_k , začátek indexace irrelevantní

Obsah

^b: bonusový materiál.

0. Úvod, návody k použití, značení, obsah

1. Matematika, logika

 1a. Průlet logikou

 (výroky a spojky; pravidla; logika v aplikacích; kvantifikátory)

 1b. Logika a matematika

 (jazyk: definice a věty; důkazy: přímý, nepřímý, sporem; 1b.5 jak dokazovat)

2. Teorie množin

 2a. Množiny

 (množiny, operace, pravidla)

 2b. Zobrazení

 (skládání a inverze; prosté, na a bijekce; velikost množin; funkce $[x]$ a $[x]$)

 2c. Mohutnost množin

 (mohutnost; spočetné a nespočetné množiny; \mathbb{N} , \mathbb{Z} , \mathbb{Q} a \mathbb{R})

3. Binární relace

 3a. Binární relace a operace s nimi

 (reprezentace maticemi; množinové operace; inverze, skládání a mocnina; operace a reprezentace)

 3b. Základní vlastnosti binárních relací

 (čtyři základní vlastnosti; vyšetřování vlastností; 3b.9 Kartézský součin)

 3c. Další vlastnosti relací^b

 (vlastnosti a operace; 3c.6 uzávěr relace; 3c.8 další vlastnosti)

 3d. n -ární relace^b

4. Speciální relace

 4a. Ekvivalence

 (ekvivalence; třídy ekvivalence; rozklad množiny)

 4b. Částečná uspořádání

 (relace \preceq a \prec ; 4b.6 Hasseův diagram; 4b.17 uspořádání a kartézský součin)

 . Minima, nejmenší prvky a podobně, dobré uspořádání

 (max. a min., nej[men/vět]ší prvek; porovnatelnost a lineární uspořádání; 4c.14 princip dobrého uspořádání)

 4d. Bonus: Další pojmy okolo uspořádání

 (horní/dolní mez, sup a inf, svaz)

5. Indukce a rekurze

 5a. Matematická indukce

 (principy slabé indukce; 5a.7 indukce a algoritmy; principy silné indukce)

 5b. Rekurze a strukturální indukce

 (induktivní definice množin; princip strukturální indukce)

6. Dělitelnost a prvočísla

 6a. Dělitelnost

 (dělitelnost; dělení se zbytkem; gcd a lcm; 6a.19 Bezout; 6a.21 Euklidův algoritmus)

 6b. Prvočísla

 (prvočísla; 6b.4 Fundamentální věta aritmetiky)

 6c. Diofantické rovnice

 (struktura řešení; homogenní případ; 6c.6 algoritmus)

7. Počítání modulo

 7a. Kongruence, počítání modulo

 (kongruence a operace; počítání modulo, opačný a inverzní prvek; \mathbb{Z}_n ;

 7a.12 Malá Fermatova věta; \mathbb{Z}_n jako třídy ekvivalence; 7a.22 Eulerova funkce a věta)

 7b. Řešení rovnic modulo

 (7b.1 lineární kongruence; 7b.8 soustavy a Čínská věta o zbytcích)

 7c. Matice a polynomy modulo

 (7c.1 matice: problém výpočtu determinantu a inverzní matice;

 7c.5 polynomy: stupeň, rozklad, kořeny)

8. Binární operace

8a. Pologrupy a monoidy

(binární operace; asociativita; mocnina; jednotkový a inverzní prvek; podmonoid; řád prvku; monoid generovaný mocninami prvku)

8b. Grupy

(grupy; podgrupy; podgrupy generované prvkem, řád prvku, mohutnost podgrup)

8c. Další struktury^b

(okruhy; obory integrity; tělesa; 8c.4 okruhy polynomů)

8d. Bonus: Racionální čísla

9. Posloupnosti a součty, řady

9a. Posloupnosti

(aritmetická a geometrická posloupnost; monotonie; limita)

9b. Porovnávání rychlosti růstu

(pojmy \ll , o , ω , Θ atd.; škála mocnin)

9c. Sumy

(operace se sumami; součty mocnin; součet geometrické posloupnosti; součiny)

9d. Řady^b

(konvergence; operace s řadami; mocninné řady)

10. Rekurentní vztahy

10a. Lineární rekurentní rovnice

(obecné a partikulární řešení; počáteční podmínky; Věta o existenci a jednoznačnosti; Věty o struktuře řešení)

10b. Rovnice s konstantními koeficienty

(charakteristická čísla a báze prostoru řešení homogenní rovnice; metoda odhadu pro speciální pravou stranu; Věta o superpozici)

10c. Další rovnice (Master theorem)

(rovnice algoritmů divide-and-conquer; The Master Theorem)

10d. Bonus: Generující funkce

(transformace posloupností na funkce, řešení rekurentních rovnic)

11. Kombinatorika (počítání)

11a. Základní principy

(sčítací, násobící a doplnkový princip; klasické permutace kombinace variace; nestandardní situace; stromy jako nástroj)

11b. Pokročilejší principy

(princip inkluze a exkluze; Dirichletův šuplíkový princip; 11b.7 krabičky)

11c. Binomická věta, kombinaciční čísla

(kombinaciční čísla; Pascalova a jiné identity; 11c.3 binomická věta a její varianty)

11d. Bonus: Generování výběrů.

12. Grafy (přehled)

12a. Co jsou grafy (poprvé)

12b. Co jsou grafy (podruhé)

12c. Procházení grafem

12d. Kreslení grafů

12e. Barvení grafu

12f. Stromy, kostra grafu

12g. Bonus: Platónovská tělesa

13. Bonus: Isomorfismy a transformace

S Návody a algoritmy:

1b.5: Jak vytvářet důkazy

2a.10: Jak psát a čist důkazy

2b.9: Jak na vlastnosti funkcí

2c.17: Jak určovat mohutnost

3b.2: Jak vyšetřovat vlastnosti relací

4b.7: Jak vytvářet Hasseův diagram

5a.12: Jak dokazovat indukcí

6a.21: Rozšířený Euklidův algoritmus

6a.22: Ruční výpočet Euklidova algoritmu

6c.6: Jak řešit diofantické rovnice

7a.11: Jak najít inverzní prvek modulo

7b.7: Jak řešit rovnice $ax = b$ modulo n

7b.13: Jak řešit soustavy lineárních kongruencí

10b.8: Jak řešit lineární rekurentní rovnice

1. Matematika, logika

Co je matematika? Rozhodně to nejsou vzorečky k učení nazpaměť—to jsou počty. Matematika je jazyk, který nám dovoluje vyjádřit, že mezi určitými objekty existují rozličné vztahy, popřípadě že tyto objekty spolu různě interagují a tyto interakce splňují rozličná pravidla. Výhodou tohoto jazyka je jeho přesnost, protože v matematice je přesně definováno (určeno), co jednotlivé pojmy a značky znamenají. Další výhodou je jeho stručnost.

Je například možné říci: „Uvažujme kvantitu takovou, že když udělám čtvercové pole o straně rovné této kvantitě, pak bude mít toto pole výměru 4 jednotky“. Matematika nám umožňuje namísto toho říct „uvažujme číslo $\sqrt{4}$.“ Mimochodem, není to umělý příklad. Ta dlouhá věta ukazuje, jak lidé s matematikou začínali, první spisky o výpočtech a algoritmech (starověký Egypt, Čína, Indie,...) takto fungovaly.

Kromě stručnosti a přesnosti má matematický jazyk ještě další podstatnou výhodu: Umožňuje relativně efektivně hledat způsoby, jak ze zachycených vztahů, interakcí a pravidel získat co nejvíce informací. Například víme, že ona kvantita musí být 2.

Nevýhodou matematického jazyka je, že mu normální člověk nerozumí. Pokud jej někdo potřebuje (vědci, inženýři), tak se jej musí (jako každý jiný jazyk) naučit, a to v zásadě klasickým způsobem. I matematický jazyk má svá slovíčka (pojmy, značení) a gramatiku (například pravidla logiky), jeho obtížnost je ale v tom, že se podstatně liší od ostatních lidských jazyků a navíc je při jeho používání nutno silně používat mozek, a to ještě jedním specifickým způsobem (v zásadě tím, co měří IQ testy), což pro hodně lidí představuje problém a pro některé problém nepřekonatelný. Jedním z cílů tohoto textu je tento jazyk čtenáře alespoň trochu naučit. Bude to od něj samozřejmě vyžadovat práci, protože nejlépe se člověk jazyk naučí tak, že jakmile jej trochu pochytí, začne jím mluvit.

Zkusme ukázat ještě jeden pohled na věc: Umět matematiku znamená především umět kolem sebe nacházet vztahy a pravidelnosti a vyjadřovat je matematickým jazykem, struktury vzniklé matematickým popisem pak prozkoumat a problém vyřešit. A naopak, je to i schopnost čist matematický jazyk, porozumět mu a překládat z něj do lidských, intuitivních pojmu, které je pak možno vztáhnout na svět kolem nás. Je to také schopnost umět odůvodnit, proč jsou metody použité k řešení problémů správné, protože základním prvkem matematiky je její spolehlivost. V matematice je přesně známo, která tvrzení platí, protože matematici vše dokazují.

Aby byly závěry matematiky spolehlivé, je potřeba přesně definovat pojmy a o nich pak tvrdit rozličné věci. Tímto se zabývá kapitola 1b. Spolehlivost matematických tvrzení se pak odvíjí od toho, že je jejich platnost nezvratně dokázána (v rámci přijatého výkladu světa, viz poznámky o axiomech v dalších kapitolách). O metodách důkazu se rozgovídáme v kapitole 1b, ale potřebujeme na to jako základní nástroj logiku. S ní tedy začneme.

1a. Průlet logikou

Znalost logiky je v matematice naprostou nutností. Zde je třeba říct, že logika jako taková je samostatný a obsáhlý obor, kterému je možné se plně věnovat několik semestrů, ale pro práci v matematice většinou stačí základní znalosti. V této kapitolce je přiblížíme pro ty, kdo se s formální logikou ještě nesetkali, ale hlouběji nepřejdeme. Jednak to není třeba a druhak ji studenti computer science často mají jako samostatný kurs, který se na ni podívá mnohem blíže, než tady vůbec potřebujeme, navíc jsou na to pěkné knihy a nemá tedy smysl to zde dělat znovu.

Co je tedy logika? Z praktického pohledu je to obor zabývající se zkoumáním situací, ve kterých dokážeme ze znalosti, zda jsou pravdivá určitá tvrzení, odvodit spolehlivě pravdivost či nepravdivost tvrzení jiných. Základem jsou výroky (značíme je malými písmeny), což je v zásadě nějaké vyjádření, o kterém lze rozhodnout, zda je pravdivé či ne. Příklady výroků: „ $13 > 23$ “, „Země je blíž ke Slunci než Jupiter“, „Právě čtu tato skripta“, „Žižka jedl 4. října 1424 jablka“. Všimněte si, že u posledního tvrzení nevíme, zda je pravdivé či ne (možné to je, umřel až o týden později), ale nějakou pravdivostní hodnotu to má, to je podstatné.

Toto pro změnu výroky nejsou: „Ahoj“, „31“, „Pavlač“, „Beze mne“, „Modrá je dobrá“, „Jsem normální“, „Prší“. Kupodivu se zrovna tvrzení „prší“ často v populárnějších rozpravách o logice používá jako příklad výroku, ale výrokem se to stane teprve ve chvíli, kdy řekneme, kde a kdy má pršet, jinak totiž nelze určit, zda je pravdivý či ne. „Tady a teď prší“ je výrok. Jenže lidé jsou líní to psát celé.

Rozmyslete si, že všechny ty výroky byly tak jednoduché, že už nešly dál zmenšit, aby ještě zůstaly výroky. Naopak „Právě ťukám do klávesnice a hraje mi Weird Al Yankovic“ se dá rozlousknout na dva výroky jednodušší. Přesně takové situace nás budou zajímat. Máme jednoduché výroky (atomární, ale nechceme zde začínat s odbornou terminologií) a zajímá nás, co se s nimi dá dělat a jak to pak dopadne. Jinými slovy, výroky různě modifikujeme a spojujeme a pak se ptáme, co se děje s pravdivostí. Přitom nás vůbec nebude zajímat, co vlastně jednotlivé výroky říkají, jen jestli jsou pravdivé či ne. Pravdivost výsledných tvrzení bude odvozována čistě z toho,

jakým způsobem jsou prvotní výroky poskládány, a z informace o jejich pravdivosti, dominantní je forma, nikoliv obsah. Proto se tomu říká formální logika.

Začneme tím nejjednodušším.

- Nechť je p výrok. Jeho **negace** se značí $\neg p$ a je to výrok, jehož pravdivostní hodnota je přesně opačná než pravdivostní hodnota p .

Takže $\neg p$ je takový výrok, který je pravdivý, když p pravdivý není, a naopak. Například negace výroku „Tady a teď prší“ je „Tady a teď neprší“. Rozhodně negací nebude „Nejsou tu teď mraky“, protože může nastat situace, kdy jsou tvrzení „Nejsou tu teď mraky“ a „Tady teď prší“ obě nepravdivá. Nás to samozřejmě nejvíce zajímá v matematice, takže například negace k „ $13 > 23$ “ není „ $13 < 23$ “, ale „ $13 \leq 23$ “.

Fungování negace se dá elegantně vyjádřit takzvanou pravdivostní tabulkou.

p	$\neg p$
0	1
1	0

V prvním sloupci si najdeme, co víme o p , a v druhém se ve stejném řádku dozvídme, jak se pak zachová $\neg p$. Jako obvykle používáme 0 pro nepravdu a 1 pro pravdu.

Výroky spojujeme logickými spojkami. Nejpoužívanější jsou tyto čtyři.

Nechť p a q jsou výroky.

- **konjunkce** značená „ $p \wedge q$ “ popř. „ $p \& q$ “ či „ p a q “ a čtená „ p a q “ je výrok, který je pravdivý právě tehdy, když jsou pravdivé oba výroky p i q .
- **disjunkce** značená „ $p \vee q$ “ popř. „ p nebo q “ a čtená „ p nebo q “ je výrok, který je pravdivý v situaci, když je pravdivý alespoň jeden z výroků p či q .
- **implikace** značená „ $p \Rightarrow q$ “ popř. „ $p \rightarrow q$ “ a čtená „jestliže p , pak q “ je výrok, který je pravdivý, když jsou pravdivé oba p i q nebo když je p nepravdivý.
- **ekvivalence** značená „ $p \Leftrightarrow q$ “ popř. „ $p \leftrightarrow q$ “ a čtená „ p právě tehdy, když q “ je výrok, který je pravdivý, když mají výroky p a q stejnou pravdivost, tedy jsou oba pravdivé či oba nepravdivé.

Fungování těchto operací se zase standardně vyjadřuje pomocí pravdivostních tabulek, které ukazují, jakou má ten který složený výrok pravdivost v závislosti na tom, co je zrovna známo o pravdivosti p a q .

p	q	$p \wedge q$	p	q	$p \vee q$	p	q	$p \Rightarrow q$	p	q	$p \Leftrightarrow q$
1	1	1	1	1	1	1	1	1	1	1	1
1	0	0	1	0	1	1	0	0	1	0	0
0	1	0	0	1	1	0	1	1	0	1	0
0	0	0	0	0	0	0	0	1	0	0	1

Zkusme si to, uvažujme například výroky p : „Dnes je pátek“ a q : „Dnes je 13. den v měsíci“. Pak výrok $p \wedge q$ říká „Dnes je pátek třináctého“ a selský rozum ve shodě s tabulkou říká, že bude pravdivý jen tehdy, když jej řekneme v pátek, který je také třináctý den v měsíci. Naopak výrok $p \vee q$ bude pravdivý každý pátek a také každého třináctého, což zahrnuje i pátky třináctého. Na konjunkci a disjunkci asi není moc co řešit, jsou v zásadě jasné. Podívejme se blíže na ostatní dvě operace.

Implikace má jednu zajímavou vlastnost, v našem příkladě $p \Rightarrow q$ znamená „Jestliže je pátek, tak je třináctého“. Podle tabulky je pravdivá v pátky třináctého, ale také od pondělí do čtvrtka a o víkendu, bez ohledu na to kolikátého je. Dostáváme se tím ke klíčové vlastnosti implikace, v případě neplatného předpokladu je implikace jako celek pravdivá. To může začátečníka zarazit, snad mu pomůže následující představa. Implikace se dá brát jako jakási forma slibu. Slibujeme, že jestliže se splní p , tak my uděláme věc q . Někdo se pak dívá, jak to proběhlo, a hodnotí, zda jsme svůj slib splnili (tedy implikace je pravdivá). Tabulka pak dává smysl: Pokud p nenastalo, tak nás slib k ničemu nezavazoval, tudíž ať už jsme q udělali či ne, tak ten slib nebyl porušen a dostává jedničku. Porušili jsme jej jedině v případě, kdyby p bylo splněno, ale my jsme neudělali q .

Toto chování je možná na první pohled divné, ale brzy uvidíme, že přesně vystihuje, co při logických úvahách často potřebujeme.

Při běžném hovoru si lidé často pletou implikaci s **ekvivalencí**. Například řeknou: „Když budeš hodný, dostaneš bonbón“, ale zároveň tím myslí, že když hodný nebude, tak bonbón nedostane, což ovšem ze zvolené formy implikace nevyplývá. Jak už jsme viděli, zlobivé dítě klidně bonbón dostat může a slib–implikace tím porušen nebude. Správné logické vyjádření tedy je „bonbón dostaneš právě tehdy, když budeš hodný“. Tím jsou obě základní fakta („hodný“ a „bonbón“) vzájemně zcela spojeny, pravdivostí si musí odpovídat, aby tento slib zůstal splněn.

Výroky sestavené pomocí základních čtyř spojek a negace je ovšem možné znovu spojovat a negovat, takže můžeme (podobně jako s čísly a operacemi v algebře) sestavovat komplikované konstrukce, i zde pořadí vyhodnocování vyznačujeme závorkami. Abychom jich trochu ušetřili, dává se negaci absolutní prioritu nad ostatními operacemi. Přestavme si tedy takový obecný výrok vzniklý z určitých atomárních výroků p, q, \dots , pak nás zajímá, jak jeho pravdivost závisí na vstupních datech neboli na pravdivostních hodnotách těch p, q, \dots . To se zase nejlépe vyjádří tabulkou. Ukážeme si to pro vcelku jednoduchý výrok $\neg p \vee q$, nejprve si uděláme pomocný sloupec pro tu negaci a pak jej „zdisjunktníme“ s q . Mimořádem, díky prioritě negace jsme nemuseli psát $(\neg p) \vee q$.

p	q	$\neg p$	$\neg p \vee q$
1	1	0	1
1	0	0	0
0	1	1	1
0	0	1	1

Tento příklad je typický, zkoumaný výrok je někdy pravdivý a někdy ne. Jsou ale výjimky. Dobrým příkladem je výrok „Teď tady prší nebo teď tady neprší“. Vzhledem k tomu, že třetí možnost není, tak jedna z těch dvou variant musí vždycky nastat a podle pravdivostní tabulky vidíme, že pak už je pravdivá celá disjunkce. Je to tedy výrok, který je za všech okolností pravdivý.

Je asi zjevné, že toto bude fungovat s libovolným výrokem p , výraz $p \vee \neg p$ bude vždy pravdivý už z principu. Logici takovýmto formálně vždy platným výrazům říkají tautologie, někdy se značí T jako Tautologie nebo taky True.

Rozmyslete si, že pro libovolný výrok p je naopak tvrzení $p \wedge \neg p$ vždy nepravdivé, třeba „Teď tady prší a neprší“. Tomu se v logice říká kontradikce a výrok, který je vždy nepravdivý, se značí F jako False.

Vraťme se k našemu příkladu. Všimněte si, že výraz $\neg p \vee q$ má přesně stejné pravdivostní hodnoty jako implikace $p \Rightarrow q$. To znamená, že z pohledu logiky jsou tyto dva výrazy naprostě rovnocenné a můžeme je zaměňovat dle libosti. Ten druhý výraz není tak výstižný, ale zase je v některých situacích praktičtější, protože disjunkce a negace jsou jednodušší operace, pohodlnější při úpravách.

Podobně se dá ekvivalence $p \Leftrightarrow q$ nahradit dvěma implikacemi $p \Rightarrow q$ a $q \Rightarrow p$, ostatně samo značení to naznačuje. Tyto implikace pak případně můžeme dále nahradit pomocí triku z předchozího odstavce.

V běžných matematických knihách by toto vyjádřili pomocí rovnosti, například takto: $(p \Rightarrow q) = (\neg p \vee q)$, ale to není úplně korektní z hlediska logického, ve formální logice se totiž musí dát pozor na správný význam rovnítka. Matematici z jiných oborů, kteří logiku používají jen jako nástroj, nebývají při jejím použití tak formální, díky čemuž ušetří spoustu času různými šikovnými zkratkami, například tím rovnítkem. Tohle je matematická kniha, takže budeme většinu těch zkrátek používat také, ale u logické rovnosti dvou výrazů uděláme výjimku. Ve formální logice se pro to zavádí značení \models , a protože se dá čekat, že studenti diskrétní matematiky formální logiku potkají, nebudeme jim v tom dělat zmátek a použijeme to také. Naše pozorování by se tedy zapsala takto:

- $(p \Rightarrow q) \models (\neg p \vee q)$
- $(p \Leftrightarrow q) \models ([p \Rightarrow q] \wedge [q \Rightarrow p])$

Z pohledu praktického se tato „rovnost“ chová stejně jako rovnost v algebře, vztah \models nám umožňuje provádět s výroky logické úpravy a zjednodušování, podobně jako pomocí rovnítka děláme úpravy s čísly a využíváme identit jako $1 + 3 = 4$. A podobně jako pro algebraické operace máme určitá pravidla, existují (a jsou užitečná) i pro operace logické.

1a.1 Pravidla

Začneme tím nejednodušším.

- $p \wedge p \models p, \quad p \vee p \models p, \quad \neg \neg p \models p;$
- $p \wedge \neg p \models F, \quad p \vee \neg p \models T;$
- $p \wedge T \models p, \quad p \vee F \models p;$
- $p \wedge F \models F, \quad p \vee T \models T.$

Většina se dá rozmyslet selským rozumem. První vztahy ukazují, že zdvojováním výroků ani zdvojením negace nic nezískáme. Druhý rádek jsme již potkali (prší a/nebo neprší). Další dva rádky kombinují obecný výrok p s výroky, které jsou vždy pravda či vždy nepravda. I zde je to jasné po kratším rozmyšlení, například výrok $p \wedge T$ je pravdivý přesně tehdy, když jsou pravdivé oba výroky p a T , ale T je pravdivý vždy, takže pravdivost obou závisí čistě na pravdivosti p . Takové identity se občas hodí při úvahách, jak ještě uvidíme například v kapitole 2.

Jak už jsme obecně vyložili v úvodu, není cílem se všechny takovéto vlastnosti učit nazpaměť. Důležité je rozumět logice natolik, aby si to potřebné dokázal člověk rozmyslet ve chvíli, kdy to potřebuje. Pak ale pomůže, když to již předtím někde viděl a že je v takovém rozmyšlení trénovaný. Platí to i pro většinu vlastností zmíněných dále.

Podívejme se na další vlastnosti logických spojek:

- $p \wedge q \models q \wedge p, \quad p \vee q \models q \vee p;$
- $p \wedge (q \wedge r) \models (p \wedge q) \wedge r, \quad p \vee (q \vee r) \models (p \vee q) \vee r;$
- $p \wedge (q \vee r) \models (p \wedge q) \vee (p \wedge r), \quad p \vee (q \wedge r) \models (p \vee q) \wedge (p \vee r).$

Odborně řečeno, první odrážka ukazuje komutativitu spojek, druhá zase asociativitu a třetí ukazuje distributivní zákon. Tím prvním jsme asi nikoho nepřekvapili, i asociativitu znáte z běžných algebraických operací. Je užitečná například proto, že nám ušetří závorky, při opakování stejně spojce je díky asociativitě vůbec nemusíme psát, třeba takto: $p \wedge q \wedge r \wedge s$. Každý si to teď může ozávorkovat dle libosti.

Distributivní zákon vlastně student dobře zná, dotyčné logické spojky jsou spřízněny s operacemi, \wedge se chová jako násobení a \vee jako sčítání a první rovnost třetí odrážky je pak vlastně jen roznásobení závorky. Druhá rovnost je pak zajímavá, ukazuje, že u logiky se dá roznásobovat i v opačném umístění operací, což pro sčítání a násobení rozhodně neplatí.

Důležitá jsou také pravidla, která nám umožňují negovat operace.

- $\neg(\neg p) \models p;$
- $\neg(p \wedge q) \models \neg p \vee \neg q;$
- $\neg(p \vee q) \models \neg p \wedge \neg q;$
- $\neg(p \Rightarrow q) \models p \wedge \neg q;$
- $\neg(p \Leftrightarrow q) \models (p \wedge \neg q) \vee (\neg p \wedge q).$

Druhý a třetí vztah se jmenují **de Morganovy zákony** a budou se nám v knize hodit, je dobré si je pamatovat. Nejsou vlastně ničím záhadným. Představme si konjunkci $p \wedge q$. Ta platí, pokud jsou pravdivé obě složky p a q . Jestli ji tedy chceme zneplatnit, tak stačí zneplatnit alesoň jednu z těch složek. Konjunkce je tedy neplatná (znegovaná), když neplatí p nebo neplatí q , což je přesně to, co je na pravé straně v druhém řádku.

Podobně se dá rozmyslet i negace implikace. Potřebujeme vystihnout situaci, kdy implikace neplatí, a to je přesně situace, kdy platí předpoklad a neplatí závěr. Dá se k tomu dojít i formálně pomocí přepisu implikace na rovnocenný výraz a pak použitím de Morganových zákonů:

$$\neg(p \Rightarrow q) \models \neg(\neg p \vee q) \models (\neg\neg p \wedge \neg q) \models (p \wedge \neg q).$$

Negace ekvivalence plyne z toho, že ekvivalence je vlastně oboustranná implikace, takže pokud se má pokazit, musí se pokazit jedna (či obě) z těch implikací, pro takovou negaci máme vzoreček o řádek výše. Dá se to také rozmyslet přímo, ekvivalence platí, pokud mají p a q stejnou pravdivostní hodnotu, negace tedy musí popisovat vztah, kde je jeden z p, q pravdivý a druhý ne. Můžeme si to odvodit i formálně, alespoň si naše nová pravidla trochu procvičíme.

$$\neg(p \Leftrightarrow q) \models \neg[(p \Rightarrow q) \wedge (q \Rightarrow p)] \models \neg(p \Rightarrow q) \vee \neg(q \Rightarrow p) \models (p \wedge \neg q) \vee (q \wedge \neg p).$$

Na závěr si ukážeme ještě dva zajímavé vztahy. Následující implikace jsou vždy pravdivé.

- $p \Rightarrow (p \vee q), \quad (p \wedge q) \Rightarrow p.$

Podíváme se na tu první, rozebereme si možnosti pro p . Víme už, že pokud p neplatí, tak je celá implikace automaticky pravdivá. Pokud by p platilo, pak také platí i $p \vee q$ a implikace zase platí. Zkusíme to ukázat ještě pomocí pravdivostní tabulky.

p	q	$p \vee q$	$p \Rightarrow (p \vee q)$
1	1	1	1
1	0	1	1
0	1	1	1
0	0	0	1

Daný výrok je tedy opravdu vždy pravdivý (je to tautologie), našim značením $[p \Rightarrow (p \vee q)] \models T$. Rozmyslete si a tabulkou ověřte i platnost druhého výroku.

K čemu takováto tautologie může být? Když máme implikaci, která je vždy pravdivá, můžeme ji použít v logických úvahách, například pokud je dána pravdivost $p \wedge q$, tak můžeme podle první tautologie s jistotou odvodit platnost p . Ale to se již dostáváme k následujícím tématu.

1a.2 Logika v aplikacích

Formální logika se nezabývá obsahem výroků, proto nám tautologie neboli výroky, jejichž platnost plyne čistě z formální logiky, málodky dají nějakou opravdu užitečnou informaci v aplikacích. Tam se snažíme logiku použít k poznávání světa (či jiných věcí) a obsah výroků začíná být důležitý. V aplikacích nás proto zajímají výroky, jejichž

pravdivost (v rámci zkoumaného světa!) vyplývá z informací, nikoliv chladné logiky. Jinak řečeno, zajímají nás situace, kdy ony rádky v tabulce, kde má dotyčný výrok nuly, jsou jen virtuální, nemohou nastat ve skutečnosti. Například výrok „Sekačka na trávu mi omylem usekla před rokem hlavu a já teď bez ní normálně chodím do školy“ je evidentně nepravdivý (alespoň tedy za současného stavu technologie). Něco pravdivého: „Jestliže je dnes 29. února, pak je rok dělitelný čtyřmi.“

Zkusme si to rozebrat. Pokud opravdu je 29. února, pak musí nutně podle našeho kalendáře platit i to o dělitelnosti roku a implikace platí. Pokud naopak 29. února není, pak není splněn předpoklad implikace a tato implikace jako celek je tedy pravdivá. To znamená, že reálně nemůže nastat situace, kdy by tato konkrétní implikace neplatila.

Právě taková vždy pravdivá tvrzení nás zajímají nejvíce a většina matematických tvrzení je tohoto typu, pravdivé implikace. Ukážeme vzájemnou propojenosť $p \Rightarrow q$ určitých věcí a ona pak čeká na uživatele. Pokud by se někomu přihodilo, že p platí, pak díky naší práci už hned ví, že mu platí i q . Například ona implikace o 29. únoru čeká na někoho, kdo si na ni jednoho krásného 29. února vzpomene, pak už rovnou ví něco o dělitelnosti roku, aniž by vlastně věděl, jaký rok přesně je. Čtenář se už setkal s implikací jinou, určitě ví, že když má kvadratickou rovnici a vyjde mu číslo $b^2 - 4ac$ kladné, tak už ta rovnice má dva reálné kořeny.

Všimněte si mimochodem, že v jiném kulturním okruhu, kde mají kalendář jinak, naše implikace pravdivá být nemusí. Když tedy v běžném hovoru mluvíme o tom, že určitý výrok je pravdivý, tak tím myslíme, že je vždy pravdivý v rámci dotyčného světa, přičemž ten svět nemusí být náš fyzický, ale třeba abstraktní matematický (viz zbytek této knihy). Pravdivost tvrzení tedy závisí na pravidlech, která fungování dotyčného světa určují. My se k tomu blíže dostaneme v povídání o axiomech v kapitole 3 o indukci.

Ted' se podíváme na to, jak se z tvrzení, o kterém víme, že je (vždy) pravdivé, dá něco získat díky znalosti formální logiky. Nejčastěji se takto informačně vytěžují **implikace**, které jsou v matematice klíčové, tak se na ně zaměříme.

Máme-li pravdivou implikaci $p \Rightarrow q$, pak části p říkáme „postačující podmínka“ a části q říkáme „nutná podmínka“. Je to vlastně rovnocenné vyjádření, říct, že „ p je postačující podmínka pro q “, je z hlediska logiky stejně, jako říct, že $p \Rightarrow q$ je vždy pravdivá, což je stejně, jako říct „ q je nutná podmínka pro p “. Jak ještě uvidíme, pro matematiky je tato terminologie velmi užitečná. Odkud se to vzalo?

Je to přirozené. Pokud vím, že je dnes 29. února, tak to postačuje k jistotě, že je rok dělitelný čtyřmi. Toto je také hlavní způsob využití implikace, který jde zpět minimálně o tisíc let. Tradiční zápis vypadá nějak takto:

- Platí: Jestliže je dnes 29. února, pak je rok dělitelný čtyřmi.
- Předpokládáme, že platí: Je 29. února.
- Závěr: Tento rok je dělitelný čtyřmi.

Ve středověku tomu říkali „modus ponens“. Využití implikace je tedy následující. Zajímá nás, zda je pravdivé tvrzení q , ale do jeho zkoumání se nám nechce. Naštěstí dostaneme k dispozici pravdivou implikaci $p \Rightarrow q$, kde p je mnohem příjemnější. Podíváme se tedy na p a pokud se ukáže pravdivým, pak je nutně pravdivé i q . Skvělé. Má to ale zádrhel. Pokud by se ukázalo, že p neplatí, tak o q nic nevíme. U našeho příkladu je to jasné:

- Platí: Jestliže je dnes 29. února, pak je rok dělitelný čtyřmi.
- Předpokládáme, že platí: Není 29. února.
- Závěr: ????

Implikace tedy dokáže posouvat informaci z p do q , ale je to nástroj nedokonalý, posílá jen jeden typ informace. Tato nespolehlivost dokáže někdy docela potrápit. Tento problém se projeví ještě jedním způsobem. Máme-li pravdivou implikaci $p \Rightarrow q$, tak rozhodně neplatí, že závěr q je postačující pro q . Zase je to vidět na našem příkladě:

- Platí: Jestliže je dnes 29. února, pak je rok dělitelný čtyřmi.
- Předpokládáme, že platí: Je rok dělitelný čtyřmi.
- Závěr: ????

Je to jasné, v roce dělitelném čtyřmi nemusí dnes být zrovna 29. února, klidně může být červen, dokonce v tom roce ani žádný 29. únor nemusí být (protože roky, které jsou kromě čtyř dělitelné i stem, ale už ne čtyřmi sty, žádný přechodný den nemají).

Toto se dá jinak vyjádřit pozorováním, že pravdivou implikaci nelze beztrestně obrátit. Mějme implikaci $p \Rightarrow q$, o které víme třeba to, že je (vždy) pravdivá. V okamžiku, kdy se ji pokusíme obrátit, tak všechny informace ztratíme, $q \Rightarrow p$ je úplně jiná věc, jejíž pravdivost nemá obecně s pravdivostí $p \Rightarrow q$ nic společného. Často tak dostaneme implikaci nepravdivou, jen v některých případech zjistíme, že je pravdivá $p \Rightarrow q$ i $q \Rightarrow p$. Pak musí mít p a q stejnou pravdivostní hodnotu a dostáváme vlastně ekvivalenci $p \Leftrightarrow q$, ale to musíme mít hodně štěstí. V obecném případě rozhodně implikace obracet nesmíme.

Teď se podíváme na vyjádření slovy nutná podmínka. Je jasné, že abychom si mohli užívat 29. února, tak musí být rok dělitelný čtyřmi, je to tedy nutné, bez toho to nejde. Nutné podmínky nebývají tak populární jako podmínky postačující, protože když víme, že je q pravdivé, tak nám to nepomůže odvodit nic o p kromě toho, že nám to umožnuje mít p pravdivé. Přesto ale bývají nutné podmínky užitečné. Když zkoumáme, za jakých okolností platí p , tak nám pravdivá implikace $p \Rightarrow q$ umožňuje omezit se pouze na situace, kde platí q , protože jinak rovnou víme, že p neplatí.

Už jsme poznamenali, že q postačující podmínkou není, tedy ze znalosti o pravdivosti q nic o p neodvodíme. Viděli jsme ale, že znalost o neplatnosti q již dává neplatnost p , takže dokážeme posílat jistou informaci z q na p . Vypadá to u našeho příkladu takto.

- Platí: Jestliže je dnes 29. února, pak je rok dělitelný čtyřmi.
- Předpokládáme, že platí: Není rok dělitelný čtyřmi.
- Závěr: Dnes není 29. února.

Za středověku tomuto typu úvahy říkali „modus tollens“. Vlastně tak dostáváme novou implikaci.

- Máme: Jestliže je dnes 29. února, pak je rok dělitelný čtyřmi.
- Dostali jsme: Jestliže není rok dělitelný čtyřmi, pak dnes není 29. února.

Uděláme si to formálně. Je-li dána implikace $p \Rightarrow q$, pak se definuje její **obměna** jako $\neg q \Rightarrow \neg p$. Pomocí pravdivostní tabulky se hravě dokáže, že implikace a její obměna mají vždy nutně stejnou pravdivostní hodnotu, jde tedy o ekvivalentní výroky a lze je libovolně zaměňovat. Přechod k obměně je občas velice užitečný a setkáme se s ním u nepřímého důkazu níže.

Čtenář si může předchozí úvahy procvičit na implikaci „Jestliže je mi přes 21, tak mi je přes 18“ neboli „Být přes 21 je postačující k tomu, abych byl přes 18“. Ta je dozajista pravdivá, ať už ji řekne kdokoliv. Je na ní zajímavé, že 21 bývá v mnoha zemích legálním věkem pro pití alkoholu, zatímco 18 bývá věkem pro získání řidičského průkazu. Tuto implikaci lze tedy vyjádřit slovy „Jestliže mohu pít, tak mohu řídit“, což je z logického pohledu pravdivá implikace, ale zní to podezřele.

Ekvivalence $p \Leftrightarrow q$ je podle definice situace, kdy mají výroky p a q vždy stejnou pravdivost. Již jsme uvedli, že pak máme implikace oběma směry. To znamená, že p je nutnou i postačující podmínkou pro q a naopak, ekvivalence je již z podstaty symetrická. V praxi je to skvělé, cokoliv se dozvíme o p platí i pro q . Jak se dá čekat, získat pravdivou ekvivalence je mnohem obtížnější než získat pravdivou implikaci.

1a.3 Predikátová logika

Opravdu užitečné výroky nejsou „absolutní“, ale mají proměnnou. Už i to „teď tady prší“ vlastně mělo dvě proměnné, místo a čas, a podle toho, kde a kdy jsme tento výrok řekli, se měnila jeho pravdivost. Výroky s proměnnými značíme $p(x)$, $p(x, y)$ a podobně. Matematika je výroků s proměnnými plná. Třeba „ $x > 13$ “ někdy platí a někdy ne, podle toho, co dáme za x . Výrok „pro $x = 23$ je $x > 13$ “ je pravdivý, výrok „pro $x = 3$ je $x > 13$ “ pravdivý není. Toto ale moc užitečné není. Z hlediska teorie nás zajímají hlavně výroky, které by měly pravdivost stálou bez nutnosti volit nějaké konkrétní x . U výroků s jednou proměnnou se tak studují dvě situace:

1. Chceme vědět, jestli náhodou takový výrok neplatí úplně vždy, bez ohledu na naši volbu proměnné. To se vyjadřuje slovy „pro každé x platí $p(x)$ “ a máme i pohodlnou zkratku „ $\forall x: p(x)$ “. Jestliže si například vezmeme výrok „ $x^2 \geq 0$ “, pak je pravdivý, ať už za x zvolíme jakékoli reálné číslo. Běžný matematik by tedy napsal toto:

$$\forall x \in \mathbb{R}: x^2 \geq 0.$$

Čteme to: Pro každé x z množiny reálných čísel platí, že $x^2 \geq 0$. Ta dvojtečka je tedy jen zkratka pro slovo „platí“. Logici by zase měli nepříjemný pocit, hlavně z té dvojtečky, a dávali by přednost zápisu

$$\forall x (x \in \mathbb{R} \Rightarrow x^2 \geq 0).$$

Pro běžné matematiky (nelogiky) je první zápis dostačující a normálně jej používají, nicméně jsou situace (viz třeba důkaz Faktu 2a.1), kdy ten přesný logický zápis pomůže. Většinou ale i zde raději použijeme ten stručnější zápis, odpovídá lépe přirozenému jazyku. Podstatné je tomu zápisu dobře rozumět.

2. Mnoho tvrzení s argumentem ovšem takto pěkných není, aby platilo vždy, třeba zrovna to $x > 13$. Matematici se pak ptají, jestli by toto nemohlo být pravda alespoň někdy. Zkoumaný výrok je „existuje x , pro které platí $p(x)$ “ a zapisujeme jej „ $\exists x: p(x)$ “. Příklad pravdivého výroku:

$$\exists x \in \mathbb{R}: x > 13.$$

Naopak $\exists x \in \mathbb{R}: x = x + 1$ je výrok nepravdivý, protože $x = x + 1$ se nedá splnit žádnou volbou reálného čísla x . I zde by logici raději viděli

$$\exists x (x \in \mathbb{R} \wedge x = x + 1)$$

a i zde je v této knize budeme opakovaně zklamávat. Značkám $\forall x$ a $\exists x$ se říká **kvantifikátory**, ten první je obecný, ten druhý existenční. Když je otočíte vzhůru nohama, dostanete písmena A a E jako „All“ a „Exists“, dobře se to pamatuje.

Zde bychom mohli uvést pravidla popisující, jak kvantifikátory interagují s logickými spojkami, ale podobně jako předtím není důležité si ta pravidla pamatovat, ale zamyslet se nad nimi a porozumět jim. Necháváme je proto jako cvičení 1a.3.

Důležité pro nás bude, jak se takové věci dokazují. Odpověď je jednoduchá: Přesně tak, jak by člověk čekal. V důkazu výroku $\forall x \in M: p(x)$ musíme nějak odůvodnit, že ať už si vezmeme z M cokoliv, vždy pro tento prvek bude platit p . Naopak k důkazu výroku $\exists x \in M: p(x)$ stačí nějaké takové x najít. Důkaz výroku $\exists x \in \mathbb{R}: x > 13$ zní: Zkuste $x = 14$.

Někdy potřebujeme naopak dokázat, že výrok neplatí, což je totéž jako dokazovat, že platí negace výroku. Pro negace výroků s kvantifikátory máme následující pravidla:

- $\neg(\forall x \in M: p(x)) \models \exists x \in M: \neg p(x);$
- $\neg(\exists x \in M: p(x)) \models \forall x \in M: \neg p(x).$

Opět je to selský rozum. Na to, abychom ukázali, že se někdo mylí, když tvrdí, že vždy platí p , mu stačí ukázat jedený případ, kdy tomu tak není, což je přesně první řádek. Tomu jednomu příkladu se pak říká **protipříklad**.

Podobně jako u implikace, i zde máme jeden zvláštní případ, který se moc nevyskytuje, ale když už na něj narazíme, měli bychom vědět, co s ním. Jak fungují kvantifikátory, když se odkazují na prázdnou množinu? Jinými slovy, bude výrok $\forall x \in \emptyset: p(x)$ pravdivý? A výrok $\exists x \in \emptyset: p(x)$? Odpověď zní, že první vždy ano, druhý vždy ne. U toho druhého je to jasné, v prázdné množině nic nenajdeme, tudíž existence speciálního prvku nemůže být splněna, ale u toho prvního to tak jasné není. Jedna možnost je si říct, že vlastně ten výrok nejde pokazit, protože nenajdeme v prázdné množině x takové, aby pro něj p neplatilo. Jestliže nejde výrok pokazit, tak musí platit opak, tedy výrok je pravdivý. Naprostě jasné je to z toho správně logického přepisu $\forall x (x \in \emptyset \implies x^2 \geq 0)$. Předpoklad dotyčné implikace je vždy nepravdivý, tudíž je celá implikace automaticky pravdivá.

Zajímavá situace je, když máme výrok s více proměnnými. Pokud je uvozujeme kvantifikátory stejného typu, pak na pořadí nezáleží a obvykle je sloučíme do jednoho (pokud vybíráme ze stejné množiny):

$$(\forall x \in \mathbb{R} \forall y \in \mathbb{R}: p(x, y)) \models (\forall y \in \mathbb{R} \forall x \in \mathbb{R}: p(x, y)) \models (\forall x, y \in \mathbb{R}: p(x, y)); \\ (\exists x \in \mathbb{R} \exists y \in \mathbb{R}: p(x, y)) \models (\exists y \in \mathbb{R} \exists x \in \mathbb{R}: p(x, y)) \models (\exists x, y \in \mathbb{R}: p(x, y)).$$

Například $\forall x, y \in \mathbb{R}: x^2 + y^2 \geq 0$ je pravdivý výrok. Naopak $\forall x, y \in \mathbb{R}: x^2 + y^2 = 5^2$ pravdivý výrok není. Volba $x = 3, y = 4$ ovšem ukazuje pravdivost výroku $\exists x, y \in \mathbb{R}: x^2 + y^2 = 5^2$.

Složitější situace je, když se smíchají kvantifikátory rozličných druhů, pak totiž na pořadí velice záleží. Základem je rozmyslet si dobré situaci pro dva kvantifikátory. Ukážeme si to na příkladech.

Výrok $\forall x \in \mathbb{R} \exists y \in \mathbb{R}: x^2 = 4y^2$ říká: „Pro každé reálné číslo x existuje reálné číslo y takové, že $x^2 = 4y^2$.“ Již samotná forma naznačuje, že y hledáme vždy k jistému konkrétnímu x . Vezmeme nějaké x a hledáme k němu y splňující specifikovanou vlastnost. Pak vezmeme jiné x a klidně se může stát (ale není to nutné), že k němu najdeme jiné y s požadovanou vlastností (nebo také nenajdeme, podle toho, zda daný výrok platí či ne). Je nás výrok vlastně platný? Ano. Když nám někdo dá x , tak stačí zvolit $y = \frac{1}{2}x$ a vlastnost je splněna, opravdu pak $x^2 = 4y^2$. Je také možné volit $y = -\frac{1}{2}x$, což ale vůbec nevadí, výrok toho neřeší. Pro něj je podstatné, aby alespoň nějaké takové y vždy pro dané x bylo. Můžeme si také všimnout, že pro některá x dostáváme stejné možnosti na y , třeba pro $x = 6$ a $x = -6$ máme vždy pro y kandidáty ± 3 . Ani to ten výrok neřeší.

U tohoto pořadí kvantifikátorů tedy pravdivost výroku znamená, že vzniká jakési přiřazení $x \mapsto y$, které ale nemusí být jednoznačné.

Teď se podíváme na opačné pořadí: Výrok $\exists y \in \mathbb{R} \forall x \in \mathbb{R}: x^2 = 4y^2$ říká: „Existuje reálné číslo y takové, že pak pro každé reálné číslo x platí $x^2 = 4y^2$.“ Zde již čeština naznačuje, že číslo y musí být univerzální, jedno číslo pro všechna x . Je zjevné, že v tomto případě takové univerzální číslo y nenajdeme. Vidíme tedy, že prohozením kvantifikátorů došlo ke změně pravdivosti výroku. Příklad pravdivého výroku tohoto typu: $\exists x \in \mathbb{R} \forall y \in \mathbb{R}: (|y| + 1)^x = 1$. Stačí totiž zvolit $x = 0$ a vlastnost bude pro všechna reálná y platit.

Pro praktickou práci v matematice je proto důležité rozumět dobře těmto kombinacím kvantifikátorů a hlavně si pamatovat, že pořadí nelze zaměňovat. Pečlivější čtenář si nicméně může všimnout, že alespoň něco říct lze, jmenovitě platí toto:

- $[\exists x \in \mathbb{R} \forall y \in \mathbb{R}: p(x, y)] \implies [\forall y \in \mathbb{R} \exists x \in \mathbb{R}: p(x, y)].$

Jinými slovy, jestliže máme univerzálně fungující prvek x , pak tento prvek bude samozřejmě také fungovat individuálně pro jednotlivce. Pro další pravidla, která jsou někdy užitečná, se podívejte na cvičení 1a.3. Jako obvykle nemá smysl se je učit, spíš si dobré rozmyslete, proč to vlastně nemůže být jinak, než je tam řečeno.

Existenční kvantifikátor má jednu užitečnou modifikaci. Když se za něj přidá vykřičník, tak se to čte „existuje právě jedno“, je to tedy spojení dvou věcí, „existuje“ a „není jich víc“. Například výrok $\exists!x \in \mathbb{R}: x + 1 = 14$ je pravdivý, tato rovnice má přesně jedno řešení, ale výroky $\exists!x \in \mathbb{R}: x^2 = 13$ a $\exists!x \in \mathbb{R}: x^2 = -13$ pravdivé nejsou. Ve formální logice tento kvantifikátor neexistuje, takže se náspravidlo musí zapsat například takto:

$$[\exists x (x \in \mathbb{R} \wedge x + 1 = 14)] \wedge \neg[\exists x, y (x \in \mathbb{R} \wedge y \in \mathbb{R} \wedge x \neq y \wedge x + 1 = 14 \wedge y + 1 = 14)].$$

Už asi chápete, proč obyčejný matematik-nelogik radostně sáhne po $\exists!$, i když nutno přiznat, že logici mají dobré důvody, proč to do formální logiky nepřibírat.

O logice by se toho dala napsat spousta. Existuje více pravidel, existují další operace (užitečné v některých aplikacích), to už vůbec nemluvíme o tématech, která jsme tu ani nenačali (zvídavému čtenáři doporučujeme přečíst si nějakou pěknou knížku), ale pro běžnou matematickou práci v zásadě stačí to, co vidíme výše.

Na to, jak se logika v matematice opravdu používá, se blíže podíváme v příští kapitole, a v praxi to pak uvidíme ve větší či menší míře ve všech kapitolách následujících.

Cvičení

Cvičení 1a.1: Připomeňme, že \mathbb{R} značí množinu všech reálných čísel a \mathbb{Z} množinu všech celých čísel. Rozhodněte, zda jsou pravdivé následující výroky:

- | | |
|--|--|
| (i) $\forall x \in \mathbb{R}: (x \geq 3 \vee x < 5);$ | (ix) $\forall x \in \mathbb{R} \exists y \in \mathbb{R}: x + y = 0;$ |
| (ii) $\exists x \in \mathbb{R}: (x \geq 3 \wedge x < 0);$ | (x) $\exists y \in \mathbb{R} \forall x \in \mathbb{R}: x + y = 0;$ |
| (iii) $\forall x \in \mathbb{Z}: (x > 3 \wedge x < 7);$ | (xi) $\forall x \in \mathbb{R} \exists y \in \mathbb{R}: x \cdot y = 0;$ |
| (iv) $\exists x \in \mathbb{Z}: (x \geq 3 \wedge x < 5);$ | (xii) $\exists y \in \mathbb{R} \forall x \in \mathbb{R}: x \cdot y = 0;$ |
| (v) $\forall x \in \mathbb{R}: (x > 3 \implies x^2 > 9);$ | (xiii) $\forall x \in \mathbb{R} \exists y \in \mathbb{R}: \frac{x}{y} = 1;$ |
| (vi) $\forall x \in \mathbb{R}: (x^2 > 9 \implies x > 3);$ | (xiv) $\exists y \in \mathbb{R} \forall x \in \mathbb{R}: \frac{x}{y} = 1;$ |
| (vii) $\forall x \in \mathbb{R}: (x^2 < 0 \implies x = 13);$ | (xv) $\forall x \in \mathbb{R} \exists y \in \mathbb{R}: x < 3y;$ |
| (viii) $\exists x \in \mathbb{R}: (x \geq 5 \implies x^2 = 40);$ | (xvi) $\exists y \in \mathbb{Z} \exists x \in \mathbb{Z}: x^2 - y^2 = 3.$ |

Cvičení 1a.2: Následující výrazy s proměnnou z \mathbb{R} upravte pomocí distributivního zákona a pak zjednodušte:

- (i) $x > 3 \wedge (e^x = x^5 \vee x = 4);$
- (ii) $x < 13 \wedge (\sin(x) < \frac{1}{3} \vee x > 14);$
- (iii) $(\sin(x) < x^3 \wedge x < 3) \vee (\sin(x) < x^3 \wedge x > 1).$

Cvičení 1a.3: Rozhodněte, zda platí obecně (tedy pro libovolné množiny M a výroky p, q) následující tvrzení o logické ekvalenci výroků. Pokud máte pocit, že některá dvojice výroků ekvivalentní není, tak najděte příklad takových výroků p, q , aby jeden z výroků platil a druhý ne.

- (i) $\forall x \in M: p(x) \wedge q(x) \models [\forall x \in M: p(x)] \wedge [\forall x \in M: q(x)];$
- (ii) $\forall x \in M: p(x) \vee q(x) \models [\forall x \in M: p(x)] \vee [\forall x \in M: q(x)];$
- (iii) $\exists x \in M: p(x) \wedge q(x) \models [\exists x \in M: p(x)] \wedge [\exists x \in M: q(x)];$
- (iv) $\exists x \in M: p(x) \vee q(x) \models [\exists x \in M: p(x)] \vee [\exists x \in M: q(x)];$
- (v) $p \wedge \forall x \in M: q(x) \models \forall x \in M: p \wedge q(x);$
- (vi) $p \vee \forall x \in M: q(x) \models \forall x \in M: p \vee q(x);$
- (vii) $p \wedge \exists x \in M: q(x) \models \exists x \in M: p \wedge q(x);$
- (viii) $p \vee \exists x \in M: q(x) \models \exists x \in M: p \vee q(x).$

Cvičení 1a.4: Znemocněte formálně následující výroky. Pro každý výrok i jeho negaci si pak zvlášť rozmyslete, zda platí či ne, abyste se přesvědčili, že vždy mají opačnou pravdivost.

- (i) $\exists x \in \mathbb{R}: x > 5;$
- (v) $(\exists x \in \mathbb{R}: x = \frac{x}{2}) \implies (\forall x \in \mathbb{R}: x = 13x);$
- (ii) $\forall x \in \mathbb{Z}: (x > 5 \vee x^2 = 14);$
- (vi) $\exists x \in \mathbb{R} \forall y \in \mathbb{R}: x < y;$
- (iii) $\exists x \in \mathbb{R}: (x < 3 \implies x = x - 1);$
- (vii) $\forall x \in \mathbb{R} \forall y \in \mathbb{R}: x^2 + y^2 \geq 0;$
- (iv) $(\forall x \in \mathbb{R}: x^2 \geq 0) \implies (\forall x \in \mathbb{R}: x < 0);$
- (viii) $\forall x \in \mathbb{R} \exists y \in \mathbb{R}: \sin(x) = \cos(y).$

Řešení:

1a.1: (i): platí; (ii): neplatí (podmínky se vylučují); (iii): neplatí (pro některá x obě nerovnosti platí, ale to nestačí); (iv): platí, $x = 3$ nebo $x = 4$; (v): platí; (vi): neplatí, protipříklad $x = -4$; (vii): platí (předpoklad není nikdy splněn, proto je implikace pravdivá); (viii): platí, $x = \sqrt{40}$ nebo třeba $x = 0$; (ix): platí, $y = -x$; (x): neplatí; (xi): platí, $y = 0$; (xii): platí, $y = 0$; (xiii): neplatí, pro $x \neq 0$ sice najdeme $y = x$, ale pro $x = 0$ to zařídí nejde; (xiv): neplatí; (xv): platí, stačí zvolit třeba $y = |x| + 1$; (xvi): platí, $x = 2$ a $y = 1$ (všimněte si, že kdyby tam bylo $x^2 - y^2 = 2$, tak už by to neplatilo).

1a.2: (i): $\models (x > 3 \wedge e^x = x^5) \vee (x > 3 \wedge x = 4) \models (x > 3 \wedge e^x = x^5) \vee x = 4.$

- (ii): $\models (x < 13 \wedge \sin(x) < \frac{1}{3}) \vee (x < 13 \wedge x > 14) \models \models (x < 13 \wedge \sin(x) < \frac{1}{3}) \vee F \models x < 13 \wedge \sin(x) < \frac{1}{3}$.
 (iii): $\models \sin(x) < x^3 \wedge (x < 3 \vee x > 1) \models \sin(x) < x^3 \wedge T \models \models \sin(x) < x^3$.

1a.3: (i): platí. Oba výrazy vyžadují platnost p i q pro všechna x .

(ii): neplatí, jde jen v jednom směru. Pokud platí výrok napravo, tak už platí i výrok nalevo. Pravý výrok totiž vynutí platnost jednoho z p, q vždy, díky tomu platí i $p \vee q$ vždy. Naopak to ale nejde, pokud platí výraz nalevo, tak je $p \vee q$ splněno vždy, ale nepřinutí to jeden z nich, aby platil vždy. Příklad: $M = \mathbb{R}, p(x): x \geq 13, q(x): x < 13$.

(iii): neplatí, jde jen v jednom směru. Pokud platí výrok nalevo, tak existuje x , pro které platí $p \wedge q$, pro toto x pak platí oba výroky. Naopak to nejde, pokud platí výrok napravo, tak jde p i q nějakou volbou x splnit, ale nikde není zaručeno, že to bude totéž x , aby tak platil i výrok nalevo. Příklad: $M = \mathbb{R}, p(x): x = 13, q(x): x = 14$.

(iv): platí. Výrok nalevo i výrok napravo požadují, aby šlo alespoň jeden p, q alespoň jednou volbou x splnit.

(v): platí. Výrok nalevo i výrok napravo požadují, aby platilo jak p , tak $q(x)$ pro všechna x .

(vi): platí. Pokud platí p , tak jsou pravdivé výroky na obou stranách. Pokud p neplatí, ale $q(x)$ vždy platí, tak jsou zase výroky na obou stranách pravdivé. Pokud p neplatí a také $q(x)$ alespoň pro jedno x neplatí, tak jsou výroky na obou stranách nepravdivé. Mají tedy vždy stejnou pravdivost.

(vii): platí. Oba výroky požadují, aby platilo jak p , tak $q(x)$ pro nějaké x .

(viii): platí. Pokud platí p , tak jsou výroky na obou stranách pravdivé. Pokud p neplatí a q platí alespoň pro jedno x , tak jsou zase výroky na obou stranách pravdivé. Pokud neplatí ani p , ani $q(x)$ pro žádné x , pak jsou výroky na obou stranách nepravdivé. Mají tedy vždy stejnou pravdivost.

1a.4: (i): negace: $\forall x \in \mathbb{R}: x \leq 5$. Výrok platí, negace ne, třeba $x = 7$.

(ii): negace: $\exists x \in \mathbb{Z}: (x \leq 5 \wedge x^2 \neq 14)$. Výrok neplatí, negace ano, třeba $x = 2$.

(iii): negace: $\forall x \in \mathbb{R}: (x < 3 \wedge x \neq x - 1)$. Výrok platí (třeba $x = 0$, pak má implikace nesplněný předpoklad a tudíž platí), negace ne.

(iv): negace: $(\forall x \in \mathbb{R}: x^2 \geq 0) \wedge (\exists x \in \mathbb{R}: x \geq 0)$. Výrok neplatí, negace ano.

(v): negace: $(\exists x \in \mathbb{R}: x = \frac{x}{2}) \wedge (\exists x \in \mathbb{R}: x \neq 13x)$. Výrok neplatí (předpoklad splněn $x = 0$, závěr ne), negace ano (první výrok splněn $x = 0$, druhý také $x = 1$).

(vi): negace: $\forall x \in \mathbb{R} \exists y \in \mathbb{R}: x \geq y$. Výrok neplatí (to by muselo existovat jedno číslo, které je nejmenší ze všech reálných), negace ano (pro dané x stačí zvolit $y = x - 1$).

(vii): negace: $\exists x \in \mathbb{R} \exists y \in \mathbb{R}: x^2 + y^2 < 0$. Výrok platí, negace ne.

(viii): negace: $\exists x \in \mathbb{R} \forall y \in \mathbb{R}: \sin(x) \neq \cos(y)$. Výrok platí (pro dané x stačí zvolit $y = \arccos(\sin(x))$), negace ne (ať zkusíme jakékoli x , vždy nám jeho volbu zkazí nějaké y , které se hodnotou cosinu trefí do $\sin(x)$).

1b. Logika a matematika

Zde se podíváme, jak požadavek na přesnost a logickou správnost ovlivňuje strukturu matematiky a naopak jak matematika ovlivňuje použití logiky. Ukážeme také praktické rady, které ocení zejména ti, kdo se budou snažit psát důkazy. Některé části možná čtenář lépe docení, když se k nim zase vrátí po nabytí zkušeností v několika dalších kapitolách.

Matematika se dělí na rozličné obory podle toho, jaké objekty se zkoumají, například (zhruba řečeno) analýza zkoumá funkce, algebra struktury s operacemi, lineární algebra lineární prostory atd. Matematika ovšem nezkoumá objekty konkrétní, ale objekty abstraktní, dá se říci typy objektů. Každý výklad určitého oboru tak musí začít úmluvou, jaké objekty se budou zkoumat. Protože chceme, aby závěry matematiky byly naprostě spolehlivé, musí být také popis zkoumaných objektů zcela přesný, aby bylo vždy jasné, co je pravda a co ne (abstraktní svět matematiky je nutně černobílý).

Rozmyslete si například, že není možné se spolehlivě rozhodnout, zda autor této knihy je normální, protože nikdo přesně neví, co to vlastně je (hodně lidí si myslí, že ví, co je to normální, ale nějak se neshodnou). Zato všichni poznají, co je to sudé číslo, protože na to je přesná specifikace.

Specifikaci nových pojmu se v matematice říká **definice**. Většinou je nový pojem charakterizován vlastností, podle které se dá jednoznačně poznat. Můžeme například říct, že číslo x je sudé, pokud existuje celé číslo k takové, že $x = 2 \cdot k$. Když nám pak někdo dá číslo, tak se prostě podíváme, zda jej je nebo není možno zapsat příslušným způsobem, a tím se dozvímme, zda je sudé či ne.

Stojí za to poznamenat, že aby taková definice fungovala, tak ještě předtím musíme mít definice udávající, co je to rovnost a co jsou celá čísla, ty asi budou zase potřebovat další definice atd. Když se matematika dělá opravdu pořádně, dostane se člověk k úplným základům. Na to je speciální obor matematiky, většina matematiků se spokojí s tím, že určité věci už považuje za známé (rovnost, rovnice, základní algebra atd.). Zkusme si tedy představit, že známe reálná čísla a víme, co je to nerovnost, a udělejme následující ukázkovou definici.

Definice.

Nechť x je reálné číslo. Řekneme, že je **kladné**, jestliže platí $x > 0$.

Zkusme si to rozebrat. První věta je uvozovací, říká nám, s jakými objekty budeme pracovat. Je to vlastně kód, myslí se tím, že x může být *libovolné* reálné číslo (někdo to občas i takto doslově napiše, ať v tom má čtenář jasno). Správný překlad do logiky by tedy byl následující: „ $\forall x \in \mathbb{R}$ “.

Pro každé reálné číslo pak můžeme či nemůžeme říct, že je kladné, podle toho, jak dopadne ona definující podmínka. Tím je tento nový pojem přesně vymezen a nemůže se stát, že by dva lidé měli totéž číslo a neshodli se v názoru na to, zda je kladné.

Všimněte si jedné podstatné věci. Již z principu musí být mezi novým pojmem a podmínkou, která jej definuje, vztah ekvivalence: Pokud podmínka funguje, je použití nového pojmu oprávněné. Pokud nefunguje, není možné jej použít. Správně by tedy v definici mělo být napsáno „Řekneme, že x je kladné, právě tehdy když platí $x > 0$ “. Jenže z nějakého důvodu je zvykem psát tam „jestliže“, což vlastně značí implikaci, takže tak, jak je to napsáno, to značí „ $x > 0 \implies x \text{ kladné}$ “. Jinými slovy, je to vlastně napsáno špatně, ale už se to tak dělá nejméně sto let a snad ve všech jazycích, tak do toho nemá smysl vrtat (občas se najde nadšenec, který si dá tu práci a píše opravdu definice jako ekvivalence, ale není jich moc).

V praxi tato nepřesnost nevadí, protože každý matematik ví, že definice se píšou takto a přitom se to bere jako ekvivalence. Je to součástí zasyvěcení do matematiky, začátečníka to ale může zaskočit (pak jsou tu ovšem studenti, kteří nad tím nepřemýšlejí, ti si ničeho nevšimli a tahle poznámku nejspíše vůbec nečtou). Vy už to tajemství znáte, vítejte v tajné lóži matematiků (na krvavý iniciační obřad si počkejte do prvního drsnějšího důkazu).

Poznámka stranou: Co když si někdo zavede jinou definici kladnosti? Pokud k tomu nemá opravdu dobrý důvod, tak jej nikdo nebude brát vážně a ta jeho definice zanikne. Pokud to bude mít dobře podloženo, pak získá následovníky a vytvoří se konkurenční typ matematiky. Naštěstí se to skoro vůbec nestává, protože lidé se při vytváření definic řídí především užitečností. Při hlubším studiu matematiky se na to dá občas narazit, pak člověk ví, že se musí při čtení knihy nejprve dobře podívat, s jakými pojmy autor pracuje. Není to ale až tak velký problém, protože na definici není podstatné jméno, ale význam dotyčného pojmu, tedy myšlenka. Matematik si tedy zjistí, jakou myšlenku dotyčným slovem autor míní, a pak už se jen dívá, co s ní v knize vyvede. Každopádně čtenář se nemusí bát, je téměř jisté, že na takovou situaci v životě nenarází.

Narázíme tím na věc, která možná čtenáře překvapí: Definice si můžeme dělat, jak se nám zlíbí. Představme si, že by ten úplně první člověk, který pojem kladnosti zavedl, namísto toho prohlásil, že kladná čísla jsou taková, která splňují $x^2 = 13$. Co by se stalo? Z hlediska logického i matematického by na tom nebylo nic špatně, jenže problém by byl jinde: Tento pojem by nebyl příliš užitečný, nikdo by jej nepoužíval a brzy by z matematického života vymizel (darwinismus v matematice). Pojmy, které potkáváme, jsou vymyšleny tak, aby nám pomáhaly při práci, přičemž to, že se dožily současnosti, ukazuje, že se jejich autoři dobře trefili.

Definicemi vlastně vytváříme imaginární světy, záleží jen na naší představivosti, kolik a jaké vytvoříme. Úkolem matematiky pak je takové světy zkoumat.

Máme tedy pojmy a posuňme se dále. Cílem matematiky je najít o těchto užitečných pojmech co nejvíce informací. Tyto informace jsou pak sdělovány ve formě tvrzení, která se rozličně jmenují. Důležitá tvrzení se jmenují „věty“, jednoduchá zase „fakta“, používá se také přímo název „tvrzení“. Někdy z jednoho tvrzení hned s minimální prací vyplýne další, tomu pak říkáme „důsledek“. Posledním zajímavým názvem je „lemma“, to používáme pro pomocná tvrzení, často do nich schováváme nudné a pracné části důkazů vět, aby lépe vynikly hlavní myšlenky (viz Lemma 3a.5 a Věta 3c.5). Tato klasifikace je samozřejmě subjektivní a co je u jednoho autora důsledek, může mít jiný jako větu a podobně.

Ukažme si příklad.

Fakt.

Nechť $x \in \mathbb{R}$. Jestliže $x > 0$, pak $x(x + 1) > 0$.

Jako obvykle vidíme uvozovací větu a už jsme si rozmysleli, že je tam schováno slůvko „libovolné“ či „každé“. Trochu přesnější přepis by tedy byl následující:

- Pro každé $x \in \mathbb{R}$ platí: Jestliže $x > 0$, pak $x(x + 1) > 0$.

Ted' už to snadno přeložíme do formálního jazyka, jak jsme jej viděli v kapitole 1a:

- $\forall x \in \mathbb{R}: (x > 0 \implies x(x + 1) > 0)$.

Každopádně jde o implikaci, nejoblíbenější matematickou strukturu.

Je také možné jít opačným směrem, k menší formálnosti. Můžeme třeba říct:

- Pro všechna kladná $x \in \mathbb{R}$ platí $x(x+1) > 0$.
- Pro všechna $x \in \mathbb{R}^+$ platí $x(x+1) > 0$.
- Pro každé $x > 0$ platí $x(x+1) > 0$.
- Všechna kladná x splňují $x(x+1) > 0$.

První z nich je stejně dobrá jako původní verze, jen zní méně „oficiálně“, což u méně důležitých tvrzení nemusí vadit. Druhá verze je také dobrá, dokonce je pěkně kompaktní, na druhou stranu může zkomplikovat život čtenáři, který není zvyklý na speciální značení \mathbb{R}^+ . Její použití tedy záleží na tom, jak často autor v knize tuto značku používá a jak hodný chce na čtenáře být.

Poslední dva výroky už jsou na hranici, mnozí matematici by je považovali za nepřípustně nepřesné, protože nespecifikují, z jakého oboru vlastně x bereme (že by to byla kladná racionální čísla?). Nicméně pokud například celou kapitolu pojednáváme o reálných číslech, pak se považuje za jasné, že bereme $x \in \mathbb{R}$, a autoři občas dají přednost čitelnosti před naprostou správností.

Již jsme mluvili o tom, že spolehlivost matematiky spočívá v důkazech. I nás Fakt je tedy třeba dokázat, ukážeme si na něm nejobvyklejší metody důkazu implikace. Nejprve se ale zamyslíme nad tím, co takový důkaz vlastně je.

1b.1 Důkazy

Jak se vlastně dokáže, že nějaký výrok sestavený pomocí logických operací je (vždy) pravdivý? Většina důkazů má stejnou myšlenku: Je třeba pomocí známých faktů nějak ukázat, že ze všech rádků příslušné pravdivostní tabulky mohou reálně nastat jen ty, které mají na konci jedničku. Na to existují různé metody, podle toho, jak je výrok sestaven, ale většina z nich skončí tím, že se musí dokázat nějaká implikace. Například ekvivalence se nejčastěji dokazuje tak, že se ukážou implikace $p \Rightarrow q$ a $q \Rightarrow p$. Proto se zde na implikaci zaměříme.

Již jsme si rozmysleli, že pro pravdivost určité konkrétní implikace $p \Rightarrow q$ je naprostě kritická situace, kdy je p splněno. Co se pak stane, rozhodne o její pravdivosti, protože v situacích, kdy p splněno není, prostě nejde dotyčnou implikaci zneplatnit, ať už se stane cokoliv. Z toho vychází nejčastější způsob, jak se implikace dokazují. Představíme si, že jsme v situaci, že je p splněno, a musíme nějak ukázat, že pak za každých okolností už nutně nastane i q . Případ, kdy p splněno není, tedy při důkazu nijak neřešíme.

Ponaučení: Přímý důkaz implikace $p \Rightarrow q$ se dělá takto: Předpokládáme, že je p splněno, a pomocí argumentů ukážeme, že pak nutně nastává i q .

Všimněte si, že tím neříkáme, že je to p opravdu splněno, jen si představujeme, k čemu by vedlo, kdyby splněno bylo. Ukažme si to na příkladu implikace „Jestliže mi useknou hlavu, tak umřu“. Na začátku důkazu budeme předpokládat, že mi usekli hlavu, a pak pomocí vědy lékařské dovodíme, že jsem mrtev. Provedli jsme teď důkaz implikace „dekapitace \Rightarrow kaput“, ale to neznamená, že jsem opravdu o kebuli přišel, jen jsme ukázali vzájemnou souvislost dvou určitých věcí.

Pravidlo, že při dokazování implikace zkoumáme jen situace, kdy je p splněno, má jednu zajímavou výjimku. Někdy (velice zřídka) potkáme situaci, že p splnit nikdy nejde. Pokud toto ukážeme, pak už celá implikace automaticky platí, viz pravdivostní tabulka. Takže implikace „Jestliže $13 > 23$, pak všichni studenti tohoto kursu vyletí“ je zaručeně pravdivá.

Tedě se podíváme na tři hlavní postupy, kterými se v matematice dokazuje.

1b.2 Přímý důkaz: Používá se k důkazu implikace a funguje přesně tak, jak to zní, prostě se vezme za dané, že platí její předpoklad, a dojde se nějakým zcela spolehlivým způsobem k platnosti jejího závěru. Ukážeme si to na důkazu implikace $x > 0 \Rightarrow x(x+1) > 0$.

Cestu od předpokladu k závěru si rozložíme na jednodušší kroky, které již budou jasné. Nejprve ukážeme důkaz superúplný, kde pečlivě zdokumentujeme všechny úvahy.

Celý dokazovaný výrok zní $\forall x \in \mathbb{R}: (x > 0 \Rightarrow x(x+1) > 0)$. Jde tedy o výrok s obecným kvantifikátorem, proto je nutno dokázat platnost oné implikace pro úplně všechna reálná čísla. Díky tomu rozhodně nebude fungovat to, co někdy zkouší začátečníci: vyberou si nějaké pěkné číslo a vyzkouší to pro něj.

Kdyby byla množina reálných čísel konečná, tak by je šlo probrat jedno po druhém. Existují důkazy, které lze redukovat na konečnou množinu případů, které se pak proberou, a pokud to pokaždé dopadne dle zadání, pak je důkaz hotov (viz poznámka po důkazu Faktu 2a.3). Nicméně nás případ to není.

My musíme ukázat platnost implikace pro všechna x , což se dělá standardně tak, že si prostě vezmeme nějaké reálné číslo x , ale nespecifikujeme jaké (ani to sami nevíme), prostě máme reálné číslo x , o kterém nevíme nic konkrétního, jen tu informaci, kterou dostaneme z dokazovaného tvrzení (popřípadě věci, které jsou platné pro všechna reálná čísla).

Mějme tedy reálné číslo x a ptáme se, zda platí implikace $x > 0 \implies x(x+1) > 0$. O její pravdivosti rozhodne, zda ve všech případech, kdy je splněn předpoklad $x > 0$, je také splněn závěr. K původnímu předpokladu $x \in \mathbb{R}$ tedy přidáme předpoklad další, že $x > 0$, a zkusíme se od nich postupnými kroky dobrat k cíli.

Jestliže $x > 0$, pak také $x + 1 > 0$. Čtenáři je to asi jasné, ale zkusme se pro úplnost zamyslet, jak by to šlo odvodit ze základních vlastností čísel. Když k rovnici $x > 0$ přičteme na obou stranách jedničku, dostaneme $x + 1 > 1$, máme také $1 > 0$, díky čemuž dostáváme řetězec nerovností $x + 1 > 1 > 0$. Pak také musí platit $x + 1 > 0$ (vlastně používáme tranzitivitu relace $>$, viz kapitola 3b).

Takže teď máme předpoklad $x > 0$, také jsme odvodili, že $x + 1 > 0$, a patří mezi základní vlastnosti, že vynásobením dvou kladných čísel získáme číslo kladné, tedy $x(x+1) > 0$.

(Je také možné argumentovat tím, že v nerovnost $x + 1 > 0$ vynásobíme obě strany kladným číslem x , čímž dostaneme tu žádanou.)

Každopádně je důkaz hotov, ukázali jsme, že jakmile je pro libovolné $x \in \mathbb{R}$ splněno $x > 0$, pak už není jiná možnost, než že $x(x+1) > 0$.

Důkaz je správný, pokud je v něm každý krok odůvodněn, někdy se odvoláváme na předpoklady z věty (bud' z preambule, nebo z předpokladu dokazované implikace), někdy na základní, již známé (a někde dokázané) vlastnosti, často také na tvrzení, která jsme dokázali dříve. Náš důkaz toto splňuje.

Takto se ale samozřejmě důkazy nepíší, ty jednodušší věci se vynechávají, protože se předpokládá, že si je čtenář domyslí. Stručnost důkazu tedy přímo závisí na tom, jak pokročilé čtenáře autor očekává. Zkusme si tu ukázat verzi důkazu vhodnou pro zcela začínajícího studenta (tedy pro tuto kapitolu):

Důkaz: Nechť x je libovolné reálné číslo. Jestliže $x > 0$, pak také $x + 1 > 0$, z těchto dvou nerovností již dostáváme $x \cdot (x+1) > 0$. □

Ten čtvereček je obvyklá značka udávající konec důkazu, aby čtenář věděl, že autor již nic dalšího nehodlá dodat. Čtenář by si v té chvíli měl rozmyslet, že to, co do té doby četl, je opravdu důkazem žádaného tvrzení. Používá se také celý černý čtvereček či zkratka Q.E.D. z latinského „quod erat demonstrandum“ neboli „což bylo dokázati“. Vlastenci dávají CBD.

Pro úplnost ještě ukážeme, jak by tento důkaz vypadal v knize pro pokročilejší studenty (kapitoly této knihy s vyšším číslem). Rovnou jich ukážeme několik.

Důkaz je zřejmý.

Důkaz je triviální.

Důkaz je snadný a přenecháme jej čtenáři jako cvičení.

1b.3 Nepřímý důkaz: I ten slouží k dokazování implikace, finta spočívá v tom, že se namísto té dané dokazuje její obměna (viz 1a), což je z logického pohledu postačující.

Vráťme se k našemu příkladu, teď musíme nejprve nahradit implikaci její obměnou:

$$\forall x \in \mathbb{R}: (x(x+1) \leq 0 \implies x \leq 0).$$

Tento výrok má zase tvar implikace a tu dokážeme přímým důkazem, tedy postupně se od předpokladu k závěru nové implikace propracujeme jednoduchými kroky.

Důkaz: Mějme libovolné $x \in \mathbb{R}$ a předpokládejme, že splňuje $x(x+1) \leq 0$. Má-li být součin dvou čísel záporný či nulový, vede to na dvě možnosti.

- a) Jedna možnost je, že $x \leq 0$ a $x+1 \geq 0$. Pak máme $x \leq 0$ a důkaz je hotov.
- b) Druhá možnost je, že $x \geq 0$ a $x+1 \leq 0$. Tyto dvě nerovnosti ale nemohou pro žádné číslo x platit zároveň, tento případ proto nikdy nenastane.

Z nerovnosti $x(x+1) \leq 0$ se tedy vždy dostáváme k případu a) a odtud k $x \leq 0$. □

Zde jsme si ukázali další užitečnou věc. Někdy se stane, že nás důkaz doveze k rozcestí, můžeme se vydat vícero směry podle toho, s jakými objekty pracujeme. Protože u obecných důkazů nikdy přesně nevíme, s čím pracujeme, je nutné projít všechny nabízející se cestičky a u všech dojít ke správnému cíli, popřípadě ukázat, že se do té či oné cestičky vůbec nedá vejít. Ukažme si ještě jednu verzi nepřímého důkazu.

Důkaz: Mějme libovolné $x \in \mathbb{R}$ a předpokládejme, že splňuje $x(x+1) \leq 0$. Důkaz dále rozdělíme na možnosti podle znaménka $x+1$:

- a) Jestliže $x+1 \leq 0$, pak $x \leq -1 < 0$ a tedy $x \leq 0$, přesně jak jsme potřebovali.
- b) Jestliže $x+1 > 0$, pak lze nerovnost $x(x+1) \leq 0$ vydělit kladným číslem $x+1$ bez změny směru nerovnosti a dostaneme zase $x \leq 0$.

Každopádně tedy $x \leq 0$. □

Jsou tvrzení, u kterých je nepřímý důkaz tou nejlepší volbou, ale tady to spíš neplatí, určitě bychom dali přednost přímému důkazu výše. Ukažme si příklad, kdy je nepřímý důkaz znatelně lepší.

Příklad 1b.a: Tvrzení: Jestliže je číslo $n > 2$ prvočíslo, pak je liché.

Přímý důkaz nevypadá moc nadějně, protože býti prvočíslem je docela komplikovaná vlastnost, není jasné, jak z ní něco vytěžit. Zkusme důkaz nepřímý, na to ale potřebujeme nejprve trochu logicky pracovat. Začneme tím, že si dotyčný výrok přepíšeme do správného logického tvaru. To se dá udělat více způsoby, nejpohodlnější je tento:

$$\forall n \in \{x \in \mathbb{N}; x > 2\}: (n \text{ je prvočíslo} \implies n \text{ je liché}).$$

Obměna pak zní

$$\forall n \in \{x \in \mathbb{N}; x > 2\}: (n \text{ je sudé} \implies n \text{ není prvočíslo}).$$

Teď toto tvrzení dokážeme přímým důkazem. Vezmeme si libovolné n z dané množiny, n je tedy přirozené číslo větší než 2. Pro něj chceme dokázat příslušnou implikaci, takže budeme navíc předpokládat, že je také sudé. To podle definice znamená, že $n = 2k$, kde k je nějaké celé číslo. Protože $n > 2$ neboli $2k > 2$, musí také platit $k > 1$. Odvodili jsme tedy, že $n = 2 \cdot k$ je možné rozložit jako součin dvou celých čísel větších než 1, což podle definice znamená, že n nemůže být prvočíslo.

Obměnu jsme dokázali, tudíž jsme dokázali i původní dané tvrzení.

△

1b.4 Důkaz sporem: Důkaz sporem je jeden z nejmocnějších nástrojů. Mějme libovolný výrok r (ne nutně implikaci). Důkaz sporem spočívá v tom, že dokážeme implikaci $\neg r \implies F$ (například přímo či nepřímo), řečeno slovy, ukážeme, že pokud by r neplatilo, tak nastane něco, co se nikdy nemůže stát, něco, co je ve sporu s naším (matematickým) světem. Podle selského rozumu to znamená, že neplatnost r nemůže nastat neboli r platí.

Formální logika to vidí podobně: Dokázali jsme platnost implikace $\neg r \implies F$. Její závěr je ale vždy nepravdivý, a jediný případ, kdy je implikace s nepravdivým závěrem pravdivá, je tehdy, když je také předpoklad nepravdivý. Tedy $\neg r$ neplatí čili r platí.

Jednou z výhod důkazu sporem je, že jej lze aplikovat i na tvrzení, které nejsou implikace. Oblíbenou situací je, když chceme dokázat, že něco neexistuje. To se přímo dokazuje špatně (dokažte, že kolem nás nelítají neviditelní a nenahmatatelní Marťané s anténkami). Důkaz sporem znamená, že začneme naopak: Předpokládáme, že to něco existuje, což je pozitivní informace, ze které se dá s trochou štěstí něco vytěžit, pokud možno nějaký kýžený nesmysl.

Jak důkaz sporem vypadá, když takto chceme dokázat implikaci $p \implies q$? Pak bychom měli dokázat implikaci $\neg[p \implies q] \implies F$ neboli $(p \wedge \neg q) \implies F$. To nám dává praktický návod: Předpokládáme, že platí předpoklad p a neplatí závěr q , a odvodíme z toho nějaký spor.

Jako příklad znova dokážeme (tentokrát sporem), že pro všechna reálná čísla platí $x > 0 \implies x(x+1) > 0$.

Důkaz: Mějme libovolné reálné číslo x a předpokládejme, že platí $x > 0$ a také $x(x+1) \leq 0$ (negace závěru). Nerovnost můžeme vydělit kladným číslem x na obou stranách a ona pořád zůstane platná, máme tedy $x+1 \leq 0$. Spojením nerovností $x+1 \leq 0 < x$ dostaneme $x+1 < x$ neboli $1 < 0$, což je spor. Důkaz je hotov.

□

Pro další ukázkou nepřímého důkazu a důkazu sporem viz důkaz Faktu 2a.3 a poznámky za ním. Tím jsme probrali hlavní metody důkazu.

S 1b.5 Jak vytvářet důkazy

Na vytváření důkazů žádný algoritmus či návod není, vždy je to otázka inspirace a hlavně zkušenosti a znalostí. Spíš než klasickému řešení příkladů se to podobá řešení hádanek či hlavolamů. Zmíníme zde několik zásad, které by mohly pomoci navést čtenáře na správnou cestu, když se dostane do problémů.

1. Vždy si nejprve dobře rozmyslete, s čím vlastně dokazované tvrzení pracuje. Někdy je to jasné, někdy to chce trochu přemýšlení. Jestliže máme dokazovat například prostotu zobrazení či rovnost obyčejných množin, pak je to vcelku jasné, pracujeme se zobrazeními či množinami. Co když ale máme dokázat například inkluzi $P(A) \subseteq P(B)$? Pak nemáme šanci uspět, dokud si nerozmyslíme, že $P(A)$ je množina, jejíž prvky jsou zase množiny, jmenovitě podmnožiny A . Takže když napišeme $x \in P(A)$, tak to x vlastně splňuje $x \subseteq A$.

Mluví-li dokazované tvrzení o řešení rekurentní rovnice, co to vlastně řešení je, jaký matematický objekt? I zde je malá šance, že se důkaz povede, pokud nám není jasné, že tímto objektem je posloupnost čísel, která po dosazení do rovnice dá pravdivý výraz.

Podobně když máme dokázat nějakou vlastnost dané relace, tak si musíme rozmyslet, jak s ní budeme pracovat. Je možné pracovat s pojmem platnosti či neplatnosti xRy , je ale také možné pracovat s relací jako s množinou R dvojic a používat množinové operace, jejichž význam je pak také třeba si rozmyslet. Každý z těchto přístupů má své slabiny i výhody, je třeba se pro jeden rozhodnout.

2. Další důležitá věc je si přesně vyjasnit, co se vlastně dokazuje. Když dokazujeme indukcí, je třeba si nejprve přesně říct, jak vypadá výrok $V(n)$ a pak se toho držet. Vyplatí se si takovéto věci napsat, mozek lépe pracuje s tím, co vidí očima. Vůbec je dobré si své úvahy někde bokem črtat. Často se stane, že člověk v myšlenkách dojde do určitého bodu a neví, jak dál. Když si ten bod napiše, občas najednou zjistí, že je to dál vlastně snadné. Pracovat s pojmy v hlavě je pro méně zkušeného velice obtížné a omezuje to možnosti.

Je dobré si v průběhu dokazování čas od času občerstvit, co se vlastně chce dokázat, protože někdy člověka myšlenky svedou jinam. Vyplatí se tedy opravdu si to hlavní napsat a občas se na to podívat.

3. Většinou se vyplácí přistupovat k dokazování strukturovaně a začínat od základů, nenechat se ohromit případnou komplikovaností zadání. Máme dokázat, že $P(A) \subseteq P(B)$? Ať už jsou ty množiny sebekomplikovanější, inkluze se dělá vždy stejně, přes prvky, takže si to napišme: Chceme ukázat, že $x \in P(A) \implies x \in P(B)$. Hned máme návod, na co se zaměřit.

Chceme ukázat, že nějaké T je prosté? Zase začneme od základů, napišeme si definici prostoty a vidíme, co je třeba udělat. Podobně když máme dokázat nějakou vlastnost relace atd atd. Ztrácíme se v indukci? Začneme od základů. Krok (1) chce dokázat pro libovolné $n \geq n_0$ výrok $V(n) \implies V(n+1)$. Co to vlastně je? Dosadíme za V do dotyčné implikace a hned máme návod.

4. Udělejte si pořádek v tom, jakou roli jednotlivé faktíky objevující se v problému hrají. Některé jsou dány již v zadání jako něco, co je možné použít. Některé se v průběhu důkazu takovým předpokladem stanou, buď z logiky důkazu (když třeba dokazujeme implikaci $p \implies q$, tak začneme předpokladem, že p platí, je to tedy další fakt, který je možné použít), nebo třeba proto, že jsme již to či ono úspěšně dokázali. Další faktíky jsou tu naopak od toho, abychom je dokázali.

Je opravdu důležité v tom mít jasno, opět často pomůže si to přehledně napsat. Pokud máte nutkání něco při dokazování použít, tak si ověřte, jestli je už to v kategorii „mám dáno, mohu použít“, častou chybou je totiž používat v důkazu to, co se vlastně má dokazovat. Takový důkaz je pak samozřejmě špatně. Naopak pokud se stane, že se zadrhnete, pak se vyplatí podívat na seznam věcí, které jsou k dispozici jako předpoklady. Použili už jsme všechno nebo je tam ještě něco, co jsme nevyužili? Tato jednoduchá věc často výrazně napoví. Pokud napišete důkaz a nepoužijete v něm všechny předpoklady, tak to je většinou znamení, že je někde chyba.

5. Poté, co důkaz dopíšete, se na něj trochu z odstupu podívejte, jakou má strukturu. Plyne správným směrem? Pokud po půl stránce výpočtu vítězoslavně podtrhnete $1 = 1$, tak je to skoro určitě špatně, protože toto jste dozajista dokazovat nechtěli. Tím se ovšem dostáváme k tématu další sekce, tak s radami skončíme.

1b.6 O jedné oblíbené chybě

Představme si studenta-začátečníka, který se snaží dokázat, že pro $n \in \mathbb{N}$ platí $\frac{n+1}{n} > 1$. S vysokou pravděpodobností student začne žádanou nerovnost upravovat, dokud se nedostane k něčemu pravdivému, například takto:

$$\begin{aligned}\frac{n+1}{n} &> 1 \\ n+1 &> n \\ 1 &> 0.\end{aligned}$$

Načež prohlásí „což platí“ a myslí si, že má hotovo. A nemá. Toto totiž ani náhodou nedokazuje, že $\frac{n+1}{n} > 1$. Proč? Kdyby to byl opravdu důkaz, pak by fungovalo i toto:

$$\begin{aligned}13 &= 23 \\ 13 - 18 &= 23 - 18 \\ -5 &= 5 \\ (-5)^2 &= 5^2 \\ 25 &= 25 \\ 0 &= 0.\end{aligned}$$

Je nicméně zjevné, že $13 = 23$ není pravda, tudíž tento postup nemůže být důkazem. Kde je chyba? Rozhodně ne v samotných krocích, ty jsou všechny korektní. Problém je v tom, že důkaz vede špatným směrem. Student totiž dokázal následující tvrzení:

$$\frac{n+1}{n} > 1 \implies 1 > 0,$$

čili jsme ukázali, že pokud platí něco, co vlastně chceme zkoumat, tak pak platí i $1 > 0$. Jenže my nechceme dokazovat, že $1 > 0$, to už víme. My naopak potřebujeme ukázat, že platí předpoklad, to ale z oné implikace nedostaneme, jak už jsme si rozmysleli v předchozí části.

Nám by pomohla opačná implikace, od známého k neznámému: $1 > 0 \implies \frac{n+1}{n} > 1$. Ten dostaneme, když předchozí postup obrátíme.

$$\begin{aligned} 1 &> 0 \\ n + 1 &> n \\ \frac{n+1}{n} &> 1. \end{aligned}$$

Všechny provedené úpravy jsou korektní, jde tedy o správný důkaz, dostal nás od známého k žádanému.

U druhého příkladu se právě toto nepovede. Dokážeme se dostat na půl cesty,

$$\begin{aligned} 0 &= 0 \\ 25 &= 25 \\ (-5)^2 &= 5^2, \end{aligned}$$

ale odebrat mocninu z rovnosti není možné, tam se pokus o obrácený chod pokazí. Můžeme zkousit obě strany odmocnit (to je korení úprava), dostaneme $\sqrt{(-5)^2} = \sqrt{5^2}$ neboli $|-5| = |5|$, což dává $5 = 5$ namísto toho, co bychom potřebovali.

Kupodivu se onen nesprávný postup „od konce“ často vídá. Proč tomu tak je? Pro méně zkušeného není snadné najít ten správný postup. Jak vlastně člověk přijde na to, že má začít zrovna nerovností $1 > 0$, aby se dostal k $\frac{n+1}{n} > 1$? Jak pak přijde na to, kterými úpravami se tam dostat? Onen „špatný postup“ na tyto otázky umí odpovědět, což je od něj pěkné. Jen si musíme být vědomi toho, že to důkaz není, je to prostě jen taková pomocná čmáranice, kterou jsme si udělali někde bokem. Pak ale musíme napsat „Důkaz:“ a znova to přepsat, tentokrát v opačném pořadí, a přitom si kontrolovat, že opravdu všechny kroky lze obrátit. Pokud to vždy vyjde, dostaneme korektní důkaz.

Přesto bych doporučil se pokud možno tomuto postupu vyhýbat, a to z několika důvodů. Za prvé, je zbytečně dlouhý, často jednu stranu (ne)rovnosti vůbec neupravujeme a jen ji opisujeme (viz příklad níže). Za druhé, tento postup nás nutí omezit se jen na ekvivalentní úpravy, díky čemuž nám jsou některé užitečné triky odepřeny - to se týká zejména důkazů nerovností, které jsou při tomto postupu pro začátečníka vyloženě zrádné, viz poznámka na konci. Jak tedy vypadá doporučovaný postup?

Začne se výrazem na jedné straně (ne)rovnosti a pomocí úprav se postupně řetězcem rovností či nerovností dojde k cílovému výrazu na pravé straně.

Jako příklad si dokážeme, že pro všechna $k \in \mathbb{N}$ platí $\frac{(k-1)^2+4k}{(k+1)^2} + \frac{k-1}{k^2-k} = \frac{k+1}{k}$. Obvykle bývá lepší začít tou složitou stranou, zkusíme ji co nejvíce zjednodušit.

$$\frac{(k-1)^2+4k}{(k+1)^2} + \frac{k-1}{k^2-k} = \frac{k^2-2k+1+4k}{k^2+2k+1} + \frac{k-1}{k(k-1)} = \frac{k^2+2k+1}{k^2+2k+1} + \frac{1}{k} = 1 + \frac{1}{k} = \frac{k+1}{k}.$$

Tím je důkaz hotov, výraz nalevo se opravdu rovná výrazu úplně napravo. Tento postup je výrazně kratší, než kdybychom použili metodu postupného upravování, protože tam bychom jen opisovali pravou stranu a pak to museli ještě jednou přepsat ve správném pořadí.

Zkušenější student většinou u jednodušších rovností a nerovností dokáže odhadnout, jak s tou jednou stranou cvičit, aby z toho vznikla strana druhá. Opravdu tento postup doporučujeme. U nerovností si ještě potřebujeme pohlídat, aby všechny kroky vedly na stejnou stranu: Například z řetězce $a < b = c \leq d < e = f$ dostáváme $a < f$, ale z $a < b = c \geq d$ neplyne o vztahu mezi a a d nic.

Někdy se samozřejmě může stát, že potřebné kroky nejsou vidět, to se u složitějších nerovností stane i ostřílenému matematikovi. Pak přijde vhod onen postup „od konce“, jen si je třeba pamatovat, že to není důkaz. Ten vznikne teprve tehdy, když se to přepíše ve správném směru neboli provede zpětný chod. Někdy se čas šetří tím, že se za ten „špatný“ postup napiše věta typu „Protože všechny provedené operace jsou ekvivalentní a postup lze obrátit, žádaná (ne)rovnost je dokázána.“ Je to ale nouzovka, nedoporučujeme to.

Tyto rady mají mnohem obecnější platnost než jen u důkazů (ne)rovností. Přestavme si, že se snažíme ukázat, že z výroku p plyne výrok z . Zkoušíme přímý důkaz a pomocí úvah se přes mezikroky dostaneme k výroku u , ale nevíme, jak dál: $p \rightarrow q \rightarrow \dots \rightarrow u$. Pak někdy stojí za pokus si vyjít vstříc z cíle a najít cestu $z \rightarrow y \rightarrow \dots \rightarrow u$. Tím došlo k propojení, ale důkaz to není, což je pěkně vidět na obrázku:

$$p \rightarrow q \rightarrow \dots \rightarrow u \leftarrow \dots \leftarrow y \leftarrow z.$$

Abychom dostali platný důkaz, je třeba ověřit, že všechny kroky při postupu od konce jsou ekvivalentní, takže lze

udělat i zpětný chod. Dostáváme pak řetězec úvah

$$p \rightarrow q \rightarrow \dots \rightarrow u \leftrightarrow \dots \leftrightarrow y \leftrightarrow z,$$

který již dává přímý důkaz $p \rightarrow \dots \rightarrow u \rightarrow \dots \rightarrow z$. Pro příklad se podívejte třeba na cvičení 5a.7 (ii).

Tím v zásadě končí tato kapitola. Pokud si chce čtenář udělat jasno v korektních a ekvivalentních úpravách a problémech s důkazy zpětným chodem, přidáváme další detaily v kapitole 14.

2. Teorie množin

Aby mohla matematika pomoci s popisem světa, musí mít struktury, které umožní zachytit různé aspekty toho, co zkoumá. V této kapitole si uvedeme ty úplně základní pojmy. Pojem množiny nám zhruba řečeno umožní zachytit situaci, kdy něco máme (popřípadě nemáme). K vystížení situace, kdy s našimi objekty něco provádíme (ubíráme, sesypáváme atd.) si zavedeme známé operace jako průnik, sjednocení a podobně. Po zevrubném prozkoumání světa množin si zavedeme další zásadní pojem, zobrazení.

Mnohé, možná většinu věcí z této kapitoly již student nejspíše někdy potkal. Využijeme proto této situace k prvnímu vážnějšímu ponoru do světa matematiky. Představíme čtenáři, jak se vytváří matematické teorie, procvičíme si důkazy a zkusíme také rozšířit rozsah naší představivosti. Díky tomu, že mnohá téma čtenář zná, bude se moci více věnovat vnímání matematického jazyka a postupů.

S Tato kapitola začíná základními věcmi, ale paradoxně právě u těch má čtenář někdy s důkazy problém, zejména pokud je tento způsob myšlení pro něj nový. Pro čtenáře, který se na to necítí, je proto naším doporučením číst tuto kapitolu spíš po povrchu, soustředit se na pochopení důležitých myšlenek a učení matematického jazyka, popřípadě si jen zlehka číst v lehkých důkazech označovaných jako rutinní, popř. poučné. Pro vniknutí do umění důkazu jsou nejlepší situace, kdy se pracuje s konkrétnějšími objekty, třeba důkaz prostoty nějakého zobrazení (viz kapitola 2b) či dokazování vlastností konkrétních relací (kapitola 3b). Až bude mít čtenář pocit, že už si s důkazy docela rozumí, může se k této kapitole zase vrátit a přečíst ji důkladněji.

2a. Množiny

Pro každého matematika představují množiny jeden ze základních vyjadřovacích prostředků. Teorie množin je ale zároveň samostatný obor matematiky, který studuje její základy, na kterých pak stojí ostatní matematické obory. Tuto hlubokou teorii zde dělat nebudem, zaměříme se na zkoumání množin na spotřební úrovni.

Základním termínem je množina, ale právě proto, že nám chybí ty hluboké základy, tak si nebudem schopni přesně specifikovat, co to vlastně je. Proto si namísto formální definice jen tak něco povíme.

Množina je neusporeádaný soubor objektů, které jsou přesně specifikovány. Tyto objekty se nazývají **prvky** dané množiny. Množina je těmito objekty jednoznačně dána, jinými slovy, pokud mají dvě množiny stejné prvky, pak je to tatáž množina.

By a **set** we mean an arbitrary collection of objects (called its **elements**).

Pokud jste to dobře pochopili (zejména to o shodnosti množin), tak už vás následující věc nepřekvapí:

Příklad 2a.a: Množina $\{b, a, a\}$ je stejná jako množina $\{b, a\}$, popřípadě množina $\{a, b\}$, protože mají stejné prvky. Jestliže se vás tedy někdo zeptá, kolik prvků má množina s pěti červenými kolečky, pak odpověď je jeden, leda že by každé to kolečko bylo nějak jiné.

△

Značení: Mějme množiny A, B . Značení $a \in A$ znamená, že objekt a je prvkem množiny A , naopak $a \notin A$ znamená, že objekt a není prvkem množiny A . Značení $A = B$ znamená, že jde o shodné množiny, naopak $A \neq B$ znamená, že množiny shodné nejsou, tedy nemají stejné prvky.

Množiny tradičně značíme velkými písmeny anglické abecedy, jejich prvky malými, pokud je to rozumně možné. Proč by to nemuselo být možné? Například $A = \{1, 2\}$ je množina, $B = \{13, 23\}$ je množina, ale lze z nich vytvořit další množinu: $M = \{\{1, 2\}, \{13, 23\}\}$. Množina M tedy má dva prvky, A a B , což jsou také množiny a už jsme je měli zapsané velkými písmeny. A M ještě může být strčeno do další množiny, nejde o nic výjimečného.

Množiny je možno zadat různými způsoby. Dva populární jsou výčtem prvků, třeba $M = \{1, 13, a, \diamond\}$, nebo značením zvaným anglicky „set builder“, kdy se nejprve odvoláme na nějakou větší známou množinu (universum) a pak uvedeme, které prvky z tohoto universa patří do naší množiny. Například množina všech sudých přirozených čísel se zapíše $M = \{x \in \mathbb{N}; x \text{ sudé}\}$.

Čímž se dostáváme k nejznámějším universům, což jsou

- přirozená čísla $\mathbb{N} = \{1, 2, 3, \dots\}$; (natural numbers)
- celá čísla $\mathbb{Z} = \{0, 1, -1, 2, -2, 3, \dots\}$; (integers)
- racionální čísla $\mathbb{Q} = \left\{ \frac{p}{q}; p \in \mathbb{Z} \wedge q \in \mathbb{N} \right\}$; (rational numbers)
- reálná čísla \mathbb{R} . (real numbers)

Používají se i různé modifikátory, malé plus či míinus omezují znaménko, tedy třeba $\mathbb{Z}^+ = \{n \in \mathbb{Z}; n > 0\} = \mathbb{N}$ či $\mathbb{Q}^+ = \{x \in \mathbb{Q}; x > 0\}$ nebo naopak $\mathbb{R}^- = \{x \in \mathbb{R}; x < 0\}$, malá nula přidá nulu, třeba $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$ nebo $\mathbb{R}_0^+ = \{x \in \mathbb{R}; x \geq 0\}$.

Nás budou zajímat hlavně první dvě množiny, protože diskrétní matematika v nich tráví většinu času. Naopak prakticky nebudeme pracovat s \mathbb{R} , o pár kapitol dál uvidíme, že těchto čísel je prostě příliš na to, aby je diskrétní matematika zvládla.

Velice důležitou množinou je prázdná množina čili množina bez prvků, $\emptyset = \{\}$.

Poznámka: Uděláme si malou exkurzi pro pokročilé a zvědavé. Intuitivní představa množin bývá taková, že si vymyslíme nějakou vlastnost p a ptáme se, pro které objekty je splněna. Pro usnadnění zápisu budeme psát $P(x)$, pokud objekt x tuto vlastnost splňuje. Když shromáždíme všechny objekty, které ji splňují, dostaneme množinu, zapsalo by se to $\{x; P(x)\}$.

Ted' překvapení: Tohle nefunguje, i když je to přirozená představa a první teorie množin (za kterou děkujeme Cantorovi, cca polovina 19. století) byla takto vystavěna. Bohužel se na přelomu 19. a 20. století ukázalo, že to vede k průsvihům. Asi nejjednodušší a také nejznámější je Russelův paradox z roku cca 1901.

Začneme tímto: Existují množiny, které jsou svými vlastními prvky. Rozhodně to zní šíleně a dá velkou námahu si nějakou takovou představit, ale jde to. Třeba takto: Uvažujme vlastnost mít nekonečně mnoho prvků (přesná definice nekonečnosti přijde, ale snad máme nějakou představu už ted'). Pokud by naše intuitivní představa o množinách byla správná, tak bychom pomocí této vlastnosti měli získat množinu M všech množin, které jsou nekonečné. Prázdná nebude, třeba $\mathbb{N} \in M$ nebo $\mathbb{R} \in M$. A ted' to přijde: Toto M samotné má nekonečně mnoho prvků, protože určitě vymyslíme nekonečně mnoho nekonečně velkých množin, stačí třeba vzít \mathbb{N} a postupně odebírat 1, 2, 3, ..., vznikají tak různé nekonečné množiny a všechny jsou v M . Proto podle definice $M \in M$, množina je svým vlastním prvkem. Takže stát se to může. Ted' jsme připraveni na to hlavní.

Definujme množinu A jako množinu všech množin, které nejsou svými vlastními prvky, naším zápisem tedy $A = \{M; M \notin M\}$. Ta určitě obsahuje spoustu objektů, v zásadě většinu množin, které si běžně představujeme, třeba množina $\{13, 14\}$ určitě není svým prvkem a tudíž leží v A . Co platí o množině A ? Kdyby byla svým vlastním prvkem, tedy $A \in A$, pak by nesplňovala podmínu z definice, proto by muselo platit $A \notin A$. Pak ale podmínu z definice splní, proto $A \in A$, pak ale ... Není tedy možné rozhodnout, zda A patří do A , což je pro teorii množin smrtící. Je to tzv. paradox, je z podobné lhůtě jako ten o pánovi, co prohlásí „Já ted' lžu“.

Bylo tedy nutno přepracovat teorii množin, jmenovitě změnit způsob, kterým se množiny tvoří, aby se tím zakázaly určité nepříjemnosti. To se povedlo a už nějakých sto let máme uznávanou teorii množin, která těmito problémy netrpí. Základní finta je v tom, že se zakáže tvořit množiny pomocí vlastností jen tak z ničeho, vždy se musí pomocí vlastnosti vybírat z již existující množiny. Při práci s množinami se tedy obvykle pohybujeme v rámci nějaké ohromné množiny U zvané universum, zvolené tak, aby nám v ní nic nechybělo. Z jejích prvků pak tvoříme nové množiny povědomým způsobem: Vymyslíme podmínu P , která se vztahuje k prvkům z U , a definujeme $M = \{x \in U; P(x)\}$. Dá se ukázat, že když tvoříme množiny takto, tak už paradoxy nelze vyrobít.

Jsou tam ještě další komplikace, ale zde to budeme úspěšně ignorovat. Ukazuje se totiž, že problémy vyvstávají, jen když se člověk v množinách hrabe trochu hlouběji, při běžné „spotřební“ práci se na paradoxy nenarazí. Spousta lidí si proto vystačí s tou intuitivní představou, kterou jsme tuto kapitolu začali, říká se tomu naivní teorie množin a my se s ní spokojíme také.

△

Je čas představit si první definici. Připomínáme, že tradičně se definice píší jako implikace, ale míni se ekvivalence (viz úvodní kapitola o logice).

!

Definice.

Nechť A, B jsou množiny. Řekneme, že A je **podmnožina** B , značeno $A \subseteq B$, jestliže jsou všechny prvky A také prvky B .

Řekneme, že A je **vlastní podmnožina** B , jestliže $A \subseteq B$, ale $A \neq B$.

Vztahu býti podmnožinou říkáme **inkluze**.

We say that a set A is a **subset** of a set B , denoted $A \subseteq B$, if all elements of A are also elements of B .

We say that A is a **proper subset** of B if $A \subseteq B$ but $A \neq B$.

Definice inkluze formálně: $A \subseteq B \iff [\forall a \in A: a \in B]$.

Na tento způsob zápisu byste si měli pomalu začít zvykat. Pokud si ještě nerozumíte s kvantifikátory, koukněte do první kapitoly.

Někteří autoři značí vlastnost býti vlastní podmnožinou jako $A \subset B$. Má to ale problém, protože jiní autoři používají z lenosti $A \subset B$ pro běžnou vlastnost inkluze (dokonce někdy i já, ale ne v této knize, na to jsem si dal pozor). Ve významu značení \subset je tedy zmatek, proto jej tady zavádět nebudeme a spokojíme se se značením $A \subseteq B$, kterému rozumí všichni stejně.

Když matematici zavedou nový pojem či vlastnost, tak hned začnou přemýšlet, jak fungují a jak se chovají. V jistém smyslu se dá říci, že toto je jednou z hlavních náplní matematiky: Dozvídat se co nejvíce o různých pojmech. Pojmy se totiž definují, protože se zdají užitečné, a když známe jejich vlastnosti, tak nám to pomůže při práci s nimi, tak jako vám například při práci s algebraickými výrazy pomáhá znalost různých identit a pravidel (třeba že se dá krátit ve zlomku). Praktickým výstupem takových znalostí pak jsou různé metody na řešení problémů.

Začneme něčím snadným.

Fakt 2a.1.

Nechť A je libovolná množina. Pak platí následující:

- (i) $A \subseteq A$;
- (ii) $\emptyset \subseteq A$.

Ted si ukážeme náš první důkaz, proto bude poněkud podrobnější.

Důkaz (rutinní, poučný): Normálně by se tento důkaz skládal ze slov „je to triviální“. Ukážeme, proč je to tak lehké.

(i): Pro každou množinu A máme dokázat tvrzení $A \subseteq A$.

Vezměme si tedy nějakou libovolnou množinu A . O této množině musíme dokázat, že $A \subseteq A$, což podle definice znamená $\forall a \in A: a \in A$. Jinými slovy, když si z ní vezmeme libovolný prvek a (viz ten kvantifikátor $\forall a$), tak je v A . To je ale triviálně pravda, všechno z A je v A , tudíž je důkaz hotov.

Zajímavý alternativní pohled na věc: To, co od A chceme, se dá přepsat do tvaru implikace:

Pro libovolný objekt x platí: $x \in A \implies x \in A$.

Přeloženo do slov, jestliže je x z A , tak je z A . Tato implikace je samozřejmě pravdivá. Obecně se dá dokázat (třeba pravdivostní tabulkou, viz kapitola 1), že implikace $p \implies p$ je vždy pravdivá.

Pro libovolnou množinu A jsme tedy dokázali (aniž bychom věděli, jak vypadá), že $A \subseteq A$.

(ii): Nechť A je libovolná množina. Ted máme dokázat, že $\emptyset \subseteq A$, podle definice tedy chceme $\forall x \in \emptyset: x \in A$, což lze ještě přesněji vyjádřit slovy $\forall x: x \in \emptyset \implies x \in A$. V kapitole 1a jsme viděli, že takové tvrzení je pravdivé vždy, důkaz je hotov.

□

S Poznámka: Nebyly to typické důkazy, první byl triviální a druhý netypický, protože vlastně jeho podstata nebyla v práci s množinami, ale ve fungování logiky. Snadno se stane, že student má něco dokázat, ale nevidí, co vlastně přesně je třeba udělat. Velice často pomůže nesnažit se rovnou psát důkaz, ale nejprve si vyjasnit, jaká je vlastně situace. V typickém případě se podíváme na to, co máme dokázat a co máme k dispozici, a snažíme se to vyjádřit jinak. Často používáme definice, jak jsme to udělali výše, někdy už třeba máme za sebou nějakou teorii, která nám dovolí zkoumané vlastnosti vyjádřit pomocí jiných, o kterých už něco víme.

Někdy stačí jen formální úprava k tomu, aby náš mozek náhle uviděl, jak věci udělat. Proto když se zadrhne, tak se může vyplatit, když si nějaký fakt zapíšeme jinak. Například to, že objekt a není v jisté množině A , lze zapsat $a \notin A$, $\neg(a \in A)$ či dokonce $a \in \overline{A}$. Může se stát, že jedno z těch vyjádření vyloženě zapadne do situace, zatímco ostatní by případný důkaz jen komplikovaly.

Následující výrok má velice příjemný důkaz, přirozený a snadný. Pokud chce student někde s důkazy začít, toto může být to správné místo.

△

Fakt 2a.2.

Nechť A, B, C jsou množiny. Jestliže $A \subseteq B$ a $B \subseteq C$, pak $A \subseteq C$.

Důkaz (rutinní, poučný): Výrok má platit pro všechny trojice množin, proto si vezmeme libovolné množiny A, B, C a chceme pro ně ukázat pravdivost implikace

$$(A \subseteq B \wedge B \subseteq C) \implies A \subseteq C.$$

Jako obvykle při důkazu implikace začneme tím, že považujeme její předpoklad za pravdivý, a musíme ukázat, že pak už bezpodmínečně dojdeme k závěru (viz kapitola 1b). V tomto případě tedy předpokládáme, že platí logická konjunkce $A \subseteq B \wedge B \subseteq C$, což znamená, že platí obě části, $A \subseteq B$ i $B \subseteq C$. Musíme ukázat, že pak také nutně platí $A \subseteq C$. Podle definice inkluze to znamená, že musíme ukázat, že pro libovolný prvek $a \in A$ platí i $a \in C$. Dejme se do toho.

Nechť $a \in A$ je libovolné. Podle našeho předpokladu, že $A \subseteq B$, pak také (podle definice inkluze) $a \in B$. Z toho podle předpokladu $B \subseteq C$ a definice inkluze zase dostaneme $a \in C$ a důkaz je hotov.

□

Poznámka o dokazování: I tento důkaz by se v „normální“ knize odbyl slovem „triviální“, my jsme si na něm zopakovali logickou spojku „a“ a připomeneme si základy dokazování. Nejprve jsme si analýzou rozebrali, co vlastně máme dělat: Ukázat $a \in A \implies a \in C$. To jsme provedli přímým důkazem ve dvou krocích

$$a \in A \implies a \in B \implies a \in C.$$

Každý z těchto částečných kroků byl pečlivě odvolán odvolávkou buď na nějaký již akceptovaný fakt (zde definici vlastnosti být podmnožinou) nebo na nějaký předpoklad, který jsme v té chvíli považovali za platný. U důkazu pokročilejších tvrzení se často také odvoláváme na již dokázaná tvrzení. V tom je podstata matematiky, pokaždé, když něco říkáme, tak to musíme mít podepřeno. V běžně psaných důkazech se ovšem detailní odvolávky vynechávají, protože se předpokládá, že si ty samozřejmě dokáže čtenář sám domyslet, komentují se jen kritické kroky. Až budete v dalších kapitolách číst stručnější důkazy, zkuste si rozmyslet, čím jsou podepřeny všechny ty „proto“, „tudíž“ a podobně, je to dobrý trénink. Až budete vy psát důkaz na písemce, tak je lepší ta odůvodnění napsat, jednak abyste si šplhli a ukázali, že víte, co děláte, druhak protože hlavně pro začátečníka je obtížné odhadnout, co se dá coby jasné vynechat.

△

Fakt 2a.3.

Nechť A, B jsou množiny. Pak $A = B$ právě tehdy, když $A \subseteq B$ a $B \subseteq A$.

Toto je zrovna jedna z věcí, které asi čtenáři přijdou naprostě jasné, a právě proto patrně neví přesně, jak tohle vlastně dokázat. Budeme následovat výše zmíněné rady a začneme od základů, nejprve si vyjasníme strukturu problému a pak si jednotlivá tvrzení přeložíme do řeči jednodušších pojmu.

Důkaz (rutinní): Vezměme dvě libovolné množiny A a B . Tvrzení, které o nich máme dokázat, je ekvivalence, což je totéž jako dvě implikace, tam i zpět. Máme tedy ukázat, že z faktu nalevo plyne fakt napravo a také naopak.

1) \implies : Předpokládejme, že $A = B$, což znamená, že tyto dvě množiny mají stejné prvky.

1a) Ukážeme, že pak $A \subseteq B$. Podle definice tedy máme ukázat, že $\forall a \in A: a \in B$. Nechť je $a \in A$ libovolné. Protože $A = B$, mají tyto množiny stejně prvky, tudíž $a \in A$ znamená také $a \in B$ a je to hotovo.

1b) Ukážeme, že pak i $B \subseteq A$. Vzhledem k symetrii situace půjde vlastně o stejný důkaz, jen se prohodí písmenka. Normálně bychom tedy v takové situaci napsali: „důkaz $B \subseteq A$ je obdobný.“ V rámci tréninku to zkusíme napsat: Nechť b je libovolný prvek z B . Protože A a B mají stejně prvky, pak také $b \in A$. Hotovo.

2) \impliedby : Předpokládejme, že $A \subseteq B$ a $B \subseteq A$. Potřebujeme dokázat, že pak $A = B$.

To je ale jasné. Všechny prvky z A jsou díky $A \subseteq B$ i v B a naopak všechny prvky z B jsou díky $B \subseteq A$ i v A . Množiny tedy mají shodné prvky. Důkaz je hotov.

□

Ta část 2) je asi nejtěžší, protože opravdu není jasné, co k tomu říct, když je to tak evidentní. Ukážeme ještě dva důkazy tohoto faktu v následující poznámce.

S 2a.4 Poznámka:

Vyzkoušíme si na implikaci

$$(A \subseteq B \wedge B \subseteq A) \implies A = B$$

nepřímý důkaz (viz kapitola 1b). Jinými slovy, budeme chtít dokázat její obměnu

$$\neg(A = B) \implies \neg(A \subseteq B \wedge B \subseteq A),$$

což se přepíše pomocí de Morganových zákonů (viz kapitola 1a) jako

$$A \neq B \implies [\neg(A \subseteq B) \vee \neg(B \subseteq A)]. \quad (*)$$

Tedž tuto implikaci dokážeme.

Důkaz: Předpokládejme tedy, že $A \neq B$. Rovnost množin je definována přes obecný kvantifikátor (všechny jejich prvky jsou sdíleny). Její negací je tedy tvrzení, že existuje prvek, který není sdílen (viz negace kvantifikátorů v kapitole 1a). Nás předpoklad $A \neq B$ tedy říká, že existuje nějaký prvek x , který je v jedné z těchto množin ale ne v druhé. Jsou dvě možnosti:

1) Jedna možnost je, že $x \in A$, ale $x \notin B$. To zapíšeme jako $\exists x \in A : x \notin B$, což podle právě probraných pravidel znamená

$$\exists x \in A : \neg(x \in B) \quad \parallel \quad \neg(\forall x \in A : x \in B).$$

Řečeno česky, není pravda, že všechny prvky z A jsou v B . To je negace vlastnosti $A \subseteq B$, čili A nemůže být podmnožinou B . Protože platí $\neg(A \subseteq B)$, platí i disjunkce $\neg(A \subseteq B) \vee \neg(B \subseteq A)$ (pro její pravdivost stačí, aby byla splněna některá ze složek). Pokud tedy nastane situace $x \in A$ ale $x \notin B$, pak je kýžená implikace (*) dokázána.

2) Druhá možnost je, že $x \in B$, ale $x \notin A$. Stejným argumentem jako v 1) pak ukážeme, že neplatí $B \subseteq A$ a tudíž i v tomto případě je ona implikace (*) pravdivá.

Žádný jiný případ už není možný, takže dokazovaná implikace (obměna) platí. □

Všimněte si, že se nám důkaz rozvětvil na dvě možnosti. To se stává, je pak ale důležité v takové situaci probrat úplně všechny možnosti a pokaždé dojít ke správnému závěru, popřípadě ukázat, že ta či ona možnost v dané situaci vlastně nemůže nastat. Anglicky se tomuto říká „exhaustion argument“ neboli „důkaz vyčerpáním“, myslí se tím všech možností, ale často také čtenáře a nezřídka i autora.

Výraznou úsporou může být, pokud jsou některé situace obdobné a jejich případy by se řešily stejně, zejména užitečná je symetrie, například v našem důkazu se dají A a B zaměnit. V běžných důkazech se to využije tak, že si prostě jednu z možností vybereme, musí se to ale správně odůvodnit. Hned si to ukážeme.

Třetí rozšířený typ důkazu implikace je důkaz sporem (viz kapitola 1b), připomeneme si jej opět na naší oblíbené implikaci ($A \subseteq B \wedge B \subseteq A \implies A = B$).

Důkaz: Předpokládejme tedy, že platí její předpoklady $A \subseteq B$ a $B \subseteq A$, ale neplatí závěr $A = B$. To znamená, že existuje nějaký bod x takový, že je v jedné množině a není v druhé. Protože je situace symetrická, můžeme předpokládat, že $x \in A$ a $x \notin B$. Ovšem z předpokladů $x \in A$ a $A \subseteq B$ také plyne, že $x \in B$. Prvek x tedy zároveň splňuje $x \notin B$ a $x \in B$, což je ve sporu. Důkaz implikace je hotov. □

Tím končíme s důkazy, které byly sice většinou lehké, ale ukazovali jsme si na nich podrobně různé triky. Další důkazy budeme postupně dělat stručnější, ke konci této kapitoly už budou v zásadě psány standardním způsobem.

△

Definice.

Nechť A je množina. Definujeme **potenční množinu** A , značeno $P(A)$, jako množinu všech podmnožin A .

Příklad 2a.b: Jestliže $A = \{a, b\}$, pak $P(A) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$.

△

Jako rozcvíčku si ukážeme jednu vlastnost, která je z matematického hlediska triviální, ale důkaz může být pro začátečníka poněkud drsný.

Fakt 2a.5.

Nechť A, B jsou množiny. Jestliže $A \subseteq B$, pak $P(A) \subseteq P(B)$.

Důkaz (rutinní): Předpokládejme, že máme libovolné množiny A, B splňující $A \subseteq B$. Potřebujeme ukázat, že $P(A) \subseteq P(B)$, což podle definice znamená, že $\forall m \in P(A): m \in P(B)$.

Zde je zásadní si rozmyslet, s jakými objekty vlastně pracujeme. Co jsou to ty m výše? $P(A)$ je množina všech podmnožin A , takže $m \in P(A)$ je vlastně nějaká podmnožina A . S tímto objektem tedy někdy pracujeme jako s prvkem (když mluvíme o $P(A)$ a $P(B)$) a jindy jako s množinou (když se budeme pohybovat v A, B). Pro začátečníka to může být zmátečné, ale důkaz je vlastně snadný, když si v tom člověk udělá v hlavě trochu pořádek, právě tak, jak jsme si to teď rozmysleli. Je čas na důkaz.

Vezměme tedy libovolný prvek m z $P(A)$. Podle definice $P(A)$ je m podmnožinou A , ale máme také předpoklad $A \subseteq B$, tudíž podle Faktu 2a.3 je $m \subseteq B$. Proto podle definice $P(B)$ platí $m \in P(B)$ a důkaz je hotov. □

! Obvykle pracujeme s více množinami a všechny jsou schovány uprostřed jedné velké množiny, universa U , ze kterého při své práci nevyskočíme. V rámci tohoto universa pak množiny všelijak kombinujeme či vytváříme nové. Asi každý čtenář se již potkal se sjednocením množin (sesypeme všechny jejich prvky do jednoho pytlíčku), průnikem (to, co je množinám společné) a doplňkem (všechny prvky mimo). Teď si ukážeme formální definice, čtenář už by je měl být schopen plynule číst a překládat si je do srozumitelné představy.

Definice.

Nechť A je množina v nějakém universu U . Definujeme její **doplňek** vzhledem k U jako

$$A^c = \overline{A} = \{x \in U; x \notin A\}.$$

Let A be a set in a universe U . We define its **complement** (with respect to U) as the set $A^c = \overline{A}$ of all elements of U that are not in A .

!

Definice.

Nechť A, B jsou množiny v nějakém universu U . Definujeme jejich

sjednocení: $A \cup B = \{x \in U; x \in A \vee x \in B\}$;

průnik: $A \cap B = \{x \in U; x \in A \wedge x \in B\}$;

rozdíl či doplněk B v A : $A - B = \{x \in U; x \in A \wedge x \notin B\}$;

kartézský součin: $A \times B = \{(a, b); a \in A \wedge b \in B\}$, zde (a, b) značí uspořádanou dvojici.

Anglickou verzi uděláme méně formální, ať si čtenář zvyká na jazyk.

Let A, B be sets in some universe U . We define their

union $A \cup B$ as the set of all elements that are in A or in B ;

intersection $A \cap B$ as the set of all elements that are both in A and B ;

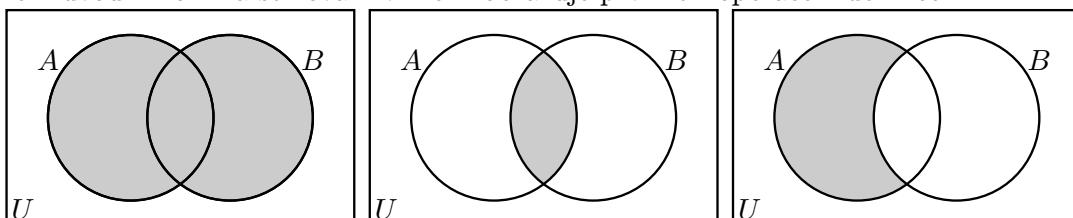
difference $A - B$ as the set of all elements that are in A but not in B ;

Cartesian product as the set of all ordered pairs (a, b) such that $a \in A$ and $b \in B$.

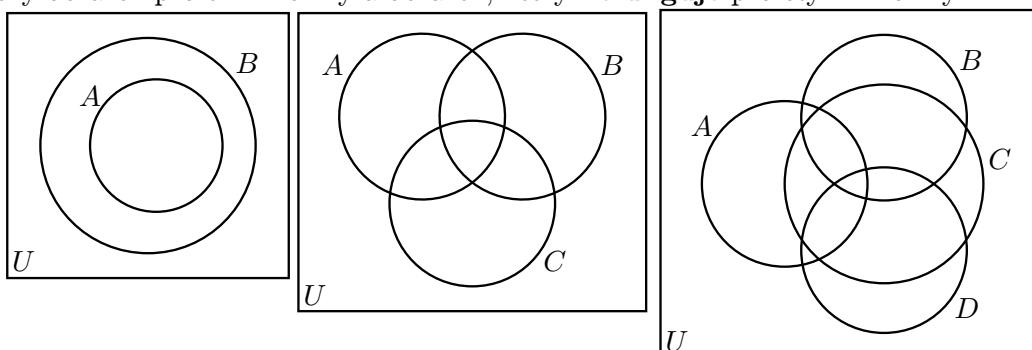
Příklady asi moc nemají smysl, všichni to znají, ale budíž: Když třeba $A = \{1, 2, 13\}$ a $B = \{13, 23\}$, pak $A \cup B = \{1, 2, 13, 23\}$, $A \cap B = \{13\}$, $A - B = \{1, 2\}$ a také $A \times B = \{(1, 13), (1, 23), (2, 13), (2, 23), (13, 13), (13, 23)\}$.

Co je doplněk A ? To není jasné, protože jsme neřekli, v jakém universu pracujeme. Nabízí se třeba universum \mathbb{N} , pak je $\bar{A} = \{3, 4, \dots, 11, 12, 14, 15, 16, \dots\}$. Jenže můžeme vzít jiné U a pak bude \bar{A} jiné. V zásadě se dá říct, že pokud nějaká situace vyžaduje, aby se dělal doplněk, tak už bývá z kontextu jasné i U , a pokud doplnky nepotřebujeme, tak nám v zásadě U nijak nechybí. Spousta lidí pracuje s množinami celá léta a ani neví, že jsou nějaká universa, i my jsme teď v pohodě vytvořili třeba $A \cup B$, aniž bychom znali U .

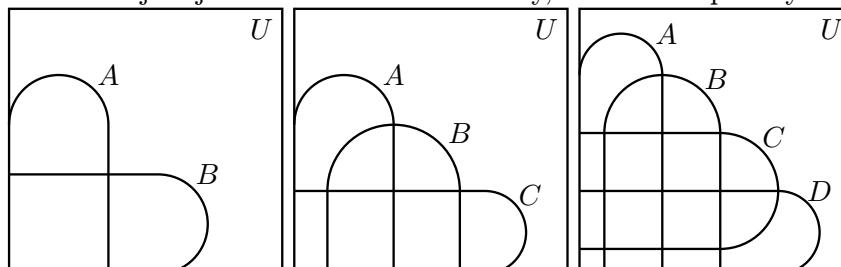
! Dobrým znázorněním vztahu mezi množinami jsou tzv. **Vennovy diagramy**. Následující obrázky ukazují standardní znázornění dvou množin a stínování v nich zobrazuje první tři operace z definice.



Někdy chceme obrázkem vyjádřit přímo určitou situaci. Následující obrázek ukazuje situaci, kdy $A \subseteq B$. Připojili jsme také klasický obrázek pro tři množiny a obrázek, který **nefunguje** pro čtyři množiny.



Proč nefunguje? Protože na něm není místo pro prvky, které jsou v A, B a D , ale nejsou v C , tedy chybí tam místo na vyznačení množiny $(A \cap B \cap D) - C$. Dá se dokázat, že nelze vytvořit obrázek ze čtyř kružnic, který by vyhovoval (jinými slovy, ať už nakreslíte čtyři kružnice jakkoliv, vždycky bude existovat určitý typ prvků, pro které ten obrázek nebude mít chlívceček). Co s tím? Jedna možnost je namísto jedné z kružnic použít klobásoid, což ale není moc estetické. Existují zajímavé alternativní obrázky, které už to pro čtyři množiny dokážou:



Tyhle obrázky to zase neumí pro pět množin, ale ještě mi to nikdy nechybělo.

! Uvedeme si teď vlastnosti operací. Rozhodně nemá smysl učit se pravidla z následujících tvrzení nazpaměť (až na pár výjimek), ani profesionální matematik by je nedokázal všechny vyjmenovat. Důležité je se nad nimi zamyslet, představit si různé situace a rozmyslet si, že by ta pravidla měla platit. Cílem je začít množinám rozumět, aby vám platnost těch pravidel přišla stejně přirozená jako platnost $7 + 5 = 5 + 7$. Když je pak člověk v situaci, kdy by nějaké to pravidlo potřeboval, tak se mu samo nabídne, jako se člověku nabízí třeba krácení ve zlomcích, aniž by o tom znal nějakou větu. Pro matematika je většina následujících tvrzení „jasná“, v jeho světě to tak prostě fungovat musí, stejně jako v našem světě když pustíme kámen, tak všichni víme, co se pak stane, nemusíme si na to pamatovat nějaké věty.

Pro získání této intuice je důležité si také rozmýšlet věci, které neplatí, aby člověk nepropadl přílišnému optimismu, jako třeba čtenář ví, že nelze napsat $\frac{1}{2+3}$ jako $\frac{1}{2} + \frac{1}{3}$. Podobně mnoho věci selhává pro množinové operace a o nejsvůdnějších by měl člověk vědět, asi nejzrádnější uvidíte za chvíli a ve cvičení 2a.2.

Přemýšlení nad pravidly je také dobrá příležitost si protrénovat logiku a důkazy.

Hned z definice operací dostaneme následující.

Fakt 2a.6.

Nechť A, B jsou množiny. Pak platí:

- (i) $A \subseteq A \cup B, B \subseteq A \cup B;$
- (ii) $A \cap B \subseteq A, A \cap B \subseteq B.$

Důkaz (rutinní): (i): Dokážeme, že $A \subseteq A \cup B$. Nechť $x \in A$ je libovolné. Protože obecně je implikace $p \implies p \vee q$ pravdivá, tak z pravdivosti výroku $x \in A$ vyplývá i pravdivost výroku $x \in A \vee x \in B$ a tedy $x \in A \cup B$. Důkaz hotov.

Důkaz $B \subseteq A \cup B$ je stejný dle symetrie.

(ii): $A \cap B \subseteq A$: Nechť $x \in A \cap B$. Pak $x \in A \wedge x \in B$, proto tedy $x \in A$. Důkaz je hotov, druhé tvrzení plyne ze symetrie. □

Všimněte si, že při důkazu (ii) jsme napsali jen „nechť $x \in A \cap B$ “. Pokládá se za samozrejmé, že se v takové situaci bere x libovolné, tudíž se šetří místem a časem a to slovo se vynechává, i zde to budeme dělat. Pokud student předvádí důkaz u zkoušky, tak ať raději to „libovolné“ napíše, ať ukáže zkoušejícímu, že ví, co se děje.

Mimochedem, mohlo by se stát, že namísto inkluze budou v těch vztazích rovnosti? A pokud ano, tak za jakých okolností? Matematici si pořád kladou takové zvědavé otázky, odpovědi na tyto dvě najdete ve cvičení 2a.1 (iii) a (iv).

! Věta 2a.7. (zákon pro počítání s množinami)

Nechť A, B, C jsou libovolné množiny z universa U . Pak platí následující:

- (i) $A \cup \emptyset = A, A \cap U = A$; (zákon identity)
- (ii) $A \cap \emptyset = \emptyset, A \cup U = U$; (zákon dominance)
- (iii) $A \cup A = A, A \cap A = A$; (idempotence)
- (iv) $\overline{\overline{A}} = A$; (zákon komplementu)
- (v) $A \cup B = B \cup A, A \cap B = B \cap A$; (komutativní zákon)
- (vi) $A \cup (B \cup C) = (A \cup B) \cup C, A \cap (B \cap C) = (A \cap B) \cap C$; (asociativní zákon)
- (vii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C), A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$; (distributivní zákon)
- (viii) $\overline{A \cup B} = \overline{A} \cap \overline{B}, \overline{A \cap B} = \overline{A} \cup \overline{B}$; (De Morganovy zákony)
- (ix) $A \cup (A \cap B) = A, A \cap (A \cup B) = A$; (zákon absorbce)
- (x) $A \cup \overline{A} = U, A \cap \overline{A} = \emptyset$. (zákon doplňku)

Doporučujeme, aby si čtenář postupně všechna pravidla prošel a pokaždé si začal kreslit Vennovy diagramy dané situace. Je velice užitečné zkousit si každou vlastnost vyvrátit, tedy nakreslit situaci, kdy neplatí. Samozrejmě se to nemůže podařit, ale naší intuici velice pomáhá, když se snažíme takovou protipříkladovou situaci vytvořit a ona se nám vždycky nějakým způsobem zvrtne, takže nakonec studovaná vlastnost platí.

Pokud si čtenář vlastnosti prošel, tak jistě zjistil, že hlavně těch prvních pět je opravdu jasných, ale pak jsou situace, které vyžadují hlubší zamýšlení a stojí za to si je pamatovat. Jde zejména o De Morganovy zákony a distributivní pravidlo („roznásobení závorky“), které dokonce funguje pro obě pozice operací (na rozdíl od násobení a sčítání, které si takto rozumí jen jedním způsobem).

Dokážeme to nejdůležitější, zbytek necháme na čtenáři, protože je to opravdu snadné.

Důkaz (rutinní, poučný): (vii): Dokážeme první vztah, druhý je obdobný. Rovnost množin se nejčastěji dokazuje přes dvě inkluze, ty se pak dokazují podle definice, tedy implikace pro náležení prvků.

1) $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$: Nechť $x \in A \cap (B \cup C)$. Pak $x \in A$ a $x \in B \cup C$. Druhý fakt nabízí dvě možnosti.

Jestliže $x \in B$, pak spolu s $a \in A$ dostaneme $x \in A \cap B$, proto $x \in (A \cap B) \cup (A \cap C)$.

Jestliže $x \in C$, pak symetricky dostaneme $x \in A \cap C$, proto $x \in (A \cap B) \cup (A \cap C)$.

Pokryli jsme všechny (obě) možnosti, důkaz je úplný.

V části 2) tento rozbor možností, které jsou v podstatě stejné, nahradíme odvolávkou na symetrii.

2) $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$: Nechť $x \in (A \cap B) \cup (A \cap C)$. Pak x leží alespoň v jednom z těch průniků, díky symetrii můžeme předpokládat, že $x \in A \cap B$. Pak $x \in A$ a také $x \in B$. To druhé ale dává $x \in B \cup C$, tedy $x \in A \cap (B \cup C)$ a důkaz je hotov.

(viii): Nechť A, B, C jsou množiny.

1) Dokážeme, že $\overline{A \cup B} = \overline{A} \cap \overline{B}$, zase přes dvě inkluze.

1a) Nechť x je libovolný prvek z $\overline{A \cup B}$. To znamená, že $x \notin A \cup B$. Prvky $x \in A \cup B$ splňují $x \in A \vee x \in B$, prvky mimo tedy splňují negaci této vlastnosti, což je podle de Morganových zákonů pro formální logiku rovno

$$\neg(x \in A \vee x \in B) \Leftrightarrow \neg(x \in A) \wedge \neg(x \in B) \Leftrightarrow x \notin A \wedge x \notin B.$$

To znamená, že $x \in \overline{A} \wedge x \in \overline{B}$, tedy $x \in \overline{A} \cap \overline{B}$. Právě jsme dokázali, že $\overline{A \cup B} \subseteq \overline{A} \cap \overline{B}$.

1b) Nechť naopak $x \in \overline{A} \cap \overline{B}$. Použijeme podobný postup jako v 1a), ale zapíšeme jej čistě formálně jako řetěz implikací.

$$\begin{aligned} x \in \overline{A} \cap \overline{B} &\implies x \in \overline{A} \wedge x \in \overline{B} \implies x \notin A \wedge x \notin B \implies \neg(x \in A) \wedge \neg(x \in B) \\ &\implies \neg(x \in A \vee x \in B) \implies \neg(x \in A \cup B) \implies x \in \overline{A \cup B}. \end{aligned}$$

Všimněte si, že všechny implikace platí i zpětně, tedy jsou to vlastně ekvivalence. To znamená, že části 1a) a 1b) šlo dokázat najednou. U snažších věcí lze někdy ekvivalenci dokázat přímo.

2) Teď dokážeme, že $\overline{A \cap B} = \overline{A} \cup \overline{B}$, tentokrát zkusíme jinou metodu. Jednak uděláme oba směry najednou, druhak zvolíme jinou formu zápisu. V předchozí části jsme pracovali s prvky, teď budeme pracovat s celými množinami a budeme upravovat podmínky, které je definují. Je asi zřejmé, že když v definici množiny podmínu příslušnosti nahradíme jinou, která je ekvivalentní (říká totéž), tak se dotyčná množina nezmění.

$$\begin{aligned} \overline{A \cap B} &= \{x \in U; x \notin A \cap B\} = \{x \in U; \neg(x \in A \cap B)\} = \{x \in U; \neg(x \in A \wedge x \in B)\} \\ &= \{x \in U; \neg(x \in A) \vee \neg(x \in B)\} = \{x \in U; x \notin A \vee x \notin B\} = \{x \in U; x \in \overline{A} \vee x \in \overline{B}\} \\ &= \overline{A} \cup \overline{B}. \end{aligned}$$

□

Všechna právě probraná pravidla z Věty mají své prakticky stejně vypadající bratříčky ve formální logice, stačí namísto \cap psát \wedge , místo \cup se píše \vee , doplněk se nahradí negací, \emptyset je F a podobně. Mezi množinovými a logickými operacemi je úzká souvislost, v důkazu výše to bylo také vidět, například de Morganovo pravidlo pro množiny jsme dokazovali pomocí de Morganova pravidla pro logické výrazy.

Poznámka: Někdy se v důkazu situace výrazně zdědění, pokud o prvku, se kterým pracujeme, víme něco navíc. Toho se dá dosáhnout například tím, že se hned na začátku prvky rozdělí do skupin podle nějakého kritéria a pak se důkaz dělá pro každou skupinu zvlášť. Někdy si toto rozdělení důkaz sám vynutí, i v důkazu výše je jedna rozdvojka.

Nejčastější dělení je podle toho, zda prvek leží či neleží v množinách z předpokladu, tedy v A, B, C, \dots . Vznikají tak skupiny, jejichž počet rychle stoupá, pro dvě množiny jsou čtyři možnosti, pro tři množiny osm, pro čtyři 16 atd., v důkazu pak musíme probrat všechny skupiny, takže tento typ důkazu není zrovna nejpoužívanější. Často ale tuto metodu používáme v situacích, kdy jen chceme zjistit, zda nějaký množinový vztah platí či ne, pak se podíváme, co se děje pro typické zástupce různých skupin.

Jako příklad dokážeme rozbořem pro typy prvků rovnost $\overline{A \cap B} = \overline{A} \cup \overline{B}$. Jsou čtyři možnosti, jaký vztah může nějaký prvek mít k množinám A, B .

a) Jestliže $x \in A$ a $x \in B$, pak i $x \in A \cap B$ a tudíž $x \notin \overline{A \cap B}$.

Pak ale také $x \notin \overline{A}$ a $x \notin \overline{B}$, tudíž $x \notin \overline{A} \cup \overline{B}$. Tyto prvky x tedy nejsou ani v množině nalevo, ani v množině napravo zkoumané rovnosti.

b) Jestliže je x takové, že $x \in A$ ale $x \notin B$, pak $x \notin A \cap B$, tudíž $x \in \overline{A \cap B}$.

Podle $x \in \overline{B}$ máme i $x \in \overline{A} \cup \overline{B}$. Tyto prvky x tedy jsou i v množině nalevo, i v množině napravo.

c) Ukažte sami, že prvky x splňující $x \notin A$ ale $x \in B$ jsou také v obou množinách, ukažte to i pro případ d), tj. prvky x splňující $x \notin A$ a $x \notin B$.

Proto jsou všechny typy prvků buď v obou zkoumaných množinách, nebo nejsou v žádné, ony množiny tedy mají stejné prvky a jsou si rovny.

Zdlouhavost takovýchto úvah lze zkrátit tabulkou. Ve sloupcích značíme množiny a v řádcích značíme pomocí 0 a 1, zda zkoumaný prvek v nich je nebo není. V záhlaví jsou dva sloupce, kterými si prvky vybíráme, tam musíme dostat všechny možné kombinace, takže tabulka pro dvě množiny bude mít 4 řádky. Tabulka našeho důkazu vypadá takto:

A	B	$A \cap B$	$\overline{A \cap B}$	\overline{A}	\overline{B}	$\overline{A} \cup \overline{B}$
1	1	1	0	0	0	0
1	0	0	1	0	1	1
0	1	0	1	1	0	1
0	0	0	1	1	1	1

Protože se sloupce zkoumaných množin shodují, jsou si tyto množiny rovny.

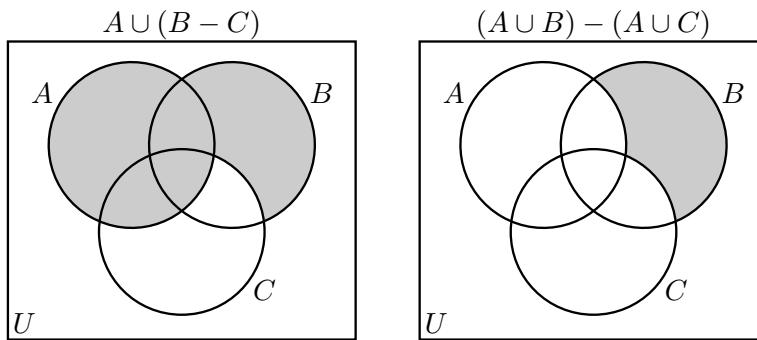
△

! **Poznámka:** Zatím jsme jen dokazovali, že něco platí. Může se ale stát, že nám někdo předhodí tvrzení, které není dobré, třeba toto:

Pro libovolné množiny A, B, C platí $A \cup (B - C) = (A \cup B) - (A \cup C)$.

(Je to pokus o distributivní zákon, zkusili jsme „roznásobit“ tu závorku.)

Jak zjistíme, zda má cenu to dokazovat? Jedna možnost je zakreslit si obě množiny ve Vennově diagramu.



Vidíme, že nejde o stejné objekty. Jak se tedy dokáže, že je dané tvrzení nepravdivé? Toto tvrzení je uvedeno obecným kvantifikátorem „pro všechny množiny“. K vyvrácení tedy stačí najít jeden protipříklad, kdy dané tvrzení selže (viz kapitola 1b). Obrázek nás inspiruje, stačí zvolit množiny tak, aby na nich byl vidět ten rozdíl v obrázku, neboli chceme mít prvek v místě, kde se obrázky liší. Zvolme tedy třeba $A = \{13\}$ a $B = C = \emptyset$, pak

$$A \cup (B - C) = \{13\} \cup \emptyset = \{13\}, \text{ zatímco } (A \cup B) - (A \cup C) = \{13\} - \{13\} = \emptyset.$$

Tento protipříklad tedy dokázal, že dané tvrzení neplatí.

Mimochodem, obrázek naznačuje, že by první množina měla vždy obsahovat tu druhou. A to je pravda, viz cvičení 2a.1 (ix).

Další možnost, jak najít protipříklad, je pomocí tabulky z poznámky výše. Pokud se sloupce zkoumaných množin neshodují, tak se podíváme na řádek, kde se liší, a vytvoříme takové množiny $A, B, C \dots$, aby tuto situaci měly neprázdnou.

△

Když mají matematici operace pro dva objekty, tak se většinou nezastaví a chtejí je pro více objektů. Ukážeme si standardní cestu, kterou k tomu dospívají. Začneme třemi množinami: Jak bychom vymysleli $A \cap B \cap C$? Protože dvě množiny pronikat umíme, nabízí se dělat tři postupně. Nejprve pronikneme $A \cap B$ a ten výsledek pak s C , formálně zapsáno to je $(A \cap B) \cap C$. To je zajímavý nápad, ale má zádrhel: Proč zrovna takto, proč nezačít třeba $B \cap C$, celkem pak $A \cap (B \cap C)$? V takové chvíli člověka zachrání hlavně asociativní zákon (což je moment, který se vyskytne opakován i v dalších kapitolách). Ten říká, že je jedno, které závorkování použijeme, takže nápad, který jsme měli, funguje docela dobře.

Jakmile umíme proniknout tři množiny, není důvod se zastavit a nepřidat množinu čtvrtou, můžeme třeba definovat $A \cap B \cap C \cap D$ jako $(A \cap B \cap C) \cap D$ a díky asociativitě zase víme, že to vyjde náležitě jako třeba $(A \cap B) \cap (C \cap D)$, což je také zajímavá možnost, protože používá jen průniky dvou množin.

Podobně pak uděláme průnik pěti, šesti, 50 atd. množin. Jak to pak ale zapsat pořádně? Nejjednodušší způsob je rekurzí či indukcí, což v zásadě znamená, že se na tu definici díváme od konce (viz ten příklad se čtyřmi

množinami):

$$\bigcap_{i=1}^{n+1} A_i = \left(\bigcap_{i=1}^n A_i \right) \cap A_{n+1}.$$

Toto je typ definice, který se používá často a zde na to máme speciální kapitolu 5a o indukci a rekurzi (berete to jako první vlaštovku či reklamu). Například chceme-li průnik pěti množin A_1 až A_5 , tak vzorec s $n = 4$ říká, že nejprve musíme umět proniknut 4 množiny,

$$A_1 \cap \dots \cap A_5 = (A_1 \cap \dots \cap A_4) \cap A_5.$$

Na to podle stejného vzorce, ale s $n = 3$, zase potřebujeme umět proniknout 3 množiny $A_1 \cap A_2 \cap A_3$, odtud už se další iterací konečně dopracujeme k průniku dvou množin $A_1 \cap A_2$, který umíme, tak ho uděláme. Načež následuje „zpětný chod“ naším rozkladem: Výsledek $A_1 \cap A_2$ pronikneme s A_3 (průnik dvou množin umíme), tento výsledek s A_4 , ten pak s A_5 .

Tento způsob je klasický, spolehlivě zobecňuje asociativní operace na více objektů. Často se povede, že operace, která tak vnikne, má dokonce nějaký rozumný význam. Když si například člověk rozmyslí, které prvky jsou v množině $(A \cap B) \cap C$, tak zjistí, že to jsou přesně ty, které jsou zároveň ve všech třech množinách, podobně se to dá rozmyslet i pro více množin. Naše definice rekurzí tak zachovala hlavní smysl, průnik se ptá na to, co je společné. Podobně bychom mohli rekurzí definovat sjednocení pro více množin a zjistilo by se, že i tato operace funguje stejně jako ta pro dva, sesypává prvky z množin do jedné společné. Nabízí se tak možnost definovat operace pro mnoho množin najednou pomocí velice čitelné podmínky.

Definice.

Nechť A_1, A_2, \dots, A_n jsou množiny ve stejném universu U . Definujeme

$$\bigcup_{k=1}^n A_k = \{x \in U; \exists k \in \{1, 2, \dots, n\} : x \in A_k\},$$

$$\bigcap_{k=1}^n A_k = \{x \in U; \forall k \in \{1, 2, \dots, n\} : x \in A_k\},$$

$$A_1 \times A_2 \times \dots \times A_n = \{(a_1, a_2, \dots, a_n); a_1 \in A_1 \wedge a_2 \in A_2 \wedge \dots \wedge a_n \in A_n\}.$$

Jestliže jde o stejné množiny, tedy $A_i = A$ pro všechna i , pak značíme $A_1 \times A_2 \times \dots \times A_n = A^n$.

Brzy uvidíme, že definice, kterou jsme nakonec zvolili, je výrazně vhodnější pro důkazy. Pokud bychom totiž operace zobecňovali na více objektů rekurzí, musely by všechny důkazy probíhat matematickou indukcí.

Adoptované definice mají ještě jednu výhodu, u průniku a sjednocení není vůbec důvod se omezovat na konečný počet množin. I když jich budeme mít nekonečně mnoho, pořád je možné se zeptat, zda existují prvky ležící úplně ve všech, popřípadě které prvky jsou v alespoň jedné.

Formálně se takové velké kolekce množin udělají tak, že se indexy k neberou z množin typu $\{1, 2, \dots, n\}$, ale dovolí se jakákoli množina indexů I , která může klidně být nekonečná. Jednoduchý příklad: Pro přirozené číslo i definujeme $A_i = \{i, i + 1\}$, pak dostáváme nekonečný soubor dvouprvkových množin $\{A_i\}_{i \in \mathbb{N}}$, například $A_{13} = \{13, 14\}$, $A_{42} = \{42, 43\}$ atd.



Definice.

Nechť A_i pro $i \in I$ jsou množiny ve stejném universu U , kde I je nějaká množina indexů. Definujeme

$$\bigcup_{i \in I} A_i = \{x \in U; \exists i \in I: x \in A_i\},$$

$$\bigcap_{i \in I} A_i = \{x \in U; \forall i \in I: x \in A_i\}.$$

Příklad 2a.c: Uvažujme $A_i = \{i, i + 1\}$ pro $i \in \mathbb{N}$. Nejprve se zamyslíme nad konečnými sjednoceními a průniky.

1) Jako inspiraci si všimneme, že $A_1 \cup A_2 = \{1, 2, 3\}$ a $A_1 \cup A_2 \cup A_3 = \{1, 2, 3, 4\}$, takže si tipneme, že pro $n \in \mathbb{N}$ je $\bigcup_{i=1}^n A_i = \{1, 2, 3, \dots, n, n + 1\}$. Důkaz:

$n + 1 \in A_n$ a pro $i = 1, \dots, n$ platí $i \in A_i$, proto $\{1, 2, 3, \dots, n, n + 1\} \subseteq \bigcup_{i=1}^n A_i$. Naopak pokud $j \in A_i$ pro nějaké $i = 1, \dots, n$, tak určitě $i \leq j \leq i + 1$, tedy $1 \leq j \leq n + 1$. Proto $\bigcup_{i=1}^n A_i \subseteq \{1, 2, 3, \dots, n, n + 1\}$.

2) Podobně si vyzkoušíme $A_1 \cap A_2$, $A_1 \cap A_2 \cap A_3$ (zkoušte to?), pak se zdá jasné, že

$$\bigcap_{i=1}^n A_i = \begin{cases} \{1, 2\}, & n = 1; \\ \{2\}, & n = 2; \\ \emptyset, & n \geq 3. \end{cases}$$

Důkaz: Pokud $n \geq 3$, pak ten průnik obsahuje prvky společné mimo jiné množinám A_1 a A_3 , tedy prvky z množiny $\{1, 2\} \cap \{3, 4\} = \emptyset$.

3) Teď se podíváme na nekonečné případy.

Protože každé $i \in \mathbb{N}$ leží alespoň v nějaké ze zúčastněných množin (konkrétně $i \in A_i$), dostáváme $\bigcup_{i \in \mathbb{N}} A_i = \mathbb{N}$.

Naopak každé číslo z \mathbb{N} se s většinou našich množin míjí (určitě $i \notin A_j$ pro $j > i$), proto $\bigcap_{i=1}^{\infty} A_i = \emptyset$.

△

V příkladu jsme použili dva způsoby specifikace indexů, $\bigcap_{i \in \mathbb{N}}$ a $\bigcap_{i=1}^{\infty}$, jsou rovnocenné a můžete si vybrat. Pokud je I jasné z kontextu, tak jeho specifikaci někdy vynecháváme a píšeme jen $\bigcup A_i$ či $\bigcap A_i$.

Příklad 2a.d: Množina indexů může být opravdu veliká. Nechť $I = \mathbb{R}$. Pro libovolné reálné číslo r definujeme C_r jako množinu všech čísel, které se při zápisu r (v desítkové soustavě) použily. Například $C_{1.07} = \{0, 1, 7\}$, také víme, že $17/6 = 2.83333\dots$, proto $C_{17/6} = \{2, 3, 8\}$, a $C_{\pi} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Dostáváme opravdu velký soubor množin $\{C_r\}$, ještě uvidíme, že je mnohem větší než ten z předchozího příkladu.

Protože je každá číslice použita v nějakém reálném čísle, je $\bigcup_{r \in \mathbb{R}} C_r = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Žádná cifra ale není ve všech číslech, proto $\bigcap_{r \in \mathbb{R}} C_r = \emptyset$.

△

Příklad 2a.e: Představme si, že I je množina všech molekul v mém těle (asi tedy bude konečná, ale neptejte se mě, kolik má prvků, ono se to i každou chvíli mění, raději do toho nebudeme vrtat).

Je-li i jedna taková molekula, pak M_i nechť je množina všech atomů, ze kterých se tato molekula skládá. Pak $\bigcup M_i$ je množina všech prvků, které jsou ve mne obsaženy, protože víme, že při tom sjednocování se ve výsledné množině opakování výskytu atomů ignorují a za každý typ atomu zůstane jen jeden zástupce.

Dobrá otázka je, jak vypadá $\bigcap M_i$. Možná tam bude uhlík, taky může být průnik prázdny, ale to je spíš otázka pro biology než matematiky.

△

Příklad 2a.f: Nechť $I = \mathbb{R}$. Pro $x \in I$ nechť L_x je množina všech lidí, která považuje x za své šťastné číslo. Pak $\bigcup L_x$ je množina všech číselně pověřivých lidí a $\bigcap L_x$ je množina všech lidí, pro které je šťastné každé číslo.

△

Máme krásnou obecnou definici a teď ukážeme, že to hlavní se tím nezkazilo.

!

Věta 2a.8. (de Morganovy zákony)

Nechť A_i pro $i \in I$ jsou množiny ve stejném universu U , kde I je nějaká množina indexů. Pak

$$\overline{\bigcup_{i \in I} A_i} = \bigcap_{i \in I} \overline{A_i} \quad \text{a} \quad \overline{\bigcap_{i \in I} A_i} = \bigcup_{i \in I} \overline{A_i}.$$

Důkaz (rutinní, poučný): Nejprve dokážeme první vztah, a to dlouze a komentovaně:

Prvek x leží v $\overline{\bigcup_{i \in I} A_i}$ právě tehdy (podle definice doplňku), když neleží v $\bigcup_{i \in I} A_i$, což je (podle definice sjednocení) právě tehdy, když není pravda, že existuje i , aby $x \in A_i$. Výraz $\neg(\exists i \in I : x \in A_i)$ je podle pravidel logiky totéž jako $\forall i \in I : \neg(x \in A_i)$, tedy x neleží v žádném A_i , což je (podle definice doplňku) právě tehdy, když leží ve všech $\overline{A_i}$, což je (podle definice průniku) právě tehdy, když leží v $\bigcap \overline{A_i}$.

Ukázali jsme, že $x \in \overline{\bigcup_{i \in I} A_i} \iff x \in \bigcap_{i \in I} \overline{A_i}$, čímž je rovnost této dvou množin dokázána.

Druhý vztah dokážeme v zásadě stejným důkazem, teď jej ale zapíšeme čistě symbolicky, abyste si procvičili překlad do lidštiny.

$$\begin{aligned} x \in \overline{\bigcap_{i \in I} A_i} &\iff x \notin \bigcap_{i \in I} A_i \iff \neg(x \in \bigcap_{i \in I} A_i) \iff \neg(\forall i \in I : x \in A_i) \\ &\iff \exists i \in I : \neg(x \in A_i) \iff \exists i \in I : x \notin A_i \iff \exists i \in I : x \in \overline{A_i} \iff x \in \bigcup_{i \in I} \overline{A_i}. \end{aligned}$$



Ten druhý důkaz byl tedy opravdu úsporný, kdyby se tak psaly matematické knihy, tak by vážily čtvrtinu. Nevýhoda by byla, že by je nikdo nedokázal rozumně číst, ani matematici ne, protože i je by zdržoval překlad z matematictiny do lidštiny, pro začátečníka by pak byly zcela nesrozumitelné. Klasické důkazy v knihách jsou tedy obvykle jakýmsi kompromisem mezi tím ukecaným a tím stručným výše.

Platí i distributivní zákon pro mnoho množin, viz cvičení 2a.5.

Poznámka: Všimněte si, že jsme nedefinovali množinový rozdíl pro více množin. Důvod je jednoduchý, odečítání není asociativní, tudíž nevíme, jak vlastně dělat $A - B - C$ (viz cvičení 2a.2 (ix)). Obdobně to ostatně funguje s čísly, umíme počítat $3 + 7 + 13$, ale co je $3/7/13$? Problém je právě v nedostatku asociativity, výrazy $(3/7)/13$ a $3/(7/13)$ nejsou stejné.



Na závěr ještě doplníme jednu definici.



Definice.

Množiny A, B se nazývají **disjunktní**, jestliže $A \cap B = \emptyset$.

To je velice užitečný pojem, až budeme dělat kombinatoriku, tak se bez něj neobejdeme.

Zjímavá otázka: Má smysl definovat tento pojem i pro více množin? Šlo by to udělat a říct, že množiny A_i jsou disjunktní, pokud mají prázdný průnik. V praxi se to ale nepoužívá, protože se ukazuje, že by tato podmínka o množinách mnoho neřekla. Proč tomu tak je?

Představme si několik množin A_1, \dots, A_n . Pokud by náhodou platilo $A_1 \cap A_2 = \emptyset$, tak už automaticky také $\bigcap A_i = \emptyset$. To znamená, že nám pak informace $\bigcap A_i = \emptyset$ vlastně vůbec neříká o množinách A_2, \dots, A_n . Pro většinu aplikací je proto tato informace nedostatečná.

Pokud chceme vyjádřit, že nějaké množiny spolu nemají nic společného, pak většinou potřebujeme jinou frázi: Chce se, aby množiny A_i byly **po dvou disjunktní**, což znamená, že $A_i \cap A_j = \emptyset$ pro libovolné $i \neq j$. Tato podmínka odpovídá situaci, kterou si ve Vennově diagramu představíme jako zcela separátní kroužky.

! 2a.9 Reprezentace množin v počítačích

Nekonečně mnoho dat počítač nespolekne, takže už z principu budeme v počítačích pracovat s množinami konečnými. Existuje pak jednoduchý způsob, jak si je reprezentovat. Začneme tím, že vezmeme konečné universum a jeho prvky si očíslujeme, $U = \{u_1, \dots, u_n\}$. Každá podmnožina A tohoto universa se pak dá jednoduše zakódovat jako binární řetězec (číslo) délky n tak, že i -tá cifra je 1, pokud $u_i \in A$, jinak je to nula.

Například v universu $U = \{u_1 = 1, u_2 = 13, u_3 = a, u_4 = \diamond, u_5 = 23\}$ se množina $A = \{13, 23\}$ zakóduje jako 01001, popřípadě 10010, podle toho, jestli jsme si vybrali kódování (čtení řetězce) zprava doleva nebo naopak. V zásadě je to jedno, jen se pak toho kódování musíme už pořád držet :-).

Jednou z velkých výhod této reprezentace je, že se pak krásně dělají množinové operace pomocí logických operací na bitech odpovídajících kódovacích řetězců. Sjednocení množin odpovídá logická disjunkce jednotlivých bitů, naopak průnik je přesně konjunkce neboli obyčejné binární násobení bitů. To jsou operace, které má počítač rád, takže je všechno v pohodě.

S 2a.10 Poznámka (jak psát a číst důkazy): Přichází cvičení a čtenář by měl začít psát důkazy. O stránce logické jsme již psali i v předchozí kapitole, zde se zaměříme na jedno pomocné hledisko. Chybně napsaný důkaz lze často odhalit i tím, že jej prostě nelze přečíst jako text. I matematický důkaz totiž musí dávat české věty (podmět, příslušek a tak podobně). Pro začátečníka je toto důležité zejména v situaci, kdy se rozhodne ušetřit čas použitím matematických symbolů.

Pořád platí, že když symboly zase nahradíme odpovídajícími slovy, musí vzniknout rozumný text. To se týká zejména symbolu \Rightarrow , který běžně používáme ve významu „z toho nalevo vyplývá to napravo“, neboť $A \Rightarrow B$ čteme například jako „platí A , proto také platí B “.

Zkusme si česky přečíst něco, co autor skripta potkal v písemce:

$$\forall x \Rightarrow x > 2.$$

Česky například „Pro každé x , z toho vyplývá, že $x > 2$ “.

Člověk ani nemusí umět matematiku, aby jej napadlo, že je něco špatně. Je tedy dobré si po zapsání svého argumentu zkoušit říct slovy. Pokud to nefunguje, je někde problém, třeba jen v zápisu.

Další úroveň, na které důkaz musí dávat smysl, je úroveň pojmová. Zacitujme opět z jedné písemky:

Protože $A \cap B$, musí být $A \subseteq B$.

Toto je implikace, tudíž by člověk čekal, že po slově „protože“ bude nějaký pravdivý fakt, ze kterého se pak něco dále dovodí. Jenže $A \cap B$ není fakt, není to něco, co může být pravda či nepravda. To je operace, jejímž výsledkem je nějaká množina. Dotyčná věta tedy nedává smysl již na úrovni pojmu, ani se nemusíme zamýšlet nad tím, co se říká v její druhé půlce. Kdyby tam ale bylo třeba „ $A \cap B$ je něco“, tak už to je výrok (buď je to pravda nebo ne, podle toho, co je to „něco“) a věta není vykloubená (což ještě neznamená, že je celá implikace pravdivá, to je ta další a rozhodující úroveň, na které to musí fungovat).

Teď přijdou cvičení a jejich řešení jsou často psána vysoce kodenzovaně. Čtenář si tak může procvičit překlad těchto úvah do češtiny, mělo by to vždy rozumně jít.

△

Cvičení

Cvičení 2a.1 (rutinní^o, zkouškové^{*}, dobré^{*}, poučné⁺): Pravidel pro množinové operace je mnohem více, než jsme uvedli v textu. Dokažte následující:

Nechť A, B, C jsou množiny v univerzu U . Pak platí:

- (i)^o $A - B \subseteq A$;
- (ii)^{**+} $A - B = A \cap \overline{B}$;
- (iii)^{*} $A \cap (B - A) = \emptyset$;
- (iv)^{*} $(A - B) \cap (B - C) = \emptyset$;
- (v)^{*} $(A - B) - C \subseteq A - C$;
- (vi)^{*+} $A \cup (B - A) = A \cup B$;
- (vii)^{*+} $(A \cup B) - C = (A - C) \cup (B - C)$;
- (viii)^{*+} $A - (B \cup C) = (A - B) \cap (A - C)$;
- (ix)^{*+} $A - (B \cap C) = (A - B) \cup (A - C)$;
- (x)^{**+} $A \cap (B - C) = (A \cap B) - (A \cap C)$;
- (xi)^{*+} $A \subseteq B$ právě tehdy, když $\overline{B} \subseteq \overline{A}$;
- (xii)^{*+} $A \subseteq B$ právě tehdy, když $A \cap B = A$;
- (xiii)^{*+} $A \subseteq B$ právě tehdy, když $A \cup B = B$;
- (xiv) $P(A) \subseteq P(B) \implies A \subseteq B$.

Poznámka: Všimněte si, že ve třech případech se jedná o distributivní zákon. Bod (vii) ukazuje, že – umí roznásobit závorku se sjednocením zprava, ale v (viii) vidíme, že zleva už to nejde, tam je třeba vzorec upravit. Bod (ix) ukazuje, že \cap umí roznásobit závorku s odčítáním, pro další kombinace operací se podívejte do následujícího cvičení.

Cvičení 2a.2 (poučné, zkouškové^{*}, dobré^{*}): Rozhodněte, zda pro libovolné množiny A, B, C platí následující vztahy. Pak buď příslušný vztah dokažte, nebo dokažte, že neplatí.

V případě, že rovnost neplatí, rozmyslete si, jestli nebude platit alespoň nějaká inkluze, a tu dokažte.

Poznámka: Některé důkazy jsou dosti trikové, ale u všech příkladů byste měli být schopni určit, zda uvedená rovnost platí, popřípadě která inkluze platí. Dobré důkazy klidně vynechte. Mimochodem, v bodech (viii)-(xii) zkoumáme platnost různých verzí distributivního zákona.

- (i)^{*} $(A - B) \cup B = A$;
- (ii)^{*} $(A \cap B) \cup (A \cap \overline{B}) = A$;
- (iii)^{*} $(A - B) - C = (A - C) - B$;
- (iv)^{**} $(A - B) - C = A - (B - C)$;
- (v)^{*} $(A - B) \cup (B - C) = A - C$;
- (vi)^{*} $P(A \cap B) = P(A) \cap P(B)$;
- (vii)^{*} $P(A \cup B) = P(A) \cup P(B)$;
- (viii)^{*} $A \cup (B - C) = (A \cup B) - (A \cup C)$;
- (ix)^{*} $A - (B \cap C) = (A - B) \cap (A - C)$;
- (x)^{*} $A - (B \cup C) = (A - B) \cup (A - C)$;
- (xi): $(A \cap B) - C = (A - C) \cap (B - C)$;
- (xii): $(A \cup B) - C = (A - C) \cup (B - C)$;
- (xiii)^{*} $(A - B) - C = (A - C) - (B - C)$.

Cvičení 2a.3 (rutinní): Nechť A, B, C, D jsou množiny. Platí $(A - B) - (C - D) = (A - C) - (B - D)$? Svou odpověď zdůvodněte.

Cvičení 2a.4 (poučné): Uvažujme množinu indexů $I = \mathbb{N}$ a množiny M_i pro $i \in I$. Najděte výsledky operací $\bigcup_{i=1}^n M_i$, $\bigcup_{i=1}^{\infty} M_i$, $\bigcap_{i=1}^n M_i$ a $\bigcap_{i=1}^{\infty} M_i$, jestliže
 (i) $M_i = \{1, 2, 3, \dots, i\}$ pro $i \in \mathbb{N}$;
 (ii) $M_i = \{i, i+1, i+2, \dots\}$ pro $i \in \mathbb{N}$.

Cvičení 2a.5 (poučné): Nechť A a A_i pro $i \in I$ jsou množiny v universu U . Dokažte, že pak platí následující:

- (i) $A \cap \bigcup_{i \in I} A_i = \bigcup_{i \in I} (A \cap A_i)$;
- (ii) $A \cup \bigcap_{i \in I} A_i = \bigcap_{i \in I} (A \cup A_i)$.

Cvičení 2a.6 (poučné): Uvažujme množinu indexů $I = \mathbb{R}^+ = (0, \infty)$ a množiny M_r pro $r \in I$. Najděte $\bigcup_{r \in I} M_r$ a $\bigcap_{r \in I} M_r$, jestliže

- (i) $M_r = (-r, 13+r)$;
- (ii) $M_r = \langle -r, 13+r \rangle$;
- (iii) $M_r = (-r, r)$.

Cvičení 2a.7 (poučné): Uvažujme množiny indexů $I = (0, 1)$ a $J = \langle 0, 1 \rangle$. Najděte $\bigcup_{r \in I} M_r$ a $\bigcap_{r \in I} M_r$, $\bigcup_{r \in J} M_r$ a $\bigcap_{r \in J} M_r$, jestliže

- (i) $M_r = (-r, 13+r)$;
- (ii) $M_r = \langle -r, 13+r \rangle$.

Řešení:

2a.1: (i): $\forall x \in A - B: x \in A \wedge x \notin B \implies x \in A$.

(ii): Zkusíme oba směry najednou: $x \in A - B \iff x \in A \wedge x \notin B \iff x \in A \wedge x \in \overline{B} \iff A \cap \overline{B}$.

(iii): Sporem, existuje $x \in A \cap (B - A)$, pak $x \in A \wedge x \in (B - A) \implies x \in A \wedge (x \in B \wedge x \notin A) \implies x \in A \wedge x \notin A$, spor.

(iv): Sporem, existuje $x \in (A - B) \cap (B - C)$, pak $x \in (A - B) \wedge x \in (B - C) \implies (x \in A \wedge x \notin B) \wedge (x \in B \wedge x \notin C) \implies x \in B \wedge x \notin B$, spor.

Nebo pomocí (ii): $(A - B) \cap (B - C) = (A \cap \overline{B}) \cap (B \cap \overline{C}) = A \cap (\overline{B} \cap B) \cap \overline{C} = A \cap \emptyset \cap \overline{C} = \emptyset$.

(v): $x \in (A - B) - C \implies (x \in A \wedge x \notin B) \wedge x \notin C \implies x \in A \wedge x \notin C \implies x \in A - C$.

Nebo pomocí (ii): $(A - B) - C = (A \cap \overline{B}) \cap \overline{C} = (A \cap \overline{C}) \cap \overline{B} \subseteq A \cap \overline{C} = A - C$.

(vi): Dokázat dvě inkluze. 1) $A \cup (B - A) \subseteq A \cup B: \forall x \in A \cup (B - A): x \in A \vee x \in (B - A)$

$\implies x \in A \vee (x \in B \wedge x \notin A) \implies x \in A \vee x \in B \implies x \in A \cup B$.

2) $A \cup B \subseteq A \cup (B - A): \forall x \in A \cup B: x \in A \vee x \in B$. Rozdělíme na případy. Pokud $x \in A$, pak $x \in A \cup (B - A)$. Pokud $x \in B$, tak zase dva případy. Jestliže $x \in B \wedge x \in A$, pak $x \in A$ a přejdeme na předchozí. Jestliže $x \in B \wedge x \notin A$, pak $x \in B - A \implies x \in A \cup (B - A)$.

Nebo pomocí (ii): $A \cup (B - A) = A \cup (B \cap \overline{A}) = (A \cup B) \cap (A \cup \overline{A}) = (A \cup B) \cap U = A \cup B$.

(vii): Zkusíme oba směry najednou pomocí distributivního zákona pro formální logiku: $x \in (A \cup B) - C \iff (x \in A \vee x \in B) \wedge x \notin C \iff (x \in A \wedge x \notin C) \vee (x \in B \wedge x \notin C) \iff x \in (A - C) \vee x \in (B - C) \iff x \in (A - C) \cup (B - C)$.

Snažší varianta pomocí (ii): $(A \cup B) - C = (A \cup B) \cap \overline{C} = (A \cap \overline{C}) \cup (B \cap \overline{C}) = (A - C) \cup (B - C)$.

(viii): Zkusíme oba směry najednou pomocí distributivního zákona a deMorganova zákona pro formální logiku:

$A - (B \cup C) \iff x \in A \wedge \neg(x \in B \cup C) \iff x \in A \wedge \neg(x \in B \vee x \in C) \iff$

$x \in A \wedge \neg(x \in B) \wedge \neg(x \in C) \iff x \in A \wedge x \in A \wedge \neg(x \in B) \wedge \neg(x \in C) \iff$

$(x \in A \wedge x \notin B) \wedge (x \in A \wedge x \notin C) \iff x \in (A - B) \wedge x \in (A - C) \iff x \in (A - B) \cap (A - C)$.

Snažší varianta pomocí (ii) a deMorgana pro množiny:

$A - (B \cup C) = A \cap \overline{B \cup C} = A \cap \overline{B} \cap \overline{C} = A \cap A \cap \overline{B} \cap \overline{C} = (A \cap \overline{B}) \cap (A \cap \overline{C}) = (A - B) \cap (A - C)$.

(ix): Zkusíme oba směry najednou pomocí distributivního zákona a deMorganova zákona pro formální logiku: $A - (B \cap C) \iff x \in A \wedge \neg(x \in B \cap C) \iff x \in A \wedge \neg(x \in B \wedge x \in C) \iff$

$x \in A \wedge (\neg(x \in B) \vee \neg(x \in C)) \iff (x \in A \wedge x \notin B) \vee (x \in A \wedge x \notin C) \iff x \in (A - B) \vee x \in (A - C) \iff (A - B) \cup (A - C)$.

Snažší varianta pomocí (ii) a deMorgana pro množiny:

$A - (B \cap C) = A \cap \overline{B \cap C} = A \cap (\overline{B} \cup \overline{C}) = (A \cap \overline{B}) \cup (A \cap \overline{C}) = (A - B) \cup (A - C)$.

(x): Tento důkaz je jedním směrem relativně snadný a používá distributivní zákon, těžší je najít směr opačný. Ukáže se, že ten snadný směr je vlastně ekvivalentní a funguje i naopak (rozmyslete si).

$$\begin{aligned} x \in (A \cap B) - (A \cap C) &\iff x \in A \cap B \wedge \neg(x \in A \cap C) \iff x \in A \cap B \wedge \neg(x \in A \wedge x \in C) \iff \\ (x \in A \wedge x \in B) \wedge (x \notin A \vee x \notin C) &\iff [(x \in A \wedge x \in B) \wedge x \notin A] \vee [(x \in A \wedge x \in B) \wedge x \notin C] \iff \\ [x \in A \wedge x \notin A \wedge x \in B] \vee [x \in A \wedge (x \in B \wedge x \notin C)] &\iff [F \wedge x \in B] \vee [x \in A \wedge x \in (B - C)] \iff \\ F \vee [x \in A \cap (B - C)] &\iff x \in A \cap (B - C). \end{aligned}$$

(xi): 1) Předpoklad $A \subseteq B$, chceme $\forall b \in \overline{B}: b \in \overline{A}$. Jedna možnost sporem: Nechť existuje $b \in \overline{B}$ takové, že $b \notin \overline{A}$. Pak $b \in A$, podle předpokladu $b \in B$. Takže $b \in \overline{B} \wedge b \in B$, spor.

Alternativa: Předpoklad říká $\forall x: x \in A \implies x \in B$. Přejdeme k ekvivalentní obměně: $\forall x: x \notin B \implies x \notin A$ neboli $\forall x: x \in \overline{B} \implies x \in \overline{A}$, přesně jak potřebujeme.

2) Předpoklad $\overline{B} \subseteq \overline{A}$, chceme $A \subseteq B$. Podle 1) plyne z předpokladu $\overline{\overline{A}} \subseteq \overline{\overline{B}}$, což je právě $A \subseteq B$.

(xii): 1) Předpoklad $A \subseteq B$, chceme $A \cap B = A$. Ukážeme dvě inkluze.

Důkaz $A \cap B \subseteq A$: $\forall x \in A \cap B: x \in A \wedge x \in B \implies x \in A$.

Důkaz $A \subseteq A \cap B$: $\forall a \in A: a \in B$ dle předpokladu. Proto $a \in A \wedge a \in B \implies a \in A \cap B$.

2) Předpoklad $A \cap B = A$, chceme $A \subseteq B$. Důkaz:

$\forall a \in A: a \in A \cap B$ dle předpokladu, proto $a \in A \wedge a \in B \implies a \in B$.

(xiii): 1) Předpoklad $A \subseteq B$, chceme $A \cup B = B$. Ukážeme dvě inkluze. Důkaz $A \cup B \subseteq B$:

$\forall x \in A \cup B: x \in A \vee x \in B$, ale předpoklad dává, že i v případě $x \in A$ je $x \in B$, proto každopádně $x \in B$.

Důkaz $B \subseteq A \cup B$: $\forall b \in B: b \in A \wedge b \in B \implies b \in A \cup B$.

2) Předpoklad $A \cup B = B$, chceme $A \subseteq B$. Důkaz: $\forall a \in A: a \in A \cup B = B$ dle předpokladu, proto $a \in B$.

(xiv): Předpoklad říká, že prvky $P(A)$ jsou i prvky $P(B)$, zde je zásadní si uvědomit, že množina $P(A)$ má jako prvky podmnožiny A . Chceme ukázat, že prvky z A jsou v B :

$a \in A \implies \{a\} \subseteq A \implies \{a\} \in P(A) \implies \{a\} \in P(B) \implies \{a\} \subseteq B \implies a \in B$.

2a.2: (i): Protipříklad: třeba $A = \emptyset$, $B = \{13\}$. Platí ale $A \subseteq (A - B) \cup B$: Nechť $a \in A$ libovolné. Dvě možnosti: $x \in B$, pak $x \in (A - B) \cup B$, nebo $x \notin B$, pak $x \in A \wedge x \notin B \implies x \in (A - B) \implies x \in (A - B) \cup B$.

Formální důkaz: $x \in A \iff x \in A \wedge T \iff x \in A \wedge (x \notin B \vee x \in B) \iff$

$(x \in A \wedge x \notin B) \vee (x \in A \wedge x \in B) \implies x \in (A - B) \vee x \in B \implies x \in (A - B) \cup B$.

(ii): Platí. Dvě inkluze. $(A \cap B) \cup (A \cap \overline{B}) \subseteq A$: $x \in (A \cap B) \cup (A \cap \overline{B}) \implies x \in (A \cap B) \vee x \in (A \cap \overline{B}) \implies x \in A \vee x \in A \implies x \in A$.

$A \subseteq (A \cap B) \cup (A \cap \overline{B})$: Nechť $x \in A$. Dvě možnosti. Pokud $x \in B$, pak $x \in A \wedge x \in B \implies x \in (A \cap B) \implies x \in (A \cap B) \cup (A \cap \overline{B})$. Nebo $x \notin B$, pak $x \in A \wedge x \notin B \implies x \in (A \cap \overline{B}) \implies x \in (A \cap B) \cup (A \cap \overline{B})$.

(iii): Platí, důkaz obou směrů najednou: $x \in (A - B) - C \iff x \in (A - B) \wedge x \notin C \iff$

$(x \in A \wedge x \notin B) \wedge x \notin C \iff (x \in A \wedge x \notin C) \wedge x \notin B \iff x \in (A - C) \wedge x \notin B \iff x \in (A - C) - B$.

(iv): Protipříklad: $A = B = C = \{1\}$. Platí ale $(A - B) - C \subseteq A - (B - C)$ (důkaz dost trikový):

$x \in (A - B) - C \implies x \in (A - B) \wedge x \notin C \implies x \in A \wedge x \notin B \wedge x \notin C \implies x \in A \wedge x \notin B$

$\implies x \in A \wedge x \notin B \vee x \in C \implies x \in A \wedge \neg(x \in B \wedge x \notin C) \implies x \in A \wedge \neg[x \in (B - C)]$

$\implies x \in A \wedge x \notin (B - C) \implies x \in A - (B - C)$.

(v): Protipříklad: třeba $A = C = \emptyset$, $B = \{13\}$. Platí ale $A - C \subseteq (A - B) \cup (B - C)$: $x \in A - C \implies x \in A \wedge x \notin C$.

Dvě možnosti. Pokud $x \in B$, tak $x \in A \wedge x \in B \wedge x \notin C \implies x \in B \wedge x \notin C \implies x \in (B - C) \implies$

$x \in (A - B) \cup (B - C)$.

Nebo $x \notin B$, pak $x \in A \wedge x \notin B \wedge x \notin C \implies x \in A \wedge x \notin B \implies x \in (A - B) \cup (B - C)$.

(vi): Platí, dokážeme dvě inkluze. 1) $P(A \cap B) \subseteq P(A) \cap P(B)$: Nechť $x \in P(A \cap B)$, pak x je vlastně podmnožina $A \cap B$. Proto $x \subseteq A \wedge x \subseteq B \implies x \in P(A) \wedge x \in P(B) \implies x \in P(A) \cap P(B)$.

2) $P(A) \cap P(B) \subseteq P(A \cap B)$: $x \in P(A) \cap P(B) \implies x \in P(A) \wedge x \in P(B) \implies x \subseteq A \wedge x \subseteq B \implies x \subseteq A \cap B \implies x \in P(A \cap B)$.

(vii): Protipříklad: třeba $A = \{1, 2\}$, $B = \{3, 4\}$. Pak množina $M = \{2, 3\}$ leží v $P(A \cup B)$, ale není ani v $P(A)$, ani v $P(B)$, tedy není v jejich sjednocení. Platí ale $P(A) \cup P(B) \subseteq P(A \cup B)$:

$x \in P(A) \cup P(B) \implies x \in P(A) \vee x \in P(B) \implies x \subseteq A \vee x \subseteq B \implies x \subseteq A \cup B \implies x \in P(A \cup B)$.

(viii): Protipříklad: $A = \{13\}$, $B = C = \emptyset$. Platí ale $(A \cup B) - (A \cup C) \subseteq A \cup (B - C)$: $x \in (A \cup B) - (A \cup C) \iff$

$(x \in A \vee x \in B) \wedge \neg(x \in A \vee x \in C) \iff (x \in A \vee x \in B) \wedge (x \notin A \wedge x \notin C) \iff$

$(x \in A \wedge x \notin A \wedge x \notin C) \vee (x \in B \wedge x \notin A \wedge x \notin C) \implies F \vee (x \in (B - C)) \iff x \in (B - C) \implies$

$x \in A \cup (B - C)$.

(ix): Protipříklad: $A = B = \{1\}$, $C = \emptyset$. Platí ale $(A - B) \cap (A - C) \subseteq A - (B \cap C)$: $x \in (A - B) \cap (A - C) \iff$

$x \in (A - B) \wedge x \in (A - C) \iff x \in A \wedge x \notin B \wedge x \in A \wedge x \notin C \iff x \in A \wedge (\neg x \in B \wedge \neg x \in C) \implies$

$x \in A \wedge (\neg x \in B \vee \neg x \in C) \iff x \in A \wedge \neg(x \in B \wedge x \in C) \iff$

$x \in A \wedge \neg(x \in B \cap C) \iff x \in A \wedge x \notin (B \cap C) \iff x \in A - (B \cap C)$.

(x): Protipříklad: $A = B = \{1\}$, $C = \emptyset$. Platí ale $A - (B \cup C) \subseteq (A - B) \cup (A - C)$: $x \in A - (B \cup C) \iff$

$x \in A \wedge x \notin (B \cup C) \iff x \in A \wedge \neg(x \in B \cup C) \iff x \in A \wedge \neg(x \in B \vee x \in C) \iff$

$x \in A \wedge x \notin B \wedge x \notin C \iff (x \in A \wedge x \notin B) \wedge (x \in A \wedge x \notin C) \iff x \in (A - B) \wedge x \in (A - C) \implies$

$$x \in (A - B) \vee x \in (A - C) \iff x \in (A - B) \cup (A - C).$$

(xi): Platí, zkusíme oba směry najednou, začneme od složitějšího: $x \in (A - C) \cap (B - C) \iff x \in (A - C) \wedge x \in (B - C) \iff x \in (A \wedge x \notin C) \wedge (B \wedge x \notin C) \iff x \in A \wedge x \notin C \wedge x \in B \wedge x \notin C \iff x \in A \wedge x \in B \wedge x \notin C \iff x \in (A \cap B) \wedge x \notin C \iff x \in (A \cap B) - C.$

(xii): Platí, zkusíme oba směry najednou: $x \in (A \cup B) - C \iff x \in (A \cup B) \wedge x \notin C \iff (x \in A \vee x \in B) \wedge x \notin C \iff (x \in A \wedge x \notin C) \vee (x \in B \wedge x \notin C) \iff x \in (A - C) \vee x \in (B - C) \iff x \in (A - C) \cup (B - C).$

(xiii): Platí, zkusíme oba směry najednou, začneme od složitějšího: $x \in (A - C) - (B - C) \iff x \in (A - C) \wedge x \notin (B - C) \iff (x \in A \wedge x \notin C) \wedge \neg(x \in B \wedge x \notin C) \iff (x \in A \wedge x \notin C) \wedge (x \notin B \vee x \in C) \iff (x \in A \wedge x \notin C \wedge x \notin B) \vee (x \in A \wedge x \notin C \wedge x \in C) \iff (x \in A \wedge x \notin C \wedge x \notin B) \vee (x \in A \wedge F) \iff (x \in A \wedge x \notin C \wedge x \notin B) \vee F \iff x \in A \wedge x \notin C \wedge x \notin B \iff (x \in A \wedge x \notin B) \wedge x \notin C \iff (x \in (A - B) \wedge x \notin C) \iff x \in (A - B) - C.$

2a.3: Neplatí, intuitivně vidíme, že vlevo odebíráme celé C , zatímco vpravo jen zmenšené C , na tomto pocitu zkusíme založit protipříklad: $A = C = D = \{13\}$, $B = \emptyset$.

2a.4: (i): $\{1, 2, 3, \dots, n\}$, \mathbb{N} , $\{1\}$, $\{1\}$.

(ii): \mathbb{N} , \mathbb{N} , $\{n, n+1, n+2, n+3, \dots\}$, \emptyset .

2a.5: (i): $x \in A \cap \bigcup_{i \in I} A_i \iff x \in A \wedge x \in \bigcup_{i \in I} A_i \iff x \in A \wedge (\exists i \in I : x \in A_i) \iff$

$\exists i \in I : (x \in A \wedge x \in A_i) \iff \exists i \in I : (x \in A \cap A_i) \iff x \in \bigcup_{i \in I} (A \cap A_i).$

(ii): $x \in A \cup \bigcap_{i \in I} A_i \iff x \in A \vee x \in \bigcap_{i \in I} A_i \iff x \in A \vee (\forall i \in I : x \in A_i) \iff \forall i \in I : (x \in A \vee x \in A_i) \iff$

$\forall i \in I : (x \in A \cup A_i) \iff x \in \bigcap_{i \in I} (A \cup A_i).$

2a.6: (i): $\mathbb{R}, \langle 0, 13 \rangle$; (ii): $\mathbb{R}, \langle 0, 13 \rangle$; (iii): $\mathbb{R}, \{0\}$.

2a.7: (i): $(-1, 14), \langle 0, 13 \rangle, (-1, 14), (0, 13)$; (ii): $(-1, 14), \langle 0, 13 \rangle, \langle -1, 14 \rangle, \langle 0, 13 \rangle$.

2b. Zobrazení

Z kapitoly 2a známe pojem množiny, který nám v zásadě umožní vyjádřit to, že něco máme či nemáme. Často jsme ale v situaci, že máme nějaké objekty a mezi těmito objekty existují určité vztahy. Abychom tuto situaci mohli zkoumat, potřebujeme matematickou strukturu, která ony vztahy dokáže zachytit. Takové struktury existují, dokonce je jich více, aby dokázaly správně zachytit různé typy vztahů.

Zde se soustředíme na vztah, který má podobu jednoduchého přiřazení. Například každý člověk má rodné číslo. Matematicky to vidíme tak, že máme množinu lidí A a množinu čísel B a každému člověku z množiny A přiřadíme právě jedno číslo z množiny B . Tím vzniká vztah. Podobně funguje třeba přiřazení, které každému místu na zemi dává souřadnici GPS, či přiřazení, které každému konkrétnímu zvířeti (savci) přiřadí jeho pohlaví.

Tím se dostáváme k pojmu zobrazení, což je jeden ze základních matematických nástrojů. Všimněme si, že tento pojem nebude schopen obsáhnout například situaci, která každému žákovi přiřadí učitele, který jej učí, protože takových učitelů je více. My bychom samozřejmě mohli definici zobrazení udělat tak, aby umožňovala přiřazovat více objektů, ale tím by vznikl pojem, který se chová úplně jinak, na takovéto situace máme jiné nástroje.

Student se s pojmem podobným zobrazení již setkal, když pracoval s funkcemi. Tato zkušenost mu zde pomůže, ale neměl by na ni spoléhat až příliš. Hodně středoškoláků si ze školy odnáší představu, že funkce je vzoreček, a právě na toto je třeba rychle zapomenout. Mnohem lepší je dívat se na funkci jako na černou skříňku, které podstrčíme číslo a ona na oplátku jiné vydá. Když máme velké štěstí, tak se tento proces dá vyjádřit vzorečkem, ale rozhodně se na to nedá spoléhat.

Jednoduchý příklad pro čtenáře, pro které je to nové: Definujme funkci f následovně. Jestliže je reálné číslo x vyjádřitelné jako desetinné číslo s konečným rozvojem, pak je hodnota $f(x)$ dánou jako ta cifra, která se v jeho zápisu vyskytuje nejčastěji; pokud by byla plichta, bere se nejmenší taková cifra. Pokud se x nedá vyjádřit pomocí konečného desetinného rozvoje, pak definujeme $f(x) = 0$. Touto definicí je f definováno pro všechna reálná čísla, třeba $f(146824834) = 4$, $f(714.397721) = 7$, $f(0.333) = 3$, naopak $f(\pi) = 0$ či třeba $f(\frac{1}{3}) = 0$. Je to naprostě normální funkce, jen ji nelze vyjádřit vzorečkem.

Smiřme se tedy s tím, že funkce je jakékoli posílátko, které bere čísla a posílá je na jiná čísla. Jak ale takovou funkci reprezentovat matematicky, když nemůžeme spoléhat na vzoreček? Nejjednodušší je představit si, že je funkce dána množinou uspořádaných párů (počáteční bod, cílový bod), což vlastně přesně odpovídá grafu. Tento způsob nás vrací zpět k množinám, což je objekt, se kterým umíme pracovat, je to tedy perspektivní představa.

Když se na funkce podíváme tímto způsobem, hned se nabídne nápad, že by se nemusela posílat jen čísla, ale i jiné objekty, můžeme si klidně představit třeba černou skříňku, které dáváme písmenka a ona na oplátku vydává třeba různá lízátka. I fungování takového skříňky by šlo (přinejmenším teoreticky) zachytit jako množinu dvojic (písmenko, lízátka). Tím se dostáváme k obecné definici.

!

Definice.

Nechť A, B jsou neprázdné množiny. Definujeme **zobrazení** z A do B jako libovolnou podmnožinu T množiny $A \times B$ splňující

$$\forall a \in A \exists !b \in B: (a, b) \in T.$$

Fakt $(a, b) \in T$ značíme $T(a) = b$.

Množina A je **definiční obor** T , značeno $D(T)$, množina B je cílová množina T . Definujeme také **obor hodnot** T jako

$$R(T) = \{b \in B; \exists a \in A: T(a) = b\} = \{T(a); a \in A\}.$$

By a **mapping** we mean any subset T of $A \times B$ satisfying the following condition: For every $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in T$. We denote this $T(a) = b$. The set A is called the **domain** of T , denoted $D(T)$, and the set B is called the **codomain** of T . We also define the **range** of T as $R(T) = \{T(a); a \in A\}$.

! Připomínáme, že $\exists !$ čteme „existuje právě jedno“ (viz kapitola 1a), tedy definice opravdu vyžaduje, aby posílátka neposílalo jeden vstupní objekt na více míst. Tuto podstatu zobrazení coby posílátka dobře vystihuje ono alternativní značení $T(a) = b$, používá se také (řídčeji) $T: a \mapsto b$. Fakt, že T je zobrazení z A do B , pak často značíme jako $T: A \rightarrow B$.

Poznamenejme, že zápis $T(a) = b$ je sice intuitivně příjemný, ale je třeba si uvědomit, že se tím rozhodně nenaznačuje, že by se a dosazovalo do nějakého vzorečku. Je to jen sugestivní zkratka pro fakt, že dvojice (a, b) leží v T . Práce se zápisem $T(a) = b$ je často příjemná, ale jsou chvíle (zejména v některých důkazech), kdy nezbývá než přejít ke skutečnému významu a používat značení $(a, b) \in T$.

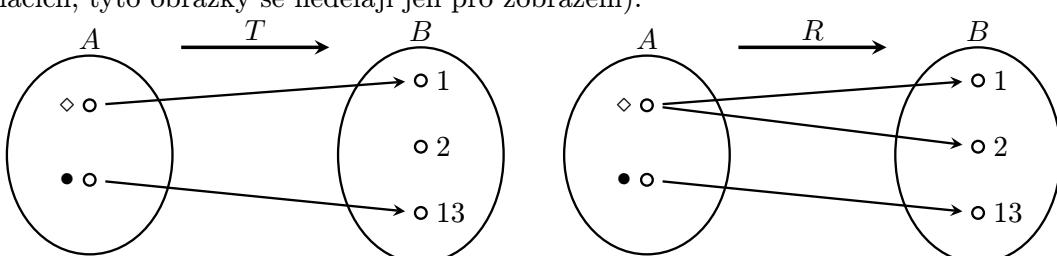
Příklad 2b.a: Uvažujme množiny $A = \{\diamond, \bullet\}$ a $B = \{1, 2, 13\}$. Pak $\widehat{R} = \{(\diamond, 13)\}$ není zobrazení $A \rightarrow B$, protože prvku $a = \bullet$ není nic přiřazeno. Také $R = \{(\diamond, 13), (\bullet, 1), (\diamond, 2)\}$ není zobrazení, protože prvku $a = \diamond$ jsou přiřazeny dvě různá b .

Jak se dá čekat, teď přijde zobrazení, třeba $T = \{(\diamond, 1), (\bullet, 13)\}$. Toto je správný formální zápis podle definice, ale máme k dispozici i možná přirozenější zápis ve formě výčtu $T(\diamond) = 1, T(\bullet) = 13$.

Toto zobrazení má definiční obor $D(T) = A$ a obor hodnot $R(T) = \{1, 13\}$.

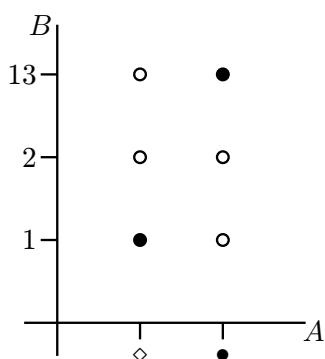
△

! Jak si takové zobrazení můžeme znázornit? Každému a je přiřazeno jediné b , toto posílání $a \mapsto b$ se přirozeně zachytí šípkami. Ukážeme obrázek pro naše T a také pro R z příkladu 2b.a (R sice není zobrazení, ale jak uvidíme v kapitole o relacích, tyto obrázky se nedělají jen pro zobrazení).



Takový obrázek je velice užitečný, například obor hodnot v něm vidíme jako všechny body z B , do kterých vede šipka. Hned také vidíme, kdy nějaká podmnožina $A \times B$ není zobrazení, buď pro ni nějaké a nemá šipku žádnou, nebo jich má více (viz R).

Alternativní možnost znázornění je vyjít z definice, tedy vnímat T jako nějakou podmnožinu kartézskeho součinu $A \times B$, který si (přinejmenším u konečných množin) tradičně znázorňujeme jako obdélníkovou síť bodů. V ní zvýrazníme dvojice ležící v T .



Tento obrázek se u diskrétních příkladů (konečné množiny a podobně) moc nepoužívá, ale velmi užitečný začne být, když $A = \mathbb{R}$, tak jej asi čtenář zná. Pak většinou mluvíme o **funkcích**. Někteří autoři používají název funkce i pro obecná zobrazení, jiní si jej rezervují jen pro zobrazení, jejichž definiční obor je v \mathbb{R} , popřípadě v \mathbb{Z} . Zde se držíme druhého způsobu, takže když pracujeme s množinami, budeme mluvit o zobrazení a upřednostňovat ten šípkový obrázek.

Příklad 2b.b: Uvažujme A coby množinu všech studentů a $B = \mathbb{R}$. Definujeme zobrazení $T: A \mapsto B$ předpisem, že pro konkrétní $a \in A$ udává $T(a)$ studijní průměr studenta a k určitému pevně zvolenému dni. Pak by T mělo být zobrazení.

△

! Příklad 2b.c: Uvažujme A coby množinu všech studentů a B množinu všech předmětů. Jestliže definujeme T jako množinu všech dvojic $(a, b) \in A \times B$ takových, že student a si v tomto semestru zapsal kurs b , pak to téměř určitě nebude zobrazení, protože se nejspíše najde nějaký sabotující student a , který si zapsal více než jeden kurs, díky čemuž se toto a vyskytne v množině T ve více dvojicích a poruší tak podmínu z definice zobrazení.

Takovéto objekty zkoumáme v kapitole 3.

Pro určité množiny studentů by to ale zobrazení být mohlo, takže obecně se nedá říct nic.

△

Jakmile umíme posílat někam prvky, tak už umíme posílat i celé množiny, prostě je pošleme po jednotlivých prvcích. Můžeme si také vzít nějaký objekt v cílové množině a zeptat se, kdo všechno je na něj poslán.

Definice.

Nechť $T: A \mapsto B$ je zobrazení. Pro $M \subseteq A$ definujeme **obraz** M jako

$$T[M] = \{b \in B; \exists a \in M : T(a) = b\} = \{T(a); a \in M\}.$$

Pro $N \subseteq B$ definujeme **vzor** N jako

$$T^{-1}[N] = \{a \in A; T(a) \in N\}.$$

Pak máme třeba $R(T) = T[D(T)]$, evidentně vždy $T^{-1}[B] = A$. Značení T^{-1} pro vzor se může plést se značením pro inverzní zobrazení (viz níže), vzor množiny se pozná podle hranaté závorky. Hledání vzoru ve smyslu množiny se dá udělat vždycky. Vrátíme-li se k příkladu 2b.a, tak $T^{-1}[\{1\}] = T^{-1}[\{1, 2\}] = \{\diamond\}$, $T^{-1}[\{2\}] = \emptyset$.

Kdy se dvě zobrazení rovnají? Není to tak jednoduché, jak to vypadá na první pohled. U funkcí je mnohý čtenář zvyklý, že se rovnají, pokud jsou dány stejným vzorečkem, ale jsou v tom tři háčky. Za prvé, tentýž vzoreček se dá vyjádřit více způsoby a čtenáře možná překvapí, že obecně neexistuje způsob, jak spolehlivě poznat, zda dva vzorečky dívají totéž. Za druhé, my už navíc víme, že na existenci vzorečků nelze spoléhat, takže se spíš musíme zaměřit na to, co zobrazení opravdu dělají, tedy kam prvky posílají. A za třetí, dokonce i kdyby byly dvě zobrazení dány stejnými vzorcí, tak ještě nemusí být stejná, pokud nepracují se stejnými výchozími a cílovými množinami. Brzy totiž uvidíme, že u zobrazení stačí změnit jednu z množin (aniž bychom měnili šipky samotné) a už tím můžeme změnit jeho vlastnosti, vznikne tím tedy vlastně jiné zobrazení. Tím se dostáváme k následující definici.

! Definice.

Nechť $T: A \mapsto B$ a $S: C \mapsto D$ jsou zobrazení. Řekneme, že jsou si rovna, značeno $T = S$, jestliže $A = C$, $B = D$ a platí $\forall a \in A: T(a) = S(a)$.

Jinak řečeno, všechny tři symboly v obrázku „ $T: A \mapsto B$ “ jsou důležité.

My jsme ovšem definovali zobrazení jako určité množiny dvojic. Pokud se k tomuto pohledu vrátíme, tak se celý problém rovnosti poněkud zjednoduší: Zobrazení T, S jsou si rovna právě tehdy, pokud $C = D$ a zobrazení jsou si rovna coby množiny dvojic.

Zobrazení a operace.

Nejjednodušší operací je proces, kdy se omezíme z původní množiny A jen na nějakou její podmnožinu. To je vysoko užitečný nástroj, například v případech, kdy se nám nelibí, co zobrazení na jisté části množiny A dělá, a situace nám umožňuje dotyčnou část ignorovat. Formálně to funguje takto.

Definice.

Nechť $T: A \mapsto B$ je zobrazení, nechť $M \subseteq A$. Definujeme **restrikci** zobrazení T na M , značeno $T|_M$, jako zobrazení z M do B definované

$$T|_M(a) = T(a) \text{ pro } a \in M.$$

Například T z příkladu 2b.a může být omezeno na podmnožinu $M = \{\diamond\}$, vznikne pak zobrazení $T|_M: M \mapsto B$ definované $T(\diamond) = 1$.

Vrátíme-li se k definici zobrazení, tedy nahlížíme-li na něj jako na nějakou množinu dvojic z $A \times B$, pak nás zajímají jen ty dvojice, které mají první souřadnici z M , proto $T|_M = T \cap (M \times B)$. Není to nic zásadního, ale je

dobré umět si takovéto věci rozmyslet. V zásadě ale se zobrazeními jako s množinami pracujeme jen výjimečně, protože ten jazyk pro ně není úplně nevhodnější, jde přeci jen o velice speciální množiny.

U zobrazení se nejvíce pracuje s operací skládání.

! Definice.

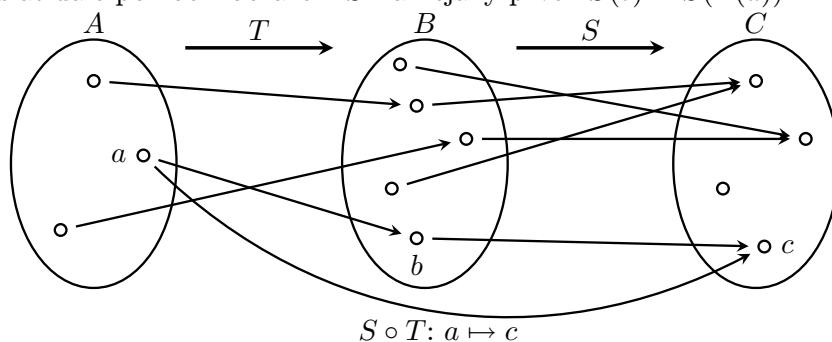
Nechť $T: A \mapsto B$ a $S: B \mapsto C$ jsou zobrazení. Definujeme jejich **složené zobrazení** či **kompozici** $S \circ T: A \mapsto C$ předpisem

$$(S \circ T) : a \mapsto S(T(a)) \text{ pro } a \in A.$$

Značíme také $S \circ T = S(T)$.

Consider mappings $T: A \mapsto B$ and $S: B \mapsto C$. We define their **composition** as the mapping $S \circ T: A \mapsto C$ defined by $(S \circ T)(a) = S(T(a))$ for $a \in A$.

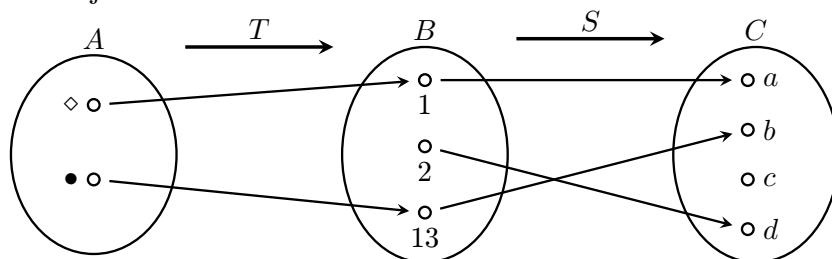
Obrázek naznačuje, oč zde jde. Zobrazení T posílá $a \in A$ na nějaké $b \in B$, ale situace je tak pěkně nastavená, že toto $b = T(a)$ lze poslat dále pomocí zobrazení S na nějaký prvek $S(b) = S(T(a))$.



Když se u každého takového dvoukroku podíváme jen na výchozí a cílový bod (tedy vynecháme prostředníka), dostáváme zobrazení $S \circ T : A \mapsto C$.

Všimněte si, že když chceme $S \circ T$ použít, tak nejdříve aplikujeme na výchozí prvek zobrazení T a na ten výsledek pak teprve S . Jinými slovy, složení $S \circ T$ se čte zprava doleva. To je poněkud nezvyklé, ale kdysi se tak matematici dohodli, protože jim přišlo, že to pěkně ladí s běžným funkčním zápisem. Když aplikujeme T na a , dostaneme $T(a)$. Tohle pak chceme dosadit do S , čili děláme $S(T(a))$, takže T je jako první, ale vpravo.

! Příklad 2b.d: Uvažujme naše známé T z příkladu 2b.a a také zobrazení S z B do $C = \{a, b, c, d\}$ dané $S(1) = a$, $S(2) = d$, $S(13) = b$. Zkusíme je složit.



Vidíme, že opravdu dostáváme zobrazení z A do C , které posílá $\diamond \mapsto a$ a $\bullet \mapsto b$. Ověříme si to první podle definice: $(S \circ T)(\diamond) = S(T(\diamond)) = S(1) = a$.

Je vidět, že $S(2)$ je pro složené zobrazení irrelevantní.

△

Poznámka: Vidíme, že nápad s navazováním funguje vždy, když je možné hodnoty T dosadit do S . Skládání lze tedy zavést obecněji pro libovolná zobrazení T, S splňující $R(T) \subseteq D(S)$. Proč jsme tedy neudělali obecnější definici, která pracuje se zobrazeními $T: A \mapsto B$ a $S: C \mapsto D$ splňujícími $B \subseteq C$? Protože ta naše je jednodušší a bude se nám s ní dále trochu lépe pracovat, přičemž jsme nic neztratili.

Pokud jsme totiž v oné obecnější situaci, tak nás při skládání stejně nezajímá, so S dělá s prvky, do kterých se T nedostane. Jinými slovy, vždy se můžeme omezit na restrikci $S|_B: B \mapsto D$ a pak už můžeme použít naši definici.

Poznámka: zkuseme udělat výjimku a vrátit se ještě k množinovému přístupu k zobrazení. Jestliže bychom chtěli s T a S pracovat podle definice, tedy jako s dvojicemi prvků, jak bude vypadat definice složeného zobrazení? Musíme specifikovat, které dvojice prvků jej tvoří, rozmyslete si, že to dopadne takto:

$$S \circ T = \{(a, c) \in A \times C; \exists b \in B : (a, b) \in T \wedge (b, c) \in S\}.$$

Jako obvykle se nám bude se zobrazeními a skládáním lépe pracovat, pokud budeme umět dělat nějaké zaručeně správné „úpravy“, jinými slovy nás zajímá, jaká pravidla pro tuto operaci platí.

Už z principu je jasné, že skládání nemůže být komutativní, protože například v tom našem příkladě pořadí $T \circ S$ vůbec nemá smysl. Když se podíváme, jak by zobrazení za sebou šla: $S: B \mapsto C, T: A \mapsto B$, tak vidíme, že zobrazení T vůbec není schopno akceptovat výsledky S jako svůj vstup. To je tedy principiální problém.

Jsou ovšem situace, kdy při obrácení pořadí se množiny správně navážou, například pokud používáme stále stejnou množinu. Ani pak ale není komutativita zaručena.

Příklad 2b.e: Uvažujme množinu $B = \{1, 2, 13\}$ a zobrazení $U, V: B \mapsto B$ definovaná takto:

$$U: 1 \mapsto 1, 2 \mapsto 13, 13 \mapsto 1.$$

$$V: 1 \mapsto 2, 2 \mapsto 13, 13 \mapsto 1.$$

Pak zobrazení $V \circ U$ posílá $1 \mapsto V(U(1)) = V(1) = 2$, zatímco $U \circ V$ posílá $1 \mapsto U(V(1)) = U(2) = 13$. Neplatí tedy $V \circ U = U \circ V$.

△

! **Příklad 2b.f:** Vraťme se teď k tomu, co student dobře zná, reálným funkcím. Uvažujme funkce $f(x) = x^2$ a $g(x) = x + 13$, obě jsou vlastně zobrazení $\mathbb{R} \mapsto \mathbb{R}$ a můžeme je tedy složit v libovolném pořadí. Složení $g \circ f$ posílá

$$x \mapsto g(f(x)) = g(x^2) = x^2 + 13,$$

nahradili jsme nejprve $f(x)$ příslušnou hodnotou a pak jsme tento výsledek použili jako vstupní hodnotu pro g . Můžeme začít i vyhodnocením g a dopadne to stejně,

$$x \mapsto g(f(x)) = f(x) + 13 = x^2 + 13.$$

Každopádně $(g \circ f)(x) = x^2 + 13$. Rozmyslete si, že v opačném pořadí skládání dostaneme $(f \circ g)(x) = (x + 13)^2$, značeno také $f(g)(x) = (x + 13)^2$. Zase vidíme, že změnou pořadí skládání dostáváme jinou funkci.

△

Popravdě řečeno, komutativita je sice příjemná, ale až tak zásadní není, takže její selhání tolík nevadí. Mnohem více nám záleží na asociativitě a tam máme štěstí.

Věta 2b.1.

Nechť $T: A \mapsto B, S: B \mapsto C$ a $R: C \mapsto D$ jsou zobrazení. Pak platí $(R \circ S) \circ T = R \circ (S \circ T)$.

Důkaz (rutinní): Nejprve si rozmyslíme, že $(R \circ S) \circ T$ a $R \circ (S \circ T)$ jsou obojí zobrazení z A do D (nakreslete si obrázek), takže se shodují výchozí a cílové množiny. Teď ukážeme, že obě zobrazení dávají stejné hodnoty na prvcích z A .

Vezměme libovolné $a \in A$. Zobrazení $(R \circ S) \circ T$ vzniká jako složení T a $R \circ S$. Podle definice se tedy a nejprve dosazuje do T a výsledný prvek pak do $R \circ S$. Dostáváme $(R \circ S)[T(a)]$ a podle definice skládání si rozmyslíme, jak složené zobrazení $R \circ S$ působí na prvek $T(a)$.

$$a \mapsto (R \circ S)[T(a)] = R(S[T(a)]) = R(S(T(a))).$$

Použili jsme hranaté závorky, abychom vizuálně oddělili úrovně, na kterých se používá definice, ale různé typy závorek jsou samozřejmě pořád jen závorky.

Stejně rozebereme $R \circ (S \circ T)$, podle definice se má R aplikovat na složení $(S \circ T)[a]$, což si pak přepíšeme pomocí definice:

$$a \mapsto R[(S \circ T)(a)] = R[S(T(a))] = R(S(T(a))),$$

tedy hodnoty jsou stejné.

□

Už jsme viděli v kapitole 2a, že asociativita nám umožňuje rozšířit definici operace indukcí ze dvou prvků na libovolný konečný počet, tedy nejprve ze dvou na tři předpisem $R_3 \circ R_2 \circ R_1 = R_3 \circ (R_2 \circ R_1)$, odtud pak na čtyři předpisem $R_4 \circ R_3 \circ R_2 \circ R_1 = R_4 \circ (R_3 \circ R_2 \circ R_1)$ a tak dále. Obecně se to dělá indukcí/rekurzí (viz kapitola 5):

Definice.

Nechť $n \in \mathbb{N}$, $n \geq 2$. Uvažujme množiny $A_1, \dots, A_n, A_{n+1}, A_{n+2}$ a zobrazení $T_i: A_i \mapsto A_{i+1}$ pro $i = 1, \dots, n+1$. Jejich složení definujeme vzorcem

$$T_{n+1} \circ T_n \circ T_{n-1} \circ \dots \circ T_1 = T_{n+1} \circ (T_n \circ T_{n-1} \circ \dots \circ T_1).$$

Je to zase lehké, při pohledu na obrázky výše člověka napadne, že by klidně těch zobrazení mohl za sebe navázat hodně a pak ignorovat vše uprostřed.

Zajímavé to začne být, když se podobné hrátky dělají jen s jedním zobrazením, které zřetězíme. Aby ale $T \circ T$ fungovalo, musí být cílová množina T podmnožinou jeho definičního oboru, nejčastěji jsou rovnou stejné a máme po starostech. Výraz $T \circ T \circ \dots \circ T$ se pak nazývá mocnina, inspirace násobením je evidentní. Uděláme si na to speciální definici.

Definice.

Nechť $T: A \mapsto A$ je zobrazení, $n \in \mathbb{N}_0$. Pak definujeme ***n*-tou mocninu** T značenou T^n takto:

- definujeme $T^1 = T$;
- pro $n \geq 1$ definujeme $T^{n+1} = T \circ T^n$;
- definujeme T^0 jako zobrazení $i_A: A \mapsto A$ definované předpisem $\forall a \in A: i_A(a) = a$.

Tomuto zobrazení říkáme **identita** nebo **identické zobrazení** na A .

Příklad 2b.g: Jak toto funguje u funkcí? Uvažujme $f(x) = x^3$ coby zobrazení $\mathbb{R} \mapsto \mathbb{R}$. Pak $f^0(x) = x$, je to identické zobrazení posílající každé číslo na sebe, snadné je i $f^1(x) = f(x) = x^3$. Dále máme $f^2(x) = f(f(x)) = (x^3)^3 = x^9$, $f^3(x) = f(f^2(x)) = f(x^9) = (x^9)^3 = x^{27}$, $f^4(x) = f(f^3(x)) = (x^{27})^3 = x^{81}$ atd.

Rozmyslete si, že pro $g(x) = \sin(x)$ bude $g^2(x) = \sin(\sin(x))$, $g^3(x) = \sin(\sin(\sin(x)))$ atd.

Zde narázíme na jednu nepříjemnost. U funkcí totiž také máme operace na výchozím prostoru \mathbb{R} (sčítání, násobení) a od nich odvozené operace s funkcemi (funkce vzájemně sčítáme, násobíme atd.). Pak také interpretujeme f^k jako součin $f \cdot f \cdots f$, například $f^2(x) = f(x) \cdot f(x) = x^3 \cdot x^3 = x^6$, $f^3(x) = f(x) \cdot f(x) \cdot f(x) = x^3 \cdot x^3 \cdot x^3 = x^9$ atd., také $f^0(x) = (x^3)^0 = 1$. Podobně máme v tomto smyslu $g^k(x) = \sin^k(x)$. Evidentně dostáváme jiné výsledky než v předchozím odstavci, ale značení je stejně. To je vysoce nepříjemné, ale naštěstí méně, než by se zdálo. V analýze hodně pracujeme s operacemi na reálných číslech a nejsme zvyklí opakovaně skládat funkce se sebou, takže tam f^k automaticky bereme jako opakované násobení. Zde nás naopak při práci se zobrazením $T: A \mapsto B$ vůbec nezajímá, co si množiny A a B dělají, často ani žádné své operace nemají (množina lidí atd.), takže T^k vždy znamená opakované skládání.

Na tento rozpor narazíme v silnější podobě u inverzních funkcí, naštěstí nás v této knize trápit nebude.

△

Příklad 2b.h: Vraťme se k příkladu se zobrazeními U, V na množině $B = \{1, 2, 13\}$. Pak máme následující:

$$U^1 = U: 1 \mapsto 1, 2 \mapsto 13, 13 \mapsto 1.$$

$$U^2 = U \circ U: 1 \mapsto 1 \mapsto 1, 2 \mapsto 13 \mapsto 1, 13 \mapsto 1 \mapsto 1, \text{ tedy } U^2(b) = 1 \text{ pro všechna } b \in B.$$

$U^3 = U \circ U^2: 1 \mapsto 1 \mapsto 1, 2 \mapsto 1 \mapsto 1, 13 \mapsto 1 \mapsto 1$, tedy $U^3 = U^2$. Rozmyslete si, že u tohoto zobrazení jsou všechny další mocniny stejné, posílají všechno z B do 1.

$$V^1 = V: 1 \mapsto 2, 2 \mapsto 13, 13 \mapsto 1.$$

$$V^2 = V \circ V: 1 \mapsto 2 \mapsto 13, 2 \mapsto 13 \mapsto 1, 13 \mapsto 1 \mapsto 2, \text{ tedy } V^2: 1 \mapsto 13, 2 \mapsto 1, 13 \mapsto 2.$$

$V^3 = V \circ V^2: 1 \mapsto 13 \mapsto 1, 2 \mapsto 1 \mapsto 2, 13 \mapsto 2 \mapsto 13$. Takže vlastně $V^3 = i_B$ je identické zobrazení na B . Rozmyslete si, že $V^4 = V$, $V^5 = V^2$, $V^6 = i_B$, $V^7 = V$, $V^8 = V^2$, $V^9 = i_B$ atd.

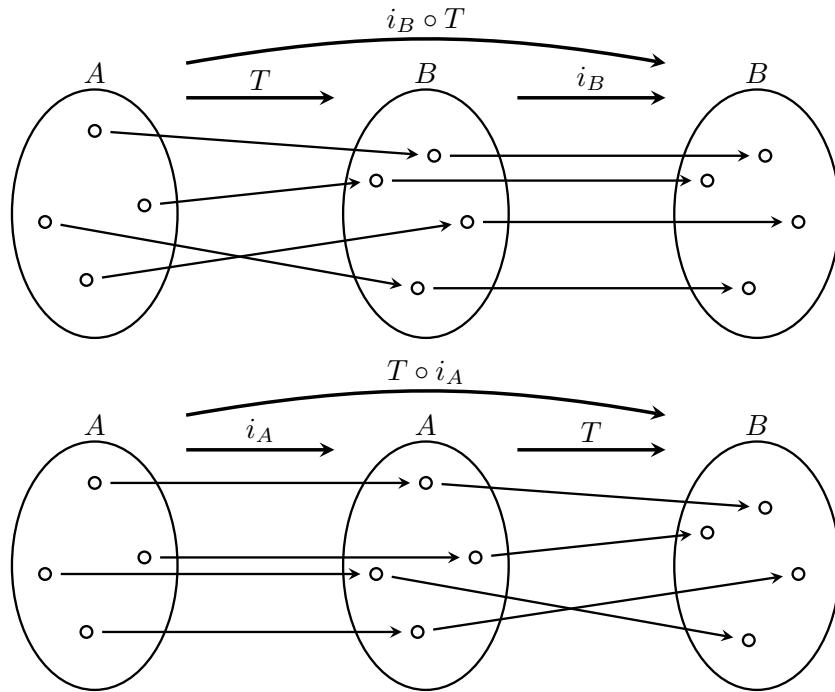
△

Poslední pozorování už plyne obecně z rovnosti $V^3 = i_B$ a následujícího faktu:

Fakt 2b.2.

Nechť $T: A \mapsto B$ je zobrazení. Pak $i_B \circ T = T$ a $T \circ i_A = T$.

Tohle by mělo být jasné, když si člověk představí správný obrázek.



Z obrázku je také jasné, proč je v jednom vzorci potřeba i_A a v druhém i_B .

Důkaz (rutinní): Nejprve dokážeme, že $i_B \circ T = T$. Protože $T: A \mapsto B$ a $i_B: B \mapsto B$, množiny správně navazují a toto skládání má smysl. Ukážeme, že zobrazení $i_B \circ T$ dělá totéž co T . Nechť $a \in A$. Pak

$$(i_B \circ T)(a) = i_B(T(a)) = T(a),$$

protože $T(a)$ je prvek z B a zobrazení i_B takové prvky nechává, jak jsou.

Podobně dokážeme druhou rovnost. □

Velice zajímavé je, když si vezmeme zobrazení $T: A \mapsto A$. Rovnosti pak dají $i_A \circ T = T \circ i_A = T$. To znamená, že zobrazení i_A se chová vůči skládání jako číslo 1 při násobení čísel. Dají se dokonce (relativně snadno pomocí asociativity) dokázat i pravidla $T^m \circ T^n = T^{m+n}$ a $(T^m)^n = T^{mn}$. Ještě se k tomu vrátíme v kapitole o binárních operacích, teď si od násobení čísel vypůjčíme inspiraci, jmenovitě to, že se k číslům x pokoušíme hledat čísla $\frac{1}{x}$ tak, aby $x \cdot \frac{1}{x} = 1$.

! Definice.

Nechť $T: A \mapsto B$ je zobrazení. Řekneme, že zobrazení $S: B \mapsto A$ je **inverzní** k T , jestliže $S \circ T = i_A$ a $T \circ S = i_B$. Pokud takové zobrazení existuje, tak řekneme, že T je **invertibilní**, a inverzní zobrazení značíme T^{-1} .

Let $T: A \mapsto B$ be a mapping. We say that a mapping $S: B \mapsto A$ is an **inverse mapping** of T if it satisfies $S \circ T = i_A$ a $T \circ S = i_B$. If such a mapping exists, then we denote it T^{-1} and say that T is **invertible**.

Co ta definice vlastně požaduje? Máme tam rovnosti dvou zobrazení, což znamená, že se chovají stejně, když do nich dosazujeme prvky. Pro začátek si vždy rozmyslíme, jaké prvky:

Máme $T: A \mapsto B$ a $S: B \mapsto A$, proto $S \circ T$ jde z A do A . Má tedy smysl jej porovnávat s i_A , které jde také z A do A . Porovnání děláme dosazováním prvků z A , takže rovnost $S \circ T = i_A$ ve skutečnosti znamená, že

$$S(T(a)) = a \text{ pro všechna } a \in A. \quad (1)$$

Podobně si rozmyslíme, že $T \circ S$ jde z B do B , a vidíme, že rovnost z definice je ekvivalentní rovnosti

$$T(S(b)) = b \text{ pro všechna } b \in B. \quad (2)$$

Definici je tedy alternativně možno formulovat prostřednictvím podmínek (1) a (2), bez použití zobrazení identity, mnoho autorů to tak dělá.

Lidově řečeno, podmínky (1) a (2) nám říkají, že se zobrazení T a S „navzájem zkrátí“, pokud se ve skládání objeví vedle sebe. Čtenář to už nejspíše viděl, například funkce $\ln(x)$ a e^x jsou navzájem inverzní, tudíž platí $e^{\ln(x)} = x$ pro $x > 0$ a $\ln(e^x) = x$ pro $x \in \mathbb{R}$.

Příklad 2b.i: Vratme se k zobrazení $V: 1 \mapsto 2, 2 \mapsto 13, 13 \mapsto 1$ z příkladu 2b.e.

Tvrdíme, že zobrazení $W: 1 \mapsto 13, 2 \mapsto 1, 13 \mapsto 2$ je inverzní k zobrazení V . Dokážeme to dosazením, přesně jak jsme si to teď rozmysleli.

$W \circ V: 1 \mapsto 2 \mapsto 1, 2 \mapsto 13 \mapsto 2, 13 \mapsto 1 \mapsto 13$. Ano, vidíme, že toto složené zobrazení je identita.

$V \circ W: 1 \mapsto 13 \mapsto 1, 2 \mapsto 1 \mapsto 2, 13 \mapsto 2 \mapsto 13$. A zase máme identitu. Takže $W = V^{-1}$.

△

Příklad 2b.j (pokračování 2b.a): Teď si zase připomeneme zobrazení T z množiny $A = \{\diamond, \bullet\}$ do $B = \{1, 2, 13\}$ definované předpisem $T(\diamond) = 1, T(\bullet) = 13$. Definujme $\widehat{T}: B \mapsto A$ předpisem $\widehat{T}(1) = \diamond, \widehat{T}(2) = \bullet, \widehat{T}(13) = \bullet$.

Pak $\widehat{T} \circ T$ jde z A do A a dělá $\widehat{T}(T(\diamond)) = \widehat{T}(1) = \diamond$ a $\widehat{T}(T(\bullet)) = \widehat{T}(13) = \bullet$, tedy $\widehat{T} \circ T = i_A$.

To vypadá nadějně. Bohužel ale $T(\widehat{T}(2)) = T(\bullet) = 13$ a jsme v háji, zobrazení $T \circ \widehat{T}$ neposílá $2 \mapsto 2$ a tím pádem to není i_B , proto také \widehat{T} není inverzní zobrazení k T .

△

Vidíme, že v definici opravdu potřebujeme mít obě rovnosti.

Příklad 2b.k: Uvažujme funkci $f(x) = 2x + 1$. Standardní algoritmus na hledání inverzní funkce funguje tak, že rovnost $y = 2x + 1$ vyřešíme pro x , dostáváme tak vzorec $g(y) = \frac{1}{2}(y - 1)$. Dokážeme, že jsme opravdu dostali inverzní funkci:

$$\begin{aligned} x \in \mathbb{R} &\implies g(f(x)) = g(2x + 1) = \frac{1}{2}([2x + 1] - 1) = x, \\ y \in \mathbb{R} &\implies f(g(y)) = f\left(\frac{1}{2}(y - 1)\right) = 2 \cdot \frac{1}{2}(y - 1) + 1 = y. \end{aligned}$$

Potvrzeno, $g \circ f = I_{\mathbb{R}}$ a $f \circ g = I_{\mathbb{R}}$, tedy g je inverzní zobrazení k f .

Jak bychom to zapsali? Zase máme problém, z hlediska teorie zobrazení píšeme $g = f^{-1}$, ale u reálných funkcí f^{-1} znamená $\frac{1}{f}$, což je $x \mapsto \frac{1}{2x+1}$ neboli úplně jiná funkce než g . Některí autoři proto používají značení f_{-1} pro inverzní funkci.

Poznamenejme ještě jednu věc, na mnohých středních školách se studenti učí přejít u inverzní funkce zase k proměnné x , tedy psali by $g(x) = \frac{1}{2}(x - 1)$. To ale není moc dobrý nápad, jednak to není třeba a druhak to dokonce posílá špatný vzkaz. My totiž pracujeme s dvěma kopiemi množiny reálných čísel, v jedné používáme pro prvky x a v druhé y .

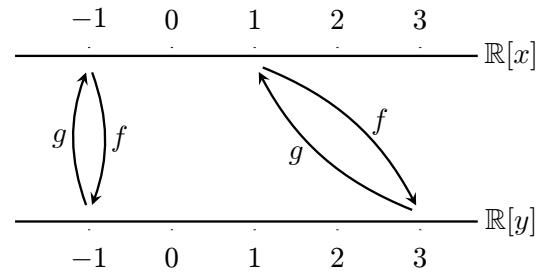
Jak vidíme, funkce g vůbec neumí s prvky x pracovat, protože má jako výchozí množinu úplně jiný svět. Nemá tedy smysl jí tyto proměnné podsouvat, naopak dosazováním y čtenáři jasné sdělujeme, jak tato funkce funguje.

△

Teď se zamyslíme, jak vlastně takové inverzní zobrazení funguje, zároveň tím vyřešíme jeden problém, který se objevil v definici. Tam jsme si pro inverzní zobrazení zavedli značení T^{-1} , ale co když je jich více? Některí autoři proto toto značení zavedou až později, když je jasné, jaká je situace.

Jaká tedy je? Obrázek výše silně napovídá, podívejme se na to obecně. Prvek $a \in A$ je zobrazením T někam poslan, jmenovitě na prvek $T(a) = b$. Aby pro zobrazení S platila rovnost $S(T(a)) = a$, tak S musí vrátit b zpět na a .

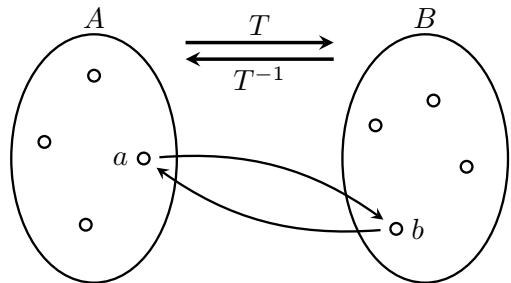
Zdá se, že T^{-1} má přesně stejné šipky jako T , jen jdou opačným směrem, což mimochodem souhlasí s předchozími třemi příklady. Také to ukazuje, že inverzní zobrazení nemá na výběr, jak jít.



Potvrďme si to oficiálně.

Fakt 2b.3.

Nechť $T: A \mapsto B$ je invertibilní zobrazení. Pak $T^{-1}(b) = a$ právě tehdy, když $T(a) = b$.



Důkaz: Je to ekvivalence, musíme dokázat implikace oběma směry.

1) \implies : Předpokládejme, že $T^{-1}(b) = a$. Je to tedy prvek z A , proto na něj můžeme aplikovat zobrazení T : $T(a) = T(T^{-1}(b))$. Jenže napravo máme $T \circ T^{-1} = i_B$, proto dostáváme $T(a) = b$.

2) \impliedby : Předpokládejme, že $T(a) = b$. Toto je prvek z B , můžeme na něj aplikovat T^{-1} : $T^{-1}(b) = T^{-1}(T(a))$. Jenže napravo máme $T^{-1} \circ T = i_A$, proto $T^{-1}(b) = a$.

□

Důsledek 2b.4.

Nechť $T: A \mapsto B$ je zobrazení. Jestliže je invertibilní, tak je jeho inverzní zobrazení T^{-1} dáno jednoznačně.

Teď si všichni matematici oddechli, definice inverzního zobrazení byla korektní.

Coby cvičení matematické představivosti se ještě jednou podíváme na zobrazení jako na množinu dvojic. Právě jsme zjistili, že jestliže je nějaké zobrazení $T \subseteq A \times B$ invertibilní, pak

$$T^{-1} = \{(b, a) \in B \times A; (a, b) \in T\}.$$

Z představy otáčení šipek lze snadno odvodit základní pozorování o inverzních zobrazeních. Začneme něčím jednoduchým. Jestliže je T invertibilní, tak umíme otočit šipky a dostaneme tím nové zobrazení. Nic by nám pak nemělo bránit v novém otočení šipek a dostaneme zase zpět to původní zobrazení. Teď to řekneme matematicky.

Fakt 2b.5.

Nechť $T: A \mapsto B$ je zobrazení. Jestliže je T invertibilní, tak je i T^{-1} invertibilní a $(T^{-1})^{-1} = T$.

Důkaz (poučný): Předpokládejme, že T je invertibilní, takže máme $T^{-1}: B \mapsto A$. Potřebujeme ukázat, že je nějaké zobrazení S , které jde naopak než T^{-1} , tedy $A \mapsto B$, a splňuje $T^{-1} \circ S = i_A$ a $S \circ T^{-1} = i_B$. Zobrazení T opravdu jde $A \mapsto B$ a když jej dosadíme do těch rovností místo S , tak dostaneme $T \circ T^{-1} = i_A$ a $T^{-1} \circ T = i_B$, což určitě platí, protože je T^{-1} je inverzní k T . T tedy splňuje požadavky na S , je to $(T^{-1})^{-1}$. \square

Umíme dělat inverzi a také umíme skládat, jak to jde dohromady?

Věta 2b.6.

Nechť $T: A \mapsto B$ a $S: B \mapsto C$ jsou zobrazení. Jestliže jsou invertibilní, tak je i $S \circ T$ invertibilní a navíc platí $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$.

Ten vzorec je opravdu zajímavý, protože obrací pořadí. To je u inverzních pojmu normální (viz Věta 8a.9), podívejme se, že to ani jinak nejde. Daná zobrazení jdou $A \xrightarrow{T} B \xrightarrow{S} C$ a když složíme, dostaneme $S \circ T: A \mapsto C$. Případné inverzní zobrazení k němu tedy musí jít $C \mapsto A$. Abychom vyšli z C , musíme začít s S^{-1} , protože T^{-1} začíná v množině B . Tím jsme inspirováni k opačnému pořadí a ověříme, že to opravdu dopadne dle očekávání: Máme „řetízek“ $C \xrightarrow{S^{-1}} B \xrightarrow{T^{-1}} A$, čili $T^{-1} \circ S^{-1}$ jde $C \mapsto A$, přesně jak potřebujeme. Ukázali jsme, že z hlediska množin má všechno ten správný smysl, ale to na rovnost $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$ nestačí. Ještě se musíme podívat, kam se posílají jednotlivé prvky.

Důkaz (rutinní, poučný): Předpokládejme, že T a S jsou invertibilní. Potřebujeme dokázat, že existuje zobrazení inverzní k $S \circ T$, tedy zobrazení $U: C \mapsto A$ splňující $U \circ (S \circ T) = i_A$ a $(S \circ T) \circ U = i_C$. Takže jedno takové najdeme, ukážeme, že $U = T^{-1} \circ S^{-1}$ funguje. Už jsme si rozmysleli, že jde $C \mapsto A$, a opakovánou aplikací asociativního zákona, Faktu 2b.2 a vlastnosti inverzní funkce dostaneme

$$(T^{-1} \circ S^{-1}) \circ (S \circ T) = T^{-1} \circ (S^{-1} \circ S) \circ T = T^{-1} \circ i_B \circ T = T^{-1} \circ (i_B \circ T) = T^{-1} \circ T = i_A,$$

$$(S \circ T) \circ (T^{-1} \circ S^{-1}) = S \circ (T \circ T^{-1}) \circ S^{-1} = S \circ i_B \circ S^{-1} = S \circ (i_B \circ S^{-1}) = S \circ S^{-1} = i_C.$$

\square

Toto tvrzení snadno zobecníme pro vícenásobné skládání.

Věta 2b.7.

(i) Nechť $n \in \mathbb{N}$, $n \geq 2$. Uvažujme množiny A_1, \dots, A_n, A_{n+1} a zobrazení $T_i: A_i \mapsto A_{i+1}$ pro $i = 1, \dots, n$. Jestliže jsou všechna tato zobrazení invertibilní, pak je invertibilní i složené zobrazení $T_n \circ \dots \circ T_1$ a

$$(T_n \circ \dots \circ T_1)^{-1} = T_1^{-1} \circ \dots \circ T_n^{-1}.$$

(ii) Nechť je $T: A \mapsto A$ invertibilní, $n \in \mathbb{N}_0$. Pak je i T^n invertibilní a $(T^n)^{-1} = (T^{-1})^n$.

Důkaz se dělá indukcí, necháme jej do příslušné kapitoly jako cvičení (viz cvičení 5a.8), protože je rutinní. Mimochodem, ten druhý vztah vlastně známe z reálných čísel, $\frac{1}{x^n} = (\frac{1}{x})^n$.

Poznámka (pokročilá): Pro invertibilní zobrazení je možné definovat mocninu i pro záporné exponenty vzorcem $T^{-n} = (T^{-1})^n$. Máme pak pro invertibilní zobrazení T definováno T^n pro všechna $n \in \mathbb{Z}$, podobně

jako máme pro nenulová čísla x definováno x^n pro všechna $n \in \mathbb{Z}$. Dá se ukázat (není to těžké, ale dlouhé a nudné), že jsme tím rozšířením na záporné exponenty nepokazili pěkné vlastnosti původní mocniny, například pořád platí $T^m \circ T^n = T^{m+n}$ a $(T^m)^n = T^{mn}$, tentokrát ovšem pro libovolná celá čísla m, n .

Ted' se vraťme k tomu, že inverzní zobrazení prostě jen obrací šipky. Tím se ovšem dostáváme k problému, že ne všechna zobrazení jsou invertibilní (ostatně ani $\frac{1}{x}$ nenajdeme pro všechna čísla x). Vidíme dva zádrhele, které by nás mohly při pokusu o obrácení šipek potkat. Pokud by se dvě šipky zobrazení T sbíhaly v jednom bodě, pak nevíme, kterou z nich si vybrat pro cestu zpět. A kdyby u nějakého prvku z B šipka chyběla, pak zase nevíme, kam jej zpětně poslat (přesně toto nás postihlo v příkladu 2b.j). Zavedeme si vlastnosti, které přesně toto popíšou.

! Definice.

Nechť $T: A \rightarrow B$ je zobrazení.

Řekneme, že T je **prosté** či **injektivní**, jestliže

$$\forall x, y \in A: [x \neq y \implies T(x) \neq T(y)].$$

Řekneme, že T je **na** či **surjektivní**, jestliže $R(T) = B$.

Řekneme, že T je **vzájemně jednoznačné** či **bijekce**, jestliže je prosté a na.

A mapping $T: A \rightarrow B$ is called **one-to-one** (often denoted **1-1**) or **injective**, if for all distinct elements $x \neq y \in A$ one has $T(x) \neq T(y)$. It is called **onto** or **surjective** if $R(T) = B$. If it is both 1-1 and onto, then we call it a **bijection**.

Podmínka $R(T) = B$ se dá také napsat $T[A] = B$ nebo podrobněji

$$\forall b \in B \exists a \in A: T(a) = b$$

a znamená, že ke každému prvku v cílové množině B vede alespoň jedna šipka. Každé zobrazení identita i_A je automaticky na, z ostatních příkladů v této kapitole jsou na jen V , $f(x) = 2x + 1$ a jeho inverze g .

Podmínka pro prostotu zase říká, že se žádné šipky zčínající v různých místech nemohou sejít v jednom bodě. Když si projdeme naše příklady, tak těch prostých je docela dost: T , S , V , $f(x) = 2x + 1$ a jeho inverze a automaticky každá i_A . Příklad U není prostý, protože prvky $x = 1$ a $x = 13$ splňují předpoklad implikace z definice ($x \neq y$), ale nesplňují její závěr ($U(x) = 1 = U(13)$), tudíž je implikace nepravdivá a U tedy není prosté.

Protože pracovat se vztahem \neq je nepříjemné, mnoho autorů (možná i většina) raději používá v definici obměnu té původní implikace, podmínka prostoty se dá také napsat

$$\forall x, y \in A: [T(x) = T(y) \implies x = y]. \quad (\text{I})$$

V praxi prostotu daného zobrazení T zkoumáme tak, že začneme s rovností $T(x) = T(y)$ a řešíme to jako rovnici, což je mnohem pohodlnější. Většinou to tak budeme dělat i zde.

Co se týče bijekce, mezi příklady máme bijekci V , funkci $f(x) = 2x + 1$ a její inverzi a pro libovolnou množinu A je identita i_A samozřejmě také bijekce. Dá se říct, že bijekce jsou z pohledu teorie množin nejlepší zobrazení.

Zde jak vidno používáme kratšího slova „bijekce“, které postupně začíná být bráno na milost, název „vzájemně jednoznačné“ je tradičnější a pěknější, nicméně delší.

Před chvílí jsme si rozmysleli, že při obrácení šipek nám vadí situace, když se u T šipky sbíhají nebo nedojdou všude. To první nám zakáže prostota, to druhé surjektivita, takže následující tvrzení by nemělo překvapit.

! Věta 2b.8.

Nechť $T: A \rightarrow B$ je zobrazení. Je invertibilní právě tehdy, když je to bijekce.

Důkaz (poučný): 1) \implies : Předpokládejme, že T je invertibilní.

Nejprve ukážeme, že T je prosté, pomocí obměny (I). Vezměme prevky $x, y \in A$ takové, že $T(x) = T(y)$. Dosadíme do zobrazení T^{-1} (máme ho k dispozici, T je invertibilní), stejný vstup musí dát stejný výsledek, proto dostaneme $T^{-1}(T(x)) = T^{-1}(T(y))$ a tedy $x = y$. Prostota je dokázána.

Ted' ukážeme, že je na. Nechť $b \in B$. Potřebujeme najít nějaký jeho vzor. Definujme $a = T^{-1}(b)$. Pak $a \in A$ a $T(a) = T(T^{-1}(b)) = b$. Surjektivita je dokázána.

2) \impliedby : Předpokládejme, že T je bijekce. Ukážeme, že je invertibilní. Nejprve definujeme zobrazení $S: B \rightarrow A$. Nechť $b \in B$. Protože T je na, tak určitě existuje nějaké a takové, že $T(a) = b$, a protože je T prosté, tak je to jediný takový prvek. Můžeme tedy definovat $S(b) = a$.

Dokážeme, že $S = T^{-1}$.

Nechť $b \in B$. Pak $S(b)$ je prvek a splňující $T(a) = b$, proto $T(S(b)) = T(a) = b$.

Nechť $a \in A$. Potřebujeme vědět, co je $S(T(a))$. Hodnota S v bodě $b = T(a)$ je definovaná jako nějaký prvek $x \in A$ takový, že $T(x) = b$. Jedním z takových prvků je a , ten to určitě splní, a díky prostotě T je také jediný takový. Proto jsme při definici S použili $S(b) = a$, tedy $S(T(a)) = a$. Důkaz je hotov. \square

Příklad 2b.l: Uvažujme zobrazení T z množiny občanů ČR do množiny desetimístných čísel definované tak, že $T(x)$ je dáné jako rodné číslo člověka x . Určitě není na, například není možné se narodit ve třináctém měsíci, tudíž desetimístná čísla začínající xx13 nebudou dosažitelná pomocí T . Dobrá otázka je, zda je toto zobrazení prosté. Skutečnost je taková, že my chceme, aby bylo prosté, dlouho jsme si to i mysleli, ale v devadesátých letech se ukázalo, že občas někdo někde něco spletl a toto zobrazení prosté nebylo. Úřady se to snažily napravit, ale kdo ví.
 \triangle

Příklad 2b.m: Prozkoumáme prostotu a surjektivitu pro několik funkcí coby zobrazení $\mathbb{R} \mapsto \mathbb{R}$.

a) $f(x) = x^3 - x$. Není problém si načrtnout graf této funkce, začíná v levém dolním rohu (utíká do mínu nekonečna), při své cestě nahoru protne osu x v bodě -1 , pak se otočí a zase jede dolů, protne osu v počátku, pak se zase otočí nahoru a uteče do nekonečna, protínaje osu x v bodě 1 . Vidíme, že tato funkce dokáže nabýt libovolné reálné hodnoty, je tedy na. Zároveň také vidíme, že není prostá, protože například $f(0) = 0$ a také $f(1) = 0$, takže se nám sešly šipky $0 \mapsto 0$ a $1 \mapsto 0$.

b) $f(x) = 2x - 1$. Tato funkce je prostá. Důkaz: použijeme alternativní podmítku (I).

Vezměme tedy libovolné $x, y \in \mathbb{R}$ takové, že $f(x) = f(y)$. Pak $2x - 1 = 2y - 1$, odsud hravě dostaneme $x = y$ a důkaz je hotov.

Tato funkce je také na. Důkaz: Nechť y je nějaký prvek z cílové množiny \mathbb{R} . Tvrdíme, že existuje jisté x_0 splňující $f(x_0) = y$. Toto tvrzení dokážeme tak, že takové x_0 najdeme. Chceme, aby $f(x_0) = y$, tedy aby $2x_0 - 1 = y$. Odtud $x_0 = \frac{y+1}{2}$. To je určitě reálné číslo, ještě potvrďme, že splňuje požadavek:

$$f(x_0) = 2x_0 - 1 = 2\frac{y+1}{2} - 1 = y.$$

Pro dané y jsme tedy našli $x_0 \in \mathbb{R}$ takové, že $f(x_0) = y$, tudíž je f na.

Toto f je proto bijekce. Všimněte si, že jsme právě ukázali, že pro libovolné $y \in \mathbb{R}$ najdeme $x = \frac{1}{2}(y+1)$ tak, aby $f(x) = y$. Našli jsme tedy vzorec obracející šipky neboli vzorec pro f^{-1} (viz příklad výše).

c) $f(x) = \operatorname{arctg}(x)$. Znalosti z analýzy ukazují, že tato funkce je prostá, ale není na.

d) $f(x) = x^2 + 1$. Grafem je klasická parabola obrácená nahoru a posunutá nahoru o 1, tato funkce tedy rozhodně není na a není ani prostá.

Důkaz, že není na: Protože vždy platí $x^2 \geq 0$, je i $f(x) \geq 1$. Nelze tedy nalézt $x \in \mathbb{R}$ takové, aby $f(x) = -13$.

Důkaz, že není prostá: Protipříkladem je třeba $f(-1) = 2 = f(1)$.

Poznámka: Co kdybychom prostotu zkoumali tradičním způsobem, tedy testováním podmíny (I)? Vyšli bychom z rovnosti $f(x) = f(y)$ neboli $x^2 + 1 = y^2 + 1$, odtud pak $x^2 = y^2$. Z tohoto ale neumíme přejít k $x = y$, což ještě nemusí nic znamenat (třeba jen nejsme dost šikovní), ale je to znamení, že máme zpozornět. Pokud si dále všimneme, že z $x^2 = y^2$ vyplývá $y = \pm x$, tak nás to navede k protipříkladu k prostotě.

\triangle

S 2b.9 Jak na vlastnosti funkcí

Tento příklad ukazuje nejčastější způsob zkoumání prostoty a surjektivity. Dostane-li student k prozkoumání nějaké zobrazení T , tak se není třeba bát toho, že je třeba na první pohled komplikované, stačí se držet definice (či její obměny):

1. Chceme-li určit, zda je T **prosté**, tak si vezmeme dva libovolné prvky $x, y \in A$ (tedy obecné prvky, nemůžeme si vybrat dva pěkné konkrétní) a napíšeme si rovnici $T(x) = T(y)$. Dosadíme z definice zobrazení T do obou stran a dostaneme rovnici, ze které se pokusíme odvodit informaci o vztahu x a y . Tento obecný začátek většinou silně napoví, jak dál.

Je-li například $T: \mathbb{N}^3 \mapsto \mathbb{N}^2$ dáné $T(r, s, t) = (r^3, s^t)$, pak prvky $x, y \in A$ jsou vlastně oba tříložkové vektory, tedy třeba $x = (r, s, t)$ a $y = (u, v, w)$ pro nějaké neznámé $r, s, t, u, v, w \in \mathbb{N}$. Základní rovnice pak dává

$$T(r, s, t) = T(u, v, w) \implies (r^3, s^t) = (u^3, v^w)$$

a je třeba se rozmyslet, co dál. Rovnost vektorů znamená rovnost souřadnic, máme tedy $r^3 = u^3$ a $s^t = v^w$. Dá se z toho něco odvodit? Třetí mocnina je prostá funkce, proto z první rovnice vyjde $r = u$. U druhé rovnice ale nic tak očividného není, takže v takovém případě je dobré začít experimenovat, zkoušet různá čísla a vzpmínat na předchozí zkušenosti. Zde se rychle ukáže, že dvojic dávajících stejnou mocninu může být více, třeba $3^4 = 9^2$. To

ukazuje, že dané zobrazení nebude prosté, a máme i protipříklad na prostotu: $T(1, 3, 4) = (1, 81) = T(1, 9, 2)$, ale neplatí $(1, 3, 4) = (1, 9, 2)$. T tedy není prosté.

Někdy ovšem z rovnice $T(x) = T(y)$ dokážeme odvodit $x = y$ (což může klidně znamenat rovnost vektorů neboli rovnost složek), pak bude zobrazení prosté.

2. Chceme-li určit, zda je T **na**, tak si vezmeme libovolný prvek $y \in B$ (z cílového prostoru) a zkusíme k němu najít $x \in A$ takové, aby $T(x) = y$. Pokud takto začneme a pak dosadíme konkrétní T , navede nás to obvykle na správnou cestu. Rovnost $T(x) = y$ je v zásadě rovnice, kterou se snažíme vyřešit pro x , což je trochu komplikováno tím, že vlastně y neznáme, potřebujeme to udělat obecně, pro všechna y . Pokud se to povede, pak je zobrazení na.

U našeho příkladu vybíráme y z cílového prostoru \mathbb{N}^2 , je to tedy nějaký vektor, třeba $y = (u, v)$, přičemž u, v neznáme, jde o nějaká libovolná čísla z \mathbb{N} . Ptáme se, zda existuje $x \in A$ neboli zda existují $r, s, t \in \mathbb{N}$ tak, aby $T(r, s, t) = (u, v)$. Tento obecný začátek nám dává rovnice $r^3 = u$, $s^t = v$ a my se ptáme, zda jsou řešitelné pro r, s, t , přesněji řečeno, zda vždy dokážeme najít r, s, t tak, aby to fungovalo (problém nemusí mít jediné řešení, to nás ale nezajímá).

U rovnice $s^t = v$ si všimneme, že řešení určitě má bez ohledu na volbu v , stačí prostě vzít $s = v$ a $t = 1$. To je dobrý začátek, pomocí T a vektoru z A dokážeme dostat do libovolné druhé souřadnice. Teď se podíváme na tu první: Existuje určitě nějaké $r \in \mathbb{N}$ takové, aby $r^3 = u$? Protože u je libovolné, zkušený student hned tuší, že je zle, protože například třetí odmocnina z 2 existuje, ale není to celé číslo. Takže nenajdeme $r \in \mathbb{N}$ takové, aby $r^3 = 2$, tím pádem ani nelze najít $(r, s, t) \in \mathbb{N}^3$ tak, aby $T(r, s, t) = (2, v)$. Zobrazení T proto není na.

Tyto postupy fungují spolehlivě (viz první cvičení v této kapitole), nicméně jsou situace, kdy je lepší hledat alternativu. U prostoty je někdy lépe vidět přímo vlastnost z definice, tedy doloží se, že když $x \neq y$, pak určitě $T(x) \neq T(y)$, ale to je vzácné.

Někdy se dá také prostota či surjektivita získat mnohem snadněji nepřímo, pomocí vlastností již prozkoumaného zobrazení, od kterého nějakým trikem přejdeme k tomu, které zkoumáme. Dobrou ukázkou jsou poslední cvičení této kapitoly.

V mnoha příkladech (zejména při práci s funkcemi) je lepší zkoumat prostotu pomocí metod matematické analýzy, ale to je jiná pohádka.

△

Vraťme se k teorii, začneme zkoumat chování nových pojmů. Jak si naše tři vlastnosti rozumí se skládáním?

Fakt 2b.10.

Nechť $T: A \mapsto B$ a $S: B \mapsto C$ jsou zobrazení. Pak platí:

- (i) Jestliže jsou T a S prosté, tak je $S \circ T$ prosté.
- (ii) Jestliže jsou T a S na, tak je $S \circ T$ na.
- (iii) Jestliže jsou T a S bijekce, tak je $S \circ T$ bijekce.

Důkaz (poučný): (i): Prostotu dokážem pomocí obměny (I) aplikované na $S \circ T$.

Nechť $x, y \in A$ splňují $(S \circ T)(x) = (S \circ T)(y)$. To se dá napsat jako $S[T(x)] = S[T(y)]$, je to tedy S aplikované na nějaké dva body. Protože je S prosté, tak odtud nutně $T(x) = T(y)$. A protože je T prosté, tak $x = y$. Prostota je dokázána.

(ii): Dokážeme podle definice, že $S \circ T$ je na. Nechť $c \in C$. Protože je S na, musí existovat $b \in B$ takové, že $S(b) = c$. Protože T je na, musí existovat $a \in A$ takové, že $T(a) = b$. Našli jsme a takové, že $(S \circ T)(a) = S(T(a)) = S(b) = c$.

(iii): Jestliže jsou T a S bijekce, tak jsou prosté, a tudíž podle (i) je i $S \circ T$ prosté.

Jestliže jsou T a S bijekce, tak jsou na, a tudíž podle (ii) je i $S \circ T$ na. Takže $S \circ T$ je bijekce.

(Všimněte si, jak jsme pěkně využili již udělané práce. To je pro matematiku typické.)

Alternativa: Jestliže jsou T a S bijekce, tak jsou dle Věty 2b.8 invertibilní, tudíž dle Věty 2b.6 je i $S \circ T$ invertibilní, tudíž bijekce.

□

Stručně řečeno, skládání nepokazí dobré vlastnosti (bude se nám to hodit v příští kapitole). Jako obvykle se tento výsledek dá zobecnit na skládání více zobrazení. Může skládání vylepšit špatné vlastnosti? Někdy ano, někdy ne, podívejte se na cvičení 2b.10. Tuto otázku lze ekvivalentně položit i jinak: Víme-li, že $S \circ T$ má nějakou vlastnost, musí ji mít nutně i složky S a T ? Cvičení odpoví.

Fakt 2b.11.

Jestliže je zobrazení T bijekce, tak T^{-1} existuje a je to také bijekce.

Toto okamžitě plyne z Věty 2b.8 a Faktu 2b.5.

Rozeberme si trochu situaci. Když jsme si hráli s našimi příklady, tak nás mohlo napadnout, že S z příkladu 2b.d nemůže být na už z principu, protože má jen tři šipky (B má jen tři prvky), ale cílová množina má 4 prvky, tudíž je nelze všechny pokrýt.

Podobně se dá rozmyslet, že když posíláme šipky a cílový prostor má méně prvků, než je šipek, tak se musí nějaké šipky potkat a je po prostotě. Shrňeme si to oficiálně.

Fakt 2b.12.

Nechť $T: A \rightarrow B$ je zobrazení a A, B mají konečně mnoho prvků.

- (i) Jestliže má B více prvků než A , pak T nemůže být na.
- (ii) Jestliže má A více prvků než B , pak T nemůže být prosté.
- (iii) Jestliže A a B nemají stejně prvků, pak T nemůže být bijekce.

Dokazovat to nebudeme, protože bychom museli hlouběji do teorie množin (například jsme zatím ani nedefinovali, co je to počet prvků množiny). Naštěstí tento fakt nebudeme v dalších důkazech používat, takže vynecháním jeho důkazu nevznikne díra v základech toho, co tu v dalších kapitolách vystavíme.

Všimněte si, že všechna tato tvrzení jsou zjevně jen implikace. Když se například v (i) podíváme na situaci, kdy je u nějakého zobrazení splněn závěr implikace (T není na), tak nelze s jistotou tvrdit, že počty prvků množin splňují předpoklad. Klidně se totiž mohlo stát, že množiny A, B vyšly tomu zobrazení vstří a B má nejvýše tolik prvků jako A , ale dotyčné zobrazení svou šanci nevyužilo a vyplývalo šipky tím, že jich spoustu poslalo do jednoho prvku.

U tvrzení (iii) je zajímavá i obměna:

(iii)* Jestliže je T bijekce, tak mají A a B stejný počet prvků.

Toto bude výchozí bod pro další kapitolu, stejně jako obměna tvrzení (ii).

I (iii)* je obecně jen implikace, tedy z rovnosti počtu prvků dvou množin nelze automaticky prohlásit všechna zobrazení mezi nim za bijekce, nicméně něco zajímavého se o této situaci říct dá. Uvažujme tedy dvě množiny A, B se shodným (konečným) počtem prvků (nakreslete si obrázek). Co se může stát nějakému zobrazení $T: A \rightarrow B$? Pokud T není prosté, tak se nějaké šipky spojí, pak ale chybí v cílové oblasti a nedojde k jejímu pokrytí, T nebude na. Pokud naopak začneme předpokladem, že T není na, tak vlastně T leze do menší množiny a nemůže být prosté.

Fakt 2b.13.

Nechť $T: A \rightarrow B$ je zobrazení, předpokládejme, že A a B mají stejný konečný počet prvků. Pak je T prosté právě tehdy, když je T na.

Toto je někdy velice užitečné, protože nám to ušetří polovinu práce s dokazováním bijekce.

! Existují situace, kdy máme zobrazení a potřebujeme pracovat s opačnými šipkami, ale nejde to, protože T není prosté. Někdy se dá tento problém obejít tak, že „vyhodíme“ prvky, které nám prostotu kazí, neboli omezíme se na množinu takovou, že už je na ní T prosté.

Je to vlastně něco, co čtenář patrně zná ze střední školy. Funkce (zobrazení) $f: \mathbb{R} \rightarrow \mathbb{R}$ definovaná jako $f(x) = x^2$ není prostá ani náhodou, třeba $f(-13) = 169 = f(13)$ je protipříklad, ale rádi bychom používali opačné šipky (odmocňovali). To se vyřeší tak, že se namísto f uvažuje její restrikce na množinu $\{x \in \mathbb{R}; x \geq 0\}$. Na této množině je už f prostá a veselé inverzníme.

Při definici rovnosti zobrazení jsme zmínili, že změnami množin se mohou podstatně změnit vlastnosti zobrazení. Právě jsme viděli, jak se dá „vyrobit“ prostota tím, že z definičního oboru odstraníme zlobivce. S ještě menším úsilím „vyrobíme“ surjektivitu. Mějme nějaké zobrazení $T: A \rightarrow B$, které není na. To znamená, že v B jsou prvky, do kterých se T nedostane. Jenže ono se dostane do všech prvků z $R(T)$, tak je to ostatně definováno, tudíž když se na T podíváme jako na zobrazení z A do $R(T)$, tak už je na. Takže $T: A \rightarrow B$ a $T: A \rightarrow R(T)$ nemohou být z hlediska teorie stejná zobrazení, protože mají jiné vlastnosti, i když jsme vlastně T samotné vůbec nemodifikovali, pořád posílá stejný prvky stejným způsobem.

Tento trik je v některých situacích velice užitečný. Vyplývá z něj totiž následující:

- Jestliže je $T: A \rightarrow B$ prosté, pak je $T: A \rightarrow R(T)$ bijekce, takže například máme i její inverzi $T^{-1}: R(T) \rightarrow A$.

Teď si ukážeme dvě funkce, které jsou v computer science docela důležité.

Definice.

Definujme následující funkce na \mathbb{R} :

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z}; n \leq x\}; \quad (\text{zaokrouhlení dolů})$$

$$\lceil x \rceil = \min\{n \in \mathbb{Z}; n \geq x\}. \quad (\text{zaokrouhlení nahoru})$$

Význam je doufejme zjevný. Například $\lfloor x \rfloor$ je největší celé číslo, které se najde „pod“ x . Pokud je tedy x celé, tak tím největším celým číslem ne větším než x je samozřejmě právě to x . Kdyby ale x celé nebylo, tak se při procházení celými čísly směrem nahoru zarazíme dřív, než k x dojdeme, jmenovitě u kladných čísel se zarazíme přesně u toho, co vidíme před desetinnou čárkou. U záporných čísel je to drobet jiné, protože na záporné části osy pořád přicházíme k x s celými čísly zleva, od menších, čtenář si teď zkusi namalovat reálnou osu a rozmyslet si, co se pak děje. Při jednom si také rozmyslí, jak funguje $\lceil x \rceil$.

Takže například $\lfloor 13 \rfloor = 13$ a $\lceil 13 \rceil = 13$, $\lfloor -13 \rfloor = -13$ a $\lceil -13 \rceil = -13$, $\lfloor 13.23 \rfloor = 13$ a $\lceil 13.23 \rceil = 14$, ale také $\lfloor -13.23 \rfloor = -14$ a $\lceil -13.23 \rceil = -13$. Opravdu to tedy zaokrouhuje dolů, tedy k menším číslům, a nahoru, tedy k větší k číslům, a to nikoliv v absolutní hodnotě, ale doleva a doprava na reálné ose.

Anglicky se těmto funkcím říká **floor** (podlaha) a **ceiling** (strop).

Bývá dobré si tu definici rozmyslet více způsoby, protože člověk to pak lépe vidí. $\lfloor x \rfloor$ je celé číslo, které má určité speciální vlastnosti. Podle čeho poznáme, že zrovna jedno konkrétní celé číslo je $\lfloor x \rfloor$?

Fakt 2b.14.

Nechť $x \in \mathbb{R}$, $n \in \mathbb{Z}$. Pak platí:

- (i) $\lfloor x \rfloor = n \iff n \leq x < n + 1$.
- (ii) $\lceil x \rceil = n \iff n - 1 < x \leq n$.
- (iii) $\lfloor x \rfloor = n \iff x - 1 < n \leq x$.
- (iv) $\lceil x \rceil = n \iff x \leq n < x + 1$.
- (v) $x - 1 < \lfloor x \rfloor \leq x \leq \lceil x \rceil < x + 1$.

Důkaz (pro úplnost): (i): \implies : Předpokládejme, že $n = \lfloor x \rfloor$. Pak $n = \max\{m \in \mathbb{Z}; m \leq x\}$. To mimo jiné znamená, že n v té množině leží, tudíž $n \leq x$. Protože je to ale maximální celé takové číslo, tak už v ní $n + 1$ neleží, proto $n + 1 > x$.

\impliedby : Předpokládejme, že celé číslo n splňuje $n \leq x < n + 1$. Pak toto n leží v množině $\{m \in \mathbb{Z}; m \leq x\}$. Z nerovnosti $x < n + 1$ ale vidíme, že $n + 1$ už v této množině neleží, proto je n největší číslo z této množiny, a tedy $n = \lfloor x \rfloor$.

(ii) se dokazuje podobně.

(iii): \implies : Jestliže je $n = \lfloor x \rfloor$, pak podle (i) je $n \leq x$ a také $x < n + 1$, což je $x - 1 < n$.

\impliedby : Nechť celé číslo n splňuje $x - 1 < n \leq x$. Z levé nerovnosti máme $x < n + 1$, proto $n \leq x < n + 1$ a podle (i) je $n = \lfloor x \rfloor$. Důkaz (iv) je podobný.

(v): Plyne z (i) až (iv), stačí do vztahů na pravých stranách dosadit namísto n příslušnou funkci. □

Ukážeme si ještě jiný způsob, jak poznat, že nějaké číslo n je jedna z těch dvou funkcí.

Fakt 2b.15.

Nechť $x \in \mathbb{R}$, $n \in \mathbb{Z}$. Pak platí:

- (i) $n = \lfloor x \rfloor$ právě tehdy, když existuje ε splňující $0 \leq \varepsilon < 1$ a $x = n + \varepsilon$.
- (ii) $n = \lceil x \rceil$ právě tehdy, když existuje ε splňující $0 \leq \varepsilon < 1$ a $x = n - \varepsilon$.

Důkaz (pro úplnost): (i): $1 \implies$: Definujme $\varepsilon = x - \lfloor x \rfloor = x - n$. Pak $x = n + \varepsilon$ a podle (i) z předchozího Faktu pak $0 \leq \varepsilon < 1$.

$2 \implies$: Předpokládejme, že $x = n + \varepsilon$ a $0 \leq \varepsilon < 1$. Pak $n \leq x$ a z $\varepsilon < 1$ máme $x < n + 1$, tudíž je splněna podmínka v (i) předchozího Faktu a $n = \lfloor x \rfloor$.

Důkaz (ii) je obdobný. □

Někteří autoři definují $\lfloor x \rfloor$ a $\lceil x \rceil$ pomocí podmínek z tohoto faktu.

Jak už jsme viděli, matematici se rádi ptají na pravidla, která by mohla platit, protože se pak lépe pracuje. Například by se mohly hodit vzorečky typu $\lfloor x+y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor$ či podobné pravidlo pro násobení, ale zrovna tohle nefunguje (viz cvičení níže). Zato platí desítky různých speciálních vzorečků. Alternativní definice z Faktu 2b.14 nám snadno dají následující identity.

Fakt 2b.16.

Nechť $x \in \mathbb{R}$. Pak platí:

- (i) $\lfloor -x \rfloor = -\lceil x \rceil$.
- (ii) $\lceil -x \rceil = -\lfloor x \rfloor$.
- (iii) $\lfloor x+n \rfloor = \lfloor x \rfloor + n$ pro všechna $n \in \mathbb{Z}$.
- (iv) $\lceil x+n \rceil = \lceil x \rceil + n$ pro všechna $n \in \mathbb{Z}$.

Důkaz (rutinní): (i): Nechť $n = \lceil x \rceil$. Pak podle Faktu 2b.14 (ii) platí $n-1 < x \leq n$. Potom také platí $-n+1 > x \geq -n$, tedy $(-n) \leq x < (-n)+1$ a podle Faktu 2b.14 (i) je $-n = \lfloor x \rfloor$.

Důkaz (ii) je podobný.

(iii): Označme si $m = \lfloor x \rfloor$. Pak podle Faktu 2b.14 (i) je $m \leq x < m+1$. Pak $m+n$ je celé číslo splňující $m+n \leq x+n < m+n+1$, tudíž podle Faktu 2b.14 (i) je $m+n = \lfloor x+n \rfloor$.

Důkaz (iv) je obdobný. □

Charakterizace z Faktu 2b.15 se zase hodí při důkazu této identity.

Fakt 2b.17.

Pro každé $x \in \mathbb{R}$ platí $\lfloor 2x \rfloor = \lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor$.

Důkaz (poučný): Označme $x = \lfloor x \rfloor + \varepsilon$. Rozebereme dva případy.

A) Předpokládejme, že $\varepsilon < \frac{1}{2}$. Pak máme i $x + \frac{1}{2} = \lfloor x \rfloor + (\varepsilon + \frac{1}{2})$ a $0 \leq \varepsilon + \frac{1}{2} < 1$, proto podle Faktu 2b.15 (ii) platí $\lfloor x + \frac{1}{2} \rfloor = \lfloor x \rfloor$.

Dále také máme $2x = 2\lfloor x \rfloor + (2\varepsilon)$ a $0 \leq 2\varepsilon < 1$, proto podle Faktu 2b.15 (ii) platí $\lfloor 2x \rfloor = 2\lfloor x \rfloor$. Dáme to dohromady:

$$\lfloor 2x \rfloor = 2\lfloor x \rfloor = \lfloor x \rfloor + \lfloor x \rfloor = \lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor.$$

B) Druhá (a poslední) možnost je, že $\frac{1}{2} \leq \varepsilon < 1$. Pak $0 \leq \varepsilon - \frac{1}{2} < 1$. Také máme $x + \frac{1}{2} = (\lfloor x \rfloor + \varepsilon) + \frac{1}{2} = (\lfloor x \rfloor + 1) + (\varepsilon - \frac{1}{2})$, proto podle Faktu 2b.15 (ii) platí $\lfloor x + \frac{1}{2} \rfloor = \lfloor x \rfloor + 1$.

Z předpokladu $\frac{1}{2} \leq \varepsilon < 1$ také vidíme, že $1 \leq 2\varepsilon < 2$, tedy $0 \leq 2\varepsilon - 1 < 1$. Také máme $2x = 2\lfloor x \rfloor + 2\varepsilon = (2\lfloor x \rfloor + 1) + (2\varepsilon - 1)$, proto podle Faktu 2b.15 (ii) platí $\lfloor 2x \rfloor = 2\lfloor x \rfloor + 1$. Dáme to dohromady:

$$\lfloor 2x \rfloor = 2\lfloor x \rfloor + 1 = \lfloor x \rfloor + (\lfloor x \rfloor + 1) = \lfloor x \rfloor + \lfloor x + \frac{1}{2} \rfloor.$$

□

Ukážeme si teď jednu aplikaci, další se budou tu a tam objevovat, viz třeba příklad 6a.b nebo Fakt 11b.6.

Příklad 2b.n: Máte flashku o velikosti 12GB. Kolik filmů o velikosti 700MB si na ni dokážete stáhnout?

Odpověď: $\lfloor \frac{12 \cdot 1024}{700} \rfloor = 17$.

△

Cvičení

Cvičení 2b.1: Který z následujících předpisů definuje zobrazení z množiny všech binárních řetězců do množiny celých čísel?

- (i) $T(r)$ je počet bitů 1 v r ;
- (ii) $T(r)$ je pozice prvního výskytu bitu 0 v r .

Cvičení 2b.2 (rutinní): Pro následující zobrazení určete jejich definiční obor a obor hodnot.

- (i) Zobrazení přiřazuje každému nezápornému celému číslu jeho poslední číslici.
- (ii) Zobrazení přiřazuje každému přirozenému číslu následující číslo.
- (iii) Zobrazení přiřazuje každému binárnímu řetězci jeho délku.
- (iv) Zobrazení přiřazuje každému binárnímu řetězci počet skupin „01“ v něm.
- (v) Zobrazení přiřazuje každému binárnímu řetězci počet jedniček mínus počet nul v něm.
- (vi) Zobrazení přiřazuje každému celému čísmu nejmenší čtverec, tj. číslo typu k^2 , který není menší než ono.
- (vii) Zobrazení dává maximum ze dvou reálných čísel.

Cvičení 2b.3: Zobrazení T přiřazuje každému studentovi jeho studijní průměr a zobrazení S přiřazuje k jednotlivým studijním průměrům stipendia. Co je zobrazení $S \circ T$?

Cvičení 2b.4 (rutinní, poučné): Pro následující dvojice funkcí $f, g: \mathbb{R} \mapsto \mathbb{R}$ najděte $g \circ f$ a $f \circ g$:

- (i) $f(x) = \sin(x)$, $g(x) = \pi x$;
- (ii) $f(x) = x$, $g(x) = e^x$;
- (iii) $f(x) = x^2$, $g(x) = 13$;
- (iv) $f(x) = 1 + x^2$, $g(x) = 1 + x^2$;
- (v) $f(x) = x^3 - 1$, $g(x) = \sqrt[3]{x+1}$;
- (vi) $f(x) = e^x$, $g(x) = \ln(|x|)$ pro $x \neq 0$ a $g(0) = 0$;
- (vii) $f(x) = 1 - x$, $g(x) = 1 - x$;
- (viii) $f(x) = 1 + x$, $g(x) = 1 + x$.

Cvičení 2b.5 (poučné): Pro následující dvojice funkcí $f, g: \mathbb{Z} \mapsto \mathbb{Z}$ rozhodněte, zda číslo 13 leží v oboru hodnot složené funkce $g \circ f$:

- (i) $f(x) = x^2 + 2$, $g(x) = 2x + 1$;
- (ii) $f(x) = x^3 + 4$, $g(x) = 2x - 11$;
- (iii) $f(x) = x^3 - 1$, $g(x) = 13x$.

Cvičení 2b.6 (poučné): Nechť $f(x) = ax + b$, $g(x) = cx + d$. Pro která a, b, c, d platí, že $f \circ g = g \circ f$?

Cvičení 2b.7 (rutinní, zkouškové, dobré^{*}): Jsou následující funkce prosté a na? Svou odpověď dokažte.

- (i) $f(n) = n + 1$ ze \mathbb{Z} do \mathbb{Z} ;
- (ii) $f(n) = n + 1$ z \mathbb{N} do \mathbb{N} ;
- (iii) $f(n) = 13n$ ze \mathbb{Z} do \mathbb{Z} ;
- (iv) $f(x) = 13x$ z \mathbb{Q} do \mathbb{Q} ;
- (v) $f(n) = n^3$ ze \mathbb{Z} do \mathbb{Z} ;
- (vi) $f(x) = x^3$ z \mathbb{R} do \mathbb{R} ;
- (vii) $f(n) = n^2 + 1$ ze \mathbb{Z} do \mathbb{Z} ;
- (viii) $f(n) = \lfloor \frac{n}{2} \rfloor$ ze \mathbb{Z} do \mathbb{Z} ;
- (ix) $f(n) = (-1)^n n$ z \mathbb{N}_0 do \mathbb{Z} ;
- (x)^{*} $f(n) = (-1)^n \lfloor \frac{n+1}{2} \rfloor$ z \mathbb{N}_0 do \mathbb{Z} ;
- (xi) $f(n) = (n+1, 2n)$ z \mathbb{N} do $\mathbb{N} \times \mathbb{N}$;
- (xii) $f(n) = (n^2, n^2 + 2n)$ ze \mathbb{Z} do $\mathbb{Z} \times \mathbb{Z}$;
- (xiii) $f(m, n) = (m^2, mn)$ ze $\mathbb{Z} \times \mathbb{Z}$ do $\mathbb{Z} \times \mathbb{Z}$;
- (xiv) $f(x, y) = (x+y, x-y)$ z $\mathbb{Q} \times \mathbb{Q}$ do $\mathbb{Q} \times \mathbb{Q}$;
- (xv) $f(m, n) = (m+n, m-n)$ ze $\mathbb{Z} \times \mathbb{Z}$ do $\mathbb{Z} \times \mathbb{Z}$;
- (xvi) $f(m, n) = 2m - n$ ze $\mathbb{Z} \times \mathbb{Z}$ do \mathbb{Z} ;
- (xvii)^{*} $f(m, n) = m^2 - n^2$ ze $\mathbb{Z} \times \mathbb{Z}$ do \mathbb{Z} ;
- (xviii) $f(m, n) = m + n + 13$ ze $\mathbb{Z} \times \mathbb{Z}$ do \mathbb{Z} ;
- (xix) $f(m, n) = m - n$ z $\mathbb{N}_0 \times \mathbb{N}$ do \mathbb{Z} .

Cvičení 2b.8 (poučné, dobré): Uvažujte zobrazení $T: \mathbb{N} \mapsto \mathbb{N}$ definované takto: $T(n) = \begin{cases} \frac{1}{2}n, & n \text{ sudé}; \\ 3n+1, & n \text{ liché}. \end{cases}$

Rozhodněte, zda je toto zobrazení bijekce \mathbb{N} na \mathbb{N} .

Poznámka: Vezměte si nějaké přirozené číslo n , dosaďte do T , pak ten výsledek zase strčte do T a tak dále, dokud nedostanete 1. Povedlo se to? Zkuste začít jiným číslem. A zase jiným. Co si o tom myslíte? Viz poznámka 5a.7.

Cvičení 2b.9 (poučné, dobré): Ukažte příklady funkcí z \mathbb{N} do \mathbb{N} (tedy vymyslete vzorečky), které by pokryly všechny kombinace vlastností prostoty a na (každá z nich má dvě možnosti, platí/neplatí, celkem tedy čtyři možné kombinace těchto vlastností).

Vymyslete čtyři obdobné funkce ze \mathbb{Z} do \mathbb{N} .

Cvičení 2b.10 (poučné, dobré, zkouškové): Nechť $T: A \mapsto B$ a $S: B \mapsto C$ jsou zobrazení. Rozhodněte, zda následující implikace platí, ty pravdivé dokažte, ty nepravdivé vyvrátte protipříkladem.

U všech implikací napište i její obměnu.

- (i) Jestliže T není prosté, tak $S \circ T$ není prosté.
- (ii) Jestliže S není prosté, tak $S \circ T$ není prosté.
- (iii) Jestliže T není na, tak $S \circ T$ není na.
- (iv) Jestliže S není na, tak $S \circ T$ není na.
- (v) Jestliže T není bijekce, tak $S \circ T$ není bijekce.
- (vi) Jestliže S není bijekce, tak $S \circ T$ není bijekce.

Cvičení 2b.11 (poučné, zkouškové, dobré^{*}): Nechť $T: A \mapsto B$ je zobrazení, $M, N \subseteq A$. Dokažte, že pak platí:

- (i) $T[M \cup N] = T[M] \cup T[N]$;
- (ii) $T[M \cap N] \subseteq T[M] \cap T[N]$.
- (iii)^{*} Je-li T prosté, pak $T[M \cap N] = T[M] \cap T[N]$.
- (iv) Ukažte, že obecně $T[M \cap N] = T[M] \cap T[N]$ neplatí.

Cvičení 2b.12 (poučné, dobré): Nechť U je universum. Pro $M \subseteq U$ definujme tzv. **charakteristickou funkcií** M jako

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

Také se jí říká indikátorová funkce, protože jedničkami indikuje, které body z U jsou v M . Dokážte následující:

- (i) Pro libovolné $M \subseteq U$: $\chi_{\overline{M}} = 1 - \chi_M$.
- (ii) Pro libovolné $M, N \subseteq U$: $\chi_{M \cap N} = \chi_M \cdot \chi_N$.
- (iii) Pro libovolné $M, N \subseteq U$: $\chi_{M \cup N} = \chi_M + \chi_N - \chi_{M \cap N}$.

Cvičení 2b.13 (rutinní): Kolik bajtů (bytes) je třeba na zakódování informace v délce 4/10/500/3000 bitů (bits)?

Cvičení 2b.14 (dobré): Nechť $a < b \in \mathbb{R}$.

- (i) Kolik celých čísel se nachází v intervalu $\langle a, b \rangle$?
- (ii) Kolik celých čísel se nachází v intervalu (a, b) ?

Cvičení 2b.15 (poučné): Dokažte, že pro $x \in \mathbb{R}$ platí $\lceil x \rceil - \lfloor x \rfloor = \begin{cases} 1, & x \notin \mathbb{Z}; \\ 0, & x \in \mathbb{Z}. \end{cases}$

Cvičení 2b.16 (poučné): Dokažte, že pro $n \in \mathbb{Z}$ platí $\lfloor n/2 \rfloor = \begin{cases} n/2, & n \text{ sudé}; \\ (n-1)/2, & n \text{ liché}. \end{cases}$

Cvičení 2b.17 (dobré): Načrtněte grafy funkcí $f_1(x) = \lfloor 2x \rfloor$, $f_2(x) = \lfloor x/2 \rfloor$, $f_3(x) = \lfloor x \rfloor + \lfloor x/2 \rfloor$, $f_4(x) = \lceil x \rceil + \lfloor x/2 \rfloor$, $f_5(x) = \lceil 2\lfloor x/2 \rfloor + \frac{1}{2} \rceil$ a $f_6(x) = \lceil x-2 \rceil + \lfloor x+2 \rfloor$.

Cvičení 2b.18 (dobré): Dokažte, že pro všechna $n \in \mathbb{Z}$ platí:

- (i) $\lfloor \lfloor n/2 \rfloor /2 \rfloor = \lfloor n/4 \rfloor$;
- (ii) $\lfloor n/2 \rfloor \cdot \lceil n/2 \rceil = \lfloor n^2/4 \rfloor$.

Cvičení 2b.19 (poučné, dobré): Dokažte či vyvraťte následující tvrzení:

- (i) $\forall x \in \mathbb{R}: \lfloor 2x \rfloor = 2\lfloor x \rfloor$.
- (ii) $\forall x, y \in \mathbb{R}: \lfloor x+y \rfloor = \lfloor x \rfloor + \lfloor y \rfloor$.
- (iii) $\forall x, y \in \mathbb{R}: \lceil x+y \rceil = \lceil x \rceil + \lceil y \rceil$.
- (iv) $\forall x, y \in \mathbb{R}: \lceil x \rceil + \lceil y \rceil - \lceil x+y \rceil$ je 0 nebo 1.
- (v) $\forall x, y \in \mathbb{R}: \lceil xy \rceil = \lceil x \rceil \cdot \lceil y \rceil$.
- (vi) $\forall x, y \in \mathbb{R}: \lfloor xy \rfloor = \lfloor x \rfloor \cdot \lfloor y \rfloor$.
- (vii) $\forall x \in \mathbb{R}: \lfloor \lceil x \rceil \rfloor = \lceil x \rceil$.
- (viii) $\forall x \in \mathbb{R}: \lceil \lfloor x \rfloor \rceil = \lfloor x \rfloor$.
- (ix) $\forall x \in \mathbb{R}: \lfloor \sqrt{\lceil x \rceil} \rfloor = \lfloor \sqrt{x} \rfloor$.
- (x) $\forall x \in \mathbb{R}: \lceil \sqrt{\lfloor x \rfloor} \rceil = \lceil \sqrt{x} \rceil$.
- (xi) $\forall x \in \mathbb{R}: \lceil \sqrt{\lceil x \rceil} \rceil = \lceil \sqrt{x} \rceil$.

Řešení:

2b.1: (i): ano. (ii): Co když je r prázdný, co když neobsahuje 0?

2b.2: (i): $D(T) = \mathbb{N}_0$, $R(T) = \{0, 1, 2, \dots, 8, 9\}$. (ii): $D(T) = \mathbb{N}$, $R(T) = \{2, 3, 4, \dots\} = \mathbb{N} - \{1\}$. (iii): $D(T)$ je množina konečných binárních řetězců, $R(T) = \mathbb{N}_0$. (iv): $D(T)$ je množina konečných binárních řetězců, $R(T) = \mathbb{N}_0$. (v): $D(T)$ je množina konečných binárních řetězců, $R(T) = \mathbb{Z}$. (vi): $D(T) = \mathbb{Z}$, $R(T) = \mathbb{N}_0$. (vii): $D(T) = \mathbb{R} \times \mathbb{R}$, $R(T) = \mathbb{R}$.

2b.3: Přiřazuje studentovi jeho stipendium.

2b.4: (i): $(g \circ f)(x) = g(f(x)) = \pi \sin(x)$, $(f \circ g)(x) = f(g(x)) = \sin(\pi x)$. (ii): $(g \circ f)(x) = g(f(x)) = e^x$, $(f \circ g)(x) = f(g(x)) = e^x$. (iii): $(g \circ f)(x) = g(f(x)) = 13$ (cokoliv dosazené do konstantní funkce je ta konstanta), $(f \circ g)(x) = f(g(x)) = 13^2$. (iv): $(g \circ f)(x) = g(f(x)) = 1 + (1+x^2)^2$, $(f \circ g)(x) = f(g(x)) = 1 + (1+x^2)^2$, je to vlastně f^2 (ve smyslu skládání zobrazení, ne ve smyslu násobení funkcí, pozor). (v): $(g \circ f)(x) = g(f(x)) = x$, $(f \circ g)(x) = f(g(x)) = x$, máme $g = f^{-1}$ a $f = g^{-1}$. (vi): $(g \circ f)(x) = g(f(x)) = x$ neboť $e^x > 0$ a tedy $g(f(x)) = \ln(|e^x|) = \ln(e^x) = x$, $(f \circ g)(x) = f(g(x)) = |x|$ pro $x \neq 0$ (takže není $g = f^{-1}$) a $(f \circ g)(0) = 1$. (vii): $(g \circ f)(x) = g(f(x)) = x$, $(f \circ g)(x) = f(g(x)) = x$, máme $g = f^{-1}$ neboli $f = f^{-1}$. (viii): $(g \circ f)(x) = g(f(x)) = x+2$, $(f \circ g)(x) = f(g(x)) = x+2 = f^2(x)$.

2b.5: Hledáme $x \in \mathbb{Z}$ tak, aby $x \xrightarrow{f} y \xrightarrow{g} 13$. Nejprve y , pak x .

- (i): $2y+1 = 13 \implies y = 6$, $x^2 + 2 = 6 \implies x = \pm 2$. Ano, $g(f)(\pm 2) = 13$, proto $13 \in R(g \circ f)$.

(ii): $2y - 11 = 13 \Rightarrow y = 12$, $x^3 + 4 = 12 \Rightarrow x = 2$. Ano, $g(f)(2) = 13$, proto $13 \in R(g \circ f)$.

(iii): $13y = 13 \Rightarrow y = 1$, $x^3 - 1 = 1 \Rightarrow x^3 = 2$ nemá řešení ze \mathbb{Z} . Proto $13 \notin R(g \circ f)$.

2b.6: Musí platit $ad + b = bc + d$ neboli $b(c - 1) = d(a - 1)$.

2b.7: (i): Je prosté: $T(x) = T(y) \Rightarrow x + 1 = y + 1 \Rightarrow x = y$. Je na: $y \in \mathbb{Z} \Rightarrow \exists x = y - 1 \in \mathbb{Z}$: $T(x) = T(y - 1) = (y - 1) + 1 = y$.

(ii): Je prosté: $T(x) = T(y) \Rightarrow x + 1 = y + 1 \Rightarrow x = y$. Není na: neexistuje $x \in \mathbb{N}$ aby $x + 1 = 1$, proto $1 \notin R(T)$.

(iii): Je prosté: $T(x) = T(y) \Rightarrow 13x = 13y \Rightarrow x = y$. Není na: neexistuje $x \in \mathbb{Z}$ aby $13x = 23$, proto $23 \notin R(T)$.

(iv): Je prosté: $T(x) = T(y) \Rightarrow 13x = 13y \Rightarrow x = y$. Je na: $y \in \mathbb{Q} \Rightarrow \exists x = \frac{1}{13}y \in \mathbb{Q}$: $T(x) = T(\frac{y}{13}) = 13 \cdot \frac{y}{13} = y$.

(v): Je prosté: $T(x) = T(y) \Rightarrow x^3 = y^3 \Rightarrow x = y$. Není na: neexistuje $x \in \mathbb{Z}$ aby $x^3 = 2$, proto $2 \notin R(T)$.

(vi): Je prosté: $T(x) = T(y) \Rightarrow x^3 = y^3 \Rightarrow x = y$. Je na: $y \in \mathbb{R} \Rightarrow \exists x = \sqrt[3]{y} \in \mathbb{R}$: $T(x) = T(\sqrt[3]{y}) = (\sqrt[3]{y})^3 = y$.

(vii): Není prosté, třeba $T(1) = T(-1)$. Není na: neexistuje $x \in \mathbb{Z}$ aby $x^2 + 1 = 0$, proto $0 \notin R(T)$.

(viii): Není prosté, třeba $T(2) = T(3)$. Je na: $y \in \mathbb{Z} \Rightarrow \exists x = 2y \in \mathbb{Z}$: $T(x) = T(2y) = \lfloor y \rfloor = y$.

(ix): Je prosté: $T(x) = T(y) \Rightarrow (-1)^x x = (-1)^y y \Rightarrow |(-1)^x x| = |(-1)^y y| \Rightarrow |x| = |y|$. Pak také $(-1)^x = (-1)^y$, tedy z $(-1)^x x = (-1)^y y$ je $x = y$. Není na: neexistuje $x \in \mathbb{Z}$ aby $(-1)^x x = 1$, proto $1 \notin R(T)$.

(x): Je prosté: $T(x) = T(y)$ pak musí mít $T(x), T(y)$ stejně znaménko, proto mají x, y stejnou paritu, tedy $y = x + 2k$. Platí také $|T(x)| = |T(y)|$ a tedy $\lfloor \frac{x+1}{2} \rfloor = \lfloor \frac{y+1}{2} \rfloor$, tedy $\lfloor \frac{x+1}{2} \rfloor = \lfloor \frac{x+1}{2} + k \rfloor = k + \lfloor \frac{x+1}{2} \rfloor$, proto $k = 0$ a $x = y$.

Je na: Nechť $y \in \mathbb{Z}$. Pokud $y \geq 0$, pak existuje $x = 2y \in \mathbb{N}_0$ a $T(x) = 1 \cdot \lfloor y + \frac{1}{2} \rfloor = y + \lfloor \frac{1}{2} \rfloor = y$.

Pokud $y < 0$, pak $-2y \geq 2$ a existuje $x = -2y - 1 \in \mathbb{N}_0$ takové, že $T(x) = (-1) \cdot \lfloor -y \rfloor = y$.

(xi): Je prosté: $T(x) = T(y) \Rightarrow (x + 1, 2x) = (y + 1, 2y) \Rightarrow 2x = 2y \Rightarrow x = y$. Není na: neexistuje $x \in \mathbb{Z}$ aby $x + 1 = 1$ a $2x = 1$, proto $(1, 1) \notin R(T)$.

(xii): Je prosté: $T(x) = T(y) \Rightarrow (x^2, x^2 + 2x) = (y^2, y^2 + 2y) \Rightarrow x^2 = y^2 \wedge x^2 + 2x = y^2 + 2y \Rightarrow 2x = 2y \Rightarrow x = y$. Není na: neexistuje $x \in \mathbb{Z}$ aby $x^2 = 0$ a $x^2 + 2x = 1$, proto $(0, 1) \notin R(T)$.

(xiii): Není prosté, třeba $T(2, 1) = (4, 2) = T(-2, -1)$. Není na: neexistují $x, y \in \mathbb{Z}$ aby $x^2 = 0$ a $xy = 1$, proto $(0, 1) \notin R(T)$.

(xiv): Je prosté: $T(x, y) = T(u, v) \Rightarrow (x + y, x - y) = (u + v, u - v) \Rightarrow x + y = u + v \wedge x - y = u - v$, sečteme: $2x = 2u \Rightarrow x = u$, odečteme: $2y = 2v \Rightarrow y = v$, proto $(x, y) = (u, v)$.

Je na: $(u, v) \in \mathbb{Q}^2 \Rightarrow \exists x = \frac{1}{2}(u + v), y = \frac{1}{2}(u - v) \in \mathbb{Q}$ a $T(x, y) = (u, v)$.

(xv): Je prosté: $T(x, y) = T(u, v) \Rightarrow (x + y, x - y) = (u + v, u - v) \Rightarrow x + y = u + v \wedge x - y = u - v$, sečteme: $2x = 2u \Rightarrow x = u$, odečteme: $2y = 2v \Rightarrow y = v$, proto $(x, y) = (u, v)$.

Není na: Soustava $x + y = 1$, $x - y = 0$ nemá řešení v \mathbb{Z} , proto $(1, 0) \notin R(T)$.

(xvi): Není prosté, třeba $T(1, 2) = 0 = T(0, 0)$. Je na: $z \in \mathbb{Z} \Rightarrow \exists x = 0, y = -z \in \mathbb{Z}$ a $T(x, y) = z$.

(xvii): Není prosté, třeba $T(1, 1) = 0 = T(0, 0)$. Na: To je moc dobrá otázka. Existují celá čísla m, n tak, aby třeba $m^2 - n^2 = 2$? Kupodivu ne. Omezíme se na nezáporná m, n . Aby vyšel výsledek kladný, musí být $m > n$, takže $m \geq n + 1$ a proto $m^2 - n^2 \geq (n + 1)^2 - n^2 = 2n + 1$. Kdyby $n = 0$, vyjde z rovnice $m^2 - n^2 = 2$ neřešitelné $m^2 = 2$, a pro $n \geq 1$ je $m^2 - n^2 \geq 3$. Takže nic.

(xviii): Není prosté, třeba $T(1, -1) = 13 = T(-1, 1)$. Je na: $z \in \mathbb{Z} \Rightarrow \exists x = z, y = -13 \in \mathbb{Z}$ a $T(x, y) = z$.

(xix): Není prosté, třeba $T(1, 1) = 0 = T(2, 2)$.

Je na: Nechť $z \in \mathbb{Z}$. Pokud $z \geq 0$, pak existuje $x = y, y = 0 \in \mathbb{N}_0$ a $T(x, y) = z$. Pokud $z < 0$, pak existuje $x = 0, y = -z \in \mathbb{N}_0$ a $T(x, y) = z$.

2b.8: Není prosté, protože $T(1) = 4 = T(8)$. Je na, pro $y \in \mathbb{N}$ existuje $x = 2y \in \mathbb{N}$ takové, že $T(x) = y$.

2b.9: $T(n) = n$ je prosté a na; $T(n) = 2n$ je prosté ale není na; $T(n) = n - 1$ pro $n \geq 2$, $T(1) = 1$ není prosté a je na, $T(n) = (n - 3)^2 + 2$ není prosté ani na.

$T(n) = \begin{cases} 2n + 1; & n \geq 0; \\ -2n; & n < 0 \end{cases}$ je prosté a na; $T(n) = \begin{cases} 2n + 3; & n \geq 0; \\ -2n; & n < 0 \end{cases}$ je prosté a není na; $T(n) = |n| + 1$ není

prosté a je na, $T(n) = n^2 + 1$ není prosté ani na.

2b.10: Obměny: (i) Jestliže je $S \circ T$ prosté, tak je T prosté. (ii) Jestliže je $S \circ T$ prosté, tak je S prosté.

(iii) Jestliže je $S \circ T$ na, tak je T na. (iv) Jestliže je $S \circ T$ na, tak je S na.

(v) Jestliže je $S \circ T$ bijekce, tak je T bijekce. (vi) Jestliže je $S \circ T$ bijekce, tak je S bijekce.

(i): Platí, T není prosté $\Rightarrow \exists x \neq y \in A: T(x) = T(y)$, pak $S(T(x)) = S(T(y))$ neboli $(S \circ T)(x) = (S \circ T)(y)$.

(ii), (iii), (v), (vi): nepravda. Třeba $A = \{1\}$, $B = \{a, b\}$, $C = \{\alpha\}$. Nechť $T: 1 \mapsto a$, $S: a, b \mapsto \alpha$. Pak T není na, S není prosté, ale $S \circ T$ je bijekce.

(iv): Platí, dokážeme tu obměnu. Zvolme $c \in C$ libovolné. Protože $S \circ T$ je na, $\exists a \in A: (S \circ T)(a) = c$ neboli $S(T(a)) = c$. Označme $b = T(a) \in B$, pak $S(b) = c$. Tedy $\forall c \in C$ najdeme $b \in B$ aby $S(b) = c$, tedy S je na.

2b.11: (i): $y \in T[M \cup N] \iff \exists x \in M \cup N: T(x) = y \iff (\exists x_1 \in M: T(x_1) = y) \vee (\exists x_2 \in N: T(x_2) = y) \iff y \in T[M] \vee y \in T[N] \iff y \in T[M] \cup T[N]$.

(ii): $y \in T[M \cap N] \iff \exists x \in M \cap N: T(x) = y \implies (\exists x_1 \in M: T(x_1) = y) \wedge (\exists x_2 \in N: T(x_2) = y) \iff y \in T[M] \wedge y \in T[N] \iff y \in T[M] \cap T[N]$.

(iii): $y \in T[M] \cap T[N] \iff y \in T[M] \wedge y \in T[N] \iff (\exists x_1 \in M: T(x_1) = y) \wedge (\exists x_2 \in N: T(x_2) = y)$. protože je T prosté, musí nutně být $x_1 = x_2$, označme $x = x_1 = x_2$ a vidíme, že $x \in M \cap N$, proto $x \in M \cap N$ a $T(x) = y$, tedy $y \in T[M \cap N]$.

(iv): $A = \{1, 2, 3\}$, $B = \{a, b\}$, $M = \{1, 2\}$, $N = \{2, 3\}$, $T(2) = b$, $T(1) = T(3) = a$.

2b.12: (i): $\chi_{\overline{M}}(x) = 1 \iff x \in \overline{M} \iff x \notin M \iff \chi_M(x) = 0 \iff (1 - \chi_M)(x) = 1$, podobně $\chi_{\overline{M}}(x) = 0 \iff (1 - \chi_M)(x) = 0$, tato dvě zobrazení tedy mají stejné hodnoty. Druhou ekvivalence není třeba dokazovat, protože obě funkce mají jen dvě možné hodnoty, 0 a 1, jestliže tedy nabývají jedničky ve stejných případech, pak nabývají i nuly ve stejných případech (těch ostatních).

(ii): $\chi_{M \cap N}(x) = 1 \iff x \in M \cap N \iff x \in M \wedge x \in N \iff \chi_M(x) = 1 \wedge \chi_N(x) = 1 \iff (\chi_M \cdot \chi_N)(x) = 1$. Poslední ekvivalence plyne z toho, že χ_X může být jen 0 nebo 1.

(iii): Důkaz se nejlépe dělá rozborém podle příslušnosti x k množinám. 1) Jestliže $x \in M \cap N$, pak $\chi_{M \cup N}(x) = 1$ a $(\chi_M + \chi_N - \chi_{M \cap N})(x) = 1 + 1 - 1 = 1$. 2) Jestliže $x \in M$, $x \notin N$, pak $\chi_{M \cup N}(x) = 1$ a $(\chi_M + \chi_N - \chi_{M \cap N})(x) = 1 + 0 - 0 = 1$. 3) Jestliže $x \in N$, $x \notin M$, pak $\chi_{M \cup N}(x) = 1$ a $(\chi_M + \chi_N - \chi_{M \cap N})(x) = 0 + 1 - 0 = 1$. 4) Jestliže $x \notin M$, $x \notin N$, pak $\chi_{M \cup N}(x) = 0$ a $(\chi_M + \chi_N - \chi_{M \cap N})(x) = 0 + 0 - 0 = 0$.

Tyto dvě funkce tedy mají vždy stejné hodnoty.

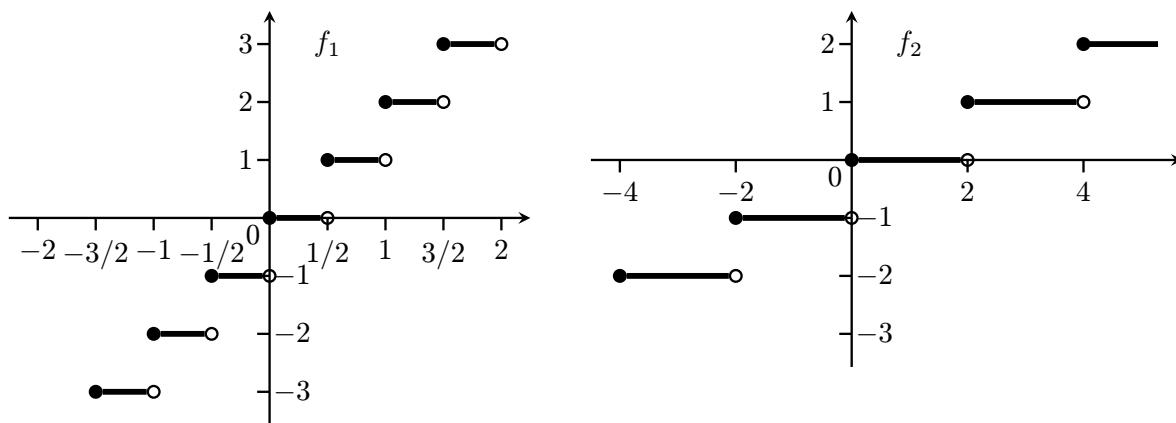
2b.13: 1 byte je 8 bits, takže: $\lceil \frac{4}{8} \rceil = 1$, $\lceil \frac{10}{8} \rceil = 2$, $\lceil \frac{500}{8} \rceil = 63$, $\lceil \frac{3000}{8} \rceil = 375$.

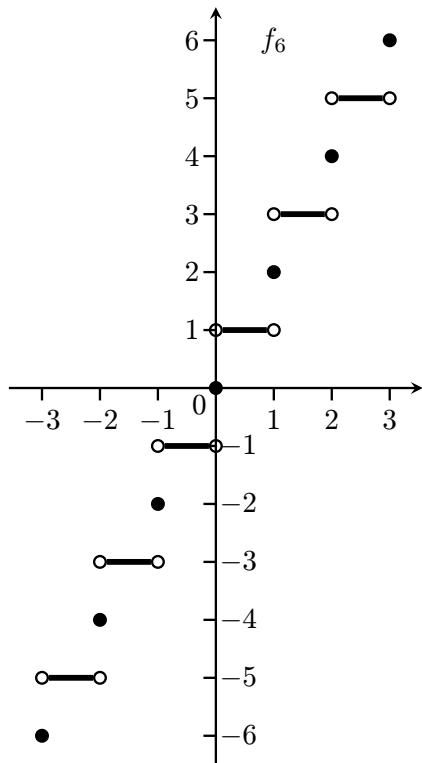
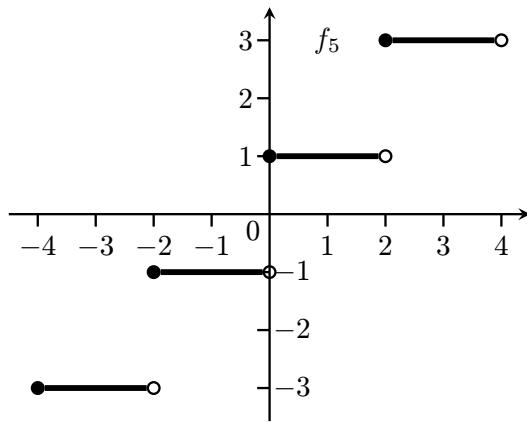
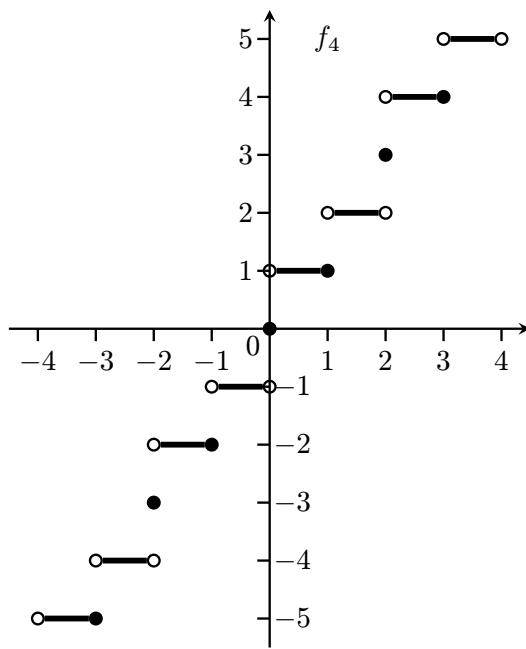
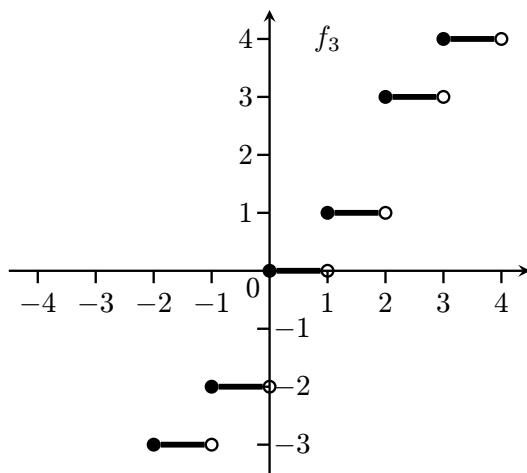
2b.14: Toto chce hodně experimentovat se zaokrouhlováním. (i): $\lfloor b \rfloor - \lceil a \rceil + 1$; (ii): $\lceil b \rceil - \lfloor a \rfloor - 1$.

2b.15: Nechť $x = n + r$, kde $r \in \langle 0, 1 \rangle$. Jestliže $r = 0$, pak $\lfloor x \rfloor = \lceil x \rceil = n$, jinak $\lfloor x \rfloor = n$ a $\lceil x \rceil = n + 1$.

2b.16: Je-li n sudé, pak $n = 2k$ pro $k \in \mathbb{Z}$ a proto $\lfloor \frac{n}{2} \rfloor = \lfloor k \rfloor = k = \frac{n}{2}$. Je-li n liché, pak $n = 2k + 1$ pro $k \in \mathbb{Z}$ a proto $\lfloor \frac{n}{2} \rfloor = \lfloor k + \frac{1}{2} \rfloor = k = \frac{n-1}{2}$.

2b.17:





2b.18: (i): Nechť $n = 4k + r$ pro $k \in \mathbb{Z}$ a $r = 0, 1, 2, 3$. Pak $\lfloor \frac{r}{2} \rfloor$ je 0 nebo 1, tedy $\lfloor \frac{1}{2} \lfloor \frac{r}{2} \rfloor \rfloor = 0$ a proto $\lfloor \frac{1}{2} \lfloor \frac{n}{2} \rfloor \rfloor = \lfloor \frac{1}{2} \lfloor 2k + \frac{r}{2} \rfloor \rfloor = \lfloor k + \frac{1}{2} \lfloor \frac{r}{2} \rfloor \rfloor = k + \lfloor \frac{1}{2} \lfloor \frac{r}{2} \rfloor \rfloor = k = \lfloor \frac{n}{4} \rfloor$.

(ii): Nechť $n = 2k + r$, kde $k \in \mathbb{Z}$ a $r = 0, 1$. Pak $\lfloor \frac{n}{2} \rfloor = k$ a $\lceil \frac{n}{2} \rceil = k + r$, proto $\lfloor \frac{n}{2} \rfloor \cdot \lceil \frac{n}{2} \rceil = k(k + r)$, zatímco $\lfloor \frac{n^2}{4} \rfloor = \lfloor \frac{4k^2 + 4kr + r^2}{4} \rfloor = k^2 + kr + \lfloor \frac{r^2}{4} \rfloor = k^2 + kr = k(k + r)$.

2b.19: (i): Neplatí, třeba: $\lfloor 2 \cdot 0.7 \rfloor = 1$, ale $2 \lfloor 0.7 \rfloor = 0$.

(ii): Neplatí, třeba $\lfloor 0.5 + 0.5 \rfloor = 1$, ale $\lfloor 0.5 \rfloor + \lfloor 0.5 \rfloor = 0$.

(iii): Neplatí, třeba $\lceil 0.4 + 0.4 \rceil = 1$, ale $\lceil 0.4 \rceil + \lceil 0.4 \rceil = 2$.

(iv): Platí, případy: pokud $x, y \in \mathbb{Z}$, pak evidentně vyjde 0. Pokud $x \in \mathbb{Z}$ a $y \notin \mathbb{Z}$, pak $x = n + r$ a $x + y = x + y + r$, kde $n \in \mathbb{Z}$, $n + y \in \mathbb{Z}$ a $0 < r < 1$, proto $\lceil x \rceil + \lceil y \rceil - \lceil x + y \rceil = n + 1 + y - (n + y + 1) = 0$. Zbývá případ $x, y \notin \mathbb{Z}$, tedy $x = n + r$, $y = m + s$, kde $m, n \in \mathbb{Z}$ a $0 < r, s < 1$. Dva případy. Pokud $r + s > 1$, pak $\lceil x \rceil + \lceil y \rceil - \lceil x + y \rceil = n + 1 + y + 1 - (n + y + 2) = 0$. Pokud $r + s \leq 1$, pak $\lceil x \rceil + \lceil y \rceil - \lceil x + y \rceil = n + 1 + y + 1 - (n + y + 1) = 1$.

(v): Neplatí, třeba $\lceil 1.1 \cdot 1.1 \rceil = \lceil 1.21 \rceil = 2$, ale $\lceil 1.1 \rceil \cdot \lceil 1.1 \rceil = 2 \cdot 2 = 4$.

(vi): Neplatí, třeba $\lfloor 4 \cdot 0.5 \rfloor = \lfloor 2 \rfloor = 2$, ale $\lfloor 4 \rfloor \cdot \lfloor 0.5 \rfloor = 4 \cdot 0 = 0$.

(vii) a (viii): Platí, protože $\lfloor x \rfloor \in \mathbb{Z}$ a $\lceil x \rceil \in \mathbb{Z}$, takže aplikace dalšího zaokrouhlení již nic neovlivní.

(ix): Neplatí, nechť $x = 1.9^2 = 3.61$, pak $\lfloor \sqrt{\lceil x \rceil} \rfloor = 2$, ale $\lceil \sqrt{x} \rceil = 1$.

- (x): Platí. Pro $x \geq 0$ nechť $n \in \mathbb{N}_0$ je číslo takové, že $n^2 \leq x < (n+1)^2$. Pak $n \leq \sqrt{x} < n+1$, proto $\lfloor \sqrt{x} \rfloor = n$. Jelikož $n^2 \in \mathbb{Z}$, bude i $n^2 \leq \lfloor x \rfloor < (n+1)^2$ a tedy $n \leq \sqrt{\lfloor x \rfloor} < n+1$, proto i $\lfloor \sqrt{\lfloor x \rfloor} \rfloor = n$.
- (xi): Platí, důkaz jako v (xi), pro dané $x \geq 0$ se vybere $n \in \mathbb{N}$ tak, aby $(n-1)^2 < x \leq n^2$.

2c. Mohutnost množin

V předchozí sekci jsme si intuitivně rozmysleli, že pokud máme konečné množiny s různým počtem prvků, tak mezi nimi nedokážeme udělat bijekci, viz Fakt 2b.12 (iii). Naopak pokud máme dvě konečné množiny se stejným počtem prvků, tak mezi nimi bijekci udělat dokážeme (stačí si prvky v obou množinách očíslovat a poslat první na první, druhý na druhý atd.) Tato pozorování se stanou východiskem pro porovnávání velikostí množin obecně.

!

Definice.

Řekneme, že množiny A, B mají stejnou **mohutnost**, značeno $|A| = |B|$, jestliže existuje bijekce z A na B .

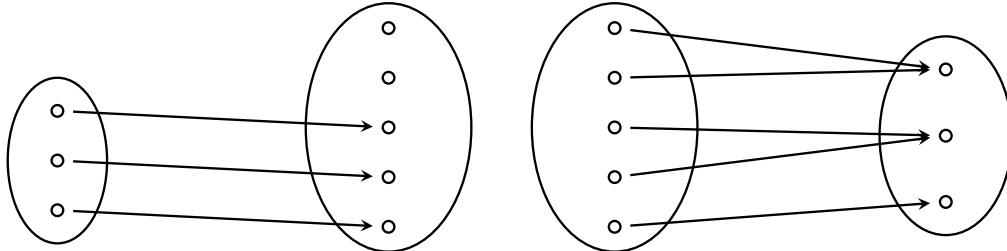
Řekneme, že mohutnost množiny A je menší nebo rovna mohutnosti množiny B , značeno $|A| \leq |B|$ nebo $|B| \geq |A|$, jestliže existuje prosté zobrazení z A do B .

We say that two sets A, B have the same **cardinality**, denoted $|A| = |B|$, if there exists a bijection from one onto the other.

We say that cardinality of A is less than or equal to cardinality of B , denoted $|A| \leq |B|$, if there exists a 1-1 mapping from A to B .

Odpovídá i druhá definice naší intuici pro konečné množiny?

Předpokládejme, že máme prosté zobrazení $T: A \rightarrow B$. Pak je $T: A \rightarrow R(T)$ bijekce, tudíž mají A a $R(T)$ stejnou mohutnost (šipky se nesbíhají, to zní rozumně). A protože $R(T) \subseteq B$, tak je přirozené, že pak považujeme B za větší (případně stejně velké) ve srovnání s A , přesně toto konec konců říká obměna tvrzení (ii) z Faktu 2b.12. A naopak, přinejmenším u konečných množin má člověk pocit, že by z menší dokázal posílat šipky do větší tak, aby se nesbíhaly, a vyrobít tak prosté zobrazení.



Obrázek napravo naznačuje ještě jeden způsob, jak poznat tu větší množinu ze dvou, vycházíme zde z obměny tvrzení (i) z Faktu 2b.12. Formálně:

Fakt 2c.1.

Nechť A, B jsou množiny.

$|A| \leq |B|$ právě tehdy, když existuje zobrazení $T: B \rightarrow A$, které je na.

Toto se občas hodí, ale většinou je výrazně snadnější pracovat s prostými zobrazeními jako v definici.

Poznamenejme, že existuje ještě alternativní a velmi rozšířené značení pro porovnávání mohutnosti. Někteří autoři píšou místo $|A| \leq |B|$ zápis $A \preceq B$ a místo $|A| = |B|$ píšou $A \sim B$, popřípadě $A \approx B$. Sám nemám jasnou preferenci, možná mírně k tomuto alternativnímu značení, ale v této knize používáme symbol \preceq pro relaci částečného uspořádání, tak jsem pro mohutnost zvolil verzi s $|A|$.

Teď dokážeme několik jednoduchých vlastností, které nás ubezpečí, že se nové pojmy chovají tak, jak bychom rádi.

Fakt 2c.2.

Nechť A je množina.

(i) $|A| = |A|$ a $|A| \leq |A|$.

(ii) Jestliže $B \subseteq A$, pak $|B| \leq |A|$.

(iii) $|A| = |B|$ právě tehdy, když $|B| = |A|$.

Důkaz (rutinní): (i): Uvažujme identitu $i_A: A \mapsto A$ (viz mocniny T^n). Toto zobrazení je bijekce, tedy je $|A| = |A|$, a je i prosté, proto $|A| \leq |A|$.

(ii): Zobrazení $i_B: B \mapsto B$ lze považovat také za zobrazení $i_B: B \mapsto A$. Tím, že se do cílové množiny přidaly prvky navíc, jsme nemohli změnit prostotu i_B (pořád posílá stejné prvky stejným způsobem), podle definice tedy $|B| \leq |A|$.

(iii): Jestliže $|A| = |B|$, pak existuje bijekce $T: A \mapsto B$. Inverzní zobrazení T^{-1} dává bijekci z B na A , tedy $|B| = |A|$. Opačný směr plyne ze symetrie. \square

Čtenáři se tato tvrzení (stejně jako mnohá další) mohou zdát samozřejmá, ale samozřejmá jsou jen pro intuitivní pojem „velikosti množiny“ tak, jak jej zná z běžného života. Zde máme „mohutnost“ definovanou pomocí zobrazení, takže platnost oněch „jasných“ věcí není automatická a je třeba je ověřit pomocí definice. Protože jsme pojem mohutnosti vymysleli dobře, budou ty běžné věci v běžných situacích fungovat, ale občas to dá překvapivě hodně práce a někdy ty jasné věci nebudou fungovat vůbec (to je reklama na zbytek kapitoly).

S mohutností (tedy porovnávání množin dle velikosti) se pracuje podobně jako s porovnáváním čísel podle velikosti $|x| \leq |y|$ a $|x| = |y|$, následující tvrzení ukážou, že tyto vztahy mají podobné vlastnosti.

Fakt 2c.3.

Nechť A, B, C jsou množiny.

- (i) Jestliže $|A| \leq |B|$ a $|B| \leq |C|$, pak $|A| \leq |C|$.
- (ii) Jestliže $|A| \leq |B|$ a $|B| = |C|$, pak $|A| \leq |C|$.
- Jestliže $|A| = |B|$ a $|B| \leq |C|$, pak $|A| \leq |C|$.
- (iii) Jestliže $|A| = |B|$ a $|B| = |C|$, pak $|A| = |C|$.

Důkaz (rutinní, poučný): (i): Z předpokladu $|A| \leq |B|$ dostáváme existenci zobrazení $T: A \mapsto B$, které je prosté. Podobně z předpokladu $|B| \leq |C|$ dostaneme prosté zobrazení $S: B \mapsto C$. Podle Faktu 2b.10 (i) je i složené zobrazení $S \circ T: A \mapsto C$ prosté, proto podle definice je $|A| \leq |C|$.

(ii): Teď se začne prostým zobrazením a bijekcí, ale ta je také prostá, tedy máme dvě prostá zobrazení a dál je to jako v (i).

(iii): Stejný důkaz, jen se použije Fakt 2b.10 (iii).

Na tento důkaz už byste opravdu měli přijít sami. \square

Fakt 2c.4.

Nechť A, B jsou množiny.

Jestliže $|A| = |B|$, pak $|A| \leq |B|$ a $|B| \leq |A|$.

Důkaz (rutinní): Jestliže $|A| = |B|$, pak existuje bijekce $T: A \mapsto B$. Tato bijekce je i prostá, proto $|A| \leq |B|$, a je na, tedy $|A| \geq |B|$. \square

Platí to i naopak? Ano, ale už to není tak lehké, což je vidět například z toho, že je to věta a navíc pojmenovaná po třech lidech, kteří se na ni museli dát dohromady.

! Věta 2c.5. (Cantor-Bernstein-Schroeder)

Nechť A, B jsou množiny. Jestliže $|A| \leq |B|$ a $|B| \leq |A|$, pak $|A| = |B|$.

Důkaz je těžký, je totiž třeba ze dvou prostých zobrazení $A \mapsto B$ a $B \mapsto A$ vyrobit bijekci. Zvědavci a puntičkáři mohou zkousit prakticky jakoukoliv tlustší knihu o teorii množin či Wikipedii. Každopádně je to věta zajímavá nejen z hlediska teoretického, ale i z hlediska praktického. Vyrábět bijekce je totiž často výrazně obtížnější než vyrobit prostá zobrazení, která potřebujeme k důkazu oněch dvou „nerovností“.

Vidíme, že porovnávání mohutnosti se opravdu silně podobá rovnosti a nerovnosti, pro další užitečné vlastnosti se podívejte na cvičení 2c.1. Zavedeme ještě jedno značení, které nám občas zjednoduší práci.

Definice.

Nechť A, B jsou množiny. Řekneme, že mohutnost A je **striktně (ostře) menší** než mohutnost B , značeno $|A| < |B|$, jestliže $|A| \leq |B|$, ale neplatí $|A| = |B|$.

Zavedeme také značení $|A| \neq |B|$ pro případ, kdy neplatí $|A| = |B|$.

Teď si v mohutnostech množin uděláme pořádek.

!

Definice.

Množina A se nazve **konečná**, jestliže $A = \emptyset$ (pak píšeme $|A| = 0$) nebo existuje takové $m \in \mathbb{N}$, aby platilo $|A| = |\{1, 2, \dots, m\}|$, pak píšeme $|A| = m$.

Jinak se množina nazve **nekonečná**.

Množina A se nazve **spočetná**, jestliže má stejnou mohutnost jako množina \mathbb{N} .

Množina A se nazve **nespočetná**, jestliže je nekonečná, ale není spočetná.

We say that a set A is **finite** if either $A = \emptyset$, then we write $|A| = 0$, or if there exists $m \in \mathbb{N}$ such that $|A| = |\{1, 2, \dots, m\}|$, then we write $|A| = m$. Otherwise we say that A is **infinite**. We say that A is **countable** if $|A| = |\mathbb{N}|$. We say that A is **uncountable** if it is infinite but not countable.

Poznámka: Někteří autoři rozumí pod pojmem „spočetná“ podmínce $|A| \leq |\mathbb{N}|$, z pohledu diskrétní matematiky to docela dává smysl, protože právě s těmito množinami se dobře pracuje například indukcí. My zde volíme obvyklejší názvosloví (spočetná znamená $|A| = |\mathbb{N}|$), podmínu $|A| \leq |\mathbb{N}|$ umíme vyjádřit slovy „ A je nejvíce spočetná“. Praktický dopad nejednoznačnosti v terminologii je, že až se budete s někým o spočetnosti bavit, tak se nejprve domluvte, co tím vlastně myslíte.

△

Příklad 2c.a: $|\{a, b, a\}| = 2$, $|\emptyset| = 0$. Množina \mathbb{N} je nekonečná a spočetná. Množina \mathbb{R} je nekonečná, ale zatím nevíme, jestli je spočetná.

△

Poučná poznámka:. Čtenáře možná překvapí, že si v zásadě můžeme definice dělat, jak chceme. Můžeme třeba zadefinovat, že konečné množiny jsou ty, které obsahují číslo 13, ostatní množiny jsou pak nekonečné. Z čistě logického hlediska by to nebylo špatně, jenže nový pojem velikosti by měl divné vlastnosti (například sjednocením konečné a nekonečné množiny bychom dostali konečnou). To v zásadě nevadí, matematici rádi vymýšlejí podivné světy a pak zkoumají, co tam vlastně platí a co ne, jenže my matematiku vytváříme také proto, aby byla užitečná, a moje alternativní definice velikosti množin je na pytel. Kdybych tu definici vážně navrhnul, matematici by se mi hlasitě smáli.

Když matematici nové pojmy vymýšlejí, tak se přitom řídí několika zásadami. Jako druhou věc po definici chtejí, aby ten nový pojem k něčemu byl. Často se jedná o pojem inspirovaný naší intuicí či zkušeností, pak se také chce, aby ten pojem s naší intuicí souhlasil. Naše diskuse a faktíky výše i níže doufajme přesvědčí čtenáře, že zde zavedený pojem mohutnosti opravdu funguje tak, jak bychom chtěli. Třeba jsme dokázali, že když je A „menší“ než B a B „menší“ než C , tak je nutně A „menší“ než C . Kdyby to náš pojem velikosti množin nesplňoval, tak bychom měli silné podezření, že jsme naší definici nevymysleli zrovna nejlépe.

Ovšem to první, co matematici při vytváření definice žádají, je její správnost logická. Říká se tomu, že se chce, aby „definice měla smysl“, což mimo jiné znamená, že musí umět rozhodnout. Například to, zda $|A| = |B|$, je jasné, prostě buď nějaká bijekce je, nebo není. U naší definice konečných a jiných množin to ovšem jasné není, čímž se konečně dostáváme k tématu této poznámky. Je velikost množiny touto definicí jasně dána? Máme například množinu A , která je bijekcí spojena s množinou $\{1, 2, 3\}$, tedy podle definice $|A| = 3$. Mohlo by se stát, že by také existovala bijekce z A na $\{1, 2, 3, 4\}$? To by bylo velice nemilé, protože pak by také $|A| = 4$ a my rozhodně nechceme, aby jedna množina mohla mít více velikostí.

Podle Faktu 2c.3 (iii) by pak ale platilo $|\{1, 2, 3\}| = |\{1, 2, 3, 4\}|$, což nevypadá moc pravděpodobně. Abychom ukázali, že naše definice funguje rozumně, musíme dokázat, že nelze vytvořit bijekci mezi $\{1, 2, 3\}$ a $\{1, 2, 3, 4\}$, podobně o dalších vzorových množinách rozdílných velikostí. To je ale spíš téma pro teorii množin, necháme to odborníkům a spokojíme se s konstatováním, že to ověřili a nepřístojnosti se nekonají.

Podobně si necháme dokázat, že ani množina \mathbb{N} se nedá bijekcí spojit s množinami typu $\{1, \dots, n\}$, čímž se potvrdí, že nejde o množinu konečnou (to jsme si oddechli). V definici je tedy vše v pořádku.

△

Teď se postupně podíváme na jednotlivé typy mohutností. Začneme množinami konečnými a ukážeme, že vše fungují tak, jak bychom čekali. Nejprve zkusíme (snadným) tvrzením čtenáře přesvědčit, že definice opravdu správně vystihla, co konečné množiny jsou.

Fakt 2c.6.

- (i) Nechť A je konečná množina, $|A| = n$. Pak ji lze zapsat jako $A = \{a_1, a_2, \dots, a_n\}$, kde a_k jsou navzájem různé prvky.
(ii) Je-li naopak $A = \{a_1, a_2, \dots, a_n\}$, kde a_k jsou navzájem různé prvky, pak A je konečná a $|A| = n$.

Důkaz (poučný): (i): Protože je to množina konečná, existuje bijekce T z nějaké množiny $\{1, 2, \dots, n\}$ na A . Definujme $a_k = T(k)$, pak z prostoty vyplývá, že jsou to navzájem různé prvky A , a ze surjektivity T vyplývá, že $A = \{a_1, a_2, \dots, a_n\}$.

(ii): Jestliže $A = \{a_1, \dots, a_n\}$, pak stačí definovat $T(a_k) = k$. To bude určitě zobrazení z A na $\{1, \dots, n\}$ a prosté je také: Jestliže jsou $x \neq y \in A$, pak existují indexy k, l takové, že $x = a_k$ a $y = a_l$. Protože $x \neq y$, musí být i $k \neq l$ a tedy $T(x) \neq T(y)$. \square

Následující věta ukazuje, že se pojmy spojené s konečnými množinami chovají v souladu s naší intuicí. Poznámejme, že důkaz je snadný, ale dlouhý, protože je třeba hlídat spoustu věcí. Pro čtenáře může být zajímavé si důkaz číst a přitom si kreslit odpovídající obrázky.

Věta 2c.7.

- (i) Jestliže je A konečná množina, pak je i každá její podmnožina B konečná a platí $|B| \leq |A|$.
Je-li navíc B podmnožina vlastní, pak $|B| < |A|$.
(ii) Nechť A, B jsou konečné množiny. Pak je i $A \cup B$ konečná a platí $|A \cup B| \leq |A| + |B|$.
Jsou-li navíc A, B disjunktní, pak $|A \cup B| = |A| + |B|$.
(iii) Nechť A, B jsou konečné množiny. Pak je $A \times B$ konečná a platí $|A \times B| = |A| \cdot |B|$.

U (i) je zajímavá i obměna, viz cvičení 2c.5.

Důkaz (poučný, asi drsný): Důkaz (i) spíš jen naznačíme. Nechť A je konečná množina. Podle definice tedy existuje $m \in \mathbb{N}$ a bijekce T z A na $\{1, \dots, m\}$. Začneme následující situací. Nechť a je libovolný prvek A a uvažujme množinu $A' = A - \{a\}$. Chceme dokázat, že je konečná a má menší mohutnost než A . Kdyby náhodou $T(a) = m$, pak je restrikce $T|_{A'}$ bijekcí z A' na $\{1, \dots, m-1\}$, což dokazuje, že A' je konečná a $|A'| = m-1 < |A|$.

Zbývá rozebrat situaci, když $T(a) = n < m$. Protože je T na, musí existovat jiný prvek $b \in A$ takový, že $T(b) = m$. Vytvoříme nové zobrazení tak, že tyto dvě šipky prohodíme. Formálně to uděláme tak, že definujeme $S(x) = \begin{cases} T(x), & x \in A' - \{b\}; \\ n, & x = b. \end{cases}$

Protože si S vybírá své hodnoty z hodnot T , dostali jsme zobrazení z A' do $\{1, \dots, m\}$, ale hodnotě m jsme se také vynuli, takže zobrazení S jde vlastně do množiny $\{1, \dots, m-1\}$.

Je prosté? Nechť $x \neq y \in A'$. Jestliže se ani jeden z x, y nerovná b , pak $S(x) = T(x)$ a $S(y) = T(y)$; ale T bylo prosté, proto $S(x) \neq S(y)$. Druhá možnost je, že jeden z nich je b , podle symetrie můžeme předpokládat, že třeba $x = b$ a $y \neq b$. Pak $S(x) = n$, mohlo by být i $S(y) = n$? Protože $y \neq b$, tak $S(y) = T(y)$, a jediný prvek z A , který dá po dosazení do T hodnotu n , byl a , ale ten v A' není a proto $y \neq a$, tedy i $S(y) \neq n$.

Takže T je prosté z A do $\{1, \dots, m-1\}$, proto $|A'| \leq m-1 < |A|$.

Ukázali jsme, že se mohutnost konečné množiny při odebrání prvku zmenší, z toho už (i) vyplýne.

(ii): Nejprve dokážeme případ, kdy jsou A a B disjunktní. Podle předpokladu jsou konečné, tedy existují čísla $m, n \in \mathbb{N}$ a bijekce $R: A \mapsto \{1, \dots, m\}$ a $S: B \mapsto \{1, \dots, n\}$.

Definujme zobrazení T na množině $A \cup B$ takto:

$$T(x) = \begin{cases} R(x), & x \in A; \\ S(x) + m, & x \in B. \end{cases}$$

Obrázkem: Jako bychom posunuli cíle šipek vedoucích z B do \mathbb{N} nahoru o m , čímž na začátku \mathbb{N} vzniklo přesně m volných míst pro původní (neposunuté) šipky z A .

Tato definice má smysl, protože každý prvek x padne přesně do jedné z těch kategorií (A či B), u žádného nemůže být spor mezi dvěma různými možnostmi—tady právě silně používáme toho, že jde o množiny disjunktní.

Tvrdíme, že jde o bijekci z $A \cup B$ na $\{1, \dots, m+n\}$.

Nejprve ukážeme, že nevyskočí pryč. Vezměme $x \in A \cup B$. Jestliže $x \in A$, pak $T(x) = R(x) \leq m \leq m+n$. Jestliže $x \in B$, pak $T(x) = S(x) + m \leq n+m$. Zobrazení T tedy opravdu jde do cílové množiny. Je na?

Nechť $k \in \{1, \dots, m+n\}$. Jsou dvě možnosti. Pokud je $k \leq m$, pak díky tomu, že je R na, dostaneme $a \in A$ takové, že $T(a) = k$. Pak ovšem $a \in A \cup B$ a $T(a) = R(a) = k$.

Pokud je $k > m$, pak $1 \leq k - m \leq n$ a S bylo také na, tudíž existuje $b \in B$ splňující $S(b) = k - m$. Pak $b \in A \cup B$ a $T(b) = S(b) + m = (k - m) + m = k$. Surjektivita T je dokázána.

Je T prosté? Vezměme $x \neq y \in A \cup B$. Jestliže obě splňují $x, y \in A$, pak $T(x) = R(x)$ a $T(y) = R(y)$. Protože R bylo prosté, musí být $T(x) \neq T(y)$.

Jestliže jsou oba prvky v B , pak podobně $S(x) \neq S(y)$ a proto $S(x) + m \neq S(y) + m$, tedy $T(x) \neq T(y)$.

Zbývá situace, že jeden prvek je z A a druhý z B , podle symetrie situace můžeme předpokládat, že $x \in A$ a $y \in B$. Pak ale $T(x) = R(x) \leq m$, zatímco $T(y) = S(y) + m > m$. Tudíž zase $T(x) \neq T(y)$ a všechny možnosti jsme vyčerpali. T je prosté.

Ukázali jsme, že existuje bijekce z $A \cup B$ na $\{1, \dots, m+n\}$. Proto je podle definice množina $A \cup B$ konečná a $|A \cup B| = m+n = |A|+|B|$.

Zbývá ukázat, že platí to obecné tvrzení pro A a B libovolné. To se udělá následujícím trikem.

Uvažujme množinu $B' = B - A$ (vyhodíme z B společné prvky s A , pokud nějaké jsou). Pak $B' \subseteq B$, proto je to podle (i) konečná množina a platí $|B'| \leq |B|$. Navíc jsou A a B' disjunktní, proto podle právě dokázaného je i $A \cup B'$ konečná a platí $|A \cup B'| = |A| + |B'|$.

Platí také $A \cup B = A \cup B'$ (viz cvičení 2a.1 (vi)), když tak si nakreslete Vennův diagram), proto máme

$$|A \cup B| = |A \cup B'| = |A| + |B'| \leq |A| + |B|.$$

(iii): Protože je A konečná množina, můžeme ji napsat jako $\{a_1, \dots, a_m\}$, kde $m = |A|$ (viz Fakt 2c.6). Pro $k \in \{1, \dots, m\}$ uvažujme $B_k = \{(a_k, b); b \in B\}$. Pak je zobrazení $T_k(b) = (a_k, b)$ bijekce z B na B_k . Prostota: $x \neq y \in B \implies (a_k, x) \neq (a_k, y) \implies T_k(x) \neq T_k(y)$.

Na: Nechť $(a_k, b) \in B_k$. Pak $b \in B$ a $T_k(b) = (a_k, b)$.

Tohle je zjevné, prostě jsme ke každému prvku z množiny B jakoby přidali značku, také si to můžeme představit, že jsme celou množinu jen posunuli, množina tím samozřejmě nemohla změnit velikost. Máme tedy $|B| = |B_k|$. Teď si uvědomíme, že B_k jsou navzájem disjunktní množiny, neboť pro $k \neq l$ se prvky z B_k liší od prvků z B_l na první souřadnici, a $A \times B = B_1 \cup \dots \cup B_m$. Můžeme teď opakováně použít výsledek z (ii) a dostaneme

$$|A \times B| = |B_1| + \dots + |B_m| = |B| + \dots + |B| = m \cdot |B| = |A| \cdot |B|.$$

Tím je důkaz hotov. □

Samozřejmě existují i verze pro více množin.

! Věta 2c.8.

(i) Jsou-li A_i pro $i = 1, 2, \dots, n$ konečné množiny, pak je i $\bigcup_{i=1}^n A_i$ konečná a $\left| \bigcup_{i=1}^n A_i \right| \leq \sum_{i=1}^n |A_i|$.

Jsou-li navíc po dvou disjunktní, tak $\left| \bigcup_{i=1}^n A_i \right| = \sum_{i=1}^n |A_i|$.

(ii) Jsou-li A_i pro $i = 1, 2, \dots, n$ konečné množiny, pak je i $A_1 \times \dots \times A_n$ konečná a

$$|A_1 \times \dots \times A_n| = |A_1| \cdots |A_n| = \prod_{i=1}^n |A_i|.$$

Důkaz je indukcí a necháme to do cvičení 5a.11. Vrátíme se ještě k situaci, když sjednocujeme množiny A a B , které nejsou disjunktní. Jakou velikost pak dostaneme? Jestliže zvlášť spočítáme prvky z A a prvky z B , tak jsme vlastně dvakrát započítali ty prvky, které jsou společné, což je třeba napravit. Teď už je asi jasné, jak se to má dělat.

Fakt 2c.9.

(i) Jsou-li A, B konečné množiny, pak $|A \cup B| = |A| + |B| - |A \cap B|$.

(ii) Jsou-li A, B, C konečné množiny, pak

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Důkaz (náznak): Pořádný důkaz by nás zase zavedl do hlubin teorie množin, v knihách se často odvoláme právě na to, kolikrát je který prvek zpočítán nalevo a napravo. Pro (i) jsme to už provedli před Faktem, pro (ii) to uděláme teď.

Na levé straně je samozřejmě každý prvek z $A \cup B \cup C$ započítán jen jednou. Musíme ukázat totéž o pravé straně. Rozdělíme si prvky podle toho, zda do A, B, C patří či nepatří. Víme, že je celkem 8 možností, z toho ta, kdy nejsou ani v jedné množině, nás teď nezajímá. Zbývá 7 ostatních.

- a) $x \in A$, ale není v žádné ostatní množině. Pak neleží ani v žádném z průniků, tudíž je napravo započítán jen jednou. Podobná úvaha platí i pro prvky jen z B a prvky jen z C .
- b) $x \in A$, $x \in B$, ale $x \notin C$. Pak x leží z těch množin napravo v A , B , $A \cap B$ a žádné jiné, tudíž je tam započítán $1 + 1 - 1 = 1$ krát. Podobná úvaha zase platí pro prvky, které jsou jen v A a C či jen v B a C .
- c) Zbývají prvky, které jsou v A , B i v C . Takové prvky jsou pak ve všech množinách napravo, tudíž jsou započítány celkem $1 + 1 + 1 - 1 - 1 + 1 = 1$ krát.

□

Je užitečné si nakreslit obecný Vennův diagram a rozmyslet si, co se děje. Dá se to zase zobecnit na konečný počet množin, ale pak to začne být docela zajímavé a necháme to do kapitoly 11b.

Ted' se podívejme na množiny nekonečné, které asi čtenáře dosud nezasvěceného do magie nekonečna notně překvapí. Začneme faktem, který říká, že nejmenší nekonečné množiny jsou ty spočetné.

Fakt 2c.10.

Nechť A je množina. Jestliže je nekonečná, pak $|\mathbb{N}| \leq |A|$.

Důkaz (náznak): Protože jde o nekonečnou množinu, určitě není prázdná. Vezměme tedy $a_1 \in A$. Pokud něco zbývá v $A - \{a_1\}$, vybereme odtud a_2 . Pokud něco zbývá v $A - \{a_1, a_2\}$, vybereme odtud a_3 a tak dále. Jsou dvě možnosti.

a) Pokud se tento proces někdy zarazí, tak to bude tím, že pro nějaké m je $A - \{a_1, \dots, a_m\}$ prázdná množina. Pak ale $A = \{a_1, \dots, a_m\}$ a podle Faktu 2c.6 (ii) by byla A konečná, což je spor s předpokladem tvrzení, že je nekonečná, čili to nemůže nastat.

b) Určitě tedy nastane druhý případ, kdy najdeme nekonečně mnoho navzájem různých prvků $a_n \in A$. Pak $T(n) = a_n$ je bijekce z \mathbb{N} na $A' = \{a_n; n \in \mathbb{N}\}$, proto $|A'| = |\mathbb{N}|$. Také máme $A' \subseteq A$, proto $|A'| \leq |A|$, zbytek plyne pomocí Faktu 2c.3 (ii).

□

Připomeňme si Větu 2c.7, která nám říkala, že se pojem velikosti chová u konečných množin přesně tak, jak bychom čekali. Následující věta ukáže, že u množin nekonečných je všechno jinak.

Věta 2c.11.

- (i) Každá nekonečná množina má vlastní podmnožinu, která má stejnou mohutnost.
- (ii) Nechť A, B jsou množiny, A je nekonečná a $|B| \leq |A|$. Pak $|A \cup B| = |A|$.
- (iii) Nechť A, B jsou množiny, A je nekonečná a $|B| \leq |A|$. Pak $|A \times B| = |A|$.
- (iv) Nechť A_i pro $i = 1, \dots, m$ nebo $i \in \mathbb{N}$ jsou množiny, kde A_1 je nekonečná, a nechť $|A_i| \leq |A_1|$ pro všechna i . Pak $\left| \bigcup_i A_i \right| = |A_1|$.
- (v) Nechť A_i pro $i = 1, \dots, m$ nebo $i \in \mathbb{N}$ jsou množiny, kde A_1 je nekonečná, a nechť $|A_i| \leq |A_1|$ pro všechna i . Pak $|A_1 \times A_2 \times \dots| = |A_1|$.

Všechny vlastnosti vypadají šíleně. Uberu z množiny prvky a ona zůstane stejně velká. Přidám si k nekonečné množině nějaké prvky (srovnejte s Větou 2c.7 (ii)) a ona je pořád stejně velká. Přidám k nekonečné množině jinou, třeba i disjunktní, třeba i stejně velkou nekonečnou, a ta množina se nezvětší. Dokonce nám (iv) říká, že to nekonečná množina ani velikostně nepozná, když k ní přidám nekonečně (ale spočetně) mnoho takto menších množin. Vlastnosti (iii) a (v) ukazují totéž pro kartéský součin, kde by to člověk čekal ještě méně.

Platí dokonce, že se podle takto divného chování nekonečné množiny poznají: Množina je nekonečná právě tehdy, jestliže má nějakou vlastní podmnožinu stejně mohutnosti.

Tvrzení (iii) lze vyjádřit ještě jinak: Když sjednotíme konečný či spočetný soubor množin, z nichž alespoň jedna je nekonečná, tak má toto sjednocení stejnou mohutnost jako největší ze zúčastněných množin. Stejná věc platí pro kartézský součin.

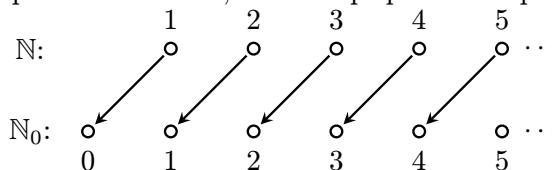
Níže ukážeme, že všechny tyto vlastnosti při bližším pohledu dávají smysl. Problém je v tom, že my se nejsme zvyklí potkávat s nekonečnými množinami, proto si náš mozek nevytvoril příslušné představy. Abychom tedy byli schopni dobře pracovat s mohutností, musíme si naši intuici vycvičit, aby jí ty divné věci příšly normální. To je jedna z věcí, která je na skutečné matematice obtížná, někdy je třeba pracovat ve světech, které se chovají zcela mimo naše představy (mohutnost je ještě v zásadě v pohodě), o to důležitější je pak hlídat si logickou správnost postupů, tvrzení a argumentů v důkazech. Pro většinu lidí je takovéto cvičení vlastního mozku příliš těžké, asi je k tomu třeba nějaká mutace. Možná nejpřekvapivější na tom ale je, že se některé šílené matematické struktury kupodivu vyskytují v převleku kolem nás (teorie relativity, kvantová mechanika).

Důkaz Věty 2c.11 tady dělat nebudeme, místo toho si ukážeme konkrétní případy, kdy k témtu jevům dochází. Pomůže nám to vycvičit naši intuici. Silně doporučujeme následující důkaz nepreskočit, protože to je spíš zamýšlení nad fungováním nekonečnosti.

Věta 2c.12.

- (i) Množina \mathbb{N}_0 je spočetná.
- (ii) Množina \mathbb{Z} je spočetná.
- (iii) Množina $\mathbb{N} \times \mathbb{N}$ je spočetná.
- (iv) Množina $\mathbb{Z} \times \mathbb{Z}$ je spočetná.

Důkaz (poučný, dobrý): (i): Ukážeme, že \mathbb{N}_0 má stejnou mohutnost jako \mathbb{N} . Potřebujeme najít nějakou bijekci $T: \mathbb{N} \mapsto \mathbb{N}_0$, často jako inspirace poslouží obrázek, v tomto případě se nápad docela nabízí.



Formálně definujeme $T(n) = n - 1$. Tvrdíme, že toto zobrazení je bijekce.

Na: Nechť $m \in \mathbb{N}_0$. Pak je m celé číslo splňující $m \geq 0$, proto je $n = m + 1$ celé číslo splňující $n \geq 1$, tedy $n \in \mathbb{N}$, a platí $T(n) = n - 1 = m$.

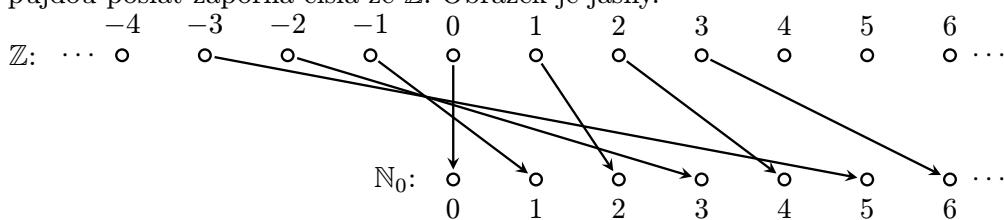
Prosté: Nechť $x, y \in \mathbb{N}$ splňují $T(x) = T(y)$. Pak $x - 1 = y - 1$, tedy $x = y$.

Poznámka: Řečeno hodně nepřesně, klíčovou vlastností nekonečných množin je, že v některém „směru“ nekončí (například rovina nekončí v mnoha směrech, přímka ve dvou). Množinu proto můžeme v takovém směru bez problémů posunout a tím si vytvořit místo pro přidání prvků, aniž by se množina velikostí zvětšila.

V následujícím důkazu množinu \mathbb{N} nejen posuneme, ale zároveň ji rozprostřeme (zředíme), čímž vznikne nekonečně mnoho volných míst.

△

(ii): Ukážeme, že $|\mathbb{Z}| = |\mathbb{N}|$. Protože už máme $|\mathbb{N}| = |\mathbb{N}_0|$, stačí podle Faktu 2c.3 (iii) dokázat, že platí $|\mathbb{Z}| = |\mathbb{N}_0|$. Vytvoříme zobrazení ze \mathbb{Z} na \mathbb{N}_0 následovně. Čísla ze \mathbb{Z}_0^+ pošleme do \mathbb{N}_0 , ale šipky roztáhneme, aby v cíli zbyla čísla, na které půjdou poslat záporná čísla ze \mathbb{Z} . Obrázek je jasný.



Vzoreček: $T(n) = 2n$ pro $n \geq 0$ a $T(n) = 2|n| - 1$ pro $n < 0$. Tvrdíme, že je to bijekce.

Na: Vezměme $m \in \mathbb{N}_0$. Jestliže je sudé, pak $n = \frac{m}{2} \in \mathbb{Z}$ a $n \geq 0$, tudíž podle definice je $T(n) = 2n = m$.

Jestliže je m liché, pak je $m + 1$ sudé, proto $\frac{m+1}{2} \in \mathbb{Z}$. Nechť $n = -\frac{m+1}{2}$. Pak $n \in \mathbb{Z}$. Z $m \geq 1$ máme $\frac{m+1}{2} \geq 1$, tudíž $n < 0$, $|n| = -n = \frac{m+1}{2}$ a podle definice T je $T(n) = 2|n| - 1 = (m + 1) - 1 = m$.

Prostota: Nechť $x, y \in \mathbb{Z}$ splňují $T(x) = T(y)$. Pokud by $x \geq 0$ a $y < 0$, tak by $T(x)$ bylo sudé a $T(y)$ liché a nemohly by se rovnat, tento případ tedy nastat nemůže. Podobně nemůže nastat případ $y \geq 0$ a $x < 0$. Zbývají dva.

Jestliže $x, y \geq 0$, pak $T(x) = T(y) \implies 2x = 2y \implies x = y$.

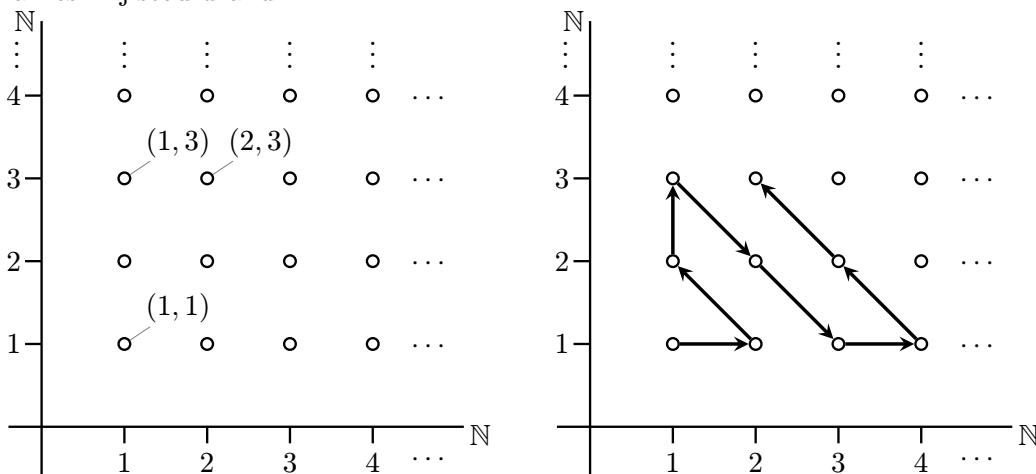
Jestliže $x, y < 0$, pak $T(x) = T(y) \implies 2|x| - 1 = 2|y| - 1 \implies |x| = |y|$. Jenže víme, že jsou obě čísla záporná, tudíž $|x| = |y| \implies -x = -y \implies x = y$. Ve všech případech tedy máme $x = y$ a prostota je také dokázána.

Alternativa: Protože $\mathbb{N} \subseteq \mathbb{Z}$, máme $|\mathbb{N}| \leq |\mathbb{Z}|$. Podle Věty 2c.5 tedy stačí dokázat, že $|\mathbb{Z}| \leq |\mathbb{N}|$, tedy najít nějaké prosté zobrazení ze \mathbb{Z} do \mathbb{N} . Tvrdíme, že $T(n) = 2^{|n|+n} 3^{|n|-n}$ takové je. Nejprve si připomeneme, že pro $n \geq 0$ je $|n| = n$, tedy $|n| + n = 2n$ a $|n| - n = 0$, zatímco pro $n < 0$ je $|n| = -n$ a tedy $|n| + n = 0$ a $|n| - n = -2n$, což je v tomto případě kladné neboli $3^{-2n} \in \mathbb{N}$. Proto vždy $2^{|n|+n} 3^{|n|-n} \in \mathbb{N}$ a navíc vidíme, že $T(0) = 1$, dále $T(n) = 2^{2n}$ pro $n > 0$ a $T(n) = 3^{-2n}$ pro $n < 0$.

Protože čísla s různými prvočíselnými rozklady nemohou být stejná (viz Věta 6b.4), tak hned vidíme, že pro $m \neq n$ také platí $T(n) \neq T(m)$ a proto je T prosté.

Tato alternativa možná není tak pěkně vidět z obrázku jako první důkaz a dá víc práce dokázat prostotu, ale zase je to zobrazení dané jen jedním vzorečkem, což někdy může být výhoda.

(iii): Tady je tradiční důkaz obrázkem. Potřebujeme vytvořit bijekci $T: \mathbb{N} \leftrightarrow \mathbb{N} \times \mathbb{N}$, čili potřebujeme říct, kam pošleme 1, kam pošleme 2 atd. Podívejme se na následující obrázek. Nejdříve jsme vlevo reprezentovali kartézský součin $\mathbb{N} \times \mathbb{N}$ a naznačili význam několika bodů, jen abychom se ujistili, že tomu rozumíme, a pak jsme vpravo nakreslili jistou dráhu.



Ta cestička nám ukazuje, jak postupně vybírat hodnoty pro T . Takže $T(1) = (1, 1)$, $T(2) = (2, 1)$, $T(3) = (1, 2)$, $T(4) = (1, 3)$, $T(5) = (2, 2)$ atd. Při tomto způsobu výběru je jasné, že se hodnoty neopakují, takže T je prosté, a z obrázku se zdá, že dříve či později ta cestička dojde na libovolné místo v té síti, takže T by mělo být i na. Abychom z toho udělali pořádný důkaz, potřebovali bychom toto T přesně definovat, což se dá, ale je to dost komplikované, takže formální důkaz raději uděláme jinak. Nebyl to nicméně ztracený čas, protože tento způsob nám názorně ukázal, že se opravdu může stát, že „dvourozměrná“ síť má stejně bodů jako jednorozměrná. Takovéto názorné představy nám pomáhají vypěstovat správnou intuici o nekonečných množinách.

Formálně korektní důkaz: Nejprve ukážeme, že $|\mathbb{N}| \leq |\mathbb{N} \times \mathbb{N}|$. Použijeme klasický trik a vyrobíme si v rámci $\mathbb{N} \times \mathbb{N}$ věrnou kopii množiny \mathbb{N} zcela přirozeným způsobem, matematicky řečeno množinu \mathbb{N} do toho kartézského součinu vnoríme. Je třeba to udělat správně formálně.

Uvažujme množinu $M = \{(n, 1); n \in \mathbb{N}\}$ (první řádek v té síti). Pak určitě $|M| = |\mathbb{N}|$, což se snadno dokáže přirozenou bijekcí $T(n) = (n, 1)$. A protože $M \subseteq \mathbb{N} \times \mathbb{N}$, tak máme $|M| \leq |\mathbb{N} \times \mathbb{N}|$, zbytek vyplýne pomocí Faktu 2c.3 (ii).

Zajímavější bude druhý směr, kdy zkusíme vnořit zdánlivě „větší“ množinu do „menší“. Potřebujeme najít prosté zobrazení $T: \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$. Uvažujme $T(m, n) = 2^m 3^n$. Evidentně pro $m, n \in \mathbb{N}$ dává $T(m, n) \in \mathbb{N}$, takže jde $\mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$, ještě zbývá ukázat, že je prosté. Nechť tedy $(m, n), (u, v) \in \mathbb{N} \times \mathbb{N}$ splňují $T(m, n) = T(u, v)$. To znamená, že $2^m 3^n = 2^u 3^v$. Jenže celá čísla se dají zapsat pomocí prvočísel jen jediným způsobem, takže musí jít o stejně výrazy, tedy $m = u$ a $n = v$ čili $(m, n) = (u, v)$. Důkaz je hotov.

(iv): Tady je asi nejlepší zkombinovat (ii) a (iii). Jedna možnost je sloučit použité triky a dokázat, že zobrazení $U(m, n) = 2^{|m|+m} 3^{|n|-m} 5^{|n|+n} 7^{|n|-n}$ je prosté $\mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{N}$.

Ukážeme ještě jeden způsob, kde se nemusíme hrabat v detailech, ale pracujeme konceptuálně. Zde využijeme přímo výsledky (ii) a (iii). Podle (ii) víme, že existuje bijekce $S: \mathbb{Z} \rightarrow \mathbb{N}$. Pomocí ní teď definujeme zobrazení $R(m, n) = (S(m), S(n))$. Tvrdíme, že je to bijekce $\mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{N} \times \mathbb{N}$.

Na: Nechť $(u, v) \in \mathbb{N} \times \mathbb{N}$. Protože je S bijekce, určitě existuje $m \in \mathbb{Z}$ splňující $S(m) = u$ a existuje $n \in \mathbb{Z}$ splňující $S(n) = v$. Pak $(m, n) \in \mathbb{Z} \times \mathbb{Z}$ a $R(m, n) = (S(m), S(n)) = (u, v)$.

Prostota: Nechť $(m, n), (u, v) \in \mathbb{Z} \times \mathbb{Z}$ splňují $R(m, n) = R(u, v)$. Pak podle definice R máme $S(m) = S(u)$ a $S(n) = S(v)$ a S je bijekce, tudíž $m = u$ a $n = v$, čili $(m, n) = (u, v)$.

Právě jsme dokázali, že $|\mathbb{Z} \times \mathbb{Z}| = |\mathbb{N} \times \mathbb{N}|$, spolu s (iii) a Faktem 2c.3 (iii) to dává kýžený výsledek. □

Všimněte si, že jsme pracovali s bijekcí S , o které jsme vůbec nevěděli, jak vlastně vypadá, jen jsme se odvolávali na její vlastnosti. Začínáme se stávat matematiky.

Bod (i) ukázal, že odebráním prvku z \mathbb{N}_0 se mohutnost nezměnila, což ukazuje, že u nekonečných množin je opravdu snadné vyrobit vlastní podmnožinu o stejné mohutnosti. Evidentně také nebude problém odebrat i více prvků. Můžeme se na to podívat i z druhé strany, že přidáním jednoho prvku (a tedy indukcí i konečně mnoha prvků) mohutnost nekonečné množiny nezměníme.

Bod (ii) ukazuje, že když dáme dohromady dvě stejně velké nekonečné množiny, tak jim zůstane původní velikost. Víme totiž, že $\mathbb{Z} = \mathbb{N}_0 \cup (-\mathbb{N})$, kde jsme označili $-\mathbb{N} = \{-n; n \in \mathbb{N}\}$. Snadno ukážeme pomocí bijekce $T(n) = -n$, že množina $-\mathbb{N}$ má stejnou mohutnost jako množina \mathbb{N} .

Asi nejzajímavější je (iii), to nám ukazuje dvě věci. Jednak je to příklad toho, že ani kartézským součinem

dvou nekonečných množin nedosáhneme větší mohutnosti, ale dá se to číst i jinak. Pro libovolné $i \in \mathbb{N}$ označme $M_i = \{(n, i); n \in \mathbb{N}\}$, takže třeba $M_1 = \{(1, 1), (2, 1), (3, 1), \dots\}$, zatímco $M_{13} = \{(1, 13), (2, 13), (3, 13), \dots\}$. Jde o disjunktní množiny, které mají všechny stejnou mohutnost jako \mathbb{N} , což se snadno dokáže bijekcemi $T_i(n) = (n, i)$. Když teď uděláme nekonečné sjednocení, dostaneme $\bigcup_{m=1}^{\infty} M_m = \mathbb{N} \times \mathbb{N}$, což je zase množina mohutnosti \mathbb{N} . Je to tedy krásný příklad na (iii) z Věty 2c.11.

Důkazy, které jsme používali, jsou nejen názorné, ale i užitečné, protože tyto nápady se při práci s mohutností používají docela často.

Z (iv) hned plyne toto:

! Věta 2c.13.

Množina racionálních čísel \mathbb{Q} je spočetná.

Důkaz (poučný): Protože $\mathbb{N} \subseteq \mathbb{Q}$, platí $|\mathbb{N}| \leq |\mathbb{Q}|$. Potřebujeme teď opačnou nerovnost, podle Věty 2c.12 (iv) nám ale vlastně stačí ukázat, že $|\mathbb{Q}| \leq |\mathbb{Z} \times \mathbb{N}|$. Vnoření \mathbb{Q} do $\mathbb{Z} \times \mathbb{N}$ pomocí prostého zobrazení se dělá relativně snadno, ale ještě jednodušší je použít alternativní přístup a najít zobrazení T z $\mathbb{Z} \times \mathbb{N}$ na \mathbb{Q} (vi Fakt 2c.1).

Definujme jej takto: $T(p, q) = \frac{p}{q}$. Surjektivita je zjevná, každé racionální číslo lze zapsat jako zlomek $\frac{p}{q}$, kde $p \in \mathbb{Z}$ a $q \in \mathbb{N}$. □

Pozorný čtenář si všimne, že už známe mohutnost \mathbb{N} , \mathbb{Z} a \mathbb{Q} , ale jednu populární množinu jsme ještě nezkoumali. Máme tu také jiný dloužek, definovali jsme nespočetné množiny, ale zatím není vůbec jasné, jestli nějaká taková množina existuje. Tohle může skončit jediným způsobem.

! Věta 2c.14.

Interval reálných čísel $(0, 1)$ je nespočetný.

Důkaz (poučný): Ukážeme, že žádné zobrazení $T: \mathbb{N} \mapsto (0, 1)$ nemůže být na. Pro účely tohoto důkazu si budeme čísla z intervalu $(0, 1)$ zapisovat jako čísla s nekonečným desetinným rozvojem, což si představíme například tak, že u čísel typu 0.347 doplníme dál nuly (teď narázíme na drobné nejasnosti s tím, že třeba $0.1000\dots = 0.0999\dots$, v případě více možných vyjádření jednoho čísla si prostě pro účely tohoto důkazu vždy jeden zápis zvolíme).

Vezměme tedy libovolné zobrazení $T: \mathbb{N} \mapsto (0, 1)$ a ukážeme, že nemůže být na. Zlobivé číslo b vytvoříme takto: Začíná „0.“ a pak doplňujeme desetinné číslice. Číslice na k -té místě se určí následovně: Podíváme se na k -tou cifru v rozvoji čísla $T(k)$ a jestliže je to „3“, tak do našeho čísla b jako k -tou cifru dáme „1“, jinak do našeho čísla dáme „3“. Dostaneme tak číslo b , které začíná „0.“ a tudíž určitě leží v $(0, 1)$. Zároveň se ale od každého $T(k)$ liší na k -té místě rozvoje, tudíž se mu nemůže rovnat. Proto neexistuje $n \in \mathbb{N}$ takové, že $T(n) = b$ a T není na.

Formálně: Zapíšeme obrazy T ve tvaru $T(k) = \sum_{i=1}^{\infty} a_{k,i} 10^{-i}$ a definujeme cifry $b_k = \begin{cases} 1, & a_{k,k} = 3; \\ 3, & a_{k,k} \neq 3, \end{cases}$ pak $b = \sum_{k=1}^{\infty} b_k 10^{-k}$ je ono divné číslo. □

Přiblížíme si obrázkem, jak tento argument funguje, na příkladě jednoho konkrétního T . Jeho hodnoty si vypíšeme do řádků nekonečné tabulky.

$$\begin{array}{r} T(1) = 0 . \boxed{1} 3 8 4 0 \dots \\ T(2) = 0 . 2 \boxed{3} 7 4 0 \dots \\ T(3) = 0 . 6 0 \boxed{0} 0 0 \dots \\ T(4) = 0 . 9 3 8 \boxed{2} 1 \dots \\ T(5) = 0 . 0 8 5 4 \boxed{3} \dots \\ \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \dots \\ \hline b = 0 . 3 1 3 3 1 \dots \end{array}$$

Procházíme diagonálou a do „našeho“ čísla b dáváme vždy něco jiného, čímž zaručíme, že se naše číslo nebude shodovat s žádným řádkem tabulky. Tomuto argumentu se říká „Cantorův diagonální argument“ a při práci s velikostí množin je velice mocný, přitom tak jednoduchý. Všimněte si, že k tomu, aby nám fungoval, stačí mít možnost volit „něco jiného“, čili v zásadě stačí mít dva různé znaky. Mohli jsme tedy namísto desetinného rozvoje použít třeba zápis ve dvojkové soustavě se znaky 0 a 1, fungovalo by to stejně.

Celá tahle záležitost je další z výzev pro naši intuici. Asi všichni žijeme v představě, že když vezmeme papír a začneme pod sebe psát čísla 1, 2, 3, …, tak nakonec (pokud budeme žít věčně) napíšeme všechna přirozená čísla. Člověk by si naivně myslel, že to jde udělat s libovolnou množinou čísel (a výše jsme viděli, že s celými i racionálními čísly ano), ale poslední důkaz ukazuje, že s intervalem $(0, 1)$ to nepůjde, protože i kdybychom opravdu měli nekonečně mnoho času, tak z té množiny obsáhneme jen zanedbatelnou část. Přitom není selským rozumem zjevné, proč by to nemělo jít. Důkaz je ale nemilosrdě jasný, nejde to, tak to musíme akceptovat a zvyknout si na to. Nespočetné množiny jsou a jsou (nepředstavitelně?) velké.

! Důsledek 2c.15.

Množina reálných čísel \mathbb{R} je nespočetná.

Důkaz (rutinní): Víme, že $|\mathbb{N}| < |\langle 0, 1 \rangle|$, také z $\langle 0, 1 \rangle \subseteq \mathbb{R}$ máme $|\langle 0, 1 \rangle| \leq |\mathbb{R}|$, proto $|\mathbb{N}| < |\mathbb{R}|$ (viz cvičení 2c.1). \square

Mimochodem, víme už, že množina \mathbb{Q} je spočetná, což znamená, že množina iracionálních čísel musí být nutně nespočetná (viz cvičení 2c.13). Mezi reálnými čísly je tedy nekonečně mnohokrát víc čísel iracionálních než zlomků.

Jakou má vlastně množina \mathbb{R} mohutnost? Snadno se ukáže, že libovolný interval typu $\langle n, n + 1 \rangle$ má stejnou mohutnost jako $\langle 0, 1 \rangle$. Protože $\mathbb{R} = \bigcup_{n=-\infty}^{\infty} \langle n, n + 1 \rangle$ je sjednocení spočetného souboru intervalů, které mají všechny stejnou mohutnost jako $\langle 0, 1 \rangle$, tak podle Věty 2c.11 (iii) platí $|\mathbb{R}| = |\langle 0, 1 \rangle|$. Mohutnost množiny reálných čísel či intervalu $\langle 0, 1 \rangle$ je další ze základních mohutností, které se objevují často.

Ve cvičení 2c.14 si například rozmyslíme, že pro libovolné $a < b$ má $\langle a, b \rangle$ stejnou mohutnost jak $\langle 0, 1 \rangle$, a protože už víme, že nekonečné množiny jeden bodík nerovnou, tak vlastně stejnou mohutnost mají všechny intervaly $\langle a, b \rangle$, $\langle a, b \rangle$, (a, b) a (a, b) pro $a < b$, přičemž za a, b připouštíme i nekonečna. Mimochodem ta podmínka $a < b$ je podstatná, vylučuje tzv. degenerované intervaly jako $\langle 13, 13 \rangle = \{13\}$ či $\langle 13, 13 \rangle = \emptyset$.

Mezi množinami spočetnými a nespočetnými je podstatný rozdíl při praktické práci. Množiny spočetné mohou být očíslovány, čili zapsány jako $A = \{a_1, a_2, \dots\}$. To se udělá jednoduše, pro spočetnou množinu A existuje bijekce z \mathbb{N} na A , tak prostě označíme $a_n = T(n)$ a už nám a_n dají celou množinu (srovnejte Fakt 2c.6). Můžeme je tedy takto alespoň potencionálně spočítat, proto se tak jmenují. V průběhu počítání přitom pracujeme s konečnými množinami, což je přesně parketa diskrétní matematiky. V kapitole o indukci dokonce uvidíme, jak se pomocí množin konečných dozvědět ledacos o spočetných množinách nekonečných.

Naopak do množin nespočetných nedokážeme pomocí postupného počítání ani pořádně nahlédnout, takže jsou povětšinou mimo dosah metod diskrétní matematiky a budeme se jim vyhýbat. Vyplatí se proto umět již na začátku rychle odhadnout, zda je daný problém rázu spočetného či nikoliv. Zkusíme si to.

! Příklad 2c.b: Množina A kladných lichých čísel je spočetná.

Protože $A \subseteq \mathbb{N}$, máme jasné $|A| \leq |\mathbb{N}|$. Stačí nám tedy dokázat, že $|\mathbb{N}| \leq |A|$, tedy najít prosté zobrazení z \mathbb{N} do A . To je ale snadné, definujeme $T(n) = 2n - 1$. Určitě pro $n \in \mathbb{N}$ dává kladná lichá čísla, takže jde do A .

Je prosté? Nechť $x, y \in \mathbb{N}$ splňují $T(x) = T(y)$. Pak $2x - 1 = 2y - 1$, tedy $x = y$. Ano, je prosté. Tím je důkaz hotov.

Mimochodem, dokonce jsme tím našli bijekci z \mathbb{N} na A .

\triangle

! Příklad 2c.c: Množina A konečných řetězců vytvořených ze znaků 0 a 1 (tzv. binárních řetězců) je spočetná.

Označme si jako A_n množinu binárních řetězců o délce n . Kolik jich je? Na každou pozici máme na výběr ze dvou znaků, celkem je tedy $2 \cdot 2 \cdots 2 = 2^n$ možností. Hlavní teď je, že A_n je konečná.

Protože máme $A = \bigcup_{n=1}^{\infty} A_n$, zajímá nás, co se stane, když sjednotíme spočetně mnoho konečných množin. Na to vlastně nemáme žádný vzorec, buď umíme sjednocovat konečně mnoho konečných množin (Věta 2c.8), nebo nekonečně mnoho nekonečných (Věta 2c.11). Zkusíme si to rozmyslet.

Určitě to bude alespoň spočetná množina, protože kdyby se z každé A_n vzal jeden prvek, tak už máme tolík prvků, kolik je v \mathbb{N} (množiny A_n jsou disjunktní a proto dostáváme různé prvky).

Na druhou stranu nečekáme, že bychom dostali množinu nespočetnou, protože víme z Věty 2c.11, že sjednocením spočetně mnoha spočetně velkých množin dostaneme spočetnou množinu, a naše množiny jsou dokonce menší. Tato úvaha je užitečná a uděláme si ji obecně.

\triangle

Fakt 2c.16.

- (i) Jestliže jsou A_n pro $n \in \mathbb{N}$ nejvýše spočetné množiny, pak je $\bigcup_{n=1}^{\infty} A_n$ nejvýše spočetná.
(ii) Jestliže jsou navíc A_n neprázdné a po dvou disjunktní, pak je $\bigcup_{n=1}^{\infty} A_n$ spočetná.

Důkaz (rutinní): (i): Přidáme si jednu množinu navíc, $A_0 = \mathbb{N}$, pak podle Věty 2c.11 (i) už $\left| \bigcup_{n=0}^{\infty} A_n \right| = |\mathbb{N}|$.

Protože $\bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=0}^{\infty} A_n$, tak $\left| \bigcup_{n=1}^{\infty} A_n \right| \leq \left| \bigcup_{n=0}^{\infty} A_n \right|$ a zbytek je dle Faktu .

(ii): Teď potřebujeme i dolní odhad. Protože jsou A_n neprázdné, existuje v každé nějaký prvek, nazvěme jej a_n . Definujeme zobrazení $T(n) = a_n$, pak určitě $T: \mathbb{N} \mapsto \bigcup_{n=1}^{\infty} A_n$.

Je to prosté zobrazení? Nechť $m \neq n \in \mathbb{N}$. Protože jsou ty množiny po dvou disjunktní, $A_m \cap A_n = \emptyset$, tak nutně $a_m \notin A_n$, tedy i $a_m \neq a_n$, což znamená $T(m) \neq T(n)$. Toto zobrazení je tedy prosté, což dokazuje, že $|\mathbb{N}| \leq \left| \bigcup_{n=1}^{\infty} A_n \right|$.

□

! Příklad 2c.d: Množina A nekonečných binárních řetězců je nespočetná.

Tato množina je evidentně nekonečná, například proto, že obsahuje řetězce 1000..., 0100..., 0010... atd., kterých je spočetně neboli nekonečně mnoho. Zbývá ukázat, že je to množina nespočetná.

Protože jde přímo o řetězce, nabízí se Cantorův diagonální trik. Dokážeme, že žádné očíslování nemůže uspět. Předpokládejme tedy, že jsme se řetězce pokusili očíslovat, můžeme je pak seřadit pod sebe. Následně vytvoříme řetězec, který v seznamu není. Nejprve pro jeden konkrétní příklad:

$$\begin{array}{ll} a_1 : & 0 . \boxed{1} 0 1 0 \dots \\ a_2 : & 0 . 0 \boxed{0} 0 0 \dots \\ a_3 : & 0 . 1 1 \boxed{0} 0 \dots \\ a_4 : & 0 . 0 0 1 \boxed{1} \dots \\ \vdots & \vdots \vdots \vdots \vdots \vdots \dots \\ b : & 0 . 0 1 1 0 \dots \end{array}$$

A teď pořádně: Nechť T je nějaké zobrazení z \mathbb{N} do A . Označme $T(n) = (a_{n,1} a_{n,2} a_{n,3} \dots)$. Definujeme pak prvek $b \in A$ předpisem $b = (1-a_{1,1} 1-a_{2,2} 1-a_{3,3} \dots 1-a_{n,n} \dots)$. Ten se liší od každého prvku $T(n) = (a_{n,1} a_{n,2} a_{n,3} \dots a_{n,n} \dots)$ na n -tém místě, tedy $b \neq T(n)$ pro všechna $n \in \mathbb{N}$, proto T není na.

Ukázali jsme, že není možné vytvořit bijekci z \mathbb{N} na A .

△

Příklad 2c.e: Uvažujme všechny nekonečné řetězce, které je možné vytvořit z malých písmen anglické abecedy. Z nich do množiny A vybereme takové řetězce, které vždy začínají opakováním jednoho konkrétního písmene, v některém místě pak přejdou na jiné a tím už dál pokračují, viz třeba řetězec $ppphhhhh\dots$. Tvrdíme, že množina A takovýchto řetězců je spočetná.

Tady je zajímavý ten dělící bod, zkuseme se odpíchnout od něj. Nechť A_n je množina všech řetězců, které mají změnu hned za pozicí n , tedy n -tý člen je ještě jako ten první, ale následující už jsou jiné (a všechny stejné). Kolik má taková množina prvků? Máme 26 možností, jak začít, a 25 možností, jak dál pokračovat, celkem $26 \cdot 25$, cíli je to množina konečná (a neprázdná). Máme také $A = \bigcup_{n=1}^{\infty} A_n$ a již z definice jsou ty množiny disjunktní (řetězec se nesmí měnit na více místech), proto podle Faktu 2c.16 je A spočetná.

Alternativa: Pro $\alpha \neq \beta \in \{a, b, c, \dots, z\}$ nechť je $A_{\alpha\beta}$ množina všech řetězců, které začínají písmenem α a končí písmenem β . Jak je taková množina velká? Písmeno β může začít od pozice 2, 3, 4, ..., taková množina je tedy spočetná. Máme také $A = \bigcup_{\alpha, \beta} A_{\alpha, \beta}$ a jde o sjednocení konečně mnoha množin, podle Věty 2c.11 (v) je A spočetná.

△

S 2c.17 Jak určovat mohutnost

Při práci s množinami je často užitečné umět rychle odhadnout, jak velká množina to je, jmenovitě určit, zda je konečná, spočetná či nespočetná. Konečné množiny asi každý hravě pozná, takže se zaměříme na množiny nekonečné. Zde je základem znát dobře množiny, které jsme zkoumali výše (\mathbb{Z} , $(0, 1)$, konečné či nekonečné řetězce

atd.) a ještě probereme níže a také pravidla o sjednocení/kartézském součinu spočetných množin atd. Pomocí těchto znalostí pak odhadujeme (či dokonce dokazujeme) mohutnosti množin jiných. Nejčastěji používáme následující tři přístupy.

1) Přímé porovnání se známou množinou.

Někdy množina svou strukturou vyloženě nabízí porovnání s jinou, nám již známou množinou. V množině všech celočíselných násobků 150 má každý prvek tvar $150k$ pro $k \in \mathbb{Z}$, což zjevně nabízí bijekci na množinu \mathbb{Z} předpisem „ $150k \leftrightarrow k$ “. Množina všech matic 2×2 nabízí okamžitou bijekci s prostorem čtyřsložkových vektorů. Množina všech vodorovných přímek v rovině nabízí bijekci na \mathbb{R} danou třeba „přímka \leftrightarrow hodnota průsečíku přímky s osou y “ atd.

Pokud je také třeba odhadnutou mohutnost dokázat, pak stačí ukázat, že ono přiřazení je opravdu bijekce.

2) Další užitečnou strategii je množinu omezit shora či zdola. U situací, kdy je zkoumaná množina nekonečná, její spočetnost dokážeme tak, že její mohutnost shora omezíme pomocí jiné zaručeně spočetné množiny, což se dá často udělat pomocí vztahu být podmnožinou, někdy pomocí prostého zobrazení (tedy už není třeba surjektivita). Například množina všech matic 4×4 v dolním trojúhelníkovém tvaru s celočíselnými prvky je určitě podmnožinou množiny všech matic 4×4 s celočíselnými prvky, která je spočetná díky bijekci na \mathbb{Z}^{16} (zde vlastně kombinujeme strategie 1 a 2). Mimochodem, je zde také možné použít přímo strategii 1 a vyrobit bijekci na množinu \mathbb{Z}^{10} (dolní trojúhelníkové matice 4×4 mají obecně 10 nenulových prvků). Záleží na tom, co je již považováno za známé, spočetnost matic konečné velikosti s celočíselnými prvky je při pokročilejší práci považována za naprostě jasnou, takže bývá jednodušší toho prostě využít.

Nespočetnost pak dokazujeme tak, že množinu omezíme zdola nějakou nespočetnou množinou, opět buď ve smyslu inkluze, nebo prostým zobrazením. Například množina všech nekonečných řetězců ze znaků $\{1, 2, a, c, \diamond\}$ je nespočetná třeba proto, že obsahuje nekonečné řetězce ze znaků $\{1, 2\}$ a o takových jsme si už dokázali, že je nespočetná (my jsme to tedy udělali pro znaky 0, 1, ale to je jen otázka obrázku, který pro ony dva symboly používáme).

Podobně pokud u matic připustíme reálné prvky, tak okamžitě dostaneme množinu nespočetnou, protože určitě obsahuje matice s jedním nenulovým prvkem vpravo nahoře a takových matic je přesně stejně jako reálných čísel evidentní bijekcí „ $r \leftrightarrow$ matice s r vpravo nahoře“.

3) Třetí oblíbenou metodou je rozložit danou množinu na množiny jednodušší, jejichž velikost už snadno rozpoznáme, a pak použít pravidla. Například množina všech čtercových matic s celočíselnými prvky se dá rozložit na spočetně mnoho množin podle velikosti, přičemž pro konkrétní velikost $k \times k$ je množina takových matic celočíselných matic také spočetná, proto je uvažovaná množina jako celek spočetná.

Také strategie 2 a 3 nabízejí v případě potřeby i důkaz a bývá často velice snadný, protože se v úvahách vlastně odvoláváme na již dokázané věty.

Čtenář si tyto strategie může nacvičit ve cvičení 2c.6 a 2c.9.

△

Škála mohutností ovšem nekončí množinou \mathbb{R} , jsou i větší množiny.

Připomeňme si, že je-li A množina, pak $P(A)$ je množina všech jejích podmnožin. Jak je velká, když je A konečná? Při vytváření podmnožin se u každého prvku $a \in A$ můžeme rozhodnout, zda jej vezmeme či ne, a každá odpověď ovlivní výsledek. Celkem je tedy možno udělat $2 \cdot 2 \cdots 2 = 2^{|A|}$ rozhodnutí.

Opravdu? Pro $A = \{1, 2\}$ máme $P(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$, celkem $4 = 2^2$ podmnožin, pro $A = \{1, 2, 3\}$ máme $P(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$, celkem $8 = 2^3$ podmnožin, dál to necháme na čtenáři, asi to opravdu takto funguje. Mimochodem, a co prázdná, jde to i tam? $P(\emptyset) = \{\emptyset\}$, takže má velikost $1 = 2^0 = 2^{|\emptyset|}$. Ano, funguje to.

Fakt 2c.18.

Jestliže je A konečná množina, pak $|P(A)| = 2^{|A|}$.

Vlastně jsme to dokázali, ten argument s výběry je obvykle považován za dostačující. Mimo jiné z toho plyne, že pro konečné množiny $|A| < |P(A)|$, následující věta nám o zobecní.

Věta 2c.19. (Cantorova)

Pro každou množinu A platí $|A| < |P(A)|$.

Důkaz (poučný, možná drsný): Ukážeme, že libovolné zobrazení $T: A \rightarrow P(A)$ nemůže být na. Nechť je tedy T nějaké takové zobrazení.

Když si vezmeme $a \in A$, pak je $T(a) \in P(A)$, čili $T(a)$ je nějaká podmnožina A . Můžeme se zeptat, jestli je a v této množině či ne. Tím vzniká test, pomocí kterého můžeme utvořit podmnožinu.

Nechť $M = \{a \in A; a \notin T(a)\}$. To je podmnožina A , proto $M \in P(A)$. Tvrdíme, že neexistuje žádné $b \in A$ takové, že $T(b) = M$, proto T není na.

Sporem: Předpokládejme, že takové b existuje. Ukážeme, že pak zároveň leží i neleží v M , což je nejde.

A opravdu: Kdyby $b \in M$, pak $b \in T(b)$, proto podle definice této množiny $b \notin M$. Kdyby naopak $b \notin M$, pak splňuje podmínku definice a $b \in M$. Existence takového prvku by tedy vedla k paradoxu, čímž se ukazuje, že to (v rámci standardní teorie množin) není možné. \square

To je velice zajímavé. Začali jsme „nejmenší“ nekonečnou množinou \mathbb{N} (mohutnost spočetné množiny), pak jsme se dozvěděli, že \mathbb{R} má striktně větší mohutnost, podle Cantora má množina všech podmnožin \mathbb{R} značená $P(\mathbb{R})$ zase striktně větší mohutnost, pak můžeme Cantora aplikovat na $P(\mathbb{R})$ a dostaneme ještě větší mohutnost a tak dále, je tedy vidět, že hierarchie velikostí množin nikdy nekončí, vždy je možné vyrobit ještě zase jednu neporovnatelně větší.

2c.20 Poznámka: Víme, že \mathbb{R} má striktně větší mohutnost než \mathbb{N} , to má ovšem podle Cantorovy věty i množina $P(\mathbb{N})$ všech podmnožin \mathbb{N} . Jaký je mezi těmito většími množinami vztah?

Každou podmnožinu M přirozených čísel \mathbb{N} je možné zakódovat pomocí nekonečného řetězce (a_k) ze znaků 0, 1 metodou $a_k = 1$ právě tehdy, pokud $k \in M$ (viz 2a.9 Reprezentace množin). Toto kódování je jednoznačné, takže máme bijekci mezi $P(\mathbb{N})$ a množinou R všech nekonečných řetězců ze znaků 0, 1.

Každý takový řetězec je ovšem možné chápát jako desítkový zápis reálného čísla z množiny $\langle 0, 1 \rangle$ ve dvojkové soustavě. Toto přiřazení je na, ale ne prosté, protože některá čísla lze vyjádřit dvěma způsoby (třeba 0.1000... = 0.0111...). Každopádně vidíme, že množina R nemůže mít větší mohutnost než $\langle 0, 1 \rangle$. Snadno se ale nahlédne, že těch nejednoznačných čísel je jen spočetně, což je vzhledem k velikosti $\langle 0, 1 \rangle$ pod úrovni rozpoznatelnosti. Mohutnost R je tedy stejná jako mohutnost $\langle 0, 1 \rangle$ neboli mohutnost \mathbb{R} .

Závěr: Množina všech podmnožin \mathbb{N} má přesně stejnou mohutnost jako množina \mathbb{R} .

\triangle

Poznámka: Zajímavá otázka je, zda je \mathbb{R} hned ta další velikost nekonečna po spočetnosti, nebo je mezi nimi třeba ještě nějaký mezíkrok. To se zkoumá už přes sto let, Cantor si myslí, že nic mezi není, tomu se říká Hypotéza kontinua, a on se to celý život marně snažil dokázat. Po něm byli i další, až se v 50. letech 20. století ukázalo, že tento fakt je zcela nezávislý na matematické teorii, přesněji řečeno se v rámci klasické teorie množin (ZFC, kterou používáme už někdy od 30. let) nedá ani ukázat, že je HC pravdivá, ani ukázat, že je nepravdivá, je prostě nerovnodnutelná. V zásadě se tedy můžeme rozhodnout, zda ji přijmeme mezi axiomy a dostaneme tím určitou teorii množin, která v sobě nebude obsahovat spory (ta HC ji nepokazí), a přijetím HC se některé věci v té teorii objeví jako pravdivé. Nebo se rozhodneme, že budeme dělat teorii množin bez HC, a pak nám ty věci zase odpadnou. Tím narázíme na problematiku axiomatiky, kterou si raději necháme do kapitoly o uspořádání. Poznamenejme jenom, že pro lidi, kteří s množinami pracují na naší úrovni, je to jedno, rozdíl mezi teoriemi s HC a bez HC nepoznáme.

\triangle

V našich předchozích úvahách jsme odvodili, že podmnožiny \mathbb{N} lze kódovat jako řetězce ze znaků 0, 1, použili jsme (zatím neformálně) pojem posloupnosti, pro číslo $k \in \mathbb{N}$ nám a_k kódovalo přítomnost v dané podmnožině. My jsme již tuto myšlenku poznali ve cvičení 2b.12, kde jsme ji zvedli obecně jako způsob kódování podmnožin dané množiny. Naše úvahy o podmnožinách \mathbb{N} naprostě stejně projdou i pro obecnou množinu M , má tolik podmnožin, kolik jsme schopni vytvořit indikátorových zobrazení. Kolik jich je? TO se dá snadno rozmyslet, tak to rovnou uděláme obecně.

Fakt 2c.21.

Nechť A, B jsou konečné množiny. Množina všech zobrazení $A \rightarrow B$ má mohutnost $|B|^{|A|}$.

Důkaz (poučný): Jak vytváříme zobrazení z A do B ? Pro každý prvek $z a \in A$ se rozhodujeme zcela svobodně, na který prvek z B jej pošleme, máme tedy $|B|$ možností. Pro každý prvek $z A$ tuto volbu opakujeme nezávisle, takže celkem máme tolik možností voleb: $|B| \cdot |B| \cdots |B|$, násobí se tolikrát, kolik je prvků v A . Je tedy celkem $|B|^{|A|}$ možností, jak vytvořit zobrazení z A do B . \square

Tento výsledek naznačí, kde se vzalo následující obecné značení.

Definice.

Nechť A, B jsou množiny. Symbolem B^A značíme množinu všech zobrazení z A do B .

Pro konečné množiny tedy máme $|B^A| = |B|^{|A|}$. Pomocí nového pojmu šikovně zachytíme naše obecné úvahy o mohutnosti množiny podmnožin.

Fakt 2c.22.

Nechť A je množina. Pak $|P(A)| = |\{0, 1\}^A|$.

Důkaz (poučný): Pro každou podmnožinu M množiny A máme zobrazení $\chi_M: A \mapsto \{0, 1\}$ dané $\chi_M(a) = \begin{cases} 1, & a \in M; \\ 0, & a \notin M \end{cases}$ (viz cvičení). Vzniká tím korespondence mezi podmnožinami množiny A a zobrazeními $A \mapsto \{0, 1\}$ neboli zobrazení $T: P(A) \mapsto \{0, 1\}^A$ definované $T(M) = \chi_M$.

Toto zobrazení je na, protože každá indikátorová funkce χ dává podmnožinu M , ze které pochází, jmenovitě množinu tvořenou těmi prvky z A , kde je χ rovna jedné. Je také prosté, protože pokud máme dvě různé podmnožiny M_1, M_2 , pak musí existovat prvek $a \in A$, který je v jedné z nich a není v druhé, a v tom prvku se pak liší i odpovídající indikátorové funkce.

Našli jsme tedy bijekci z $P(A)$ na $\{0, 1\}^A$ a důkaz je hotov. □

Spojíme-li poslední dvě tvrzení, tak vidíme, že pro konečnou množinu A dostáváme $|P(A)| = |\{0, 1\}^A| = 2^{|A|}$, což souhlasí s našimi předchozími závěry.

Poznámka: Vráťme se k Větě 2c.12 a podívame se na ni trochu jinak. Operace s přirozenými čísly se musí v matematice také nějak vytvořit a dělá se to právě v teorii množin velice zhruba takto: Chcete vědět, kolik je $3+2$? Je to velikost množiny $\{1, 2, 3\} \cup \{a, b\}$. Chcete vědět, kolik je $3 \cdot 2$? Je to velikost množiny $\{1, 2, 3\} \times \{a, b\}$. Iterací násobení se pak člověk naučí i m^n , ale dá se to také (viz výše) dělat i přes množinu všech zobrazení z $\{1, \dots, n\}$ do $\{1, \dots, m\}$.

Co by se stalo, kdybychom si zavedli i nekonečno jako kvantitu označující velikost nekonečných množin? Můžeme pak psát $|A| = \infty$ (jakoby číslo), jednotlivá tvrzení z Věty 2c.12 nám pak dávají následující pravidla:

- (i) $\infty + n = \infty$,
- (ii) $\infty + \infty = \infty$,
- (iii) $\infty \cdot n = \infty$ (to se dá i indukcí z (ii) jako opakování sčítání),
- (iv) $\infty \cdot \infty = \infty$,
- (v) $\infty^n = \infty$ (to se dělá z (iv) opakováním násobením).

Cantorova věta ovšem ukazuje, že když mocníme na nekonečno, dostaneme víc: $2^\infty > \infty$, tedy i $\infty^\infty > \infty$.

Upřímně řečeno, v okamžiku, kdy si člověk na nekonečna zvykne, tak mu to začne připadat v zásadě normální a přesně toto by očekával, ostatně se nám podobné vzorečky vylíhnou i v analýze.

V teorii množin se zavádí „kardinální čísla“, což jsou symboly pro mohutnosti množin. Začínají $1, 2, 3, \dots$, po probrání všech přirozených čísel pak přijde velikost spočetných množin značená \aleph_0 a pak přijdou další (větší nekonečna), dají se pak pro ně také zavést počítací pravidla. Jde o hlubokou a náročnou látku, která je samozřejmě zajímavá, ale tohle není kniha o teorii množin.

△

Pro doplnění si představíme ještě jeden pojem.

Cvičení

Cvičení 2c.1 (poučné, zkouškové): Dokažte následující tvrzení:

Nechť A, B, C jsou množiny.

- (i) Jestliže $|A| < |B|$ a $|B| \leq |C|$, pak $|A| < |C|$.
- (ii) Jestliže $|A| \leq |B|$ a $|B| < |C|$, pak $|A| < |C|$.

Cvičení 2c.2 (rutinní): Dokažte, že pro množiny A, B platí $|A \cap B| \leq |A \cup B|$. Kdy je tam rovnost pro konečné množiny?

Cvičení 2c.3 (dobré, poučné): Dokažte, že jestliže $|A| = |B|$, pak $|P(A)| = |P(B)|$.

Cvičení 2c.4 (rutinní, poučné): Dokažte, že jestliže $|A| = |B|$ a $|C| = |D|$, pak $|A \times C| = |B \times D|$.

Cvičení 2c.5 (rutinní, poučné): Dokažte, že jestliže $B \subseteq A$ a B je nekonečná, tak je i A nekonečná.

Cvičení 2c.6 (rutinní, zkouškové): Rozhodněte, zda jsou následující množiny spočetné či ne. Pokud ano, dokažte to.

- (i) Množina záporných celých čísel;
- (ii) množina sudých celých čísel;
- (iii) množina celých násobků 13;
- (iv) množina celých čísel větších než 23;
- (v) množina lichých záporných celých čísel;
- (vi) množina celých čísel, která nejsou násobkem tří;
- (vii) množina racionálních čísel, která jsou mezi 0 a $\frac{1}{2}$;
- (viii) množina všech binárních řetězců neobsahujících 0;
- (ix) množina všechna kladných racionálních čísel, jež nelze napsat pomocí jmenovatele menšího než 4;
- (x) množina reálných čísel neobsahujících 0 v desetinném rozvoji;
- (xi) množina reálných čísel obsahujících pouze konečný počet číslic 1 v zápisu v desítkové soustavě;
- (xii) množina reálných čísel, jejichž zápis v desítkové soustavě obsahuje pouze číslice 1;
- (xiii) množina reálných čísel, jejichž zápis v desítkové soustavě obsahuje pouze číslice 1 nebo 3.

Cvičení 2c.7 (poučné): Uvažujte následující předpis: $T(p, q) = \frac{p}{q}$. Dostáváme tak bijekci z $\mathbb{Z} \times \mathbb{N}$ na \mathbb{Q} ?

Cvičení 2c.8 (poučné): Uvažujte množinu $M = \{n^m; n, m \in \mathbb{N} - \{1\}\}$. Definuje předpis $T(n^m) = (m, n)$ bijekci z M na $(\mathbb{N} - \{1\}) \times (\mathbb{N} - \{1\})$?

Cvičení 2c.9 (poučné, zkouškové): Rozhodněte, zda jsou následující množiny spočetné či ne. Pokud ano, dokažte to.

- (i) Množina matic 2×2 s celočíselnými prvky;
- (ii) množina polynomů, které mají celočíselné koeficienty;
- (iii) množina přímek v rovině;
- (iv) množina přímek vedoucích skrz bod (13, 23);
- (v) množina přímek vedoucích skrz bod (13, 23) s celočíselnými směrnicemi;
- (vi) množina trojúhelníků, jejichž vrcholy mají celočíselné souřadnice;
- (vii) množina trojúhelníků, jejichž strany mají celočíselné délky.

Cvičení 2c.10 (rutinní): Dokažte, že množina všech slov je nejvýše spočetná.

Cvičení 2c.11 (rutinní): Dokažte, že množina všech programů v jistém programovacím jazyce je spočetná.

Cvičení 2c.12 (poučné): Dokažte, že nadmnožina nespočetné množiny je nespočetná.

Cvičení 2c.13 (poučné): Rozhodněte, zda platí následující tvrzení, odpověď dokažte:

Je-li A nespočetná a B spočetná množina, pak musí být $A - B$ nespočetná.

Cvičení 2c.14 (poučné): Dokažte podle definice, že libovolný konečný interval (a, b) pro $a < b$ má stejnou mohutnost jako $(0, 1)$.

Cvičení 2c.15 (poučné): Dokažte podle definice, že množina \mathbb{R} má stejnou mohutnost jako interval $(0, \infty)$.

Cvičení 2c.16 (poučné): Dokažte podle definice, že množina \mathbb{R} má stejnou mohutnost jako interval $(-\frac{\pi}{2}, \frac{\pi}{2})$.

Cvičení 2c.17 (dobré, poučné): Uvažujte množinu $M = \{(a, b) \in \mathbb{N} \times \mathbb{N}; a > b\}$. Na této množině definujeme zobrazení $S(a, b) = (a-1)(a-2) + 2b$. Abychom viděli, jak vlastně S vypadá, uděláme si tabulku jeho hodnot pro kousek M . V řádcích bude a a ve sloupcích b , všimněte si, že pro dané a jsou v M jen dvojice s $1 \leq b < a$. To znamená, že nejmenší možné a , které se v M vyskytuje, je $a = 2$.

$b \rightarrow$	1	2	3	4	5
$a = 2 :$	2				
$a = 3 :$	4	6			
$a = 4 :$	8	10	12		
$a = 5 :$	14	16	18	20	
$a = 6 :$	22	24	26	28	30

Vidíme několik evidentních věcí, toto cvičení bude po vás chtít důkaz toho nejdůležitějšího: Že hodnoty v řádcích rostou a že při přeskoku na další řádek ještě dále vzrostou. Dokažte tedy následující:

- (i) Pro každé $a \geq 2$ a pro každé $1 \leq u < v < a$ platí $S(a, u) < S(a, v)$.
- (ii) Pro každé $a \geq 2$ platí $S(a, a-1) < S(a+1, 1)$.

Poznámka: Pomocí (i) a (ii) už se pak indukcí a pář jednoduchými úvahami dokáže, že S je prosté zobrazení z M do \mathbb{N} . Ještě zajímavější je následující: Zobrazení dané vzorcem $\frac{1}{2}S(a, b)$ je prosté zobrazení z M na \mathbb{N} .

Cvičení 2c.18 (poučné): Použijte to, že je $\frac{1}{2}S(a, b)$ prosté zobrazení z M na \mathbb{N} (viz předchozí cvičení), k důkazu, že zobrazení dané vzorcem $T(m, n) = (m + n - 2)(m + n - 1)/2 + m$ je bijekce z $\mathbb{N} \times \mathbb{N}$ na \mathbb{N} .

Dostáváme tedy přímý předpis vzorečkem pro bijekci $\mathbb{N} \times \mathbb{N} \mapsto \mathbb{N}$ a dokonce je to vzoreček jednoduchý, polynom.

Cvičení 2c.19 (poučné): Dokažte, že zobrazení $U(n) = (3n + 1)^2$ je prosté ze \mathbb{Z} do \mathbb{N} .

Cvičení 2c.20 (poučné): Dokažte pomocí předchozích cvičení, že zobrazení dané předpisem

$$V(m, n) = ((3m + 1)^2 + (3n + 1)^2 - 2)((3m + 1)^2 + (3n + 1)^2 - 1)/2 + (3m + 1)^2$$

je prosté ze $\mathbb{Z} \times \mathbb{Z}$ do \mathbb{N} .

Poznámka: Není známo, zda existuje polynom ve dvou proměnných, který by byl bijekce z $\mathbb{Q} \times \mathbb{Q}$ na \mathbb{N} .

Řešení:

2c.1: (i): Předpoklad dává prosté zobrazení $T: A \mapsto B$ a bijekci $S: B \mapsto C$. Pak je $S \circ T$ prosté $A \mapsto C$ a proto $|A| \leq |C|$.

Kdyby bylo $|A| = |C|$, pak by existovala bijekce $U: A \mapsto C$ a tudíž by $S^{-1} \circ U$ byla bijekce $A \mapsto B$, spor s $|A| < |B|$. Proto $|A| < |C|$.

(ii): Předpoklad dává bijekci $T: A \mapsto B$ a prosté zobrazení $S: B \mapsto C$. Pak je $S \circ T$ prosté $A \mapsto C$ a proto $|A| \leq |C|$.

Kdyby bylo $|A| = |C|$, pak by existovala bijekce $U: A \mapsto C$ a tudíž by $U \circ T^{-1}$ byla bijekce $B \mapsto C$, spor s $|B| < |C|$. Proto $|A| < |C|$.

2c.2: Definujeme $T: A \cap B \mapsto A \cup B$ jako $T(a) = a$. To je evidentně prosté. Obecně platí $A \cap B \subseteq A \cup B$, takže aby platilo $|A \cap B| = |A \cup B|$, muselo by platit $A \cap B = A \cup B$, což je jen když $A = B$.

2c.3: Předpoklad dává bijekci $T: A \mapsto B$. Definujeme $S: P(A) \mapsto P(B)$ jako $S(M) = T[M]$ pro $M \subseteq A$ neboli $M \in P(A)$. S je prosté, protože $S(M) = S(N) \implies T[M] = T[N] \implies T^{-1}T[M] = T^{-1}T[N] \implies M = N$. S je na, pro $N \in P(B)$ je $N \subseteq B$, definujeme $M = T^{-1}[N]$, pak $S(M) = TT^{-1}[N] = N$.

2c.4: Předpoklad dává bijekce $T: A \mapsto B$ a $S: C \mapsto D$. Definujeme $U: A \times C \mapsto B \times D$ jako $U(a, b) = (T(a), S(b))$. Prosté: $U(a, b) = U(x, y) \implies (T(a), S(b)) = (T(x), S(y)) \implies T(a) = T(x) \wedge S(b) = S(y) \implies a = x \wedge b = y \implies (a, b) = (x, y)$, použila se prostota T, S .

Na: Nechť $(x, y) \in C \times D$. T, S jsou na, proto $\exists a \in A: T(a) = x$ a $\exists b \in B: S(b) = y$. Pak $(a, b) \in A \times B$ a $U(a, b) = (T(a), S(b)) = (x, y)$.

2c.5: Je možný důkaz sporem, pomůže Věta 2c.7 (i). Nebo nepřímý, nejprve napsat jako „Nechť $B \subseteq A$. Jestliže je B nekonečná, pak je A nekonečná“, načež tuto implikaci obměnit.

2c.6: (i): Spočetná, je to nekonečná podmnožina spočetné množiny \mathbb{Z} . Alternativa: Přímý důkaz, uvažujme $T(n) = -n$, to je zobrazení z množiny záporných celých čísel do \mathbb{N} , evidentně je na i prosté. Pro úplnost prostota: $T(n) = T(m) \implies -n = -m \implies n = m$.

(ii): Spočetná, je to nekonečná podmnožina spočetných celých čísel. Přímý důkaz bijekcí: zobrazení $T(n) = 2n$ je bijekce ze spočetné množiny \mathbb{Z} na množinu sudých celých čísel. Na: Je-li m sudé, pak $m = 2n$ pro nějaké $n \in \mathbb{Z}$ a $T(n) = m$. Prosté: $T(n) = T(m) \implies 2n = 2m \implies n = m$.

(iii): Spočetná, je to nekonečná podmnožina spočetných celých čísel. Přímý důkaz bijekcí: zobrazení $T(n) = 13n$ je bijekce ze spočetné množiny \mathbb{Z} na množinu celých násobků 13. Na: Je-li m celý násobek třinácti, pak $m = 13n$ pro nějaké $n \in \mathbb{Z}$ a $T(n) = m$. Prosté: $T(n) = T(m) \implies 13n = 13m \implies n = m$.

(iv): Spočetná, je to nekonečná podmnožina spočetné množiny \mathbb{N} . Alternativa: Přímý důkaz, uvažujme $T(n) = n - 23$, to je zobrazení z množiny celých čísel větších než 23 do \mathbb{N} , neboť pak je $T(n)$ celé a $T(n) \geq 1$. Evidentně je na i prosté. Pro úplnost prostota: $T(n) = T(m) \implies n - 23 = m - 23 \implies n = m$.

(v): Spočetná, nechť $T(n) = -(2n + 1)$. Je to zobrazení z \mathbb{N} na lichá záporná čísla, je prosté: $T(n) = T(m) \implies 2n + 1 = 2m + 1 \implies n = m$. Chcete-li zobrazení naopak, zvolte $S(m) = T^{-1}(m) = \frac{1-m}{2}$.

(vii): Spočetná, je obsažena v \mathbb{Q} (tudíž je nejvýše spočetná) a je nekonečná. Zdola se dá velikost odhadnout třeba i tak, že daná množina obsahuje množinu $\{\frac{1}{n}; n \in \mathbb{N}\}$, která je určitě spočetná, protože máme bijekci $T(n) = \frac{1}{n}$ mezi touto množinou a \mathbb{N} .

(viii): Spočetná, jsou to vlastně konečné řetězce jedniček, které se liší délkom, takže můžeme definovat $T(r)$ jako počet jedniček, je to prosté zobrazení na \mathbb{N} .

(ix): Spočetná, je podmnožinou \mathbb{Q} , tudíž nejvýše spočetná, a je nekonečná. Dolní odhad lze udělat i tak, že v dané množině najdeme podmnožinu $\{\frac{2k+1}{8}; k \in \mathbb{N}\}$, podmnožina to určitě je (zlomky nelze zkrátit, tudíž je opravdu nelze napsat se jmenovatelem menším než 4) a spočetná také (bijekce $k \mapsto \frac{2k+1}{8}$).

(x): Nespočetná, daná množina určitě obsahuje například množinu všech reálných čísel, která mají desetinný rozvoj složený z nekonečně mnoha jedniček a dvojek, ta je nespočetná Cantorovým diagonálním argumentem nebo proto, že je díky bijekci stejně velká, jako množina nekonečných řetězců ze znaků 1, 2 neboli množina zobrazení $\mathbb{N} \mapsto \{1, 2\}$, jejíž nespočetnost byla v kapitole dokázána.

(xi): Nespočetná, obsahuje v sobě množinu čísel, která 1 nemají vůbec, a ta je nespočetná, viz předchozí příklad po záměně znaků $0 \mapsto 1$.

(xii): Spočetná, každé takové číslo je jednoznačně dánou dvojicí $(m, n) \in \mathbb{N}_0 \times (\mathbb{N}_0 \cup \{\infty\})$, kde m je počet jedniček před desetinnou tečkou a n počet jedniček za ní. Pozor, dvojice $(0, 0)$ nedává žádné číslo, neboť z ní vyleze jen desetinná tečka, je třeba to ošetřit v definici.

(xiii): Nespočetná. Stačí vzít taková čísla mezi 0 a 1, představit si, že by šlo o spočetnou množinu, a aplikovat na ně Cantora, tedy podívat se na diagonálu a vyměnit 3 a 1.

2c.7: Je evidentně na, ale není prosté $T(2, 4) = \frac{2}{4} = \frac{1}{2} = T(1, 2)$. Takže není bijekce.

2c.8: Je evidentně na, ale není prosté $T(9, 2) = 81 = 3^4 = T(3, 4)$. Takže není bijekce.

2c.9: (i): Spočetná, jasná bijekce $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \mapsto (a_{11}, a_{12}, a_{21}, a_{22}) \in \mathbb{Z}^4$ („seřazení matic do řady“).

(ii): Nejprve ukázat, že množina P_n polynomů stupně n má stejnou mohutnost jako \mathbb{Z}^{n+1} , pomocí $a_n x^n + \dots + a_1 x + a_0 \mapsto (a_n, a_{n-1}, \dots, a_0)$, pak spočetné sjednocení spočetných je spočetné.

(iii): Nespočetná, obsahuje nespočetnou podmnožinu všech vodorovných přímek $\{y = a; a \in \mathbb{R}\}$. Šlo by také zkoušet popsat přímky rovnicemi $ax + by = c$ a pak to porovnat se zaručené nespočetnou množinou zobrazením $ax + by = c \mapsto (a, b, c)$, ale tam je problém s přiřazením, protože jednu přímku lze popsat více rovnicemi. To se dá obejít požadavkem, že čísla a, b, c mají být co nejvíce zkrácena, tedy jejich největší společný dělitel má být 1.

(iv): Nespočetná, tyto přímky lze jednoznačně popsat pomocí směrnice k a těchto směrnic je tolik, kolik je reálných čísel, tedy nespočetně. Vzniká tak bijekce $T: k \mapsto y = k(x - 13) + 23$, což ale není na, chybí svislá přímka, tu doplníme tak, že ji vezmeme jako obraz $T(\infty)$, pak T jde z množiny $R \cup \{\infty\}$, která je nespočetná.

(v): Spočetná, viz (iv), $T: k \mapsto y = k(x - 13) + 23$ je bijekce ze \mathbb{Z} na danou množinu.

(vi): Spočetná, jasná bijekce na \mathbb{Z}^6 .

(vii): Nespočetná, vezmu jeden takový trojúhelník a pak ho mohu posouvat ve směru osy x na tolik pozic, kolik je reálných čísel, vznikne nespočetně mnoho trojúhelníků.

2c.10: Slova jsou konečné řetězce nad českou abecedou, tedy nad konečným počtem symbolů (řekněme, že je jich 82). Množina řetězců z 82 znaků o délce k má 82^k znaků, je tedy spočetná. Konečné řetězce vzniknou sjednocením těchto množin přes všechna $k \in \mathbb{N}$, je tedy spočetná. Takže množina všech konečných řetězců nad 82 písmeny je spočetná a slova tvoří její podmnožinu, tudíž je nejvýše spočetná.

Patrně bude dokonce konečná, protože neexistuje české slovo libovolné délky, třeba padesátipísmenné slovo asi nenajdeme.

2c.11: Každý program lze považovat za konečný řetězec znaků ASCII. Množina programů je tedy podmnožinou množiny konečných řetězců nad konečnou abecedou, což je spočetná množina, viz výše.

2c.12: Toto je jen obměna tvrzení z kapitoly, že podmnožina nejvýše spočetné množiny je nejvýše spočetná.

2c.13: Nejlépe sporem. Kdyby byla $A - B$ spočetná, pak by byla spočetná i $B \cup (A - B) = A \cup B$ a tudíž i $A \subseteq A \cup B$, což je spor.

2c.14: Nechť $T(x) = a + (b - a)x$, pak je to bijekce z $(0, 1)$ na (a, b) . 1) Definice má smysl, pro $0 < x < 1$ je $a < T(x) < b$, tedy opravdu t je do (a, b) . Je na: Dáno $y \in (a, b)$, pak existuje $x = \frac{y-a}{b-a} \in (0, 1)$ takové, že $T(x) = y$. Prostě: $T(x) = T(y) \implies a + (b - a)x = a + (b - a)y \implies x = y$.

2c.15: $T(x) = e^x$ je bijekce $\mathbb{R} \mapsto (0, \infty)$.

2c.16: $T(x) = \arctg(x)$ je bijekce $\mathbb{R} \mapsto (-\frac{\pi}{2}, \frac{\pi}{2})$.

2c.17: (i): Pokud $u < v$, pak $S(a, u) = (a-1)(a-2) + 2u < (a-1)(a-2) + 2v = S(a, v)$.

(ii): $S(a, a-1) = (a-1)(a-2) + 2(a-1) = a^2 - a < a^2 - a + 2 = a(a-1) + 2 = S(a+1, 1)$.

2c.18: Nechť $R: \mathbb{N} \times \mathbb{N} \mapsto M$ je dané $R(m, n) = (m+n, m)$. Je to bijekce. Prostě: $T(m, n) = T(u, v) \implies (m+n, m) = (u+v, u) \implies m+n = u+v \wedge m = u \implies m = u \wedge n = v \implies (m, n) = (u, v)$. Je na: Pro dané $(x, y) \in M$ je $x, y \in \mathbb{N}$ a $x > y$, proto $(m, n) = (y, x-y) \in \mathbb{N} \times \mathbb{N}$ a $T(m, n) = (x, y)$.

Proto je také bijekcí $T = \frac{1}{2}S \circ R$ a $R(m, n) = \frac{1}{2}S(m+n, m) = \frac{1}{2}(m+n-1)(m+n-2) + m$ přesně dle zadání.

2c.19: Pro $n \in \mathbb{Z}$ nelze mít $3n+1=0$, proto $U(n) \in \mathbb{N}$

Nechť $U(x) = U(y)$. Pak $(3x+1)^2 = (3y+1)^2$, tedy $|3x+1| = |3y+1|$. Jaké jsou možnosti? Rozebereme si to podle znamének. Pokud by byla různá, tak jednu absolutní hodnotu odstraníme a druhou nahradíme mínusem, dostaneme tedy $3x+1 = -(3y+1)$. Do dává $3(x+y) = -2$, ale to nejde, protože $3(x+y)$ je celé číslo dělitelné třemi.

Znaménka tedy musí být stejná, pak se dají absolutní hodnoty odstranit, kdyby náhodou byla obě záporná, tak se obě absolutní hodnoty nahradí mínusy a ty se zkrátí, čili každopádně dostaneme $3x+1 = 3y+1$, tedy $x = y$ a prostota je dokázána.

2c.20: Definujme $W(m, n) = (U(m), U(n))$, ak je W prosté $\mathbb{Z} \times \mathbb{Z} \mapsto \mathbb{N} \times \mathbb{N}$, viz např. cvičení 2c.4. Podle předchozích cvičení je pak $T \circ W$ prosté $\mathbb{Z} \times \mathbb{Z} \mapsto \mathbb{N}$ a $(T \circ W)(m, n) = T((3n+1)^2, (3n+1)^2) = V(m, n)$.

3. Binární relace

Vzájemné vztahy mezi objekty se vyskytují velice často. Některá čísla se rovnají, jiná ne, některá jsou menší než jiná. Někteří lidé se spolu znají a jiní ne, některé jídlo vyžaduje jistou surovинu a jiné ne. Takové situace teď budeme chtít zachytit. Existující vztahy lze vyjádřit matematicky vytvářením uspořádaných dvojic. Vyjdeme ze dvou množin objektů, třeba A jsou druhy ovoce a B jsou vitamíny, a to, že určité ovoce obsahuje určitý vitamín, zachytíme tak, že si příslušnou dvojici (ovoce, vitamín) schováme do nějaké množiny R (jako relace).

Takto jsme to dělali již u zobrazení, ale tam jsme měli pro výslednou množinu dvojic výrazné omezení, které zrovna u našeho příkladu s ovocem neplatí, protože jedno ovoce může mít více vitamínů. Zde tedy žádná omezení dělat nebude, čímž vznikne naprostě odlišný typ objektu. Budou nás proto na relacích zajímat jiné věci, než jsme zkoumali u zobrazení, ta už měla svou vlastní kapitolu.

3a. Binární relace a operace s nimi

! Definice.

Nechť A, B jsou množiny. Libovolná podmnožina $R \subseteq A \times B$ se nazývá **relace z A do B** .
 Jestliže $(a, b) \in R$, pak to značíme aRb a řekneme, že a je v **relaci s b** vzhledem k R .
 Jestliže $(a, b) \notin R$, pak řekneme, že a **není v relaci s b** vzhledem k R .

By a **relation** from a set A to a set B we mean an arbitrary subset R of the Cartesian product $A \times B$.

When $(a, b) \in R$, we also denote it aRb for short and say that a is **related to b** by R .

Oba způsoby značení, aRb i $(a, b) \in R$, jsou zcela rovnocenné a autor obvykle volí to, o kterém si myslí, že v daném kontextu rychleji předá čtenáři sdělovanou myšlenku. První značení je kratší a v mnoha situacích výrazně snažší, jsou ale věci, které se jím vyjadřují obtížně. Občas tedy nezbyde než si připomenout, že vlastně „ aRb “ není nic jiného než pohodlná zkratka pro $(a, b) \in R$, protože konec konců z matematického hlediska relace nejsou nic jiného než množiny dvojic.

Příklad 3a.a: Nechť A je množina všech živých lidí a B je množina všech existujících zájmů a koníčků. Definujeme relaci R jako množinu všech dvojic (a, b) takových, že a má b jako koníčka. Pokud bychom chtěli použít ten druhý zápis, definovali bychom tuto relaci takto: „ aRb právě tehdy, jestliže se a zabývá koníčkem b .“

Například $(\text{pH, čtení}) \in R$ neboli $(\text{pH})R(\text{čtení})$, zde je evidentně první zápis lepší.

Toto určitě nebude zobrazení, sice je to podmnožina $A \times B$, ale nesplňuje podmínu z definice, protože například autor této knihy má více než jeden koníček. Někteří lidé naopak nemají žádného.

△

Příklad 3a.b: Uvažujme množiny A a B a zobrazení T z A do B . Pak je to relace z A do B . Je to vlastně samozřejmé, podle definice je každé zobrazení podmnožinou $A \times B$ splňující určité podmínky, a všechny podmnožiny $A \times B$ jsou relace. V jistém smyslu se dá říct, že relace jsou zobecněním zobrazení, protože nám umožňují poslat z jednoho a více šipek nebo také žádnou.

△

Často porovnáváme objekty stejného typu, tedy množiny A a B jsou stejné.

! Definice.

Nechť A je množina. Řekneme, že R je relace na A , jestliže je to relace z A do A .

By a **relation** on a set A we mean an arbitrary relation from A to A .

Příklad 3a.c: V kapitole 2 jsme vlastně zavedli dvě relace, když jsme porovnávali množiny pomocí rovnosti a inkluze. Jak se to udělá formálně? Relace srovnávají objekty z určitých souborů A a B , zde chceme porovnávat množiny, a to mezi sebou, takže si zvolíme nějaký soubor \mathcal{A} množin, které hodláme porovnávat. Pro množiny $M, N \in \mathcal{A}$ pak můžeme zavést relaci R předpisem $(M, N) \in R$ právě tehdy, když $M = N$. Dostáváme tak relaci na \mathcal{A} . Formálně jako množina dvojic zapsáno například takto:

$$R = \{(M, N); M, N \in \mathcal{A} \wedge M = N\}.$$

Takto jsme vyhověli definici a udělali to správně, ale v praxi se prostě řekne, že uvažujeme relaci $M = N$ na souboru množin \mathcal{A} . Často se \mathcal{A} dostane tak, že vezmeme nějakou univerzální množinu U a jako \mathcal{A} bereme všechny

podmnožiny U , tedy $\mathcal{A} = P(U)$. Pak vlastně porovnáváme všechny množiny, které v univerzu U dokážeme vytvořit, běžně říkáme, že uvažujeme relaci $M = N$ na podmnožinách U .

Podobně můžeme ne zcela správně, ale srozumitelně říct, že uvažujeme relaci $M \subseteq N$ na souboru množin \mathcal{A} , formálně bychom to zavedli třeba takto:

$$R = \{(M, N) \in \mathcal{A} \times \mathcal{A}; M \subseteq N\}.$$

Ale psát MRN a překládat si to pokaždé jako $M \subseteq N$ je opravdu zbytečná ztráta času, proto se to nedělá (je ale důležité vědět, co se za tím srozumitelným značením $M \subseteq N$ schovává, když na něj budeme chtít aplikovat nějaké relační triky).

△

Příklad 3a.d: Nechť je U nějaké universum a uvažujme množinu $\mathcal{A} = P(U)$ všech jeho podmnožin. V kapitole 2c jsme definovali dvě relace na \mathcal{A} , jmenovitě relaci rovnosti množin $|M| = |N|$ a relaci $|M| \leq |N|$.

△

Ani v případě známých číselných množin a známých vztahů obvykle nedodržujeme přesně formální jazyk. Například řekneme „uvažujme relaci $=$ na \mathbb{R} “ a míňme tím, že uvažujeme relaci R na \mathbb{R} danou podmínkou xRy právě tehdy, když $x = y$. Evidentně by jen komplikovalo naši práci, kdybychom místo $=$ psali R .

Příklad 3a.e: Nechť A je množina všech počítačů na světě. Definujeme relaci R na A takto: aRb pro $a, b \in A$ jestliže jsou spolu a, b propojeny tak, aby si mohly vyměňovat informace.

△

Příklad 3a.f: Nechť R je relace na množině \mathbb{R} definovaná podmínkou: xRy právě tehdy, když $x + y = 7$.

Například $3R4$, protože $3 + 4 = 7$. Také $8R(-1)$, $\frac{13}{3}R\frac{8}{3}$ nebo třeba $(7 - \sqrt{99})R\sqrt{99}$.

Pokud je relace zadána vztahem, který nám není úplně jasný, vyplatí se najít si příklady dvojic, které v dotyčné relaci jsou, a také dvojice, které napak v relaci nejsou.

Formálně bychom napsali

$$R = \{(x, y) \in \mathbb{R}^2; x + y = 7\}.$$

△

Reprezentace relací.

Často si relace z A do B přibližujeme tak, že je znázorníme orientovaným grafem: Nakreslíme prvky množiny A , prvky množiny B a za každou dvojici $(a, b) \in R$ uděláme orientovanou šipku z a do b .

Evidentně budeme mít problém, když je některá z množin nekonečná. Jsou nicméně případy, kdy alespoň konečný kus nekonečného grafu ledacos naznačí, reprezentace relací pomocí grafů totiž hovoří přímo k naší intuici.

Jde-li o relaci na A , tedy z A do A , tak je zvykem nakreslit množinu A jen jednou a do ní vpisovat šipky z a do b za dvojice (a, b) z relace. U příkladu s počítači bychom tak dostali pěkné znázornění propojenosti počítačů.

Relace (s konečnými množinami) se dá také reprezentovat **tabulkou**. Každý prvek z A má svůj rádek, sloupce jsou pro prvky z B a pokud relace aRb existuje, tak uděláme křížek.

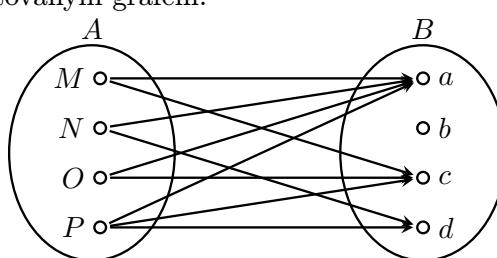
Příklad 3a.g: Uvažujme malou školu se studenty **Mordred**, **Nazgûl**, **Othello** a **Pippin**, škola nabízí kurzy algoritmizace, banalit, cyklistiky a diskrétní matematiky. Mordred si zapsal algoritmizaci a cyklistiku, Othello totéž, Nazgûl zkusil algoritmizaci a diskrétku a Pippin algoritmizaci, cyklistiku a diskrétku.

Zapišeme to jako relaci. Budeme mít množinu studentů $A = \{M, N, O, P\}$ a množinu předmětů $B = \{a, b, c, d\}$, informace zachytíme v následující relaci R z A do B :

$$R = \{(M, a), (M, c), (N, a), (N, d), (O, a), (O, c), (P, a), (P, c), (P, d)\}.$$

Vyjádření tabulkou a orientovaným grafem:

	a	b	c	d
M	×		×	
N	×			×
O	×		×	
P	×		×	×



Mimochodem, tato relace není zobrazením, protože některé prvky z A (jmenovitě všechny) se vyskytují ve více dvojicích.

△

Jak tento příklad naznačuje, relace jsou jedním z možných matematických modelů pro reprezentaci databází.

! 3a.1 Reprezentace relací v počítačích.

V tabulce výše se dají namísto křížků dělat 0 a 1, čímž vznikne něco jako matici. Matice jsou standardním matematickým objektem, se kterými se v počítačích běžně pracuje, čímž dostaváme jednu z populárních reprezentací relací v počítačích. Když si ale takovou matici představíme, tak si jistě všimneme, že v ní chybí informace o tom, které konkrétní prvky nějaká konkrétní jednička spojuje.

Nedílnou součástí maticového zápisu relace je tedy také domluva, kterému objektu odpovídá který řádek a sloupec matice. Matematicky řečeno, je třeba zvolit očislování množin A a B a pak se ho držet. Když začneme s konkrétními objekty $A = \{a_1, a_2, \dots, a_m\}$ a $B = \{b_1, b_2, \dots, b_n\}$, tak se stačí odvolávat na jejich indexy, takže vlastně pracujeme s relací z $\{1, 2, \dots, m\}$ do $\{1, 2, \dots, n\}$.

! Definice.

Nechť $A = \{a_1, a_2, \dots, a_m\}$ a $B = \{b_1, b_2, \dots, b_n\}$ jsou množiny. Pro relaci R z A do B definujeme **matici relace** $M_R = (m_{ij})_{i,j=1}^{m,n}$ předpisem

$$m_{ij} = \begin{cases} 1, & (a_i, b_j) \in R; \\ 0, & (a_i, b_j) \notin R. \end{cases}$$

Hned vidíme, že pro relace na množině bude tato matice čtvercová. Protože matice vznikající z relací jsou speciální a budeme se chtít omezit jen na ně, dáme jim také speciální jméno.

Definice.

Jako **01-matice** budeme označovat matice, které mají pouze prvky 0 či 1.

By a **zero-one matrix** we mean a matrix whose elements are only 0 or 1.

Příklad 3a.h (pokračování 3a.g): Vrátíme se k naší malé třídě. Pokud zachováme přirozené pořadí studentů

$A = \{M, N, O, P\}$ a předmětů $B = \{a, b, c, d\}$, pak se diskutovaná relace zachytí maticí $M_R = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}$.
△

Příklad 3a.i: Uvažujme relaci R z $A = \{1, 3, 4\}$ do $B = \{2, 3, 4, 5\}$ definovanou předpisem $(a, b) \in R$ právě tehdy, když $a > b$. Rozmyslíme si, že pak dostaváme relaci $R = \{(3, 2), (4, 2), (4, 3)\}$.

Jestliže si k ní chceme udělat matici, tak si musíme uvědomit, že při tom používáme pořadového čísla prvku v množině. To například znamená, že dvojice $(3, 2)$ je vlastně druhý prvek z A spojený s prvním prvkem z B , takže se

to v matici objeví jako 1 na pozici $(2, 1)$. Podobně si rozmyslíme další dvojice a dostaváme $M_R = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$.

Je to vlastně lehké, je to jako bychom dělali tabulkou, představíme si prvky z A napsané podél levé hrany matice a prvky z B podél horní hrany matice a pak už je to jasné.
△

Všimněte si, že jsme dostali matici, která je tzv. řídká, tedy má většinou nuly. Pamatovat si takovou informaci jako matici je plýtváním a stojí za to zvážit jiný způsob uchování v počítači, například jako seznam uspořádaných dvojic, populární je také reprezentace spojovým seznamem. Tyto způsoby mají zase nevýhodu v tom, že maticový zápis si docela dobře rozumí s operacemi nad relacemi a s vlastnostmi relací, zatímco se seznamem dvojic se některé věci dělají hůře. Jako obvykle je to něco za něco, to je ale téma pro kurs o databázích.

Příklad 3a.j: Nechť A je množina všech lidí. Pro $a, b \in A$ definujeme aRb jestliže je b rodič a (pro rýpaly: pro účely této definice tím míníme, že mu poskytl(a) genetickou informaci).

Zrovna toto je relace, která by měla velice řídkou matici. Pokud bychom tak chtěli zapsat relaci pro Prahu, šlo by o matici s cca milionem řádků i sloupců, tedy asi 10^{12} dat k uložení, ale ve skutečnosti jsou v každém řádku pouze dvě jedničky (či ještě méně, pokud už rodič umřel či není z Prahy), takže ve skutečnosti je to nejvýše cca $2 \cdot 10^6$ dat. Určitě to tedy bude chtít pro uložení i manipulaci vymyslet nějaký efektivní způsob.

△

! 3a.2 Operace

Interakce mezi reálnými objekty se při matematickém vyjádření projeví jako operace a relace nejsou výjimkou. Nechť R_1, R_2 jsou relace z nějaké množiny A do nějaké množiny B . Protože R_1, R_2 jsou podmnožiny stejného universa, jmenovitě $A \times B$, můžeme na ně aplikovat všechny běžné **množinové operace**. Interpretace bývá většinou jasná na první pohled (přinejhorším na druhý).

Příklad 3a.k: Nechť je A množina všech měst (v České republice, aby jich nebylo tolik). Nechť R_1 je relace definovaná tak, že aR_1b právě tehdy, jestli se dá z a do b dostat autobusem, a R_2 je relace definovaná tak, že aR_2b právě tehdy, jestli se dá z a do b dostat vlakem.

Pak relace $R_1 \cup R_2$ udává, zda se dá z a do b dostat autobusem či vlakem, relace $R_1 \cap R_2$ udává, zda se dá z a do b dostat vlakem i autobusem, a relace $R_1 - R_2$ udává, zda se dá z a do b dostat autobusem, ale ne vlakem. Relaci $R_2 - R_1$ si rozmyslíte sami.

△

Co to znamená, že $R_1 \subseteq R_2$? Podle definice to znamená, že platí implikace $(a, b) \in R_1 \implies (a, b) \in R_2$, tedy $aR_1b \implies aR_2b$. Slovy, jestliže jsou dva prvky v relaci vzhledem k R_1 , pak už jsou nutně i v relaci vzhledem k R_2 . Intuitivně řečeno, relace R_1 je „silnější“ ve smyslu, že už vynucuje tu druhou. Takovéto hierarchie jsou občas zajímavé, třeba relace „být synem/dcerou“ je podmnožinou relace „být příbuzný“.

Ještě zbývá operace doplňku, na to je třeba říct, co je vlastně universum. Zde je kandidát jasný, maximální možná relace $A \times B$.

Definice.

Nechť R je relace z nějaké množiny A do nějaké množiny B . Definujeme její **doplněk** či **komplementární relaci (complementary relation)** jako relaci

$$\bar{R} = \{(a, b) \in A \times B; (a, b) \notin R\}.$$

Zjednodušeně řečeno, pokud vnímáme aRb jako „zná ho“, pak $a\bar{R}b$ je „nezná ho“.

Je přirozené chtít také relace obracet.

Definice.

Nechť R je relace z nějaké množiny A do nějaké množiny B . Definujeme **relaci inverzní k R** , značeno R^{-1} , jako relaci z B do A danou

$$R^{-1} = \{(b, a); (a, b) \in R\}.$$

Let R be a relation from a set A to a set B . We define its **inverse relation**, denoted R^{-1} , as the relation from B to A defined by

$$R^{-1} = \{(b, a); (a, b) \in R\}.$$

Význam je jasný, podle definice $bR^{-1}a$ právě tehdy, když aRb . Jde tedy o obrácení šipek v grafu, například pro relaci R s dvojicemi (dítě,rodič) bude inverzní relace R^{-1} obsahovat dvojice (rodič,dítě). Všimněte si, že na rozdíl od zobrazení zde nejsou žádné podmínky na výslednou množinu dvojic, takže obracet můžeme libovolné relace.

Příklad 3a.l:

Uvažujme množinu $A = C^2(\mathbb{R})$ všech reálných funkcí, které mají na \mathbb{R} spojité derivace alespoň druhého rádu. Definujme relaci R na A podmínkou fRg pokud $f' = g$. Je to tedy relace, která spojuje funkce a jejich derivace, například $\sin(x)R\cos(x)$ nebo x^2R2x . Protože je derivace jednoznačně dána, jde také o zobrazení. Rozdíl bude podstatný ve chvíli, kdy zatoužíme tento vztah otočit. Protože více různých funkcí může mít stejnou derivaci, nejde o zobrazení prosté a tudíž k němu neexistuje inverzní zobrazení.

Pokud se ale na tento vztah podíváme jako na relaci, pak problém není a hravě najdeme inverzní relaci R^{-1} . Podle definice,

$$fR^{-1}g \iff gRf \iff g' = f \iff g \in \int f.$$

Inverzní relace k R tedy spojuje funkce s jejich primitivními funkcemi.

Relaci samozřejmě vůbec nevadí, že k jedné funkci může být přiřazeno více primitivních funkcí.

△

Zvídavého čtenáře napadne, že se může stát, že potřebujeme aplikovat operaci inverzní relace na nějaký komplikovanější výraz, třeba na relaci $R \cup S$. Platí pak nějaká pravidla, která by nám situaci ulehčila, v ideálním případě pomohla získat $(R \cup S)^{-1}$ čistě na základě znalosti R^{-1} a S^{-1} ? Určitě ano, ale jak už jsme psali, není to něco, co bychom se měli učit nazepaměť, spíš je to příležitost si vyzkoušet, zda novým pojmem dobře rozumíme. Jinými slovy, zveme čtenáře k nahlédnutí do cvičení, jmenovitě 3a.5 a 3a.6.

I u relací má smysl je spojovat, pokud na sebe navazují.

! Definice.

Nechť R je relace z nějaké množiny A do nějaké množiny B a S je relace z B do nějaké množiny C . Definujeme jejich **složení** $S \circ R$ jako relaci z A do C definovanou pro $a \in A$, $c \in C$ jako

$$(a, c) \in S \circ R \text{ právě tehdy, když existuje } b \in B: [(a, b) \in R \wedge (b, c) \in S].$$

If R is a relation from a set A to a set B and S is a relation from B to a set C , we define the **composite** of R and S , denoted $S \circ R$, as the relation consisting of ordered pairs (a, c) such that there exists $b \in B$ so that $(a, b) \in R$ and $(b, c) \in S$.

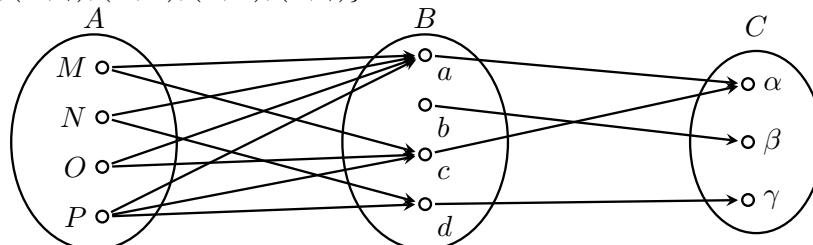
Smysl této operace je jednoduchý. Pokud na sebe dvě relace „navazují“ $A \xrightarrow{R} B \xrightarrow{S} C$, pak složením vynecháme prostředníka. Zkráceným zápisem: V situaci aRb, bSc vložíme do složené relace $S \circ R$ dvojici (a, c) . Všimněte si, že zase máme obrácené pořadí, při složení $S \circ R$ se dělá nejdříve R . Už jsme o tom mluvili u zobrazení, kde pro to byl docela dobrý důvod. Tady ten důvod ovšem odpadá, jedinou omluvou je, že by bylo podivné, kdyby to pro relaci bylo jinak než pro zobrazení, která jsou vlastně také relacemi. To je ale chabá výmluva, takže někteří autoři u relací používají „přirozené“ pořadí (tedy naopak než my zde), což je docela zmatek, některé obory computer science si pak raději zavádějí vlastní značení (které má pořadí rozumně), například toto: $R;S$. Protože ale žádné z těch značení není vnímáno jako všeobecně přijímané, budeme se (s nechutí) držet toho asi standardního $S \circ R$. Jednu věc si ale ulehčíme: Pro navazující kroky aRb a bSc budeme prostě psát $aRbSc$.

! Příklad 3a.m (pokračování 3a.g): Připomeňme, že $A = \{M, N, O, P\}$ jsou studenti, $B = \{a, b, c, d\}$ kurzy a relace $R = \{(M, a), (M, c), (N, a), (N, d), (O, a), (O, c), (P, a), (P, c), (P, d)\}$ říká, který student si zapsal jaký kurs.

Teď přidejme množinu učitelů $C = \{\alpha, \beta, \gamma\}$ a relaci $S = \{(a, \alpha), (b, \beta), (c, \alpha), (d, \gamma)\}$, která říká, který kurs je učen kterým učitelem.

Dvojice (x, z) patří do $S \circ R$ právě tehdy, jestliže existuje kurs y takový, že si student x zapsal y a vyučující z ho učí. To znamená, že $S \circ R$ je relace, která přiřazuje vyučující ke studentům. Řečeno jinak, pro zvolené $x \in A$ nám množina $\{z \in C; (x, z) \in S \circ R\}$ říká, kdo všechno bude studenta v tomto semestru učit.

Podle řetízků $MRaS\alpha$, $MRcS\alpha$, $NRaS\alpha$, $NRdS\gamma$, $ORaS\alpha$, $ORcS\alpha$, $PRaS\alpha$, $PRcS\alpha$, $PRdS\gamma$ dostáváme $S \circ R = \{(M, \alpha), (N, \alpha), (N, \gamma), (O, \alpha), (P, \alpha), (P, \gamma)\}$.



Definovali jsme také inverzní relaci, pro R dostáváme

$$R^{-1} = \{(a, M), (c, M), (a, N), (d, N), (a, O), (c, O), (a, P), (c, P), (d, P)\},$$

což nám pro určitý kurs říká, kdo na něj chodí.

△

Je zřejmé, že o komutativitě této operace nemůže být řeč už z principu: Základní podmínkou pro skládání je, aby cílová množina první relace byla stejná jako výchozí množina druhé relace, což se ovšem snadno ukáže, když zkusíme pořadí relací zaměnit: $B \xrightarrow{S} C$, $A \xrightarrow{R} B$. Ale i v případě, že množiny navazují, ještě nemusí opačné pořadí při složení relací dát totéž. Komutativitu proto u skládání neřešíme.

Kdo přečetl kapitolu 2b, tak už ví, co přijde: Zavedeme skládání více navazujících relací, například tří v situaci $A \xrightarrow{R} B \xrightarrow{S} C \xrightarrow{T} D$. Jak už jsme viděli u zobrazení, základním předpokladem, aby to vůbec udělat šlo, je dokázat asociativitu.

Fakt 3a.3.

Nechť R je relace z nějaké množiny A do nějaké množiny B , S je relace z B do nějaké množiny C a T je relace z C do nějaké množiny D . Pak $(T \circ S) \circ R = T \circ (S \circ R)$.

Důkaz (rutinní): 1) Dokážeme, že $(T \circ S) \circ R \subseteq T \circ (S \circ R)$.

Uvažujme $(a, d) \in (T \circ S) \circ R$. Podle definice to znamená, že $\exists b \in B: aRb$ a $(b, d) \in T \circ S$. Z toho druhého faktu pak máme, že $\exists c \in C: bSc$ a cTd . Z dvojice aRb a bSc vyplývá, že $(a, c) \in S \circ R$, spolu s cTd dostaneme $(a, d) \in T \circ (S \circ R)$.

2) Opačnou inkluzi dokážeme obdobně. □

Opakovaným použitím tohoto faktu dostaneme například to, že výrazy $(U \circ T) \circ (S \circ R)$, $((U \circ T) \circ S) \circ R$ či třeba $U \circ ((T \circ S) \circ R)$ dávají totéž, takže podobně jako u násobení můžeme závorkování vynechat a prostě napsat $U \circ T \circ S \circ R$.

Souvislost mezi inverzní relací a skládáním je jako u zobrazení, srovnejte s Větou 2b.6.

Věta 3a.4.

Nechť R je relace z množiny A do množiny B a S je relace z B do množiny C . Pak $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$.

Důkaz (rutinní): Relace $S \circ R$ jde z A do C , takže $(S \circ R)^{-1}$ jde z C do A . To je stejný směr, kterým jde složená relace $R^{-1} \circ S^{-1}$, vznikla totiž z řetízku $C \xrightarrow{S^{-1}} B \xrightarrow{R^{-1}} A$. Množiny tedy souhlasí a víme, s jakými relacemi budeme pracovat.

Relace jsou množiny dvojic, takže jejich rovnost dokážeme argumentem, že mají stejné prvky (dvojice). Zkusíme to dokázat oběma směry najednou, tedy ekvivalentními kroky.

$(c, a) \in (S \circ R)^{-1}$ právě tehdy, když $(a, c) \in S \circ R$, což je právě tehdy, když $\exists b \in B: aRb$ a bSc . Podle definice inverzních relací to je právě tehdy, když $\exists b \in B: cS^{-1}b$ a $bR^{-1}a$, což je právě tehdy, když $(c, a) \in R^{-1} \circ S^{-1}$. □

Poznámka o matematice: Všimněte si, že důkaz poslední Věty byl v zásadě stejný jako důkaz obdobné věty pro zobrazení. Dělali jsme tedy totéž dvakrát, což naznačuje, že to bylo zbytečné. My totiž víme, že zobrazení jsou speciálním případem relací. Pokud bychom tedy zařadili nejprve tuto kapitolu a kapitolu 2b až po ní, tak už by vzorec $(S \circ R)^{-1}$ pro zobrazení automaticky vyplynul z toho pro relace.

Dokonce je možné zajít ještě dál. Čtenář možná zná z lineární algebry, že i u matic platí podobný vztah mezi inverzní maticí a násobením matic (prohození pořadí). To naznačuje, že tento jev ve skutečnosti neplyne z nějakého speciálního talentu relací či zobrazení, ale bude za tím nějaký hlubší důvod, obecný princip. Právě hledání skutečných důvodů, tedy prvotních principů, je jedním ze zajímavých úkolů matematiky.

V tomto případě je ono prohazování pořadí konkrétní inkarnací obecného principu, který dokážeme jako Větu 8a.9. Z čistě logického pohledu je tedy to, co zde děláme, mnohdy zbytečné. Správný postup by byl nejprve udělat binární operace (kapitola 8) a tam dokázat vlastnosti pro velice obecnou skupinu objektů, pak udělat kapitoly o relacích, kde některé věty prostě vyplynou přímo z té obecné kapitoly o operacích, a protože jsou relace přeci jen trochu speciální, dokázaly by se i nějaké věci navíc. Nakonec bychom udělali kapitolu o zobrazeních a tam by spousta vět zase vyplynula z vět o relacích. Bylo by to celé mnohem kratší, důkazy by byly přehlednější a navíc z matematického pohledu by byla situace jasnější, protože bychom hned viděli, že za některé věci nevděčíme přímo zobrazením (či relacím), ale mnohem obecnějším principům, které relace a zobrazení shodou okolností také sdílejí.

Takto se píší pokročilé knihy pro matematiky, ale začátečník by se ztratil hned v té první obecné kapitole, protože by se ještě neuměl v matematickém jazyce orientovat a hlavně by netušil, co a proč se dělá. Proto jsme i zde věrní zásadám Komenského (od speciálního k obecnému, od jednoduchého ke složitému, od známého k neznámému) čtenáře k onomu abstraktnímu světu spíš pomalu přibližujeme. Vrcholem tohoto snažení bude již zmíněná kapitola o binárních operacích, kde se opravdu vyrádíme ve zcela umělých světech.

△

Lze zapřemýšlet, jaká pravidla platí, když se spolu potkají množinové operace a skládání. Zase tak často se to ale při práci s relacemi nevyskytuje, a když už to člověk potřebuje, tak si (pokud věci rozumí) většinou velice rychle rozmyslí, co platí a co ne. Proto zde nebudeme text přetěžovat tvrzeními a odkážeme čtenáře na cvičení 3a.12. Přemýšlením nad tím, která pravidla platí, si člověk dobrě procvičí mozek i kreslení obrázků a lépe pochopí relace a manipulace s nimi.

Docela zajímavé výsledky dostaneme, když zkusíme relaci skládat samu se sebou.

Definice.

Nechť R je relace na nějaké množině A . Pak definujeme její **mocninu** rekurzivně jako

$$(0) R^1 = R;$$

$$(1) R^{n+1} = R \circ R^n \text{ pro } n \in \mathbb{N}.$$

Let R be a relation on some set A . We define **powers** of R recursively by $R^1 = R$ and $R^{n+1} = R \circ R^n$ for $n \in \mathbb{N}$.

Díky asociativitě víme, že je úplně jedno, jak dáme ve výrazu $R \circ R \circ \dots \circ R$ závorky, takže je ani nebudeme psát.

Příklad 3a.n: Vraťme se k relaci na množině A všech lidí danou jako aRb jestliže je b rodič a . Co je R^2 ?

Aby $(a, c) \in R^2$, musí existovat člověk $b \in A$ takový, že aRb a bRc , tedy b je rodič a a c je rodič b . Vidíme, že R^2 je relace, která spojuje vnu(č)ky a s jejich dědy a babičkami c . Obdobně si rozmyslíme, že R^3 spojuje pravnu(č)ky s prarodiči atd.

△

Příklad 3a.o: Uvažujme množinu A všech autobusových zastávek v Opavě a relace R nám říká, kam se dostaneme přímým spojem. Přesně, aRb jestliže existuje linka, která staví nejprve v a a někdy poté v b . Pak nám R^2 říká, kam se dostaneme s přesně jedním přestupem, R^3 nám říká, kam se dostaneme s přesně dvěma přestupy, a tak dále. Když to dáme všechno dohromady, tak vidíme, že $\bigcup_{n=1}^{\infty} R^n$ je relace, která říká, odkud kam se nějakým způsobem dostaneme autobusem.

△

Naše pozorování o mocnině si teď zformuluujeme obecně.

Lemma 3a.5. (o cestách)

Nechť R je relace na nějaké množině A . Uvažujme prvky $a, b \in A$. Pak $(a, b) \in R^n$ právě tehdy, když existují prvky $c_1, c_2, \dots, c_{n+1} \in A$ takové, že $c_1 = a$, $c_{n+1} = b$ a $(c_i, c_{i+1}) \in R$ pro všechna $i = 1, \dots, n$.

Řetězci jako v Lemmatu budeme říkat „trasa délky n z a do b “, někdy jej budeme zapisovat $aRc_2Rc_3R\dots Rc_nRb$. Mimochodem, rádi bychom říkali cesta, ale problém je, že v teorii grafů je „cesta“ zavedený pojem, který nepovoluje opakování zastávek, ale zde potřebujeme svobodu volby. Raději jsme zvolili termín trasa, který v teorii grafů vůbec není.

Důkaz: Důkaz povedeme indukcí.

(0) Pro $n = 1$ to platí, protože zvolíme $c_1 = a$, $c_2 = b$ a je to.

(1) Předpokládejme, že tvrzení o tom, že R^n je dáno právě trasami o délce n , platí pro jisté $n \in \mathbb{N}$. Chceme ukázat, že platí i pro $n + 1$.

1) Vezměme nějaké $a, b \in A$. Jestliže $(a, b) \in R^{n+1}$, pak podle definice mocniny $(a, b) \in R \circ R^n$, což podle definice skládání znamená, že existuje $x \in A$ s vlastností $(a, x) \in R^n$ a xRb . Teď aplikujeme indukční předpoklad na dvojici $(a, x) \in R^n$ a dostaneme trasu délky n neboli prvky $\tilde{c}_1, \dots, \tilde{c}_{n+1} \in A$ takové, že $\tilde{c}_1 = a$, $\tilde{c}_{n+1} = x$ a $\tilde{c}_i R c_{i+1}$ pro všechna i . Definujeme prvky c_i takto: $c_i = \tilde{c}_i$ pro $i = 1, \dots, n + 1$ a $c_{n+2} = b$. Pak zjevně c_i splňují nároky na ně kladené, tedy tvoří trasu délky $n + 1$ z a do b . Takže prvky z R^{n+1} jsou opravdu dány trasami.

2) Předpokládejme naopak, že existují prvky $c_1, \dots, c_{n+2} \in A$ takové, že $c_1 = a$, $c_{n+2} = b$ a $c_i R c_{i+1}$ pro všechna $i = 1, \dots, n + 1$. Označme $x = c_{n+1}$. Pak řetězec c_1, \dots, c_{n+1} splňuje požadavky z tvrzení o trasách pro délku n z a do x , proto podle indukčního předpokladu $(a, x) \in R^n$. Víme také, že $(x, b) = (c_{n+1}, c_{n+2}) \in R$. Podle definice skládání tedy $(a, b) \in R \circ R^n = R^{n+1}$.

V částech 1) a 2) jsme dokázali, že existence trasy délky $n + 1$ z a do b je ekvivalentní tomu, že $(a, b) \in R^{n+1}$, čímž je indukční krok hotov.

□

Uděláme si teď formálně i to pozorování o relaci udávající všechna možná spojení. Je totiž důležitá a má své vlastní jméno.

Definice.

Nechť R je relace na nějaké množině A . Definujeme její **connectivity relation** jako $R^* = \bigcup_{n=1}^{\infty} R^n$.

Dostali jsme se k ní u autobusů, ale je evidentní, že to má aplikace i v mnoha jiných oborech, například jestliže relace R udává, které počítače jsou navzájem přímo spojeny, pak R^* je relace udávající, které počítače se dokážou spolu domluvit, pokud uvažujeme i spojení přes prostředníky.

Dobrá otázka samozřejmě je, jak toto dělat v praxi, kde jsou nekonečna trochu problém. Dá se ukázat, že pokud má množina A m prvků, pak se dá libovolná trasa (viz Lemma o cestách) zkrátit tak, aby nebyla delší než m .

To znamená, že dostaneme $R^* = \bigcup_{n=1}^m R^n$, což (pokud to děláme přes reprezentující matice a maticovou mocninu) vyžaduje hrůzných $2m^3(m - 1)$ operací. Existuje algoritmus, který to zvládne za $2m^3$ operací, což je pořád ještě docela dost, ale o řád lepší než přímý útok. Tím už se ale dostaváme nebezpečně blízko k algoritmizaci, tak to vysvětlování necháme odborníkům.

Co se od takové mocniny R^n dá čekat? V zásadě cokoliv. Všimněte si například, že nikde není zaručeno, že v $R \circ R$ zůstanou původní dvojice z R . Ze prvků $(a, b) \in R$ se do $R \circ R$ dostanou jen ty, u kterých lze trasu $a \mapsto b$ udělat i přes prostředníka $aRxRb$. Klidně se může stát, že začneme s relativně bohatou relací, ale po umocnění už zbyde velice málo.

Jednoduchý extrémní příklad: Uvažujme relaci R na množině $A = \{1, 2, 3\}$ danou jako $R = \{(1, 2), (3, 2)\}$. Pak nelze vytvořit žádný navazující řetězec $aRbRc$, takže $R^2 = \emptyset$.

Je ovšem také možný jev přesně opačný: Začneme s relací, pákrát umocníme a dostaneme maximální možnou relaci $A \times A$, kdy je každý s každým v relaci. Někdy se stane, že se výsledek s každou mocninou změní, jindy zase po určité mocnině již k ničemu novému nedojdeme a všechny další mocniny vypadají stejně. Jsou dokonce relace, které se umocňováním vůbec nemění, viz například Věta 3b.6.

Velikost mocniny tedy v jistém smyslu říká, jak dobře na sebe dvojice v R navazují. Někdy hodně napoví už prvním skládání.

Fakt 3a.6.

Nechť R je relace na nějaké množině A . Jestliže $R^2 \subseteq R$, pak $R^n \subseteq R$ pro všechna $n \in \mathbb{N}$.

Důkaz (poučný): Předpokládejme, že platí $R^2 \subseteq R$. Důkaz inkluze $\forall n \in \mathbb{N}: R^n \subseteq R$ povedeme indukcí.

(0) Pro $n = 1$ máme zkoumat inkluzi $R \subseteq R$, ta evidentně platí.

(1) Uvažujme libovolné $n \in \mathbb{N}$ a předpokládejme, že $R^n \subseteq R$. Chceme dokázat, že $R^{n+1} \subseteq R$.

Nechť $(a, b) \in R^{n+1}$. Protože $R^{n+1} = R \circ R^n$, musí existovat $x \in A$ takové, že $(a, x) \in R^n$ a $(x, b) \in R$. Podle indukčního předpokladu ovšem také $(a, x) \in R$, proto jak (a, x) , tak (x, b) leží v R . Proto $(a, b) \in R^2$, tedy podle předpokladu tvrzení také $(a, b) \in R$. □

Na závěr ještě jedna definice.

Definice.

Nechť R je relace na množině A , nechť $B \subseteq A$. Definujeme restrikci R na B jako relaci $R \cap (B \times B)$.

Je to velice jednoduché, z relace R prostě vyhodíme všechny dvojice, ve kterých se objevují prvky mimo B . Řečeno jinak, označíme-li tuto restrikci S , pak pro prvky $a, b \in B$ platí aSb přesně tehdy, platí-li pro ně aRb . Z toho důvodu se také nezavádí pro restrikci nějaké speciální značení, je to prostě pořád R , jen ji používáme pouze na některé prvky. Čtenář to zná již dálno, například máme relaci $x \leq y$ na \mathbb{R} , ale můžeme ji používat třeba jen pro celá čísla.

! 3a.7 Operace a reprezentace

Pro relaci R z konečné množiny A do konečné množiny B jsme zavedli její reprezentaci odpovídající 01-maticí M_R . Jak se operace, které jsme právě probrali, odrazí v řeči matic? Velice pěkně.

! Fakt 3a.8.

Uvažujme relaci R z množiny A do množiny B s reprezentující maticí M_R .

(i) Relace R^{-1} je reprezentovaná transponovanou maticí M_R^T .

(ii) Relace \bar{R} je reprezentovaná maticí danou $m_{\bar{R},ij} = 1 - m_{R,ij}$.

Zde je důležité si uvědomit, že matice M závisí silně na tom, jak si prvky množin A a B očíslovujeme. V tomto Faktu i následujících obdobných situacích tedy předpokládáme, že zachováváme jedno pevně zvolené číslování prvků A a B .

Důkaz (rutinní): (i): Předpokládejme, že $|A| = m$ a $|B| = n$, takže M_R je matice $m \times n$. Relace R^{-1} jde z B do A , proto bude mít její matice rozměr $n \times m$, což M_R^T má. Teď ukážeme shodnost prvků matic.

Protože matice M_R^T a $M_{R^{-1}}$ obsahují pouze jedničky a nuly, tak stačí dokázat, že se shodují jedničky, a nuly už budou automaticky také souhlasit. Takže:

$$m_{R^{-1},ij} = 1 \iff (b_i, a_j) \in R^{-1} \iff (a_j, b_i) \in R \iff m_{R,ji} = 1.$$

Důkaz (ii) je snadný a necháme jej jako cvičení 3a.11. □

Teď se podíváme na operace, které spojují dvě relace. K nim budeme potřebovat operace spojující dvě matice, již při základním pohledu je ale zřejmé, že nemá smysl zkoušet ty běžné. My teď totiž pracujeme jen s 01-maticemi, ale například sčítání je schopné vyrobit ve výsledné matici i jiná čísla než 0 a 1. Ke správnému nápadu nás navede, když si uvědomíme, že ty 0 a 1 vlastně reprezentují logické hodnoty (pravda, nepravda), konkrétně m_{ij} označuje pravdivostní hodnotu výroku „ $(a_i, b_j) \in R$ “. Má tedy smysl používat operace založené na operacích logických (Booleanovských).

Definice.

Nechť A, B, C jsou 01-matice typu $m \times n$.

Řekneme, že C je **join** matic A a B , značeno $C = A \vee B$, jestliže $c_{ij} = a_{ij} \vee b_{ij}$ pro všechna i, j .

Řekneme, že C je **meet** matic A a B , značeno $C = A \wedge B$, jestliže $c_{ij} = a_{ij} \wedge b_{ij}$ pro všechna i, j .

Není těžké ukázat, že tyto operace jsou komutativní a asociativní a spojuje je distributivní zákon:

- $A \vee B = B \vee A$, $A \wedge B = B \wedge A$;
- $(A \vee B) \vee C = A \vee (B \vee C)$, $(A \wedge B) \wedge C = A \wedge (B \wedge C)$;
- $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$, $A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$.

Databázové systémy (třeba i SAP) mívají tyto operace implementované.

Fakt 3a.9.

Nechť R_1, R_2 jsou relace z A do B s maticemi M_1, M_2 . Pak $M_1 \wedge M_2$ je matice relace $R_1 \cap R_2$ a $M_1 \vee M_2$ je matice relace $R_1 \cup R_2$.

Potřebujeme ještě jednu operaci.

Definice.

Nechť A je 01-matice typu $m \times k$ a B je 01-matice typu $k \times n$. Definujeme **Booleanovský součin (Boolean product)** těchto matic jako matici $C = A \odot B$ typu $m \times n$ danou

$$c_{ij} = (a_{i1} \wedge b_{1j}) \vee (a_{i2} \wedge b_{2j}) \vee \dots \vee (a_{ik} \wedge b_{kj}).$$

Všimněte si, že je to vlastně stejně jako vzorec pro běžné maticové násobení, jen se místo násobení dala konjunkce a místo sčítání disjunkce.

Věta 3a.10.

Nechť A, B, C jsou konečné množiny, R je relace z A do B s maticí M_R a S je relace z B do C s maticí M_S . Pak $M_R \odot M_S$ je matice relace $S \circ R$.

Důkaz (z povinnosti): Předpokládejme, že $A = \{a_1, \dots, a_m\}$, $B = \{b_1, \dots, b_k\}$, $C = \{c_1, \dots, c_n\}$. Pak M_R je matice $m \times k$ a M_S je matice $k \times n$, proto má smysl je násobit $M_R \odot M_S$. Dále máme $m_{R,il} = 1$ jestliže $a_i R b_l$ a $m_{S,lj} = 1$ jestliže $b_l S c_j$.

Ukážeme, že $M_{S \circ R} = M_R \odot M_S$. Protože jde o 01-matice, stačí ukázat, že obě matice mají 1 na stejných místech.

Vezměme prvek m_{ij} matice $M_R \odot M_S$. Tento prvek je 1 právě tehdy, jestliže je

$$(m_{R,i1} \wedge m_{S,1j}) \vee (m_{R,i2} \wedge m_{S,2j}) \vee \dots \vee (m_{R,ik} \wedge m_{S,kj}) = 1.$$

Jde o logickou disjunkci mnoha členů, ta je rovna jedné právě tehdy, když existuje $l \in \{1, \dots, k\}$ takové, že $m_{R,il} \wedge m_{S,lj} = 1$, tedy $\exists l \in \{1, \dots, k\}: (m_{R,il} = 1 \wedge m_{S,lj} = 1)$. To je podle definice těchto matic právě tehdy, pokud $\exists b_l \in B: (a_i R b_l \wedge b_l S c_j)$. Toto zase nastane právě tehdy, pokud $(a_i, c_j) \in S \circ R$, což je právě tehdy, pokud

$m_{S \circ R,ij} = 1$. Rovnost matic je dokázána. □

Booleanovský součin je tedy ta správná operace. Má všechny vlastnosti jako běžné maticové násobení:

- $(M \odot N) \odot P = M \odot (N \odot P)$;
- $M \odot (N \wedge P) = (M \odot N) \wedge (M \odot P)$ a $M \odot (N \vee P) = (M \odot N) \vee (M \odot P)$;
- není obecně komutativní.

Dokonce má i jednotkový prvek, jmenovitě standardní jednotkovou $n \times n$ matici $E_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$.

Platí tedy následující:

- Jestliže je A nějaká 01-matici typu $m \times n$, pak $A \odot E_n = E_m \odot A = A$.

Když máme asociativitu, obvykle toho využijeme k definování mocniny.

Definice.

Nechť A je čtvercová 01-matici $n \times n$.

Definujeme její Booleanovské mocniny rekurzí jako $A^{[0]} = E_n$ a $A^{[n+1]} = A \odot A^{[n]}$ pro $n \in \mathbb{N}_0$.

Z předchozí věty pak okamžitě máme následující tvrzení.

Důsledek 3a.11.

Nechť R je relace na množině A a M_R její matice. Pak pro $n \in \mathbb{N}$ platí $M_{R^n} = M_R^{[n]}$.

U relací na konečných množinách se tedy dá hodně věcí udělat pomocí maticových operací. Nejde ale o nástroj univerzální. Matice jsou jednak drahé na paměť (o tom už jsme mluvili), ale hlavně náročné na výpočetní výkon, zejména násobení je opravdu výpočetně drahá operace. Pokud je tedy relace řídká, vyplatí se hledat efektivní algoritmy pro práce s nimi či někdy dokonce zvážit zcela jiné metody uložení relace. Mimochodem i v mnoha dalších oborech se musí pracovat s velkými maticemi a lidé si musí najít svůj způsob, jak si ekonomicky poradit s těmi řídkými.

Cvičení

Cvičení 3a.1 (rutinní): Uvažujme relace na množině $A = \{0, 2, 4, 6\}$ definované pro $a, b \in A$ takto:

- (i) aR_0b jestliže $a < b$; (ii) aR_1b jestliže $a + 10 < 3b$; (iii) aR_2b jestliže $2a \leq b$.

Pro každou z nich napište danou relaci (tedy jako množinu s výpisem prvků), nakreslete její graf a napište její reprezentující matici. Pak pro každou z nich najděte inverzní relaci. Na závěr najděte $R_1 \cup R_2$, $R_1 \cap R_2$, $R_1 - R_2$ a $R_2 - R_1$.

Cvičení 3a.2 (rutinní, poučné): Uvažujme relaci R z množiny $A = \{1, 7, 8, 10, 20\}$ do množiny $B = \{a, d, s, t\}$ definovanou pro $a \in A$ a $b \in B$ takto: $(a, b) \in R$ jestliže se písmeno b objeví ve slovním vyjádření čísla a .

Napište danou relaci (tedy jako množinu s výpisem prvků) a nakreslete její graf, pak najděte její reprezentující matici a na závěr její inverzní relaci.

Cvičení 3a.3 (rutinní): Uvažujme množinu A studentů a množinu B učitelů určitého ústavu (vzdělávacího). Pro $a \in A$, $b \in B$ definujme relaci R předpisem aRb jestliže a měl učitele b na přednášku a relaci S předpisem aSb jestliže měl a učitele b na cvičení. Interpretujte relace $R \cup S$, $R \cap S$, $R - S$, \overline{R} a $\overline{R \cup S}$.

Cvičení 3a.4 (rutinní): Uvažujme množinu A studentů a množinu B právě nabízených předmětů. Pro $a \in A$, $b \in B$ definujeme relaci R_1 předpisem aR_1b jestliže si student a v tomto semestru zapsal předmět b a relaci R_2 předpisem aR_2b jestliže student a v tomto semestru dostal kredit za absolvování předmětu b . (Předpokládáme, že tuto relaci děláme na konci semestru, kdy již proběhly zkoušky.)

- a) Jak se otázka „platí $R_2 \subseteq R_1$ “ řekne českou větou?
b) Interpretujte relace $R_1 \cap R_2$, $R_1 \cup R_2$, $R_1 - R_2$, $R_2 - R_1$, R^{-1} .

Cvičení 3a.5 (poučné, zkouškové): Nechť R je relace z množiny A na množinu B . Dokažte, že $(R^{-1})^{-1} = R$.

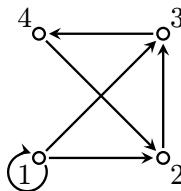
Cvičení 3a.6 (rutinní, zkouškové): Nechť R, S jsou relace z množiny A do množiny B . Dokažte, že pak

- (i) $(R \cup S)^{-1} = R^{-1} \cup S^{-1}$; (ii) $(R \cap S)^{-1} = R^{-1} \cap S^{-1}$; (iii) $(R - S)^{-1} = R^{-1} - S^{-1}$.

Cvičení 3a.7 (rutinní): Uvažujme dvě relace na množině $A = \{1, 2, 3, 4\}$, relaci $R = \{(1, 1), (1, 4), (2, 1), (3, 4)\}$ a relaci $S = \{(1, 4), (1, 3), (4, 3), (3, 2)\}$.

Najděte relaci $S \circ R$ a $R \circ S$.

Cvičení 3a.8 (rutinní):

Uvažujte relaci danou následujícím grafem: 

Cvičení 3a.9 (dobré): Nechť A je množina lidí. Definujme relaci R na A předpisem aRb jestliže je a rodič b a relaci S předpisem aRb jestliže je a je sourozencem b . Co je $S \circ R$ a $R \circ S$?

Poznámka: Definujeme „sourozence“ jako ty lidi, kteří mají společné oba rodiče. To mimo jiné znamená, že každý je sám sobě sourozencem.

Pokud bychom sourozence definovali jinak, pak by se změnil i výsledek těch skládání. Zkuste prozkoumat jiné zajímavé definice.

Cvičení 3a.10 (rutinní, poučné, *dobré): Uvažujme následující relace na \mathbb{R} : R_1 je relace $>$, R_2 je relace \geq , R_3 je relace $=$, R_4 je relace \neq , R_5 je relace $<$, R_6 je relace $|x| = |y|$.

Určete, čemu se rovnají složené relace

- | | | |
|-------------------------|------------------------|---------------------------|
| (i) $R_2 \circ R_1$; | (iv) $R_1 \circ R_4$; | (vii)* $R_4 \circ R_4$; |
| (ii) $R_1 \circ R_2$; | (v) $R_1 \circ R_5$; | (viii)* $R_1 \circ R_6$; |
| (iii) $R_1 \circ R_3$; | (vi) $R_2 \circ R_2$; | (ix)* $R_6 \circ R_1$. |

Cvičení 3a.11 (rutinní): Uvažujme relaci R na množině A s reprezentující maticí M_R . Dokažte, že relace \bar{R} je reprezentovaná maticí danou $m_{\bar{R},ij} = 1 - m_{R,ij}$.

Cvičení 3a.12 (poučné, zkouškové až drsné):

a) Nechť S je relace z množiny A do množiny B , nechť R_1, R_2 jsou relace z B do množiny C . Dokažte, že pak

- (i) $(R_1 \cup R_2) \circ S = (R_1 \circ S) \cup (R_2 \circ S)$;
- (ii) $(R_1 \cap R_2) \circ S \subseteq (R_1 \circ S) \cap (R_2 \circ S)$;
- (iii) $(R_1 \circ S) - (R_2 \circ S) \subseteq (R_1 - R_2) \circ S$.

b) Nechť R_1, R_2 jsou relace z množiny A do množiny B , nechť S je relace z B do množiny C . Dokažte, že pak

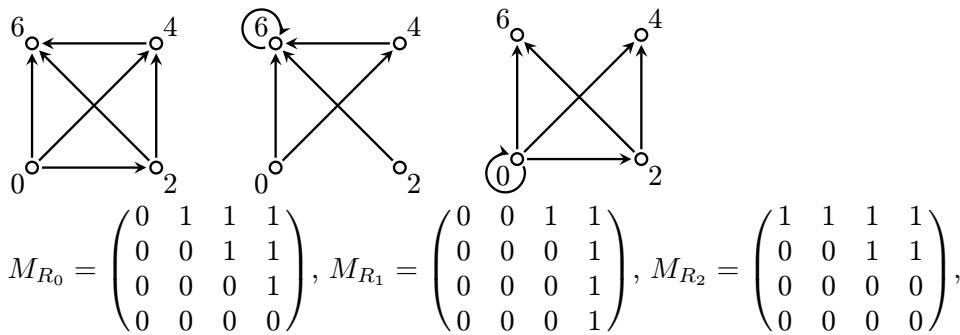
- (i) $S \circ (R_1 \cup R_2) = (S \circ R_1) \cup (S \circ R_2)$;
- (ii) $S \circ (R_1 \cap R_2) \subseteq (S \circ R_1) \cap (S \circ R_2)$;
- (iii) $(S \circ R_1) - (S \circ R_2) \subseteq S \circ (R_1 - R_2)$.

Řešení:

3a.1: $R_0 = \{(0, 2), (0, 4), (0, 6), (2, 4), (2, 6), (4, 6)\}$, $R_0^{-1} = \{(2, 0), (4, 0), (6, 0), (4, 2), (6, 2), (6, 4)\}$;

$R_1 = \{(0, 4), (0, 6), (2, 6), (4, 6), (6, 6)\}$, $R_1^{-1} = \{(4, 0), (6, 0), (6, 2), (6, 4), (6, 6)\}$;

$R_2 = \{(0, 0), (0, 2), (0, 4), (0, 6), (2, 4), (2, 6)\}$, $R_2^{-1} = \{(0, 0), (2, 0), (4, 0), (6, 0), (4, 2), (6, 2)\}$.



$$M_{R_0} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, M_{R_1} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, M_{R_2} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$R_1 \cup R_2 = \{(0, 0), (0, 2), (0, 4), (0, 6), (2, 4), (2, 6), (4, 6), (6, 6)\}$, $R_1 \cap R_2 = \{(0, 4), (0, 6), (2, 6)\}$,

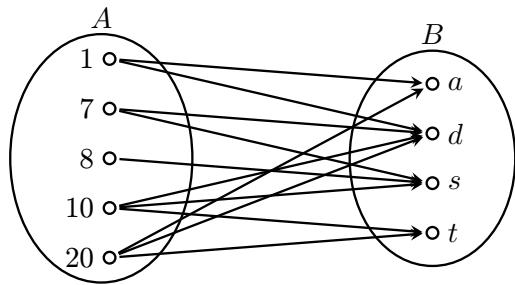
$R_1 - R_2 = \{(4, 6), (6, 6)\}$, $R_2 - R_1 = \{(0, 0), (0, 2), (2, 4)\}$.

3a.2:

$$R = \{(1, a), (1, d), (7, d), (7, s), (8, s), (10, d), (10, s), (10, t), (20, a), (20, d), (20, t)\}$$

$$M_R = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$$

$$R^{-1} = \{(a, 1), (a, 20), (d, 1), (d, 7), (d, 10), (d, 20), (s, 7), (s, 8), (s, 10), (t, 10), (t, 20)\}$$



3a.3: $R \cup S$: a potkal b coby vyučujícího, $R \cap S$: a měl b jako cvičícího i přednášejícího, $R - S$: a měl b jako přednášejícího, ale ne cvičícího, \bar{R} : a neměl b jako přednášejícího, $\bar{R} \cup \bar{S}$: a vůbec oficiálně neměl b , takže může předstírat, že ho nezná, a nemusí ho zdravit.

3a.4: (i): Dá se kredit získat jen za předměty, které si člověk zapsal?

3a.5: Nechť $a \in A, b \in B$. $(a, b) \in (R^{-1})^{-1} \iff (b, a) \in R^{-1} \iff (a, b) \in R$.

3a.6: (i): Nechť $a \in A, b \in B$. $(b, a) \in (R \cup S)^{-1} \iff (a, b) \in R \cup S \iff [(a, b) \in R \vee (a, b) \in S]$

$\iff [(b, a) \in R^{-1} \vee (b, a) \in S^{-1}] \iff (b, a) \in R^{-1} \cup S^{-1}$;

(ii): Nechť $a \in A, b \in B$. $(b, a) \in (R \cap S)^{-1} \iff (a, b) \in R \cap S \iff [(a, b) \in R \wedge (a, b) \in S]$

$\iff [(b, a) \in R^{-1} \wedge (b, a) \in S^{-1}] \iff (b, a) \in R^{-1} \cap S^{-1}$;

(iii): Nechť $a \in A, b \in B$. $(b, a) \in (R - S)^{-1} \iff (a, b) \in R - S \iff [(a, b) \in R \wedge (a, b) \notin S]$

$\iff [(b, a) \in R^{-1} \wedge (b, a) \notin S^{-1}] \iff (b, a) \in R^{-1} - S^{-1}$.

3a.7: Najdeme řetězce $1R1S3, 1R1S4, 1R4S3, 2R1S3, 2R1S4, 3R4S3$, proto

$$S \circ R = \{(1, 3), (1, 4), (2, 3), (2, 4), (3, 3)\};$$

najdeme řetězce $1S3R4, 3S2R1, 4S3R4$, proto $R \circ S = \{(1, 4), (3, 1), (4, 4)\}$.

3a.8: $R = \{(1, 1), (1, 2), (1, 3), (2, 3), (3, 4), (4, 2)\}; R^2 = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 4), (3, 2), (4, 3)\}$.

3a.9: $S \circ R$: a je rodič b (b nemusí mít sourozence, protože je sám sobě sourozencem), tedy $S \circ R = R$; $R \circ S$: a je strýc/teta nebo rodič b .

3a.10: (i): $(x, z) \in R_2 \circ R_1 \iff \exists y: xR_1y \wedge yR_2z \iff \exists y: x > y \wedge y \geq z \iff x > z$, tedy $R_2 \circ R_1 = R_1$; (ii):

R_1 ; (iii): R_1 ; (iv): $(x, z) \in R_1 \circ R_4 \iff \exists y: x \neq y \wedge y > z$, to lze pro libovolnou dvojici (x, y) , proto $R_1 \circ R_4 = \mathbb{R}^2$;

(v): \mathbb{R}^2 ; (vi): R_2 ; (vii): \mathbb{R}^2 ; (viii): $\{(x, z); |x| < z \vee -|x| < z\}$; (ix): $\{(x, z); x < |z|\}$.

3a.11: $m_{\bar{R},ij} = 1 \iff (a_i, a_j) \in \bar{R} \iff (a_i, a_j) \notin R \iff m_{R,ij} = 0$, obdobně $m_{\bar{R},ij} = 0 \iff m_{R,ij} = 1$.

3a.12: Nechť $a \in A, c \in C$.

a)(i): 1) $(R_1 \cup R_2) \circ S \subseteq (R_1 \circ S) \cup (R_2 \circ S)$: $(a, c) \in (R_1 \cup R_2) \circ S \implies \exists b \in B: [(a, b) \in S \wedge (b, c) \in R_1 \cup R_2]$

$\implies \exists b \in B: [(a, b) \in S \wedge ((b, c) \in R_1 \vee (b, c) \in R_2)]$

$\implies \exists b \in B: [(a, b) \in S \wedge (b, c) \in R_1] \vee \exists b \in B: [(a, b) \in S \wedge (b, c) \in R_2]]$

$\implies [(a, c) \in R_1 \circ S \vee (a, c) \in R_2 \circ S] \implies (a, c) \in (R_1 \circ S) \cup (R_2 \circ S)$;

2) $(R_1 \circ S) \cup (R_2 \circ S) \subseteq (R_1 \cup R_2) \circ S$: $(a, c) \in (R_1 \circ S) \cup (R_2 \circ S) \implies [(a, c) \in (R_1 \circ S) \vee (a, c) \in (R_2 \circ S)]$

$\implies [\exists b_1 \in B: ((a, b_1) \in S \wedge (b_1, c) \in R_1) \vee \exists b_2 \in B: ((a, b_2) \in S \wedge (b_2, c) \in R_2)]$

$\implies \exists b \in B: [(a, b) \in S \wedge (b, c) \in R_1 \cup R_2] \implies (a, c) \in (R_1 \cup R_2) \circ S$.

a)(ii): $(a, c) \in (R_1 \cap R_2) \circ S \implies \exists b \in B: [(a, b) \in S \wedge (b, c) \in R_1 \cap R_2]$

$\implies \exists b \in B: [((a, b) \in S \wedge (b, c) \in R_1) \wedge ((a, b) \in S \wedge (b, c) \in R_2)] \implies$

$(\exists b \in B: [(a, b) \in S \wedge (b, c) \in R_1] \wedge \exists b \in B: [(a, b) \in S \wedge (b, c) \in R_2]) \implies [(a, c) \in R_1 \circ S \wedge (a, c) \in R_2 \circ S]$

$\implies (a, c) \in (R_1 \circ S) \cap (R_2 \circ S)$;

Poznámka: Rovnost nastat nemusí, třeba pro $S = \{(a, b_1), (a, b_2)\}$, $R_1 = \{(b_1, c)\}$, $R_1 = \{(b_2, c)\}$, kde $b_1 \neq b_2$, dostáváme $(R_1 \cap R_2) \circ S = \emptyset \circ S = \emptyset$, ale $(R_1 \circ S) \cap (R_2 \circ S) = \{(a, c)\}$.

a)(iii): $(R_1 \circ S) - (R_2 \circ S) \subseteq (R_1 - R_2) \circ S$: $(a, c) \in (R_1 \circ S) - (R_2 \circ S)$

$\implies [(a, c) \in (R_1 \circ S) \wedge (a, c) \notin (R_2 \circ S)]$

$\implies [\exists b_1 \in B: ((a, b_1) \in S \wedge (b_1, c) \in R_1) \wedge \neg \exists b_2 \in B: ((a, b_2) \in S \wedge (b_2, c) \in R_2)]$

$\implies \exists b_1 \in B: [(a, b_1) \in S \wedge (b_1, c) \in R_1 \wedge (b_1, c) \notin R_2] \implies \exists b_1 \in B: [(a, b_1) \in S \wedge (b_1, c) \in R_1 - R_2]$

$\implies (a, c) \in (R_1 - R_2) \circ S$.

Poznámka: Rovnost nastat nemusí, třeba pro $S = \{(a, b_1), (a, b_2)\}$, $R_1 = \{(b_1, c)\}$, $R_1 = \{(b_2, c)\}$, kde $b_1 \neq b_2$, dostáváme $(R_1 - R_2) \circ S = \{(a, c)\}$, ale $(R_1 \circ S) \cap (R_2 \circ S) = \{(a, c)\} - \{(a, c)\} = \emptyset$.

b) Důkazy jsou obdobné.

3b. Základní vlastnosti binárních relací

Zde se omezíme pouze na relace na množině A . Používají se často a lidé si brzy všimli, že velice ocení, pokud se u zpracovávané relace mohou spolehnout na určité způsoby fungování, jinak řečeno, pokud dotyčná relace zachovává určitá pravidla. Dali jím jména a vznikly z toho populární vlastnosti relací. Je jich hodně, ale nejčastěji se zkoumají tyto čtyři:

!

Definice.

Nechť R je relace na množině A .

Řekneme, že R je **reflexivní**, jestliže pro všechna $a \in A$ platí aRa .

Řekneme, že R je **symetrická**, jestliže pro všechna $a, b \in A$ platí $(aRb \implies bRa)$.

Řekneme, že R je **antisymetrická**, jestliže pro všechna $a, b \in A$ platí $[(aRb \wedge bRa) \implies a = b]$.

Řekneme, že R je **tranzitivní**, jestliže pro všechna $a, b, c \in A$ platí $[(aRb \wedge bRc) \implies aRc]$.

V anglické verzi si procvičíme ten druhý zápis relací.

Consider a relation R on a set A .

It is called **reflexive** if $(a, a) \in R$ for all $a \in A$.

It is called **symmetric** if the following is true for all $a, b \in A$: $(a, b) \in R \implies (b, a) \in R$.

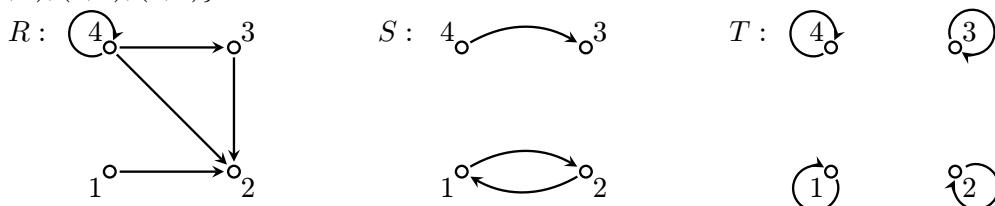
It is called **antisymmetric** if the following is true for all $a, b \in A$: $[(a, b) \in R \wedge (b, a) \in R] \implies a = b$.

It is called **transitive** if the following is true for all $a, b, c \in A$: $[(a, b) \in R \wedge (b, c) \in R] \implies (a, c) \in R$.

Poznamenejme, že některí čeští autoři používají namísto „antisymmetrie“ termínu „slabá antisymmetrie“, pro silnou viz část 3c.8 Další vlastnosti.

! Významu těchto definic nejlépe porozumíme, když se podíváme na několik relací na konečných množinách, které si znázorníme grafy. Uvažujme proto následující tři relace na množině $A = \{1, 2, 3, 4\}$:

- $R = \{(1, 2), (3, 2), (4, 2), (4, 3), (4, 4)\}$;
- $S = \{(1, 2), (2, 1), (4, 3)\}$;
- $T = \{(1, 1), (2, 2), (3, 3), (4, 4)\}$.



Reflexivita je jasná, aby byla relace reflexivní, musí být každý prvek množiny A v relaci sám se sebou (reflexe = odraz třeba v zrcadle), což snadno ověříme ve výpisu dvojic výše, vyhovuje jen relace T . V grafu to znamená, že pro reflexivitu vyžadujeme smyčky u všech bodů, což opravdu splňuje právě T .

Symetrie znamená podívat se na všechny možné dvojice prvků z A a pro každou z nich ověřit platnost jisté vlastnosti, která má formu implikace. To znamená, že pokud není splněn její předpoklad, tedy pokud prvky a, b nejsou v spolu relaci, pak se implikace již nemůže pokazit a tudíž takovéto případy nerovnou s platností symetrie, nemusíme je zkoumat. Kritické je, co se stane s dvojicemi, když $(a, b) \in R$. U těchto dvojic pak potřebujeme, aby v R byla i dvojice v opačném pořadí.

U relace R se tedy budeme ptát třeba na dvojici $a = 1$ a $b = 2$, která splňuje předpoklad $1R2$, ale nesplňuje závěr $2R1$. Implikace proto pro tuto dvojici neplatí. Na symetrii ale potřebujeme platnost pro všechny dvojice (je tam použit obecný kvantifikátor), takže se tato konkrétní dvojice stává protipříkladem a relace R není symetrická.

Pohrajme si trochu s relací S . Dvojice $a = 1, b = 2$ implikaci $1S2 \implies 2S1$ splňuje, stejně jako dvojice $a = 2, b = 1$ splňuje $2S1 \implies 1S2$. Rovněž s dvojicí $a = 1, b = 4$ nejsou potíže, ta nesplňuje předpoklad implikace (neplatí $1S4$) a tudíž to symetrii nemůže ohrozit. Čtenář ale asi od začátku vidí, že při probírání dvojic dříve či později dorazí k dvojici $a = 4, b = 3$, pro kterou příslušná implikace neplatí, a proto ani S není symetrická.

Měli bychom mimochodem zkoušet i dvojice jako $a = 1, a = 1$ (v definici symetrie se neříká nic o tom, že by a, b měly být různé), ale to se obvykle nedělá, protože je to jasné. Pokud aSa neplatí, pak to symetrii neohrozí, a pokud náhodou aSa , pak určitě i aSa (tedy jsme je prohodili, smyčky jsou už z podstaty obousměrné). Smyčky tedy symetrii neovlivní, můžeme je při zkoumání symetrie ignorovat.

Tedě už bychom měli symetrii rozumět, vyžaduje následující: Pokud je v grafu někde šipka, pak tam musí být i šipka v opačném směru. Pokud někde šipka není, tak to symetrii nezajímá.

Proto by také mělo být jasné, že relace T je symetrická, smyčky symetrii nerovnou s jiné dvojice, které bychom měli kontrolovat, tam nejsou. Tento způsob symetrie není zrovna typický, pro hezčí graf symetrické relace se čtenář

může podívat na cvičení 3b.1 (ii). Všimněte si, že T by zůstala symetrickou, i kdybychom umazali i ty smyčky, čímž by vznikla relace prázdná, bez dvojic. Symetrii nezajímá, kolik spojnic se v relaci vyskytuje, podstatné je jen zdvojování těch, které (pokud vůbec) tam jsou.

Antisimetrie má také formu implikace, takže už víme, že nás budou zajímat jen takové dvojice a, b , které splňují aRb a bRa . V relaci R je jediná taková dvojice, jmenovitě $a = 4$ a $b = 4$. Pro ni pak opravdu platí $a = b$, relace R je proto antisymetrická. Vidíme první věc, antisimetrii smyčky nevadí a nevadí jí také jednoduché ani žádné spojnice mezi body.

U relace S je jedna dvojice splňující předpoklad, $a = 1$ a $b = 2$. Ta splňuje aSb a bSa , ale nesplňuje závěr $a = b$, relace S proto není antisymetrická. Opět by už mělo být jasné, oč v této vlastnosti jde: Antisimetrie vylučuje možnost dvojitých šipek mezi různými prvky (smyčky jí nevadí). Z toho hned vyplývá, že relace T je antisymetrická. Tam kromě smyček žádné dvojice splňující aTb a bTa nejsou, tudíž se implikace z definice nemůže pokazit. Opět je to spíše netypický případ, R je obvyklejší antisymetrická relace.

Poznamenejme, že intuitivně antisimetrie zní jako popření symetrie, takže se nabízí také definice typu „pokud je šipka aRb , pak nesmí být šipka bRa “. Není to zas tak špatný nápad, viz asymetrie v části 3c.8, ale taková definice by vyloučila smyčky, což se příliš nehodí, proto se používá definice, kterou jsme tu uvedli a která smyčky připouští. Možná není napsaná nejasnější a čtenář by dal přednost ekvivalentnímu vyjádření „pokud $a \neq b$ a aRb , pak nesmí platit bRa “, ale to není tak vhodné z hlediska výrokové logiky a (kupodivu) to bývá i méně praktické, proto se to nepoužívá, stejně jako se nepoužívá obměna naší definice: Jestliže $a \neq b$, pak alespoň jedna z dvojic $(a, b), (b, a)$ není v R .

Z předchozího odstavce vyplývá, že jsme vědomě zvolili pro antisimetrii jinou definici než opak symetrie. Nejde tedy o opačné vlastnosti, což vidíme i výše, T je zároveň symetrická i antisymetrická, naopak S není ani symetrická, ani antisymetrická. Posléze uvidíme, že aby byla relace zároveň symetrická i antisymetrická, tak už to musí být v zásadě T , takže jde o výjimku. Naopak porušení obou vlastností najednou je velice snadné, stačí zařadit jednu šipku dvojitou (obousměrnou) a jednu jednoduchou, čímž se pokazí antisimetrie i symetrie.

Tranzitivita také znamená, že ověřujeme platnost implikace, tedy zajímají nás jen situace, kdy nám tři body dávají navazující šipky aRb a bRc . Pak potřebujeme, aby existovala i zkratka aRc . Jak je na tom relace R ? To už dá (hlavně u větších množin) více práce, zde asi čtenář brzo odhalí navazující trojici $4R3$ a $3R2$. Prvky $a = 4, b = 3, c = 2$ tedy splňují předpoklad, je nutné ověřit, zda platí i závěr $4R2$, a on platí. Otázka na tělo: Je tam ještě jiná podobná trojice?

Ano, trojice $a = b = 4, c = 3$, protože v definici tranzitivity se možnost shodných bodů nevylučovala. Měli bychom tedy také ověřit platnost implikace pro případ $a = b = c = 4$. Je nicméně jasné, že u dvojic tohoto typu je podmínka tranzitivity automaticky splněna, takže dvojice zahrnující smyčky nemohou tranzitivitu pokazit. Závěr: R je tranzitivní.

Je tranzitivní S ? Na první pohled by si mnohý čtenář mohl myslet, že je tranzitivní automaticky, protože tam nevidí dvojkroky, tudíž není nutné hledat zkratky. Jenže dvojkrok tam jeden je, jmenovitě $a = 1, b = 2, c = 1$. Opravdu $1S2$ a $2S1$, tudíž tranzitivita vyžaduje také existenci jednokroku $1S1$, ale ten tam nemáme, podobně chybí $2S2$. Relace S proto není tranzitivní.

Smysl tranzitivity je tedy v tom, že kdykoliv se v grafu vyskytuje navazující dvoukrok, pak tam musí existovat také zkratka jedním krokem, přičemž musíme být opatrní, abychom něco nevynechali. Ani tranzitivita neřeší, co se děje, pokud dvoukroky nejsou, takže už je jasné, že relace T bude tranzitivní.

Teď už bychom měli intuitivně rozumět základním vlastnostem relací a vidět je v grafech na první pohled, s výjimkou tranzitivity, která se u houštinovitějších grafů dělá pohledem nesnadno. Právě u ní se ale ještě chvíli zdržíme. Pokud pravidlo o zkracování dvoukroků aplikujeme opakováně na delší trasy, dostaneme následující: Jakomile se někam dostaneme vícero kroků, tak musí existovat i přímá cesta jednou šipkou. Potvrďme si oficiálně, že tato na pohled silnější podmínka je ekvivalentní tranzitivitě.

Fakt 3b.1.

Nechť R je relace na množině A . Je tranzitivní právě tehdy, když pro libovolné $n \in \mathbb{N}$, $n \geq 2$ a prvky $a_1, \dots, a_n \in A$ takové, že a_iRa_{i+1} pro všechna $i = 1, 2, \dots, n-1$, platí také a_1Ra_n .

Důkaz (rutinní): 1) Předpokládejme, že daná podmínka platí. Pak ji lze použít pro $n = 2$, což říká, že kdykoliv $a_1Ra_2Ra_3$, pak a_1Ra_3 , což je přesně definice tranzitivity.

2) Předpokládejme, že R je tranzitivní. Platnost podmínky dokážeme indukcí na n .

(0) Pro $n = 2$ podmínka platí, je to definice tranzitivity.

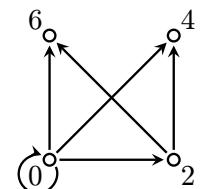
(1) Mějme $n \in \mathbb{N}$ splňující $n \geq 2$. Předpokládejme, že když libovolné prvky $a_1, \dots, a_n \in A$ splňují a_iRa_{i+1} pro všechna $i = 1, 2, \dots, n-1$, pak a_1Ra_n . Platí to i pro $n+1$ prvků?

Vezměme prvky $a_1, \dots, a_n, a_{n+1} \in A$ takové že $a_i Ra_{i+1}$ pro všechna $i = 1, 2, \dots, n$. Aplikací indukčního předpokladu na prvních n zjistíme, že $a_1 Ra_n$, spolu s $a_n Ra_{n+1}$ a tranzitivitou dostaneme $a_1 Ra_{n+1}$ a důkaz je hotov. \square

! Klasickou úlohou v oboru relací je takzvané **vyšetřování vlastností**, což znamená, že rozhodneme, která ze základních čtyř vlastností pro danou relaci platí a která ne, a své odpovědi také dokážeme. Budeme se přitom muset naučit jiné triky než v předchozím rozboru, protože nám málokterá relace přijde coby graf na malé množině. Typické relace jsou na velkých množinách, často nekonečných, a jsou dány nějakou logickou podmínkou. Postup vhodný pro tyto situace si ukážeme v následujícím příkladě.

! **Příklad 3b.a:** Připomeňme příklad ze cvičení 3a.1: Uvažujme relaci na množině $A = \{0, 2, 4, 6\}$ definovanou pro $a, b \in A$ takto: aRb jestliže $2a \leq b$. Cvičení nám dalo, že $R = \{(0, 0), (0, 2), (0, 4), (0, 6), (2, 4), (2, 6)\}$, a dostali jsme i obrázek.

Nejprve jako opakování vyhodnotíme vlastnosti z grafu. Protože existují prvky bez smyček, R není reflexivní. Existují dva prvky, které jsou spojeny, ale jen jedním směrem, R proto není symetrická. S výjimkou smyčky u 0 se nikdy nestane, že by šly šipky oběma směry, proto je R antisymetrická (antisimetrii smyčky nevadí). Na tranzitivitu musíme zkousit procházení grafem dvojkroky a hlídat, jestli vždy existuje zkratka. To je pravda, relace je tedy tranzitivní.



Ted' se podíváme, jak bychom ke stejným závěrům došli bez znalosti obrázku. Úplně stejným předpisem $2a \leq b$ bychom totiž mohli definovat relaci na množině se 100 prvky (pak je procházení grafu kvůli tranzitivitě práce na dlouhé zimní večery) nebo dokonce na \mathbb{N} či \mathbb{R} nebo jiné nekonečné množině, pak už graf ani nemáme. Zkusíme tedy vlastnosti vyšetřit znova, tentokráté čistě z definice pomocí matematiky a logiky.

Reflexivita: Platí pro každé $a \in A$, že aRa ? To by znamenalo $2a \leq a$. Víme, že takováto nerovnost je pravdivá jen pro reálná čísla $a \leq 0$, takže v naší množině A jsou určitě čísla, pro která to neplatí. Třeba neplatí $2 \cdot 6 \leq 6$, proto neplatí $6R6$. Tímto protipříkladem jsme ukázali, že relace R není reflexivní.

Symetrie: Ptáme se, zda pro všechna $a, b \in A$ platí $(aRb \implies bRa)$. Přeložíme to: Platí pro všechna $a, b \in A$ implikace $2a \leq b \implies 2b \leq a$? Intuice se na to dívá podezřívavě, první nerovnost totiž nutí a být malé ve srovnání s b a pak to chceme naopak. Zkusíme na tomto pocitu založit protipříklad. Třeba $2 \cdot 3 \leq 6$ platí, tedy $3R6$, ale $2 \cdot 6 \leq 3$ neplatí, proto nemáme $6R3$ a tudíž tato relace není symetrická.

Pokud odpověď neuhádneme intuicí, pak bývá dobrý nápad zkousit symetrii dokázat a pokud se někde zadrhneme, většinou to naznačí, kde hledat protipříklad. V tomto případě chceme z nerovnice $2a \leq b$ odvodit $2b \leq a$. Zkusíme to:

$$2a \leq b \implies 4a \leq 2b \implies 2b \geq 4a.$$

Předpoklad nám dal $2b \geq 4a$, ale my potřebujeme $2b \leq a$. Tyto nerovnosti jdou opačnými směry, takže je vysoko nepravděpodobné, že by se povedl přechod $2b \geq 4a \implies 2b \leq a$. To je silná indikace, že symetrie neplatí.

Antisymetrie: Ptáme se, zda pro všechna $a, b \in A$ platí, že když aRb a bRa , pak $a = b$. Zase si doplníme z definice, co ty relace znamenají: Platí, že pro libovolná čísla $a, b \in A$ z předpokladů $2a \leq b$ a $2b \leq a$ již odvodíme, že $a = b$? Pokud obě nerovnosti platí, tak můžeme eliminovat jednu proměnnou, třeba dosadit b z první nerovnosti do druhé, dostaneme $4a \leq a$. To platí jen pro čísla $a \leq 0$, což vzhledem k našemu A znamená $a = 0$. Symetricky dostaneme i $4b \leq b$, takže máme $a = 0$ a $b = 0$, tedy opravdu $a = b$.

Shrnuto: Ukázali jsme algebraicky, že pro libovolná $a, b \in A$ z nerovností $2a \leq b$ a $2b \leq a$ vyplývá $a = b$, čímž je antisymetrie R dokázána.

Alternativní důkaz: Všechny prvky z A splňují $a \leq 2a$. Podmínka $2a \leq b$ tedy znamená i $a \leq b$, obdobně $2b \leq a$ znamená $b \leq a$. Z obou podmínek tedy máme $a \leq b$ a $b \leq a$, proto $a = b$.

Tranzitivita: Ptáme se, zda pro všechny $a, b, c \in A$ platí, že jestliže aRb a bRc , pak také aRc . Opět přepíšeme podle definice R : Ptáme se, zda pro $a, b, c \in A$ platí, že když $2a \leq b$ a $2b \leq c$, pak také $2a \leq c$. Zvolíme podobný postup, zkousíme algebraicky eliminovat z předpokládaných nerovnic b tak, že jej z první dosadíme do druhé. Dostaneme $4a \leq c$. Protože ale pro prvky z A platí $2a \leq 4a$, tak máme nerovnice $2a \leq 4a \leq c$, tedy $2a \leq c$, přesně jak jsme potřebovali. Tranzitivita dokázána.

\triangle

S 3b.2 Jak zkoumat vlastnosti

Při vyšetřování konkrétních relací daných vzorcem (což byl příklad předchozí i většina dalších) doporučujeme následovat postup předvedený výše: U každé vlastnosti si nejprve přepsat její obecnou definici do aktuálního znění, tedy místo obecných výrazů typu aRb psát konkrétní podmínky z definice R . Když se pak člověk na takový konkrétní přepis podívá, většinou je hned jasné, zda je schopen dokázat jeho pravdivost či naopak ukázat protipříkladem, že někdy selhává. U vlastností, které platí, je dobré na konci úvah zkontovalovat, že opravdu vznikl

správný důkaz, občas stojí za to jej znova napsat načisto a pořádně. Správný důkaz musí začínat specifikací, že něco dokazujeme pro *libovolnou* volbu prvků (jednoho pro reflexivitu, dvou pro (anti)symetrii a tří pro tranzitivitu), čímž vyhovíme obecnému kvantifikátoru. Argument samotný pak musí probíhat ve správném směru, tedy začít předpokladem a skončit závěrem. Zkušený relační vyšetřovač dokáže u snadnějších relací dopředu odhadnout, jak se důkaz bude vyvíjet, a rovnou jej píše ve finálním tvaru.

Hned na začátku je ale často velice užitečné udělat ještě něco jiného. Než se člověk pustí do vlastností, měl by si udělat jasno v tom, jak relace vlastně funguje. Jinými slovy, je třeba se podívat, s jakými objekty relace pracuje, a pak si zkoušet vymyslet několik dvojic těchto objektů, které v relaci jsou, a naopak několik dvojic, které v relaci nejsou. Možná to zní triviálně, ale zkušenosť ukazuje, že se to vyplatí, zejména u relací, které pracují s trochu komplikovanějšími objekty. My si teď testování relací vyzkoušíme. Začneme relacemi snažšími (na číslech), ale až čtenář začne mít pocit, že je mu to jasné, tak by se měl podívat i na relace zajímavější, viz cvičení 3b.7.

! Příklad 3b.b:

Vyšetříme základní čtyři vlastnosti pro relaci $R = \{(a, b) \in \mathbb{Z}^2; a \leq b\}$.

Jinými slovy, zkoumáme běžnou nerovnost pro celá čísla. Zkusíme si zápis relace pomocí dvojic, ať si jej protrénujeme.

Reflexivita: Vezměme libovolné $a \in \mathbb{Z}$. Platí $(a, a) \in R$? Podle definice R to znamená $a \leq a$, což je splněno. Takže R je reflexivní. Do „oficiálního“ důkazu bychom samozřejmě ty úvahy nepsali, vypadalo by třeba takto:

Nechť $a \in \mathbb{Z}$ je libovolné. Pak $a \leq a$, proto $(a, a) \in R$.

Symetrie: Nechť $a, b \in \mathbb{Z}$ jsou takové, že $(a, b) \in R$, platí pak $(b, a) \in R$? Podle definice R tedy chceme, aby z $a \leq b$ vyplývalo $b \leq a$, ale toto nevypadá nadějně, máme podezření, že by to nemuselo platit. Zkusíme tedy dokázat, že R není symetrická, pomocí protipříkladu. Určitě $(13, 23) \in R$, neboť $13 \leq 23$, ale neplatí $23 \leq 13$ a tudíž $(23, 13) \notin R$.

Antisimetrie: Nechť $a, b \in \mathbb{Z}$ splňují $(a, b) \in R$ a $(b, a) \in R$. Platí pak $a = b$? Podle definice R předpoklad znamená, že $a \leq b$ a $b \leq a$, z čehož už hned plyne, že opravdu $a = b$. Antisimetrie dokázána.

Tranzitivita: Vezměme libovolné $a, b, c \in \mathbb{Z}$. Chceme ukázat, že jestliže $(a, b) \in R$ a $(b, c) \in R$, pak $(a, c) \in R$. Nechť tedy $(a, b) \in R$ a $(b, c) \in R$. Podle definice R nás předpoklad znamená, že $a \leq b$ a $b \leq c$, z čehož hned dostaneme $a \leq c$ a tedy $(a, c) \in R$. Tranzitivita dokázána.

△

Do důkazů jsme v příkladě vkládali naše úvahy, aby čtenář viděl, jak k problému přistupujeme. Do finální verze bychom je samozřejmě nepsali. Protože i u dalších relací přemýslíme stejně, budeme už důkazy psát stručněji, mimo jiné budeme upřednostňovat zápis aRb .

S 3b.3 Poznámka: Čtenář si jistě všiml, že jsme v důkazech určité části podtrhávali. Když si vytáhne například z důkazu, že je R antisymetrická, jen ty podtržené části, dostane přímo definici antisimetrie. Tak se u přímého důkazu (který obvykle u vlastností používáme) pozná, že je správně strukturálně postaven, u ostatních vlastností to platí samozřejmě také. Čtenář si to může hlídat i v následujících důkazech.

Dodržování správné struktury je velice užitečné v situaci, kdy se čtenář snaží sám nějaký důkaz vlastnosti napsat. Například každý (přímý) důkaz tranzitivity by měl vypadat nějak takto: Nechť $a, b, c \in A$ jsou libovolné, předpokládejme, že aRb a aRc . Pak ... , a proto aRc . Pokud se důkaz tváří, že přímo tranzitivitu dokazuje, ale tuto strukturu nemá, pak je nejspíš chybně. Pokud ji má, tak také ještě nemusí být správně, chyba může být někde v těch spojovacích textech (argumentech), ale alespoň má šanci. Je dobré si vždy před vymýšlením důkazu rozmyslet, jakou by měl mít strukturu, protože to často napoví, jak jej udělat. Rozhodně pak doporučujeme u již hotového důkazu platnosti nějaké vlastnosti zkoušit podtrháním částí získat její definici.

△

! Příklad 3b.c:

Vyšetříme základní čtyři vlastnosti pro relaci $R = \{(a, b) \in \mathbb{Z}^2; a < b\}$.

Zkoumáme tedy běžnou ostrou nerovnost, tentokráté zvolíme zápis aRb .

R: Nechť $a \in \mathbb{Z}$. Pak neplatí $a < a$, proto neplatí aRa . Tedy R není reflexivní.

Pro úplnost protipříklad: neplatí $3 < 3$, proto neplatí $3R3$.

S: Nechť $a, b \in \mathbb{Z}$ splňují aRb . Pak $a < b$, tedy neplatí $b < a$ a tedy neplatí bRa . Proto R není symetrická.

Pro úplnost protipříklad: $1R3$ neboť $1 < 3$, ale neplatí $3R1$.

A: Nechť $a, b \in \mathbb{Z}$ splňují aRb a bRa . Pak $a < b$ a $b < a$, což ale nemůže nastat nikdy. Předpoklad zkoumané implikace $(aRb \wedge bRa) \implies a = b$ tedy není nikdy splněn, proto celá implikace vždy automaticky platí. Tato relace je antisymetrická.

T: Nechť $a, b, c \in \mathbb{Z}$ splňují aRb a bRc . Pak $a < b$ a $b < c$, proto $a < c$ a tedy aRc . Tato relace je tranzitivní.

△

! Příklad 3b.d: Vyšetříme základní čtyři vlastnosti pro relaci $R = \{(a, b) \in \mathbb{R}^2; a = b\}$.

Takže zkoumáme relaci rovnosti. Zde tyto známé relace definujme formálně korektně, aby bylo lépe vidět myšlenky v důkazech, ale běžně bychom prostě řekli „uvažujme relaci $=$ na \mathbb{R} “.

R: Nechť $a \in \mathbb{R}$. Pak platí $a = a$, proto aRa . Tedy R je reflexivní.

S: Nechť $a, b \in \mathbb{R}$ splňují aRb . Pak $a = b$, tedy také $b = a$ neboli bRa . Proto je R symetrická.

A: Nechť $a, b \in \mathbb{R}$ splňují aRb a bRa . Pak $a = b$ a $b = a$, prostě $a = b$. Tato relace je antisymetrická.

T: Nechť $a, b, c \in \mathbb{R}$ splňují aRb a bRc . Pak $a = b$ a $b = c$, proto $a = c$ a tedy aRc . Tato relace je tranzitivní.

△

Poznámka: V příkladě 3b.d jsme měli relaci rovnosti mezi celými čísly a ukázali jsme její vlastnosti. Není těžké si rozmyslet (je to vlastně stejné), že relace rovnosti je vždy reflexivní, symetrická, antisymetrická a tranzitivní, ať už porovnávám jakékoli objekty, třeba rovnost množin, rovnost lidí, rovnost lineárních prostorů, cokoliv chcete. Je to v zásadě jediný typ relace, který je zároveň symetrický i antisymetrický, viz cvičení 3b.11.

△

! Příklad 3b.e: Vyšetříme základní čtyři vlastnosti pro relaci $R = \{(a, b) \in \mathbb{Z}^2; a + b \leq 13\}$.

R: Nechť $a \in \mathbb{Z}$. Pak $a + a \leq 13$ někdy platí, ale někdy ne, například $a = 7$ je protipříklad proti reflexivitě R .

S: Nechť $a, b \in \mathbb{Z}$ splňují aRb . Pak $a + b \leq 13$, tedy také $b + a \leq 13$ a proto bRa . Tato relace je symetrická.

A: Nechť $a, b \in \mathbb{Z}$ splňují aRb a bRa . Pak $a + b \leq 13$ a $b + a \leq 13$, toho $a = b$ nedostaneme. Například dvojice $a = 3, b = 7$ splňuje $3R7$ a $7R3$, ale nesplňuje $3 = 7$, je proto protipříkladem proti antisymetrii R .

T: Nechť $a, b, c \in \mathbb{Z}$ splňují aRb a bRc . Pak $a + b \leq 13$ a $b + c \leq 13$. Potřebujeme ukázat, že $a + c \leq 13$, abychom tím dostali aRc . To ale není nijak jasné, chybí nám vhodné nástroje (rádi bychom z oněch dvou předpokladů vyloučili b , aby zůstal jen vztah s a, c , ale u nerovnic například nefunguje rozumně eliminace). Zkusíme se proto zamyslet. Potřebujeme $a + c \leq 13$, tedy čísla by neměla být moc velká. Podmínka $a + b \leq 13$ může znamenat, že jedno číslo je dost velké a druhé malé, nemáme zde žádnou kontrolu, takže zkusíme vymyslet protipříklad: $11 + 2 \leq 13$, tedy $11R2$, také $2 + 7 \leq 13$, tedy $2R7$, ale neplatí $11 + 7 \leq 13$, proto neplatí $11R7$. Takže z řetízku $11R2R7$ nelze udělat $11R7$ a relace proto není tranzitivní.

Tady by byl možná čitelnější zápis přes dvojice, máme $(11, 2) \in R$ a $(2, 7) \in R$, ale neplatí $(11, 7) \in R$.

△

! Příklad 3b.f: Uvažujme nějaký soubor množin \mathcal{M} a na něm relaci býti podmnožinou. Podle Faktu 2a.1 (i) je reflexivní, podle Faktu 2a.2 je tranzitivní. Fakt 2a.3 ukazuje antisymetrii. Zbývá vyšetřit symetrii. Platí vždy, že $A \subseteq B$ implikuje $B \subseteq A$? Obecně ne, třeba $\{13\} \subseteq \{13, 23\}$, ale neplatí $\{13, 23\} \subseteq \{13\}$, relace \subseteq tedy obecně není symetrická.

Mohla by být symetrická pro nějaké speciální \mathcal{M} ? K vytvoření protipříkladu nám stačí mít dvě množiny tak, aby $A \subseteq B$ a zároveň $A \neq B$. Pokud soubor množin \mathcal{M} takovéto množiny neobsahuje, pak by relace \subseteq byla symetrická. Takovýchto situací existuje mnoho. Nejjednodušší je vzít jako \mathcal{M} soubor jen s jednou množinou. Například $\mathcal{M} = \{\{13\}\}$ obsahuje jedinou množinu, $\{13\}$. Relace \subseteq pak splňuje všechny čtyři vlastnosti.

Aby to nebyla taková nuda, i na souboru $\mathcal{M} = \{\{13\}, \{23\}\}$ je inkluze symetrickou relací, protože v tomto souboru nenajdeme dvě různé množiny, pro které by byl splněn předpoklad $A \subseteq B$. Rozmyslete si, že relace \subseteq je symetrická i na nekonečném souboru množin

$$\mathcal{M} = \{\{n, n+1\}; n \in \mathbb{N}\} = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \dots\}.$$

Nejtypičtější případ \mathcal{M} ale je, když máme nějaké neprázdné universum U a pracujeme se všemi jeho podmnožinami, tedy $\mathcal{M} = P(U)$. Pak relace \subseteq určitě není symetrická.

△

! Příklad 3b.g: Připomeňme porovnávání množin podle mohutnosti, budeme porovnávat množiny z nějakého souboru \mathcal{M} .

1) Relace $|A| \leq |B|$ je podle Faktu 2c.2 (i) reflexivní, podle Faktu 2c.3 (i) také tranzitivní. Není ale obecně symetrická, například $|\{13\}| \leq |\{13, 23\}|$, ale neplatí $|\{13, 23\}| \leq |\{13\}|$. Z toho je vidět, že zkoumaná relace je symetrická pouze v případě, že všechny množiny z \mathcal{M} mají stejnou mohutnost. V typickém případě $\mathcal{M} = P(U)$ tedy symetrii nemáme.

Není obecně ani antisymetrická, neboť $|\{13, 23\}| \leq |\{13, 14\}|$ a $|\{13, 14\}| \leq |\{13, 23\}|$, ale neplatí $\{13, 14\} = \{13, 23\}$. Teď vidíme, že aby byla zkoumaná relace antisymetrická, musela by \mathcal{M} mít pro každou mohutnost jen jednu množinu. To neplatí v tradičním případě $\mathcal{M} = P(U)$, proto zde tato relace není antisymetrická.

Pro dostatečně bohatý soubor množin, například $\mathcal{M} = P(U)$, tedy tato relace není ani symetrická, ani antisymetrická.

2) Relace $|A| = |B|$ je podle Faktu 2c.2 (i) reflexivní, podle 2c.2 (iii) symetrická a podle Faktu 2c.3 (iii) také tranzitivní. Není ale obecně antisymetrická, například máme $|\{13, 23\}| = |\{13, 14\}|$ a $|\{13, 14\}| = |\{13, 23\}|$, ale neplatí $\{13, 14\} = \{13, 23\}$. Zkuste si zase udělat rozbor, pro jaké soubory množin \mathcal{M} bychom zde antisymetrii dostali.

Mimochodem, před chvílí jsme říkali, že rovnost objektů je vždy symetrická i antisymetrická. Tento příklad tomu neprotiřečí, protože zde neporovnáváme přímo objekty, ale nějaké jejich vlastnosti (mohutnost). Jde tedy o něco jiného.

△

Příklad 3b.h: Nechť A je množina všech lidí. Definujeme relaci aRb jestliže a a b mají stejnýho otce či stejnou matku.

Protože každý má stejnýho otce či matku sám se sebou, je tato relace reflexivní. Evidentně je také symetrická, protože ve výroku „ a a b mají stejnou matku/otce“ je role a a b symetrická. Ani náhodou nebude antisymetrická: Jestliže mají a a b stejnou matku/otce a také b a a mají stejnou matku/otce, pak rozhodně nemusí jít o stejnou osobu, tedy neplatí $a = b$.

Tranzitivita: Nechť aRb a bRc . Pak mají a a b stejnýho rodiče a také b a c mají stejnýho rodiče. Znamená to, že pak i a a c sdílejí rodiče? Neznamená. Představme si následující situaci. Paní M_1 má s panem O_1 dítě a . Poté si pan O_1 udělá dítě b s paní M_2 a stává se tak spojovacím můstkom, díky kterému aRb (mají shodného otce O_1). Paní M_2 už má z předchozího manželství s panem O_2 dítě c , takže mají b a c společnou matku M_2 a tudíž bRc . Jenže a a c nemají ani společnou matku, ani společného otce a tudíž neplatí aRc , tato relace tedy není tranzitivní.

Poznámka: I zde lze vlastnosti ovlivnit tím, že měníme množinu, na které tu relaci děláme. Pokud si například vezmeme množinu lidí, kde jsou všichni jedináčci, tak se naše relace náhle stane antisymetrickou. Možná zajímavější je, že pokud v naší množině A má každý otce/matku ve stálém svazku bez „bokovek“.

△

V příkladech jsme se setkali s tím, že množina, na které relaci máme, může ovlivnit vlastnosti. S tím souvisí pojem restrikce, kdy máme relaci na množině A , ale aplikujeme ji na nějakou podmnožinu B . Pak máme následující.

!

Fakt 3b.4.

Nechť R je relace na množině A , nechť S je restrikce R na nějakou $B \subseteq A$. Jestliže má R některou z výše definovaných čtyř vlastností, tak ji má S také.

Důkaz (rutinní): Předpokládejme, že R je tranzitivní. Nechť $a, b, c \in B$ jsou takové, že $(a, b) \in S$ a $(b, c) \in S$. Protože $S \subseteq R$, pak $(a, b) \in R$ a $(b, c) \in R$ a z tranzitivity R plyne $(a, c) \in R$. Ale také $(a, c) \in B \times B$, proto $(a, c) \in S$. Tedy i S je tranzitivní.

Ostatní vlastnosti viz cvičení 3b.13.

□

Zjednodušeně řečeno, přechodem k podmnožině se vlastnosti nepokazí, ale mohou zlepšit. Když jsme tedy dokázali, že rovnost = je reflexivní, symetrická a tranzitivní jako relace na \mathbb{R} , pak má tyto vlastnosti i tehdy, když ji uvažujeme na \mathbb{N} nebo třeba na množině všech prvočísel.

Zde je třeba jisté opatrnosti, mluvíme o podmnožině základní množiny A , na které relace R „žije“. Ona je totiž i samotná relace R množinou (dvojic), tudíž můžeme uvažovat její podmnožiny. Když přejdeme k restrikci, dostáváme skutečně podmnožinu relace R , jmenovitě množinu těch dvojic z R , které nepoužívají prvky mimo B . Restrikce tedy vytváří podmnožiny relace, ale ne podmnožiny ledajaké, jsou vybírané speciálním způsobem. Proto se zachovávají vlastnosti relace. Jakmile začneme vytvářet podmnožiny R jinak, tak už zachování vlastností garantovat nelze.

Příklad: $A = \{1, 2, 3\}$, uvažujme relaci $R = \{(1, 2), (2, 1), (1, 3), (3, 1)\}$, která je symetrická. Když se omezíme na množinu $B = \{1, 2\}$, tak z relace zbyde její restrikce $S = \{(1, 2), (2, 1)\}$. I ta je symetrická, to souhlasí.

Pokud ale začneme mluvit o podmnožinách relace obecně, pak si z relace R můžeme vybrat třeba podmnožinu $T = \{(1, 2), (2, 1), (1, 3)\}$ a máme relaci, která už není symetrická.

Podobný příklad jsme již viděli, relace „být příbuzný“ je symetrická, ta má jako podmnožinu relaci „být rodičem“ a ta už symetrická není.

Poslední poznámka: Reflexivita, symetrie i tranzitivita vyžadují přítomnost určitých šipek, to se snadno pokazí odebráním nevhodně vybraných dvojic z relace. Naopak antisimetrie chce, aby určité šipky nebyly, což je požadavek v opačném směru. Nepřekvapí proto, že když máme antisymetrickou relaci a přejdeme k její podmnožině, tak zase dostaneme antisymetrickou relaci (viz důkaz (iii) Faktu 3c.2).

Zkoumání, jak si vlastnosti poradí s operacemi, necháme na další kapitolu, teď se podíváme, jak poznat platnost či neplatnost vlastností nepřímo, pomocí jiných pojmu. U některých vlastností se nám bude hodit jedna speciální relace.

! Definice.

Nechť A je množina. Definujeme její **diagonální relaci** (**diagonal relation**) jako relaci

$$\Delta(A) = \{(a, a) \in A \times A; a \in A\}.$$

Je to tedy relace ze všech „smyček“ a odpovídá relaci rovnosti na množině A , je totiž dána pro prvky $x, y \in A$ předpisem $(x, y) \in \Delta(A) \iff x = y$. Její reprezentace maticí je jednotková matici E_n .

! Věta 3b.5.

Nechť R je relace na nějaké množině A .

- (i) R je reflexivní právě tehdy, když $\Delta(A) \subseteq R$.
- (ii) R je symetrická právě tehdy, když $R = R^{-1}$.
- (iii) R je antisymetrická právě tehdy, když $R \cap R^{-1} \subseteq \Delta(A)$
- (iv) R je tranzitivní právě tehdy, když $R^2 \subseteq R$.

Tato věta by se v typickém matematickém textu dokázala větou „Důkaz je zřejmý“. A taky je, proto jsme jej nechali jako cvičení 3b.15.

Spojením (iv) a Faktu 3a.6 dostaneme, že pro tranzitivní relace platí $R^n \subseteq R$. To se dá zajímavě vylepšit, protože reflexivita zase dává inkluzi opačnou (viz cvičení 3b.16). Dostáváme tak následující tvrzení.

Věta 3b.6.

Nechť R je relace na množině A . Jestliže je R reflexivní a tranzitivní, tak $R^n = R$ pro všechna $n \in \mathbb{N}$.

Věta 3b.5 spojuje vlastnosti relací s množinovými operacemi. Ty jsou zase díky Faktům 3a.8 a 3a.9 a Důsledku 3a.11 svázány s operacemi. Závěr se nabízí, jen ještě potřebujeme zjistit, jak se pomocí matic pozná inkluze relací.

Definice.

Nechť M, N jsou dvě matice $m \times n$. Píšeme $M \leq N$, jestliže $m_{ij} \leq n_{ij}$ pro všechna $i = 1, \dots, m$ a $j = 1, \dots, n$.

Zde je třeba upozornit, že nejde o standardní nerovnost mezi maticemi, taková neexistuje. Přesněji, neexistuje pojem nerovnosti mezi maticemi, který by byl univerzálně užitečný a tudíž univerzálně přijímaný. Některé obory matice nerovností neporovnávají vůbec, některé pak mají svou vlastní definici, která jim vyhovuje. Nám vyhovuje ta, kterou jsme zavedli.

Fakt 3b.7.

Nechť jsou R_1 a R_2 relace na stejně množině A s maticemi M_1 a M_2 . Pak $R_1 \subseteq R_2$ právě tehdy, když $M_1 \leq M_2$.

Důkaz necháváme jako cvičení 3b.17. Je založen na tom, že každý prvek a nějaké 01-matice má povoleny pouze hodnoty 0 a 1, takže nerovnost $1 \leq a$ již nutně značí $a = 1$, naopak $a \leq 0$ značí $a = 0$.

Kombinací výše citovaných vět okamžitě dostáváme následující:

Věta 3b.8.

Nechť je R relace na nějaké n -prvkové množině A a nechť je M její reprezentující matici.

- (i) R je reflexivní právě tehdy, když $E_n \leq M$.
- (ii) R je symetrická právě tehdy, je-li M symetrická.
- (iii) R je antisymetrická právě tehdy, když $M \wedge M^T \leq E$.
- (iv) R je tranzitivní právě tehdy, když $M^{[2]} \leq M$.

Důkaz je opět snadný a necháme jej z větší části jako cvičení 3b.18, jen podotkněme jako návod, že v části (i) ta nerovnost nutí M mít na diagonále jedničky.

3b.9 Kartézský součin.

Před chvílí jsme se naučili „porovnávat“ matice podle toho, co dělají jejich prvky. Obecně jsme někdy v situaci, kdy pracujeme s komplikovanými objekty, které jsou poskládány ze složek. My už nějaké relace mezi složkami máme a rádi bychom z toho odvodili relaci o celých objektech.

Takto obecně odpověď existuje, protože ony složky se mohou na chování celého objektu podílet mnoha různými způsoby, což přirozeně ovlivní, jakým způsobem se vlastnosti složek přenáší na celek. Uděláme tedy obvyklou věc, zaměříme se jen na určitý typ struktury. Velice oblíbený objekt se složkami je vektor. Položíme si tedy otázku, zda bychom uměli vytvořit relaci mezi vektory (tedy prvky kartézského součinu) za předpokladu, že už máme relace pracující s jejich složkami. Kupodivu ani s vektory to není tak jednoduché. Představíme si zde jednu populární a jednoduchou myšlenku, která se v zásadě nabízí sama. Vzápětí ukážeme, že tento přirozený nápad má jisté nedostatky, které mohou (a nemusí) mrzet.

Definice.

Nechť $n \in \mathbb{N}$, pro $i \in \{1, 2, \dots, n\}$ nechť R_i je relace z nějaké množiny A_i do množiny B_i . Označme $A = A_1 \times A_2 \times \dots \times A_n$ a $B = B_1 \times B_2 \times \dots \times B_n$. Definujeme relaci R z A do B zvanou **součinové uspořádání (product order)** následovně: Pro $(a_1, \dots, a_n) \in A$, $(b_1, \dots, b_n) \in B$ platí $(a_1, \dots, a_n)R(b_1, \dots, b_n)$ právě tehdy, když $a_i R_i b_i$ pro všechna $i = 1, \dots, n$.

Takže chceme-li vědět, zda jsou dva vektory v relaci, tak se podíváme, zda to platí pro všechny složky. Nejčastěji se pracuje se situací, kdy $A_i = B_i$, jinými slovy, máme relace R_i na množinách A_i a chceme porovnávat vektory z $A_1 \times \dots \times A_n$.

Příklad 3b.i: Uvažujme jako R_1 relaci $<$ na $A_1 = \mathbb{R}$ a jako R_2 relaci dělitelnosti na $A_2 = \mathbb{N}$, aR_2b v případě, že a dělí b neboli b je násobek a (viz kapitola 6). Pracujeme pak na množině $A = \mathbb{R} \times \mathbb{N}$, kde vzniká součinová relace R . Máme například $(-1, 4)R(\pi, 12)$ nebo $(2, 13)R(e, 26)$, ale už neplatí $(1, 6)R(2, 8)$, protože 6 nedělí 8. Neplatí také $(1, 9)R(0, 18)$, kde selhává relace $<$ u první složky, a už vůbec ne $(13, 13)R(11, 23)$.

Nějaké závěry z téhle relace dělat nebudeme, byl to jen takový příklad, ať se ujistíme, že tomu opravdu dobře rozumíme.

△

Asi nejčastěji jsou všechny množiny A_i, B_i stejné.

Příklad 3b.j: Uvažujme relaci $=$ na \mathbb{R} . Pokud zvolíme $A_1 = A_2 = \dots = A_n = B_1 = \dots = B_n = \mathbb{R}$, pak $A = A_1 \times A_2 \times \dots \times A_n = \mathbb{R}^n = B$, vznikne tedy relace na klasickém prostoru reálných vektorů. Podle definice jsou dva vektory $(x_1, \dots, x_n), (y_1, \dots, y_n)$ spolu v součinové relaci R právě tehdy, pokud pro všechna i máme $x_i = y_i$.

Jinými slovy, v tomto příkladě nám vzniká běžná rovnost vektorů.

△

Součinovou relaci nejčastěji používáme právě takto, když jsou všechny relace R_i vlastně stejně. Pro výslednou relaci R se pak dokonce často používá stejná značka jako pro původní R_i , což je zrovna případ rovnosti vektorů.

Je snadné ukázat, že pokud mají všechny relace R_i nějakou z vyšetřovaných vlastností, tak ji má příslušná součinová relace také.

Fakt 3b.10.

Nechť $n \in \mathbb{N}$, pro $i \in \{1, 2, \dots, n\}$ nechť R_i je relace na nějaké množině A_i . Uvažujme odpovídající součinovou relaci R na $A = A_1 \times A_2 \times \dots \times A_n$. Pak platí:

- (i) Jestliže jsou pro všechna $i \in \{1, 2, \dots, n\}$ relace R_i reflexivní, pak je i relace R reflexivní.
- (ii) Jestliže jsou pro všechna $i \in \{1, 2, \dots, n\}$ relace R_i symetrická, pak je i relace R symetrická.
- (iii) Jestliže jsou pro všechna $i \in \{1, 2, \dots, n\}$ relace R_i antisymetrická, pak je i relace R antisymetrická.
- (iv) Jestliže jsou pro všechna $i \in \{1, 2, \dots, n\}$ relace R_i tranzitivní, pak je i relace R tranzitivní.

Důkaz (rutinní): R: Nechť $(a_i) = (a_1, \dots, a_n) \in A$. Protože jsou všechna R_i reflexivní, tak $a_i R_i a_i$ pro všechna i , tedy dle definice součinové relace i $(a_i)R(a_i)$.

Ostatní tři vlastnosti s důvěrou přenecháme čtenáři, viz cvičení 3b.19

□

Příklad 3b.k: Uvažujme relaci $<$ na \mathbb{R} . Součinová relace s volbou $A_1 = A_2 = \dots = A_n = \mathbb{R}$ nám dá jakési porovnávání vektorů z \mathbb{R}^n , které zase můžeme značit $<$.

Například při volbě $n = 2$ můžeme psát $(-1, 4) < (0, 13)$, protože platí $-1 < 0$ a také $4 < 13$.

△

Zatímco příklad s rovností ukázal, že pojem součinové relace má své využití, tento příklad naopak ukazuje, že v některých případech součinová relace rozhodně není to pravé. Kde je problém?

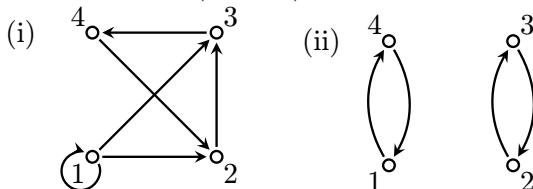
Pokud vektory používáme v geometrii, tak to dává špatné výsledky. Například v součinové nerovnosti platí $(-4, -3) < (1, 0)$, jenže to je zcela proti intuici. U vektorů je totiž důležitá hlavně magnituda a zde máme $\|(-4, -3)\| = 5$ a $\|(1, 0)\| = 1$, takže ten první vektor cítíme jako větší. Proto se v analýze vektory podle složek zásadně neporovnávají, podobný problém je s porovnáváním matic podle „velikosti“.

Zde ovšem nejsme v analýze ale v diskrétní matematice, máme i my se součinovou nerovností problém? Ano, a rovněž podstatný. Součinovým uspořádáním totiž vzniká jen velice málo srovnání, protože aby se shodly všechny složky, musíme mít velké štěstí. Třeba vektory $(1, 2)$ a $(2, 1)$ touto relací nedokážeme porovnat. To může být smrtící. Typickým příkladem je řazení lidí podle abecedy. Budeme-li slova považovat za vektory písmen, přičemž pro jednotlivá písmena máme srovnání dle abecedy, pak součinová relace není schopna porovnat například jména uk a Gek. V první složce přijde nejdřív to druhé slovo ($<G$), ve druhé složce je to naopak ($u>e$) a tento rozpor součinová relace nerozechodí.

Jenže my potřebujeme umět řadit slova nějak za sebe, bude tedy třeba vymyslet jiný způsob, jak z informace o složkách vyrábět relaci na celých vektorech. Na to si ale počkáme do kapitoly o částečném uspořádání, viz 4b.17.

Cvičení

Cvičení 3b.1 (rutinní): Uvažujte relace dané následujícími diagramy.



Vyšetřete pro ně čtyři základní vlastnosti.

Poznámka: Vyšetřit znamená zjistit, zda určitá vlastnost platí, a tuto odpověď dokázat.

Cvičení 3b.2 (rutinní): Uvažujme následující relace na množině $A = \{1, 2, 3, 4\}$:

$$(i) R = \{(1, 1), (1, 4), (2, 1), (3, 4)\}; \quad (ii) R = \{(1, 4), (1, 3), (4, 3), (2, 2)\}.$$

a) Vyšetřete pro ně čtyři základní vlastnosti.

b) Nakreslete jejich grafy, ověřte si na nich výsledky z a).

Cvičení 3b.3 (rutinní): Pro relace zadané následujícími maticemi vyšetřete čtyři základní vlastnosti.

$$(i) M = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}; \quad (ii) M = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad (iii) M = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Cvičení 3b.4 (rutinní): Uvažujte relaci R na množině lidí A danou aRb jestliže a a b sdílí křestní jméno.

Vyšetřete, kterou ze základních čtyř vlastností má.

Cvičení 3b.5 (rutinní, zkouškové, * dobré): Pro následující relace na \mathbb{Z} vyšetřete čtyři základní vlastnosti.

- | | |
|--|--|
| (i) aRb jestliže $ a = b $; | (v) aRb jestliže $a = b + 1$; |
| (ii) aRb jestliže $a \geq b$; | (vi)* aRb jestliže $a \geq b^2$ (viz následující cvičení); |
| (iii) aRb jestliže $a \neq b$; | (vii) aRb jestliže a a b mají nějakého společného dělitele různého od 1. |
| (iv) aRb jestliže $a - b = 2k$ pro nějaké $k \in \mathbb{Z}$; | |

Cvičení 3b.6 (rutinní, zkouškové, * dobré): Pro následující relace na \mathbb{R} vyšetřete čtyři základní vlastnosti.

- | | |
|--|--|
| (i) xRy jestliže $y - x \in \mathbb{Z}$; | (v) xRy jestliže $x = y^2$; |
| (ii) xRy jestliže $x - y \in \mathbb{Q}$; | (vi)* xRy jestliže $x \geq y^2$ (viz předchozí cvičení); |
| (iii) xRy jestliže $xy \geq 0$; | (vii) xRy jestliže $ x \leq y $. |
| (iv) xRy jestliže $xy \geq 1$; | |

Cvičení 3b.7 (rutinní, zkouškové, dobré): Vyšetřete čtyři základní vlastnosti pro následující relace:

- (i) Relace R na množině \mathbb{R}^2 definovaná takto: $(u, v)R(x, y)$ jestliže $u^2 - y = x^2 - v$,
tedy formálně $R = \{((u, v), (x, y)) \in \mathbb{R}^2 \times \mathbb{R}^2; u^2 - y = x^2 - v\}$.

- (ii) Relace R na množině \mathbb{R}^2 definovaná takto: $(u, v)R(x, y)$ jestliže $u^2 - y = v^2 - x$, tedy formálně $R = \{((u, v), (x, y)) \in \mathbb{R}^2 \times \mathbb{R}^2; u^2 - y = v^2 - x\}$.
- (iii) Relace R na množině \mathbb{R}^2 definovaná takto: Uvažujme množinu $N = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 = 13\}$ (mimochedem, kružnice okolo počátku s poloměrem $\sqrt{13}$). Definujme $R = \{((u, v), (x, y)) \in \mathbb{R}^2 \times \mathbb{R}^2; (u, v) - (x, y) \in N\}$.
- (iv) Relace R na množině \mathbb{R}^2 definovaná takto: Uvažujme množinu $N = \{(x, y) \in \mathbb{R}^2; x + y = 0\}$ (mimochedem, vedlejší diagonála). Definujme $R = \{((u, v), (x, y)) \in \mathbb{R}^2 \times \mathbb{R}^2; (u, v) - (x, y) \in N\}$.
- (v) Relace \mathcal{R} na množině F všech zobrazení $\mathbb{Z} \mapsto \mathbb{Z}$ definovaná jako $T\mathcal{RS}$ jestliže $T(0)S(0) = 2$.
- (vi) Relace \mathcal{R} na množině F všech zobrazení $\mathbb{Z} \mapsto \mathbb{Z}$ definovaná jako $T\mathcal{RS}$ jestliže $T(1) = S(2)$.
- (vii) Relace \mathcal{R} na množině F všech funkcí $\mathbb{R} \mapsto \mathbb{R}$ definovaná jako $f\mathcal{R}g$ jestliže $f(x) \geq g(y)$ pro všechna $x \in \mathbb{R}$.
- (viii) Relace \mathcal{R} na množině $M_{2 \times 2}$ všech 2×2 reálných matic definovaná jako $A\mathcal{R}B$ jestliže $|A| = |B|$ (stejný determinant).
- (ix) Relace \mathcal{R} na množině $M_{2 \times 2}$ všech 2×2 reálných matic definovaná jako $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \mathcal{R} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ jestliže $a_{11} = b_{22}$.
- (x) Relace R na množině P všech reálných polynomů definovaná jako $p(x)Rq(x)$ jestliže mají p a q stejný stupeň.
- (xi) Relace R na množině P všech reálných polynomů definovaná jako $p(x)Rq(x)$ jestliže mají p a q stejné reálné kořeny včetně násobnosti.
- (xii) Relace R na množině P všech reálných polynomů definovaná jako $p(x)Rq(x)$ jestliže mají p a q stejné komplexní kořeny včetně násobnosti.
- Viz též cvičení 4a.5 a 4a.6.

Cvičení 3b.8 (rutinní): Nechť A je množina řetězců z 26 malých anglických písmen. Určete vlastnosti relací daných na A následujícími podmínkami:

- (i) $\alpha R \beta$ jestliže řetězce α a β nemají žádná společná písmena;
- (ii) $\alpha R \beta$ jestliže řetězce α a β sdílejí nějaká písmena;
- (iii) $\alpha R \beta$ jestliže řetězce α a β nemají stejnou délku;
- (iv) $\alpha R \beta$ jestliže je řetězec α delší než β .

Cvičení 3b.9 (poučné, dobré): Uvažujme množinu $A = \{a, b, c\}$. Máme čtyři vlastnosti (reflexivita, symetrie, antisimetrie, tranzitivita), každá může a nemusí být splněna, což dává celkem $2^4 = 16$ možností. Pro každou možnost vytvořte nějaký příklad relace na A .

Cvičení 3b.10 (poučné): Dokažte, že pro libovolnou množinu A je $\Delta(A)$ reflexivní, symetrická, antisimetrická a tranzitivní.

Cvičení 3b.11 (poučné): Dokažte, že pro libovolnou množinu A a relaci R platí, že jestliže je R symetrická i antisimetrická, pak $R \subseteq \Delta(A)$.

Cvičení 3b.12 (poučné, zkouškové): (i) Dokažte, že pro každou relaci R na libovolné množině A jsou $R^{-1} \circ R$ a $R \circ R^{-1}$ symetrické relace.

(ii) Jestliže pro každé $a \in A$ existuje nějaké $b \in A$ takové, že $(a, b) \in R$, pak jsou $R^{-1} \circ R$ i $R \circ R^{-1}$ reflexivní relace.

Cvičení 3b.13 (rutinní): Nechť R je relace na množině A , nechť S je restrikce R na nějakou $B \subseteq A$. Dokažte následující (viz Fakt 3b.4):

- (i) Jestliže R je reflexivní, pak S je reflexivní.
- (ii) Jestliže R je symetrická, pak S je symetrická.
- (iii) Jestliže R je antisimetrická, pak S je antisimetrická.

Cvičení 3b.14 (dobré, zkouškové): Nechť R je relace na množině A , která je reflexivní a tranzitivní. Definujme relaci S na A předpisem aSb právě tehdy, jestliže aRb a bRa (rozmyslete si, že $S = R \cap R^{-1}$).

Dokažte, že pak relace S je reflexivní, symetrická a tranzitivní.

Poznámka: Tímto způsobem lze z relace \leq vytvořit relaci $=$ na číslech.

Cvičení 3b.15 (rutinní, zkouškové): Nechť R je relace na nějaké množině A . Dokažte následující (viz Věta 3b.5):

- (i) R je reflexivní právě tehdy, když $\Delta(A) \subseteq R$.
- (ii) R je symetrická právě tehdy, když $R = R^{-1}$.
- (iii) R je antisimetrická právě tehdy, když $R \cap R^{-1} \subseteq \Delta(A)$
- (iv) R je tranzitivní právě tehdy, když $R^2 \subseteq R$.

Cvičení 3b.16 (rutinní, poučné, zkouškové): Dokažte následující tvrzení pro relaci R na množině A . Jestliže je R reflexivní, pak $R \subseteq R^n$ pro všechna $n \in \mathbb{N}$.

Cvičení 3b.17 (rutinní, poučné, zkouškové): Nechť jsou R_1 a R_2 relace na stejně množině A s maticemi M_1 a M_2 . Dokažte, že $R_1 \subseteq R_2$ právě tehdy, když $M_1 \leq M_2$ (viz Fakt 3b.7).

Cvičení 3b.18 (rutinní, poučné, zkouškové): Nechť je R relace na nějaké n -prvkové množině A a nechť je M její reprezentující matice. Dokažte následující (viz Věta 3b.8):

- (i) R je reflexivní právě tehdy, když $E_n \leq M$.
- (ii) R je symetrická právě tehdy, je-li M symetrická.
- (iii) R je tranzitivní právě tehdy, když $M^{[2]} \leq M$.

Cvičení 3b.19 (rutinní): Nechť pro $i = 1, \dots, n$ je R_i relace na množině A_i . Nechť R je odpovídající součinová relace na $A = A_1 \times \dots \times A_n$. Dokažte následující:

- (ii) Jestliže jsou všechny R_i symetrické, pak je i R symetrická.
- (ii) Jestliže jsou všechny R_i antisymetrické, pak je i R antisymetrická.
- (iii) Jestliže jsou všechny R_i tranzitivní, pak je i R tranzitivní.

Viz Fakt 3b.10

Cvičení 3b.20 (poučné, dobré): Najděte chybu v následujícím „důkazu“, že jestliže je nějaká relace R na A symetrická a tranzitivní, pak už musí být i reflexivní:

Nechť $a \in A$. Vezměme $b \in A$ takové, aby aRb . Podle symetrie pak bRa , máme řetězec $aRbRa$, tedy podle tranzitivity aRa .

Řešení:

3b.1: (i): R: ne, chybí smyčka třeba u 2; S: ne, protože třeba $1R2$, ale neplatí $2R1$; A: ano, jediný případ s aRb a bRa je $a = b = 1$, což je v pořádku (alternativa: pro $a \neq b$ nikdy nevedou šipky oběma směry); T: ne, je $1R3$ a $3R4$, ale není $1R4$.

(ii): R: ne, chybí smyčka třeba u 1; S: ano, pro každou šipku je i zpět; A: ne, jsou dvojité šipky mezi dvěma různými prvky; T: ne, je $1R4$ a $4R1$, ale není $1R1$.

3b.2: a) (i): R: ne, chybí třeba $(2, 2)$; S: ne, protože třeba $(1, 4) \in R$, ale $(4, 1) \notin R$; A: ano, jediný případ s $(a, b) \in R$ a $(b, a) \in R$ je $a = b = 1$, což je v pořádku; T: ne, $(2, 1) \in R$ a $(1, 4) \in R$, ale $(2, 4) \notin R$.

a) (ii): R: ne, chybí třeba $(1, 1)$; S: ne, protože třeba $(1, 4) \in R$, ale $(4, 1) \notin R$; A: ano, jediný případ s $(a, b) \in R$ a $(b, a) \in R$ je $a = b = 2$, což je v pořádku; T: ano, jediný navazující případ je $(1, 4) \in R$ a $(4, 3) \in R$, opravdu $(1, 3) \in R$.

3b.3: Testy: Reflexivita znamená, že matice musí mít 1 všude na diagonále. Symetrie: matice musí být symetrická. Antisimetrie: porovnáváme dvojice čísel symetrické podle diagonály, nesmí být obě zároveň 1. Tranzitivita: Jedna možnost je použít Booleanovský součin. Druhá možnost je vypsat si z matice všechny nediagonální jedničky jako dvojice v relaci a zkoumat tranzitivitu na nich.

(i): R,T; (ii): S; (iii): R,A,T.

3b.4: R,S,T.

3b.5: (i): R: Pro každé $a \in \mathbb{Z}$ platí $|a| = |a|$, proto aRa . Reflexivní.

S: Libovolné $a, b \in \mathbb{Z}$ splňující aRb , to dává $|a| = |b|$, proto $|b| = |a|$ a tedy bRa . Symetrická.

A: Libovolné $a, b \in \mathbb{Z}$ splňující aRb a bRa , to dává $|a| = |b|$ a $|b| = |a|$, z toho asi $a = b$ nedostaneme. Protipříklad: $|-13| = |13|$, proto $13R(-13)$ a $(-13)R13$, ale neplatí $-13 = 13$, tedy R není antisymetrická.

T: Libovolné $a, b, c \in \mathbb{Z}$ splňující aRb a bRc , to dává $|a| = |b|$ a $|b| = |c|$, z toho hned máme $|a| = |c|$ a tedy aRc . R je tranzitivní.

(ii): R: ano, pro $a \in A$ je $a \geq a$, tedy aRa ; S: ne, $2R1$ neboť $2 \geq 1$, ale neplatí $1 \geq 2$ tedy neplatí $1R2$;

A: ano, $aRb \wedge bRa \implies a \geq b \wedge b \geq a \implies a = b$; T: ano, $aRb \wedge bRc \implies a \geq b \wedge b \geq c \implies a \geq c \implies aRc$.

(iii): R: ne, třeba neplatí $1 \neq 1$ proto neplatí $1R1$; S: ano, $aRb \implies a \neq b \implies b \neq a \implies bRa$;

A: ne, třeba $1R2 \wedge 2R1$, ale neplatí $1 = 2$; T: ne, třeba $1R2$ a $2R1$, ale neplatí $1R1$.

(iv): R: ano, $a - a = 2 \cdot 0 \implies aRa$ pro každé a ; S: ano, $aRb \implies a - b = 2k \implies b - a = 2(-k) \implies bRa$;

A: ne, třeba $1R3$ a $3R1$, přesto neplatí $1 = 3$;

T: ano, $aRb \wedge bRc \implies a - b = 2k \wedge b - c = 2l \implies a - c = 2(k + l) \implies aRc$.

(v): R: ne, neplatí $13 = 13 + 1$ a proto neplatí $13R13$; S: ne, $2R1$ ale neplatí $1R2$;

A: ano, $aRb \wedge bRa \implies a = b + 1 \wedge b = a + 1 \implies b = b + 2 \implies 0 = 2$ spor, předpoklad nikdy nenastane, proto implikace vždy platí; T: ne, třeba $3R2$ a $2R1$, ale neplatí $3R1$.

(vi): Není R viz $a = 2$; není S viz $4R2$; A: $aRb \wedge bRa \implies a \geq b^2 \wedge b \geq a^2$. Pokud $a = 0$, tak to dává $0 \geq b^2 \implies b = 0 = a$. Pokud $a \neq 0$, pak $|a| \geq 1$, také $a \geq b^2 \geq 0$ a proto $a \geq 1$, podobně $b \geq 1$. Počítáme: $a \geq b^2 \wedge b \geq a^2 \implies a \geq b^2 \geq a^4 \implies a \geq a^4 \implies 1 \geq a^3$, spolu s $a \geq 1$ to dává $a = 1$. Pak $1 \geq b^2 \geq 1 \implies b = 1$ a zase $a = b$. Relace je antisymetrická.

T: Pro $b \in \mathbb{Z}$ platí $b^2 \geq b$ (viz A), proto $aRb \wedge bRc \implies a \geq b^2 \wedge b \geq c^2 \implies a \geq b \geq c^2 \implies a \geq c^2 \implies aRc$.
Je tranzitivní.

(vii): R: Má každé $a \in \mathbb{Z}$ nějakého společného dělitele samo se sebou jiného než 1? Skoro ano, neplatí to pro $a = 1$.
Takže R není reflexivní.

S: Nechť $a, b \in \mathbb{Z}$ splňují aRb . Pak existuje $c > 1$, které dělí a i b, to pak dělí i b a a, tedy bRa . R je symetrická.

A: $aRb \wedge bRa$ dává společného dělitele, není šance vynutit $a = b$. Protipříklad: $2R4$ a $4R2$ (společný dělitel 2), proto není antisymetrická.

T: a, b mají společného dělitele > 1 , b, c mají společného dělitele > 1 , z toho nic společného pro a, c neplyne.
Protipříklad: $2R6$ a $6R3$, ale neplatí $2R3$. Není tranzitivní.

3b.6: (i): R,S,T, viz příklad 4a.e;

(ii): R ano $x - x = 0 \in \mathbb{Q}$, S ano $y - x \in \mathbb{Q} \implies x - y = -(y - x) \in \mathbb{Q}$, T ano $y - x \in \mathbb{Q} \wedge (z - y) \in \mathbb{Q} \implies (z - x) = (y - x) + (z - y) \in \mathbb{Q}$; není A viz $1R2$ a $2R1$;

(iii): R ano $xx = x^2 \geq 0$, S ano $xy \geq 0 \implies yx \geq 0$; není A viz $1R2$ a $2R1$; není T viz $(-1)R0$ a $0R1$;

(iv): Není R viz $x = 0$, S ano $xy \geq 1 \implies yx \geq 1$; není A viz $2R1$ a $1R2$; není T viz $\frac{1}{2}R4$ a $4R1$;

(v): Není R viz $x = 2$; není S viz $4R2$; A ano $x = y^2 \wedge y = x^2 \implies x, y \geq 0 \wedge x = x^4 \wedge y = y^4 \implies x = y = 1 \vee x = y = 0$; není T viz $16R4$ a $4R2$;

(vi): Není R viz $x = 2$; není S viz $4R2$; není A viz $x = 0.1, y = 0.2$ neboť $0.1 \geq (0.2)^2$ a $0.2 \geq (0.1)^2$ ale neplatí $0.1 = 2$; není T viz $(0.5)R(0.7)$ neboť $0.5 \geq (0.7)^2 = 0.49$, $(0.7)R(0.8)$ neboť $0.7 \geq 0.64$, ale není $0.5 \geq 0.64$ (tohle asi bylo drobet zákerné).

(vii): R ano $|x| \leq |x|$, T ano $|x| \leq |y| \wedge |y| \leq |z| \implies |x| \leq |z|$; není S viz $1R2$, není A viz $1R(-1)$ a $(-1)R2$.

3b.7: (i): R: ano $u^2 - v = u^2 - v \implies (u, v)R(u, v)$; S: $(u, v)R(x, y) \implies u^2 - y = x^2 - v$

$\implies x^2 - v = u^2 - y \implies (x, y)R(u, v)$ ano; A: ne viz třeba $(1, 4)R(2, 1)$ a $(2, 1)R(1, 4)$;

T: ano; $(s, t)R(u, v) \wedge (u, v)R(x, y) \implies s^2 - v = u^2 - t \wedge u^2 - y = x^2 - v$ sečist $s^2 - v + u^2 - y = u^2 - t + x^2 - v \implies s^2 - y = x^2 - t \implies (s, t)R(x, y)$.

(ii): R: ne viz třeba $(2, 3)$, neplatí $2^2 - 3 = 3^2 - 2$; S: ne viz třeba $(2, 1)R(1, 4)$ ale neplatí $(1, 4)R(2, 1)$; A: ne viz třeba $(1, 0)R(0, 1)$ a $(0, 1)R(1, 0)$; T: ne viz třeba $(1, 4)R(2, 1)$ a $(2, 1)R(1, 4)$ ale neplatí $(1, 4)R(1, 4)$.

(iii): přepis: $(u, v)R(x, y) \iff (u - x)^2 + (v - y)^2 = 13$; R: ne $(u - u)^2 + (v - v)^2 = 0 \neq 13$;

S: ano $(u, v)R(x, y) \implies (u - x)^2 + (v - y)^2 = 13 \implies (x - u)^2 + (y - v)^2 = 13 \implies (x, y)R(u, v)$; A: ne třeba $(4, 3)R(1, 1)$ a $(1, 1)R(4, 3)$; T: ne třeba $(4, 3)R(1, 1)$ a $(1, 1)R(4, 3)$ ale neplatí $(4, 3)R(4, 3)$.

(iv): přepis: $(u, v)R(x, y) \iff (u - x) + (v - y) = 0$; R: ano $(u - u) + (v - v) = 0$;

S: ano $(u, v)R(x, y) \implies (u - x) + (v - y) = 0 \implies (x - u) + (y - v) = 0 \implies (x, y)R(u, v)$; A: ne třeba $(1, 3)R(2, 2)$ a $(2, 2)R(1, 3)$; T: ano $(s, t)R(u, v) \wedge (u, v)R(x, y) \implies (s - u) + (t - v) = 0 \wedge (u - x) + (v - y) = 0$ sečist rovnice: $(s - x) + (t - y) = 0 \implies (s, t)R(x, y)$.

(v): R: ne, to by muselo každé zobrazení splňovat $T(0)T(0) = 2$, ale například zobrazení $T(n) = n + 1$ má $T(0)T(0) = 1 \cdot 1 = 1$;

S: ano $T\mathcal{RS} \implies T(0)S(0) = 2 \implies S(0)T(0) = 2 \implies S\mathcal{RT}$; A: ne třeba $T(n) = n + 1$, $S(n) = 3n + 2$, pak $T(0)S(0) = 1 \cdot 2 = 2 = S(0)T(0)$, tedy $T\mathcal{RS}$ a $S\mathcal{RT}$, ale není $T = S$; T: ne třeba $T(n) = n + 1$, $S(n) = 3n + 2$, $U(n) = (n + 1)^2$, pak $T\mathcal{RS}$ a $S\mathcal{RU}$, ale neplatí $T\mathcal{RU}$, protože $T(0)U(0) = 1$.

(vi): R: ne, to by muselo každé zobrazení splňovat $T(1) = T(2)$, ale například zobrazení $T(n) = n$ má $T(1) = 1$ a $T(2) = 2$;

S: ne, třeba $T(n) = n + 1$ a $S(n) = n$, pak $T(1) = 1 = S(2)$, proto $T\mathcal{RS}$, ale neplatí $S(1) = T(2)$; A: ne třeba $T(n) = (2n - 3)^2$, $S(n) = 1$ (konstantní zobrazení), pak $T(1) = 1 = S(2)$ a $S(1) = 1 = T(2)$, tedy $T\mathcal{RS}$ a $S\mathcal{RT}$, ale není $T = S$; T: ne třeba $T(n) = n + 1$, $S(n) = n$, $U(n) = n - 1$, pak $T\mathcal{RS}$ a $S\mathcal{RU}$, ale neplatí $T\mathcal{RU}$, protože $T(1) = 2$ a $U(2) = 1$.

(vii): R: ano, libovolná funkce f splňuje pro každé $x \in \mathbb{R}$ nerovnost $f(x) \geq f(x)$; S: ne, třeba $f(x) = x + 13$, $g(x) = x$ splňují $f\mathcal{R}g$ ale ne $g\mathcal{R}f$; A: ano, $f\mathcal{R}g$ a $g\mathcal{R}f$ znamená $f(x) \geq g(x)$ a $g(x) \geq f(x)$ pro všechna x neboli $f(x) = g(x)$ pro všechna x neboli $f = g$; T: ano, $f\mathcal{R}g$ a $g\mathcal{R}h$ dává pro všechna $x \in \mathbb{R}$: $f(x) \geq g(x)$ a $g(x) \geq h(x)$ neboli $f(x) \geq h(x)$, takže $f\mathcal{R}h$.

(viii): R: ano $|A| = |A|$; S: ano $A\mathcal{RB} \implies |A| = |B| \implies |B| = |A| \implies B\mathcal{RA}$; A: ne třeba matice ze samých nul a nenulová matice s opakujícími se řádky mají obě nulový determinant;

T: ano $A\mathcal{RB} \wedge B\mathcal{RC} \implies |A| = |B| \wedge |B| = |C| \implies |A| = |C| \implies A\mathcal{RC}$.

(ix): R: ne, třeba u matice $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ se nerovnají levý horní a pravý dolní roh, proto neplatí $A\mathcal{RA}$;

S: ne, třeba pro $A = \begin{pmatrix} 13 & 2 \\ -2 & 7 \end{pmatrix}$ a $B = \begin{pmatrix} 1 & -1 \\ 3 & 13 \end{pmatrix}$ platí $A\mathcal{RB}$, ale ne $B\mathcal{RA}$;

A: ne, třeba $A = \begin{pmatrix} 13 & 1 \\ -1 & 23 \end{pmatrix}$ a $B = \begin{pmatrix} 23 & 2 \\ 3 & 13 \end{pmatrix}$ splňují $A\mathcal{RB}$ a $B\mathcal{RA}$, ale nesplňují $A = B$;

T: ne, třeba $A = \begin{pmatrix} 13 & 1 \\ -1 & 23 \end{pmatrix}$, $B = \begin{pmatrix} 23 & 2 \\ 3 & 13 \end{pmatrix}$ a $C = \begin{pmatrix} 14 & -3 \\ 5 & 23 \end{pmatrix}$ splňují ARB a BRC , ale nesplňují ARC .

(x): R: ano $\text{st}(p) = \text{st}(p)$; S: ano $pRq \implies \text{st}(p) = \text{st}(q) \implies \text{st}(q) = \text{st}(p) \implies qRp$; A: ne třeba $p = x$ a $q = 2x + 1$; T: ano $pRq \wedge qRr \implies \text{st}(p) = \text{st}(q) \wedge \text{st}(q) = \text{st}(r) \implies \text{st}(p) = \text{st}(r) \implies pRr$;

(xi): R: ano; S: ano; A: ne třeba $p = x - 1$ a $q = 2x - 2$; T: ano;

(xii): R: ano; S: ano; A: ne třeba $p = x - 1$ a $q = 2x - 2$; T: ano;

3b.8: (i): S; (ii): R,S; (iii): S; (iv): A,T.

3b.9: Nic nemá $R = \{(a, b), (b, a), (b, c)\}$: Chybí (a, a) proto není R, k bRc chybí cRb proto není S, je tam aRb a bRa pro $a \neq b$ proto není A, je tam $aRbRc$ ale ne aRc proto není T.

Dále jen zkratky, třeba $R_{R,A}$ bude relace, která je reflexivní a antisymetrická, ale není nic jiného (to se také musí zajistit a ověřit).

$$R_T = \{(a, b), (b, a), (b, c), (a, a), (b, b), (a, c)\}, R_A = \{(a, b), (b, c)\}, R_{A,T} = \{(a, b)\}, R_S = \{(a, b), (b, a)\},$$

$$R_{S,T} = \{(a, a), (a, b), (b, a), (b, b)\}, R_{S,A,T} = \{(a, a)\}, R_R = \{(a, a), (b, b), (c, c), (a, b), (b, a), (b, c)\},$$

$$R_{R,T} = \{(a, a), (b, b), (c, c), (a, b), (b, c), (a, c)\}, R_{R,A} = \{(a, a), (b, b), (c, c), (a, b), (b, c)\},$$

$$R_{R,A,T} = \{(a, a), (b, b), (c, c), (a, b)\}, R_{R,S} = \{(a, a), (b, b), (c, c), (a, b), (b, a), (b, c), (c, b)\},$$

$$R_{R,S,T} = \{(a, a), (b, b), (c, c), (a, b), (b, a)\}, R_{R,S,A,T} = \{(a, a), (b, b), (c, c)\}.$$

Relaci, která je symetrická a antisymetrická, ale není tranzitivní, nelze vytvořit, viz cvičení 3b.11.

3b.10: R: přímo z definice $(a, a) \in \Delta(A)$; S: Jestliže $(a, b) \in \Delta(A)$, pak $b = a$, proto $(b, a) = (a, a) = (a, b)$ a tedy $(b, a) \in \Delta(A)$; A: Nechť $(a, b) \in \Delta(A) \wedge (b, a) \in \Delta(A)$, už z toho prvního je $a = b$;

T: Nechť $(a, b) \in \Delta(A) \wedge (b, c) \in \Delta(A)$, pak $a = b = c$ a tedy $(a, c) = (a, a) \in \Delta(A)$.

3b.11: Nechť $(a, b) \in R$. Ze symetrie také $(b, a) \in R$, tedy z antisimetrie $a = b$ a proto $(a, b) = (a, a) \in \Delta(A)$.

3b.12: (i) Pro $R^{-1} \circ R$: Nechť $(a, b) \in R^{-1} \circ R$. Pak $\exists x \in A$ aby $(a, x) \in R$ a $(x, b) \in R^{-1}$. Odtud podle definice inverzní relace $(x, a) \in R^{-1} \wedge (b, x) \in R \implies (b, x) \in R$ a $(x, a) \in R^{-1}$, tedy $(b, a) \in R^{-1} \circ R$.

Důkaz pro $R \circ R^{-1}$ je obdobný.

(ii): Nechť $a \in A$. Pak $\exists b \in A$: $(a, b) \in R$. Pak $(b, a) \in R^{-1}$, máme $aRbR^{-1}a$, tedy $(a, a) \in R^{-1} \circ R$.

3b.13: (i): Nechť $a \in B$. Pak $a \in A$, proto $(a, a) \in R$, a jelikož také $(a, a) \in B \times B$, tak $(a, a) \in S$.

(ii): Nechť $a, b \in B$ jsou takové, že $(a, b) \in S$. Protože $S \subseteq R$, pak $(a, b) \in R$ a ze symetrie R dostaneme $(b, a) \in R$. Ale také $(b, a) \in B \times B$, proto $(b, a) \in S$.

(iii): Nechť $a, b \in B$ jsou takové, že $(a, b) \in S$ a $(b, a) \in S$. Protože $S \subseteq R$, pak $(a, b) \in R$ a $(b, a) \in R$, z antisimetrie R pak hned dostaneme $a = b$.

3b.14: R: Pro $a \in A$ je díky reflexivitě S jak aSa , tak aSa (opačné pořadí), proto aRa .

S: Nechť aRb . Pak podle definice aSb a bSa , proto také bSa a aSb a tedy bRa .

T: Nechť aRb a bRc . Pak podle definice aSb a bSa a také bSc a cSa . Jinými slovy, platí aSb a bSc , ptoto z tranzitivity S máme aSc , podobně i cSa a tedy aRc .

3b.15: (i): 1) Je-li R reflexivní, pak $\forall a \in A$: $(a, a) \in R$, tedy $\Delta(A) \subseteq R$.

2): Jestliže naopak $\Delta(A) \subseteq R$, pak $\forall a \in A$: $(a, a) \in \Delta(A) \subseteq R$, tedy $(a, a) \in R$ a R je reflexivní.

(ii): 1) Nechť je R je symetrická. Chceme ukázat $R = R^{-1}$.

Nechť $a, b \in A$ splňují $(a, b) \in R$. Pak podle symetrie také $(b, a) \in R$ a proto dle definice inverzní relace $(a, b) \in R^{-1}$. Takže $R \subseteq R^{-1}$.

Naopak nechť $(a, b) \in R^{-1}$. Pak $(b, a) \in R$ a podle symetrie $(a, b) \in R$. Takže také $R^{-1} \subseteq R$.

2) Teď předpokládejme, že $R = R^{-1}$, potřebujeme ukázat, že R je symetrická.

Jestliže $a, b \in A$ jsou takové, že $(a, b) \in R$, pak podle $R^{-1} \subseteq R$ nutně musí být $(a, b) \in R^{-1}$ a tedy podle definice inverzní relace $(b, a) \in R$. R je tedy opravdu symetrická.

(iii): 1) Nechť je R je antisymetrická. Potřebujeme ukázat, že $R \cap R^{-1} \subseteq \Delta(A)$.

Vezměme libovolné $(a, b) \in R \cap R^{-1}$. Pak $(a, b) \in R$ a také $(a, b) \in R^{-1}$, tedy podle definice $(b, a) \in R$. Protože (a, b) i (b, a) jsou v R , podle antisymetrie nutně $a = b$ a proto $(a, b) \in \Delta(A)$.

2) Teď předpokládejme $R \cap R^{-1} \subseteq \Delta(A)$, potřebujeme ukázat, že R je antisymetrická. Nechť tedy $a, b \in A$ jsou takové, že $(a, b) \in R$ a $(b, a) \in R$. Z toho druhého máme $(a, b) \in R^{-1}$, takže $(a, b) \in R \cap R^{-1}$. To je ale podmnožina $\Delta(A)$, takže $(a, b) \in \Delta(A)$. To je ovšem možné jen tehdy, když $a = b$.

(iv): 1) Nejprve předpokládejme, že R je tranzitivní, potřebujeme ukázat, že $R^2 \subseteq R$.

Vezměme tedy libovolné $(a, b) \in R^2 = R \circ R$. Pak podle definice skládání musí existovat $x \in A$ takové, že $(a, x) \in R$ a $(x, b) \in R$. Podle tranzitivity potom také $(a, b) \in R$.

2) Předpokládejme naopak, že $R^2 \subseteq R$, chceme z toho vyvodit tranzitivitu.

Vezměme teď libovolné $a, b, c \in A$ takové, že $(a, b) \in R$ a $(b, c) \in R$. Pak podle definice skládání $(a, c) \in R \circ R = R^2$. Předpoklad pak dává $(a, c) \in R$.

3b.16: Indukcí: (0) $n = 1$: $R \subseteq R = R^1$.

(1) Předpoklad: $R \subseteq R^n$.

Nechť $(a, b) \in R$. Pak $(a, b) \in R^n$, podle reflexivity $(b, b) \in R$ a proto $(a, b) \in R \circ R^n = R^{n+1}$. Proto $R \subseteq R^{n+1}$.

- 3b.17:** 1) Předpoklad $R_1 \subseteq R_2$. Kdyby $m_{1,ij} = 0$, pak $m_{1,ij} \leq m_{2,ij}$, neboť jde o 01-matice. Kdyby $m_{1,ij} = 1$, pak $(a_i, a_j) \in R_1$, proto i $(a_i, a_j) \in R_2$ a $m_{2,ij} = 1$, tedy každopádně $m_{1,ij} \leq m_{2,ij}$.
 2) Předpoklad $M_1 \leq M_2$. Když $(a_i, a_j) \in R_1$, tak $m_{1,ij} = 1$, z $m_{1,ij} \leq m_{2,ij}$ také $m_{2,ij} = 1$ (je to 01-matice), proto $(a_i, a_j) \in R_2$. Dokázáno $R_1 \subseteq R_2$.

3b.18: (i): Předpoklad R je reflexivní. Pro $i \neq j$ je $e_{ij} = 0$, proto $e_{ij} \leq m_{ij}$. Pro $i = j$ je $(a_i, a_i) \in R$, proto $m_{ii} = 1$ a $e_{ij} \leq m_{ij}$.

Předpoklad $E_n \leq M$. Pro každé i je $m_{ii} \geq e_{ii} = 1$, tedy $m_{ii} = 1$ a $(a_i, a_i) \in R$.

(ii): Předpoklad R symetrická. Když $m_{ij} = 1$, pak $(a_i, a_j) \in R$, proto $(a_j, a_i) \in R$ a $m_{ji} = 1$, tedy $m_{ij} = m_{ji}$. Když $m_{ij} = 0$, pak $(a_i, a_j) \notin R$, proto $(a_j, a_i) \notin R$ a $m_{ji} = 0$, tedy každopádně $m_{ij} = m_{ji}$.

Předpoklad M symetrická. Když $(a_i, a_j) \in R$, tak $m_{ij} = 1$, pak $m_{ji} = 1$ a $(a_j, a_i) \in R$.

(iii): Dle Věty 3b.5 je R tranzitivní právě tehdy, když $R^2 \subseteq R$, což je dle (i) právě tehdy, když $M_{R^2} \leq M$, což je dle Věty 3a.10 právě tehdy, když $M^{[2]} \leq M$.

3b.19: S: $(a_i)S(b_i) \implies a_i R_i b_i$ pro všechna i . Protože jsou všechny R_i symetrické, máme $b_i R_i a_i$ pro všechna i , tedy $(b_i)R(a_i)$.

A: $(a_i)R(b_i)$ a $(b_i)R(a_i) \implies \forall i: a_i R_i b_i$ a $b_i R_i a_i$. Protože jsou všechny R_i antisymetrické, máme $a_i = b_i$ pro všechna i , tedy $(a_i) = (b_i)$.

T: $(a_i)R(b_i)$ a $(b_i)R(c_i) \implies a_i R_i b_i$ a $b_i R_i c_i$ pro všechna i , z tranzitivit R_i tedy $a_i R_i c_i$ pro všechna i , proto $(a_i)R(c_i)$.

3b.20: Jak víme, že k danému a existuje b , které je s ním v relaci?

3c. Další vlastnosti relací

V této sekci se blíže podíváme na operace a vlastnosti. Je to kapitola spíše doplňková, nicméně pro čtenáře, kteří se budou relacemi zabývat hlouběji, stále velice užitečná. Většina tvrzení nebude obtížná a není ani třeba se je učit nazepamět, opět je cílem zamyslet se nad fungováním relací, ujistit se, že jim rozumíme, a potrénoval si vytváření důkazů.

Připomeňme definici diagonální relace: $\Delta(A) = \{(a, a) \in A \times A; a \in A\}$. Jakožto podmnožina $A \times A$ splňuje podmínu na zobrazení, takže je to i zobrazení, jmenovitě identita i_A na množině. V teorii relací hraje obdobnou roli jako u zobrazení.

Fakt 3c.1.

Nechť R je relace z množiny A do množiny B . Pak $R \circ \Delta(A) = R$ a $\Delta(B) \circ R = R$.

Platí také, že $[\Delta(A)]^n = \Delta(A)$ pro všechna $n \in \mathbb{N}$, důkazy jsou snadné a necháváme je jako cvičení 3c.1 a 3c.2.

Připomeňme, že je-li $T: A \rightarrow A$ zobrazení, které je invertibilní, tak nám složení $T^{-1} \circ T$ i $T \circ T^{-1}$ dalo vždy identitu na A . Jak to vypadá u relací? Protože inverzní relace existuje vždycky, máme méně omezení a tudíž se dá čekat, že také máme větší rozsah výsledků. A je tomu tak, u relací se o $R^{-1} \circ R$ či $R \circ R^{-1}$ dá říct jedině to, tyto nová relace nejsou prázdné, pokud $R \neq \emptyset$. Jinak je možné cokoliv, rozhodně nemůžeme čekat, že by vyšla diagonální relace.

Příklad 3c.a:

Uvažujme množinu $A = \{1, 3, 5, 7, 11, 13\}$ a definujme relaci R na A předpisem aRb jestliže a dělí b (viz kapitola 6a). Máme

$$R = \{(1, 1), (1, 3), (1, 5), (1, 7), (1, 11), (1, 13)\}, \quad R^{-1} = \{(1, 1), (3, 1), (5, 1), (7, 1), (11, 1), (13, 1)\}.$$

Pak jediné navazující dvojice, které jsme schopni z R a R^{-1} vytvořit, jsou typu $1RaR^{-1}1$, proto $R^{-1} \circ R = \{(1, 1)\}$.

V opačném pořadí je výsledná relace naopak bohatší, protože pro libovolná $a, b \in A$ máme $aR^{-1}1Rb$. Máme proto $R \circ R^{-1} = A \times A$.

Tento příklad ukázal, že není težké získat přímo extrémní výsledky, minimální možný (nejmenší neprázdná množina) a maximální možný (relace s úplně všemi dvojicemi).

△

Pro relace jsme zavedli různé operace (množinové, inverze, skládání). Jak se chovají vůči základním čtyřem vlastnostem? Doporučujeme, aby si čtenář nejprve na rozličných konkrétních případech zkoukal rozmyslet, jak to funguje a proč asi následující tvrzení platí.

Fakt 3c.2.

Nechť R_1 a R_2 jsou relace na množině A .

- (i) Jestliže jsou R_1 a R_2 reflexivní, tak jsou $R_1 \cup R_2$ a $R_1 \cap R_2$ reflexivní a $R_1 - R_2$ nikdy není reflexivní.
- (ii) Jestliže jsou R_1 a R_2 symetrické, tak jsou $R_1 \cup R_2$, $R_1 \cap R_2$ a $R_1 - R_2$ symetrické.
- (iii) Jestliže jsou R_1 a R_2 antisymetrické, tak jsou $R_1 \cap R_2$ a $R_1 - R_2$ antisymetrické.
- (iv) Jestliže jsou R_1 a R_2 tranzitivní, tak je $R_1 \cap R_2$ tranzitivní.

Důkaz necháme zčásti jako cvičení 3c.3, ukážeme jen pár zajímavějších momentů. Protože zde používáme množinové operace, bude rozumnější pracovat s relacemi jako s množinami (což je konec konců jejich definice), tedy psát správné $(a, b) \in R$ namísto pohodlné zkratky aRb .

Důkaz (rutinní): (i): $R_1 - R_2$ není reflexivní: Nechť $a \in A$. Protože je R_1 reflexivní, je $(a, a) \in R_1$. Podobně ale také $(a, a) \in R_2$, proto (a, a) neleží v $R_1 - R_2$.

(ii): $R_1 \cap R_2$ symetrická: Nechť $a, b \in A$ splňují $(a, b) \in R_1 \cap R_2$. To znamená, že $(a, b) \in R_1$ a $(a, b) \in R_2$. Obě relace jsou symetrické, proto $(b, a) \in R_1$ a $(b, a) \in R_2$, tedy $(b, a) \in R_1 \cap R_2$.

$R_1 - R_2$ symetrická: Nechť $a, b \in A$ splňují $(a, b) \in R_1 - R_2$. Pak $(a, b) \in R_1$ a ta je symetrická, proto i $(b, a) \in R_1$. Potřebujeme ukázat, že je také v $R_1 - R_2$. Mohlo by se stát, že $(b, a) \in R_2$? Nemohlo. Kdyby totiž $(b, a) \in R_2$, tak z její symetrie je i $(a, b) \in R_2$, což je ve sporu s $(a, b) \in R_1 - R_2$. Takže $(b, a) \notin R_2$ a máme $(b, a) \in R_1 - R_2$.

(iii) $R_1 - R_2$ antisymetrická: Nechť $(a, b) \in R_1 - R_2$ a $(b, a) \in R_1 - R_2$. Pak $(a, b) \in R_1$ a $(b, a) \in R_1$, ta je antisymetrická a proto $a = b$. □

Proč chybí v (iii) sjednocení? Zkusíme si takové tvrzení dokázat, ať vidíme, kde je zádrhel. Nechť $a, b \in A$ jsou takové, že $(a, b) \in R_1 \cup R_2$ a $(b, a) \in R_1 \cup R_2$. To znamená, že $(a, b) \in R_1$ nebo $(a, b) \in R_2$, také $(b, a) \in R_1$ nebo $(b, a) \in R_2$. Obě relace jsou antisymetrické, jenže my neumíme zajistit, aby (a, b) i (b, a) byly obě v jedné relaci, kde bychom pak mohli tu antisimetrii použít. Klidně mohly každá přijít z jiné relace, takže to vypadá, že antisimetrii dokázat neumíme. Zkusíme na tomto problému založit protipříklad, stačí k tomu $A = \{1, 2\}$.

$R_1 = \{(1, 2)\}$ je antisymetrická, $R_2 = \{(2, 1)\}$ také, ale $R_1 \cup R_2 = \{(1, 2), (2, 1)\}$ už není. Takže je dobře, že jsme v tvrzení (iii) vynechali sjednocení. Neplatí ale, že se to vždy zaručeně pokazí, protože se může stát, že se ty dvě relace dobře sejdou a sjednocení antisymetrické bude (konec konců, stačí vzít $R_1 = R_2$).

Podobně si rozmyslete, kde se zarazí důkaz tranzitivity pro sjednocení a rozdíl (cvičení 3c.4), a ověřte, že protipříkladem pro sjednocení jsou třeba relace $R_1 = \{(1, 2)\}$ a $R_2 = \{(2, 3)\}$ na $A = \{1, 2, 3\}$, protipříkladem pro rozdíl třeba $R_1 = \{(1, 2), (2, 3), (1, 3)\}$ a $R_2 = \{(1, 3)\}$.

Fakt 3c.3.

Nechť R je relace na množině A . Jestliže má některou ze čtyř základních vlastností, tak ji má R^{-1} také.

Důkaz (rutinní): Nechť je R tranzitivní. Vezměme $a, b, c \in A$ takové, že $(a, b) \in R^{-1}$ a $(b, c) \in R^{-1}$. Pak ale $(c, b) \in R$ a $(b, a) \in R$, tudíž podle tranzitivity R platí $(c, a) \in R$ čili $(a, c) \in R^{-1}$. Takže R^{-1} je tranzitivní.

Ostatní vlastnosti jsou snad ještě lehčí a necháme to jako cvičení 3c.5. □

Zbývá poslední operace.

Fakt 3c.4.

Nechť R a S jsou relace na množině A .

Jestliže jsou R a S reflexivní, pak je i $S \circ R$ reflexivní.

Důkaz (rutinní): Nechť $a \in A$. Protože jsou R i S reflexivní, dostáváme aRa a aSa , takže podle definice $(a, a) \in S \circ R$. □

Určitě jste si všimli, že ve Faktu chybí symetrie, antisimetrie a tranzitivita. Kde je problém? Zkusme začít důkaz pro symetrii. Nechť $a, b \in A$ splňují $(a, b) \in S \circ R$. Pak existuje $x \in A$ takové, že aRx a xSb . Obě relace jsou symetrické, máme tedy bSx a xRa , tedy $(b, a) \in R \circ S$. Skoro, ale vedle, máme opačné pořadí v tom skládání; potřebovali bychom $(b, a) \in S \circ R$. Zkusme na tomto průšvihu založit protipříklad.

Relace $R = \{(1, 2), (2, 1)\}$ a $S = \{(2, 3), (3, 2)\}$ jsou symetrické, ale $S \circ R = \{(1, 3)\}$ není. Takže se symetrie při skládání opravdu obecně nezachovává.

Mimochodem, pokud by bylo $R = S$, tak už náš problém odpadá, takže pro mocninu by náš pokus o důkaz prošel, viz níže.

Teď ta antisymetrie. Zkusme si vzít $a, b \in A$ takové, že $(a, b) \in S \circ R$ a $(b, a) \in S \circ R$. Z toho prvního podle definice skládání dostaneme $\exists x \in A: aRx$ a xSc . Z toho druhého zase dostaneme $\exists y \in A: bRy$ a ySa . Máme tedy aRx a bRy , ale nemůžeme čekat, že by platilo $a = b$ a $x = y$, abychom mohli použít antisymetrii relace R . Je to tedy podezřelé a další bádání ukáže, že oprávněně. Uvažujme relaci $R = \{(1, 2), (2, 4), (4, 3), (3, 1)\}$. Tato relace je antisymetrická. Snadno se ale nahlédne, že $R \circ R = R^2 = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$, což není ani náhodou relace antisymetrická (shodou okolností je ale symetrická).

Mimochodem jsme tím ukázali, že antisymetrie se nezachová ani umocňováním.

Pro tranzitivitu si důkaz zkuste sami, abyste viděli, kde se zadrhne. Tradiční protipříklad, abyste věřili, že to nefunguje: $R = \{(1, 2), (3, 4)\}$ a $S = \{(2, 3), (4, 5)\}$ jsou relace tranzitivní (zádné navazující dvojice ani jedna z nich nemá, tudíž tranzitivita platí triviálně), složením ale dostaneme $Z = S \circ R = \{(1, 3), (3, 5)\}$, kde lze vytvořit řetízek $1Z3Z5$, který na jeden krok nezvládneme, dvojice $(1, 5)$ v té složení není.

Skládání se tedy obecně k vlastnostem moc pěkně nechová, ale naštěstí často skládáme jednu relaci se sebou a pak je to veseléji.

Věta 3c.5.

Nechť R je relace na množině A .

- (i) Jestliže je R reflexivní, pak je R^n reflexivní pro všechna $n \in \mathbb{N}$.
- (ii) Jestliže je R symetrická, pak je R^n symetrická pro všechna $n \in \mathbb{N}$.
- (iii) Jestliže je R tranzitivní, pak je R^n tranzitivní pro všechna $n \in \mathbb{N}$.

Důkaz (drsný, poučný): (i): Plyne okamžitě indukcí pomocí předchozího Faktu, viz Cvičení 3c.6.

Zbývající důkazy budou silně založeny na Lemmatu 3a.5 o cestách.

(ii): Nechť $a, b \in A$ splňují $(a, b) \in R^n$. Pak existuje trasa délky n z a do b , řekněme $aRc_2Rc_3R \cdots Rc_nRb$. Protože je R symetrická, lze všechny dvojice c_iRc_{i+1} obrátit na $c_{i+1}Rc_i$ a dostaneme trasu $bRc_nRc_{n-1}R \cdots Rc_2Ra$ délky n z b do a , proto $(b, a) \in R^n$.

(iii): Nechť $a, b, c \in A$ splňují $(a, b) \in R^n$ a $(b, c) \in R^n$. Pak existuje trasa délky n z a do b , řekněme $aR\hat{c}_2R \cdots R\hat{c}_nRb$, a trasa délky n z b do c , řekněme $bR\tilde{c}_2R \cdots R\tilde{c}_nRc$. Protože $\hat{c}_{n+1} = b = \tilde{c}_1$, lze tyto trasy navázat a dostaneme trasu délky $2n$, $aR\hat{c}_2R \cdots R\hat{c}_nR\tilde{c}_1R\tilde{c}_2R \cdots R\tilde{c}_nRc$.

Tato trasa obsahuje celkem $2n$ relací, proto ji lze rozdělit na dvojice $\hat{c}_1R\hat{c}_2R\hat{c}_3, \hat{c}_3R\hat{c}_4R\hat{c}_5, \dots, \hat{c}_{n-1}R\hat{c}_nRc$. Na každou z nich aplikujeme tranzitivitu a dostaneme dvojice $\hat{c}_1R\hat{c}_3, \hat{c}_3R\hat{c}_5, \dots, \hat{c}_{n-1}Rc$. Ty tvoří trasu z a do c , proto $(a, c) \in R^n$ a je to hotovo.

Poznámka: Doporučujeme čtenáři, aby si rozmyslel, jak to zkracování probíhá v místě, kde se napojují původní dvě trasy, je třeba rozlišit případy n sudé a n liché. □

Připomeňme, že antisymetrie se mocninou zachovat nemusí.

3c.6 Uzávěry relace

Tato sekce je spíš doplňková, ale na druhou stranu občas docela zábavná.

Představte si, že máme relaci, která nemá určitou vlastnost, která nás zajímá. Naskýtá se nápad tu relaci doplnit přidáním některých dvojic tak, aby už tuto vlastnost měla. Například pokud relace není reflexivní, tak jí chybí některá z dvojic (a, a) , dodáním to napravíme. Protože ale neradi plýtváme, chceme to udělat přidáním co nejmenšího počtu prvků do relace. Zformulujeme to přesně.

Definice.

Uvažujme nějakou relaci R na nějaké množině A . Je-li P některá z vlastností relace, pak definujeme **P -uzávěr** relace R jako relaci S , která má vlastnost P , obsahuje R (tj. $R \subseteq S$) a S je nejmenší taková relace (přesně, je-li T relace s vlastností P splňující $R \subseteq T$, pak nutně $S \subseteq T$).

U některých vlastností je vcelku jasné, co udělat, u tranzitivity to dá trochu práce.

Věta 3c.7.

Nechť R je relace na množině A .

- (i) Její **reflexivní uzávěr (reflexive closure)** je dán jako $R \cup \Delta(A)$.
- (ii) Její **symetrický uzávěr (symmetric closure)** je dán jako $R \cup R^{-1}$.
- (iii) Její **tranzitivní uzávěr (transitive closure)** je dán jako $\bigcup_{n=1}^{\infty} R^n$.

Důkaz takového tvrzení se skládá ze dvou kroků, jednak musíme dokázat, že uvedená relace má opravdu požadovanou vlastnost, a pak musíme také dokázat, že to nejde s menší relací, tedy že jakmile máme nějakou „nadrelaci“ dané R s vlastností P , tak už je ta naše v ní také. Tyto důkazy provedeme, ale nebude z nich patrné, jak jsme vlastně k těm relacím došli, což je nepříjemné po toho, kdo se chce naučit takové věci vymýšlet. Proto se na problém nejprve podíváme neformálně, a to u tranzitivity, protože ty první dvě vlastnosti jsou relativně snadné.

Abychom z dané relace R vytvořili relaci tranzitivní, tak tam určitě musíme přidat všechny dvojice (a, c) takové, že existuje dvojkrok $aRbRc$ pro nějaké $b \in A$, to jsou ale přesně dvojice z R^2 . Přidáváme tedy k R relaci R^2 . Označme $S = R \cup R^2$. Může to být hledaný uzávěr?

Jedna podmínka splněna je, $R \subseteq S$. Platí i následující: Jestliže je T nějaká relace, která je tranzitivní a $R \subseteq T$, pak podle našich úvah musí být i $R^2 \subseteq T$, tedy $S \subseteq T$. To znamená, že S splňuje podmínu minimality z definice. Je ale tranzitivní?

Obecně to bohužel neplatí. Potřebujeme zkontolovat všechny dvojkroky $aSbSc$ neboli dvojice $(a, b) \in S$ a $(b, c) \in S$. Pro dvojice z S máme dva zdroje. Pokud by obě byly z R , tak už jsme je dříve uvažovali a máme zajištěno, že $(a, c) \in S$. Problém ale je, když přichází alespoň jedna z R^2 . Taková dvojice vůbec nemusela být v R , tudiž nelze obecně zaručit, že jsme do S doplnili i výsledek dvojkroku.

Například u relace $R = \{(1, 2), (2, 3), (3, 4)\}$ chybí zkratky $(1, 3)$ a $(2, 4)$, ale když je doplníme, vznikne relace, ve které lze uvažovat dvojkrok $(1, 3)$ a $(3, 4)$, jehož zkratka $(1, 4)$ v doplněné relaci zase chybí.

V dalším kole bychom se tedy měli podívat na tyto nově vzniklé páry a doplnit příslušné zkratky. Navazující původní a nový pár se najde jako prvek R^3 , dva navazující nové páry se najdou jako prvky $R^2 \circ R^2 = R^4$, takže i tyto musíme přidat. Tím ale mohly vzniknout nové páry atd. až do zblbnutí. Výsledný vzorec je nasnadě. Pokud vám něco připomíná, podívejte se na connectivity relation po Lemmatu 3a.5.

Důkaz (poučný, drsný): (ii): Nejprve je třeba ukázat, že $R \cup R^{-1}$ je symetrická relace, to necháme jako cvičení 3c.8. Teď dokážeme tu minimality.

Nechť T je nějaká symetrická relace, která obsahuje R . Potřebujeme dokázat, že $R \cup R^{-1} \subseteq T$. Inkluzi $R \subseteq T$ už máme od začátku, zbývá dokázat, že $R^{-1} \subseteq T$.

Nechť $(a, b) \in R^{-1}$. Pak $(b, a) \in R$, proto i $(b, a) \in T$. Ale T je symetrická, proto i $(a, b) \in T$. Důkaz hotov.

(iii): Označme $R^* = \bigcup_{n=1}^{\infty} R^n$. Už podle definice je jasné, že $R \subseteq R^*$.

Je R^* tranzitivní? Předpokládejme, že $(a, b) \in R^*$ a $(b, c) \in R^*$. Pak existují $m, n \in \mathbb{N}$ takové, že $(a, b) \in R^m$ a $(b, c) \in R^n$. Podle Lemmatu 3a.5 o cestách proto existuje trasa délky m z a do b , nazvěme její prvky $\tilde{c}_1, \dots, \tilde{c}_{m+1}$, a existuje trasa délky n z b do c , nazvěme její prvky $\hat{c}_1, \dots, \hat{c}_{n+1}$. Všimněte si, že $\tilde{c}_{m+1} = b = \hat{c}_1$. Teď tyto trasy napojíme: Definujeme $c_i = \tilde{c}_i$ pro $i = 1, \dots, m+1$ a $c_i = \hat{c}_{i-m}$ pro $i = m+2, \dots, m+n+1$. Je snadné ověřit, že pak c_1, \dots, c_{m+n+1} tvoří trasu délky $m+n$ z a do c a proto $(a, c) \in R^{m+n} \subseteq R^*$. R^* tedy je tranzitivní.

Je R^* minimální relace s touto vlastností? Nechť T je relace taková, že $R \subseteq T$ a T je tranzitivní. Dokážeme indukcí, že $R^n \subseteq T$ pro všechna $n \in \mathbb{N}$.

(0) $n = 1$: Evidentně $R \subseteq T$, je to jedna z vlastností T .

(1) Nechť $n \in \mathbb{N}$ je libovolné, předpokládejme, že $R^n \subseteq T$. Chceme dokázat, že $R^{n+1} \subseteq T$.

Nechť (a, b) je libovolný prvek z $R^{n+1} = R \circ R^n$. Pak existuje $x \in A$ takové, že $(a, x) \in R^n$ a $(x, b) \in R$. Podle indukčního předpokladu pak $(a, x) \in T$ a podle (0) také $(x, b) \in T$. A protože je T tranzitivní, nutně $(a, b) \in T$. Tím je důkaz hotov. \square

Ne všechny vlastnosti se dají získat přes uzávěr, například pro vytvoření antisimetrické relace z nějaké R bychom spíš potřebovali z R prvky odebírat než přidávat. Antisimetrický uzávěr bychom tedy sice definovat mohli, ale pak bychom museli konstatovat, že obecně neexistuje.

Další fakta o uzávěrech najdete ve cvičeních od 3c.8. Jednoduché příklady jsou ve cvičení 3c.7, pro jeden užitečný uzávěr se můžete podívat na cvičení 4a.17.

3c.8 Další vlastnosti

Zkoumané vlastnosti vycházejí z praktického použití relací, kdy člověk zjistí, že by se mu velice hodilo, kdyby se relace chovaly určitým způsobem. Pokud stejnou věc potká vícekrát, vyplatí se jí dát jméno. Tak jsme přišli

k těm čtyřem základním vlastnostem, které se zkoumají nejčastěji. Nejsou ale jediné, při různých aplikacích se vyskytnou i další. Uvedeme si pár populárnějších.

Definice.

Nechť R je relace na množině A .

Řekneme, že R je **antireflexivní** či **ireflexivní** (**irreflexive**), jestliže pro všechna $a \in A$ platí $(a, a) \notin R$.

Řekneme, že R je **asymetrická** (**asymmetric**), jestliže pro všechna $a, b \in A$ platí $(aRb \implies \neg(bRa))$.

Řekneme, že R je **dichotomická** (**dichotomic**), jestliže pro všechna $a, b \in A$ platí $(aRb \vee bRa)$.

Řekneme, že R je **trichotomická** (**trichotomic**), jestliže pro všechna $a, b \in A$ platí právě jedna z možností $aRb, bRa, a = b$.

Antireflexivní relace nejsou opakem reflexivních, jsou opakem zahaněným do extrému. Na to, aby relace nebyla reflexivní, stačí jediná chybějící dvojice (a, a) . Antireflexivita to bere od podlahy a rovnou zakáže všechny takovéto dvojice. Typickým příkladem je relace $<$ na číslech. Náhodně vytvořená relace bude mít s vysokou pravděpodobností jen některé z dvojic (a, a) , takže nebude ani reflexivní, ani antireflexivní.

Pro nás nejzajímavější bude asymetrie (které někteří autoři říkají silná antisimetrie). Protože zakazuje oboustranné šipky, tak již implikuje antisimetrii (takže asymetrie je silnější vlastnost než antisimetrie). Dále se všimněte, že pokud by existovalo $(a, a) \in R$, tak to už poruší podmínu asymetrie. Tato vlastnost tedy zakazuje smyčky. Dostáváme tak jeden směr v následujícím faktu.

Fakt 3c.9.

Nechť R je relace na množině A . Je asymetrická právě tehdy, pokud je antisymetrická a antireflexivní.

Druhý směr vyplýne stejně snadno a zkuste si jej rozmyslet. Pokud vám všechny tyto úvahy přijdou snadné (a ony jsou), tak to je dobré znamení, že relacím rozumíte.

Mezi antisimetrii a asymetrií je zajímavá souvislost, kterou vyjádříme v tvrzení, které se nám bude trošku hodit později. Zkuste si při jeho čtení představovat nerovnosti \leq a $<$.

Fakt 3c.10.

Uvažujme množinu A .

(i) Nechť R je antisymetrická relace na A . Definujme relaci S tímto předpisem: aSb jestliže aRb ale $a \neq b$. Pak je S asymetrická.

(ii) Nechť S je asymetrická relace na A . Definujme relaci R tímto předpisem: aRb jestliže aSb nebo $a = b$. Pak je R antisymetrická a reflexivní.

Formálně řečeno, v (i) je $S = R - \Delta(A)$, ve (ii) je $R = S \cup \Delta(A)$. Pokud to vidíte, skvělé. Pro další detaily viz kapitola 4b.

Někdy potřebujme relaci přinutit, aby vytvářela hodně spojení. Není úplně jasné, jak to udělat nejlépe, poslední dvě vlastnosti z definice výše ukazují dva možné přístupy.

Dichotomie je vlastnost, která zaručí vzájemnou propojenosť všech prvků množiny, ale neřeší, jakým způsobem, hlavně aby tam něco bylo. Všimněte si, že lze zvolit i $a = b$, takže dichotomie už vynutí reflexivitu. Dobrým příkladem je relace \leq na číslech.

I trichotomie vynucuje propojení prvků, ale zároveň je zvědavá, jak se to dělá. Všimněte si, že dvojice $a, b = a$ již splňuje $a = b$, proto z té trojice nesmí platit nic dalšího neboli neplatí aRa ; trichotomické relace jsou tedy antireflexivní. Také zakážeme oboustranné šipky mezi různými prvky, jde tedy o relaci asymetrickou, proto i antisymetrickou. Typickým příkladem je relace $<$ na číslech.

Pro příklad relace, která není dichotomická ani trichotomická, stačí vzít relaci R na \mathbb{Z}^2 definovanou $(u, v)R(x, y)$ jestliže $u < x \wedge v < y$. Pak dvojice $a = (1, 2)$, $b = (2, 1)$ není porovnatelná.

Některí autoři mají jinou definici trichotomie, používají v definici obyčejnou disjunkci, takže připouštějí, že z těch tří možností platí i více najednou. Snadno se rozmyslí, že takováto alternativní definice je ekvivalentní podmínce, aby pro všechna $a \neq b$ platilo $aRb \vee bRa$. Jinak řečeno, tato definice se neplete do (anti)reflexivity, což mnoho autorů shledává sympatickým. Je také vidět, že každá dichotomická relace by byla i trichotomická dle této jiné definice. Některé podrobnosti pak následně fungují jinak, ale protože zde nebudeme s trichotomií pracovat, nemusí nás to trápit.

K problematice vynucování spojení se vrátíme v kapitole 4c.

Existuje samozřejmě mnohem více roztodivných vlastností, ale to už jsou specializované věci. Uvedeme jednu na ukázku: Relace R se nazývá hustá, jestliže pro libovolné x, y splňující xRy existuje nějaké z takové, že $xRzRy$. Třeba relace $<$ je hustá na \mathbb{Q} či \mathbb{R} , ale na \mathbb{Z} už hustá není, protože $13 < 14$ a mezi nimi už další prvek není. Pro další zajímavou vlastnost se podívejte na cvičení 4a.18.

Cvičení

Cvičení 3c.1 (poučné, zkouškové): Nechť R je relace z množiny A do množiny B . Dokažte, že pak $R \circ \Delta(A) = R$ a $\Delta(B) \circ R = R$.

Cvičení 3c.2 (poučné, zkouškové): Dokažte, že pro každou množinu A a $n \in \mathbb{N}$ platí $|\Delta(A)|^n = \Delta(A)$.

Cvičení 3c.3 (rutinní, poučné, zkouškové): Nechť R_1, R_2 jsou relace na A . Dokažte následující (viz Fakt 3c.2):

- (i) Jestliže jsou R_1, R_2 reflexivní, pak je i $R_1 \cap R_2$ reflexivní.
 - (ii) Jestliže jsou R_1, R_2 reflexivní, pak je i $R_1 \cup R_2$ reflexivní.
 - (iii) Jestliže jsou R_1, R_2 symetrické, pak je i $R_1 \cup R_2$ symetrická.
 - (iv) Jestliže jsou R_1, R_2 antisymetrické, pak je i $R_1 \cap R_2$ antisymetrická.
 - (v) Jestliže jsou R_1, R_2 tranzitivní, pak je i $R_1 \cap R_2$ tranzitivní.

Cvičení 3c.4 (poučné): Nechť R_1, R_2 jsou relace na A .

- (i) Kde se zadrhne důkaz tvrzení „Jestliže jsou R_1, R_2 tranzitivní, pak je i $R_1 \cup R_2$ tranzitivní“?
(ii) Kde se zadrhne důkaz tvrzení „Jestliže jsou R_1, R_2 tranzitivní, pak je i $R_1 - R_2$ tranzitivní“?

Cvičení 3c.5 (rutinní, zkouškové): Nechť R je relace na A . Dokažte následující (viz Fakt 3c.3):

- (i) R je reflexivní právě tehdy, když je i R^{-1} reflexivní.
 - (ii) R je symetrická právě tehdy, když je i R^{-1} symetrická.
 - (iii) R je antisymetrická právě tehdy, když je i R^{-1} antisymetrická.

Cvičení 3c.6 (poučné): Nechť R je relace na A . Dokažte, že jestliže je reflexivní, tak také R^n je reflexivní pro všechna $n \in \mathbb{N}$.

Cvičení 3c.7 (rutinní): Uvažujme následující relace na množině $A = \{1, 2, 3, 4\}$:

- $$(i) \quad R_1 = \{(1, 1), (1, 4), (2, 1), (3, 4)\}; \quad (ii) \quad R_2 = \{(1, 4), (1, 3), (4, 3), (2, 2)\}$$

a) Doplňte co nejúsporněji relaci R_1 tak, aby byla tranzitivní (jinými slovy, přidejte k ní nějaké další dvojice tak, aby výsledná relace byla tranzitivní, a přitom se přidal nejmenší možný počet dvojic, kterým lze tranzitivitu dosáhnout, viz 3c.6) uzávěr relace.

b) Doplňte co nejúsporněji relaci R_2 tak, aby byla symetrická a tranzitivní.

Cvičení 3c.8 (poučné, zkouškové): Nechť R je nějaká relace na množině A . Dokažte, že $R \cup R^{-1}$ je symetrická relace.

Cvičení 3c.9 (poučné, dobré): Dokažte, že symetrický uzávěr reflexivního uzávěru relace R je roven reflexivnímu uzávěru symetrického uzávěru R .

Cvičení 3c.10 (poučné, dobré): Dokažte, že tranzitivní uzávěr symetrického uzávěru relace R obsahuje symetrický uzávěr tranzitivního uzávěru R , ale nemusí se rovnat.

Cvičení 3c.11 (poučné, dobré): Nechť R, S jsou relace na stejně množině A . Předpokládejme, že pro obě existují uzávěry vzhledem k jisté vlastnosti V , označme je \widehat{R} a \widehat{S} . Dokažte, že jestliže $R \subseteq S$, pak $\widehat{R} \subseteq \widehat{S}$.

Cvičení 3c.12 (poučné, dobré): Nechť R je relace na množině A . Pokud je R tranzitivní, tak už je asymetrická (a proto i antisymetrická).

Nápověda: Důkaz sporem ide docela někně.

Řešení:

3c.1: Nechť $a \in A, b \in B$.

- 1) $R \circ \Delta(A) \subseteq R$: $(a, b) \in R \circ \Delta(A) \implies \exists x \in A: [(a, x) \in \Delta(A) \wedge (x, b) \in R]$. Ale $(a, x) \in \Delta(A)$ znamená $a = x$, proto $(a, b) \in R$.

2) $B \subseteq R \circ \Delta(A)$: Nechť $(a, b) \in B$. Protože $(a, a) \in \Delta(A)$, máme $a\Delta(A)aBb$ a tedy $(a, b) \in R \circ \Delta(A)$.

3) Dôkaz $\Delta(B) \circ B \equiv B$ je obdobný.

3c. 2: Díjkaz indukci:

- (0) $n = 1$: $[\Delta(A)]^1 = \Delta(A)$ platí.
 (1) Předpoklad: $[\Delta(A)]^n = \Delta(A)$. Pak $[\Delta(A)]^{n+1} = [\Delta(A)]^n \circ \Delta(A) = \Delta(A) \circ \Delta(A) = \Delta(A)$ podle cvičení 3c.1.

3c.3: (i): Nechť $a \in A$. R_1 reflex tedy $(a, a) \in R_1$, R_2 reflex tedy $(a, a) \in R_2$, proto $(a, a) \in R_1 \cap R_2$.

(ii): Nechť $a \in A$. R_1 reflex tedy $(a, a) \in R_1 \subseteq R_1 \cup R_2$.

(iii): Nechť $(a, b) \in R_1 \cup R_2$. Pak $(a, b) \in R_1 \vee (a, b) \in R_2$. Kdyby $(a, b) \in R_1$, pak ze symetrie R_1 bude $(b, a) \in R_1 \subseteq R_1 \cup R_2$. Kdyby $(a, b) \in R_2$, pak ze symetrie R_2 bude $(b, a) \in R_2 \subseteq R_1 \cup R_2$. Každopádně $(b, a) \in R_1 \cup R_2$.

(iv): Nechť $(a, b) \in R_1 \cap R_2$ a $(b, a) \in R_1 \cap R_2$. Pak $(a, b) \in R_1 \wedge (b, a) \in R_1$, z antisymetrie R_1 tedy $a = b$.

Poznámka: Důkaz ukazuje, že stačí, aby jedna z R_1, R_2 byla antisymetrická, a už je takový i průnik.

(v): Nechť $(a, b) \in R_1 \cap R_2$ a $(b, c) \in R_1 \cap R_2$. Pak $(a, b) \in R_1 \wedge (b, c) \in R_1$, z tranzitivnosti R_1 tedy $(a, c) \in R_1$. Podobně $(a, b) \in R_2 \wedge (b, c) \in R_2$, z tranzitivnosti R_2 tedy $(a, c) \in R_2$. Proto $(a, c) \in R_1 \cap R_2$.

3c.4: (i): Nechť $a, b, c \in A$ splňují $(a, b) \in R_1 \cup R_2 \wedge (b, c) \in R_1 \cup R_2$. Pak $(a, b) \in R_1$ nebo $(a, b) \in R_2$, a $(b, c) \in R_1$ nebo $(b, c) \in R_2$. Abychom mohli použít tranzitivitu R_1 , museli bychom nějak zajistit, aby $(a, b), (b, c)$ byly oba v R_1 nebo oba v R_2 . To ale nedokážeme, nemáme nic, čím bychom je mohli přinutit vybrat si zrovna jednu z relací. Tím je také inspirován protipříklad.

(ii): Nechť $a, b, c \in A$ splňují $(a, b) \in R_1 - R_2 \wedge (b, c) \in R_1 - R_2$. Pak $(a, b) \in R_1$ a $(b, c) \in R_1$, proto dle tranzitivnosti R_1 i $(a, c) \in R_1$. Ještě potřebujeme $(a, c) \notin R_2$, ale to neumíme zajistit. Víme jen, že $(a, b) \notin R_2$ a $(b, c) \notin R_2$, to ale nikterak nebrání (a, c) v tom, aby v R_2 bylo. Zase je tímto inspirován protipříklad v diskusi po Faktu 3c.2.

3c.5: (i): Nechť R reflexivní. Pro $a \in A$ pak $(a, a) \in R$, proto po prohození prvků $(a, a) \in R^{-1}$, R^{-1} je reflexivní. Nechť naopak R^{-1} reflexivní. Pro $a \in A$ pak $(a, a) \in R^{-1}$, tedy po prohození $(a, a) \in R$.

(ii): Nechť R je symetrická. Jestliže $(a, b) \in R^{-1}$, pak $(b, a) \in R$, ze symetrie $(a, b) \in R$ a tedy $(b, a) \in R^{-1}$. R^{-1} je symetrická.

Nechť R^{-1} je symetrická. Jestliže $(a, b) \in R$, pak $(b, a) \in R^{-1}$, ze symetrie $(a, b) \in R^{-1}$ a tedy $(b, a) \in R$. R je symetrická.

(iii): Nechť R je antisymetrická. Jestliže $(a, b) \in R^{-1}$ a $(b, a) \in R^{-1}$, pak $(b, a) \in R$ a $(a, b) \in R$, z antisymetrie $a = b$. R^{-1} je symetrická.

Nechť R^{-1} je antisymetrická. Jestliže $(a, b) \in R$ a $(b, a) \in R$, pak $(b, a) \in R^{-1}$ a $(a, b) \in R^{-1}$, z antisymetrie $a = b$. R je symetrická.

3c.6: Indukcí. (0): $n = 1$: jestliže je R reflexivní, tak $R^1 = R$ je reflexivní.

(1): Předpoklad R^n reflexivní. Podle Věty 3c.4 je i $R^n \circ R = R^{n+1}$ reflexivní.

3c.7: a) $R_1 \cup \{(2, 4)\}$.

b) Nejprve doplníme na symetrickou relaci: $R_2 \cup \{(4, 1), (3, 1), (3, 4)\}$. Teď zjistíme, co chybí pro platnost tranzitivnosti (pořadí nejprve symetrie, pak tranzitivita je výhodné, protože pokud teď případně další spojnice budeme rovnou dávat v obou směrech, tak již symetrii nepokazíme). Odpověď:

$R_2 \cup \{(4, 1), (3, 1), (3, 4), (1, 1), (3, 3), (4, 4)\}$.

3c.8: Nechť $(a, b) \in R \cup R^{-1}$. Pak $(a, b) \in R \vee (a, b) \in R^{-1}$. Když $(a, b) \in R$, tak $(b, a) \in R^{-1} \subseteq R \cup R^{-1}$. Když $(a, b) \in R^{-1}$, tak $(b, a) \in R \subseteq R \cup R^{-1}$. Proto každopádně $(b, a) \in R \cup R^{-1}$.

3c.9: $(\Delta \cup R)^{-1} = \Delta^{-1} \cup R^{-1} = \Delta \cup R^{-1}$, využilo se cvičení 3a.6.

3c.10: (a, b) je v symetrickém tranzitivního, pak (a, b) v tranzitivním nebo (b, a) v tranzitivním. Proto trasa nějaké délky z a do b v R , ta je pak i v symetrickém uzávěru R , nebo trasa z b do a v R , pak otočením trasa z a do b v R^{-1} , tedy i v symetrickém uzávěru, každopádně trasa z a do b v symetrickém uzávěru a proto (a, b) v tranzitivním uzávěru symetrického.

$R = \{(a, b), (a, c)\}$, tranzitivní symetrického je $\{(a, a), (a, b), (a, c), (b, a), (b, b), (b, c), (c, a), (c, b), (c, c)\}$, symetrický tranzitivního je $\{(a, b), (a, c), (c, a), (c, b)\}$.

3c.11: \widehat{R} je nejmenší relace s vlastností P obsahující R . Ale \widehat{S} má také P a obsahuje R , proto z minimality \widehat{R} máme $\widehat{R} \subseteq \widehat{S}$.

3c.12: Předpoklad: R je antisymetrická, tranzitivní a není asymetrická. Proto $\exists a, b \in A$ tak, že aRb a bRa . Tranzitivita pak dává aRa , což je ve sporu s antireflexivitou.

3d. n -ární relace

Toto je jen rychlý pohled, aby čtenář věděl, že není třeba končit u binárního.

Definice.

Nechť A_1, \dots, A_n jsou množiny. Libovolná podmnožina $R \subseteq A_1 \times \dots \times A_n$ se nazývá **n -ární relace** na těchto množinách.

Číslu n říkáme **stupeň** této relace.

Consider sets A_1, \dots, A_n . By an **n -ary relation** on these sets we mean any subset of $A_1 \times \dots \times A_n$.

The number n is called the **degree** of this relation.

Takovéto relace nám umožňují reprezentovat vztahy mezi více objekty.

Příklad 3d.a: Nechť A je množna všech studentů zapsaných na fakultě, $B = \{\text{Bc, MSc, Ph.D.}\}$, $C = \{\text{EaI, STM, OI, KaR, KME, EEM}\}$ a $D = \{1, 2, 3, 4\}$. Pak lze studentskou populaci popsat relací R arity 4 definovanou jako $(a, b, c, d) \in R$ právě tehdy, jestiže student a studuje na b v programu c a je v ročníku d .

△

Jak příklad jemně naznačuje, tyto relace se používají při práci s databázemi. Volba vhodného modelu silně ovlivní, jak rychle databáze reaguje na možné požadavky (a které je vůbec schopna splnit). Jedním z používaných modelů je právě relační model.

Toto použití pak inspiruje operace, které definujeme. Určitě budeme chtít operaci, která umožní vybírat dle parametru, čímž vznikne podrelace (například z relace studentů chceme podrelaci, která zahrnuje jen studenty programu OI). Je možno dělat tzv. projekce, což v zásadě znamená, že z relace vynecháme některou složku (například můžeme vytvořit relaci, která nezahrnuje obor, jen titul a ročník), relace se dají spojovat a podobně. To už je ale dosti specializované téma pro kurs databází.

4. Speciální relace

Už jsme zmínili jednu skupinu relací, které jsou tak svébytné, že se zkoumají zvlášť, jmenovitě zobrazení. Jsou ještě dvě zajímavé skupiny relací, které si zaslouží vlastní kapitolku: ekvivalence a uspořádání.

4a. Ekvivalence

! Definice.

Nechť R je relace na nějaké množině A . Řekneme, že R je **ekvivalence**, jestliže je reflexivní, symetrická a tranzitivní.

A relation on a set is called an **equivalence relation** if it is reflexive, symmetric and transitive.

Čím je taková relace speciální? Jedna výhoda je vidět hned, dá se snadno znázornit grafem v případě konečné množiny. Protože víme, že je reflexivní, je zbytečné kreslit smyčky, jsou všude automaticky. Protože je symetrická, není třeba kreslit všechny dvojité šipky sem a tam. Vyplývá z toho, že nám k zakreslení stačí neorientovaný graf bez smyček, který díky své jednoduchosti lépe ukáže, které prvky množiny jsou v relaci a které ne.

Zjednodušení situace se odráží i v terminologii: Jestliže aRb , pak říkáme, že a a b jsou ekvivalentní, přičemž v tomto vyjádření už díky symetrii na pořadí oněch prvků nezáleží. Díky reflexivitě víme, že každý prvek je ekvivalentní sám sobě. Někdy se pro ekvivalence dvou prvků používá speciální značení $a \sim b$, ale nejde o univerzální konvenci, tak ji tady nebudeme používat.

Příklad 4a.a: Při pohledu na příklady a cvičení z kapitoly 3b zjistíme, že ekvivalence jsme poznali docela dost: Rovnost čísel $x = y$ na libovolné množině čísel, vlastně libovolná rovnost nějakých objektů, relace na číslech daná vztahem $|x| = |y|$, relace na množinách daná shodou mohutnosti $|A| = |B|$.

V kapitole 3a jsme také hovořili o relaci dané mezi počítací definicí aRb jestliže jsou a a b navzájem schopny komunikovat (třeba přes prostředníka), i to je ekvivalence.

Uvažujme relaci definovanou na zastávkách ve městech definicí aRb jestliže se dá z a do b dojet hromadnou dopravou. Taková relace bude určitě reflexivní (nastoupím a hned vystoupím) a tranzitivní. Je to tedy ekvivalence? Rozhodne to symetrie. Jestliže umím dojet hromadnou z a do b , může se stát, že bych z b do a dojet neuměl? To je dobrá otázka, člověk by to nečekal, ale dějí se u nás i podivnější věci. Takže tady nevíme, zda máme ekvivalence. △

Mnohé z příkladů vycházejí z jedné základní myšlenky. Máme objekty z množiny A , u kterých zjišťujeme jistý pro nás důležitý parametr (velikost, IP číslo sítě, pohlaví, operační systém, výrobce, ...). Obecně se to dá vyjádřit tak, že existuje zobrazení $T: A \mapsto B$, kde množina B říká, jaké hodnoty měřená vlastnost může mít. Pokud se rozhodneme objekty sdružovat podle toho, zda mají dotyčnou vlastnost stejnou, formálně aRb právě tehdy, když $T(a) = T(b)$, tak už vždy vznikne ekvivalence, bez ohledu na volbu T (viz cvičení 4a.7).

V situaci, kdy objekty porovnáváme podle určité vlastnosti, se pak jejich množina rozpadne na skupiny, každá zahrnuje se shodnou vlastností. V případě relace na knihách daná zobrazením kniha → autor tak vznikne skupina knih od autora xy , skupina knih od autora cw atd.

Tento proces rozpadu množiny na skupiny prozkoumáme teoreticky.

! Definice.

Nechť R je relace ekvivalence na nějaké množině A . Pro $a \in A$ definujeme **třídu ekvivalence** prvku a (**equivalence class of a**) vzhledem k R jako

$$[a]_R = \{b \in A; aRb\}.$$

Alternativní značení: $[a]_R$ se píše také $R[a]$ nebo dokonce jen $[a]$, pokud je relace jasná z kontextu.

Díky reflexivitě víme, že taková třída není nikdy prázdná, protože $a \in [a]_R$. Co se od takové třídy dá čekat? Podívejme se na ekvivalence danou mohutností množin. Vezmeme-li si konkrétní množinu A , tak odpovídající třída ekvivalence jsou všechny množiny M splňující $|A| = |M|$, takže se nám vlastně na hromádce ocitnou všechny množiny stejné mohutnosti. Původní množinu A pak lze brát jako jakéhosi typického zástupce množiny této velikosti. Tato ekvivalence nám tedy množiny rozdělí do zřetelně oddělených skupin podle velikosti.

Následující věta nám shrne nejdůležitější vlastnosti tříd ekvivalence.

! Věta 4a.1.

Nechť R je relace ekvivalence na nějaké množině A , nechť $a \in A$.

- (i) Pro každé $b, c \in [a]_R$ platí bRc .
- (ii) Pro každé $b \in [a]_R$ a $c \in A$ platí, že jestliže bRc , pak $c \in [a]_R$.
- (iii) Pro každé $b \in [a]_R$: $[a]_R = [b]_R$.
- (iv) Pro každé $a, b \in A$ platí: aRb právě tehdy, když $[a]_R = [b]_R$.
- (v) Pro všechna $a, b \in A$ platí, že buď $[a]_R = [b]_R$, nebo $[a]_R \cap [b]_R = \emptyset$.

Jaký obrázek nám tu vzniká? Třída ekvivalence je skupina prvků, ve které jsou dle (i) všechny navzájem v relaci, v grafu ji tedy vidíme jako skupinu prvků, která je křížem krážem propojena. Zároveň z ní ale nevede žádná cestička jinam, protože jakákoli spojnice zevnitř skupiny k nějakému prvku jej podle (ii) okamžitě vtáhne do této skupiny.

Množina A se tak rozpadne na samostatné, od sebe izolované skupiny. To je znovu potvrzeno v části (v).

Každou skupinu (třídu) jsme dostali pomocí jakéhosi prvku a , kterého lze považovat za jejího zástupce, ale (iii) ukazuje, že libovolný člen této třídy bude také rovnocenným zástupcem, je úplně jedno, kterého si vybereme. A konečně (iv) nám říká, že se podle příslušnosti ke třídám dá poznat, kdo je s kým v relaci, což je nápad, který použijeme posléze. A teď už důkaz, vyplatí se kreslit si obrázky.

Důkaz (poučný): (i): Jestliže $b, c \in [a]_R$, pak aRb a aRc . Podle symetrie také bRa , dvojkrok $bRaRc$ pak podle tranzitivity dává bRc .

(ii): Z $b \in [a]_R$ máme aRb , spolu s bRc a tranzitivitou dostanem aRc , tedy $c \in [a]_R$.

(iii): Vezměme libovolné $c \in [b]_R$. Pak bRc a podle (ii) tedy $c \in [a]_R$. Dokázali jsme $[b]_R \subseteq [a]_R$.

Nechť naopak $c \in [a]_R$, pak aRc . Z $b \in [a]_R$ máme aRb , ze symetrie bRa . Dostali jsme dvojkrok $bRaRc$, podle tranzitivity bRc a tedy $c \in [b]_R$. Dokázali jsme $[a]_R \subseteq [b]_R$.

(iv): Jde o snadný důsledek (iii). Jestliže aRb , pak $b \in [a]_R$ a proto podle (iii) máme $[a]_R = [b]_R$. Nechť naopak $[a]_R = [b]_R$. Pak $b \in [b]_R = [a]_R$, tedy $b \in [a]_R$ a aRb přímo podle definice $[a]_R$.

(v): Vezměme nějaké $a, b \in A$. Jestliže $[a]_R \cap [b]_R = \emptyset$, pak jsme hotovi. Jinak existuje $c \in [a]_R \cap [b]_R$. Podle (iii) to pak znamená $[a]_R = [c]_R = [b]_R$. □

V definici teď zachytíme obecně myšlenku rozdělení množiny na disjunktní části.

! Definice.

Uvažujme množinu A . Jejím **rozkladem** rozumíme libovolný soubor \mathcal{S} neprázdných podmnožin množiny A takový, že $A = \bigcup_{M \in \mathcal{S}} M$ a pro všechna $M \neq N \in \mathcal{S}$ jsou M, N disjunktní.

Consider a set A . By its **partition** we mean an arbitrary collection \mathcal{S} of non-empty subsets of A such that $A = \bigcup_{M \in \mathcal{S}} M$ and M, N are disjoint for all $M \neq N \in \mathcal{S}$.

Podmínka o disjunktnosti se často vyjadřuje pomocí obměny: pro všechna $M, N \in \mathcal{S}$ má platit: $M \cap N \neq \emptyset \implies M = N$. Tato forma je někdy výhodnější pro použití například v důkazech.

Poznamenejme, že na velikost \mathcal{S} se v této definici nekladou žádné nároky, klidně může být nekonečná, dokonce může jít i o nespočetnou kolekci podmnožin (a pak je tam nespočetné sjednocení, ještě to uvidíme).

! Věta 4a.2.

Nechť A je množina.

(i) Jestliže je R ekvivalence na A , pak $\{[a]_R\}_{a \in A}$ je rozklad množiny A .

(ii) Jestliže je \mathcal{S} nějaký rozklad množiny A , pak existuje relace ekvivalence R na A taková, že \mathcal{S} jsou přesně třídy ekvivalence vzhledem k R .

Důkaz (z povinnosti, ale poučná myšlenka): (i): Protože $a \in [a]_R$, je určitě $A = \bigcup_{a \in A} [a]_R$. Druhá podmínka vyplývá z Věty 4a.1 (v).

(ii): Definujme relaci R na A takto: aRb právě tehdy, když existuje $M \in \mathcal{S}$ takové, že $a, b \in M$.

Jinými slovy, prohlásíme, že všechny prvky z jedné množiny rozkladu jsou navzájem ekvivalentní, ale žádné jiné dvojice už nevytvoříme.

1) Nejprve ukážeme, že R je ekvivalence.

R: Jestliže $a \in A$, pak z $A = \bigcup_{M \in \mathcal{S}} M$ musí existovat $M \in \mathcal{S}$ takové, že $a \in M$. Pak $a \in M$ a také $a \in M$, tedy aRa . Relace je reflexivní.

S: Jestliže aRb , pak $a, b \in M$ pro nějakou množinu $M \in \mathcal{S}$. Pak ovšem také $b, a \in M$ a bRa . R je symetrická.

T: Předpokládejme, že aRb a bRc . Pak podle definice R existuje množina $M \in \mathcal{S}$ taková, že $a, b \in M$, a existuje množina $N \in \mathcal{S}$ (zatím musíme připustit možnost, že jiná) taková, že $b, c \in N$. Teď použijeme předpoklad, že \mathcal{S} je rozklad, abychom ukázali, že $M = N$. Našli jsme prvek $b \in M \cap N$, tedy $M \cap N \neq \emptyset$. Již jsme diskutovali, že pro rozklad z toho plyne $M = N$. Takže $a, b, c \in M$, proto i $a, c \in M$ a aRc . R je tedy tranzitivní.

2) Teď ukážeme, že množiny z rozkladu odpovídají třídám ekvivalence R .

Vezměme třídu rozkladu $[a]_R$. Pak existuje $M \in \mathcal{S}$ taková, že $a \in M$. Pro všechny $b \in M$ pak podle definice aRb , tedy $b \in [a]_R$. Dokázali jsme, že $M \subseteq [a]_R$. Naopak jestliže $b \in [a]_R$, pak aRb , proto $a, b \in N$ pro nějakou množinu $N \in \mathcal{S}$. Prvek a leží v M i v N , proto (viz výše) $M = N$, tudíž $b \in M$. Ukázali jsme, že $[a]_R \subseteq M$. Každá třída ekvivalence je tedy rovna nějaké množině z rozkladu.

Formálně, $\{[a]_R\}_{a \in M} \subseteq \mathcal{S}$.

Ještě musíme ukázat, že každá množina z rozkladu odpovídá nějaké třídě ekvivalence. Vezměme tedy $M \in \mathcal{S}$ a libovolný prvek $a \in M$ (jsou to neprázdné množiny dle definice rozkladu). Pak podle předchozího odstavce $M = [a]_R$ a je to. \square

Z praktického pohledu tedy relace ekvivalence představuje jeden z možných pohledů na situaci, kdy množinu rozparcelujeme na kousky.

Příklad 4a.b: Asi nejnámější ekvivalence je relace rovnosti. Podívejme se na ni například na \mathbb{R} . V příkladě 3b.d jsme už ukázali reflexivitu, symetrii a tranzitivitu. Jak vypadají třídy ekvivalence? Pro libovolné $x \in \mathbb{R}$ platí $[x]_R = \{x\}$, takže to je asi ta nejnudnější ekvivalence, s nejmenšími možnými třídami ekvivalence. Vzniká tím rozklad $\mathbb{R} = \bigcup_{x \in \mathbb{R}} \{x\}$, což jen potvrzuje, že tato ekvivalence je nuda.

Mírně zajímavější je relace definovaná na \mathbb{R} předpisem xRy právě tehdy, když $|x| = |y|$. V této ekvivalenci jsou třídy ekvivalence

$$[x]_R = \{y \in \mathbb{R}; xRy\} = \{y \in \mathbb{R}; |x| = |y|\} = \{x, -x\},$$

což je v případě $x = 0$ jednoprvková množina, $[0]_R = \{0\}$, jinak dvouprvková.

\triangle

! Příklad 4a.c: V teorii čísel se počet různých prvočíselných dělitelů čísla n značí $\omega(n)$, například $\omega(12) = 2$, protože prvočísla 2 a 3 dělí 12. Dvanáctku dělí i jiná čísla, ale to nejsou prvočísla (ani 1 není).

Uvažujme následující relaci R na množině $A = \{28, 29, 30, 31, 32, 33, 34, 35\}$: aRb jestliže a, b mají stejný počet různých prvočíselných dělitelů, tedy $\omega(a) = \omega(b)$. Ukážeme, že jde o ekvivalenci:

Reflexivita: Evidentně $\omega(a) = \omega(a)$ pro libovolné číslo a , proto aRa .

Symetrie: Nechť $a, b \in A$. Jestliže aRb , pak $\omega(a) = \omega(b)$, proto i $\omega(b) = \omega(a)$ a tedy bRa .

Tranzitivita: Nechť $a, b, c \in A$. Jestliže aRb a bRc , pak $\omega(a) = \omega(b)$ a $\omega(b) = \omega(c)$. Proto i $\omega(a) = \omega(c)$ a tedy aRc .

Všimněte si, že jsme vlastně dokázali, že takto vzniklá relace je ekvivalence na libovolné podmnožině přirozených čísel, nejen na té naší. Teď už se na ni podíváme, nejprve si pro všechna čísla z A určíme hodnotu parametru ω , abychom viděli, které prvky jsou se kterými v relaci, pak si nakreslíme si zjednodušený graf, jak jsme to diskutovali výše.

$$\omega(28) = \omega(2^2 \cdot 7) = 2,$$

$$\omega(29) = 1,$$

$$\omega(30) = \omega(2 \cdot 3 \cdot 5) = 3,$$

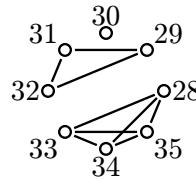
$$\omega(31) = 1,$$

$$\omega(32) = \omega(2^5) = 1,$$

$$\omega(33) = \omega(3 \cdot 11) = 2,$$

$$\omega(34) = \omega(2 \cdot 17) = 2,$$

$$\omega(35) = \omega(5 \cdot 7) = 2.$$



Třídy ekvivalence jsou $[28]_R = \{28, 33, 34, 35\} = [33]_R = [34]_R = [35]_R$, $[29]_R = \{29, 31, 32\} = [31]_R = [32]_R$ a $[30]_R = \{30\}$. Odpovídající rozklad je $S = \{\{28, 33, 34, 35\}, \{29, 31, 32\}, \{30\}\}$.

Již dříve jsme zmínili, že jakékoli porovnávání objektů pomocí shodnosti nějakého ukazatele je ekvivalence. Důkaz je shodný jako ten, který jsme zde viděli, viz cvičení 4a.7.

\triangle

Příklad 4a.d: Hledáme relaci R na množině $A = \{1, 2, 3, 13, 14, 23\}$, která by zahrnovala vztahy $1R2$, $1R3$, $13R23$ a byla by to ekvivalence, přičemž chceme nejmenší takovou relaci (relace jsou množiny dvojic, my chceme

relaci s nejmenším možným počtem dvojic, aby ještě vyhovovala zadání). Jinými slovy, hledáme ekvivalentní uzávěr relace $\{(1, 2), (1, 3), (13, 23)\}$, viz sekce 3c.6.

Ekvivalence má být reflexivní, proto nutně musíme doplnit všechny dvojice typu (a, a) . Má být také symetrická, takže nelze nepřidat dvojice $(2, 1), (3, 1)$ a $(23, 13)$. Tím jsme dostali první obrázek (jako obvykle pro zjednodušení nekreslíme smyčky a obousměrnou orientaci šipek).

Tedě je třeba doplnit hrany tak, aby vznikla relace tranzitivní, tedy je třeba uzavřít všechny nedokončené trojúhelníky. Takový je tam jen jeden. Je dobré si rozmyslet, že jsme tímto doplněním neporušili symetrii.

Dostaneme tak $R = \{(1, 1), (2, 2), (3, 3), (13, 13), (14, 14), (23, 23), (1, 2), (2, 1), (1, 3), (3, 1), (13, 23), (23, 13), (2, 3), (3, 2)\}$.

Třídy ekvivalence jsou $[1]_R = [2]_R = [3]_R = \{1, 2, 3\}$, $[13]_R = [23]_R = \{13, 23\}$, $[14]_R = \{14\}$.

Všimněte si, že kdybychom zkusili nejprve doplnit chybějící spojnice na tranzitivitu a pak teprve symetrické šipky, tak už nedostaneme ekvivalenci. Původní relace doplněná o prvky (a, a) je

$$\{(1, 1), (2, 2), (3, 3), (13, 13), (14, 14), (23, 23), (1, 2), (1, 3), (13, 23)\}.$$

V ní sice navazující dvoukroky vytvořit jdou, třeba $1R3R3$, ale žádný nás nenutí doplnit nějaké dvojice do relace, tato relace už je tranzitivní. Když v druhém kroku dodáme další dvojice k vytvoření symetrické relace, přibude tam $(2, 1)$ a náhle máme dvoukrok $2R1R3$, ke kterému neexistuje zkratka $(2, 3)$. Jinak řečeno, vidíme, že symetrizace relace může zničit její tranzitivitu, viz cvičení 3c.10 a 4a.17.

△

Příklad 4a.e: Definujme následující relaci na \mathbb{R} : Pro $x, y \in \mathbb{R}$ platí xRy právě tehdy, pokud $y - x \in \mathbb{Z}$.

Bývá dobré si novou relaci nejprve trochu vyzkoušet, máme třeba $1R4$ neboť $4 - 1 \in \mathbb{Z}$, $1R7$, ale i $7R1$ či $7R7$, zajímavější je $0.76R14.76, 396.76R0.76$ ale i $0.76R(-1.24)$, neboť $(-1.24) - 0.76 = -2 \in \mathbb{Z}$.

Je reflexivní: $x - x = 0 \in \mathbb{Z}$ a tedy i xRx pro všechna $x \in \mathbb{R}$.

Je symetrická: Vezměme libovolné $x, y \in \mathbb{R}$. Jestliže xRy , pak $y - x \in \mathbb{Z}$, potom také $x - y = -(y - x) \in \mathbb{Z}$ a máme yRx .

Je tranzitivní: Vezměme libovolné $x, y, z \in \mathbb{R}$. Jestliže xRy a yRz , pak $y - x \in \mathbb{Z}$ a $z - y \in \mathbb{Z}$, proto také $z - x = (z - y) + (y - x) \in \mathbb{Z}$ a máme xRz .

Jak vypadají třídy ekvivalence? Podle definice máme třeba pro $\pi \in \mathbb{R}$:

$$[\pi]_R = \{y \in \mathbb{R}; \pi Ry\} = \{y \in \mathbb{R}; y - \pi \in \mathbb{Z}\} = \{y \in \mathbb{R}; \exists n \in \mathbb{Z}: y - \pi = n\} = \{\pi + n; n \in \mathbb{Z}\}.$$

Třída ekvivalence π je tedy na reálné ose vidět jako nekonečný náhradník teček vzdálených od sebe o 1, přičemž jedna z nich je v π . Tedě už nepřekvapí $[\frac{1}{4}]_R = \{\frac{1}{4} + n; n \in \mathbb{Z}\}$ a $[13]_R = \mathbb{Z}$. Pro $x \in \mathbb{R}$ tedy dostáváme $[x]_R$ jako kopii množiny \mathbb{Z} posunutou o x doprava. Výrok $\mathbb{R} = \bigcup_{x \in \mathbb{R}} [x]_R$ vlastně tedy říká, že \mathbb{R} lze získat sjednocením posunutých množin \mathbb{Z} . V tom rozkladu se ovšem každá třída vyskytuje mnohokrát, dokonce nekonečně mnohokrát, například $[0]_R = [1]_R = [-1]_R = \dots = \mathbb{Z}$. To je obvyklé a vede to k následující otázce.

Máme rozklad $\{[a]_R\}_{a \in A}$, ale mnohé množiny se v něm opakují. Rádi bychom vybrali z každé třídy jednoho zástupce, čímž by vznikla množina M prvků z A taková, že $\{[a]_R\}_{a \in M}$ jsou již různé množiny, tedy každá třída je tam obsažena jen jednou, a přitom je to stále rozklad A . Často lze takovou množinu M vybrat tak, že má nějakou zajímavou strukturu, a může nám to říct něco o struktuře samotné množiny A .

Jak by se dal jeden takový zástupce z každé třídy nějak rozumně vybrat v našem příkladě? Rozmyslíme si, že $\{[x]_R\}_{0 \leq x < 1}$ je rozklad \mathbb{R} a rozsah $0 \leq x < 1$ už nelze dále zmenšit.

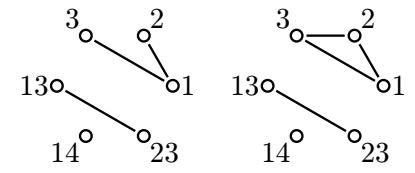
Za prvé, vezmeme-li libovolné $y \in \mathbb{R}$, tak určitě existuje $x \in (0, 1)$ a $n \in \mathbb{Z}$ takové, že $y = n + x$ (toto by mělo být jasné, x je desetinná část y a $n = \lfloor y \rfloor$). Pak $y \in [x]_R$. Toto ukazuje, že $\bigcup_{0 \leq x < 1} [x]_R = \mathbb{R}$.

Mohli bychom ještě nějakou množinu z tohoto rozkladu odebrat? Ukážeme, že ne, protože pro $x \neq y \in (0, 1)$ už platí $[x]_R \neq [y]_R$. Dokážeme to nepřímo, tedy přes obměnu (rovnosti se zpracovávají lépe než nerovnosti): Kdyby $[x]_R = [y]_R$, pak xRy , což značí $y - x \in \mathbb{Z}$. Jenže zároveň $x, y \in (0, 1)$, což znamená $|x - y| < 1$. Situace, kdy $|x - y| < 1$ a $y - x \in \mathbb{Z}$, je možná jenině pro $x = y$.

Pro $x, y \in (0, 1)$ jsme dokázali, že $[x]_R = [y]_R \implies x = y$, tudíž platí i obměna $x \neq y \implies [x]_R \neq [y]_R$.

Z důkazu je jasné, že podobně si můžeme jako množinu indexů vzít libovolný polouzavřený interval délky 1 a zase dostaneme rozklad množiny, který už nelze dále zmenšit. Jsou samozřejmě i jiné možnosti, třeba $(0, \frac{1}{2}) \cup (\frac{3}{2}, 2)$ by fungovalo, jdou vymyslet i šílenější množiny indexů.

Jedna štourová otázka nakonec: Šlo by ty zástupce vybírat tak, aby v každém intervalu typu $(n, n+1)$ pro $n \in \mathbb{Z}$ byl jen jeden? Takže pro jednu třídu bychom vybrali něco z $(0, 1)$, pro další něco z $(1, 2)$, pro další z $(-1, 0)$, pro další z $(2, 3)$ atd. Odpověď zní, že to možné není. Každá třída $[a]_R$ je totiž spočetná, ale \mathbb{R} spočetná není, tudíž



je nespočetně mnoho různých tříd ekvivalence v rozkladu. Musíme proto vybrat nespočetně mnoho zástupců, viz ten interval $\langle 0, 1 \rangle$, což uvažovaným stylem nejde, intervalů typu $\langle n, n + 1 \rangle$ je samozřejmě jen spočetně mnoho.

△

Ekvivalence a rozklady jsou užitečné z jednoho důvodu, umožňují schovávat nepodstatné detaily. Představme si, že zkoumáme nějaké objekty a zajímají nás nějaké podstatné vlastnosti V . Můžeme pak definovat ekvivalence podmírkou, že objekty sdílejí vlastnosti V . Všechny objekty s konkrétní variantou podstatných vlastností se pak dostanou do jedné skupiny a my máme možnost je zkoumat najednou, například tak, že pozkoumáme jednoho konkrétního zástupce této skupiny.

Příklad 4a.f: Toto je mírně až silně nerealistický případ ilustrující obecné povídání předchozího odstavce.

Nechť A je množina všech počítačů. Definujeme na ní relaci R takto: aRb jestliže a a b jedou pod stejným operačním systémem (včetně verze, patche, opravného balíku atp.)

Poznámka zvaná reality check: Aby tohle fungovalo, musíme předpokládat, že je tento operační systém jednoznačně dán; dá se to zařídit například tak, že pokud má nějaký počítač možnost multiple boot, tak se pro účely tohoto příkladu považuje za více počítačů, podle toho, do jakého OS naběhne. Budeme také ignorovat případy, kdy si někdo nastartuje počítač do operačního systému A a spustí si v něm emulátor systému B , ve kterém dále pracuje. Konec poznámky omezující realitu tak, abychom při definici této relace nemuseli přemýšlet, jak se vypořádat se vším, co si lidi dokážou zbastlit.

Snadno se ověří, že jde o ekvivalence. Reflexivita je snad jasná, symetrie už z definice. Tranzitivita: Jestliže aRb a bRc , pak mají a, b stejný OS a b, c mají stejný OS. Podle naší poznámky má b jen jeden OS, je tedy společný pro a, c a máme aRc .

Tato ekvivalence nám rozloží množinu všech počítačů do tříd podle toho, na jakém OS jedou (desítky podverzí různých generací Windblows, různí tučňáci s ještě různějšími X-Win nápady, rozličné verze DOSu pro staromilce (osobně doporučuji 4DOS), Nexty, Applí operační systémy,...)). Náš pohled na počítače se tím prudce zjednoduší a pokud za námi někdo přijde pro radu, tak se prostě zeptáme, do jaké třídy jeho hromádka silikonu patří, a budeme mu s vysokou pravděpodobností schopni poradit, aniž bychom přesně věděli, co má vlastně za krabičku.

To je samozřejmě další bod, ve kterém tento školní případ narází na realitu :-), počítače se již nějakou dobou chovají často nevypočítatelně a jeví známky toho, že mají vlastní osobnost (většinou zlomyslnou).

△

Nicméně i když nedokonalá, pořád je to dostatečně rozumně fungující metoda na to, aby zjednodušila náš život, takže ji vidíme všude kolem nás (třeba když se nám pokazí auto, tak je v servisu zajímá, jaké je značky, ne konkrétní výrobní série, a většinou to stačí; studenti jsou rozděleni do tříd podle ročníků atd atd.). Dokonce se dá říct, že náš mozek přímo funguje v těchto nedokonalých ekvivalentích, každý si během dětství vytváříme obraz světa tak, že si věci rozdělujeme do škatulek (psi, kočky, květiny, dospělí, děti, upíři, samopaly) a pak podvědomě hodnotíme podle toho, do které skupiny co patří, jakkoliv je to někdy nespravedlivé. To už je život.

Samozřejmě ve světě matematiky věci fungují mnohem vyhraněněji a hranice nejsou tak zamlžené, takže se ekvivalence stávají mocným nástrojem k vytváření skupin, které pak zkoumáme najednou (zobrazení prostá a ta ostatní, množiny konečné, spočetné a ty ostatní, ...).

Příklad 4a.g:

1) Pravděpodobně všichni čtenáři se s třídami ekvivalence setkali v geometrii. Existuje tam pojem vázaných vektorů, což jsou zhruba řečeno šipky například v rovině, podle toho, v kolikarozměrném světě pracujeme. Každá taková šipka někde začíná a někde končí a je jich opravdu dost. V aplikacích se ale ukazuje, že to opravdu důležité není začátek a konec, ale směr a velikost. Jinými slovy, pokud mají dva vektory tyto dva atributy stejné, tak už se mohou lišit umístěním a nějak zvlášť nám to nevadí.

Formálně se tedy zavede ekvivalence, dva vázané vektory jsou ekvivalentní, pokud mají stejný směr i velikost. Vzniklé třídy ekvivalence jsou to, čemu říkáme volné vektory, přičemž všechny vektory z jedné konkrétní třídy považujeme v zásadě za stejné, což se odráží mimo jiné i v tom, že mají všechny stejné souřadnice. V konkrétních situacích si pak ze třídy vybíráme vhodného zástupce, například pokud chceme znát souřadnice příslušné k dané třídě vektorů, tak si z ní vybereme vektor, který začíná v počátku, a jeho konec nám dá souřadnice. Když chceme dva volné vektory (tedy vlastně dvě třídy ekvivalence) sečist, tak si u prvního vybíráme zástupce začínajícího v počátku, ale u druhého už vybíráme zástupce začínajícího tam, kde první skončil.

Pro studenty začínající s vektory je někdy těžké si na taková kouzla zvyknout, ale pokud si dobře rozumí s ekvivalencemi, tak to vlastně není nic zvláštního.

2) Myšlenka, že do pytlíku schováme spoustu objektů a pak již manipulujeme s pytlíkem jako celkem, se přirozeně objevuje i v analýze. Neurčitý integrál funkce f je jistá množina funkcí, ale my s ní běžně pracujeme jako s jedním objektem.

Formálně bychom to zařídili takto. Nechť F je množina všech reálných funkcí, které jsou diferencovatelné na vnitřku definičního oboru. Pak definujme relaci $F \sim G$ právě tehdy, když $F' = G'$. Snadno se ukáže, že tato relace je ekvivalence. Když se podíváme na jednu třídu ekvivalence $[F]_\sim$, tak hned vidíme, že je to vlastně neurčitý integrál k funkci F' . Kdykoliv pracujeme se symbolem $\int F' dx$, tak vlastně pracujeme s touto třídou ekvivalence. Jsou situace (například při integraci per partes), kdy potřebujeme nějakou primitivní funkci a vybíráme si dle libosti, což zase odpovídá naší zkušenosti s ekvivalencemi, že v zásadních aplikacích na konkrétní volbě zástupce třídy vůbec nezáleží.

△

Zatímco ne každý se potkal s volnými a vázanými vektory, existuje jeden objekt, který potkali úplně všichni a který jsou vlastně také třídy ekvivalence. Jsou to zlomky, viz kapitola 8d. Teď jedna skutečná aplikace.

Příklad 4a.h: Uvažujme množinu A řetězců nad anglickou abecedou, zvolme pevně nějaké $n \in \mathbb{N}$. Definujeme relaci \sim na A pro slova v, w předpisem $v \sim w$ bud' jestliže jsou obě slova kratší než n znaků a jsou stejná, nebo jestliže mají obě slova alespoň n znaků a na prvních n znacích se shodují.

Použili jsme pro tuto relaci značku \sim , aby se nám R nepletlo do písmenek v řetězcích. Je snadné ověřit, že jde o ekvivalence.

Například jestliže $n = 3$, pak máme de \sim de a žádné jiné slovo už s „de“ ekvivalentní není, ale s delším řetězcem „pec“ jsou ekvivalentní třeba „pec“, „pecka“, „pecen“, „pecxyzancdefghij“ a nekonečně mnoho dalších.

Takže $[pec]_\sim$ jsou všechny řetězce, které začínají „pec“, zatímco pro libovolný dvou či jednoznakový řetězec w máme $[w]_\sim = \{w\}$.

Tento příklad není uměle vymyšlen, jazyk C dovoloval libovolnou délku názvu proměnných, ale při rozlišování používal jen několik prvních písmen. Množina tříd ekvivalence pak vlastně udává, které různé proměnné lze používat.

△

Užitečnost takového shlukování (a také to, že s každou třídou pak pracujeme jako s jedním samostatným objektem) vedly k zavedení pojmu, který to shrnuje. My zde s ním pracovat nebudeme, uvádíme jej pro úplnost, kdyby na něj čtenář někdy narazil.

Definice.

Nechť R je relace ekvivalence na množině A . Množině $\{[a]_R; a \in A\}$ říkáme **faktorová množina podle ekvivalence R** a značíme ji A/R .

Jisté vysvětlení značení: Pokud by byla množina A konečná a všechny třídy ekvivalence měly stejnou mohutnost n , pak má množina A/R mohutnost $|A|/n$.

V matematice je někdy klíčový obrat, kdy původní množinu rozdělíme na třídy ekvivalence a s těmi pak pracujeme jako s objekty, aniž bychom se příliš vrtaliv tom, co se vlastně děje uvnitř takových tříd. Již jsme viděli příklady s vektory nebo s primitivními funkcemi, další (který vlastně známe všichni) je v kapitole 8d o racionálních číslech, další důležitý v kapitole 7, za přečtení stojí i poznámka 4b.9.

Ekvivalence a operace

Z představy, kterou o ekvivalence máme, by mělo být vidět, že pokud z grafu ekvivalence odebereme nějaké vrcholy včetně spojnic k nim vedoucích, tak z toho zase zůstane graf rozdělený na izolovaná hnizda.

Fakt 4a.3.

Nechť R je relace ekvivalence na nějaké množině A . Pak restrikce R na libovolnou podmnožinu A je zase ekvivalence.

Důkaz plyne okamžitě z Faktu 3b.4.

V praxi to třeba znamená, že když jsme tady dokázali, že relace $=$ je ekvivalence na \mathbb{R} , tak už můžeme rovnost používat pro libovolnou podmnožinu $M \subseteq \mathbb{R}$, třeba na \mathbb{N} , a zase to bude ekvivalence. Mnohé z příkladů v tomto textu jsou vyráběny právě restrikcí známé relace na nějakou pěknou malou množinu, aby ten příklad moc nenarostl.

U operací množinových se stačí odvolut na Fakt 3c.2, viz cvičení 4a.14. Jak je na tom skládání? Když složíme dvě různé ekvivalence, pak ekvivalence vzniknout nemusí, problém je s tranzitivitou, viz diskuse po Faktu 3c.4.

Zbývá tedy situace, kdy skládáme stejnou ekvivalenci samu se sebou. Bylo by svůdné využít Větu 3c.5 k tvrzení, že když je R ekvivalence, tak je R^n zase ekvivalence. Ve skutečnosti je to ovšem ještě jednodušší, z Věty 3b.6 totiž okamžitě vyplýne toto:

Důsledek 4a.4.

Nechť R je relace ekvivalence na množině A . Pak pro všechna $n \in \mathbb{N}$ je $R^n = R$.

Ekvivalence tedy nemá smysl umocňovat a už vůbec se tím nemohou změnit jejich vlastnosti. Podobně se nemá smysl ptát, co s ekvivalence R udělá, když ji obrátíme, protože ekvivalence jsou symetrické a tudíž zase $R^{-1} = R$, nic nového nedostaneme.

V sekci 3b.9 jsme představili součinovou relaci, kdy dva vektory porováváme prostřednictvím jejich souřadnic. Z výsledků dotyčné sekce okamžitě plyne následující.

Věta 4a.5.

Nechť $(A_1, R_1), \dots, (A_n, R_n)$ jsou množiny s relacemi ekvivalence. Pak příslušná součinová relace R na množině $A = A_1 \times \dots \times A_n$ je ekvivalence na A .

O ekvivalencech a rozkladech se toho samozřejmě dá říct více, pro pár zajímavých nápadů doporučujeme cvičení, například cvičení 4a.15.

Cvičení

Cvičení 4a.1 (rutinní): Určete, zda relace určené následujícími maticemi jsou ekvivalence:

$$(i) M = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}; \quad (ii) M = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}; \quad (iii) M = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad (iv) M = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}; \quad (v) M = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Cvičení 4a.2 (rutinní): Rozhodněte, která z následujících relací na množině A lidí je ekvivalence, odpověď zdůvodněte:

- (i) aRb znamená, že a, b se někdy potkali;
- (ii) aRb znamená, že a, b mají společného známého;
- (iii) aRb znamená, že a, b mají stejnou nejoblíbenější knihu (zde předpokládejte, že každý má jen jednu takovou).

Cvičení 4a.3 (rutinní): Na množině $A = \{1, 2, 3, 4, 5, 6\}$ uvažujte následující relaci: aRb jestliže mají slovní vyjádření pro a a b stejný počet písmen.

Vypište prvky této relace. Ukažte, že je to ekvivalence, pak nakreslete její zjednodušený graf. Pro všechny prvky určete jejich třídy ekvivalence a najděte odpovídající rozklad množiny.

Cvičení 4a.4 (rutinní, poučné, zkouškové): Ve cvičení 3b.4 se vlastně ukázalo, že následující relace na \mathbb{Z} jsou ekvivalence:

- (i) aRb jestliže $|a| = |b|$;
- (ii) aRb jestliže $a - b = 2k$ pro nějaké $k \in \mathbb{Z}$.

Pro každou z těchto relací najděte $[13]_R$ a poté popište obecně, jak třídy ekvivalence vypadají.

Cvičení 4a.5 (rutinní, poučné, zkouškové): Rozhodněte, která z následujících relací na $\mathbb{Z} \times \mathbb{Z}$ je ekvivalence:

- (i) $(a_1, a_2)R(b_1, b_2)$ jestliže $a_1 b_2 = a_2 b_1$;
- (ii) $(a_1, a_2)R(b_1, b_2)$ jestliže $a_1 + b_2 = a_2 + b_1$;
- (iii) $(a_1, a_2)R(b_1, b_2)$ jestliže $a_1 - a_2 = b_1 - b_2$;
- (iv) $(a_1, a_2)R(b_1, b_2)$ jestliže $a_1 - a_2 = b_2 - b_1$.

Pro ty, které jsou ekvivalence, najděte $[(1, 2)]_R$.

Cvičení 4a.6 (dobré, poučné): Uvažujme množinu A všech zobrazení ze \mathbb{Z} do \mathbb{Z} . Která z následujících relací je ekvivalence? Pro ty, které jsou, určete třídy ekvivalence.

- (i) $\{(T, S); T(1) = S(1)\}$;
- (ii) $\{(T, S); T(0) = S(0)$ nebo $T(1) = S(1)\}$;
- (iii) $\{(T, S); T(0) = S(0)$ a $T(1) = S(1)\}$;
- (iv) $\{(T, S); T(n) - S(n) = 1$ pro všechna $n \in \mathbb{Z}\}$;
- (v) $\{(T, S); \exists C \in \mathbb{R} \forall n \in \mathbb{Z} : T(n) - S(n) = C\}$;
- (vi) $\{(T, S); T(1) = S(0)$ a $T(0) = S(1)\}$.

Cvičení 4a.7 (poučné): Nechť $T: A \mapsto B$ je zobrazení. Definujme relaci na A předpisem aRb právě tehdy, když $T(a) = T(b)$. Dokažte, že je to ekvivalence, a určete její třídy ekvivalence.

Poznámka: Tím obecně dokážeme, že porovnávání objektů pomocí shodnosti nějakého ukazatele dává ekvivalence. Všechny ekvivalence lze takto reprezentovat, jako úrovnové množiny nějakého zobrazení.

Cvičení 4a.8 (poučné): Nechť R je relace na množině všech trojúhelníků v rovině definovaná takto: aRb jestliže lze trojúhelník a získat z trojúhelníka b posunem, rotací, zrcadlením či kombinací několika těchto transformací. Dokažte, že je to relace ekvivalence.

Poznámka: Toto nám umožní redukovat všechny trojúhelníky v rovině jen na informaci o délkách stran a úhlech.

Cvičení 4a.9 (poučné): Uvažujme šachovnici 2×2 , nechť A je množina všech možných obarvení polí této šachovnice bílou a černou (rozumí se tím, že každé pole je vybarveno právě jednou z těchto dvou barev). Definujeme relaci R na A předpisem: aRb jestliže lze obarvení b získat z obarvení a rotací či zrcadlením šachovnice či kombinací několika těchto transformací. (Zpřesňující poznámka: připomíná se jen takové rotace a zrcadlení, aby byla zase šachovnice jako celek v základní pozici, čili rotace o násobky pravého úhlu a zrcadlení okolo os souměrnosti šachovnice). Dokažte, že jde o relaci ekvivalence, a najděte množinu nějakých obarvení tak, aby byla nejmenší možnou množinou zástupců tříd ekvivalence.

Poznámka: Geometrie a ekvivalence se spolu potkávají velice často, například volné vektory jsou třídy ekvivalence vázaných vektorů.

Cvičení 4a.10 (poučné): Ve cvičení 3b.7 se ukázalo, že následující relace jsou ekvivalence. Určete, jak pro ně vypadají třídy ekvivalence.

- (i) Relace R na množině A všech reálných polynomů definovaná jako $p(x)Rq(x)$ jestliže mají p a q stejný stupeň.
- (ii) Relace R na množině A všech reálných polynomů definovaná jako $p(x)Rq(x)$ jestliže mají p a q stejné reálné kořeny včetně násobnosti.
- (iii) Relace R na množině A všech reálných polynomů definovaná jako $p(x)Rq(x)$ jestliže mají p a q stejné komplexní kořeny včetně násobnosti.

Cvičení 4a.11 (rutinní): Uvažujte následující relace definované na množině všech (konečných) binárních řetězců. Pro každou z nich dokažte, že je ekvivalence, a určete, jak vypadají $[1001]_R$ a $[0]_R$.

- (i) aRb jestliže řetězce a, b mají stejný počet jedniček;
- (ii) aRb jestliže řetězce a, b mají stejnou paritu výskytu jedniček.

Cvičení 4a.12 (rutinní): Který z následujících souborů podmnožin množiny $A = \{1, 2, 3, 4, 5, 6, 7\}$ je jejím rozkladem? Nakreslete pak graf odpovídající ekvivalence.

- (i) $S = \{\{1, 2, 5\}, \{6, 7\}, \{3\}\};$
- (ii) $S = \{\{1, 2, 5\}, \{6, 7\}, \{3, 4\}\};$
- (iii) $S = \{\{1, 2, 4, 5\}, \{6, 7\}, \{3, 4\}\}.$

Cvičení 4a.13 (rutinní, poučné): Který z následujících souborů je rozkladem množiny všech binárních řetězců?

- (i) Řetězce začínající 00, řetězce začínající 01, řetězce začínající 11, řetězce začínající 10;
- (ii) Řetězce začínající 00, řetězce začínající 1;
- (iii) Řetězce obsahující 00, řetězce obsahující 01, řetězce obsahující 11, řetězce obsahující 10;
- (iv) Řetězce začínající 00, řetězce začínající 01, řetězce začínající 1.

Cvičení 4a.14 (rutinní, poučné): Který z následujících souborů podmnožin jsou rozkladem množiny $\mathbb{Z} \times \mathbb{Z}$?

- (i) množina párů (x, y) takových, že x nebo y je liché; množina párů (x, y) takových, že x je sudé; množina párů (x, y) takových, že y je sudé;
- (ii) množina párů (x, y) takových, že x nebo y je liché; množina párů (x, y) takových, že x a y jsou sudé;
- (iii) množina párů (x, y) takových, že $xy > 0$; množina párů (x, y) takových, že $xy < 0$.

Cvičení 4a.15 (poučné): Dokažte: Jsou-li R_1, R_2 relace ekvivalence na téže množině A , pak $R_1 \cap R_2$ je také ekvivalence, ale \overline{R}_1 ani $R_1 - R_2$ nejsou nikdy ekvivalence.

Cvičení 4a.16 (poučné, dobré): Nechť S_1, S_2 jsou dva rozklady téže množiny A . Řekneme, že S_1 je **zjemnění (refinement)** rozkladu S_2 , jestliže každá množina z S_1 je podmnožinou nějaké množiny z S_2 .

Intuitivně, S_1 vzniklo tak, že se dále rozdělily nějaké množiny z S_2 .

Dokažte následující: Nechť R_1, R_2 jsou relace ekvivalence na množině A . Pak $R_1 \subseteq R_2$ právě tehdy, když je rozklad $\{[a]_{R_1}\}$ zjemněním rozkladu $\{[a]_{R_2}\}$.

Intuitivně: Čím více dvojic v relaci ekvivalence, tím větší třídy—ano, to zní logicky.

Cvičení 4a.17 (poučné, dobré): Jestliže uděláme tranzitivní uzávěr symetrického uzávěru reflexivního uzávěru relace, dostaneme nutně ekvivalenci?

Cvičení 4a.18 (poučné, dobré): Nechť R je relace na množině A . Řekneme, že je **cirkulární (circular)**, jestliže pro každé $a, b, c \in A$ platí: Jestliže aRb a bRc , pak cRa .

Dokažte: Relace r je ekvivalence právě tehdy, když je reflexivní a cirkulární.

Řešení:

4a.1: Nejsnáze poznáme reflexivitu, matice musí mít 1 všude na diagonále. Tím jsme vyloučili (ii). Symetrii poznáme tak, že matice musí být symetrická. Kontrola dvojic souměrných podél diagonály vyřadí případy (i) a (iv). Zbývají případy (iii) a (v) a kontrola tranzitivity. Jedna možnost je použít Booleanovský součin. Druhá možnost je vypsat si z matice všechny nediagonální jedničky jako dvojice v relaci a zkoumat tranzitivitu na nich. U (v) se tím přijde na navazující dvojice (1, 2) a (2, 3), kdy zkratka (1, 3) v relaci není, tedy (v) není tranzitivní. U (iii) i tento test uspěje, je to tedy ekvivalence.

4a.2: (i): Není tranzitivní, a potkal b , b potkal c nemusí znamenat a potkal c .

(ii): Není tranzitivní.

(iii): Je ekvivalence. Označme $k(a)$ nejoblíbenější knihu a . Pak $k(a) = k(a)$ tedy aRa , reflexivita.

Dále $aRb \Rightarrow k(a) = k(b) \Rightarrow k(b) = k(a) \Rightarrow bRa$, symetrie.

Dále $aRb \wedge bRc \Rightarrow k(a) = k(b) \wedge k(b) = k(c) \Rightarrow k(a) = k(c) \Rightarrow aRc$ tranzitivita.

4a.3: Je praktické zavést funkci $p(n)$ značící počet písmen v čísle n . Pak $p(1) = 5$, $p(2) = 3$, $p(3) = 3$, $p(4) = 5$, $p(5) = 3$, $p(6) = 4$. Odtud snadno

$$R = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (1, 4), (4, 1), (2, 3), (3, 2), (2, 5), (5, 2), (3, 5), (5, 3)\}.$$

$$R: a \in A \Rightarrow p(a) = p(a) \Rightarrow aRa.$$

$$S: a, b \in A: aRb \Rightarrow p(a) = p(b) \Rightarrow p(b) = p(a) \Rightarrow bRa.$$

$$T: a, b, c \in A: aRb \wedge bRc \Rightarrow p(a) = p(b) \wedge p(b) = p(c) \Rightarrow p(a) = p(c) \Rightarrow aRc.$$

$$[1]_R = [4]_R = \{1, 4\}, [2]_R = [3]_R = [5]_R = \{2, 3, 5\}, [6]_R = \{6\}.$$

$$S = \{\{1, 4\}, \{2, 3, 5\}, \{6\}\}.$$

4a.4: (i): $[13]_R = \{a \in \mathbb{Z}; 13Ra\} = \{a \in \mathbb{Z}; |13| = |a|\} = \{-13, 13\}$; $[n]_R = \{n, -n\}$.

(ii): $[13]_R = \{a \in \mathbb{Z}; 13Ra\} = \{a \in \mathbb{Z}; \exists k \in \mathbb{Z}: 13 - b = 2k\} = \{a \in \mathbb{Z}; \exists k \in \mathbb{Z}: b - 13 = 2k\}$

$$= \{a \in \mathbb{Z}; \exists k \in \mathbb{Z}: b = 13 + 2k\} = \{13 + 2k; k \in \mathbb{Z}\};$$

$$[n]_R = \{n + 2k; k \in \mathbb{Z}\}.$$

4a.5: (i): R: $(a_1, a_2)R(a_1, a_2) \iff a_1 a_2 = a_2 a_1$ vždy pravda.

S: $(a_1, a_2)R(b_1, b_2) \Rightarrow a_1 b_2 = a_2 b_1 \Rightarrow b_1 a_2 = b_2 a_1 \Rightarrow (b_1, b_2)R(a_1, a_2)$, ano.

T: $(a_1, a_2)R(b_1, b_2) \wedge (b_1, b_2)R(c_1, c_2) \Rightarrow a_1 b_2 = a_2 b_1 \wedge b_1 c_2 = b_2 c_1$. Z první $b_2 = \frac{a_2 b_1}{a_1}$ do druhé

$b_1 c_2 = \frac{a_2 b_1}{a_1} c_1 \Rightarrow a_1 c_2 = a_2 c_1 \Rightarrow (a_1, a_2)R(c_1, c_2)$. Ale co když $a_1 = 0$? Rozbor situace, podezření, není tranzitivní!

Protipříklad: $(1, 1)R(0, 0)$, $(0, 0)R(2, 1)$, ale $(1, 1)R(2, 1)$ neplatí.

Není ekvivalence.

(ii): R: $(a_1, a_2)R(a_1, a_2) \iff a_1 + a_2 = a_2 + a_1$ vždy pravda.

S: $(a_1, a_2)R(b_1, b_2) \Rightarrow a_1 + b_2 = a_2 + b_1 \Rightarrow b_1 + a_2 = b_2 + a_1 \Rightarrow (b_1, b_2)R(a_1, a_2)$, ano.

T: $(a_1, a_2)R(b_1, b_2) \wedge (b_1, b_2)R(c_1, c_2) \Rightarrow a_1 + b_2 = a_2 + b_1 \wedge b_1 + c_2 = b_2 + c_1$. Sečteme obě rovnice: $a_1 + b_2 + b_1 + c_2 = a_2 + b_1 + b_2 + c_1$, zkrátíme, $a_1 + c_2 = a_2 + c_1 \Rightarrow (a_1, a_2)R(c_1, c_2)$, tranzitivní. Je to ekvivalence.

$[(1, 2)]_R = \{(b_1, b_2) \in \mathbb{Z}^2; (1, 2)R(b_1, b_2)\} = \{(b_1, b_2) \in \mathbb{Z}^2; 1 + b_2 = 2 + b_1\} = \{(b_1, b_2) \in \mathbb{Z}^2; b_2 = 1 + b_1\} = \{(k, k + 1); k \in \mathbb{Z}\}.$

(iii): R: $(a_1, a_2)R(a_1, a_2) \iff a_1 - a_2 = a_1 - a_2$ vždy pravda.

S: $(a_1, a_2)R(b_1, b_2) \Rightarrow a_1 - a_2 = b_1 - b_2 \Rightarrow b_1 - b_2 = a_1 - a_2 \Rightarrow (b_1, b_2)R(a_1, a_2)$, ano.

T: $(a_1, a_2)R(b_1, b_2) \wedge (b_1, b_2)R(c_1, c_2) \Rightarrow a_1 - a_2 = b_1 - b_2 \wedge b_1 - b_2 = c_1 - c_2 \Rightarrow a_1 - a_2 = c_1 - c_2 \Rightarrow (a_1, a_2)R(c_1, c_2)$, tranzitivní. Je to ekvivalence.

$[(1, 2)]_R = \{(b_1, b_2) \in \mathbb{Z}^2; (1, 2)R(b_1, b_2)\} = \{(b_1, b_2) \in \mathbb{Z}^2; 1 - 2 = b_1 - b_2\} = \{(b_1, b_2) \in \mathbb{Z}^2; b_2 = 1 + b_1\} = \{(k, k + 1); k \in \mathbb{Z}\}.$

(iv): R: $(a_1, a_2)R(a_1, a_2) \iff a_1 - a_2 = a_2 - a_1$ není vždy pravda.

Protipříklad: neplatí $(1, 2)R(1, 2)$, protože neplatí $1 - 2 = 2 - 1$. Není to ekvivalence.

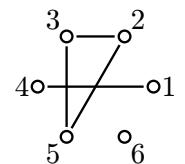
4a.6: (i): Ano. R: $T(0) = T(0)$. S: $T(0) = S(0) \Rightarrow S(0) = T(0)$. T: $T(0) = S(0) \wedge S(0) = U(0) \Rightarrow T(0) = U(0)$.

Zvolme si číslo $c \in \mathbb{Z}$. Pak třída ekvivalence odpovídající tomuto číslu je množina všech zobrazení, jejichž grafy procházejí bodem $(1, c)$. Dá se to tedy představit, že se grafy v tomto bodě setkají, ale doleva i doprava už se rozvíhají dle libosti. Zvolíme-li jiné c , dostaneme jinou třídu, jedna třída se tedy posouvá nahoru a dolů podle volby c .

(ii): Není T; stačí zvolit T tak, aby $T(0) = 1$ a $T(1) = 3$, S tak aby $S(0) = 1$ (pak $(T, S) \in R$) a $S(1) = 2$ a U tak, aby $U(0) = 5$ a $U(1) = 2$ (pak $(S, U) \in R$), ale neplatí $(T, U) \in R$.

(iii): Ano. Podobné jako (i), třída ekvivalence jsou všechna zobrazení, jejichž grafy procházejí společným bodem $(0, c)$ a také $(1, d)$.

(iv): Není R, S, T.



(v): Ano. Třída ekvivalence vznikne tak, že si vezmeme jedno zobrazení a odpovídající třída se skládá ze všech posunů jeho grafu nahoru a dolů.

(vi): Není R,T.

4a.7: R: $T(a) = T(a) \implies aRa$;

S: $aRb \implies T(a) = T(b) \implies T(b) = T(a) \implies bRa$;

T: $aRb \wedge bRc \implies T(a) = T(b) \wedge T(b) = T(c) \implies T(a) = T(c) \implies aRc$;

$[a]_R = \{b \in A; aRb\} = \{b \in A; T(a) = T(b)\} = T^{-1}[T(a)]$.

4a.8: R: trojúhelník získáme z něj samého žádnou transformací.

S: získali jsme b z a pomocí transformací, tak je uděláme pozpátku a naopak a dostaneme $z b$ to a .

T: z a získáme transformacemi b , z toho pak transformacemi c , spojíme je, dostaváme z a transformacemi c .

4a.9: Ekvivalence: Podobné předchozímu.

Množina: viz pět obarvení napravo, ostatní z nich dostaneme ale jedno z druhého ne.

b	b	c	b	c	c	c	b	c	c	c	c
b	b	b	b	b	b	b	c	c	b	c	c

4a.10: (i): $[p(x)]_R$ jsou všechny polynomy stejněho stupně jako p .

(ii): $[p(x)]_R$ jsou všechny polynomy, které mají stejné reálné kořeny včetně násobnosti. Víc se říci nedá.

(iii): $[p(x)]_R$ jsou všechny polynomy, které mají stejné komplexní kořeny včetně násobnosti. To ale znamená, že mají stejné kořenové faktory $(x - \lambda)$ v rozkladu, mohou se tedy lišit jen konstantou před kořenovými faktory.

Závěr: Jde o násobky p , $[p(x)]_R = \{a \cdot p(x); a \in \mathbb{R} - \{0\}\}$.

Proč to nešlo u (ii)? Protože v reálném oboru rozklad není zaručen, například $p(x) = (x - 13)(x^2 + 1)$ a $q(x) = (x - 13)(x^2 + 4)$ mají stejné reálné kořeny, ale jeden není násobkem druhého.

4a.11: (i) Zavedte si $j(r)$ jako počet jedniček v řetězci r , pak $aRb \iff j(a) = j(b)$ a důkaz je stejný jako třeba ve cvičení 4a.3.

$[1001]_R$ jsou všechny řetězce s přesně dvěma jedničkami. $[0]_R$ jsou všechny řetězce bez jedniček.

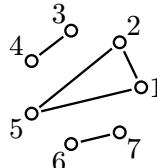
(ii) Zavedte $j(a)$ jako počet jedniček, pak $aRb \iff j(a) - j(b) = 2k$ pro $k \in \mathbb{Z}$, je to kombinace cvičení 4a.3 a cvičení 4a.4 (ii).

$[1001]_R$ jsou všechny řetězce se sudým počtem jedniček, ditto $[0]_R$.

4a.12: (i): chybí 4

(ii): ano viz obrázek

(iii): překryv, $\{1, 2, 4, 5\} \cap \{3, 4\} \neq \emptyset$.



4a.13: (i): ano; (ii): ne, chybí řetězce začínající 01; (iii): ne, množiny se překrývají, třeba řetězec 001 je v prvních dvou; (iv): ano.

4a.14: (i): ne, třeba pár (lichá, sudá) je v první i třetí množině; (ii): ano, každý pár někam patří, žádný není v obou; (iii): ne, nepokrývá páry, ve kterých je nula.

4a.15: Průnik: viz Fakt 3c.2.

Nechť $a \in A$. Pak $(a, a) \in R_1$, proto $(a, a) \notin \overline{R}_1$, tedy \overline{R}_1 není reflexivní, tudíž ani ekvivalence.

Nechť $a \in A$. Pak $(a, a) \in R_1$ a $(a, a) \in R_2$, proto $(a, a) \notin R_1 - R_2$, tedy $R_1 - R_2$ není reflexivní, tudíž ani ekvivalence.

4a.16: \implies : Nechť $R_1 \subseteq R_2$. Tvrdíme, že vždy $[a]_{R_1} \subseteq [a]_{R_2}$: $b \in [a]_{R_1} \implies aR_1b \implies aR_2b \implies b \in [a]_{R_2}$.

\Leftarrow : Nechť je $\{[a]_{R_1}\}$ zjemnění $\{[a]_{R_2}\}$. Nechť aR_2b . Pak $b \in [a]_{R_1}$. Protože musí existovat množina z $\{[a]_{R_2}\}$ taková, že $[a]_{R_1}$ je její podmnožinou, existuje $c \in A$ takové, že $[a]_{R_1} \subseteq [c]_{R_2}$. Pak tedy i $a, b \in [c]_{R_2}$ a proto aR_2b .

4a.17: Ano. Vznikne tak ekvivalentní uzávěr. Pozor, na pořadí záleží, viz příklad 4a.d a cvičení 3c.10.

4a.18: Ekvivalence je reflexivní. Je cirkulární? Nechť $aRbRc$. Z tranzitivity aRc , pak ze symetrie cRa .

Nechť je R reflexivní a cirkulární. Symetrie: Nechť aRb . Z reflexivity i bRb , proto cirkularita dává bRa . Tranzitivita: Nechť $aRbRc$. Cirkularita dá cRa , symetrie pak aRc .

4b. Částečná uspořádání

Jedním z častých úkolů je srovnat prvky jeden za druhým dle nějakého kritéria či alespoň mezi nimi udělat trochu pořádek. Můžeme třeba při pohledu na seznam úkolů rozpoznat, že určité z nich musí být udělány před jinými, někdy z toho vznikne již jediné možné pořadí pro všechny, jindy je těch omezení méně a zbude jistá míra volnosti. Když se podíváme na seznam doporučených předmětů, které by měl student computer science absolvovat, tak jsou určitá omezení zvaná prerekvizity, ale také určitá volnost. Při vaření asi nemá smysl dát sůl do hrnce, dát to na oheň rozpálit a pak teprve nalít vodu (i když sranda by to asi byla[†]), na druhou stranu bývá jedno, v jakém pořadí nastrouháme jednotlivé brambory.

[†] Jakékoli škody na majetku či zdraví vzniklé v souvislosti s tímto odstavcem jsou čistě na zodpovědnost čtenáře.

Společným prvkem těchto příkladů je, že umíme srovnávat určité jednotlivé dvojice podle nějakého kritéria a snažíme se nějak uspořádat celou množinu. Asi nejpohodlnějším srovnáním, se kterým se potkáváme, je srovnávání (celých) čísel na reálné ose, protože každé dvě umíme porovnat, a když nám někdo dá konečnou množinu čísel, tak je umíme seřadit podle velikosti. To je jakýsi ideál, ale rovnou se smířme s tím, že velice zřídka dosažitelný, většinou jsme rádi za alespoň nějaká srovnání.

Abychom mohli rozumně pracovat, budeme od srovnávací relace chtít nějaké vlastnosti. Praxe ukázala, který typ srovnání se používá nejčastěji, nikterak překvapivě je zde silná inspirace nerovností $x \leq y$ (či $x \geq y$).

! Definice.

Nechť R je relace na nějaké množině A . Řekneme, že R je **částečné uspořádání**, jestliže je reflexivní, antisymmetrická a tranzitivní.

V tom případě řekneme, že dvojice (A, R) je **částečně uspořádaná množina**.

A relation R on a set A is called a **partial ordering** or a **partial order** if it is reflexive, antisymmetric and transitive.

In that case we say that the pair (A, R) is a **partially ordered set**, often we just say **poset**.

Poznámka k anglické terminologii: V češtině se bohužel neujalo „čumnožina“ ani „čužina“.

Často říkáme jen „uspořádání“. Uvidíme, že uspořádání může být víc (částečné, lineární, dobré), takže když řekneme „uspořádání“, myslíme tím jen to nejobyčejnější, tedy částečné, pokud nemáme z kontextu dánou víc vlastností.

Příklad 4b.a: V kapitole 3b jsme viděli, že relace $x \leq y$ a $x \geq y$ jsou částečná uspořádání na \mathbb{N} , na \mathbb{Z} , na \mathbb{Q} , na \mathbb{R} a vlastně na libovolné podmnožině reálných čísel.

△

! Značení: Aby se zdůraznila srovnávací povaha relace částečného uspořádání, používá se často namísto symbolu aRb symbol $a \preceq b$. I my to zde zavedeme, pak také hovoříme o částečně uspořádané množině (A, \preceq) .

Příklad 4b.b: Uvažujme relaci R danou předpisem $|x| \leq |y|$ třeba na množině \mathbb{Z} . Tato relace je zjevně reflexivní, protože $|x| \leq |x|$ pro libovolné číslo, tedy xRx .

Tranzitivita je také splněna: Jsou-li x, y, z čísla ze \mathbb{Z} splňující xRy a yRz , pak $|x| \leq |y|$ a $|y| \leq |z|$, odtud $|x| \leq |z|$ a tedy xRz .

Bohužel ale selže antisimetrie, protože například $(-13)R13$ a $13R(-13)$, přitom neplatí $-13 = 13$. Tato relace proto není částečným uspořádáním. Přesto je to bezesporu užitečná relace, často porovnáváme objekty podle jejich velikosti neboli magnitudy. I pro takovéto relace tedy budeme chtít vytvořit teoretický kabát, jeden možný přístup ukážeme v Poznámce 4b.9.

△

Ted si ukážeme dva nejdůležitější příklady částečných uspořádání.

Příklad 4b.c: Nechť \mathcal{M} je nějaký soubor množin a uvažujme relaci na \mathcal{M} danou pro $A, B \in \mathcal{M}$ jako $A \preceq B$ právě tehdy, když $A \subseteq B$. Prostě nás zajímá relace inkluze na množinách. Je snadné nahlédnout, že jde o částečné uspořádání (viz Fakt 2a.1 (i), Fakt 2a.2 a Fakt 2a.3), takže (\mathcal{M}, \subseteq) je částečně uspořádaná množina.

Toto bude pro nás velice důležitý příklad, protože na rozdíl od nerovností nám nenaznačuje věci, které nemusí být pravda. My jsme totiž zvyklí, že umíme porovnat pomocí \leq libovolná dvě (reálná) čísla, ale to rozhodně neplatí o všech uspořádáních. Pokud například náš soubor množin \mathcal{M} obsahuje množiny $\{13, 23\}$ a $\{13, 33\}$, tak je pomocí \subseteq porovnat neumíme, neplatí ani $\{13, 23\} \subseteq \{13, 33\}$, ani $\{13, 33\} \subseteq \{13, 23\}$.

Relace inkluze je tedy velice vhodným příkladem, protože nás upozorňuje, že na některé věci nelze obecně spoléhat. Dokonce je to v jistém smyslu příklad univerzální, co se naučíme na inkluzi, bude fungovat pro všechna uspořádání.

Nejčastěji pracujeme se situací, kdy si množiny k porovnání nevybíráme, ale rovnou bereme všechny možné, neboli máme nějaké universum prvků U a my uvažujeme množinu všech podmnožin $\mathcal{M} = P(U)$. Pokud má to universum alespoň dva různé prvky u, v , tak už vznikají neporovnatelné množiny $\{u\}$ a $\{v\}$.

△

Poznámka: Již jsme viděli několikrát, že obecné pojmy vznikají v matematice často tak, že se vyjde z jednoho konkrétního a užitečného příkladu a zkusi se zachytit jeho podstatné rysy tak, aby vznikla bohatší kategorie zahrnující více objektů, které by se všechny (díky těm vlastnostem) chovaly v zásadě jako ten příklad, který nás inspiroval. Dá se říct, že částečná uspořádání vznikla (také) po inspiraci nerovností \leq a inkluzí \subseteq .

Zejména pro začátečníka se při zobecňování skrývá velké nebezpečí, protože je často v pokušení používat v argumentech věci, které zná z oblíbeného inspiračního příkladu, ale neuvědomí si, že obecně již platit nemusí.

Například v naší současné situaci je spousta věcí, které jsou „jasné“, protože se s nimi čtenář setkává u nerovností celý život, ale ony obecně platit nemusí. V důkazech je tedy třeba být opatrný. Když student napiše „ p platí, protože q “, tak by se měl zamyslet, zda ten argument q opravdu má zaručen pro obecné objekty, se kterými pracuje, nebo to říká jen ze zvyku. V takové situaci se právě hodí dobré znát „méně pěkné“ příklady typu relace \subseteq , aby člověka zastavily v rozletu, pokud myšlenky míří chybným směrem.

△

Příklad 4b.d: Uvažujme relaci $|$ na množině \mathbb{N} danou předpisem $a | b$ právě tehdy, když a dělí b , tedy pokud existuje $k \in \mathbb{Z}$ takové, že $b = k \cdot a$. Například $2 | 6, 3 | 6$, ale neplatí $4 | 6$. O této relaci máme celou další kapitolu, kde dokážeme, že je to částečné uspořádání, viz Věta 6a.5.

I zde existují dvojice čísel, která nejsou porovnatelná, například neplatí ani $4 | 6$, ani $6 | 4$, je to tedy další velice dobrý příklad. Většinou se budeme kvůli stručnosti omezovat jen na nějakou menší podmnožinu přirozených čísel.

Poznamenejme, že relace dělitelnosti již není uspořádání, pokud připustíme i záporná celá čísla, například na množině $\mathbb{N} \cup \{-13\}$. Pak totiž platí $13 | (-13)$ a $(-13) | 13$, ale nemáme $-13 = 13$, tudíž relace $|$ už není na této množině antisymetrická.

△

Zmínili jsme, že relaci dělitelnosti budeme často používat jen na podmožinách \mathbb{N} . Tím se jako obvykle dostáváme k otázce, co s uspořádáním dělají operace.

! Fakt 4b.1.

Nechť (A, \preceq) je částečně uspořádaná množina. Pak restrikce \preceq na libovolnou podmnožinu množiny A je zase částečné uspořádání.

Zde je přechod k podmnožině nejen příjemným způsobem vytváření příkladů, ale dokonce užitečným teoretickým nástrojem, zejména v důkazech.

Další tvrzení vyplýne okamžitě z Faktu 3c.3. Připomeňme, že $a \preceq^{-1} b$ právě tehdy, když $b \preceq a$.

! Fakt 4b.2.

Jestliže je (A, \preceq) částečně uspořádaná množina, pak je i (A, \preceq^{-1}) částečně uspořádaná množina.

Inverzní relace má u uspořádání významnější roli než obvykle, ostatně čtenář zná spřízněnost nerovností \leq a \geq či vlastnosti býti podmnožinou a býti nadmnožinou. I většina pojmu, které dále probereme, si velice dobře rozumí s přechodem k \preceq^{-1} .

Díky své užitečnosti má inverzní relace u uspořádání zvláštní název. Je-li dána částečně uspořádaná množina (A, \preceq) , pak se (A, \preceq^{-1}) říká **duální uspořádání (dual order)**. Zde to nebudeme používat, protože nepůjdeme do teorie tak hluboko, aby se to vyplatilo, už tak je tu dost názvů.

Čtenář by si jako cvičení měl rozmyslet, co se stane, když se dvě uspořádání spojí pomocí operací (množinových či skládání). Necháváme to jako cvičení 4b.11, odpovědi jsou snadné po přečtení Faktů 3c.2 a 3c.4 a diskuse kolem.

V této souvislosti je zajímavé si rozmyslet, jaká je interpretace průniku dvou uspořádání. Máme dva způsoby porovnávání prvků z A a my se to rozhodneme hrát na jistotu, prohlásíme, že prvek b je „opravdu větší“ než a , jestliže zvítězí v obou výchozích porovnáních, jinak raději neřekneme nic.

Existuje ještě jedno široce používané srovnání.

Příklad 4b.e: Uvažujme relaci $<$ na množině \mathbb{R} . V příkladě 3b.c jsme ukázali, že tato relace je antisymetrická a tranzitivní, ale není reflexivní. Nejde tedy o částečné uspořádání. Tato relace je ovšem antireflexivní a asymetrická, viz část 3c.8. Další vlastnosti.

△

Tato bezesporu užitečná relace se tedy nevejde do skupiny, kterou zde zkoumáme, ale je inspirací pro skupinu jinou, která je také zajímavá, například zahrnuje vztah předek-potomek.

Definice 4b.3.

Uvažujme relaci R na množině A . Řekneme, že R je **ostré uspořádání (strict ordering)**, jestliže je antireflexivní a tranzitivní.

Snadno se ukáže, že taková relace už je i asymetrická, viz cvičení 3c.12, tudíž i antisymetrická.

Víme naopak, že asymetrické relace jsou automaticky antireflexivní, takže jsme do definice mohli dát i podmínu, že dotyčná relace je asymetrická a tranzitivní.

Z praktického pohledu jsou tedy ostrá uspořádání relace antireflexivní, asymetrické, antisymetrické a tranzitivní, ale do definice coby správní matematici nechceme dávat zbytečné předpoklady.

Pro ostrá uspořádání se dá udělat teorie s věcmi obdobnými téma, které budeme pro částečná uspořádání dělat v této a příští kapitole, ale je to spíš specializovanější oblast a zde se jí nebudeme zabývat, tento pojem je pro nás pomocný. Přijde vhod při zkoumání vztahu mezi relacemi \leq a $<$, protože víme, že jsou úzce svázány, dokážeme snadno vyjádřit jednu pomocí druhé. My si teď ukážeme, že toto lze udělat i obecně.

Definice.

Nechť \preceq je částečné uspořádání na množině A . Definujeme relaci \prec na A takto: Pro $a, b \in A$ platí $a \prec b$ právě tehdy, když $a \preceq b$ ale $a \neq b$.

Této relaci budeme říkat **odvozená relace (derived relation)**.

Pro dvojici $a \prec b$ říkáme, že a je **předchůdce (predecessor)** prvku b a b je **následník (successor)** prvku a .

Potvrďme si, že se tato obecná situace opravdu chová povědomým způsobem.

Lemma 4b.4.

Nechť \preceq je částečné uspořádání na množině A a \prec je odvozená relace.

- (i) Nechť $a, b \in A$. Jestliže $a \prec b$, pak $a \preceq b$.
- (ii) Nechť $a, b \in A$. Nemůže platit najednou $a \preceq b$ a $b \prec a$.
- (iii) Nechť $a, b, c \in A$. Jestliže $(a \preceq b \text{ a } b \prec c)$ nebo $(a \prec b \text{ a } b \preceq c)$, pak nutně $a \prec c$.

Důkaz (rutinní): (i): Toto je jasné z definice \prec .

(ii): Kdyby platilo $a \preceq b$ a $b \prec a$, tak z toho druhého máme i $b \preceq a$. Z antisymmetrie \preceq pak $a = b$, to je ale ve sporu s předpokladem $b \prec a$.

(iii): Předpokládejme, že $a \preceq b$ a $b \prec c$. Pak podle (i) také $b \preceq c$, máme tedy řetězec $a \preceq b \preceq c$ a podle tranzitivity \preceq také $a \preceq c$. Zbývá ukázat, že nemůže nastat $a = c$.

To dokážeme sporem. Kdyby $a = c$, pak se z $a \preceq b$ stane $c \preceq b$, což je podle (ii) ve sporu s $b \prec c$.

Důkaz pro případ $a \prec b$ a $b \preceq c$ je samozřejmě v zásadě stejný a přenecháme jej čtenáři jako cvičení. □

Takovýchto lemmátek bychom mohli vymyslet spoustu, my jsme vybrali jedno, které se nám ještě bude silně hodit. Odvozená relace \prec není nějak zvlášť důležitá z hlediska teorie, ale ušetří nám spoustu práce při psaní důkazů, podobně jako nám nerovnost $<$ usnadňuje život, ačkoliv bychom si dokázali vystačit jen s relací \leq . Čímž se dostáváme k tomu, jak jsou obecně svázána částečná a ostrá uspořádání. Již jsme to nakousli ve Faktu 3c.10, teď to uděláme pořádně.

Věta 4b.5.

(i) Nechť (A, \preceq) je částečně uspořádaná množina. Pak je odvozená relace \prec antireflexivní, asymetrická a tranzitivní (je to ostré uspořádání).

(ii) Uvažujme relaci R na množině A , která je asymetrická a tranzitivní. Definujme relaci \preceq na A předpisem $a \preceq b$ právě tehdy, když aRb nebo $a = b$. Pak je (A, \preceq) částečně uspořádaná množina.

Jestliže byla navíc R antireflexivní, pak je relace \prec odvozená od \preceq zase rovna R .

Důkaz (poučný): (i): Už definice \prec vylučuje případ $a \prec a$ pro jakékoli $a \in A$. Je to tedy relace antireflexivní.

Asymetrie: Dokážeme, že pro žádné dva prvky $a, b \in A$ nemůžeme mít zároveň $a \prec b$ a $b \prec a$. Sporem: Kdyby tomu tak bylo, tak také $a \preceq b$ a $b \preceq a$, pak podle antisymmetrie $a = b$, což je ve sporu s $a \prec b$.

Připomeňme, že asymetrie je silnější než antisymetrie, takže \prec je samozřejmě také antisymetrická.

Tranzitivita: Jestliže pro $a, b, c \in A$ platí $a \prec b$ a $b \prec c$, pak také $a \preceq b$ a $b \preceq c$ a podle Lemmatu 4b.4 (iii) platí $a \prec c$.

(ii): Reflexivita: jasná z definice, $a \preceq a$ pro všechna $a \in A$.

Antisymetrie: Jestliže $a, b \in A$ splňují $a \preceq b$ a $b \preceq a$, pak jsou dvě možnosti, buď $a = b$ nebo $a \neq b$. V prvním případě jsme hotovi, o druhém ukážeme, že nemůže nastat. Kdyby totiž platilo $a \neq b$, pak podle definice $a \preceq b$ musí platit $a \prec b$, podobně také $b \prec a$, což je ale ve sporu s předpokladem, že \prec je asymetrická.

Tranzitivita: Nechť $a, b, c \in A$ splňují $a \preceq b$ a $b \preceq c$, pak jsou tři možnosti:

Jestliže $a = b$, pak $b \preceq c$ znamená $a \preceq c$ a jsme hotovi. Jestliže $b = c$, pak $a \preceq b$ znamená $a \preceq c$ a zase jsme hotovi. (Všimněte si, že tyto dvě možnosti se navzájem nevylučují, mohou platit obě najednou). Zbývá možnost, že $a \neq b$ a $b \neq c$. Pak ovšem podle definice \preceq podmínky $a \preceq b$, $b \preceq c$ znamenají $a \prec b$ a $b \prec c$, což podle tranzitivity \prec dává $a \prec c$, tedy i $a \preceq c$.

Když z této \preceq odvodíme \prec , tak vlastně máme (v množinovém zápisu) $\prec = \preceq - \Delta(A)$. Zároveň $\preceq = R \cup \Delta(A)$. Rovnost $(R \cup \Delta(A)) - \Delta(A) = R$ platí v případě, že $R \cap \Delta(A) = \emptyset$, tedy v případě, že R je antireflexivní. \square

Bod (ii) je pěkně vidět v akci ve cvičení 4b.5 (ii), viz také 4b.2 a další.

! 4b.6 Hasseův diagram

Vlastnosti částečných uspořádání nám umožňují zásadním způsobem redukovat jejich graf. Uvažujme tedy konečnou množinu A a částečné uspořádání \preceq na ní.

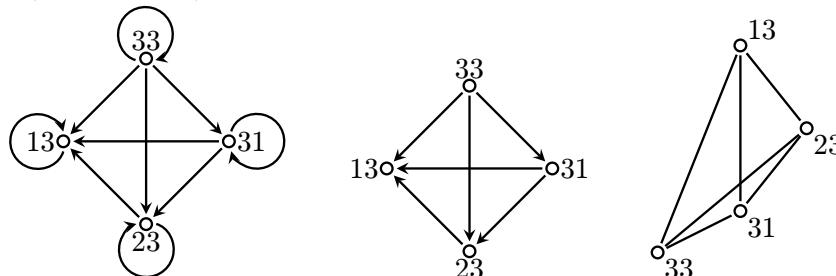
Prvním krokem ke zjednodušení grafu je, že podobně jako u ekvivalencí nemusíme kreslit smyčky, stejně víme, že jsou všude. Mimochodem, to, co zbyde, je přesně graf odvozené relace \prec .

Víme dále, že mezi dvěma různými body vede nejvíše jedna šipka, ale její orientace je zásadní, takže to nelze jen tak vynechat. Orientaci ale lze něčím nahradit, jmenovitě pozici v prostoru. To je novinka, zatím jsme se při kreslení grafů nezabývali otázkou, jak konkrétně jsou jednotlivé tečky reprezentující prvky z A rozmištěny.

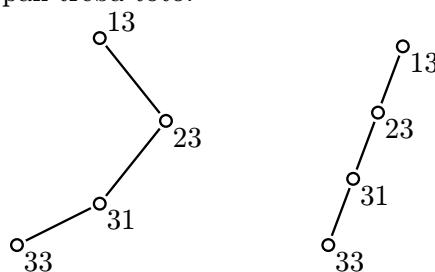
Základní myšlenka reprezentace, která se jmeneje **Hasseův diagram**, spočívá právě v chytrém uspořádání vrcholů grafu tak, aby šipky vždy směrovaly směrem vzhůru (ne nutně svisle, třeba šikmo, ale vzhůru). Tím intuitivně naznačíme, že prvky, které jsou na obrázku výše, jsou větší ve smyslu \preceq . Pak můžeme šipky vynechat a kreslit jen obyčejné spojnice, protože směr je již dán úmluvou z geometrie. Zkusíme si tyto první kroky na jednoduchém příkladě.

Uvažujme částečně uspořádanou množinu $(\{13, 23, 31, 33\}, \geq)$.

Nakreslíme nějaký graf, pak vynecháme smyčky a zkusíme graf přeorganizovat tak, aby zbylé šipky ukazovaly nahoru. Nakonec vynecháme šipky, „největší“ prvek je nahoře, ostatní jdou postupně dolů. Zde je třeba dát pozor na význam zadání, pracujeme s relací \geq , nikoliv s \leq . Takže třeba $23 \geq 13$ neboli $23R13$, tedy šipka vede od 23 směrem k 13, ne naopak (13 je „větší“).



Jak vidíte, dostali jsme houštinku. Hodilo by se další zjednodušení, ke kterému použijeme tranzitivitu. Díky ní víme, že když z prvku do prvku vede dlouhá cesta, tak musí vést i kratší, takže když tu kratší nenamalujeme, tak se v zásadě nic nestane, stejně vidíme, že se do cíle nakonec dostaneme. Jinými slovy, neztrácíme informaci, pokud vynecháme kratší spojnice tam, kde již máme spojení. Chceme-li u grafu částečného uspořádání vědět, zda vede šipka z a do b (tedy zda $a \preceq b$), stačí se podívat, jestli z a do b nevede cesta. Posledním krokem našeho zjednodušování je tedy vynechání těch hran, které nejsou potřeba. Kdykoliv vidíme na grafu cestu a její zkratku, tak vymažeme tu zkratku. Dostaváme pak třeba toto.



Ukázali jsme dvě možnosti, abychom naznačili, že geometrických uspořádání Hasseova diagramu může být více. Tady jsme ještě relativně omezováni, protože máme nadprůměrně dobře se chovající relaci, u těch zajímavějších už se může stát, že máme opravdu dost prostoru pro fantasií.

Právě předvedený postup ukazuje podstatu Hasseova diagramu, ale není perspektivní, protože vyznat se v houštině původního grafu je náročná práce a rovněž hledání správné geometrické konfigurace tak, že se spoleháme na inspiraci, není zrovna nejlepší nápad. Trochu se tomu dá pomoci tím, že nejprve odstraníme šipky z tranzitivity a teprve poté uspořádáme body geometricky, ale existuje postup, který nám umožní nakreslit Hasseův rovnou ze zadání.

S Algoritmus 4b.7. pro vytváření Hasseova diagramu částečného uspořádání (A, \preceq) pro konečnou (malou) množinu A .

1. Vypište si seznam všech dvojic $x \prec y$ přidružené ostré relace.

2. Hledejte prvky $a \in A$, které se v odvozené relaci nikdy nevyskytují ve dvojcích napravo, tedy v pozici $x \prec a$. (Tj. hledejte body, do kterých nevhází žádná šipka s výjimkou té reflexivní smyčky.)

Tyto prvky zakreslete do vznikajícího diagramu jako první řádek. Pak je škrtněte z množiny A a škrtněte také ze seznamu dvojcí všechny, ve kterých se dotyčné prvky vyskytují.

3. Pokud už v množině A nic nezbylo, je diagram hotov.

Jinak projděte částečně odmazaný seznam dvojcí a hledejte prvky a ze zmenšené množiny A , které nikdy nejsou napravo jako $x \prec a$. Tyto prvky zakreslete do vznikajícího diagramu jako druhý řádek počítáno zdola a vyškrtněte je z množiny A .

Nakreslete spojnice nahoru z vrcholů prvního řádku do vrcholů druhého řádku tam, kde jsou spolu v relaci (tedy tam, kde existují jako dvojice v seznamu dvojcí $x \prec y$, neboli kde by spolu byly spojeny v běžném grafu relace). Tyto dvojice pak vyškrtněte ze seznamu $x \prec y$.

Dostáváte zkrácený seznam prvků z A , které ještě nejsou ve vznikajícím diagramu, a zkrácený seznam dvojcí $x \prec y$, ve kterých se nevyskytují žádné prvky z prvních dvou řádků diagramu.

4. Pokud už v A nic nezbylo, je diagram hotov.

Jinak projděte částečně odmazaný seznam dvojcí a hledejte prvky a ze zmenšené množiny A , které nikdy nejsou napravo jako $x \prec a$. Tyto prvky zakreslete do vznikajícího diagramu jako nový řádek nahoru a vyškrtněte je z množiny A .

Nakreslete spojnice z bodů v dolních řádcích do bodů v novém horním řádku tam, kde v relaci existují jako dvojice, ale pouze v tom případě, že tuto cestu nelze ukutečnit pomocí již zakreslených spojnic a to čistě směrem vzhůru (pokud se někam dostanete tak, že v průběhu cesty musíte i dolů, pak se to nepočítá). Zde je důležité postupovat shora dolů, tedy nejprve zakreslovat spojnice mezi horním řádkem a tím bezpostředně pod ním, pak mezi horním a tím o dva níže, až nakonec se zkoumají možné spojnice mezi horním a dolním řádkem.

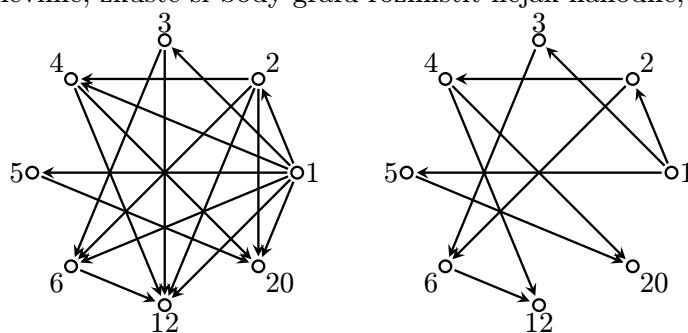
Z onoho postupně se zkracujícího seznamu dvojcí $x \prec y$ vymažte všechny, ve kterých se vyskytují prvky z právě přidané horní řady diagramu.

Jděte znovu na bod 4.

△

Je také možné kreslit diagram shora, tedy vždy hledat body, do kterých šipky pouze vchází. To, že algoritmus kreslením zdola či shora funguje, je založeno na pojmu minimum a maximum, viz poznámka 4c.4.

Příklad 4b.f: Vytvoříme Hasseův diagram pro množinu $A = \{1, 2, 3, 4, 5, 6, 12, 20\}$ uspořádanou dělitelností. Nejprve si sami udělejte intuitivní postup, kdy nejprve z grafu bez smyček odebíráme přepony v trojúhelnících a pak uhádneme tvar, kdy všechny šipky vedou vzhůru. Abychom si správně nasimulovali situaci, kdy pracujeme s relací, o které dopředu moc nevíme, zkuste si body grafu rozmištít nějak náhodně, třeba takto:



Překreslete si ten obrázek vpravo, ale bez označení vrcholů čísly, aby vám to nenapovídalo, a zkuste graf překroutit tak, aby všechny šipky šly směrem vzhůru. Jde to najít, ale asi to není nejlepší metoda.

Odvodíme správný tvar algoritmem. Nejprve si vypíšeme odpovídající ostré uspořádání: $1|2, 1|3, 1|4, 1|5, 1|6, 1|12, 1|20, 2|4, 2|6, 2|12, 2|20, 3|6, 3|12, 4|12, 4|20, 5|20, 6|12$.

Hledáme prvky A , které se v seznamu nevyskytují v žádné dvojici vpravo, tedy které už nikdo jiný nedělí. Takový tam je jeden, 1, ten nakreslíme do první (dolní) řady diagramu. Vyškrtneme prvek 1 i všechny dvojice, ve kterých se vyskytuje.

Jsou v množině $A_1 = \{2, 3, 4, 5, 6, 12, 20\}$ čísla, která v tom novém kratším seznamu $2|4, 2|6, 2|12, 2|20, 3|6, 3|12, 4|12, 4|20, 5|20, 6|12$ nikdy nejsou napravo? Vlastně vidíme, že bychom se bez seznamu obešli, prostě hledáme v A_1 čísla, která žádné jiné z A_1 nedělí.

Ano, jsou to čísla 2, 3 a 5. Ty přijdou do diagramu nad 1, a protože 1 dělí všechny tři, nakreslíme k nim spojnice z 1. Odebereme 2, 3, 5 z množiny a odpovídající dvojice ze seznamu, vznikne seznam $6|12$.

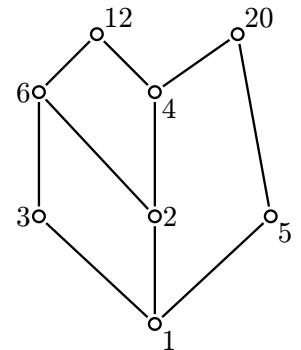
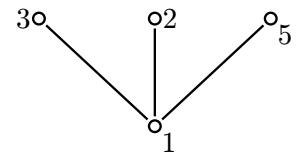
Jsou v množině $A_2 = \{4, 6, 12, 20\}$ čísla, která nikdo jiný z A_2 nedělí? Ano, jsou to čísla 4 a 6, ty přijdou do diagramu nad 2, 3 a 5. Spouštíme spojnice. Ze 4 do předposledního řádku spojujeme jen do 2, protože nejsou dvojice $3|4$ a $5|4$. Ze 6 spustíme spojnice do 2 a 3. Je třeba spustit spojnici i o řádek níž, tedy ze 4, popř. z 6 do jedničky? Ne, tam už cesta vede.

Vymažeme z A_1 prvky 4, 6, ze seznamu odstraníme všechny dvojice zahrnující 4 nebo 6.

Jsou v množině $A_3 = \{12, 20\}$ čísla, která nikdo jiný z A_3 nedělí? Ano, jsou to obě čísla 12 a 20. Mimochodem, najdeme je také jako čísla, která se ve zkráceném seznamu nevyskytují napravo, ale ten je již prázdny, takže opravdu bereme vše.

Čísla 12 a 20 přijdou do diagramu nahoru a doplníme spojnice do předposlední řady tam, kde je dělitelnost. Zamyslíme se nad spouštěním spojnic do druhé řady shora. Máme $3|12$ a $2|12$, ale cesty $3 \rightarrow 12$ i $4 \rightarrow 12$ už existují, tak nespojujeme. Máme také $2|20$ a $5|20$. Cesta z 2 do 20 nahoru už existuje, ale spojení $5 \rightarrow 1 \rightarrow 20$ neplatí, protože nevede čistě vzhůru, taže tuto spojnici je nutné dodělat. Ještě si rozmyslíme možná spojení do spodního řádku (nejsou třeba, už propojeno) a jsme hotovi.

△



Poznámka: Viděli jsme, že nám algoritmus nechával možnost volby, pokud jsme v jednom řádku měli více prvků. Změna jejich pořadí dokáže výrazně ovlivnit celkový vzhled výsledného diagramu. Často jsou také verze diagramu, ke kterým nás náš algoritmus ani zavést neumí, například prvek 5 mohl být klidně o řádek výše nebo dokonce někde mezi. Z čisté matematického pohledu na konkrétní podobě samozřejmě nezáleží, ale pro uživatele bude určitě příjemnější pracovat s grafem přehledným. Protože je vytváření diagramu pomocí algoritmu nudné, je možné si to zpestřit právě zařazením dodatkového kritéria, snahou o co nejhezčí graf. To naříkla známená, že se snažíme o graf, ve kterém se spojnice nekříží. V příkladu se nám to povedlo, ale ne vždy je to možné.

△

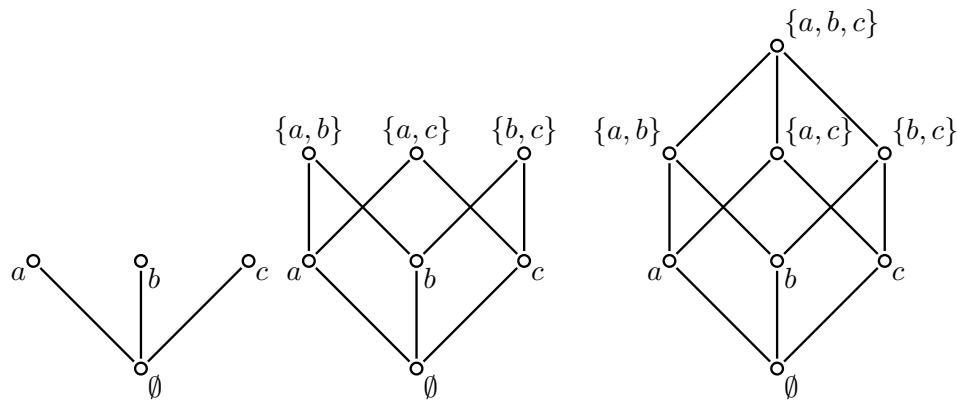
Příklad 4b.g: Vytvoříme Hasseův diagram pro částečně uspořádanou množinu $(P(\{a, b, c\}), \subseteq)$. Použijeme algoritmus růstu zdola. Aby se nám lépe pracovalo, raději si danou množinu vypíšeme, zajímá nás tedy relace inkluze na množině $A = P(\{a, b, c\}) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$.

Zkusíme to bez vypisování všech ostrých uspořádání. Je nějaká množina v A , aby žádná jiná nebyla její podmnožinou? Ano, \emptyset , bude dole.

Jsou v $A_1 = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$ množiny takové, že jiné množiny z A_1 už nejsou jejich podmnožinami? Ano, všechny jednoprvkové. Ty přijdou do druhého řádku a spojíme je s prázdnou množinou.

Jsou v $A_2 = \{\{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$ množiny takové, že jiné množiny z A_2 už nejsou jejich podmnožinami? Ano, všechny dvouprvkové. Ty přijdou do třetího řádku a spojíme je s jednoprvkovými tam, kde je mezi nimi vztah inkluze. S prvním řádkem (prázdnou množinou) není třeba přímo spojovat, protože se od prázdné ke každé dvouprvkové dostaneme již existujícími spojnicemi.

Nakonec přidáme nahoru $\{a, b, c\}$ a spojíme se všemi množinami v předposledním řádku, jiných spojnic již netřeba.

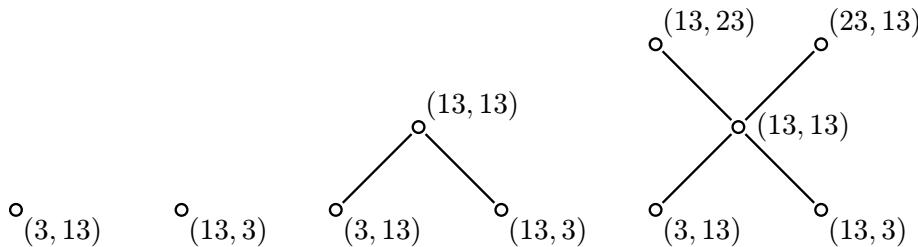


Zrovna u tohoto příkladu nelze vytvořit Hasseův diagram tak, aby se spojnice nekřížily, viz 12c.

△

Jako inspiraci si na stejné množině zaveděte relaci \supseteq , vytvořte diagram a pak si jako cvičení rozmyslete, že obecně Hasseův diagram uspořádání \preceq^{-1} získáme jednoduše tak, že otočíme Hasseův diagram \preceq vzhůru nohama.

Příklad 4b.h: Vytvoříme Hasseův diagram pro množinu $\{(3, 13), (13, 3), (13, 13), (13, 23), (23, 13)\}$ a uspořádání nerovností \leq po složkách (součinové uspořádání, viz 3b.9), tedy $(u, v) \preceq (x, y)$ právě tehdy, když $u \leq x$ a $v \leq y$. Algoritmus dává



△

Hasseův diagram je velice užitečný, dá se použít k rychlému rozpoznávání různých vlastností uspořádání, o kterých se dozvímeme v následující kapitole. Je to i díky tomu, že vztah „relace \implies diagram“ lze v jistém smyslu „obrátit“. Když nakreslíme hromádku bodů a některé z nich spojíme úsečkami tak, aby ve výsledném obrázku nebyly vodorovné spojnice, tak už tento obrázek jednoznačně určuje nějaké částečné uspořádání.

4b.8 Bonus: Covering relation. Zajímavá otázka je, co to vlastně dostaneme, když vyjdeme z nějakého částečného uspořádání a vytvoříme z něj výrazně menší podmnožinu dvojic zakreslenou v diagramu. V této nové relaci je každý prvek spojen jen s „bezprostředními“ sousedy z původní relace, tedy s těmi prvky, ke kterým se v původní relaci dostává jedině přímo, bez mezikroku. Definice vypadá takto:

Definice.

Nechť (A, \preceq) je částečně uspořádaná množina a \prec je příslušná odvozená relace. Definujeme relaci \triangleleft na A předpisem $a \triangleleft b$ jestliže $a \prec b$ a neexistuje $z \in A$ takové, že $a \prec z$ a $z \prec b$.

Relaci \triangleleft říkáme **covering relation** pro relaci \preceq .

Jestliže $a \triangleleft b$, tak říkáme, že prvek b pokrývá prvek a , popřípadě že prvek a je pokryt prvkem b . Alternativní terminologie říká, že prvek a je **bezprostřední předchůdce (immediate predecessor)** prvku b , popř. že prvek b je **bezprostřední následník (immediate successor)** prvku a .

Podívejme se na příklad 4b.f. Prvek 3 má dva následníky, 6 a 12, ale jen jeden z nich je bezprostřední, jmenovitě 6.

Když sestrojíme Hasseův diagram nějakého uspořádání \preceq a pro pořádek přidáme ke spojnicím šipky, ať je formálně zaznačena i orientace vzhůru, tak dostáváme právě graf relace \triangleleft . Bez důkazu jsme tvrdili, že z diagramu dokážeme odvodit zpětně původní relaci, což vlastně říká, že z covering relation \triangleleft zase dokážeme nějakým postupem dostat zpět \preceq . Trochu to připomíná situaci s odvozeným uspořádáním, kdy nám fungovalo kolečko $\preceq \mapsto \prec \mapsto \triangleleft$, ale není to úplně stejně.

Kolečko $\preceq \mapsto \prec \mapsto \triangleleft$ totiž funguje jen někdy. Zajímavé je, že se nezadrhne krok $\triangleleft \mapsto \preceq$, ten je spolehlivý, problém může nastat již ve fázi $\triangleleft \mapsto \prec$.

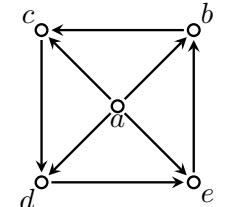
Pokud je množina A konečná, pak lze sestrojit Hasseův diagram (viz poznámka 4c.4) a tudíž vznikne i \triangleleft , tady to funguje. Ale u nekonečných množin se již na to nedá spoléhat, stačí se podívat na relaci \leq . Když ji uvažujeme na

\mathbb{N} či na \mathbb{Z} , pak má každý prvek n svého bezprostředního následníka $n+1$ i svého bezprostředního předchůdce $n-1$ (zde s výjimkou $n=1$ při práci v \mathbb{N}). To, že nějaký prvek nemá bezprostředního předchůdce či následníka, nijak nevadí, ostatně jsme to zažili již v příkladě 4b.f, kde třeba prvek 12 nemá bezprostředního následníka. Podstatné je, že vznikla relace \triangleleft dostatečně bohatá na to, aby v sobě nesla informaci o celé původní relaci.

Ted' se ale podívějme na relaci \leq na \mathbb{Q} , třeba jak to vypadá s prvkem 0. Ten má spoustu následníků i předchůdců, ale přesto nemá bezprostředního následníka, protože pro jakéhokoliv kandidáta, tedy kladné r , vždycky najdeme mezíkrok, $0 < \frac{r}{2} < r$, podobně dopadnou předchůdci. Nula přitom není ničím speciální, víme, že pro libovolný zlomek je otázka „jaký je bezprostředně větší zlomek?“ nezodpověditelná. Pro (\mathbb{Q}, \leq) je tedy covering relation prázdná! Pak z ní samozřejmě nejde zpětně odvodit původní relaci \leq .

Mimochodem, podobná definice bezprostředního předchůdce/následníka se dá udělat pro libovolnou relaci R , ale tam už jsou vůbec problémy s existencí, a to i pro konečné množiny, mimo jiné proto, že na rozdíl od uspořádání jsou v obecnějších relacích možné cykly.

Prvek a nemá bezprostředního následníka, protože ke všem prvkům b, c, d, e se dostane jak přímo, tak oklikou přes mezíkrok. Pro obecné relace se proto covering relation nezavádí.



4b.9 Poznámka (pokročilá): Představme si relaci R na množině A , která je reflexivní a tranzitivní, ale není mrška antisymetrická (takovým se říká **kvaziuspořádání (quasi-ordering)**). Typickým příkladem jsou relace dané $|x| \leq |y|$ pro čísla či $|A| \leq |B|$ pro množiny, takže jde o docela běžný případ. Dá se i zde nějak převést situace na uspořádání? Ano.

Základním trikem je zadefinovat relaci S předpisem aSb právě tehdy, jestliže aRb a bRa (rozmyslete si, že $S = R \cap R^{-1}$). V našich dvou příkladech vzniknou relace $|x| = |y|$ a $|A| = |B|$. Pro reflexivní a tranzitivní R je pak vždy takto vytvořená S ekvivalence (viz cvičení 3b.14) a pomocí ní začneme předstírat, že prvky, které nám kazí antisimetrii, jsou vlastně vždy jedna věc, čímž jakoby antisymetrie začne fungovat.

Matematicky řečeno, podíváme se na příslušný rozklad A podle S a považujeme každou třídu za jeden objekt, tomuto jsme říkali faktorová množina A/S . Na ní zavedeme nové uspořádání \preceq , které bude přirozeným způsobem odvozeno od původního uspořádání R .

Formálně: Definujeme $[a]_S \preceq [b]_S$ právě tehdy, když aRb .

Hlavním problémem takové definice je, že výsledek závisí na volbě zástupců tříd. Co kdybychom vybrali jiné zástupce? Nechť $c \in [a]_S$ a $d \in [b]_S$, potřebujeme ukázat, že také cRd . Použijeme tranzitivitu. Pro začátek máme aRb . Protože $c \in [a]_S$, je aSc , což podle definice znamená cRa . Podobně pro d odvodíme bRd . Získali jsme tak řetězec $cRaRbRd$, máme tedy opravdu cRd .

Dokázali jsme tím, že tato definice má smysl, tedy že dává stejně výsledky pro libovolné volby zástupců tříd. Máme proto relaci mezi třídami a ukážeme, že je to částečné uspořádání.

Reflexivita: R je reflexivní, proto aRa a tedy i $[a]_S \preceq [a]_S$.

Antisimetrie: Předpokládejme, že $[a]_S \preceq [b]_S$ a $[b]_S \preceq [a]_S$. Podle definice \preceq tedy aRb a bRa . Podle definice S tedy aSb a dostáváme $[a]_S = [b]_S$, přesně jak jsme to potřebovali.

Tranzitivita: Předpokládejme, že $[a]_S \preceq [b]_S$ a $[b]_S \preceq [c]_S$. Podle definice \preceq tedy aRb a bRc . Podle tranzitivity R dostáváme aRc a proto $[a]_S \preceq [c]_S$.

Podobný trik, kdy se s třídami pracuje jako s objekty, se v matematice používá relativně často, my jej tu uvidíme v kapitole o počítání modulo a pak v kapitole 8d neboli Bonusu o racionalních číslech, kde dokonce zavádíme uspořádání v zásadě zde popsaným způsobem.

△

Cvičení

Cvičení 4b.1 (rutinní): Které z následujících relací jsou částečná uspořádání na $\{1, 2, 3, 4\}$? Pro každé uspořádání nakreslete Hasseův diagram.

- (i) $\{(1, 1), (2, 2), (3, 3), (4, 4), (1, 2), (1, 3), (2, 3), (3, 4)\}$;
- (ii) $\{(1, 1), (2, 2), (3, 3), (4, 4)\}$;
- (iii) $\{(1, 1), (2, 2), (4, 4), (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$;
- (iv) $\{(1, 1), (2, 2), (3, 3), (4, 4), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$;
- (v) $\{(1, 1), (2, 2), (3, 3), (4, 4), (1, 2), (2, 3), (1, 3), (3, 1), (3, 2)\}$.

Cvičení 4b.2 (rutinní): Nechť A je množina všech lidí. Která z následujících relací je částečné uspořádání?

- | | |
|--|--------------------------------|
| (i) a a b mají společného přítele; | (iii) a je vyšší než b ; |
| (ii) a je předek b nebo a je b ; | (iv) a neváží více než b . |

Cvičení 4b.3 (rutinní): Rozhodněte, které z následujících relací na \mathbb{Z}^2 jsou uspořádání.

- (i) $(u, v)R(x, y)$ jestliže $u \leq x$ a $v = y$;
 (ii) $(u, v)R(x, y)$ jestliže $u \leq x$ a $v < y$;
 (iii) $(u, v)R(x, y)$ jestliže $u \leq x$ a $v \geq y$.

Cvičení 4b.4 (rutinní): Pro číslo $n \in \mathbb{N}$ definujme $m(n)$ jako největší cifru použitou při desítkovém zápisu n , například $m(13756) = 7$. Rozhodněte, které z následujících relací jsou částečná uspořádání na \mathbb{N} :

- (i) xRy jestliže $m(x) \leq m(y)$;
 (ii) xRy jestliže $m(x) < m(y)$;
 (iii) xRy jestliže $m(x) < m(y)$ nebo $x = y$.

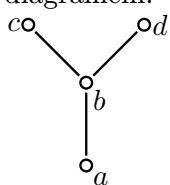
Cvičení 4b.5 (rutinní, poučné): Ukažte přímým vyšetřením vlastnosti, že následujících relace jsou uspořádání (viz Věta 4b.5 (ii)).

- (i) Relace \preceq na $M_{2 \times 2}$, množině reálných matic 2×2 : $A \preceq B$ jestliže $|A| < |B|$ (determinanty) nebo $A = B$.
(ii) Relace \preceq na \mathbb{N} : $x \preceq y$ jestliže je počet jedniček v binárním zápisu x menší než počet jedniček v binárním zápisu y nebo $x = y$.
(iii) Relace \preceq na P , množině všech reálných polynomů: $p \preceq q$ jestliže je stupeň p menší než stupeň q nebo $p = q$.
Nápověda: U (ii) si zavedete vhodné značení pro počet jedniček.

Cvičení 4b.6 (rutinní): Určete, zda relace určené následujícími maticemi jsou částečná uspořádání:

(i) $M = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$; (iii) $M = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$; (v) $M = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$.
(ii) $M = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$; (iv) $M = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$.

Cvičení 4b.7 (rutinní): Napište výčtem dvojic prvků, která relace uspořádání je dána následujícím Hasseovým diagramem.



Cvičení 4b.8 (rutinní): Nakreslete Hasseův diagram

- (i) pro $(\{13, 23, 31, 33, 43\}, \geq)$;
(ii) pro množinu množin $\mathcal{A} = \{\emptyset, \{1\}, \{2\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\}$ uspořádanou relací býti podmnožinou.

Cvičení 4b.9 (rutinní): Nakreslete Hasseův diagram pro $(A, |)$ (relace dělitelnosti), kde

(i) $A = \{1, 2, 3, 4, 5, 6, 7, 8\}$; (iii) $A = \{2, 3, 4, 5, 6, 30, 60\}$; (v) $A = \{1, 2, 4, 8, 16, 32, 64\}$;
 (ii) $A = \{1, 2, 3, 5, 11, 13\}$; (iv) $A = \{1, 2, 3, 6, 12, 24\}$; (vi) $A = \{2, 4, 6, 12, 24, 36\}$.

Viz také cvičení 4c.3.

Cvičení 4b.10 (poučné): Dokažte, že jestliže je relace R na množině A částečné uspořádání a ekvivalence zároveň, pak nutně $R \subseteq \Delta(A)$.

Cvičení 4b.11 (poučné):

Nechť \preceq_1, \preceq_2 jsou částečná uspořádání na téže množině A . Dokažte, že pak

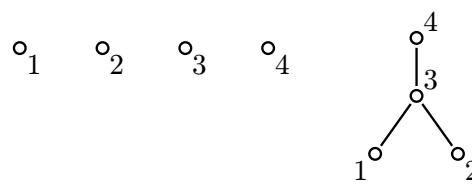
- (i) $\preceq_1 \cap \preceq_2$ je také částečné uspořádání na A ;
 (ii) $\preceq_1 \cup \preceq_2$ nemusí být částečné uspořádání na A ;
 (iii) $\preceq_1 - \preceq_2$ nikdy není částečné uspořádání na A ;

(iv) $\preceq_1 \circ \preceq_2$ nemusí být částečné uspořádání na A ;
 (v) $\preceq_1 \circ \preceq_1 = \preceq_1^2$ je částečné uspořádání na A .

Řešení:

4b.1:

- (i): není tranzitivní, viz $(2, 3)$ a $(3, 4)$.
 (ii): uspořádání (i ekvivalence).
 (iii): není reflexivní, viz $(3, 3)$.
 (iv): uspořádání.
 (v): není antisymetrická, viz $(1, 3)$ a $(3, 1)$.



4b.2: (i): není A,T; (ii): uspořádání; (iii): není R; (iv): není A;

4b.3: (i): upořádání; (ii): není R; (iii): uspořádání.

4b.4: (i): Je R, T , není A . (ii): Je A, T , není R . (iii) Je uspořádání.

4b.5: (i): R : přímo z definice. A : Nechť $A \preceq B \wedge B \preceq A$. Dvě možnosti. Kdyby $A \neq B$, pak z definice dostáváme $|A| < |B| \wedge |B| < |A|$, spor. Tudíž je to varianta $A = B$. T : Nechť $A \preceq B \wedge B \preceq C$. Čtyři možnosti:

a) V obou případech vzniklo \preceq z první podmínky. Pak $|A| < |B| \wedge |B| < |C|$, proto $|A| < |C|$ a $A \preceq C$.

b) Vzniklo to jako $|A| < |B|$ a $B = C$. Pak $|A| < |C|$ a $A \preceq C$.

c) Varianta $A = B$ a $|B| < |C|$ dává také $A \preceq C$.

d) poslední možnost je, že to $A \preceq B \wedge B \preceq C$ vzniklo jako $A = B$ a $B = C$, pak $A = C$ a tedy $A \preceq C$.

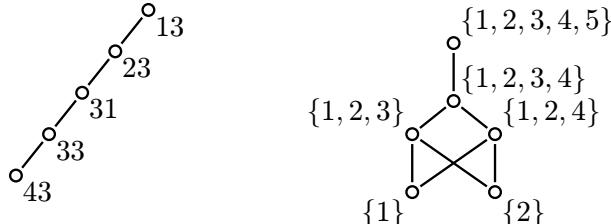
Takže vždy $A \preceq C$ a tranzitivita je prokázána.

(ii) a (iii) se dělají obdobně, jen místo determinantu se pracuje s jinou funkcí.

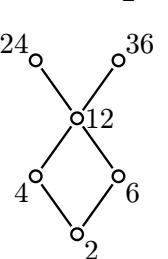
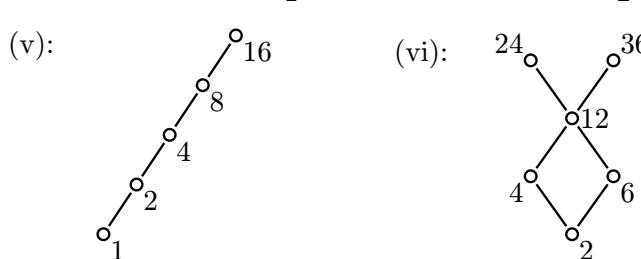
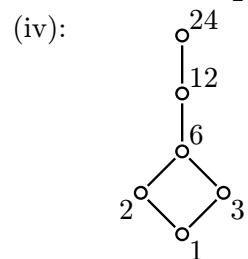
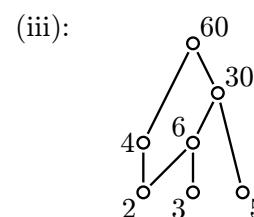
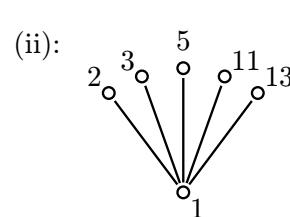
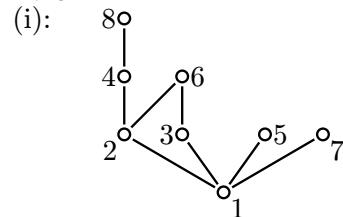
4b.6: Nejsnáze poznáme reflexivitu, matice musí mít 1 všude na diagonále. Tím jsme vyloučili (ii). Antisimetrii poznáme tak, že matice nesmí mít 1 zároveň na dvou polích symetrických podle diagonály. Projedeme matice a vyloučíme také (iii) a (v). Zbývají příklady (i) a (iv), kde je třeba ověřit tranzitivitu. Jedna možnost je použít Booleanovský součin. Druhá možnost je vypsat si z matice všechny nediagonální jedničky jako dvojice v relaci a zkoumat tranzitivitu na nich. U (iv) se takto najdou navazující dvojice $(1, 2)$ a $(2, 3)$, ke kterým chybí $(1, 3)$, dotyčná relace tedy není tranzitivní. V případě (i) se tranzitivita potvrdí a je to částečné uspořádání.

4b.7: $\{(a, a), (b, b), (c, c), (d, d), (a, b), (a, c), (a, d), (b, c), (b, d)\}$.

4b.8:



4b.9:



4b.10: Viz cvičení 3b.11.

4b.11: (i) viz Fakt 3c.2.

(ii) Diskuse po Faktu 3c.2 ukazuje, že sjednocením se nemusí zachovat tranzitivita ani antisymmetrie, protipříklady tam jsou, jeden snadný na $A = \{1, 2, 3\}$: $\preceq_1 = \{(1, 1), (2, 2), (3, 3), (1, 2)\}$ a $\preceq_2 = \{(1, 1), (2, 2), (3, 3), (2, 1), (2, 3)\}$ jsou obě uspořádání, ale sjednocení $\preceq_1 \cup \preceq_2 = \{(1, 1), (2, 2), (3, 3), (1, 2), (2, 1), (2, 3)\}$ není ani antisymetrické, ani tranzitivní.

(iii) Protože jsou dvojice (a, a) v obou relacích (reflexivita), nemůže iž z principu relace $\preceq_1 - \preceq_2$ obsahovat žádnou takovou dvojici, tudíž nemůže být reflexivní.

(iv) Stačí upravit příklad po Faktu 3c.4 na tranzitivitu.

$\preceq_1 = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (1, 2), (3, 4)\}$ a $\preceq_2 = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (2, 3), (4, 5)\}$ jsou relace uspořádání, ale složením dostaneme $R = \preceq_1 \circ \preceq_2 = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (1, 3), (3, 5)\}$, kde lze vytvořit řetízek $1R3R5$, který na jeden krok nezvládneme, dvojice $(1, 5)$ v té složenině není. Relace proto není tranzitivní, tudíž ani uspořádání.

Pokud (kromě dvojic $a \preceq a$) dáme do první relace $(1, 2), (2, 5)$ a do druhé $(5, 3), (3, 1)$, tak to opět budou uspořádání, ale složení bude obsahovat dvojice $(1, 5), (5, 1)$, címž se poruší pro změnu antisymmetrie.

(v) U skládání to vypadá na první pohled nevesele, protože antisymmetrie se dá mocninanou pokazit, viz diskuse po Faktu 3c.4. Zde ale máme i vlastnosti jiné a ty to zachrání, protože podle Věty 3b.6 pro každé částečné uspořádání máme $R^2 = R$.

4c. Minima, nejmenší prvky a podobně, dobré uspořádání

V této kapitole se budeme v zásadě zajímat o to, jak u částečně uspořádané množiny vypadají její „horní konec“ a „dolní konec“.

!

Definice.

Nechť (A, \preceq) je částečně uspořádaná množina a \prec odpovídající odvozená relace. Nechť M je neprázdná podmnožina A .

Řekneme, že prvek $m \in A$ je **nejmenší prvek** množiny M , jestliže $m \in M$ a pro všechna $x \in M$ platí $m \preceq x$.

Řekneme, že prvek $m \in A$ je **největší prvek** množiny M , jestliže $m \in M$ a pro všechna $x \in M$ platí $x \preceq m$.

Řekneme, že prvek $m \in A$ je **minimální prvek** množiny M , jestliže $m \in M$ a neexistuje $x \in M$: $x \prec m$.

Značíme to $m = \min(M)$.

Řekneme, že prvek $m \in A$ je **maximální prvek** množiny M , jestliže $m \in M$ a neexistuje $x \in M$: $m \prec x$.

Značíme to $m = \max(M)$.

Let (A, \preceq) be a poset. Let M be a non-empty subset of A .

We say that an element $m \in A$ is a **least element** of the set M , if $m \in M$ and $m \preceq x$ for all $x \in M$.

We say that an element $m \in A$ is a **greatest element** of the set M , if $m \in M$ and $x \preceq m$ for all $x \in M$.

We say that an element $m \in A$ is **minimal** in the set M (or a minimum of M), if $m \in M$ and there is no $x \in M$ such that $x \prec m$. We denote it $m = \min(M)$.

We say that an element $m \in A$ is **maximal** in the set M (or a maximum of M), if $m \in M$ and there is no $x \in M$ such that $m \prec x$. We denote it $m = \max(M)$.

Řečeno lidově, nejmenší prvek množiny je takový, že všichni ostatní jsou „nad ním nebo rovny“, zatímco minimální prvek je takový, že nikdo není „pod ním“. Evidentně nejde o totéž, jinak bychom neměli dva pojmy. Brzy ten rozdíl uvidíme na vlastní oči.

Příklad 4c.a: Uvažujme (\mathbb{N}, \leq) a podmnožinu $M = \{n \in \mathbb{N}; n > 12\} = \{13, 14, 15, \dots\}$. Pak nejmenší prvek M je $m = 13$, protože $13 \leq x$ pro všechna $x \in M$ a $13 \in M$. Také minimální prvek M je $m = 13$, protože v M neexistuje x , které by splňovalo $x < 13$.

Největší ani maximální prvek neexistují. Důkaz: Představme si, že by m bylo maximálním prvkem M . Pak by žádné prvky $x \in M$ nesměly splňovat $m < x$, ale některé to splňují, stačí si vzít $x = m + 1$. Kdyby nějaké $m \in M$ bylo největším prvkem, pak by pro všechna $x \in M$ muselo platit $x \leq m$, ale $m + 1$ to nesplňuje.

△

Tento příklad ukazuje, že množiny s nekonečným koncem jsou problémy. Není to ale problém jediný.

Příklad 4c.b: Uvažujme (\mathbb{Q}^+, \leq) a podmnožinu $M = \{x \in \mathbb{Q}; 0 < x < 1\}$. Tato množina nikam do nekonečna neutíká, ale nemá maximum ani minimum, nejmenší ani největší prvek.

Ukážeme, že nemůže mít nejmenší ani minimální prvek: Vezměme libovolného kandidáta $m \in M$. Pak $\frac{m}{2} \in M$ a $\frac{m}{2} < m$, tedy $\frac{m}{2}$ je protipříklad k tvrzení, že m je minimální, zároveň neplatí $m \leq \frac{m}{2}$ a proto není m ani nejmenší. Podobně ukážeme, že nelze najít největší ani maximální prvek M : Vezměme libovolného kandidáta $m \in M$. Pak $x = 1 - \frac{1-m}{2} \in M$ a $x > m$ (to se snadno ověří algebrou, ale je to vidět i geometricky, x jsme vytvořili tak, že jsme se podívali, jak daleko je m od 1, a x je o polovinu blíže).

△

Dostáváme se k zajímavé otázce, jaký je opravdu rozdíl mezi nejmenším a minimálním prvkem. Pro příklad roz hodně nepůjdeme k relaci \leq , protože tam to vyjde nastejno, jak brzy uvidíme. Musíme zkousit něco zajímavějšího.

!**Příklad 4c.c:** Uvažujme množinu množin $\mathcal{A} = \{\{1\}, \{2\}, \{1, 2\}, \{1, 2, 3\}\}$ a částečné uspořádání \subseteq . Jako podmnožinu M vezmeme přímo toto \mathcal{A} .

Prvek $\{1, 2, 3\}$ je největším prvkem \mathcal{A} , protože všechny prvky $N \in \mathcal{A}$ splňují $N \subseteq \{1, 2, 3\}$. Je také maximálním prvkem \mathcal{A} , protože v \mathcal{A} neexistuje množina N , která by splňovala $\{1, 2, 3\} \subseteq N$ a $\{1, 2, 3\} \neq N$.

Prvek $\{1\}$ je minimálním prvkem \mathcal{A} , protože neexistuje prvek N v \mathcal{A} , který by splňoval $N \subseteq \{1\}$ a $N \neq \{1\}$. Podobně je i $\{2\}$ minimálním prvkem \mathcal{A} . Nejmenší prvek \mathcal{A} ale neexistuje. Takový nejmenší prvek by totiž musel být podmnožinou všech množin z \mathcal{A} , například by musel být podmnožinou $\{1\}$ a také podmnožinou $\{2\}$, což splňuje jen prázdná množina, která ale není prvkem \mathcal{A} .

△

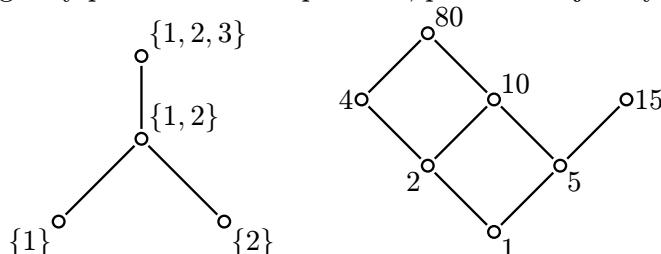
Příklad 4c.d: Uvažujme množinu $A = \{1, 2, 4, 5, 10, 15, 80\}$ uspořádanou relací dělitelnosti, tedy $a|b$ jestliže a dělí b . Existuje nějaký nejmenší prvek? Ano, číslo 1 dělí všechna čísla z A , tj. $1|a$ pro všechna $a \in A$, je to tedy nejmenší prvek A . Je to také prvek minimální, protože nejde najít $a \in A$ takové, že dělí 1 a přitom to není.

Největší prvek neexistuje, takový prvek m by totiž musel splňovat podmínu, že všechna čísla $a \in A$ jej dělí. Mluvíme tedy o společných násobcích čísel z A , ale žádný z nich v A není.

Při hledání maximálních prvků se díváme po číslech takových, že už nedělí žádné jiné číslo z A . Třeba 4 není maximální, protože $4|40$. Maximální členy jsou dva, jmenovitě 15 a 80.

△

! Podívejme se na Hasseovy diagramy posledních dvou příkladů, porovnáme je s výsledky, které jsme odvodili.



Intuitivně to funguje takto: Aby byl prvek největším v množině M , tak musí všechny ostatní prvky z M ležet pod ním a být s ním spojeny nějakou cestou. Aby byl prvek maximálním v M , pak stačí, aby žádný prvek z M nebyl nad ním a spojen cestou, což se zdá jakoby snaží splnit a ukážeme, že to je pravda. Symetricky, nejmenší prvek je takový, že všechny ostatní prvky M jsou nad ním a spojeny cestou (žádný takový v levém obrázku výše nevidíme), zato minimálnímu pruku M stačí, aby nebyl nikdo pod ním a spojen cestou. Lidově řečeno, minima a maxima jsou horní a dolní konce Hasseova diagramu, zatímco nejmenší prvek je kořen, do kterého se sbíhají všechny cesty, do největšího prveku se zase všechny cesty musí sbíhat nahore.

Člověk by řekl, že diagram bude mít nějaký konec vždycky, zato s tím sbíháním je to těžší (jak už jsme viděli). Vidíme také hierarchii mezi min/max a nejmenším/největším prvkem, pokud se třeba dole sbíhají všechny cesty, tak je to také spodní konec množiny. Šťouravější student by si ještě mohl všimnout, že jestliže existují dvě různá maxima (popř. minima), tak musí být navzájem nesrovnatelná. Tato pozorování patřičně zformulujeme a dokážeme, ale než se do toho dáme, zjednodušíme si situaci pomocí chytrého lemmátka.

Oč zde půjde? Velice nepřesně řečeno, v zásadě ukážeme, že bude stačit umět zacházet s jedním koncem Hasseova diagramu, třeba tím dolním, protože v situaci, kdy chceme něco provést nahore, jej prostě překlopíme vzhůru nohama (neboli přejdeme k inverzní relaci), provedeme to dole a zase jej překlopíme zpět. Ted' to řekneme pořádně.

Lemma 4c.1.

Nechť (A, \preceq) je částečně uspořádaná množina, uvažujme neprázdnou podmnožinu $M \subseteq A$.

(ia) m je minimum M vzhledem k \preceq právě tehdy, jestliže je m maximem M vzhledem k \preceq^{-1} .

(ib) m je maximum M vzhledem k \preceq právě tehdy, jestliže je m minimem M vzhledem k \preceq^{-1} .

(iia) m je nejmenším prvkem M vzhledem k \preceq právě tehdy, jestliže je m největším prvkem M vzhledem k \preceq^{-1} .

(iib) m je největším prvkem M vzhledem k \preceq právě tehdy, jestliže je m nejmenším prvkem M vzhledem k \preceq^{-1} .

Důkaz (rutinní): (ia): Předpokládejme, že m je minimum M vzhledem k \preceq . Ukážeme, že je to maximum M vzhledem k \preceq^{-1} . Sporem: Není-li to maximum, pak existuje $x \in M$ takové, že $m \preceq^{-1} x$ a $x \neq m$. Pak ale $x \preceq m$ a $x \neq m$, tedy $x \in M$ a $x \prec m$, což je ve sporu s $m = \min_{\preceq}(M)$.

Naopak předpokládejme, že m je maximum M vzhledem k \preceq^{-1} . Ukážeme, že je to minimum M vzhledem k \preceq . Sporem: Není-li to minimum, pak existuje $x \in M$ takové, že $x \preceq m$ a $x \neq m$. Pak ale $m \preceq^{-1} x$ a $x \neq m$, tedy $x \in M$ a $m \prec^{-1} x$, což je ve sporu s $m = \max_{\preceq^{-1}}(M)$.

(ib) se dokazuje obdobně.

(iia) m je nejmenší prvek M vzhledem k \preceq právě tehdy, když $m \preceq x$ pro všechna $x \in M$, což je právě tehdy, když $x \preceq^{-1} m$ pro všechna $x \in M$, což je právě tehdy, když m je největší prvek M vzhledem k \preceq^{-1} .

(iib) se dokazuje obdobně. □

Odtěď tedy víme, že minima a maxima mají stejné obecné vlastnosti, totéž je pravda o největších a nejmenších prvcích. Proto vždy stačí dělat důkaz jen pro jeden z nich, pro druhý bývá analogický.

!

Věta 4c.2.

Nechť je (A, \preceq) částečně uspořádaná množina, uvažujme neprázdnou podmnožinu $M \subseteq A$. Pak platí následující:

(i) Jestliže existuje nejmenší prvek M , pak je jediný.

Jestliže existuje největší prvek M , pak je jediný.

(ii) Jestliže $m_1 = \min(M)$, $m_2 = \min(M)$ a $m_1 \preceq m_2$, pak $m_1 = m_2$.

Jestliže $m_1 = \max(M)$, $m_2 = \max(M)$ a $m_1 \preceq m_2$, pak $m_1 = m_2$.

(iii) Jestliže je m nejmenší prvek M , pak $m = \min(M)$ a jiné minimum už není.

Jestliže je m největší prvek M , pak $m = \max(M)$ a jiné maximum už není.

Důkaz (rutinní): (i): Nechť m_1, m_2 jsou nejmenší prvky M , pak jsou mimo jiné z M . Protože je m_1 nejmenší z M , musí být $m_1 \preceq m_2$. Protože je m_2 nejmenší z M , musí být $m_2 \preceq m_1$. Antisimetrie pak dává $m_1 = m_2$, takže dva různé nejmenší prvky nelze mít.

Důkaz pro největší prvek je symetrický.

(ii): Sporem: Předpokládejme, že $m_1 \neq m_2$. Pak z předpokladu $m_1 \preceq m_2$ máme $m_1 \prec m_2$ a také máme $m_1 \in M$, což je ve sporu s předpokladem $m_2 = \min(M)$. Druhé tvrzení se dokáže symetricky.

(iii): Předpokládejme, že m je nejmenší prvek M . Pak pro všechny prvky $x \in M$ platí $m \preceq x$, tudíž pro ně už podle Lemmatu 4b.4 (ii) nemůže platit $x \prec m$. Proto je m minimum M .

Nechť n je také minimální prvek M . Protože je m nejmenším prvkem M a $n \in M$, platí nutně $m \preceq n$ a podle (ii) tedy $m = n$.

Podobně se dokazuje jedinečnost největšího prvku a pak i maxima.

□

Zajímavá je obměna tvrzení z (iii). Jestliže existuje více minimálních prvků M , pak neexistuje nejmenší prvek M , obdobně pro maxima a největší prvek.

! Jaká je tedy situace? Ani minimum, ani maximum, ani nejmenší či největší prvek vůbec nemusí existovat. Maximum a minimum mají lepší šanci na existenci, existují ve více případech než nejmenší a největší prvky, které jsou zase lepší z hlediska použití. Nejmenší/největší prvek je jen jeden (pokud tedy vůbec existuje), zatímco minim/maxim může být klidně více.

Co víme o existenci těchto prvků? Hned první příklad ukázal, že když vezmeme nekonečnou množinu, tak na existenci minim/maxim či nejmenšího/největšího prvku nelze spoléhat, dokonce nepomohou ani pokročilejší vlastnosti z další části. Pokud tedy chceme existenci těchto prvků vynutit, musíme se uchýlit ke konečným množinám.

Než se do toho dáme, všimněte si, že v definici maxima, minima, nejmenšího ani největšího prvku se nikde neodkazujeme na to, co se děje v A mimo M . Podobně ani v důkazu výše se vlastně vůbec nepracovalo s prvky mimo M . To znamená, že tyto pojmy vlastně závisí jen na M a restrikci \preceq na M , tedy na částečně uspořádané množině (M, \preceq) , nikoliv na tom, v jaké nadmnožině se M nachází. Na tom je založena oblíbená finta, kdy se v situaci $M \subseteq A$ rovnou prohlásí, že pracujeme s (M, \preceq) .

!

Věta 4c.3.

Nechť (A, \preceq) je částečně uspořádaná množina. Jestliže je M konečná neprázdná podmnožina A , pak existuje $\min(M)$ a $\max(M)$.

Důkaz (poučný): Podle právě provedené úvahy stačí dokázat, že pro libovolnou konečnou uspořádanou množinu (M, \preceq) existují $\min(M)$ a $\max(M)$. Jako obvykle dokážeme jen jednu věc, třeba existenci minima, a to indukcí.

Pro $n \in \mathbb{N}$ uvažujme $V(n)$: Jestliže je (M, \preceq) částečně uspořádaná množina o n prvcích, pak má minimum.

(0) Jednoprvková uspořádaná množina $M = \{m\}$ má určité minimum, jmenovitě m , protože v M nemohou být x takové, aby $x \prec m$, to totiž zahrnuje také podmínku $x \neq m$ a množina takovýchto prvků je prázdná.

(1) Nechť $n \in \mathbb{N}$ je libovolné a předpokládejme, že všechny n -prvkové množiny mají minimum. Potřebujeme ukázat, že totéž platí pro všechny množiny s $n + 1$ prvky. Uvažujme proto uspořádanou množinu (M, \preceq) , kde $|M| = n + 1$. Zvolme libovolný prvek $y \in M$ a podívejme se na $M' = M - \{y\}$. Když uděláme restrikci \preceq na M' , dostaneme n -prvkovou uspořádanou množinu, proto podle indukčního předpokladu existuje její minimum $m' \in M'$ vzhledem k \preceq . Teď porovnáme m' a y a rozobereme jednotlivé možnosti.

Jestliže $m' \preceq y$, tak tvrdíme, že $m' = \min(M)$. Na to musíme ukázat, že žádný prvek $x \in M$ nesplňuje $x \prec m'$. Pro $x \in M'$ to plyne z $m' = \min(M')$, zbývá případ $x = y$. Tam předpokládáme $m' \preceq y$, proto podle Lemmatu 4b.4 (ii) nemůže nastat $y \prec m'$. Minimum nalezeno.

Jestliže $y \preceq m'$, tak tvrdíme, že toto y je minimálním prvkem M . Dokážeme to sporem, předpokládejme, že existuje nějaké $x \in M$ takové, že $x \prec y$. Pak máme $x \prec y \preceq m'$, tedy podle Lemma 4b.4 (iii) $x \prec m'$, zároveň $x \in M - \{y\} = M'$ a máme spor s $m' = \min(M')$.

Zbývá varianta, že m' a y jsou neporovnatelné, tvrdíme, že pak m' je minimum celého M . Schválně zkusme vybrat nějaké x z $M - \{m'\}$. Jsou dvě možnosti. Buď $x = y$, pak je toto x neporovnatelné s m' , tedy rozhodně neplatí $x \prec m'$. Nebo $x \neq y$, pak $x \in M - \{y\} = M'$, a protože je m' minimum M' , zase nemůže být $x \prec m'$.

V každém případě jsme tedy nalezli minimum M , čímž je důkaz (1) dokončen.

Podle principu matematické indukce jsem tím dokázali existenci minima pro všechny konečné uspořádané množiny.

Alternativní důkaz: Zvolme $a_1 \in M$. Jestliže to není minimum M , tak existuje $a_2 \in M_1 = M - \{a_1\}$ takové, že $a_2 \prec a_1$. Jestliže a_2 není minimum, tak existuje $a_3 \in M - \{a_2\}$ takové že $a_3 \prec a_2$. Všimněme si, že $a_3 \neq a_1$, protože z tranzitivity \prec máme $a_3 \prec a_1$. Takže víme, že $a_3 \in M_2 = M - \{a_1, a_2\}$.

Jestliže a_3 není minimum, zvolme $a_4 \in M - \{a_3\}$ takové, že $a_4 \prec a_3$, zase $a_4 \in M_3 = M - \{a_1, a_2, a_3\}$. Atd., dříve či později musíme dostat minimum, protože M je konečná.

Tento alternativní důkaz vypadá snadněji, ale to je tím, že jsme nedokazovali některé kritické body, jinak by to pěkně narostlo. Vrátíme se k tomu v kapitole o indukci. □

4c.4 Poznámka: Všimněte si, že při kreslení Hasseova diagramu zdola jsme používali minima množiny. Tato věta zaručuje, že existují, tedy že dotyčný algoritmus bude fungovat. Existence minim a maxim také zaručí existenci covering relation. Vezměme prvek $a \in A$ a uvažujme množinu $\{x \in A; a \prec x\}$. Jestliže je prázdná, pak je a maximum A a tudíž přirozeně nemůže mít bezprostředního následníka. Pokud tato množina prázdná není, tak všechna její minima jsou bezprostředními následníky a . Podobně hledáme bezprostřední předchůdce jako maxima množiny $\{x \in A; x \prec a\}$.

△

Věta nám pro konečné množiny zaručuje existenci minim a maxim, ale pořád zůstává problém s největším a nejmenším prvkem. Pokud chceme jejich existenci vynutit, musíme zabránit tomu, aby bylo více minim či maxim. Jako nástroj se nabízí Věta 4c.2 (ii). Začneme se tedy ptát, jestli v dané částečně uspořádané množině (A, \preceq) dokážeme porovnávat prvky A pomocí \preceq .

! Definice.

Nechť (A, \preceq) je částečně uspořádaná množina. Řekneme, že $a, b \in A$ jsou **porovnatelné**, jestliže $a \preceq b$ nebo $b \preceq a$. Řekneme, že $a, b \in R$ jsou **neporovnatelné**, jestliže ani $a \preceq b$ ani $b \preceq a$ neplatí.

Let (A, \preceq) be a poset. We say that $a, b \in A$ are **comparable** if $a \preceq b$ or $b \preceq a$.

We say that $a, b \in A$ are **incomparable** if neither $a \preceq b$ nor $b \preceq a$.

Například pracujeme-li s relací \subseteq , pak množiny $\{13, 23\}$ a $\{3, 13\}$ porovnatelné nejsou, zato množiny $\{13, 23\}$ a $\{3, 13, 23\}$ porovnatelné jsou.

Podobně uvažujeme-li pro přirozená čísla relaci aRb jestliže a dělí b , pak čísla 6 a 12 porovnatelná jsou, zato čísla 6 a 9 porovnatelná nejsou.

! Definice.

Nechť (A, \preceq) je částečně uspořádaná množina. Řekneme, že \preceq je **lineární uspořádání**, popřípadě **úplné uspořádání**, jestliže jsou každé dva prvky z A porovnatelné.

Let (A, \preceq) be a poset. We say that \preceq is a **total order** or a **linear order** if every two elements from A are comparable.

! Příklad 4c.e: (\mathbb{Z}, \leq) je lineárně uspořádaná množina. (\mathbb{N}, \geq) je lineárně uspořádaná množina.

Na druhou stranu $(P(X), \subseteq)$ není lineárně uspořádaná množina pro $|X| > 1$ (viz příklad 4b.c), také relace dělitelnosti není lineární uspořádání na \mathbb{N} .

△

Poznámka: V této souvislosti si možná připomenete vlastnosti dichotomie a trichotomie z části Další vlastnosti (3c.8), které otázku porovnatelnosti kladou pro obecné relace. Částečné uspořádání je lineární právě tehdy, když je dichotomické.

Máme-li částečné uspořádání \preceq , které je dichotomické, tak od něj odvozená relace \prec je nutně trichotomická. Naopak začneme-li s asymetrickou, tranzitivní a trichotomickou relací \prec , tak od ní odvozená relace \preceq je lineární uspořádání.

△

Již tradičně se zeptáme, kdy se linearita zachovává.

Fakt 4c.5.

Je-li (A, \preceq) lineární uspořádání, pak jeho restrikce na libovolnou podmnožinu A je také lineární uspořádání.

Důkaz (poučný): Zde budeme protínat relaci s množinou, proto bude výhodnější zápis pomocí uspořádaných dvojic, bude také lepší použít pro \preceq písmeno R , takže namísto $a \preceq b$ píšeme $(a, b) \in R$.

Nechť B je podmnožina A , nechť S je restrikce R na B , připomeňme, že $S = R \cap (B \times B)$. Vezměme teď libovolné $a, b \in B$. Pak $a, b \in A$, proto dle linearity \preceq je splněn výrok „ $(a, b) \in R$ nebo $(b, a) \in R$ “. Ale z $(a, b) \in R$ plyne díky $(a, b) \in B \times B$ také $(a, b) \in S$, podobně z $(b, a) \in R$ plyne $(b, a) \in S$. Je tedy pravdivý výrok „ $(a, b) \in S$ nebo $(b, a) \in S$.“ \square

Jinými slovy, všechno při starém, používáním uspořádání na menší množině nic neztratíme.

Fakt 4c.6.

Nechť (A, \preceq) je lineárně uspořádaná množina. Pak je \preceq^{-1} také lineární uspořádání na A .

Důkaz je tak snadný, že to snad ani nestojí za těchto třináct slov.

Zajímavější jsou množinové operace. Víme, že průnikem dvou částečných uspořádání zase dostaneme částečné uspořádání, ale linearita už se nezachová. Například $\leq_i \geq$ jsou lineární uspořádání na \mathbb{Z} , ale jejich průnikem dostaneme $\Delta(\mathbb{Z})$ neboli relaci danou vztahem $a = b$, což rozhodně není lineární uspořádání.

Teď už je čas na nějaký pěkný výsledek.

!

Věta 4c.7.

Nechť (A, \preceq) je lineárně uspořádaná množina a M je její neprázdná podmnožina.

Jestliže je $m = \min(M)$, pak je to i nejmenší prvek M .

Jestliže je $m = \max(M)$, pak je to i největší prvek M .

Důkaz (rutinní): Předpokládejme, že $m = \min(M)$. Nechť $x \in M$ je libovolný, potřebujeme ukázat, že $m \preceq x$. Protože je \preceq lineární, musí platit $m \preceq x$ nebo $x \preceq m$. V případě toho prvního je důkaz hotov. Co kdyby platilo $x \preceq m$? Protože je m minimální, tak se nesmí stát $x < m$, což znamená, že $x = m$ a tedy zase $m \preceq x$. Důkaz je hotov.

Druhá část se dokáže symetricky. \square

U lineárně uspořádaných množin tedy minimum a nejmenší prvek jedno jsou, podobně pro maximum a největší prvek. Tím se vysvětluje, proč se v analýze pro podmnožiny \mathbb{R} definuje maximum a minimum, přičemž v definici je podmínka z největšího a nejmenšího prvku. Je to v analýze tradiční, ale z hlediska relací je to samozřejmě špatně. Naštěstí to díky linearitě uspořádání \leq na \mathbb{R} vyjde v tomto konkrétním případě nástejno.

A teď už okamžitý důsledek Věty 4c.3 a posledního tvrzení.

!

Věta 4c.8.

Nechť (A, \preceq) je lineárně uspořádaná množina. Každá její neprázdná konečná podmnožina má nejmenší a největší prvek.

Silně pokročilá a nedůležitá, nicméně možná zajímavá poznámka: Na existenci nejmenšího a největšího prvku nepotřebujeme plnou sílu uspořádání. Dobře to ukazuje následující tvrzení:

Věta 4c.9.

Nechť je R relace na množině A , která je tranzitivní. Nechť M je neprázdná množina prvků z A taková, že pro každé dva různé prvky $a, b \in M$ platí aRb nebo bRa . Pak existuje prvek $m \in M$ takový, že pro všechna $x \in M - \{m\}$ platí mRx .

Důkaz (poučný): Důkaz povedeme indukcí podle $|M|$.

$V(n)$: Každá množina velikosti n , na které je tranzitivní relace R splňující podmínu vzájemné porovnatelnosti všech různých prvků M , má prvek m s vlastností, že mRx pro všechna $x \in M - \{m\}$.

(0) $V(1)$ je triviálně splněno.

(1) Vezměme libovolné $n \in \mathbb{N}$ a předpokládejme, že $V(n)$ platí. Uvažujme teď množinu M o $n+1$ prvcích uspořádanou relací R splňující příslušné požadavky (tranzitivita, porovnatelnost). Zvolme nějaké $y \in M$, označme $M' = M - \{y\}$. Restrikce R na M' je pořád tranzitivní a porovnává všechny prvky, takže podle indukčního předpokladu existuje $m' \in M'$ takové, že $m'Rx$ pro všechna $x \in M' - \{m'\}$.

Protože $m', y \in M$ a $y \neq m'$, musí být podle předpokladu porovnatelné, platí tedy $m'Ry$ nebo yRm' .

Jestliže $m'Ry$, pak $m'Rx$ pro všechny $x \in (M' - \{m'\}) \cup \{y\} = M - \{m'\}$, máme tedy hledaný prvek.

Jestliže yRm' , pak tvrdíme, že y je ten hledaný prvek. Vezměme libovolné $x \in M - \{y\} = M'$. Jestliže $x = m'$, pak yRm' říká yRx a vše je v pořádku. Jestliže $x \neq m'$, pak $x \in M' - \{m'\}$ a proto $m'Rx$, také yRm' a podle tranzitivity yRx , jak bylo potřeba. \square

Poznámka: Z toho již hravě odvodíme předchozí větu: Nechť (A, \preceq) je lineárně uspořádaná množina, M její konečná podmnožina. Pak \preceq je tranzitivní relace, která díky linearitě porovnává všechny prvky M . Podle Věty 4c.9 tedy existuje prvek $m \in M$ takový, že pro $x \in M - \{m\}$ máme $m \preceq x$. Díky reflexivitě ovšem máme i $m \preceq m$, takže $m \preceq x$ pro všechna $x \in M$ a m je tedy nejmenší prvek.

\triangle

Věta nám poskytla existenci nejmenšího a největšího prvku za předpokladu, že máme lineární uspořádání. Následující tvrzení ukáže, že to platí i naopak, tedy bez linearity už se na takové prvky spoléhat nelze.

Fakt 4c.10.

Nechť (A, \preceq) je částečně uspořádaná množina. Jestliže platí, že každá dvouprvková podmnožina A má nejmenší prvek, pak (A, \preceq) už je nutně lineárně uspořádaná množina.

Důkaz (poučný): Nechť $x, y \in A$. Jestliže $x = y$, pak $x \preceq y$ z reflexivity. Jinak je $\{x, y\}$ dvouprvková podmnožina A , tudíž podle předpokladu musí mít nejmenší prvek. Pokud je jím x , tak platí $x \preceq y$, a pokud je jím y , tak platí $y \preceq x$. Tyto dva prvky jsou tedy porovnatelné. \square

Všechny konečné lineárně uspořádané množiny jsou v jistém smyslu stejné.

! Věta 4c.11.

Nechť (A, \preceq) je konečná částečně uspořádaná množina. Je to lineární uspořádání právě tehdy, jestliže lze prvky A napsat jako $A = \{a_1, \dots, a_n\}$ tak, aby $a_1 \prec a_2 \prec \dots \prec a_n$.

Důkaz (poučný): 1) \implies : Indukcí dokážeme tvrzení $V(n)$, že jestliže je n -prvková množina lineárně uspořádaná, pak ji lze příslušným způsobem seřadit.

(0) $V(1)$ evidentně platí, $A = \{a_1\}$.

(1) Vezměme libovolné $n \in \mathbb{N}$ a předpokládejme, že $V(n)$ platí. Uvažujme teď nějakou lineárně uspořádanou množinu (A, \preceq) o $n+1$ prvcích. Protože je to konečná lineárně uspořádaná množina, tak musí mít největší prvek m . Uvažujme $M' = M - \{m\}$. Ukázali jsme, že restrikce lineárního uspořádání je zase lineární uspořádání, takže (M', \preceq) je lineárně uspořádaná množina o n prvcích, tudíž ji podle indukčního předpokladu můžeme uspořádat jako $M - \{m\} = \{a_1 \prec a_2 \prec \dots \prec a_n\}$. A jelikož je m největší prvek M , tak určitě $a_n \preceq m$, také $a_n \neq m$, máme tedy $a_n \prec m$. Můžeme tudíž položit $a_{n+1} = m$ a jsme hotovi.

2) \impliedby : Předpokládejme, že částečně uspořádanou množinu (A, \preceq) lze napsat jako $A = \{a_1 \prec a_2 \prec \dots \prec a_n\}$. Chceme ukázat, že je lineárně uspořádná. Ale to je snadné, kdykoliv nám někdo dá $a, b \in A$, tak musí existovat i, j takové, že $a_i = a$ a $a_j = b$. Možnost $i = j$ odpovídá $a = b$, pak $a \preceq b$ díky reflexivitě. Pokud by bylo $i < j$, pak je a v tom řetězci někde před b , tudíž dle tranzitivity $a \preceq b$. Případ $i > j$ pak symetricky a naprostě stejně vede na $b \preceq a$. Čili ať už nastane jakýkoliv případ, a, b jsou porovnatelné. \square

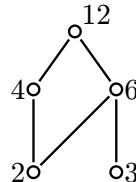
! Řečeno jinak, konečná uspořádaná množina je lineárně uspořádaná právě tehdy, jestliže její Hasseův diagram vypadá jako jedna šnůrka s korálky zdola nahoru. Zajímavá věc je, že každý Hasseův diagram lze na takovou šnůrku upravit. Jak se to dělá? Jestliže pro dané uspořádání linearita nefunguje, tak to způsobil nedostatek vhodných srovnání. Naskytá se tedy nápad napravit to tím, že prostě nějaké dvojice do relace přidáme. Musí se to ovšem dělat opatrně, aby se nové dvojice nedostaly do rozporu s původním porovnáváním, tedy aby vzniklá relace byla pořád uspořádání.

!

Definice.

Nechť (A, \preceq) je částečně uspořádaná množina. Relace \preceq_L na A se nazývá **lineární rozšíření** (linear extension) relace \preceq , jestliže je (A, \preceq_L) lineárně uspořádaná množina a $\preceq \subseteq \preceq_L$, tedy pro všechna $a, b \in A$ splňující $a \preceq b$ platí i $a \preceq_L b$.

Příklad 4c.f: Uvažujme množinu $A = \{2, 3, 4, 6, 12\}$ uspořádanou dělitelností $|$. Tato možina není lineárně uspořádaná, jak ostatně napoví Hasseův diagram.



Nebo si prostě hned všimneme, že 4 a 6 jsou dělitelností neporovnatelné. Uvažujme teď tutéž množinu, ale s uspořádáním \leq . Pak ji lze krásně uspořádat $2 < 3 < 4 < 6 < 12$, je to tedy již lineárně uspořádaná množina, a zároveň toto uspořádání nejde nikde proti tomu původnímu. Například relace $3 | 6$ ríká, že 3 je před 6 vzhledem k $|$, a toto je v novém uspořádání zachováno. Zkuste si projít všechny porovnatelné dvojice z $(A, |)$ (je jich 8) a přesvědčte se, že mají stejný pořadí i v novém srovnání.

Nebo dokažte obecně, že jestliže $a, b \in \mathbb{N}$ a $a | b$, pak $a \leq b$, a máte důkaz, že pro libovolnou podmnožinu N přirozených čísel uspořádanou pomocí dělitelnosti dostanete její lineární rozšíření pomocí \leq .

△

Linearizaci lze najít pro každou konečnou uspořádanou množinou.

!

Věta 4c.12.

Pro každou konečnou částečně uspořádanou množinu (A, \preceq) existuje lineární rozšíření \preceq_L daného uspořádání \preceq .

Důkaz (poučný, drsný): Důkaz provedeme indukcí.

Tvrzení $V(n)$: Každé uspořádání na n -prvkové množině lze lineárně rozšířit.

(0) $V(1)$ evidentně platí.

(1) Mějme libovolné $n \in \mathbb{N}$ a předpokládejme, že lineární rozšíření lze najít pro všechny n -prvkové množiny. Teď nechť (M, \preceq) je uspořádaná množina o $n+1$ prvcích. Dokázali jsme, že konečné uspořádané množiny mají maximální prvek, tak si jedno $m = \max(M)$ vezměme a uvažujme množinu $M' = M - \{m\}$, kterou uspořádáme restrikcí \preceq . Dostaneme n -prvkovou částečně uspořádanou množinu, proto podle indukčního předpokladu pro ni existuje lineární rozšíření $\preceq_{L'}$. Můžeme tedy psát $M' = \{a_1 \preceq_{L'} a_2 \preceq_{L'} \dots \preceq_{L'} a_n\}$. Přidejme definici, že $m \preceq_L m$ a $a_i \preceq_L m$ pro všechna $a_i \in M'$ a dostaneme relaci \preceq_L na M . Ukážeme, že je to lineární rozšíření \preceq .

Je to ovšem relace, jejíž část jsme dodlávali sami na koleně, tudíž není vůbec jasné, co všechno jsme náhodou nezkazili, takže teď nevíme nic, ani základní vlastnosti, musíme dokázat všechno.

Bude se nám při tom hodit, když si ujasníme jednu věc. V okamžiku, kdy napíšeme $a \preceq_L b$, tak jsou dvě možnosti. Buď $a, b \in M'$, pak ovšem to \preceq_L pochází z indukčního předpokladu, tudíž má všechny tři vlastnosti uspořádání (R,A,T) a můžeme je použít. Pokud by ale byl některý z prvků roven m , tak už jde o \preceq_L , které jsme my následně vyrobili definicí. Pak je situace zajímavá v tom, že o něm zatím žádné vlastnosti neznáme, zato ale přesně víme, co se děje. Jmenovitě, všechny prvky $a \in M$ splňují $a \preceq_L m$, ale situace, kdy $m \preceq_L b$, může nastat jedině, pokud také $b = m$, protože pro žádné jiné $x \neq m$ jsme dvojici $m \preceq_L x$ do naší relace nezařadili.

1) (M, \preceq_L) je částečné uspořádání:

Reflexivita: $m \preceq_L m$ platí dle definice, pro $x \in M'$ platí $x \preceq_L x$ z toho, že \preceq_L je uspořádání na M' .

Antisimetrie: Nechť $a, b \in M$ splňují $a \preceq_L b$ a $b \preceq_L a$. Jsou tři možnosti.

Jestliže $a, b \in M'$, pak jde o relaci \preceq_L na M' a ta už přišla coby uspořádání, proto $a = b$.

Jestliže $a = m$ a $b = m$, pak $a = b$ a je to zase v pořádku.

Poslední možnost je, že jeden z a, b je z M' a druhý je m , to ale nemůže nastat, protože u takových dvojic jsme \preceq_L definovali přímo sami předpisem a pro každou dvojici a_i, m jsme do \preceq_L zařadili jen jedno srovnání.

Tranzitivita: Nechť $a, b, c \in M$ splňují $a \preceq_L b \preceq_L c$. Zase musíme rozebrat možnosti podle toho, odkud prvky jsou.

Jestliže $a, b, c \in M'$, pak zde používáme \preceq_L , jak přišlo z indukce, je to tedy uspořádání a z jeho tranzitivity $a \preceq_L c$.

Jiná možnost je, že některý z prvků je m . Pokud by to bylo c , pak buď také $a = m$ a $a \preceq_L c$ je z reflexivity, nebo $a \in M'$ a pak $a \preceq_L c$, protože tak jsme srovnání mezi m a prvky z M' definovali.

Pokud by $b = m$, tak dle předchozí diskuse už z $b \preceq_L c$ plyne $c = m$, tudíž $b = c$ a předpoklad $a \preceq_L b$ říká i $a \preceq_L c$.

Poslední možností je, že $a = m$. Pak z $a \preceq_L b$ plyne i $b = m$ a můžeme použít závěr předchozího odstavce. Každopádně \preceq_L je tranzitivní.

2) \preceq_L je lineární: To je snadné, máme $M = \{a_1 \prec_L \dots \prec_L a_n \prec_L m\}$, tudíž jde o lineární uspořádání.

3) Platí $\preceq \subseteq \preceq_L$:

Nechť $a, b \in M$ splňují $a \preceq b$. Jestliže $a, b \in M'$, pak $a \preceq_L b$, protože jde o relaci z indukčního předpokladu.

Jestliže je $b = m$, pak tedy zkoumáme situaci $a \preceq m$ a i zde máme $a \preceq_L m$, tak jsme to definovali pro všechna $a \in M$.

Zbývá případ $a = m$. My jsme ale m zvolili jako maximum M , to znamená, že jediný prvek M , který splňuje $m \preceq b$, je $b = m$. Máme tedy $a = b = m$ a $a \preceq_L b$ z toho, že jsme definovali i $m \preceq_L m$.

Tím je důkaz hotov. □

! Tento proces je docela užitečný a má své jméno. Pro matematiky je zajímavý teoreticky, protože s lineárním uspořádáním se lépe pracuje, říkají tomu **linearizace částečného uspořádání**. Pro lidi z computer science je to praktický nástroj a říkají tomu **topologické uspořádávání**. Pojem topologie computerscientisti používají k označení prostorového rozmístění prvků a zde vlastně nejde o nic jiného, než vzít třeba i docela košatý Hasseův diagram a zmáčknout jej chytře z boku tak, aby vznikla jedna svislá šňůrka, a to tak, aby se hrany, které jdou nahoru, nepřeklopily během procesu do opačného směru. (Matematici používají pojem „topologie“ pro něco úplně jiného, proto mají svůj název.)

Důkaz Věty zároveň slouží jako praktický návod na algoritmus takového procesu, a protože jde o důkaz indukcí, půjde o algoritmus rekursivní. Abychom ukázali, že i zde funguje symetrie mezi horním a dolním koncem, zkusíme v algoritmu budovat linearizaci zdola na rozdíl od důkazu, kde jsme to dělali shora.

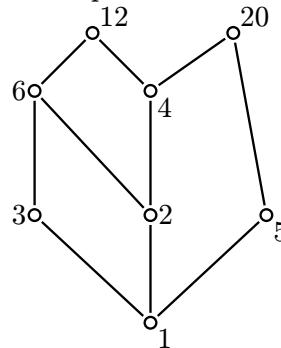
```
procedure topological sort ((A, ≤))
k := 0;
while A ≠ ∅ do
    k := k + 1;
    ak := min(A);
    A := A - {ak};
output: (a1 ≤ a2 ≤ a3 ≤ ... ≤ an);
```

Tento postup docela dobře koresponduje s algoritmem pro tvoření Hasseova diagramu. Lze toho využít. Pokud máme Hasseův diagram uspořádání vytvořený pomocí standardního algoritmu zdola, pak linearizaci provedeme velice snadno: Nejprve vezmeme prvky z dolního řádku a seřadíme je libovolně. Za ně zařadíme prvky z druhého řádku a seřadíme je libovolně. A tak dále, až je množina linearizována.

! **Příklad 4c.g:** Uvažujme částečně uspořádanou množinu $(\{1, 2, 3, 4, 5, 6, 12, 20\}, |)$, viz příklad 4b.f.

Pak $a_1 = \min(\{1, 2, 3, 4, 5, 6, 12, 20\}) = 1$, a_2 je nějaké $\min(\{2, 3, 4, 5, 6, 12, 20\})$, třeba $a_2 = 3$, a_3 je nějaké $\min(\{2, 4, 5, 6, 12, 20\})$, třeba $a_3 = 2$, a_4 je nějaké $\min(\{4, 5, 6, 12, 20\})$, třeba $a_4 = 4$, a_5 je nějaké $\min(\{4, 6, 12, 20\})$, třeba $a_5 = 6$, a_6 je nějaké $\min(\{4, 12, 20\})$, třeba $a_6 = 4$, a_7 je nějaké $\min(\{12, 20\})$, třeba $a_7 = 12$, a a_8 je nějaké $\min(\{20\})$, tedy $a_8 = 20$,

Dostáváme linearizaci $1 \prec_L 3 \prec_L 2 \prec_L 5 \prec_L 6 \prec_L 4 \prec_L 12 \prec_L 20$. Podíváte-li se na Hasseův graf z příkladu 4b.f., vidíte, že jsme prostě brali postupně řádky zleva doprava.



Jinou linearizací by bylo srovnat čísla podle velikosti, což ukazuje, že linearizace rozhodně není jednoznačná.

△

! Poznámka k příkladu: Při praktickém tvoření linearizace rukou se vyplatí vypsat si všechny dvojice z odpovídající relace \prec . V onom příkladě jsou to $1 \prec 2, 1 \prec 3, 1 \prec 4, 1 \prec 5, 1 \prec 6, 1 \prec 12, 1 \prec 20, 2 \prec 4, 2 \prec 6, 2 \prec 12, 2 \prec 20, 3 \prec 6, 3 \prec 12, 4 \prec 12, 4 \prec 20, 5 \prec 20, 6 \prec 12$.

V prvním kroku pak hledáme minimum, tedy takové číslo, které se v žádné z těchto dvojic neobjeví napravo. Najdeme jedničku, pak ze seznamu vymažeme všechny dvojice, které jedničku obsahují, a opakujeme tento postup. Vyzkoušejte si, že dostáváte tytéž situace jako v předchozím řešení.

K čemu toto může být? Máte úkoly u_1, \dots, u_n , které musíte splnit, ale existují mezi nimi jisté závislosti, třeba že u_7 se musí dělat až po u_3 a podobně. Cílem je seřadit tyto úkoly tak, aby člověk mohl dělat jeden po druhém a přitom zachoval podmínky. Jde tedy vlastně o docela užitečnou záležitost z oboru plánování, asi si umíte představit, že něco takového se může hodit programu, který zadává úkoly procesoru, inženýrovi navrhujícímu výrobní linku a spoustě dalším lidem. Je to samostatný obor, ke kterému jsme zde jen přičichli na té nejsnažší úrovni, opravdu zajímavé to začne být, když povolíme paralelní zpracování a podobné hrátky, započítáme do toho délku splnění úkolu, ceny operací a ceny prostojů a ceny přepravy od úkolu k úkolu a chceme ještě něco minimalizovat a vůbec se na tom dá vyřádit (něco takového určitě řeší v Airbusu, kde z politických důvodů vyrábí každý kousek letadla někde jinde, ale nakonec z toho musí s co nejmenšími náklady a námahou vylézt letadlo, pokud možno kompletní). Ke slovu přijdou pokročilé matematické metody, z nichž některé určitě potkáte v kursu o optimalizaci.

Z našeho pohledu to má jen jeden malý zádrhel, taková praktická vyjádření mají jen málokdy podobu částečného uspořádání, většinou dostaneme jen několik ostrých srovnání typu $u_2 \prec u_5$ (úkol 2 se musí dělat před úkolem 5). Formálně bychom to řešili tak, že bychom uvažovali nejmenší možnou relaci \preceq takovou, že už je to částečné uspořádání a přitom obsahuje podmínky \prec . Toto ovšem není vždy možné udělat, protože ty podmínky \prec nemusí doplnění na uspořádání umožňovat. Jednoduchý příklad: Kdybychom měli $u_1 \prec u_2, u_2 \prec u_4$ a $u_4 \prec u_1$, tak při snaze o doplnění na tranzitivní relaci přidáme $u_1 \prec u_4$ (viz první dvě priority) a hned máme spor se třetí. To ale vlastně není na škodu, v takovém případě je stejně nemožné najít pořadí úkolů, které vyhoví zadaným podmínkám, protože si podmínky odporuji, čili se při pokusu o vytvoření uspořádání dozvímme, že úloha nemá řešení.

V praxi dokonce ani nemusíme uspořádání vytvářet, prostě jen na zadané priority aplikujeme stejný algoritmus, jaký jsme použili v předchozím příkladě. Nejprve najdeme úkol, který se nikdy nevyskytuje ve srovnáních napravo, to bude ten první. Pak ze seznamu vymažeme všechny dvojice, kde tento úkol je, a krok s vyhledáváním opakujeme. A tak dále, buď se podaří seřadit všechny úkoly, nebo se to někde zadrhne (nenajdeme vhodný prvek), což je znamení, že úloha není řešitelná.

Příklad 4c.h: Vyvíjí se program, je nutno udělat komponenty p_1 až p_7 , přičemž p_7 lze udělat až po p_4 a p_6 ; p_6 až po p_5 a p_2 ; p_4 až po p_1, p_2 a p_3 ; p_2 až po p_1 a p_3 . Najdeme nějaké pořadí, v jakém je dělat.

Minimální komponenta je taková, že se nic nemusí dělat před ní. Nejprve si pro názornost přepíšeme předpoklady do tvaru relace: $p_4 \prec p_7, p_6 \prec p_7, p_5 \prec p_6, p_2 \prec p_6, p_1 \prec p_4, p_2 \prec p_4, p_3 \prec p_4, p_1 \prec p_2$ a $p_3 \prec p_2$.

Hledáme minimum, tedy komponentu, která se nevyskytuje ve srovnáních napravo. Vidíme p_1 , tím začneme. Teď odebereme všechna srovnání, ve kterých se p_1 vyskytuje, zbude seznam

$$p_4 \prec p_7, p_6 \prec p_7, p_5 \prec p_6, p_2 \prec p_6, p_2 \prec p_4, p_3 \prec p_4, p_3 \prec p_2.$$

Hledáme minimum mezi p_2, \dots, p_7 , najdeme třeba p_3 . Dostáváme tedy $p_1 \prec_L p_3$, nový seznam bez p_3 :

$$p_4 \prec p_7, p_6 \prec p_7, p_5 \prec p_6, p_2 \prec p_6, p_2 \prec p_4.$$

Další minimum je třeba p_2 . Dostáváme tedy $p_1 \prec_L p_3 \prec_L p_2$, nový seznam bez p_2 : $p_4 \prec p_7, p_6 \prec p_7, p_5 \prec p_6$.

Další minimum je třeba p_5 . Dostáváme tedy $p_1 \prec_L p_3 \prec_L p_2 \prec_L p_5$, nový seznam bez p_5 : $p_4 \prec p_7, p_6 \prec p_7$.

Další minimum je třeba p_4 . Dostáváme tedy $p_1 \prec_L p_3 \prec_L p_2 \prec_L p_5 \prec_L p_4$, nový seznam bez p_4 : $p_6 \prec p_7$.

Ted už dorazíme celou linearizaci, máme $p_1 \prec_L p_3 \prec_L p_2 \prec_L p_5 \prec_L p_4 \prec_L p_6 \prec_L p_7$.

Pro srovnání si zkuste vytvořit Hasseův diagram, viz cvičení 4c.7.

△

Shrňme si, jakou máme situaci. Jestliže chceme mít zaručeny nejmenší prvky, potřebujeme na to linearitu. Pro konečné množiny už to stačí, ale pro nekonečné ani to stačit nemusí. Tam už žádné prostředky na vynucení nemáme, prostě bud máme štěstí a nejmenší prvky jsou, nebo nejsou. Případy s existujícími nejmenšími prvky jsou důležité, tak jim dáme jméno.

Definice.

Nechť (A, \preceq) je částečně uspořádaná množina. Řekneme, že (A, \preceq) je **dobře uspořádaná množina**, jestliže každá neprázdná podmnožina $M \subseteq A$ má nejmenší prvek.

Let (A, \preceq) be a poset. We say that (A, \preceq) is a **well-ordered set** if every non-empty subset of A has a least element.

Někteří autoři definují dobré uspořádání jako uspořádání, které je lineární a podmnožiny mají nejmenší prvky. Není v tom žádný rozdíl, protože i naše definice už linearitu automaticky zahrnuje.

! Fakt 4c.13.

Každé dobré uspořádání je také lineární.

Vyplývá to okamžitě z Faktu 4c.10.

Všimněte si, že zde ztrácíme onu symetrii mezi „horním a dolním koncem“. Jestliže máme dobré uspořádání \preceq a hledáme největší prvek podmnožiny M , pak jej můžeme hledat jako nejmenší prvek M vůči \preceq^{-1} , ale o tomto inverzním uspořádání obecně nevíme, zda je dobré (viz níže).

Díky Větě 4c.8 víme, že každá konečná lineárně uspořádaná množina je dobré uspořádaná. Teď si představíme nejdůležitější nekonečnou dobré uspořádanou množinu.

! 4c.14. Princip dobrého uspořádání (well-ordering principle).

(\mathbb{N}, \leq) je dobré uspořádaná množina.

! 4c.15 Poznámka (důležitá): Toto tvrzení asi čtenáře nikterak nepřekvapilo, minima různých množin přirozených čísel hledáme běžně. Zajímavé ovšem je, že jsme tento užitečný fakt neuvedli jako větu. Proč? Pro odpověď musíme začít od jiného konce.

V předchozích kapitolách jsme už viděli, jak matematika roste coby strom. Něco dokážeme, pak pomocí toho dokážeme něco těžšího, čímž se naše znalosti rozvětví, ale z těch nových věcí zase dostaneme další znalosti, a tak se matematika postupně košatí, větve se zase spojují a zaplétají a zase větví, vzniká tak velice komplikovaný keř. Zajímavá otázka: Co najdeme, když se po větvích pustíme opačným směrem? Každý fakt je dokázan pomocí jistých jednodušších faktíků, ty jsou zase dokázány pomocí jiných věcí. Takže sestupujeme stále níže až ke kořenům. Co tam najdeme?

V zásadě nic :-). Neexistují totiž žádné věci, které by byly pravdivé samy od sebe. Můžete si například myslit, že bychom mohli jako jeden základ vzít $1 + 1 = 2$, ale pak si člověk položí otázku, co je vlastně 1, a rychle znejistí. Když tedy nejsou žádné absolutní pravdy, z čeho vlastně vycházíme?

Z takzvaných axiomů. Matematici si podobné otázky kladli někdy v 19. století, pořádně prozkoumali onen strom matematického vědění, co z čeho plyne a co by potřebovali znát, aby věci fungovaly, až se nakonec domluvili na seznamu kritických vlastností. Dohodli se, že tyto věci budeme považovat za pravdivé, říká se jim axiomy, a pak se uvidí, jak daleko s tím dokážeme zajít. Existuje skupina axiomů, která je všeobecně přijímaná a stojí na ní současná matematika, pro zajímavost je to třeba axiom, že vůbec nějaké množiny existují, nebo axiom, že když máme dvě množiny A a B , tak když je použijeme coby prvky v objektu $\{A, B\}$, tak ta nová věc je také množina. Axiomy jsou tedy věci, o kterých nevíme jistě, zda jsou pravdivé, ale přijde nám vhodné a praktické je za pravdivé považovat.

Tím se dostáváme k principu dobrého uspořádání. Jeho platnost se pomocí standardních axiomů dokázat nedá. Protože to ale přijde matematikům rozumné, tak si jeho fungování přibírájí jako další axiom. V kapitole o indukci uvidíme, že přibráním tohoto axiomu si zároveň otevříme dveře k mnoha dalším užitečným matematickým trikům, nejde tedy o žádnou kontroverzní volbu. Pokud se čtenář nechce hlouběji procházet filosofií matematiky, může to prostě brát jako další daný fakt, jako věří i $1 + 1 = 2$.

Mimochodem, proč se matematici rozhodli právě pro ty axiomy, které máme? V zásadě je možné použít libovolné axiomy, které nejsou vzájemně ve sporu, a dostaneme z nich nějakou teorii. Problém je, že taková teorie by nejspíše byla na nic, protože by se pravidla takového světa silně lišila od našich. Protože bychom přeci jen chtěli, ať nám matematika pomáhá v poznávání zrovna našeho světa, tak se matematici snaží volit axiomy tak, aby se vzniklé výsledky podobaly tomu, co vidíme kolem nás. Už od poloviny 20. století máme vyjasněno, na kterých axiomech matematika stojí, matematici jsou s nimi (až na pár odvážlivců) spokojeni a výsledky, které pomocí nich dostáváme, fungují. A to je to hlavní.

△

Poznámka (pro hračičky): Protože axiomy hrají základní roli a silně ovlivní to, jak výsledná teorie vypadá, tak se samozřejmě pilně zkoumají. Jedním z témat je vzájemná souvislost. Pokud se třeba ukáže, že nějaký axiom už se dá dokázat na základě jiných, tak je vlastně navíc, jen zbytečně kalí vodu. Ale nejde jen o to. Dobrá otázka také je, jestli když si nějaký axiom přidáme, tak tím nezpůsobíme rozpor. Když to přeženu, asi bychom z axiomů $1 + 1 = 2$ a $1 + 1 = 3$ daleko nedošli. To je důležité v situaci, kdy zvažujeme, zda ještě další axiom nepřibrat. Dobrým příkladem je axiom výběru (AC, axiom of choice). Víme, že je na ostatních základních

axiomech nezávislý, což znamená, že s nimi ani není ve sporu, ani z nich nevyplývá, takže když jej přibereme, tak tím teorii obohatíme. Většina matematiků jej bere, protože je velice užitečný, ale existují i takoví, kteří jej odmítají zabudovat do základů matematické teorie. Mají tedy teorii svou, bez (AC), a jak se dá čekat, mají tam také méně tvrzení, protože zrovna (AC) patří mezi velice oblíbené nástroje při důkazech.

Podobně také víme, že žádné průsvihy nevzniknou, když si mezi axiomy přibereme jeden o platnosti hypotézy kontinua, o které jsme mluvili v kapitole 2c (viz poznámka 2c.20 a zamýšlení za ní).

A protože jsou matematici hraví a zvídaví, tak se také ptají, co se stane, když naopak nějaký axiom ubero. S tím je spojena zajímavá příhoda. Na axiomech stojí i geometrie a někdy na začátku 19. století si někteří matematici řekli, že zkusí jeden axiom vyhodit, co to udělá. Ne že by si mysleli, že to není pravda, ale jedna věc je to, co matematici vidí kolem sebe, a druhá jsou axiomy. A tak ho vyhodili a dostali docela zajímavou geometrii, ve které se děly podivné věci, například už se úhly v trojúhelníku nenasčítaly do π . Byla to taková veselá hříčka, a najednou se o sto let později ukázalo, že ve skutečnosti přesně takto funguje vesmír ve velkém měřítku (obecná teorie relativity a podobné Einsteinoviny). Hrátky s axiomy se tedy mohou zajímavě zvrhnout.

△

Příklad 4c.i: Teď použijeme princip dobrého uspořádání k důkazu, že všechna přirozená čísla lze popsat nejvíce 10 českými slovy.

Zkusíme to sporem, předpokládejme, že to nejde. Pak je množina M přirozených čísel, která nejdou popsat nejvíce 10 českými slovy, neprázdná. Podle principu dobrého uspořádání tedy musí existovat její nejmenší prvek, nazveme jej m . Pak je m vlastně nejmenší přirozené číslo, které nelze popsat nejvíce desíti českými slovy, takže jsme jej popsal přesně deseti českými slovy a máme spor.

△

Poznámka: Další nepovinný náhled za oponu matematiky: Všimněte si, že jsme dokázali popsatelnost všech čísel deseti slovy, ale tento důkaz nám nedal žádný návod, jak to vlastně dělat. Já třeba nevím, jak bych popsal číslo 1946240936592523658376, jen vím, že to nějak musí jít. Jsou principiálně dva druhy důkazů, že něco existuje. Důkazy *konstruktivní* to dokazují tak, že dotyčný objekt přímo najdou (či doloží, jak jej najít). Pak jsou důkazy *nekonstruktivní*, kdy se existence dovodí nějakým trikem, aniž bychom ale nějaký exemplář předvedli. Někteří matematici s těmito důkazy mají filosofický problém a neuznávají je za platné. Mají tak svou vlastní teorii matematiky, která je o něco chudší než ta obecně přijímaná, protože k některým výsledkům se konstruktivně neumí dostat, a komplikovanější, protože k mnoha výsledkům se nakonec dopracovali, ale trnitější cestou. Je to ale silně okrajový fenomén.

△

Dobře uspořádané množiny jsou velice užitečné, mimo jiné tím, že na nich lze používat matematickou indukci (viz kapitola 5b). Ukážeme si jich teď více, ale začneme pro srovnání několika případy, kdy dobrota naopak selže.

Příklad 4c.j: (\mathbb{Z}, \leq) je lineárně uspořádaná množina, ale není dobře uspořádaná, neboť neprázdná podmnožina $M = \mathbb{Z}$ určitě nemá nejmenší prvek, tedy prvek m , který by splňoval $m \leq n$ pro všechna $n \in \mathbb{Z}$.

△

Příklad 4c.k: (\mathbb{Q}, \leq) je lineárně uspořádaná množina, ale není dobře uspořádaná, neboť neprázdná podmnožina $M = \{x \in \mathbb{Q}; 0 < x < 1\}$ určitě nemá nejmenší prvek.

Tento příklad je zajímavý tím, že M „neutíká“ nikam do nekonečna.

△

Příklad 4c.l: Nechť X je libovolná alespoň dvouprvková množina, třeba konečná, uvažujme množinu jejích podmnožin $A = P(X)$ uspořádanou inkluzí. Pak nejde o lineární uspořádání, tím pádem ani nemůže být dobré, viz Fakt 4c.13 a příklad 4b.c.

Podobně nedostáváme lineární uspořádání u relace dělitelnosti, pokud vhodně zvolíme množinu A . Například na množině $A = \{1, 2, 4, 8, 16\}$ dává dělitelnost dobré uspořádání, ale na $A = \{1, 2, 3, 6\}$ už ne: Podmnožina $\{2, 3\}$ nemá vůči dělitelnosti nejmenší prvek.

Konečné příklady nedobrých uspořádání se tedy vyrábějí velice snadno.

△

Příklad 4c.m: (\mathbb{N}, \geq) je lineárně uspořádaná množina, ale není dobře uspořádaná, neboť neprázdná podmnožina $M = \mathbb{N}$ určitě nemá nejmenší prvek. Opravdu? Takový prvek m by musel splňovat $m \geq n$ pro všechna $n \in \mathbb{N}$, takže neexistuje.

△

Poslední příklad ukazuje, že když vezmeme dobře uspořádanou množinu (\mathbb{N}, \leq) a přejdeme k duálnímu uspořádání $(\mathbb{N}, \leq^{-1}) = (\mathbb{N}, \geq)$, tak se dobrota může ztratit. Přechod k inverznímu uspořádání tedy rozhodně není něco, co bychom chtěli u dobře uspořádaných množin dělat, a ztrácíme onu příjemnou symetrii mezi pohledy „shorna“ a „zdola“. Naopak přechod k podmnožině je jako obvykle bezproblémový.

Fakt 4c.16.

Nechť (A, \preceq) je dobře uspořádaná množina. Pak pro libovolnou neprázdnou podmnožinu $B \subseteq A$ je restrikce \preceq na B dobré uspořádání.

Důkaz (rutinní): Podle Faktu 4c.5 už víme, že tato restrikce je lineární uspořádání. Zbývá dokázat, že je dobré. Nechť M je neprázdná podmnožina B . Pak je to i neprázdná podmnožina A a tudíž existuje její nejmenší prvek m . Podmínka, že $m \preceq x$ pro všechna $x \in M$, je ale zcela nezávislá na tom, v jaké nadmnožině M je, relace \preceq a její restrikce na B se na M shodují, tudíž je m také minimem M v (B, \preceq) . \square

To je zase případ, kdy je to tak lehké, až člověk přemýšlí, co vlastně napsat. Dobře uspořádané množiny tedy můžeme získávat pomocí podmnožin (\mathbb{N}, \leq) (což tady děláváme), dalším zajímavým zdrojem je kartézský součin.

4c.17 Uspořádání a kartézský součin

V obecné kapitole o relacích jsme zavedli uspořádání na kartézském součinu množin, viz 3b.9. Vysoko užitečné to začne být zejména u uspořádání. Připomeňme, že máme-li množiny s relacemi $(A_1, \preceq_1), \dots, (A_n, \preceq_n)$, pak součinové uspořádání porovnává vektory z $A_1 \times \dots \times A_n$ předpisem $(a_i) \preceq (b_i)$ právě tehdy, pokud $a_i \preceq_i b_i$ po všechna i .

Z obecného tvrzení 3b.10 okamžitě dostáváme důsledek.

Fakt 4c.18.

Jestliže jsou $(A_1, \preceq_1), \dots, (A_n, \preceq_n)$ částečně uspořádané množiny, pak je i $A_1 \times \dots \times A_n$ se součinovým uspořádáním částečně uspořádaná množina.

To vypadá moc pěkně, ale ve skutečnosti se to používá relativně zřídka, protože u tohoto typu uspořádání jsou vážné problémy s porovnatelností vektorů. Například v součinu $(\mathbb{N}, \leq) \times (\mathbb{N}, \leq)$ sice platí třeba $(1, 13) \leq (3, 23)$, ale neuměli bychom porovnat řekně vektory $(1, 2)$ a $(2, 1)$. Řečeno matematicky, součinové uspořádání bývá zřídka lineární. To bývá v mnoha aplikacích zásadní problém.

Chce to tedy jiný nápad. Jako inspirace poslouží uspořádání, které se používá ve slovnících, my totiž na základě schopnosti srovnat jednotlivé „souřadnice“, tj. individuální písmena, umíme porovnat pořadí libovolných dvou slov. To vypadá nadějně.

Definice.

Uvažujme částečně uspořádané množiny $(A_1, \preceq_1), \dots, (A_n, \preceq_n)$. Definujeme **lexikografické uspořádání (lexicographic ordering)** \preceq_L na $A = A_1 \times \dots \times A_n$ následovně: Pro $a = (a_1, \dots, a_n), b = (b_1, \dots, b_n) \in A$ platí $a \preceq_L b$ právě tehdy, jestliže $a_i = b_i$ pro všechna $i = 1, \dots, n$ (tedy $a = b$), nebo existuje index k takový, že $a_i = b_i$ pro všechna i splňující $1 \leq i < k$ a $a_k \prec_k b_k$.

V některých situacích je praktičtějšíjiná definice, kdy se nejprve zavede „ostrá nerovnost“ na A , značená tradičně $(a_1, \dots, a_n) \prec_L (b_1, \dots, b_n)$, podmínkou „existuje index k takový, že $a_i = b_i$ pro všechna i splňující $1 \leq i < k$ a $a_k \prec_k b_k$ “. Toto je asymetrická a tranzitivní relace, z ní se pak \preceq_L udělá již standardním způsobem, podmínkou „ $a \prec_L b$ nebo $a = b$ “. Naopak pokud z \preceq_L vyrobíme standardní odvozenou relaci \prec , tak to bude ta ze začátku tohoto odstavce (viz Věta 4b.5).

Co to znamená v praxi? Máme-li porovnat dva prvky, tak je začneme porovnávat po souřadnicích, začneme od první a ignorujeme je, dokud jsou stejné. Jakmile narazíme na rozdílné souřadnice, použijeme příslušné srovnání k rozhodnutí. Není to nic nového, asi nás nepřekvapí, že ve slovníku je „sada“ dříve než „salát“. Ignorujeme shodné první znaky „sa“ a ten další rozhodl, ostatní jsme pak také ignorovali. Než se pustíme do podrobnějšího zkoumání, potvrďme si, že lexikografické uspořádání je opravdu uspořádání.

Věta 4c.19.

Uvažujme částečně uspořádané množiny $(A_1, \preceq_1), \dots, (A_n, \preceq_n)$. Pak je $A = A_1 \times \dots \times A_n$ spolu s lexikografickým uspořádáním \preceq_L částečně uspořádaná množina.

Důkaz (z povinnosti): Reflexivita: Vezměme libovolné $a = (a_1, \dots, a_n) \in A$. Pak $a = a$, proto $a \preceq_L a$.

Antisimetrie: Vezměme libovolné $a = (a_1, \dots, a_n), b = (b_1, \dots, b_n) \in A$ takové, že $a \preceq_L b$ a $b \preceq_L a$. Z toho prvního srovnání máme podle definice dvě možnosti. Jedna je, že $a = b$, čímž máme, co potřebujeme, a důkaz antisimetrie je hotov. Ukážeme, že druhá možnost $a \neq b$ nastat nemůže.

Takže předpokládejme, že $a \neq b$, pak z $a \preceq_L b$ máme $a \prec_L b$ a existuje k takové, že $a_i = b_i$ pro $i < k$ a $a_k \prec_L b_k$. Podobně máme $b \prec_L a$ a existuje l splňující $a_i = b_i$ pro $i < l$ a $b_l \prec_L a_l$. Když $a_i = b_i$ pro $i < k$ ale $a_l \neq b_l$, tak nutně $k \leq l$, symetrickým argumentem pak odvodíme, že musí být $k = l$. Máme tedy $a_k \prec_L b_k$ a zároveň $b_k \prec_L a_k$, což je ve sporu s asymetrií \prec_L . Tato varianta tedy vůbec nemůže nastat, jediná možnost je ta, kdy se rovnají všechny souřadnice.

Tranzitivita: Vezměme libovolné $a = (a_1, \dots, a_n), b = (b_1, \dots, b_n), c = (c_1, \dots, c_n) \in A$ takové, že $a \preceq_L b$ a $b \preceq_L c$. Jestliže $a = b$, pak $b \preceq_L c$ je vlastně $a \preceq_L c$ a jsme hotovi. Jestliže $b = c$, pak $a \preceq_L b$ je vlastně $a \preceq_L c$ a zase jsme hotovi.

Nejjednodušší případ je ten poslední, když $a \neq b$ a $b \neq c$. Pak podle definice existuje k takové, že $a_i = b_i$ pro $i < k$ a $a_k \prec_L b_k$, a také existuje l splňující $b_i = c_i$ pro $i < l$ a $b_l \prec_L c_l$. Nechť $m = \min(k, l)$. Pak pro $i < m$ máme $a_i = b_i = c_i$.

Co platí pro m ? Jsou dvě možnosti, podle toho, jestli je to minimum rovno k nebo l (nevylučují se, v případě $m = k = l$ budou platné oba argumenty, takže to nevadí). Jak vypadá případ $m = k \leq l$? Podle předpokladu $a_k \prec_L b_k$, tedy $a_m \prec_L b_m$. Co bude s l ? Pro $m < l$ máme $b_m = c_m$, pro $m = l$ máme $b_m \prec_L c_m$, každopádně $b_m \preceq_L c_m$. Jsme tedy v situaci $a_m \prec_L b_m \preceq_L c_m$ a Lemma 4b.4 (iii) dává $a_m \prec_L c_m$, což spolu se závěrem předchozího odstavce dává $a \preceq_L c$. Podobně to dopadne, pokud $m = l \leq k$. □

V typickém případě jsou všechny množiny A_i stejné a porovnáváme prvky z $A \times \dots \times A$, kde máme jen jedno uspořádání.

Příklad 4c.n: Uvažujme \mathbb{Z}^6 s lexikografickým uspořádáním daným relací \leq . Pak máme například $(1, 2, 3, 9, 9, 9) \preceq_L (1, 2, 4, 1, 1, 1)$ a také $(1, 2, 3, 9, 9, 9) \prec_L (1, 2, 4, 1, 1, 1)$, rozhodla třetí souřadnice a další už nehrály roli.

△

Příklad 4c.o: Co vše je menší než $(3, 4)$ v množině $(\mathbb{N}, \leq) \times (\mathbb{N}, \leq)$ uspořádané lexikograficky?

Aby platilo $(x, y) \prec_L (3, 4)$, tak bud' musí rozhodnout první souřadnice, tedy musí být $x = 1$ či $x = 2$ a zbytek libovolný, nebo rozhoduje až druhá souřadnice. Proto

$$\{(x, y) \in \mathbb{N}^2; (x, y) \preceq_L (3, 4)\} = \{(1, y); y \in \mathbb{N}\} \cup \{(2, y); y \in \mathbb{N}\} \cup \{(3, 1), (3, 2), (3, 3)\}.$$

△

Teď si potvrdíme, v čem je hlavní výhoda lexikografického uspořádání.

Věta 4c.20.

Nechť $(A_1, \preceq_1), \dots, (A_n, \preceq_n)$ jsou částečně uspořádané množiny, uvažujme $A = A_1 \times \dots \times A_n$ spolu s lexikografickým uspořádáním \preceq_L .

(i) Jestliže jsou všechny (A_i, \preceq_i) lineárně uspořádané, tak je i (A, \preceq_L) lineárně uspořádaná.

(ii) Jestliže jsou všechny (A_i, \preceq_i) dobře uspořádané, tak je i (A, \preceq_L) dobře uspořádaná.

Důkaz (drsný, poučný): (i): Předpokládejme, že všechny (A_i, \preceq_i) jsou lineárně uspořádané. Vezměme libovolné $a = (a_1, \dots, a_n), b = (b_1, \dots, b_n) \in A$. Bud' jsou si rovny, pak $a \preceq_L b$ a tyto prvky jsou porovnatelné. Nebo existují j taková, že $a_j \neq b_j$, nechť k je z nich nejmenší. Pak $a_i = b_i$ pro $i < k$ a $a_k \neq b_k$. Protože je (A_k, \preceq_k) lineární, tak určitě buď $a_k \preceq_k b_k$ nebo $b_k \preceq_k a_k$. V prvním případě pak máme $a \preceq_L b$, v druhém $b \preceq_L a$. Lexikografické uspořádání je tedy lineární.

(ii): Důkaz provedeme indukcí na n .

(0) Pro $n = 1$ máme jednu množinu, jejíž součin se sebou je zase ona, tedy je dobře uspořádaná.

(1) Zvolme nějaké $n \in \mathbb{N}$ a předpokládejme, že kartézský součin libovolných n dobře uspořádaných množin je dobrý v lexikografickém uspořádání. Teď mějme $n + 1$ dobře uspořádaných množin $(A_1, \preceq_1), \dots, (A_{n+1}, \preceq_{n+1})$ a jejich kartézský součin A s lexikografickým uspořádáním \preceq_L .

Uvažujme $\widehat{M}_1 = \{a_1; (a_1, \dots, a_{n+1}) \in M\}$ (ze všech prvků M si vytáhneme první souřadnici). To je neprázdná podmnožina A_1 , takže podle předpokladu musí existovat její nejmenší prvek m_1 . Máme tedy prvek takový, že kdykoliv $(a_1, \dots, a_{n+1}) \in M$, pak $m_1 \preceq_1 a_1$. Navíc tento prvek pochází z nějakého $(m_1, a_2, \dots, a_{n+1}) \in M$. Nechť

$$M_1 = \{(a_2, a_3, \dots, a_{n+1}) \in M; (m_1, a_2, a_3, \dots, a_{n+1}) \in M \wedge a_1 = m_1\}.$$

Zde bereme všechny prvky z M , jejichž první souřadnice je m_1 , tuto první souřadnici vynecháme a zbyvající dáme do M_1 . Pak M_1 je neprázdná podmnožina kartézského součinu $A_2 \times \cdots \times A_{n+1}$. To je n množin, proto podle indukčního předpokladu jde o dobře uspořádanou množinu vzhledem k příslušnému lexikografickému uspořádání, tudíž tato množina musí mít nejmenší prvek (m_2, \dots, m_{n+1}) . Nechť $m = (m_1, m_2, \dots, m_{n+1})$. Tvrdíme, že jde o nejmenší prvek M vzhledem k lexikografickému uspořádání na A .

Vezměme tedy libovolné $a = (a_1, \dots, a_{n+1}) \in M$. Protože $a_1 \in \widehat{M}_1$, musí podle volby m_1 platit $m_1 \preceq_1 a_1$. Pokud platí dokonce $m_1 \prec_1 a_1$, pak již $m \preceq_L a$ podle definice lexikografického uspořádání (rozhodla hned první souřadnice).

Pokud $m_1 \prec_1 a_1$ neplatí, tak nutně $m_1 = a_1$. Pak ovšem $(a_2, \dots, a_{n+1}) \in M_1$ a tudíž $(m_2, \dots, m_{n+1}) \preceq (a_2, \dots, a_{n+1})$ v lexikografickém uspořádání $A_2 \times \cdots \times A_{n+1}$. To zase nabízí dvě možnosti.

Jedna je, že $m_i = a_i$ pro všechna $i = 2, \dots, n+1$, ale my teď předpokládáme i $m_1 = a_1$, tedy $m_i = a_i$ pro všechna i a máme $m \preceq_L a$. Druhá je, že existuje nějaké k takové, že $m_i = a_i$ pro $i = 2, \dots, k-1$ a $m_k \prec_k a_k$. To ale znamená, že $m_i = a_i$ pro $i = 1, \dots, k-1$ a $m_k \prec_k a_k$, tedy $m \preceq_L a$. Důkaz hotov. \square

Naopak se dá ukázat, že stačí, aby jen jediná složka takového kartézského součinu nebyla lineárně či dobře uspořádaná, a už to pokazí pro celý součin, například $(\mathbb{Z}, \leq) \times (\mathbb{N}, \leq)$ s lexikografickým uspořádáním není dobře uspořádaná.

Stojí za zmínku, že pro analýzu ani lexikografické uspořádání není tím pravým, sice porovnávat vektory o stejné délce. Ve slovníku ale porovnáváme slova různých délek. Jak by se taková definice udělala obecně? Základní myšlenka je, že když máme dva vektory nestejně délky, tak ten delší zkrátíme. Pro zjednodušení zápisu budeme předpokládat, že všechny složky pocházejí z jedné množiny.

Z pohledu computer science je nicméně možnost uspořádávat složitější množiny lexikograficky velice užitečná, v diskrétní matematice se s ním lze setkat často.

Příklad 4c.p: Lexikografické uspořádání, které jsme definovali, nám umožňuje porovnávat vektory o stejné délce. Ve slovníku ale porovnáváme slova různých délek. Jak by se taková definice udělala obecně? Základní myšlenka je, že když máme dva vektory nestejně délky, tak ten delší zkrátíme. Pro zjednodušení zápisu budeme předpokládat, že všechny složky pocházejí z jedné množiny.

Mějme uspořádanou množinu (A, \preceq) a uvažujme množinu $B = \bigcup_{n=1}^{\infty} A^n$. Nejprve definujeme ostrou relaci takto:

Pro $(a_i)_{i=1}^m, (b_j)_{j=1}^n \in B$ nechť $k = \min(m, n)$. Platí $(a_i)_{i=1}^m \prec (b_j)_{j=1}^n$ právě tehdy, když buď $(a_i)_{i=1}^k \prec_L (b_i)_{i=1}^k$, nebo $(a_i)_{i=1}^k = (b_i)_{i=1}^k$ a $m < n$.

Pak standardním způsobem definujeme $(a_i)_{i=1}^m \preceq (b_j)_{j=1}^n$ jestliže buď $(a_i)_{i=1}^m \prec (b_j)_{j=1}^n$ nebo $(a_i)_{i=1}^m = (b_j)_{j=1}^n$.

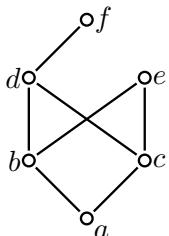
Tímto dostaneme částečné uspořádání na B . Není problém tuto definici převést na uspořádání pro řetězce, prostě se řetězce považují za vektory, například bereme *ahoj* jako (a, h, o, j) . Na abecedě máme přirozené částečné uspořádání a právě popsanou procedurou z něj dostaneme obvyklé slovníkové uspořádání, takže třeba $\text{auto} \preceq \text{autobus} \preceq \text{autobusek} \preceq \text{auvajs}$.

\triangle

Cvičení

Cvičení 4c.1 (rutinní): Najděte dva neporovnatelné prvky v $A = \{2, 4, 6, 8\}$ uspořádané dělitelností a v $(P(\{\diamond, \bullet, \odot\}), \subseteq)$.

Cvičení 4c.2 (rutinní): Uvažujte uspořádanou množinu danou následujícím Hasseovým diagramem.



Najděte maximum, minimum, největší a nejmenší prvek množiny $M = \{a, b, c, d, e\}$, pokud existují.

Cvičení 4c.3 (rutinní): Uvažujte uspořádanou množinu $(\{3, 5, 9, 15, 24, 45\}, |)$, tedy relace dělitelnosti.

(i) Nakreslete její Hasseův diagram.

(ii) Najděte její maxima, minima, největší a nejmenší prvek, pokud existují.

(iii) Najděte maxima, minima, největší a nejmenší prvek podmnožiny $M = \{3, 9, 15\}$, pokud existují.

Cvičení 4c.4 (rutinní): Uvažujte množinu množin $\mathcal{A} = \{\emptyset, \{1\}, \{2\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\}$ uspořádanou relací býti podmnožinou (viz cvičení 4b.8).

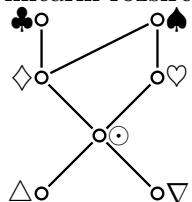
(i) Najděte maximum, minimum, největší a nejmenší prvek množiny $M = \{\{1\}, \{2\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 3, 4\}\}$, pokud existují.

(ii) Najděte nějaké lineární rozšíření (A, \subseteq) .

Cvičení 4c.5 (rutinní): Nakreslete Hasseův diagram nějaké konečné uspořádané množiny, která

- (i) nemá největší prvek, má nejmenší prvek;
- (ii) má největší prvek, nemá nejmenší prvek;
- (iii) nemá největší prvek, nemá nejmenší prvek.

Cvičení 4c.6 (rutinní): Uvažujte uspořádání dané následujícím Hasseovým diagramem. Najděte nějaké jeho lineární rozšíření.



Cvičení 4c.7 (rutinní): Uvažujme částečně uspořádanou množinu $A = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7\}$, jejíž uspořádání je dáno následujícím odvozeným ostrým uspořádáním: $p_4 \prec p_7, p_6 \prec p_7, p_5 \prec p_6, p_2 \prec p_6, p_1 \prec p_4, p_2 \prec p_4, p_3 \prec p_4, p_1 \prec p_2$ a $p_3 \prec p_2$.

Sestavte Hasseův diagram pro tuto uspořádanou množinu.

Cvičení 4c.8 (rutinní): Uvažujte množinu $A = \{2, 3, 10, 30, 60, 90\}$ uspořádanou relací dělitelnosti $a | b$. Najděte nějaké její lineární rozšíření.

Cvičení 4c.9 (rutinní): Uvažujte množinu vektorů $A = \{(3, 23), (13, 23), (23, 3), (23, 13), (23, 23)\}$ uspořádanou součinovým uspořádáním založeným na \geq , tedy $(u, v) \preceq (x, y)$ jestliže $u \geq x$ a $v \geq y$. Najděte nějaké její lineární rozšíření.

Cvičení 4c.10 (rutinní): Která z následujících množin je dobře uspořádaná?

- | | |
|---|---|
| (i) $(\{n \in \mathbb{Z}; n \geq -13\}, \leq)$ | (v) (\mathbb{Q}, \leq) |
| (ii) $(\{n \in \mathbb{Z}; n > -13\}, \leq)$ | (vi) (\mathbb{Q}^+, \leq) |
| (iii) $(\{n \in \mathbb{Z}; n > -13\}, \geq)$ | (vii) $(\{x \in \mathbb{Q}^+; x = \frac{p}{q}, p, q \in \mathbb{N}, q < 100\}, \leq)$ |
| (iv) $(\{n \in \mathbb{Z}; n \text{ sudé}\}, \leq)$ | |

Cvičení 4c.11 (rutinní): Nechť (A, \preceq) je částečně uspořádaná množina. Dokažte, že \preceq je lineární právě tehdy, když je \preceq^{-1} lineární.

Cvičení 4c.12 (rutinní): Nechť $A = \{a, b, c\}$, uspořádejme ji podle abecedy. Uvažujte $A \times A$ s lexikografickým uspořádáním. Najděte všechny prvky z $A \times A$, které jsou vzhledem k lexikografickému uspořádání

- | | | |
|--------------------------|---------------------------|----------------------------|
| (i) menší než (a, c) ; | (ii) větší než (a, c) ; | (iii) menší než (b, b) . |
|--------------------------|---------------------------|----------------------------|

Cvičení 4c.13 (rutinní): Uspořádejte podle lexikografického uspořádání binární řetězce 0, 01, 11, 001, 010, 011, 0001, 0101.

Cvičení 4c.14 (rutinní): (i) Nakreslete Hasseův diagram pro $(\{1, 2, 3\}, \leq)^2$ v lexikografickém uspořádání.
(ii) Nakreslete Hasseův diagram pro $(\{1, 2, 3\}, \leq)^2$ v součinovém uspořádání („po složkách“).

Řešení:

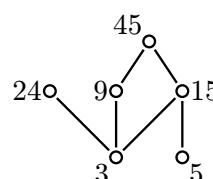
4c.1: Třeba 4 a 6. Třeba $\{\bullet\}$ a $\{\odot\}$.

4c.2: Max d, e , největší neex., min a , nejmenší a .

4c.3:

(ii): Max 24,45, největší neex., min 3,5, nejmenší neex.

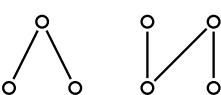
(iii): Max 9,15, největší neex., min 3, nejmenší 3.



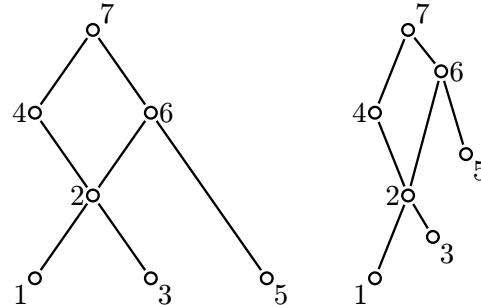
4c.4: (i): Max $\{1, 2, 3, 4\}$, největší $\{1, 2, 3, 4\}$, min $\{1\}$, $\{2\}$, nejmenší neex.

(ii): Třeba $\emptyset \prec_L \{1\} \prec_L \{2\} \prec_L \{1, 2, 4\} \prec_L \{1, 2, 3\} \prec_L \{1, 2, 3, 4\} \prec_L \{1, 2, 3, 4, 5\}$.

4c.5:

4c.6: Například $\triangle \prec_L \nabla \prec_L \odot \prec_L \diamond \prec_L \heartsuit \prec_L \clubsuit \prec_L \spadesuit$ nebo $\nabla \prec_L \triangle \prec_L \odot \prec_L \diamond \prec_L \heartsuit \prec_L \clubsuit \prec_L \spadesuit$ nebo ...

4c.7: Nalevo Hasseův diagram sestrojený dle algoritmu, napravo jeho topologická úprava, která odpovídá linearizaci v příkladu 4c.h.

4c.8: Například $3 \prec_L 2 \prec_L 10 \prec_L 30 \prec_L 90 \prec_L 60$ nebo $2 \prec_L 10 \prec_L 3 \prec_L 30 \prec_L 60 \prec_L 90$ nebo dle velikosti nebo ...4c.9: Třeba $(23, 23) \prec_L (23, 13) \prec_L (23, 3) \prec_L (13, 23) \prec_L (3, 23)$ nebo $(23, 23) \prec_L (13, 23) \prec_L (3, 23) \prec_L (23, 13) \prec_L (23, 3)$ nebo $(23, 23) \prec_L (13, 23) \prec_L (23, 13) \prec_L (3, 23) \prec_L (23, 3)$ nebo ...

4c.10: (i), (ii), (vii).

Re: (vii): V množině je omezená velikost jmenovatele, proto uvažované body nejsou husté, ale jednotlivé body množiny jsou od sebe vzdáleny. Protože $p, q > 0$, jde o kladná čísla a nemohou utéci do mínus nekonečna. Pro libovolné $K > 0$ je množina $\{(p, q) \in \mathbb{N}^2; q < 100 \wedge p \leq Kq\}$ konečná, mezi odpovídajícími zlomky $\frac{p}{q}$ tedy vždy najdeme nejmenší.

4c.11: \Rightarrow : Nechť \preceq lineární. Zvolme $a, b \in A$. Pak $a \preceq b$ (potom $b \preceq^{-1} a$) nebo $b \preceq a$ (potom $a \preceq^{-1} b$), každopádně jsou a, b porovnatelné pomocí \preceq^{-1} .

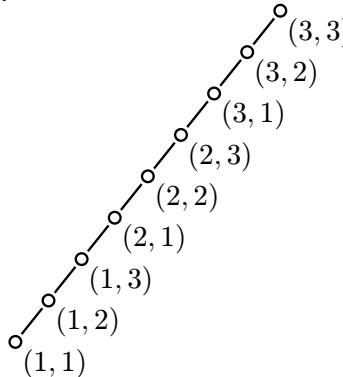
\Leftarrow : Obdobně.

4c.12: (i): $(a, a), (a, b);$ (ii): $(b, a), (b, b), (b, c), (c, a), (c, b), (c, c);$ (iii): $(a, a), (a, b), (a, c), (b, a).$

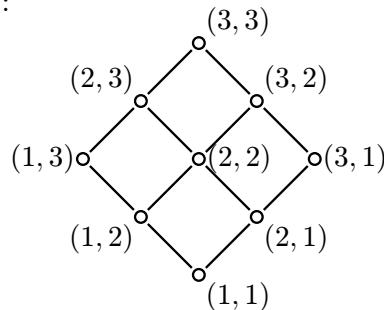
4c.13: 0,0001,001,01,010,0101,011,11

4c.14:

(i):



(ii):



4d. Bonus: Další pojmy okolo uspořádání

Když jsme hledali nejmenší a největší prvky, měli jsme dva problémy. Jednou překážkou byl nedostatek vzájemné porovnatelnosti, ten jsme vyřešili pojmem lineárního zobrazení. Druhý problém nastal, když nám množina „utíkala“ jako třeba \mathbb{Z} při porovnávání pomocí \leq . Zatím jsme to řešili tak, že jsme se omezili na konečné množiny, teď utíkání zkusíme obecně pojmenovat a pak zakázat; uvídíme, jestli dostaneme něco rozumného.

Definice.

Nechť (A, \preceq) uspořádaná množina. Řekneme, že je **fundovaná (well-founded)**, jestliže neexistuje „nekonečná klesající posloupnost“, tj. nekonečná posloupnost $\{a_i\}_{i=1}^{\infty}$ prvků z A taková, že $a_{i+1} \prec a_i$ pro všechna i .

Problémem pro fundovanost jsou tedy posloupnosti typu $a_1 \succ a_2 \succ a_3 \succ \dots \succ a_i \succ a_{i+1} \succ \dots$. Pomocí tranzitivnosti dostáváme, že pro všechna $j > i$ máme $a_j \prec a_i$, ostrá nerovnost také znamená, že jde o navzájem různé prvky. U konečných množin nekonečně mnoho různých prvků nemáme, z toho hned vyplývá, že konečné uspořádané množiny jsou automaticky fundované.

Není pravda, že pokud zabráníme takovým posloupnostem, tak automaticky dostáváme existenci nejmenších prvků, tedy dobré uspořádání, protože může zlobit ta porovnatelnost.

Příklad 4d.a: Uvažujme uspořádanou množinu (\mathbb{Z}, \preceq) , kde $x \preceq y$ jestliže $x = y$ nebo $|x| < |y|$.

Tvrdíme, že je to částečné uspořádání, které je well-founded, ale není lineární, tím spíše ne dobré.

Reflexivita je jasná, tranzitivita v zásadě také, jak je na tom antisimetrie? Předpokládejme, že $x \preceq y$ a $y \preceq x$. Kdyby nebylo $x = y$, tak by z definice muselo platit $|x| < |y|$ a $|y| < |x|$, což není možné. Máme tedy uspořádání.

Uvažujme teď nějakou nekonečnou posloupnost $\{a_i\}_{i=1}^{\infty}$ ze \mathbb{Z} takovou, že $a_{i+1} \preceq a_i$ pro všechna i . Tvrdíme, že nemůže nastat případ, že by všechna ta srovnání byla ostrá, tedy \prec . Uvažujme množinu $M = \{|a_i|; i \in \mathbb{N}\}$. To je neprázdná podmnožina \mathbb{N}_0 , což je vzhledem k \leq dobře uspořádaná množina. Existuje tedy j takové, že $|a_j|$ je nejmenší prvek z M . To znamená, že $|a_j| \leq |a_i|$ pro všechna i . Tvrdíme, že $a_i = a_j$ pro všechna $i > j$.

Z tranzitivity máme pro $i > j$ relaci $a_i \preceq a_j$, teď jsou dvě možnosti. Jedna je $|a_i| < |a_j|$, ale to nejde, takže musí platit ta druhá, $a_i = a_j$.

Dokázali jsme, že (\mathbb{Z}, \preceq) nemá nekonečné klesající posloupnosti, je tedy fundovaná.

Proč není lineárně uspořádaná? Protože třeba prvky 13 a -13 tímto uspořádáním neumíme porovnat.

△

Dobré uspořádání je definováno pomocí existence nejmenších prvků. Pro fundovaná uspořádání existuje alternativní charakterizace podobného typu.

Věta 4d.1.

Nechť je (A, \preceq) částečně uspořádaná množina. Toto uspořádání je fundované právě tehdy, když pro každou neprázdnou podmnožinu $M \subseteq A$ existuje $\min(M)$.

Důkaz (poučný, náznak): 1) \implies : Předpokládejme, že \preceq je fundované. Nechť M je neprázdná podmnožina A . Vezměme libovolné $a_1 \in M$. Pokud je to minimální prvek, jsme hotovi. Pokud ne, musí existovat $a_2 \in M$ splňující $a_2 \prec a_1$. Pokud je to minimální prvek, jsme hotovi. Pokud ne, musí existovat $a_3 \in M$ splňující $a_3 \prec a_2$. Pokračujeme tak dále, a protože není možná nekonečná klesající posloupnost, tak se tento proces musí zastavit, tedy najdeme minimální prvek M .

2) \impliedby : Fundovanost dokážeme sporem. Nechť existuje nekonečná klesající posloupnost $\{a_i\}_{i=1}^{\infty}$. Označme $M = \{a_i; i \in \mathbb{N}\}$. To je neprázdná podmnožina A , proto podle předpokladu existuje její minimální prvek, třeba a_i . Máme $a_{i+1} \in M$, z minimality a_i proto $a_i \preceq a_{i+1}$, což je ale ve sporu s $a_{i+1} \prec a_i$, viz Lemma 4b.4 (ii). □

Možná máte pocit, že slova „proces musí zastavit“ nejsou zrovna korektním matematickým vyjádřením. A máte pravdu, korektní důkaz se musí dělat podrobněji a je citelně delší, dokonce dojde na axiom výběru, takže je to úrovní znatelně výš než toto skriptum. Však jsme v úvodu důkazu vyhrožovali, že je to jen náznak. Ten základní nápad ale myslím stál za přečtení, s podobnou myšlenkou se ještě potkáme v kapitole o indukci.

Teď ukážeme, že když se postaráme o porovnatelnost a zakážeme utíkání, už nejmenší prvky najdeme.

Věta 4d.2.

Uvažujme částečně uspořádanou množinu (A, \preceq) .

Tato množina je dobré uspořádaná právě tehdy, když je lineárně uspořádaná a fundovaná.

Důkaz (poučný): 1) \implies : Podle Faktu 4c.13 víme, že dobré uspořádané množiny jsou automaticky lineárně uspořádané, tedy pro všechny neprázdné podmnožiny poskytují nejmenší prvky, což jsou podle Věty 4c.2 (iii) minima.

2) \impliedby : Nechť M je neprázdná podmnožina A . Podle předchozí věty nám fundovanost dává minimální prvek této množiny. Protože je uspořádání lineární, tak je podle Věty 4c.7 toto minimum i nejmenším prvkem. □

Pojem fundovanosti se dá použít i pro relace, které nejsou částečnými uspořádáními. Zejména populární je tento pojem pro ostrá uspořádání neboli přímo pro relace typu \prec , viz Věta 4b.5. Zajímavou souvislost s indukcí lze nalézt v poznámce za Větu 5a.13.

Kromě maxim, nejmenších prvků a podobně jsou ještě další příbuzné pojmy. Toto je již mimo rámec skripta, protože tím začíná samostatný obor, tak jen naznačíme pár věcí jako lákadlo.

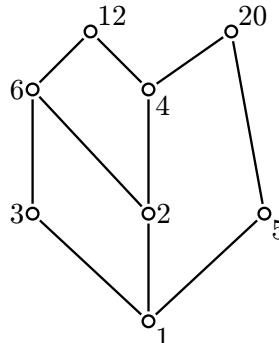
Definice.

Nechť (A, \preceq) je částečně uspořádaná množina, M je její neprázdná podmnožina.

Prvek $m \in A$ se nazývá **horní mez (upper bound)** množiny M , jestliže $x \preceq m$ pro všechna $x \in M$.

Prvek $m \in A$ se nazývá **dolní mez (lower bound)** množiny M , jestliže $m \preceq x$ pro všechna $x \in M$.

Příklad 4d.b: Uvažujme množinu $A = \{1, 2, 3, 4, 5, 6, 12, 20\}$ uspořádanou dělitelností (viz příklad 4b.f). Měli jsme pro ni následující Hasseův diagram.



Podívejme se na množinu $M = \{4, 5, 6\}$. Číslo 1 dělí všechna čísla z M , je tedy dolní mez M . Neexistuje ale číslo z A takové, že by jej dělila všechna čísla z M , proto tato množina nemá horní mez.

△

V Hasseově diagramu hledáme dolní meze tak, že vyjdeme ze všech prvků M po spojnicích dolů a čekáme, zda se někde všechny cesty sejdou. Při hledání horní meze naopak hledáme prvek, ve kterém by se sešly cesty z prvků M směrem nahoru.

Jak vidíme, existence mezí není zaručena, na druhou stranu jich může existovat i více. Je zde důležité si všimnout jednoho podstatného rozdílu oproti minimům a spol., při hledání mezí se díváme i mimo množinu M , tudíž již záleží na tom, jaká je nadmnožina, ve které pracujeme. Například v předchozím příkladě jsme nenašli horní mez, ale kdybychom tutéž množinu M uvažovali třeba v uspořádané množině $(\{1, 2, 3, 4, 5, 6, 12, 20, 60\}, |)$, tak už by prvek 60 byl horní mezí pro $\{4, 5, 6\}$. Nelze tedy aplikovat trik, který nám celou dobu věrně sloužil, že stačí pracovat s uspořádanými množinami a umíme to již i s podmnožinami.

Příklad 4d.c: Uvažujme množinu (\mathbb{N}, \leq) . Pak libovolné číslo $n \in \mathbb{N}$ splňující $n > 13$ je horní mezí množiny $M = \{3, 13\}$.

△

Fakt 4d.3.

Nechť (A, \preceq) je částečně uspořádaná množina a M její neprázdná podmnožina.

Jestliže je m největší prvek M , pak je to i jeho horní mez.

Jestliže je m nejmenší prvek M , pak je to i jeho dolní mez.

Důkaz je jasný, například nejmenší prvek m splňuje $m \preceq x$ pro všechna $x \in M$, což je přesně definice dolní meze. Souvislost mezi těmito dvěma pojmy lze vidět i jinak.

Fakt 4d.4.

Nechť (A, \preceq) je částečně uspořádaná množina a M její neprázdná podmnožina.

Horní mez množiny M existuje právě tehdy, pokud existuje nějaká nadmnožina $N \subseteq A$ množiny M taková, že N má největší prvek.

Dolní mez množiny M existuje právě tehdy, pokud existuje nějaká nadmnožina $N \subseteq A$ množiny M taková, že N má nejmenší prvek.

Důkaz (poučný): Dokážeme první tvrzení, druhé funguje symetricky.

1) Jestliže je m horní mez množiny M , pak uvažujme $N = M \cup \{m\}$. Protože je m horní mezí, platí $x \preceq m$ pro všechna $x \in M$, také $m \preceq m$, proto $x \preceq m$ pro všechna $x \in N$, také $m \in N$, tedy m je největší prvek N .

2) Naopak nechť existuje množina N taková, že $M \subseteq N$ a N má největší prvek m . Pak pro libovolné $x \in M$ platí i $x \in N$ a m je největší v N , tedy $x \preceq m$. m je tedy horní mezí M .

□

Plyne z toho například to, že v konečných lineárně uspořádaných množinách vždy najdeme horní a dolní mez. Pokud horní a dolní meze existují, hledáme mezi nimi některé speciální.

Definice.

Nechť (A, \preceq) je částečně uspořádaná množina, M je její neprázdná podmnožina.

Jestliže existuje horní mez M , pak definujeme **supremum** množiny M , také zvané **nejmenší horní mez (least upper bound)**, značeno $\sup(M)$ nebo l. u. b.(M), jako nejmenší prvek množiny $\{x \in A; x \text{ horní mez } M\}$, pokud existuje.

Jestliže existuje dolní mez M , pak definujeme **infimum** množiny M , také zvané **největší dolní mez (greatest lower bound)**, značeno $\inf(M)$ nebo g. l. b.(M), jako největší prvek množiny $\{x \in A; x \text{ dolní mez } M\}$, pokud existuje.

Podobně jako u mezí, existuje-li největší prvek množiny M , pak je jejím supremem, symetricky nejmenší prvek je infimum. Proto nám linearita uspořádání zaručí existenci infim a suprem pro konečné množiny, jinak ale nemusí existovat.

Příklad 4d.d: Uvažujme (\mathbb{N}, \leq) . To je dokonce dobře uspořádaná množina, ale množina M sudých přirozených čísel nemá ani horní mez, ani supremum.

Příklad 4d.e: Uvažujme množinu $\mathcal{A} = \{\{a\}, \{b\}, \{a, b, c\}, \{a, b, d\}\}$ uspořádanou inkluzí. Nechť $\mathcal{M} = \{\{a\}, \{b\}\}$ je její podmnožina. Pak má tato množina horní meze $\{a, b, c\}$ i $\{a, b, d\}$, protože obě množiny z \mathcal{M} jsou podmnožinami těchto dvou, ale \mathcal{M} nemá supremum, protože ony dvě horní meze jsou neporovnatelné a tudíž neumíme najít nejmenší prvek množiny $\{\{a, b, c\}, \{a, b, d\}\}$.

△

Pro množiny, u kterých hledání suprem a infim není až tak velký problém, máme speciální jméno.

Definice.

Uvažujme uspořádanou množinu (A, \preceq) . Řekneme, že je to **svaz (lattice)**, jestliže pro všechna $x, y \in A$ existují $\sup(\{x, y\})$ a $\inf(\{x, y\})$.

Příklad 4d.f: Množina $A = \{1, 2, 3, 4\}$ uspořádaná dělitelností není svaz, protože například $M = \{2, 3\}$ nemá v A ani horní mez, natož supremum.

△

Příklad 4d.g: Nechť U je množina, uvažujme $(P(U), \subseteq)$. Tato uspořádaná množina je svaz.

Stačí si rozmyslet, že jsou-li dány $M, N \in P(U)$, tedy jsou to podmnožiny U , pak $\inf(\{M, N\}) = M \cap N \in P(U)$ a $\sup(\{M, N\}) = M \cup N \in P(U)$.

Zkusíme třeba to infimum. Evidentně $M \cap N \subseteq M$ a $M \cap N \subseteq N$, takže je to dolní mez. Je největší? Nechť P je jiná dolní mez. Pak splňuje $P \subseteq M$ a $P \subseteq N$, proto i $P \subseteq M \cap N$, což jsme přesně potřebovali.

△

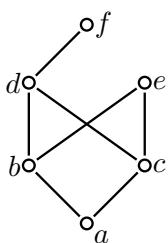
Příklad 4d.h: Množina \mathbb{N} uspořádaná dělitelností je svaz, $\sup(\{x, y\})$ najdeme jako nejmenší společný násobek, $\inf(\{x, y\})$ jako největší společný dělitel.

△

Svazy mají svou vlastní teorii a používají se třeba v práci s Booleovými algebrami či namátkou při modelování bezpečných informačních toků, kdy se zavádí uspořádání na různých stupních oprávnění.

Cvičení

Cvičení 4d.1 (rutinní): Uvažujte uspořádanou množinu danou následujícím Hasseovým diagramem.



Najděte pro množinu $M = \{b, c\}$ její horní meze a supremum, dolní meze a infimum, pokud existují.

Cvičení 4d.2 (rutinní): Uvažujte množinu množin $\mathcal{A} = \{\emptyset, \{1\}, \{2\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\}$ uspořádanou relací býti podmnožinou (viz cvičení 4b.8).

Pro $\mathcal{M} = \{\{1\}, \{2\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 3, 4\}\}$ najděte horní meze a supremum, dolní meze a infimum, pokud existují.

Řešení:

4d.1: Horní meze d, e, f , supremum neex., dolní mez a , infimum a .

4d.2: Horní meze $\{1, 2, 3, 4\}$ a $\{1, 2, 3, 4, 5\}$, supremum $\{1, 2, 3, 4\}$, dolní meze \emptyset , $\{1\}$ a $\{2\}$, infimum neex.

5. Indukce a rekurze

Na matematickou indukci jsme již v této knize několikrát narazili v důkazech. Je obtížné si představit matematiku bez tohoto silného nástroje, který si v této kapitole oficiálně představíme. Začneme oním jednoduchým typem důkazu indukcí, který už mnozí z čtenářů znají a který jsme zatím používali, a postupně se propracujeme k složitějším verzím, protože indukce znamená víc, než asi čtenář dosud tušil.

5a. Matematická indukce

Když student slyší „matematická indukce“, obvykle si představí důkaz, kterým ukazujeme pravdivost rozličných vzorečků. Začneme tím, že si přesně formulujeme, jak vlastně takový důkaz funguje.

! 5a.1. Slabý princip matematické indukce.

Nechť $n_0 \in \mathbb{Z}$, nechť $V(n)$ je vlastnost celých čísel, která má smysl pro $n \geq n_0$.

Předpokládejme, že jsou splněny následující předpoklady:

(0) $V(n_0)$ platí.

(1) Pro každé $n \in \mathbb{Z}$, $n \geq n_0$ je pravdivá následující implikace: Jestliže platí $V(n)$, pak platí i $V(n+1)$.

Potom $V(n)$ platí pro všechna $n \in \mathbb{Z}$, $n \geq n_0$.

Weak principle of mathematical induction. Let $n_0 \in \mathbb{Z}$. Let $V(n)$ be a property of integers that makes sense for $n \geq n_0$. Assume that the following conditions are satisfied:

(0) $V(n_0)$ is true.

(1) For every $n \geq n_0$ the following implication is valid: If $V(n)$ is true, then $V(n+1)$ is true.

Then $V(n)$ is true for all $n \geq n_0$.

The part (0) is called the base step, (1) is called the induction step.

Pokud tedy chceme dokázat univerzální platnost nějaké vlastnosti V , stačí dokázat pravdivost tvrzení (0) a (1). Proč to stačí? Půjčíme si představu žebříku. Tzv. **základní krok** (0) říká, že umíme vylezt na první příčku žebříku. Tzv. **indukční krok** (1) říká, že když už někde jsme, tak umíme vylezt o příčku výš. Podstatný je ten obecný kvantifikátor v (1), indukční krok je splněn pro libovolné místo na žebříku. Selský rozum říká, že pak už se dostaneme na žebříku všude.

Zkusme to matematicky, vezmeme pro jednoduchost $n_0 = 1$. Podle základního kroku platí $V(1)$. Indukční krok dává pro volbu $n = 1$ pravdivou implikaci $V(1) \implies V(2)$, my už ovšem ze základního kroku víme, že $V(1)$ platí, tudíž podle této implikace platí i $V(2)$. Pak zase můžeme použít indukční krok s $n = 2$, kde z pravdivosti $V(2)$ dostaneme pravdivost $V(3)$. Další použití indukčního kroku (s $n = 3$) dá pravdivost $V(4)$, pak $V(5)$ a tak dále. Člověk by si řekl, že se tohle nemůže pokazit, dojdeme takto libovolně daleko a V funguje pro všechna čísla. Princip výše potvrzuje, že tato představa je správná.

Jiná dobrá představa je padající domino. (1) říká, že kdykoliv nějaké domino spadne, tak spadne i to další. (0) říká, že jsme shodili to první. Zkušenosť říká, že pak by to mělo spadnout všechno. Zároveň vidíme, že nic nelze vynechat. Pokud nedokážeme (0), tak vlastně na začátku nic neshodíme a domino nepadají. Pokud není pravda, že (1) platí pro všechna relevantní n , tak se pro některé n padání domina zadrhne.

! **Příklad 5a.a:** Indukce byla na intuitivní úrovni používána již indickými a arabskými matematiky kolem 10. století. Poprvé byla přesně formulována Pascalem v roce 1665, ale intuitivní použití v Evropě sahá do 16. století. Údajně první byl F. Maurolico, který tak dokázal, že součet prvních n lichých čísel (myšleno kladných) je n^2 . Předvedeme na této úloze správný postup a okomentujeme jej. Jaké jsou jednotlivé kroky?

- Zformulujeme přesně tvrzení a oznámíme, jak jej dokážeme.

Pro $n \in \mathbb{N}$ je $V(n)$ tvrzení, že $1 + 3 + 5 + \dots + (2n - 1) = n^2$.

Dokážeme to pomocí matematické indukce.

Zde je dobré si rozmyslet, že $1 + 3 + \dots + (2n - 1)$ opravdu dává prvních n lichých (kladných) čísel.

- Dokážeme základní krok.

(0) Nechť $n = 1$. Vlastnost $V(1)$ zní $1 = 1$, což je pravda.

• Dokážeme indukční krok. Vezmeme libovolné $n \in \mathbb{N}$ (obecné, ne nějaké konkrétní) a dokážeme, že pro něj platí implikace $V(n) \implies V(n+1)$. To se typicky dělá přímým důkazem, takže předpokládáme, že pro naše zvolené n platí $V(n)$, tomu se říká „indukční předpoklad“. Pomocí něj pak dokážeme platnost $V(n+1)$. Základním trikem je při tom dekompozice, kdy se $V(n+1)$ pokusíme rozložit tak, aby se objevilo něco, na co lze aplikovat $V(n)$.

(1) Nechť $n \in \mathbb{N}$ je libovolné. Předpokládejme, že $1 + 3 + 5 + \dots + (2n - 1) = n^2$.

Chceme pomocí této rovnosti nějak ukázat, že pro naše konkrétní n platí také $1+3+5+\cdots+(2(n+1)-1) = (n+1)^2$, tedy že $1+3+5+\cdots+(2n+1) = (n+1)^2$. Obvykle bývá lepší začít tou delší či více komplikovanou stranou a upravit ji tak, aby se objevilo něco, co je i ve $V(n)$, pomocí tohoto předpokladu se pak propracujeme ke druhé straně ve $V(n+1)$.

Zde je to snadné, dokazovaná rovnost obsahuje $1+2+\cdots+(2n-1)$. Jdeme na to.

$$\begin{aligned} 1+3+5+\cdots+(2n+1) &= 1+3+5+\cdots+(2n-1)+(2n+1) \\ &= [1+3+5+\cdots+(2n-1)]+(2n+1) \\ &= n^2+(2n+1) = (n+1)^2. \end{aligned}$$

• *Uděláme závěr.*

Důkaz je hotov.

Když z toho důkazu výše vynecháme vysvětlující části psané kurzívou, dostaneme důkaz tak, jak se běžně zapisuje. Všimněte si, jak i u dalších důkazů zachováváme tuto strukturu.

△

S 5a.2 Poznámka: Začínající studenti jsou zvyklí dokazovat neznámé rovnosti tak, že si je napíší a pak upravují, dokud nedostanou něco známého. U předchozího příkladu by například napsali

$$\begin{aligned} 1+3+5+\cdots+(2n+1) &= (n+1)^2 \\ [1+3+5+\cdots+(2n-1)^2]+(2n+1) &= (n+1)^2 \\ n^2+(2n+1) &= (n+1)^2 \\ n^2+2n+1 &= n^2+2n+1 \\ 0 &= 0. \end{aligned}$$

Bohužel, toto není důkaz platnosti $V(n+1)$, ale důkaz platnosti rovnosti $0 = 0$, jde totiž špatným směrem. Navíc ani nejde o důkaz korektní, vychází totiž z rovnosti, jejíž platnost v té chvíli není známá. Správný důkaz samozřejmě vychází z něčeho, co je známé, a dojde k tomu, co chceme dokázat. V tomto případě vznikne správný důkaz tak, že ty řádky znova přepíšeme, ale v opačném pořadí. Při tom přepisování ale musíme pečlivě hlídat, zda kroky, které jsme předtím dělali, platí i v opačném „správném“ směru. U tohoto příkladu to platí, neboť všechny provedené úpravy byly ekvivalentní, ale ne vždy tomu tak je.

I v případě, že kroky obrátit jdou, jde o zbytečnou komplikaci, mnohem kratší je dokázat zadanou rovnost přímým výpočtem, tedy začít výrazem na jedné straně a postupně se propracovat k výrazu na druhé straně, přesně jak jsme to udělali v příkladě 5a.a. Nejen že je to kratší, zejména u nerovností jde často o jediný rozumný přístup (viz příklad 5a.e a poznámka 5a.4), proto jej doporučujeme.

Podrobněji o tomto problému pojednává poznámka 1b.6.

△

Příklad 5a.b: Než začneme, tak si rozmyslíme, že čísla typu 11, 1001, 100001, ... se dají zapsat způsobem $10^{2n+1} + 1$ pro $n \in \mathbb{N}_0$. Číslo n nám vlastně říká, kolik dvojic 00 je uprostřed.

Dokážeme pro $n \in \mathbb{N}_0$ toto $V(n)$: Číslo $10^{2n+1} + 1$ je dělitelné 11.

Připomeňme, že to vlastně znamená následující: Číslo $10^{2n+1} + 1$ lze zapsat jako $11k$ pro nějaké celé číslo k .

(0) Pro $n = 0$ to platí: $10^1 + 1 = 11 = 11 \cdot 1$.

(1) Mějme libovolné $n \geq 0$, předpokládejme platnost $V(n)$, tedy že $10^{2n+1} + 1 = 11k$ pro nějaké $k \in \mathbb{N}$. Potřebujeme ukázat platnost $V(n+1)$, tedy že i číslo $10^{2(n+1)+1} + 1 = 10^{2n+3} + 1$ je násobkem 11. Abychom mohli využít indukční předpoklad, musíme si nejprve v tomto čísle najít číslo z $V(n)$ chytrým přepsáním. Máme $10^{2n+3} + 1 = 10^{2n+1+2} + 1 = 100 \cdot 10^{2n+1} + 1$, ještě potřebujeme přidat na správné místo +1, což uděláme oblíbeným trikem, že si to přidáme tam, kde to chceme mít, ale pak to musíme také odebrat. Po malé úpravě pak už můžeme použít indukční předpoklad.

$$10^{2n+3} + 1 = 100 \cdot (10^{2n+1} + 1 - 1) + 1 = 100 \cdot (10^{2n+1} + 1) - 100 + 1 = 100 \cdot 11k - 99 = 11(100k - 9),$$

kde číslo $100k - 9$ je celé. Odvodili jsme, že $10^{2(n+1)+1} + 1$ je násobek 11 a tedy $V(n+1)$ platí. Důkaz je hotov.

Alternativa při dekompozici je, že si nejprve upravíme vztah z $V(n)$ na $10^{2n+1} = 11k - 1$ a pak už stačí si ve výrazu z $V(n+1)$ vyrobit 10^{2n+1} , což jsme snadno udělali výše. Nakonec to vyjde nástejno.

△

Matematická indukce je mocný nástroj, který lze aplikovat i na jiné situace než dokazování vzorečků. Zhruba řečeno, o indukci začínáme přemýšlet, když zkoumáme situaci, kterou lze rozložit do etap, ve kterých se nějak přirozeně vyskytuje parametr n coby přirozené číslo, a existuje nějaký vztah mezi etapou současnou a následnou,

čí jinak řečeno mezi současnou a tou předchozí (záleží, jak v dané chvíli zrovna přemýšíme, zda dopředu nebo se vracíme do minulosti).

Příklad 5a.c: Uvažujme turnaj, jehož účastníci hrají každý s každým. Pro zjednodušení budeme předpokládat, že každý s každým hraje pouze jednou, a budeme značit $x \succ y$ fakt, že hráč x porazil hráče y .

Když se hry dohrají, rádi bychom srovnali hráče podle výkonnosti. To se obvykle dělá podle bodů, protože dělat rozumné pořadí jen na základě výsledků je nemožné. Jedním z možných problémů jsou tzv. cykly, nelze rozumně uspořádat tři hráče podle výkonnosti, pokud první porazil druhého, ten třetího, ale třetí zase porazil prvního. Obecněji, cyklem délky n rozumíme situaci, kdy máme hráče h_1, \dots, h_n takové, že $h_1 \succ h_2 \succ \dots \succ h_n$, ale $h_n \succ h_1$. (Na něčem podobném je založena známá skautská desková hra, kdy slon pobije tygra, tygr vlka, vlk psa, pes kočku, kočka myš, ale myš zažene slona).

Dokážeme indukcí vlastnost $V(n)$: Jestliže se ve výsledcích turnaje najde cyklus délky n , pak se tam najde i cyklus délky 3.

Tato vlastnost má smysl jen pro $n \geq 3$, protože ze dvou (a méně) hráčů cyklus při vši snaze nevyrobíme.

(0) $n = 3$: triviální, už máme cyklus délky 3.

(1) Mějme libovolné $n \in \mathbb{N}$, $n \geq 3$. Předpokládejme platnost $V(n)$, zajímá nás platnost $V(n+1)$. Uvažujme proto nějaký cyklus $h_1 \succ h_2 \succ \dots \succ h_n \succ h_{n+1}$, kde také $h_{n+1} \succ h_1$. Potřebujeme se nějak dostat k indukčnímu předpokladu, tedy k cyklu délky n . To se dělá tak, že se zeptáme, jak dopadl souboj h_1 a h_3 . Jestliže $h_3 \succ h_1$, tak máme 3-cyklus $h_1 \succ h_2 \succ h_3$ a je hotovo.

Jestliže naopak $h_1 \succ h_3$, tak lze h_2 z původního cyklu vynechat a dostaneme nový cyklus délky n $h_1 \succ h_3 \succ h_4 \succ \dots \succ h_n \succ h_{n+1}$, v něm podle indukčního předpokladu umíme najít 3-cyklus. Tím je důkaz (1) hotov.

Podle (0) a (1) platí $V(n)$ pro všechna $n \geq 3$.

Právě jsme viděli aplikaci indukce v oblasti zvané teorie grafů.

△

! Poznámka: Všimněte si jednoho důležitého momentu. Oficiálně se indukce tváří, že v ní jde o „krok nahoru“. Známe současnost a ptáme se, zda pomocí ní dokážeme také zvládnout další etapu. Když ji ale vymýslíme, tak nás často zajímá opačná otázka: „Pokouším se vyřešit úlohu na nějaké úrovni. Pomohlo by mi, pokud bych věděl, že tu úlohu umím rozřešit o úroveň níže?“ Jestliže si na tuto otázku odpovím kladně a vymyslím způsob, jak vyřešit daný problém za předpokladu, že znám řešení předchozí situace, tak jsem zároveň přišel na způsob, jak provést důkaz v (1). Jde tedy o rekurzivní způsob přemýšlení, indukce a rekurze jsou tak propojeny, že je těžké rozlišit, kde jedna končí a druhá začíná.

Funguje to i naopak. Pokud najdeme důkaz (1), tak se obvykle stává návodem, jak vytvořit algoritmus k praktickému řešení studovaného problému, od indukčního důkazu bývá tedy jen krůček k rekurzivnímu algoritmu.

△

Příklad 5a.d: Uvažujme čtvercovou šachovnici se stranou o velikosti 2^n polí. (Jinak řečeno, uvažujme čtverec zformovaný z $(2^n)^2$ malých čtverců, kde $n \in \mathbb{N}$.) Začerněme jedno z polí. Tvrdíme, že to, co zbyde, lze zcela pokrýt dlaždicemi složenými ze 3 čtverečků ve tvaru L (tzv. trimin, viz obrázek) tak, aby se nepřekrývaly.

Chceme dokázat $V(n)$: Popsané pokrytí je možné pro čtverec o straně 2^n bez jednoho pole. provedeme to matematickou indukcí.

(0) Je-li $n = 1$, pak jde o čtverec o straně $2^1 = 2$. Po začernění jednoho pole ve čtverci 2×2 zbude právě jedno trimino, které samozřejmě triminy vydláždíme.

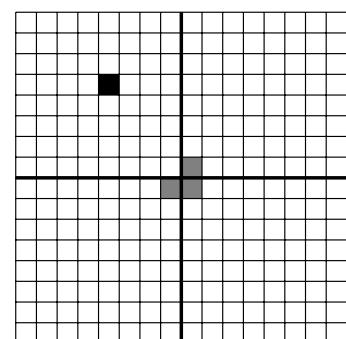
(1) Vezměme libovolné $n \in \mathbb{N}$ a předpokládejme, že umíme vydláždit „čtverec 2^n minus pole“. Potřebujeme ukázat, že pak lze dláždit i „čtverec 2^{n+1} minus pole“.

Postupujeme následovně. Šachovnici o straně 2^{n+1} (bez pole) rozdělíme na stejně velké čtvrtiny. Pak vyjmeme jedno trimino hned u středu tak, aby zmizela právě tři ze čtyř polí okolo středu, a to ta, která nejsou ve stejné čtvrtině jako to chybějící pole. Tímto způsobíme, že teď v každé čtvrtině chybí jedno pole, a každá z těch čtvrtin je čtverec o straně 2^n .

Podle indukčního předpokladu dokážeme každou z těchto čtvrtin pokrýt triminy, pak ještě doplníme jedno navíc do vynechaného středu a jsme hotovi, pokryli jsme čtverec o straně 2^{n+1} bez pole.

Důkaz je hotov.

Praktické pokrytí konkrétní šachovnice můžeme udělat opakovanou aplikací rekurzivního kroku $V(n+1) \mapsto V(n)$, tedy opakováním čtvrcením, dokud nedojdeme k velikosti 2×2 , a pak následným zpětným chodem. Mimochodem, dobrá otázka: Kolik je pro dané n potřeba trimin? Pokud tento počet označíme jako t_n , pak nám naše dekompozice



dává následující rovnici: $t_{n+1} = 4t_n + 1$. Takovéto rovnice se naučíme řešit v kapitole o rekurentních rovnicích, viz příklad .

△

S 5a.3 Poznámka (a velice důležitá!): Pro správné chápání indukce je kritické si uvědomit, že jsme v příkladech nikdy přímo nedokazovali, že by $V(n)$ platilo. Sice se tam říkalo „předpokládejme, že $V(n)$ platí,“ ale to byl jen předpoklad dokazované implikace, který může a nemusí být pravda. Také jsme v důkazu řekli slova „ $V(n+1)$ platí,“ ale to bylo pouze za předpokladu, že platí i $V(n)$. Je veliký rozdíl mezi tvrzením „ $V(n+1)$ platí“ a tvrzením „ $V(n+1)$ platí, pokud platí i $V(n)$ “. Z druhého totiž o pravdivosti $V(n)$ či $V(n+1)$ samotných nic nevyplývá, jen že jsou jejich pravdivosti nějak svázány. V kroku (1) tedy dokazujeme, že se pravdivost přenáší z $V(n)$ na $V(n+1)$, pokud tam nějaká je. Chcete vidět příklad?

Uvažujme vlastnost $V(n)$: $n > 13$. Tvrdíme, že pro všechna $n \in \mathbb{N}$ platí následující implikace:

Jestliže $V(n)$ platí, pak i $V(n+1)$ platí.

Důkaz je snadný. Vezměme si nějaké $n \in \mathbb{N}$. Jestliže platí předpoklad $n > 13$, pak také $n+1 > 13+1 = 14$, tedy platí i $n+1 > 13$.

Takže vidíme, že implikace $V(n) \implies V(n+1)$ platí pro všechna $n \in \mathbb{N}$, ale určitě není pravda, že by $n > 13$ pro přirozená čísla. Finta je v tom, že třeba $n = 2$ nesplňuje $V(2)$, tudíž nám (pravdivá) implikace $V(2) \implies V(3)$ neřekne vůbec nic o pravdivosti $V(3)$.

K důkazu vlastnosti V tedy pravdivost (1) sama o sobě nestačí, to jsme jen postavili domina jedno za druhé. Teprve když doplníme kritický krok (0), tedy šťouchneme do prvního, tak se celá mašinerie rozběhne a pomocí (1) už tato pravdivost dojede libovolně daleko čili všude. U našeho příkladu se to povede, když dokážeme (0): Číslo $n = 14$ splňuje $n > 13$. Tím se proces nastartuje a vlastnost V pak už platí pro všechna $n \geq 14$. K tomuto tématu se vrátíme v poznámce 5a.5.

Využijme této situace k diskusi dalšího problému, který studenti s indukcí mívají. Často zapomenou v kroku (1) napsat kvantifikátor, tedy namísto správného $\forall n$: $[V(n) \implies V(n+1)]$ dokazují jen $V(n) \implies V(n+1)$. Není pak jasné, co tím vlastně myslí.

Ještě horší je, pokud ten kvantifikátor špatně umístí. Někdy začnou důkaz části (1) takto: „Předpokládejme, že pro všechna n platí $V(n)$.“ Dokazují pak implikaci $\forall n: V(n) \implies V(n+1)$, která jednak není formálně správně (jaké n se bere ve vlastnosti $V(n+1)$?), navíc to ani nedává smysl, protože v zásadě není co dokazovat. Jestliže předpokládáme, že V platí pro všechna čísla, pak samozřejmě musí platit pro $n+1$. Pokud budeme předpokládat, že všichni lidé mají hranaté hlavy, pak automaticky má hranatou hlavu i Habala. Taková implikace je tedy tautologií, platí vždy bez ohledu na to, jaká je vlastně vlastnost V .

Z pohledu indukce je ovšem daleko závažnější, že tato implikace je na nic. Potřebujeme implikaci, která ze známé platnosti V pro jedno číslo dokáže tuto platnost posunout o krok dál. Předpokladem oné „zprzněné“ implikace je ovšem výrok $\forall n: V(n)$, o jehož platnosti nevíme nic, přesně tohle naopak chceme dokázat. Taková implikace je proto zcela k ničemu.

Správný zápis důkazů není jen formalita, pomáhá nám to správně pochopit smysl toho, co se děje.

△

! Viděli jsme (a ještě uvidíme), že indukci lze použít i v „nematematických“ situacích. Její speciální schopností je, že nám umožnuje pracovat s konečně mnoha objekty, ale nakonec z té práce získat informaci o nekonečné množině. Z toho ale také plyne omezení jejího pole působnosti, protože takto obsáhneme jen věci očíslovatelné, jinými slovy situace spočetné. Jakmile máme nespočetnou situaci, tak je zbytečné o indukci uvažovat, do tak velkých množin ani pořádně nenakoukne.

Další problém je, že i situace spočetné ještě nemusí být zvládnutelné indukcí, protože sice očíslovatelné jsou, ale to číslování nemusí být v souladu s problémem, který chceme řešit, takže mezi případem s číslem n a případem s číslem $n+1$ nemusí být přímá souvislost. Tím ovšem padá naděje udělat indukční krok (1). Typické je to u racionálních čísel. Je pravda, že množinu \mathbb{Q} jde očíslovat jak $\{r_n\}$, ale to číslování nemá žádnou rozumnou pravidelnost (že by třeba r_n šly podle velikosti či tak něco), proto se v \mathbb{Q} s indukcí víceméně nepracuje. Ale dokonce i když číslování funguje rozumně a my vidíme cestu, jak jít od případu n k případu $n+1$, tak se může stát, že indukce selže, viz poznámka 5a.6.

Dostí zásadní problém u indukce je, že v okamžiku, kdy ji používáme, již musíme dopředu znát odpověď, kterou jsme museli získat nějak jinak. Klasický příklad je důkaz indukcí, že $1 + 2 + \dots + n = \frac{n(n+1)}{2}$, viz cvičení 5a.1 (ii), který je velice snadný, ale onen výraz na pravé straně se musel najít jinak, viz Fakt 9c.3 (ii) a obecnější trik v příkladě 10b.j.

Z pohledu studenta je velký problém indukce v tom, že vypadá tak snadno. U lehkých problémů tomu tak opravdu je, ale ve skutečnosti se v indukci skrývají kritická místa, která při zanedbání mohou pořádně potrápit. Začneme několika příklady s nerovnostmi, které studentům tradičně činí potíže.

Příklad 5a.e: Dokážeme, že pro každé $n \in \mathbb{N}$, $n \geq 3$ platí $V(n)$: $n^2 > n + 5$.

(0) Jestliže $n = 3$, tak vlastnost $V(3)$ říká $3^2 > 8$, což určitě platí.

(1) Předpokládejme, že pro jisté libovolné $n \in \mathbb{N}$ splňující $n \geq 3$ máme $n^2 > n + 5$. Potřebujeme ukázat, že $(n+1)^2 > (n+1) + 5$, tedy že $(n+1)^2 > n + 6$. Použijeme doporučenou metodu, tedy začneme třeba levou stranou a zkusíme od ní dojít ke straně pravé, cestou někde použijeme indukční předpoklad. Na to si budeme muset v levé straně nějak vyrobit n^2 , což lze například roznásobením té druhé mocniny.

$$(n+1)^2 = n^2 + 2n + 1 > (n+5) + 2n + 1 = n + 6 + 2n.$$

My se ale potřebujeme dostat k výrazu $n + 6$ a to takovým způsobem, abychom použili jen rovnost či nerovnosti \geq a $>$. Zde je to naštěstí snadné, díky $n \geq 3$ určitě máme $2n \geq 0$, lze tedy ve výraze napravo $2n$ vynechat (či nahradit nulou, dívejte se na to, jak chcete) a tím jej zmenšit.

$$(n+1)^2 = n^2 + 2n + 1 > (n+5) + 2n + 1 = n + 6 + 2n \geq n + 6.$$

Opravdu jsme dostali, co bylo třeba, a důkaz je hotov.

Potvrdilo se, že u indukce s matematickými vzorečky se obvykle vyplatí začít s dekompozicí od komplikovanější části dokazovaného vztahu.

△

S 5a.4 Poznámka: Vracíme se k poznámce po prvním příkladu. I zde jsme při důkazu nerovnosti použili metodu postupných kroků. Začali jsme výrazem na jedné straně dokazovaného, pak jsme jej postupně upravovali a dospěli tak k žádanému. Všechny kroky byly aritmeticky v pořádku, v zásadě jsme používali následujícího principu: Máme-li dvě čísla $a + b$ a teď jedno z nich nahradíme menším, třeba namísto a dáme menší c , pak je celek menší, tedy $a + b > c + b$. Podobně pokud máme $c \leq a$, pak po nahrazení dostaneme $a + b \geq c + b$. Tímto způsobem se nerovnosti v indukci dokazují docela spolehlivě, jen někdy dá trochu problém se strefit do cílového výrazu.

Naopak vůbec nefunguje ona populární metoda postupných úprav žádané nerovnosti. Zkusme si to pro nás příklad: Napíšeme žádanou nerovnost a zkusíme ji upravit.

$$\begin{aligned} (n+1)^2 &> n+6 \\ n^2 + 2n + 1 &> n+6. \end{aligned}$$

A tím jsme skončili, indukční předpoklad sice říká, že $n^2 > n + 5$, ale my nemůžeme v naší nerovnosti nahradit část n^2 výrazem $n + 5$, protože to není korektní úprava! Jinými slovy, pokud by někdo jako další řádek napsal $n + 5 + 2n + 1 > n + 6$, tak by to měl špatně, podrobněji o tomto viz poznámka 1b.6. U nerovností tedy tento postup rozhodně nedoporučujeme. Pro další detaily viz poznámka 1b.6 a zejména příklad na konci kapitoly 14.

△

Příklad 5a.f: Dokážeme, že pro každé $n \in \mathbb{N}_0$ platí $V(n)$: $n < 2^n$.

(0) Nechť $n = 0$. Určitě $0 < 2^0$, tedy $V(0)$ platí.

(1) Předpokládejme, že pro jisté (libovolné) $n \in \mathbb{N}_0$ máme $n < 2^n$. Potřebujeme ukázat, že $n + 1 < 2^{n+1}$. Víme už, že postupné úpravy této nerovnice nefungují, musíme začít na jednom konci, třeba na pravém, a propracovat se k druhému. Nejprve potřebujme 2^{n+1} upravit tak, abychom mohli použít indukční hypotézu.

$$2^{n+1} = 2 \cdot 2^n > 2 \cdot n.$$

Tedž se nějak potřebujem dostat dalšími úpravami k $n + 1$, a aby celý řetěz úprav platil, můžeme používat pouze kroky s rovností = či nerovnostmi \geq a $>$. Je vůbec možné se dostat od $2n$ k $n + 1$ tímto způsobem, jinými slovy, platí vůbec $2n \geq n + 1$? Ano, pokud $n \geq 1$, ale to my nemáme. Tím se ukazuje, že jsme svou indukci začali příliš brzy. Je sice pravda, že $V(0)$ platí, ale indukční krok $V(n) \implies V(n+1)$ budeme umět dokázat až od $n = 1$. Začneme tedy s důkazem znova, přičemž indukce začne až pro $n \geq 1$, případ $n = 0$ vyřešíme zvlášť.

Takže nejprve přímo ověříme, že $V(n)$ platí pro $n = 0$ (už jsme provedli výše). Tedž pro $n \geq 1$ použijeme důkaz indukci:

(0) $V(1)$ platí, určitě $1 < 2^1$.

(1) Předpokládejme, že pro jisté $n \in \mathbb{N}$ máme $n < 2^n$. Potřebujeme ukázat, že $n + 1 < 2^{n+1}$. Podle indukční hypotézy a z toho, že $n \geq 1$, dostaneme

$$2^{n+1} = 2 \cdot 2^n > 2 \cdot n = n + n \geq n + 1.$$

Důkaz je hotov.

△

5a.5 Poznámka:

Pro správné chápání indukce je třeba ocenit, že tvrzení (0) a (1) jsou nezávislé věci, každá si dělá něco jiného a jejich spojením pak teprve vznikne platný důkaz.

V příkladě 5a.e jsme měli situaci, kdy v (1) ten indukční krok $V(n) \implies V(n+1)$ platil pro všechna $n \geq 0$, ale nebylo nám to nic platné. Protože $V(0)$, $V(1)$ ani $V(2)$ neplatí, nešla tato implikace využít s malými čísly a běh indukce se dal začít až od $n = 3$.

Naopak v příkladě 5a.f jsme mohli v kroku (0) použít $n = 0$, ale nebylo to k ničemu, protože indukční krok $V(n) \implies V(n+1)$ se ruměl rozběhnout až pro vyšší n .

Pro další příklady podobného typu se podívejte na cvičení 5a.4 a 5a.5.

△

Indukce sice vypadá jako jednoduchá věc, ale pokud jí člověk dobře nerozumí, popřípadě pokud není dostatečně opatrný, může snadno udělat chybu.

! Příklad 5a.g: „Dokážeme“ indukcí, že v každé (neprázdné) třídě mají vždy všichni žáci stejně pohlaví (muž/žena).

Pro $n \in \mathbb{N}$ uvažujme $V(n)$: V každé třídě s n studenty mají všichni studenti stejně pohlaví.

(0) Případ $n = 1$ je zřejmý, ve třídě s 1 studentem to platí.

(1) Nechť $n \in \mathbb{N}$ je libovolné. Předpokládejme, že shoda pohlaví platí pro všechny třídy s n studenty. Teď mějme nějakou třídu T s $n + 1$ studenty. Zvolme nějakého studenta $a \in T$ a uvažujme třídu $A = T - \{a\}$. Tato má n studentů, tudíž podle indukčního předpokladu mají všichni stejně pohlaví. Zbývá ukázat, že i a musí mít stejně pohlaví jako studenti z A . Protože $n + 1 \geq 2$, lze najít $b \in T$ takové, že $b \neq a$. Uvažujme třídu $B = T - \{b\}$. I ta má n studentů, i v této třídě musí mít všichni stejně pohlaví.

Teď zvolme nějaké $c \in T - \{a, b\}$, čili studenta, který je v A i v B . Protože $c \in A$, mají všichni z A stejně pohlaví jako c . Protože také $c \in B$, mají i všichni z B stejně pohlaví jako c . A protože $T = A \cup B$, mají všichni z T stejně pohlaví jako c . Důkaz je hotov.

Kde je chyba? Rozhodně ne v indukci, struktura důkazu je zcela správně. Problémy musíme hledat v jednotlivých argumentech. První odstavec části (1) je správně, tomu není co vytknout. Problém je v druhém odstavci: Jak víme, že se takové c dá najít? Kdyby $T = \{a, b\}$, pak žádné c není, tudíž celý důkaz padá. Z toho je vidět, že implikace $V(n) \implies V(n+1)$ platí pro všechna $n \geq 2$, ale neplatí pro $n = 1$ a to už stačí, aby celý důkaz indukcí neplatil.

Zároveň se tím ukazuje, proč v matematice vyžadujeme, aby byl každý krok v důkazu opravdu něčím podepřen. Když matematik v onom důkazu čte „zvolme nějaké c “, tak se okamžitě ptá: Opravdu můžeme? Chrání se tím před chybami z přehlédnutí.

△

! Příklad 5a.h: „Dokážeme“ indukcí, že pro všechna $x, y \in \mathbb{N}$ platí $x = y$ (tedy všechna přirozená čísla jsou si rovna).

Použijeme indukci podle toho, jak jsou x a y velká, což nám říká hodnota $\max(x, y)$. Tu tedy použijeme v indukci jako krokovací parametr. Formálně:

Pro $n \in \mathbb{N}$ uvažujeme $V(n)$: Jestliže $x, y \in \mathbb{N}$ a $\max(x, y) = n$, pak $x = y$.

(0) Jestliže $x, y \in \mathbb{N}$ a $\max(x, y) = 1$, pak $1 \leq x \leq 1$ a $1 \leq y \leq 1$, tedy opravdu $x = 1 = y$.

(1) Předpokládejme, že pro jisté (libovolné) $n \in \mathbb{N}$ platí $V(n)$, potřebujeme ukázat, že platí i $V(n+1)$. Mějme tedy nějaké $x, y \in \mathbb{N}$ takové, že $\max(x, y) = n + 1$. Pak $\max(x - 1, y - 1) = n$, tudíž dle indukčního předpokladu $x - 1 = y - 1$, proto $x = y$. Důkaz je hotov.

Kde je chyba tady? Tu často odhalíme, když si nějaký podezřelý případ zkusíme projet rekurzivním algoritmem, který je vlastně v indukci schovaný. Jak třeba ukážeme, že $2 = 4$? Máme $\max(2, 4) = 4$, podle indukčního kroku se pak odvoláváme na případ $\max(1, 3) = 3$, z toho zase na případ $\max(0, 2) = 2$ a hned máme problém, protože nás postup s nulou nepočítal. Kde je tento problém schován v našem „důkazu“? Právě provedený zpětný chod naznačuje, že je to někde v aplikaci indukčního předpokladu. Pokud chceme v důkazu něco použít, musíme si hlídat, zda ona věc nemá nějaké zabudované podmínky. U indukčního předpokladu to vždy bývá to, že jej můžeme použít jen pro naše konkrétní n , ale už ne pro jiná čísla. Někdy má ale indukční předpoklad zabudovány další podmínky. Projděme si to v našem příkladě.

Používáme jej se zvoleným n , to je v pořádku. Pak je tam ovšem omezení, na které páry čísel jej můžeme aplikovat. Je dvojice $x - 1, y - 1$ v pořádku? Určitě platí $\max(x - 1, y - 1) = n$, to je základní algebra, takže tato podmínka je v pořádku. Pak je tam ale ještě jedna věc: máme mít $x - 1 \in \mathbb{N}$ a $y - 1 \in \mathbb{N}$. A jak jsme viděli, v tom je právě zádrhel. Pro $x = 1$, popř. $y = 1$ se dostaneme k nule, která už není v \mathbb{N} a indukční předpoklad nejde použít. Tím je celý důkaz špatně.

△

Jako domácí úkol matematickou indukcí dokažte, že do autobusu jezdícího z kolejí do školy se vejde libovolný počet lidí.

5a.6 Poznámka stranou: Někdy indukce narazí i v situaci, která na první pohled vypadá jako pro ni stvořená, rizikovým faktorem bývají nerovnosti. Ukážeme si to na následujícím příkladě: Zkusíme indukcí dokázat, že pro každé $n \in \mathbb{N}$ platí vlastnost $V(n)$: $\frac{n-1}{n} < 1$ (což je dozajista pravda).

(0) Pro $n = 1$ máme $V(1)$: $0 < 1$, což platí.

(1) Mějme libovolné $n \in \mathbb{N}$ a předpokládáme, že platí $\frac{n-1}{n} < 1$. Chceme dokázat, že pak platí i $V(n+1)$, tedy $\frac{n}{n+1} < 1$. Začneme s výrazem na levé straně, potřebujeme jej upravit tak, aby šlo použít indukční předpoklad.

$$\frac{n}{n+1} = \frac{n^2}{(n+1)(n-1)} \cdot \frac{n-1}{n} < \frac{n^2}{(n+1)(n-1)} = \frac{n^2}{n^2-1}.$$

A máme problém, rozhodně už se nám nepodaří tento řetězec výrazů zakončit potřebným krokem $\frac{n^2}{n^2-1} < 1$, protože to neplatí.

Kde je zádrhel? Označme $a_n = \frac{n-1}{n}$. My jakoby víme, že $a_n < 1$, a chceme to použít k důkazu $a_{n+1} < 1$. Když se na ta čísla podíváme blíže, tak zjistíme, že a_{n+1} je mnohem blíže k 1 než výraz a_n . Pokud tedy při úpravách členu a_{n+1} použijeme onen větší rozdíl (což právě při aplikaci indukčního předpokladu děláme), dostaneme se okamžitě nad jedničku je to v háji.

Závěr je, že vlastnost $V(n)$ indukcí takto přímo dokázat nejde.

Pro další příklad viz cvičení 5a.7.

△

Ted' se zase vrátíme ke správným důkazům a ukážeme trochu jiné použití.

! Příklad 5a.i:

V tomto příkladě zabrousíme do algoritmizace.

Definujeme rekuzivní proceduru, záměrně se nedržíme nějakého konkrétního jazyka.

```
procedure factorial(n: nezáporné celé číslo)
  if n = 0 then factorial(n):= 1
  else factorial(n):= n·factorial(n - 1);
```

Tvrdíme, že výstup procedury je $n!$.

Dokážeme to indukcí: $V(n)$ je tvrzení, že výstup $\text{factorial}(n)$ je $n!$.

(0) $n = 0$: ano, podle specifikace je $\text{factorial}(0) = 1 = 0!$.

(1) Nechť $n \in \mathbb{N}_0$. Předpokládejme, že $V(n)$ platí, tedy výstup $\text{factorial}(n)$ je $n!$. Pak výstup $\text{factorial}(n+1)$ je roven $(n+1) \cdot \text{factorial}(n)$, což je $(n+1) \cdot n! = (n+1)!$.

Důkaz je hotov.

Připomněli jsme si faktoriál, který se indukcí definuje, obvykle takto:

```
(0)  $0! = 1$ .
(1)  $(n+1)! = n! \cdot (n+1)$  pro  $n \geq 0$ .
```

Není to první takový objekt, se kterým jsme se setkali, již v prvních kapitolách se induktivně definovala mocnina zobrazení T^n či složení konečně mnoha zobrazení. Definování objektů pomocí indukce se hlouběji věnujeme v další kapitole, berte tento příklad jako reklamu.

△

! 5a.7 Poznámka:

Využijme tento příklad k malé exkurzi do oblasti algoritmů. Když nějaký algoritmus navrhnete, tak bychom správně měli dokázat, že vždy dělá to, co má. Pokud je to algoritmus rekurentní, pak je k tomu nevhodnějším nástrojem indukce, ostatně jsme se o vzájemné provázanosti rekurze a indukce již dříve zmínili.

Obvykle se nezačíná tím, co jsme dělali v příkladu výše, tedy zkoumáním výsledku po doběhnutí algoritmu, ale nejprve se musí dokázat, že algoritmus vůbec doběhne. Zkusme si představit, že onen algoritmus výše poštěveme na číslo 1.7. Nás algoritmus zjistí, že to není nula, tudíž zavolá sám sebe znova, tentokrát se vstupem 0.7. To zase není nula, tudíž se zavolá se vstupem -0.3 . A tak dále, program nikdy neskončí.

Čtenář patrně namítne, že jsme špatní programátoři, protože jsme zapomněli doplnit vstupní filtr. To je pravda, jenže jsme měli snadný algoritmus. Pokud je komplikovanější, tak vůbec nemusí být jasné, co je tím vhodným vstupním filtrem.

Snad každý rekurentní algoritmus má množinu základních hodnot, které umí rovnou, vypíše výsledek a skončí. Pokud dostane hodnotu jinou, tak ji všelijak upravuje a sám sebe spouští znova a znova, dokud se netrefí do jedné z těch základních hodnot. Neexistuje obecná metoda, jak zjistit, kterými vstupními daty je možné začít, aby se nakonec do těch základních dostali. Ukážeme dva příklady, veselý a opravdový matematický, přitom velice jednoduchý.

1) Mnoho autorů knih o algoritmizaci či programování neodolá a dá do Rejstříku rádek „Rekurze – viz Rekurze.“ Jde o rekurzivní algoritmus bez ukončovací podmínky. Takovou chybu ovšem udělá jen začátečník, chytřejší autoři tam píšou toto:

„Rekurze – pokud to ještě nechápete, viz Rekurze.“

Zde podmínka pro ukončení je, ale je asi hned jasné, že u některého čtenáře nemusí dojít k její realizaci.

2) Abychom si ukázali pořádný příklad, představíme si nejprve zobrazení $T: \mathbb{N} \mapsto \mathbb{N}$ dané předpisem

$$T(n) = \begin{cases} \frac{1}{2}n, & n \text{ sudé;} \\ 3n + 1, & n \text{ liché,} \end{cases}$$

viz cvičení 2b.8. Ted' se podíváme, co se stane, když jej začneme aplikovat opakováně (neboli když uvažujeme mocniny T^m tohoto zobrazení). Když třeba začneme s $n = 13$, tak to je liché, proto $T(13) = 3 \cdot 13 + 1 = 40$. To je sudé, takže další aplikace T dává $T^2(13) = T(40) = \frac{1}{2}40 = 20$. To je zase sudé, tedy $T^3(13) = T(20) = 10$, pak $T^4(13) = 5$ a tak dále, dostáváme řetězec

$$13 \mapsto 40 \mapsto 20 \mapsto 10 \mapsto 5 \mapsto 16 \mapsto 8 \mapsto 4 \mapsto 2 \mapsto 1,$$

tedy $T^9(13) = 1$. Lidé si myslí, že ať už začneme jakýmkoliv n , vždycky dřív nebo později dojdeme k 1. Zatím to ale nikdo neuměl ani dokázat, ani vyvrátit, takže se to prostě neví. A ted' koukněte na tohle:

```
procedure T(n: přirozené číslo)
a := T(n);
if a > 1 then T(a);
```

Toto je jednoduchý rekurzivní program, který, pokud jej zavolám jako $T(13)$, skončí po devíti cyklech. Terminační podmínu má, jenže z toho, co jsme si o zobrazení T řekli, vyplývá, že není známo, zda k ní pro všechna vstupní data tento program někdy dojede. To je smutné, současná věda neumí u tohoto algoritmu dokázat, že vždy skončí.

Při zkoumání problému ukončování běhu algoritmu je dobrým nástrojem tzv. *variant*, což je nějaký parametr, který nabývá přirozených čísel a při každém volání rekurze se zmenšuje. Protože neexistuje nekonečná klesající posloupnost přirozených čísel (viz níže), musí algoritmus, u kterého se nám podaří takový variant identifikovat, také někdy skončit. Například u procedury *factorial* může jako variant sloužit n , zatímco u té procedury s T zatím nikdo nějaký variant nevymyslel.

Protože dokázat terminaci algoritmu je často vysoce náročný úkol, zkoumá se u nich takzvaná *parciální korektnost*, což jsou výroky typu „Pokud vůbec algoritmus skončí, tak se stane toto.“ Například ten zajímavý algoritmus s T je parciálně korektní ve smyslu „Jestliže skončí, tak dá jedničku.“ Ale to už opravdu zabíháme do teorie algoritmů, kde je ovšem indukce jedním z oblíbených nástrojů.

△

! Po toliku příkladech už asi není třeba vysvětlovat, že indukce je velice důležitá. Je proto kritické si položit otázku, nakolik je možné jí věřit. Čtenář si jistě všiml, že jsme princip matematické indukce neuvedli jako větu. Důvod je jednoduchý, je to totiž jeden z axiomů matematiky, viz poznámka 4c.15. Většinou se ale do seznamů základních axiomů nezahrnuje, protože je rovnocenný axiomu jinému, který jsme tu už měli (viz Princip 4c.14).

Věta 5a.8.

Princip matematické indukce je ekvivalentní s principem dobrého uspořádání.

Důkaz (drsný, poučný): 1) Předpokládejme, že princip matematické indukce platí. Chceme ukázat, že (\mathbb{N}, \leq) je dobře uspořádaná množina.

Nechť je tedy M nějaká neprázdná podmnožina \mathbb{N} . Definujme vlastnost V takto:

$$V(n): \{1, 2, \dots, n\} \cap M = \emptyset.$$

Kdyby toto platilo pro všechna n , tak by pro všechna n platilo $n \notin M$, tedy $M = \emptyset$, což je ve sporu s předpokladem, že M je neprázdná.

Vlastnost V tedy neplatí pro nějaké $n \in \mathbb{N}$. Uvažujme tato dvě tvrzení:

(0) $V(1)$ platí.

(1) Pro každé $n \in \mathbb{N}$: Jestliže $V(n)$ platí, pak i $V(n+1)$ platí.

Kdyby platilo (0) a (1), tak by dle principu matematické indukce platilo V pro všechna n , my už ale víme, že to nejde. To znamená, že alespoň jedno z těchto dvou tvrzení není pravdivé.

Jestliže není pravda $V(1)$, tak není pravda $\{1\} \cap M = \emptyset$. To znamená, že $1 \in M$. Jelikož $M \subseteq \mathbb{N}$, tak $1 \leq x$ pro všechna $x \in M$, tedy 1 je nejmenší prvek M .

Druhá možnost je, že neplatí (1). To znamená, že existuje n takové, že neplatí implikace $V(n) \implies V(n+1)$. Pro toto speciální n tedy platí, že $V(n)$ je pravda a $V(n+1)$ je nepravda neboli $\{1, 2, \dots, n\} \cap M = \emptyset$ a $\{1, 2, \dots, n, n+1\} \cap M \neq \emptyset$. To mimo jiné říká, že $n+1 \in M$.

Protože M neobsahuje čísla $1, 2, \dots, n$, tak pro všechna $x \in M$ máme $n+1 \leq x$. Toto a závěr předchozího odstavce ukazují, že $n+1$ je nejmenší prvek M .

Rozborem možností jsme ukázali, že M má za všech okolností nejmenší prvek.

2) Opačný směr plyne z věty 5a.13, která zobecňuje indukci i na jiné množiny než \mathbb{N} . □

Dobrá otázka: Proč jsme nepoužili obvyklý trik a nedokazovali indukcí $V(n)$: každá n -prvková podmnožina \mathbb{N} má minimum? Protože definice dobrého uspořádání zahrnuje i minima nekonečných podmnožin, takže by to nestačilo. S tím se ale dá vyrovnat, možných přístupů je víc, takže lze porůznu najít i jiné důkazy této věty. Ten náš je zajímavý tím, že jakoby staví indukci na hlavu.

Teď si ukážeme ještě jednu aplikaci indukce, bude to reklama na novou verzi.

Příklad 5a.j: Uvažujme množinu M měst, vesnic a vůbec osídlení vybranou porůznu po světě. Po světě také existují rozličné železniční sítě, které do některých z osídlení dosáhnou. Pro $n \in \mathbb{N}$ uvažujme tvrzení $V(n)$:

Libovolná množina M s n sídly se dá rozložit na podmnožiny M_i takové, že $M = \bigcup M_i$, mezi osídleními ze dvou různých M_i, M_j nevede železniční spojení, naopak v každé M_i jsou vždy všechna sídla navzájem propojena.

Ctenáře doufajme napadlo, že vlastně mluvíme o rozkladu množiny na třídy ekvivalence, stačí uvažovat relaci na M danou spojením a dokázat, že jde o ekvivalenci. V této kapitole ale zkusíme žádaný rozklad vyrobit přímo, bez pomoci jiné teorie.

Jak bychom takové skupiny M_i vytvářeli? Vezmeme libovolné osídlení $a_1 \in M$ a do množiny M_1 dáme všechna osídlení z M , do kterých se z a_1 dostaneme vlakem. Je pak jasné, že se dostaneme i mezi libovolnými dvěma osídleními z této množiny, přinejhorším to vezmeme s přestupem v a_1 . A co ostatní osídlení? K těm jsme se nedostali z a_1 a díky přestupům je jasné, že se k nim nedostaneme ani z ostatních osídlení v M_1 , takže tato M_1 je opravdu odříznuta od zbytku M . Logicky bychom dále vzali nějaké a_2 z toho zbytku a tak dále, to volá po indukci.

(0) Jestliže $n = 1$, pak máme množinu $M = \{a\}$ s jedním osídlením, což je zároveň množina M_1 , neboť z a do a se dostanu a nikam jinam ne.

(1) Mějme libovolné $n \in \mathbb{N}$ a předpokládejme, že se nám každá n -prvková množina osídlení rozpadne dle předpisu. Uvažujme teď nějakou množinu s $n+1$ osídleními. Vezměme prvek a_1 a vytvořme množinu M_1 , přesně jako jsme to dělali výše. Teď uvažujme $M' = M - M_1$. A máme velký problém, protože vůbec nevíme, jestli M' má n prvků. Pokud ne (což se dá mimochodem čekat), tak je nám indukční hypotéza na nic, my jsme totiž předpokládali její platnost jen pro naše n , pro jiná čísla ne.

△

Jaké je z toho poučení? Že občas potřebujeme něco lepšího než obyčejnou indukci.

! 5a.9. Silný princip matematické indukce.

Nechť $n_0 \in \mathbb{Z}$, nechť $V(n)$ je vlastnost celých čísel, která má smysl pro $n \geq n_0$.

Předpokládejme, že následující předpoklady jsou splněny:

(0) $V(n_0)$ platí.

(1) Pro každé $n \in \mathbb{Z}$, $n \geq n_0$ je pravdivá následující implikace: Jestliže platí $V(k)$ pro všechna $k = n_0, n_0 + 1, \dots, n$, pak platí i $V(n+1)$.

Potom $V(n)$ platí pro všechna $n \in \mathbb{Z}$, $n \geq n_0$.

Tomuto principu se také říká **úplná indukce**. Anglicky se tomu říká **Strong principle of mathematical induction**. Jeho interpretace je následující. (0) říká, že umíme vylézt na první příčku žebříku. (1) říká, že v případě, že umíme vylézt na prvních n příček, tak umíme vylézt i o jednu výše. Princip pak tvrdí, že umíme vylézt na celý žebřík. Ukážeme si, jak nám to pomůže. Nejprve se vrátíme k příkladu výše.

Příklad 5a.k (pokračování 5a.j): Dokážeme vlastnost $V(n)$ o rozkladu množiny osídlení pomocí silného principu indukce.

(0) Jestliže $n = 1$, pak $V(1)$ platí, to už jsme dělali.

(1) Nechť $n \geq 1$ a předpokládejme, že umíme příslušným způsobem rozložit všechny množiny osídlení o velikosti mezi 1 až n včetně. Mějme teď množinu M s $n+1$ osídleními. Zvolíme prvek a_1 a vybereme do množiny M_1 osídlení a_1 a také všechna osídlení, do kterých se z a_1 dostaneme vlakem. Uvažujme $M' = M - M_1$. Pak je počet osídlení v M' menší než v M , čili je to určitě číslo mezi 1 až n včetně. Podle indukčního předpokladu je tedy možné M' rozdělit na navzájem izolované, ale uvnitř propojené podmnožiny M_2, M_3, \dots . Pak je M_1, M_2, M_3, \dots žádaný rozklad množiny M a důkaz je hotov.

△

Příklad 5a.l: Ukážeme si aplikaci z teorie her.

Mějme dvě hromádky zápalek a dva hráče. Ti se střídají, v každém tahu si hráč vybere jednu hromádku a z ní pak odebere alespoň jednu zápalku. Hráč, který vezme poslední zápalku, vyhraje.

Ukážeme, že pokud je na začátku v obou hromádkách stejně zápalek, tak má druhý hráč výherní strategii (tedy algoritmus, který vede vždy na výhru, ať už dělá první hráč cokoliv).

Dále ukážeme, že pokud se počet zápalek v hromádkách různí, tak má první hráč výherní strategii.

Poznámka: Je to tedy jedna z her, u které je již na začátku rozhodnuto, jak to dopadne, pokud hráči hrají alespoň trochu intelligentně. Kupodivu i takové hry se hrají, například v Severní Americe tolik populární piškvorky na čtverci 3×3 zvané tic-tac-toe.

1) Nejprve dokážeme $V(n)$: Pokud je na začátku v obou hromádkách n zápalek, pak má druhý hráč výherní strategii.

(0) $n = 1$: První hráč musí vzít jednu zápalku, nemůže vzít obě (jsou na různých hromádkách), druhý pak vezme druhou neboli poslední a vyhrál.

(1) Nechť $n \geq 1$. Předpokládejme, že platí $V(1), V(2), \dots, V(n)$, tedy že druhý hráč má výherní strategii na hry, kde je na začátku na obou hromádkách stejně zápalek, a to nejvýše n . Chceme ukázat, že pak má i výherní strategii pro situaci s $n + 1$ zápalkami na obou hromádkách.

Takže máme dvě hromádky po $n + 1$ zápalkách. Nechme udělat prvního hráče první tah, odebere $r \geq 1$ zápalek z jedné hromádky. Pokud $r = n + 1$, tak už vzal všechny, druhý hráč odebere druhou hromádku a vyhrál. Pokud $r < n + 1$, tak v první hromádce zbylo $n + 1 - r$ zápalek, druhý hráč pak na to reaguje tak, že odebere r zápalek z hromádky druhé. Teď je v obou hromádkách $n + 1 - r$ zápalek a je na tahu první hráč, čili jakoby hra začínala znova, a protože mají obě hromádky po $n + 1 - r$ zápalkách, kde $1 \leq n + 1 - r \leq n$, má podle indukčního předpokladu druhý hráč výherní strategii.

2) Pokud není na hromádkách stejně, tak první hráč prvním tahem odebere z větší hromádky tolik, aby srovnal počty. Teď je na obou hromádkách stejně a druhý hráč jakoby začíná, čímž se role prohodily a první hráč má výherní strategii.

Jako obvykle nám naše důkazy zároveň daly algoritmus k řešení problému, v tomto případě strategii pro výhru. Pokud může, hráč prostě vždy dorovnává hromádky na stejný počet a nakonec vyhraje.

Teorie her je zajímavá oblast matematiky s aplikacemi v mnoha oborech, u některých to nepřekvapí (ekonomie, diplomacie, vojenské vědy), u některých možná ano (genetika).

△

Zdá se tedy, že silný princip indukce je opravdu lepší, protože nám pomohl v situacích, kdy slabý selhal. Ve skutečnosti ale silnější není, je jen pohodlnější pro uživatele. Pro názornou ukázku se vrátíme k poslednímu příkladu.

Příklad 5a.m: Uvažujme hru se zápalkami, ale definujme jinou vlastnost.

$W(n)$: Pokud je na začátku v obou hromádkách stejně zápalek, a to mezi 1 a n , pak má druhý hráč výherní strategii.

Dokážeme indukci platnost pro $n \in \mathbb{N}$:

(0) $n = 1$: Chceme ukázat existenci výherní strategie druhého hráče pro případ, kdy obě hromádky mají jednu zápalku, to už jsme udělali v předchozím řešení.

(1) Nechť $n \in \mathbb{N}$. Dokazujeme platnost implikace $W(n) \implies W(n + 1)$.

Indukční předpoklad je, že druhý hráč má výherní strategie na všechny hry se shodným počtem zápalek na obou hromádkách, a to počtem mezi 1 a n .

Potřebujeme ukázat, že má výherní strategie na všechny hry se shodným počtem zápalek na obou hromádkách, a to počtem mezi 1 a $n + 1$.

Vezměme tedy dvě shodné hromádky. Pokud je počet zápalek mezi 1 až n , pak přímo aplikujeme indukční předpoklad $W(n)$ a máme výherní strategii pro druhého hráče. Pokud je ten počet roven $n + 1$, tak v prvním kole poté, co první hráč odebral nějaké zápalky, druhý hráč dorovná hromádky na stejný počet, který je již nejvýše n . Pokud je to nula, druhý hráč vyhrál, takže ví jak na to, pokud ne, použije indukční předpoklad a má výherní strategii.

△

Tento důkaz byl komplikovanější, takže vidíme, že silný princip nám opravdu může ulehčit práci. Myšlenku použitou v příkladě lze použít obecně jako argument, že z pohledu teoretického jsou oba principy indukce rovnocenné.

Věta 5a.10.

Slabý a silný princip matematické indukce jsou ekvivalentní.

Nejprve musíme pořádně říct, co tím vlastně myslíme. Jednoduše řečeno to znamená, že množina věcí, které lze dokázat slabým principem, je úplně stejná jako množina věcí, které jdou dokázat silným principem.

Důkaz (poučný, drsný): 1) Nejprve předpokládejme, že vlastnost V pro čísla $n \geq n_0$ lze dokázat slabým principem, tedy že jsme její platnost dokázali argumentem, že splňuje následující tvrzení:

(s0) $V(n_0)$ platí.

(s1) Pro všechna $n \geq n_0$: Jestliže platí $V(n)$, tak platí i $V(n+1)$.

Musíme ukázat, že ji lze dokázat také silným principem, tedy chceme ukázat, že platí následující vlastnosti:

(S0) $V(n_0)$ platí.

(S1) Pro všechna $n \geq n_0$: Jestliže platí $V(n_0), V(n_0+1)$ až $V(n)$, tak platí i $V(n+1)$.

Hned vidíme, že (S0) je pro V splněno, protože je to totéž jako (s0).

Je pro V splněno (S1)? Vezměme si nějaké libovolné $n \geq n_0$ a předpokládejme, že platí $V(n_0)$ až $V(n)$. Takže mimo jiné platí i $V(n)$ a o vlastnosti V víme, že splňuje (s1). Podle toho tedy platí $V(n+1)$, čímž je pravdivost (S1) dokázána.

Takže vlastnost V splňuje podmínky (S0) a (S1) a tudíž je její platnost dokázána pro všechna $n \geq n_0$ pomocí silného principu indukce.

2) Teď předpokládejme, že vlastnost V lze pro $n \geq n_0$ dokázat silným principem, tedy že splňuje (S0) a (S1). Musíme ukázat, že ji lze dokázat také slabým principem.

Splňuje (s0)? Ano, protože je to totéž jako (S0).

Splňuje (s1)? Nejspíše ne. Máme ukázat platnost implikace $V(n) \implies V(n+1)$ pomocí znalosti (S1), začneme tedy předpokládat, že $V(n)$ platí, ale to nám nestačí k tomu, abychom dokázali použít (S1), protože nevíme nic o platnosti $V(n_0)$ až $V(n-1)$.

Důkaz tedy takto jednoduše, jak tomu bylo v 1), nepůjde, musíme použít trik. Definujme novou vlastnost $W(n)$ takto: $W(n)$ platí, jestliže platí $V(k)$ pro $k = n_0, \dots, n$. Dokážeme teď pomocí slabé indukce tuto vlastnost W .

(s0) Nechť $n = n_0$. $W(n_0)$ znamená, že platí $V(n_0)$, což je pravda dle (S0). Takže (s0) platí.

(s1) Nechť $n \geq n_0$ a předpokládejme, že platí $W(n)$. Podle definice této vlastnosti to znamená, že platí $V(n_0)$ až $V(n)$, odtud ale díky platnosti (S1) pro V odvodíme, že platí $V(n+1)$. Platí tedy $V(n_0)$ až $V(n)$ a také $V(n+1)$, tedy platí $W(n+1)$. Dokázali jsme pravdivost implikace $W(n) \implies W(n+1)$, tedy (s1) platí pro W .

Podle slabého principu indukce dostáváme, že W platí pro všechna $n \geq n_0$, proto podle definice W platí i $V(n)$ pro všechna $n \geq n_0$. Takže jsme V dokázali i pomocí slabého principu.

□

! Kdy budeme chtít použít silnou indukci? Rozhodne rekurentní analýza. Pokoušíme se vyřešit daný problém na určité úrovni. Pokud jej dokážeme vždy vyřešit čistě pomocí znalosti předchozí etapy, pak si vystačíme se slabou indukcí. Pokud ale potřebujeme informaci i dále z minulosti, zejména pokud vlastně ani nevíme přesně, jak daleko do minulosti máme zajít, pak musíme použít silnou indukci.

Existují ale situace, které jsou ještě trochu jiné. Tam sice musíme jít dál do minulosti, takže slabý princip indukce nelze přímo aplikovat, ale víme, že pokaždé musíme jít zpět jen o přesně specifikovaný (a stále stejný) počet kroků. Tato situace je velice specifická, protože pak na ni nelze přímo aplikovat ani silný princip indukce. Abychom to ukázali, připomeneme si jeden klasický příklad.

Jak vyložíme podrobněji v příkladě 10b.b, jistý Fibonacci zkoumal králíky a došel k zajímavému závěru, že chce-li předpovědět, kolik párů bude mít příští rok, tak stačí sečíst počet párů z předchozích dvou let. Takže ke znalosti počtu párů v roce 50 potřebujeme znát počty v letech 48 a 49, ke znalosti z let 89 a 90 zase spočteme stav v roce 91 atd. Princip je velice jednoduchý, tak zkusme začít. Řekněme, že je teď rok 1 a víme, že máme jeden pár králíků. Kolik jich budeme mít v příštím roce? Zatímco u slabé a silné indukce stačí znát jednu výchozí hodnotu, tady je to evidentně málo, protože na výpočet potřebujeme znát dvě předchozí hodnoty!

Takže přidejme výchozí informaci, že příští rok (rok 2) budeme mít také jeden pár, a dál už to jde počítat. V roce 3 budou $1 + 1 = 2$ páry, v roce 4 bude dle stavu v letech 2 a 3 celkem $1 + 2 = 3$ páry, v roce 5 bude $2 + 3 = 5$ párů a tak dále.

Vidíme tedy, že pokud ke znalosti další hodnoty potřebujeme vždy znát m hodnot předchozích, tak k rozbehnutí procesu potřebujeme také znát m hodnot počátečních. Máme tedy situaci, která je indukční, ale nehodí se k ani jednomu ze zatím probraných principů. Abychom se s touto situací uměli vyrovnat, představíme si ještě jednu verzi indukce.

!

5a.11. Modifikovaný princip matematické indukce.

Nechť $n_0 \in \mathbb{Z}$, nechť $V(n)$ je vlastnost celých čísel, která má smysl pro $n \geq n_0$. Nechť $m \in \mathbb{N}$.

Předpokládejme, že následující předpoklady jsou splněny:

(0) $V(n_0), V(n_0 + 1), V(n_0 + 2), \dots, V(n_0 + m - 1)$ platí.

(1) Pro každé $n \in \mathbb{Z}$, $n \geq n_0 + m - 1$ je pravdivá následující implikace: Jestliže platí $V(k)$ pro všechna $k = n - m + 1, n - m + 2, \dots, n$, pak platí i $V(n + 1)$.

Potom $V(n)$ platí pro všechna $n \in \mathbb{Z}$, $n \geq n_0$.

Indexy ve formulaci principu asi na první pohled vypadají trochu divoce, pomůžeme si představíme konkrétní aplikaci. Pro jednoduchost zvolíme $n_0 = 1$, tedy pracujeme na \mathbb{N} , a rozmyslíme si situaci, která se odvolává na $m = 3$ předchozí výsledky. K nastartování procesu pak potřebujeme znát situaci pro $n = 1, 2, 3$, což opravdu odpovídá indexům $n_0, n_0 + 1, \dots, n_0 + m - 1$.

Indukční krok by nás měl zavést o krok dál, než známe. Po kroku (0) má poslední známá situace (to, čemu v indukčním kroku říkáme n) index 3, pak očekáváme informaci o $n + 1 = 4$, což je první zatím neznámá situace, souhlasí to. Takže indukční kroky nás zajímají pro $n \geq 3$ neboli pro $n \geq n_0 + m - 1$, to odpovídá zápisu z principu. My ovšem nepotřebujeme znát jen tuto poslední situaci, ale i několik předchozích tak, aby jich celkem bylo m , takže první situace potřebná pro indukční krok kupředu musí mít index $n - m + 1$.

Obecná formulace tedy dává smysl. Naštěstí se nemusíme ty rozsahy indexů učit, důležité je znát princip a v konkrétním příkladě pak rozličné hodnoty vyplývající z dané situace.

Poznamenejme, že „modifikovaný princip“ je pracovní název, abychom se na něj mohli v tomto textu odvolávat, tento princip nemá univerzálně přijímaný název.

I tento princip je ve skutečnosti ekvivalentní slabému principu. V jednom směru je to zjevné. Pokud výše použijeme hodnotu $m = 1$, tak dostáváme přímo slabý princip indukce, takže ten náš modifikovaný vlastně zahrnuje slabý princip, tudíž toho dokáže přinejmenším stejně.

Důkaz opačným směrem, že věci dosažitelné modifikovaným principem jdou i pomocí slabého, se dělá podobně jako u Věty 5a.10. Definuje se pomocná vlastnost W , kdy $W(n)$ platí právě tehdy, pokud platí $V(n)$, $V(n+1)$ až $V(n+m-1)$. Detaily necháme na čtenáři. Teď si ukážeme příklad, kdy je modifikovaný princip indukce přirozeným nástrojem.

Příklad 5a.n: Dokážeme, že pomocí mincí s hodnotami 3 a 5 dokážeme přímo vyplatit libovolnou korunovou částku větší než 7. (Tím myslíme, že tuto částku rovnou dáme, bez nějakých fíglů s vrácením, to pak jde zaplatit jakákoliv částka.)

Poznámka: Tříkoruny opravdu existovaly, od roku 1953 papírová, od roku 1965 kovová, ta pak byla roku 1972 zrušena (protože si Němci stěžovali, že ji lze v jejich automatech používat místo mnohem hodnotnější mince germánské).

Takže pro $n \geq 8$ dokážeme $V(n)$: Je možné vyplatit n korun tříkorunami a pětikorunami.

Použijeme modifikovaný princip indukce, který používá zpětného chodu o tři (tedy $m = 3$). Budeme proto potřebovat také tři počáteční hodnoty.

(0) Snadno ověříme, že platí $V(8)$, $V(9)$ a $V(10)$.

Poznámka: Další hodnotu, kterou potřebujeme ověřit, je 11, tedy první indukční krok musí mít $n+1 = 11$ neboli $n = 10$. Tím je dán rozsah pro další část.

(1) Nechť $n \geq 10$, předpokládejme, že platí $V(n-2)$, $V(n-1)$ a $V(n)$. Potřebujeme ukázat, že platí i $V(n+1)$.

Ale to je snadné. Jestliže $n \geq 10$, pak $n-2 \geq 8$, proto podle indukčního předpokladu dokážeme vyplatit $n-2$. Pak ještě přihodíme tříkorunu a vyplatili jsme $n+1$, přesně jak jsme potřebovali.

Z (0) a (1) vyplývá pravdivost $V(n)$ pro všechna celá čísla $n \geq 8$.

Teoreticky víme, že by tento příklad měl jít řešit i slabou indukcí. Zde to jde dokonce i bez nějaké pomocné vlastnosti W jako v obecném důkazu, dokážeme (slabou) indukcí přímo naši V definovanou výše.

(0) Dokážeme vyplatit $8 = 3 + 5$, tedy $V(8)$ platí.

(1) Mějme libovolné $n \geq 8$ a předpokládejme, že $V(n)$ platí. Chceme ukázat, že platí i $V(n+1)$.

Podle indukčního předpokladu umíme vyplatit n . Jestli je v tom vyplácení také pětikoruna, tak ji nahradíme dvěma tříkorunami a vyplatili jsme $n+1$, hotovo. Pokud by těch n bylo vyplaceno samými tříkačkami, tak díky $n \geq 8$ musí být nejméně tři. Vezmeme tedy tři konkrétní tříkoruny (celkem 9) a nahradíme je dvěma pětikorunami (celkem 10) a máme vyplaceno $n+1$.

Tím jsme vyčerpali všechny možnosti a důkaz (1) je hotov.

Z (0) a (1) vyplývá pravdivost V pro všechna celá čísla $n \geq 8$.

První způsob byl jednodušší, což není překvapující, lépe se hodil k podstatě problému.

△

Tím jsme probrali všechny základní podoby principu indukce a shrneme si praktické použití.

S Algoritmus 5a.12. pro dokazování klasickou indukcí.

1. Ujasněte si, co vlastně chcete dokazovat, napište to jako vlastnost $V(n)$ závisející na celočíselném parametru n , kde se n bere pro všechna $n \geq n_0$ a n_0 je startovací hodnota.
2. Napište si tvrzení $V(n+1)$ a zkuste najít způsob, jak v tomto tvrzení najít/vytvořit situaci z $V(n)$ či dalších předchozích $V(k)$.
3. Podle 2. se rozhodněte, kterou verzi indukce použijete. Pokud vám k $V(n+1)$ stačí $V(n)$, je slabý princip nejlepší. Pokud potřebujete více předchozích hodnot, ale vždy stejný počet $n - m + 1, n - m + 2$ až n , modifikovaný princip může být tím nejlepším. Pokud se vracíte do minulosti nepravidelně, je to příklad na silný princip indukce.
4. Rozmyslete si, které hodnoty musíte znát na počátku, aby se proces indukce mohl rozbehřout. Pak si rozmyslete, která hodnota n vám pro $n + 1$ dá první neznámou situaci. Tím je určen rozsah pro indukční krok.
5. Proveďte vlastní důkaz:
 - a) Dokažte platnost $V(n)$ pro počáteční hodnoty rozmyšlené v bodě 4.;
 - b) Zvolte libovolné n z rozsahu rozmyšleného pro indukční krok v bodě 4., stanovte indukční předpoklad a s jeho pomocí dokažte platnost $V(n+1)$.
6. Zkontrolujte, že důkaz v části 5 b) je správný, tedy vychází z toho, co předpokládáte jako pravdivé, a po korektních krocích končí tím, co chcete dokázat.

△

Předchozí příklady tento algoritmus snad dostatečně ilustrovaly, pokud čtenáři ještě nejsou některé úvahy úplně jasné, tak si zkusí spočítat příklady třeba ze cvičení 5a.15, 6a.13, 5b.4 a 5b.5.

Základní blok této kapitoly uzavřeme doplněním směru, který jsme vynechali v důkazu věty 5a.8. Připomeňme, že jsme již dokázali, že ze slabého principu indukce plyne platnost principu dobrého uspořádání. ve Větě 5a.10 jsme pro změnu odvodili, že slabý princip plyne ze silného. Abychom kolečko uzavřeli, potřebujeme ukázat, že silný princip indukce plyne z principu dobrého uspořádání. My ukážeme dokonce něco mnohem obecnějšího.

Věta 5a.13. (o dobře uspořádané indukci)

Nechť (A, \preceq) je dobře uspořádaná množina. Nechť $V(a)$ je vlastnost prvků $a \in A$.

Předpokládejme, že je splněna následující podmínka zvaná **indukční krok**:

Pro všechna $a \in A$ platí: Jestliže $V(x)$ platí pro všechna $x \in A$ splňující $x \prec a$, pak platí i $V(a)$.

Pak platí $V(a)$ pro všechna $a \in A$.

Důkaz (poučný): Nepřímý důkaz neboli dokážeme obměnu implikace „indukční krok \implies platnost V “. Předpokládejme, že není pravda, že V platí pro všechna $a \in A$. Ukážeme, že pak neplatí ani indukční krok.

Jestliže není $V(x)$ vždy splněno, pak je množina $M = \{y \in A; V(y) \text{ neplatí}\}$ neprázdná a díky dobrému uspořádání má svůj nejmenší prvek, nazvěme jej a . Označme $X = \{x \in A; x \prec a\}$. Indukční krok pro naše a lze teď přepsat takto:

(I) Jestliže $V(x)$ platí pro všechna $x \in X$, pak platí i $V(a)$.

Všimněte si, že a coby nejmenší prvek M splňuje $a \in M$, tedy $V(a)$ neplatí. Abychom tedy ukázali neplatnost této implikace, stačí ukázat, že je splněn její předpoklad. To uděláme rozborem podle toho, jaké je X .

Jestliže je X prázdná, tak je předpoklad implikace automaticky splněn a implikace neplatí.

Druhá možnost je, že X prázdná není. Tyto prvky ovšem nemohou být z M , protože kdyby bylo nějaké $x \in M \cap X$, tak a jako nejmenší prvek M splňuje $a \preceq x$, $x \in X$ zase dává $x \prec a$ a máme spor, viz Lemma 4b.4 (ii).

To znamená, že prvky z X nejsou v M , jinak řečeno, $V(x)$ pro ně platí. Takže zase je předpoklad implikace (I) splněn a implikace tím pádem neplatí.

Více možností pro X (prázdná-neprázdná) není, takže ve všech případech (I) neplatí a důkaz je hotov.

□

Opravdu již tato věta dává hledanou implikaci? Množina (\mathbb{N}, \leq) je dobře uspořádaná, tudíž podle právě dokázané věty k důkazu platnosti nejaké vlastnosti V na \mathbb{N} stačí dokázat následující implikaci:

(I) Nechť $n \in \mathbb{N}$. Jestliže V platí pro všechna $k < n$, pak platí i pro $V(n)$.

Co to znamená? Pokud posuneme index o jedničku (substituce $n' = n - 1$, chcete-li, ted' $n' \in \mathbb{N}_0$), pak to říká následující: „Jestliže V platí pro všechna $k \leq n$, pak platí i pro $n + 1$ “, což je přesně krok (1) ze silné indukce.

Počkat, řekne teď pozorný student, a co krok (0)? Ten je v tom schován také, ale trikem. Co když tu původní implikaci (I) aplikujeme na $n = 1$ (popřípadě tu přepsanou na $n = 0$)? Dostáváme výrok „Jestliže V platí pro všechna k splňující $k \in \mathbb{N}$, $k < 1$, tak platí i $V(1)$ “. Jsme tedy v situaci, kdy se snažíme dokázat $V(1)$ za pomocí předchozích případů, jenže ony žádné takové nejsou. Nezbývá tedy, než dokázat $V(1)$ přímo, čímž vznikne základní krok (0). Máme tedy celou silnou indukci.

Obecný princip silné indukce, tak jak jsme jej formulovali, pak dostáváme obdobnou úvahou aplikovanou na množinu $(\{n_0, n_0 + 1, n_0 + 2, \dots\}, \leq)$, pro $n_0 \in \mathbb{Z}$. Poučení tedy je, že vlastně existuje jen jeden princip indukce, velmi obecný (tak to vidí lidé zabývající se základy matematiky), ale pro pohodlí praktického uživatele si z něj vytváříme různé podverze, další ještě přibude v následující kapitolce.

Zajímavé je, že indukci můžeme používat i na jiných množinách než \mathbb{N} . I pak se v konkrétních případech ověřování indukční implikace rozpadá fakticky na dva případy:

(0) Nechť m je nejmenší prvek množiny A . Pak je množina $\{x \in A; x \prec a\}$ prázdná, což znamená, že předpoklad implikace „Jestliže $V(x)$ platí pro všechna $x \in A$ splňující $x \prec m$, pak platí i $V(m)$ “ je vždy splněn automaticky. Proto k důkazu její platnosti musíme ukázat, že $V(a)$ platí vždycky, tedy dokazujeme to bez pomoci ostatních $V(k)$ jinými slovy to je ten základní krok.

(1) Jestliže a není nejmenší prvek množiny A , pak je $\{x \in A; x \prec a\}$ neprázdná a my máme při důkazu implikace k dispozici indukční předpoklady, přesně jak jsme zvyklí u indukčního kroku.

Z praktického pohledu tedy děláme věci jako obvykle, nicméně matematici znalecky ocení, že se nám to ve větě podařilo chytře vyjádřit jednou implikací. Je to elegantní, je to přesné, je to záhadné, tak to máme rádi.

Poznámka: Indukci lze dokonce použít i na množinách, na kterých nemáme plnou sílu částečného uspořádání. Například lze dokázat, že princip matematické indukce platí na **ostře usporádané** množině (A, \prec) právě tehdy, když je tato množina fundovaná (viz 4d).

Poznámka: Jak jsme již viděli, indukci lze používat i u jiných množin než \mathbb{N} . V mnoha případech je vcelku zjevné, jak princip indukce modifikovat.

A) Chceme-li dokázat vlastnost V o sudých nezáporných číslech, uděláme to takto:

(0) $V(0)$ platí.

(1) $V(n) \implies V(n+2)$ platí pro $n \geq 0$.

Pokud bychom chtěli jen kladná sudá čísla, začali bychom dvojkou.

B) Chceme-li dokázat vlastnost V o kladných lichých číslech, uděláme to takto:

(0) $V(1)$ platí.

(1) $V(n) \implies V(n+2)$ platí pro $n \geq 0$.

C) Chceme-li dokázat vlastnost V o číslech typu 13^n , uděláme to takto:

(0) $V(1)$ platí.

(1) $V(n) \implies V(13n)$ platí pro $n \geq 1$.

Opravdu? Dle (0) platí $V(1)$. Podle (1) pak platí i $V(13 \cdot 1) = V(13)$. A znova (1) s předpokladem $V(13)$ dává platnost $V(13 \cdot 13) = V(13^2)$. A znova (1) s předpokladem $V(13^2)$ dává $V(13 \cdot 13^2) = V(13^3)$ a tak dále.

Pozor, pokud uděláme toto:

(0) $V(0)$ platí,

(1) $V(n) \implies V(13n)$ platí,

tak tím dokážeme jen $V(0)$! Proč? Podle (0) dostaneme platnost $V(0)$. Pak aplikujeme (1) a dostaneme platnost pro $V(13 \cdot 0) = V(0)$, oops.

△

Tím už se ale vlastně dostáváme ke strukturální indukci, což je téma příští kapitolky.

5a.14 Poznámka: Pro doplnění ještě uvedeme jeden ekvivalentní indukční princip, který se někdy používá. Česky se mu říká „sestupná indukce“, ale od indukce zatím probrané se liší tím, že sestupnou indukcí platnost vlastnosti vyvracíme.

! 5a.15. Princip sestupné indukce (Infinite descent proof).

Nechť $V(n)$ je vlastnost přirozených čísel.

Předpokládejme, že je splněn následující předpoklad:

(1) Pro každé $n \in \mathbb{N}$ je pravdivá implikace:

Jestliže platí $V(n)$, pak existuje $k \in \mathbb{N}$ takové, že $k < n$ a $V(k)$ platí.

Potom $V(n)$ neplatí pro žádné $n \in \mathbb{N}$.

Argument, proč toto funguje, je následující: Kdyby náhodou $V(n)$ platilo pro nějaké n , tak by to podle (1) muselo platit i pro menší číslo, takže podle (1) pro ještě menší číslo a tak dále, jenž problém máme v tom, že v \mathbb{N} není pro takovýto nekonečný řetězec zmenšujících se čísel místo, protože je ta množina „dole“ useknutá. Řečeno formálně, princip sestupné indukce je ekvivalentní faktu, že neexistuje nekonečná klesající posloupnost přirozených čísel, čímž jsme u pojmu fundovanosti z kapitoly 4d. není těžké dokázat, že i tento princip je ekvivalentní ostatním principům indukce a principu dobrého uspořádání.

Jako příklad si ukážeme, jak se dá tímto principem zapsat známý Euklidův důkaz, že $\sqrt{2}$ není racionální číslo. Definujme tuto vlastnost:

$V(n)$ říká, že existuje přirozené číslo p splňující $\sqrt{2} = \frac{p}{n}$.

To, že $\sqrt{2} \notin \mathbb{Q}$, je ekvivalentní právě tomu, že $V(n)$ není splněno pro žádné $n \in \mathbb{N}$.

Abychom to dokázali, ukážeme pravdivost (1). Pokud by platilo $V(n)$, tak $\sqrt{2} = \frac{p}{n}$. Pak $2 = \frac{p^2}{n^2}$ neboli $p^2 = 2n^2$. To znamená, že 2 dělí p^2 , a protože je 2 prvočíslo, musí dělit rovnou p . Máme tedy $p = 2a$ pro nějaké $a \in \mathbb{N}$. Pak $4a^2 = 2n^2$ a stejným argumentem ukážeme, že také 2 dělí n , tudíž $n = 2k$ pro nějaké $k \in \mathbb{N}$. Pak jde ve zlomku zkrátit a máme $\sqrt{2} = \frac{a}{k}$. Našli jsme tedy $k \in \mathbb{N}$ takové, že $k < n$ a $V(k)$ platí.

Tím je tedy (1) ověřeno a podle principu sestupné indukce to již dokazuje, že žádné $V(n)$ nemůže platit.

Je samozřejmě možné pomocí tohoto principu také vlastnosti dokazovat jednoduchým trikem, kdy si jako V vezmeme negaci dokazované vlastnosti, pak V vyvrátíme a tím ta původní vlastnost musí platit.

△

A to už bylo opravdu to poslední z této kapitoly.

Cvičení

Cvičení 5a.1 (rutinní, zkouškové): Dokažte, že následující vzorce platí pro všechna $n \in \mathbb{N}$:

- (i) $2 + 4 + 6 + \cdots + (2n) = n(n+1)$;
- (ii) $1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1)$;
- (iii) $1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{1}{6}n(n+1)(2n+1)$;
- (iv) $\frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \frac{1}{5 \cdot 7} + \cdots + \frac{1}{(2n-1) \cdot (2n+1)} = \frac{n}{2n+1}$;
- (v) $1 \cdot 1! + 2 \cdot 2! + \cdots + n \cdot n! = (n+1)! - 1$;
- (vi) $1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!} \leq 2 - \frac{1}{n!}$;
- (vii) $1 + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{n^2} \leq 2 - \frac{1}{n}$;
- (viii) $n! < n^n$ (toto pro $n \geq 2$).

Cvičení 5a.2 (poučné): Pro reálná (i komplexní) čísla (a dokonce pro vektory) je důležitá známá „trojúhelníková nerovnost“ $|x+y| \leq |x| + |y|$. Dokažte její následující zobecnění:

Jestliže $x_1, \dots, x_n \in \mathbb{R}$, pak $\left| \sum_{k=1}^n x_k \right| \leq \sum_{k=1}^n |x_k|$.

Cvičení 5a.3 (poučné): Najděte chybu v následujícím „důkazu“, že pro všechna nenulová reálná a a pro všechna $n \in \mathbb{N}_0$ platí $a^n = 1$.

(0) Pro $n = 0$ evidentně platí $a^0 = 1$.

(1) Nechť $n \geq 0$, předpokládejme $a^n = 1$. Pak $a^{n+1} = \frac{a^n \cdot a^n}{a^{n-1}} = \frac{1 \cdot 1}{1} = 1$.

Cvičení 5a.4 (poučné): Uvažujte vlastnost $V(n)$: $1 + 2 + 3 + \cdots + n = \frac{1}{2}(n-1)(n+2)$.

Ukažte, že $V(n) \implies V(n+1)$ pro všechna $n \in \mathbb{N}$.

Platí $V(n)$?

Cvičení 5a.5 (poučné): Uvažujte vlastnost $V(n)$: $2 + 4 + 6 + \cdots + 2n = n^2 + n - 13$.

Ukažte, že $V(n) \implies V(n+1)$ pro všechna $n \in \mathbb{N}$.

Platí $V(n)$?

Cvičení 5a.6 (poučné): Pokusíme se sečítat všechna čísla typu $\frac{1}{k}$. Definujme

$$H(n) = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

Dokažte, že pro všechna $n \in \mathbb{N}$ platí následující:

(i) $H(1) + H(2) + \cdots + H(n) = (n+1)H(n) - n$.

(ii) $H(2^n) \leq 1 + n$.

(iii) $H(2^n) \geq 1 + \frac{n}{2}$.

Nápověda: $\frac{1}{2^n+1} \geq \frac{1}{2^{n+1}}$, $\frac{1}{2^n+2} \geq \frac{1}{2^{n+1}}$, \dots , $\frac{1}{2^n+2^n-1} \geq \frac{1}{2^{n+1}}$, $\frac{1}{2^n+2^n} = \frac{1}{2^{n+1}}$.

Poznámka: Z (iii) hned vidíme, že nekonečný součet $1 + \frac{1}{2} + \frac{1}{3} + \cdots$ nemůže být konečné číslo.

Cvičení 5a.7 (poučné):(i) Uvažujme $V(n): \frac{1}{2} \cdot \frac{3}{4} \cdots \frac{2n-1}{2n} < \frac{1}{\sqrt{3n}}$.Ukažte, že $V(1)$ platí.Ukažte, že standardní indukční důkaz $V(n) \implies V(n+1)$ nelze provést.(ii) Uvažujme $W(n): \frac{1}{2} \cdot \frac{3}{4} \cdots \frac{2n-1}{2n} < \frac{1}{\sqrt{3n+1}}$. Dokažte, že platí pro $n \geq 2$.(iii) Dokažte, že $V(n)$ platí pro všechna $n \in \mathbb{N}$.**Cvičení 5a.8** (rutinní, poučné): Nechť $n \in \mathbb{N}$, $n \geq 2$. Uvažujme množiny A_1, \dots, A_n, A_{n+1} a zobrazení $T_i: A_i \mapsto A_{i+1}$ pro $i = 1, \dots, n$. Dokažte, že jestliže jsou všechna tato zobrazení invertibilní, pak je invertibilní i $T_n \circ \cdots \circ T_1$ a $(T_n \circ \cdots \circ T_1)^{-1} = T_1^{-1} \circ \cdots \circ T_n^{-1}$ (viz Věta 2b.6 a 2b.7).**Cvičení 5a.9** (rutinní, poučné): Nechť $T: A \mapsto A$ je invertibilní zobrazení. Dokažte, že pro $n \in \mathbb{N}$ platí $(T^n)^{-1} = (T^{-1})^n$ (viz Věta 2b.7).**Cvičení 5a.10** (rutinní, poučné): Nechť $n \in \mathbb{N}$, $n \geq 2$. Uvažujme množiny A_1, \dots, A_n, A_{n+1} a zobrazení $T_i: A_i \mapsto A_{i+1}$ pro $i = 1, \dots, n$. Dokažte následující:(i) Jestliže jsou všechna tato zobrazení prostá, pak je prosté i $T_n \circ \cdots \circ T_1$.(ii) Jestliže jsou všechna tato zobrazení na, pak je na i $T_n \circ \cdots \circ T_1$.(iii) Jestliže jsou všechna tato zobrazení bijekce, pak je bijekce i $T_n \circ \cdots \circ T_1$.

(Viz Fakt 2b.10.)

Cvičení 5a.11 (zkouškové): Dokažte, že jsou-li A_i pro $i = 1, 2, \dots, n$ konečné množiny, pak je i $\bigcup_{i=1}^n A_i$ konečná a $\left| \bigcup_{i=1}^n A_i \right| \leq \sum_{i=1}^n |A_i|$.Jsou-li navíc navzájem disjunktní, tak $\left| \bigcup_{i=1}^n A_i \right| = \sum_{i=1}^n |A_i|$.

(Viz Věta 2c.7 a 2c.8.)

Cvičení 5a.12 (poučné): Dokažte tzv. Bernoulliho nerovnost: Jestliže $h > -1$, pak $(1+h)^n \geq 1+hn$ pro všechna $n \in \mathbb{N}_0$.**Cvičení 5a.13:** Dokažte indukcí, že $(a-b)$ dělí $(a^n - b^n)$ pro všechna $n \in \mathbb{N}$.**Cvičení 5a.14** (zkouškové): Uvažujme matici $A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$, kde $a, b \in \mathbb{R}$. Dokažte, že pak pro všechna $k \in \mathbb{N}$ platí $A^k = \begin{pmatrix} a^k & 0 \\ 0 & b^k \end{pmatrix}$.**Cvičení 5a.15** (poučné): Hra Nim se hraje takto. Na začátku je hromádka s n zápalkami. Hrají střídavě dva hráči, každý z nich v jednom tahu odebere 1, 2 nebo 3 zápalky. Hráč, na kterého zbyde poslední zápalka, prohrává. Dokažte, že jestliže n po dělení 4 dá zbytek 1, tak má druhý hráč výherní strategii. Jinak má výherní strategii první hráč.

Nápověda: Rozmyslete si, že jestliže je na tahu soupeř a je tam přesně 5 zápalek, tak ať už udělá cokoliv, dokážete zahrát tak, aby v dalším tahu prohrál.

Cvičení 5a.16 (poučné): Teď si trochu pohrajeme se šířením neoficiálních informací. Představme si n osob, každá z nich zná na začátku určitou informaci, kterou ostatní neznají. Tyto osoby začnou spolu porůznu po dvou hovořit, a vždy když dvě osoby spolu hovoří, tak si sdělí vše, co zrovna znají.Označme jako $G(n)$ nejmenší počet takových vzájemných rozhovorů nutný k tomu, aby nakonec všichni věděli všechno.Je snadné si rozmyslet, že $G(1) = 0$, $G(2) = 1$, $G(3) = 3$, $G(4) = 4$.Dokažte, že $G(n) \leq 2n - 4$ pro všechna n .Poznámka: Dá se ukázat, že ve skutečnosti je tam rovnost, není možné to provést za méně než těch $2n - 4$ hovorů. To už je ale těžký problém.**Řešení:****5a.1:** (i): (0) $V(1)$ říká $2 = 1 \cdot 2$, platí. (1) Nechť $n \in \mathbb{N}$. Předpoklad: $2 + 4 + 6 + \cdots + (2n) = n(n+1)$.Dokázat: $2 + 4 + 6 + \cdots + (2n+2) = (n+1)(n+2)$. Dekompozice:

$$2 + 4 + 6 + \cdots + (2n+2) = [2 + 4 + 6 + \cdots + (2n)] + (2n+2) = [n(n+1)] + (2n+2) = n^2 + 3n + 2 = (n+1)(n+2).$$

(ii): (0) $V(1)$ říká $1 = \frac{1}{2}1 \cdot 2$, platí. (1) Nechť $n \in \mathbb{N}$. Předpoklad: $1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1)$.

Dokázat: $1 + 2 + 3 + \dots + (n+1) = \frac{1}{2}(n+1)(n+2)$. Dekompozice:

$$1 + 2 + 3 + \dots + (n+1) = [1 + 2 + 3 + \dots + n] + (n+1) = [\frac{1}{2}n(n+1)] + (n+1) = \frac{1}{2}(n^2 + 3n + 2) = \frac{1}{2}(n+1)(n+2).$$

(iii): (0) $V(1)$ říká $1^2 = \frac{1}{6}1 \cdot 2 \cdot 3$, platí. (1) Nechť $n \in \mathbb{N}$. Předpoklad: $1^2 + 2^2 + 3^2 + \dots + n^2 = \frac{1}{6}n(n+1)(2n+1)$.

Dokázat: $1^2 + 2^2 + 3^2 + \dots + (n+1)^2 = \frac{1}{6}(n+1)(n+2)(2n+3)$. Dekompozice:

$$\begin{aligned} 1^2 + 2^2 + 3^2 + \dots + (n+1)^2 &= [1^2 + 2^2 + 3^2 + \dots + n^2] + (n+1)^2 = [\frac{1}{6}n(n+1)(2n+1)] + (n+1)^2 = \\ &= \frac{1}{6}(2n^3 + 9n^2 + 13n + 6) = \frac{1}{6}(n+1)(n+2)(2n+3). \end{aligned}$$

(iv): (0) $V(1)$ říká $\frac{1}{3} = \frac{1}{3}$, platí. (1) Nechť $n \in \mathbb{N}$. Předpoklad: $\frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \dots + \frac{1}{(2n-1) \cdot (2n+1)} = \frac{n}{2n+1}$.

Dokázat: $\frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \dots + \frac{1}{(2n+1) \cdot (2n+3)} = \frac{n+1}{2n+3}$. Dekompozice: $\frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \dots + \frac{1}{(2n+1) \cdot (2n+3)} =$

$$= [\frac{1}{1 \cdot 3} + \frac{1}{3 \cdot 5} + \dots + \frac{1}{(2n-1) \cdot (2n+1)}] + \frac{1}{(2n+1) \cdot (2n+3)} = [\frac{n}{2n+1}] + \frac{1}{(2n+1) \cdot (2n+3)} = \frac{2n^2 + 3n + 1}{(2n+1)(2n+3)} = \frac{n+1}{2n+3}.$$

(v): (0) $V(1)$ říká $1 \cdot 1 = 2 - 1$, platí. (1) Nechť $n \in \mathbb{N}$. Předpoklad: $1 \cdot 1! + 2 \cdot 2! + \dots + n \cdot n! = (n+1)! - 1$.

Dokázat: $1 \cdot 1! + 2 \cdot 2! + \dots + n \cdot n! + (n+1) \cdot (n+1)! = (n+2)! - 1$. Dekompozice:

$$\begin{aligned} 1 \cdot 1! + 2 \cdot 2! + \dots + n \cdot n! + (n+1) \cdot (n+1)! &= [1 \cdot 1! + 2 \cdot 2! + \dots + n \cdot n!] + (n+1) \cdot (n+1)! = [(n+1)! - 1] + (n+1) \cdot (n+1)! = \\ &= (n+1)! + (n+1) \cdot (n+1)! - 1 = (n+2)(n+1)! - 1 = (n+2)! - 1. \end{aligned}$$

(vi): (0) $V(1)$ říká $1 \leq 2 - 1$, platí. (1) Nechť $n \in \mathbb{N}$. Předpoklad: $1 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!} \leq 2 - \frac{1}{n!}$.

Dokázat: $1 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{(n+1)!} \leq 2 - \frac{1}{(n+1)!}$. Dekompozice: $1 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{(n+1)!} =$

$$= [1 + \frac{1}{2!} + \frac{1}{3!} + \dots + \frac{1}{n!}] + \frac{1}{(n+1)!} \leq [2 - \frac{1}{n!}] + \frac{1}{(n+1)!} = 2 - \frac{(n+1)-1}{(n+1)!} = 2 - \frac{n}{(n+1)!} \leq 2 - \frac{1}{(n+1)!}.$$

(vii): (0) $V(1)$ říká $1 \leq 2 - 1$, platí. (1) Nechť $n \in \mathbb{N}$. Předpoklad: $1 + \frac{1}{4} + \frac{1}{9} + \dots + \frac{1}{n^2} \leq 2 - \frac{1}{n}$.

Dokázat: $1 + \frac{1}{4} + \frac{1}{9} + \dots + \frac{1}{(n+1)^2} \leq 2 - \frac{1}{n+1}$. Dekompozice: $1 + \frac{1}{4} + \frac{1}{9} + \dots + \frac{1}{(n+1)^2}$

$$= [1 + \frac{1}{4} + \frac{1}{9} + \dots + \frac{1}{n^2}] + \frac{1}{(n+1)^2} \leq [2 - \frac{1}{n}] + \frac{1}{(n+1)^2} = 2 - \frac{(n+1)^2 - n}{n(n+1)^2} = 2 - \frac{n^2 + n + 1}{n(n+1)^2} \leq 2 - \frac{n^2 + n}{n(n+1)^2} = 2 - \frac{1}{n+1}.$$

(viii): (0) $V(2)$ říká $2 < 4$, platí. (1) Nechť $n \geq 2$. Předpoklad: $n! < n^n$. Dokázat: $(n+1)! < (n+1)^{n+1}$.

Dekompozice: $(n+1)! = (n+1)n! < (n+1)n^n < (n+1)(n+1)^n = (n+1)^{n+1}$.

5a.2: (0) Pro $n = 1$ určitě platí $|x_1| = |x_1|$, platí. (1) Nechť $n \in \mathbb{N}$. Předpokládejme, že platí $\left| \sum_{k=1}^n x_k \right| \leq \sum_{k=1}^n |x_k|$.

$$\text{Pak } \left| \sum_{k=1}^{n+1} x_k \right| = \left| x_{n+1} + \sum_{k=1}^n x_k \right| \leq |x_{n+1}| + \left| \sum_{k=1}^n x_k \right| \leq |x_{n+1}| + \sum_{k=1}^n |x_k| = \sum_{k=1}^{n+1} |x_k|.$$

5a.3: V indukčním kroku je použito $a^{n-1} = 1$, což ale nemusí platit, indukční předpoklad dává jen $a^n = 1$.

5a.4: Předpokládejme $1 + 2 + 3 + \dots + n = \frac{1}{2}(n-1)(n+2)$. Pak

$$1 + 2 + 3 + \dots + (n+1) = [1 + 2 + 3 + \dots + n] + (n+1) = \frac{1}{2}(n-1)(n+2) + (n+1) = \frac{1}{2}(n^2 + 3n) = \frac{1}{2}n(n+3).$$

$V(n)$ ale neplatí pro žádné $n \in \mathbb{N}$.

5a.5: Předpokládejme $2 + 4 + 6 + \dots + 2n = n^2 + n - 13$. Pak $2 + 4 + 6 + \dots + 2(n+1)$

$$= [2 + 4 + 6 + \dots + 2n] + (2n+2) = [n^2 + n - 13] + 2n+2 = (n^2 + 2n + 1) + (n+1) - 13 = (n+1)^2 + (n+1) - 13.$$

$V(n)$ ale neplatí pro žádné $n \in \mathbb{N}$.

5a.6: (i): indukcí: (0) $n = 1$ dává $1 = 2 \cdot 1 - 1$.

(1) Předpoklad $H(1) + H(2) + \dots + H(n) = (n+1)H(n) - n$. Pak $H(1) + H(2) + \dots + H(n+1) =$

$$= [H(1) + H(2) + \dots + H(n)] + H(n+1) = (n+1)\left[1 + \frac{1}{2} + \dots + \frac{1}{n}\right] - n + \left(1 + \frac{1}{2} + \dots + \frac{1}{n+1}\right) =$$

$$= (n+1)\left(1 + \frac{1}{2} + \dots + \frac{1}{n+1}\right) - (n+1)\frac{1}{n+1} - n + \left(1 + \frac{1}{2} + \dots + \frac{1}{n+1}\right) = (n+2)\left(1 + \frac{1}{2} + \dots + \frac{1}{n+1}\right) - n - 1 =$$

$$= (n+2)H(n+1) - (n+1).$$

(ii): Indukcí: (0) $n = 1$: $H(2) \leq 1 + 1$ znamená $1 + \frac{1}{2} \leq 2$, pravda.

(1) Předpoklad $H(2^n) \leq 1 + n$. Pak $H(2^{n+1}) = 1 + \frac{1}{2} + \dots + \frac{1}{2^{n+1}} = [1 + \frac{1}{2} + \dots + \frac{1}{2^n}] + \frac{1}{2^{n+1}} + \frac{1}{2^{n+2}} + \dots + \frac{1}{2^{n+1}}$

$$\leq 1 + n + \frac{1}{2^n} + \frac{1}{2^n} + \dots + \frac{1}{2^n} = 1 + n + 2^n \frac{1}{2^n} = 1 + (n+1), \text{ neboť těch zlomků je přesně } 2^n, \text{ jdou od } \frac{1}{2^{n+1}} \text{ po } \frac{1}{2^{n+1}} = \frac{1}{2 \cdot 2^n} = \frac{1}{2^{n+2}}.$$

(iii): Indukcí: (0) $n = 1$: $H(2) \geq 1 + \frac{1}{2}$ znamená $1 + \frac{1}{2} \geq 1 + \frac{1}{2}$, pravda.

(1) Předpoklad $H(2^n) \geq 1 + \frac{n}{2}$. Pak $H(2^{n+1}) = 1 + \frac{1}{2} + \dots + \frac{1}{2^{n+1}} = [1 + \frac{1}{2} + \dots + \frac{1}{2^n}] + \frac{1}{2^{n+1}} + \frac{1}{2^{n+2}} + \dots + \frac{1}{2^{n+1}}$

$$\geq 1 + \frac{n}{2} + \frac{1}{2^{n+1}} + \frac{1}{2^{n+1}} + \dots + \frac{1}{2^{n+1}} = 1 + \frac{n}{2} + 2^n \cdot \frac{1}{2^{n+1}} = 1 + \frac{n+1}{2}.$$

5a.7: (i): $V(1)$: $\frac{1}{2} < \frac{1}{\sqrt{3}}$ platí.

Zkusíme dokázat indukční krok: Předpoklad $\frac{1}{2} \cdot \frac{3}{4} \dots \frac{2n-1}{2n} < \frac{1}{\sqrt{3n}}$. Pak

$$\frac{1}{2} \cdot \frac{3}{4} \dots \frac{2n+1}{2n+2} = \frac{1}{2} \cdot \frac{3}{4} \dots \frac{2n-1}{2n} \cdot \frac{2n+1}{2n+2} < \frac{1}{\sqrt{3n}} \cdot \frac{2n+1}{2n+2}. \text{ Chceme, aby platilo } \frac{1}{\sqrt{3n}} \cdot \frac{2n+1}{2n+2} \leq \frac{1}{\sqrt{3(n+1)}} \text{ neboli}$$

$(2n+1)\sqrt{3(n+1)} \leq (2n+2)\sqrt{3n}$, odtud umocněním $n+1 \leq 0$. Toto neplatí nikdy, a protože kroky byly ekvivalentní, nemohla platit ani výchozí nerovnost. Indukční krok se tedy dokázat nepovede.

(ii): (0) $W(2)$: $\frac{1}{2} \cdot \frac{3}{4} < \frac{1}{\sqrt{7}}$ platí.

Předpokládejme $\frac{1}{2} \cdot \frac{3}{4} \dots \frac{2n-1}{2n} < \frac{1}{\sqrt{3n+1}}$. Pak $\frac{1}{2} \cdot \frac{3}{4} \dots \frac{2n+1}{2n+2} = \frac{1}{2} \cdot \frac{3}{4} \dots \frac{2n-1}{2n} \cdot \frac{2n+1}{2n+2} < \frac{1}{\sqrt{3n+1}} \cdot \frac{2n+1}{2n+2}$. Chceme, aby platilo $\frac{1}{\sqrt{3n+1}} \cdot \frac{2n+1}{2n+2} \leq \frac{1}{\sqrt{3(n+1)+1}}$.

Zkusíme postup od konce: přepíšeme nerovnost na $(2n+1)\sqrt{3n+4} < (2n+2)\sqrt{3n+1}$, odtud umocněním a po úpravě $0 \leq n$. To platí, neboť zde máme $n \geq 2$. Všechny kroky byly ekvivalentní včetně umocnění, protože se umocňovala kladná čísla. Postup lze proto obrátit a z pravidlivého faktu $0 \leq n$ lze korektně dojít k žádané nerovnosti, pomocí ní dokončíme důkaz indukčního kroku:

$$\frac{1}{2} \cdot \frac{3}{4} \cdots \frac{2n+1}{2n+2} < \cdots < \frac{1}{\sqrt{3n+1}} \cdot \frac{2n+1}{2n+2} \leq \frac{1}{\sqrt{3(n+1)+1}}.$$

(iii): $V(1)$ už bylo ověřeno v (i). Pokud $n \geq 2$, pak pomocí $W(n)$ máme $\frac{1}{2} \cdot \frac{3}{4} \cdots \frac{2n-1}{2n} < \frac{1}{\sqrt{3n+1}} < \frac{1}{\sqrt{3n}}$.

5a.8: Indukce: (0) $n = 2$: $T_2 \circ T_1$ je invertibilní a $(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$ platí podle Věty 2b.6.

(1) Předpoklad: Pokud T_1, \dots, T_n splňují předpoklady, pak $T_n \circ \cdots \circ T_1$ je invertibilní a $(T_n \circ \cdots \circ T_1)^{-1} = T_1^{-1} \circ \cdots \circ T_n^{-1}$.

Mějme T_1, \dots, T_n, T_{n+1} splňující předpoklady, označme $S = T_n \circ \cdots \circ T_1$, podle indukčního předpokladu je S invertibilní a máme vzorec pro S^{-1} . Podle Věty 2b.6 je pak i $T_{n+1} \circ S$ invertibilní a $(T_{n+1} \circ S)^{-1} = S^{-1} \circ T_{n+1}^{-1}$, když dosadíme, dostaneme, že $T_{n+1} \circ T_n \circ \cdots \circ T_1$ je invertibilní a $(T_{n+1} \circ T_n \circ \cdots \circ T_1)^{-1} = T_1^{-1} \circ \cdots \circ T_n^{-1} \circ T_{n+1}^{-1}$.

5a.9: (0) $V(1)$ říká $(T^1)^{-1} = (T^{-1})^1$ neboli $T^{-1} = T^{-1}$, což platí.

(1) Předpokládejme, že $(T^n)^{-1} = (T^{-1})^n$ pro T invertibilní. Mějme teď invertibilní T , označme $S = T^n$. Pak je dle indukčního předpokladu S invertibilní a $S^{-1} = (T^{-1})^n$. Podle Věty 2b.6 je potom $T \circ S$ invertibilní a $(T \circ S)^{-1} = S^{-1} \circ T^{-1}$, po dosazení $(T^{n+1})^{-1} = (T \circ S)^{-1} = S^{-1} \circ T^{-1} = (T^{-1})^n \circ T^{-1} = (T^{-1})^{n+1}$.

5a.10: Hromadný důkaz pro (i) až (iii).

Indukční krok: Předpoklad pro n . Nechť dány $T_1, T_2, \dots, T_n, T_{n+1}$, které mají příslušnou vlastnost (prosté pro (i), na pro (ii), bijekce pro (iii)). Podle indukčního předpokladu má i $S = T_n \circ \cdots \circ T_1$ příslušnou vlastnost, pak podle Faktu 2b.10 tu vlastnost má i $T_{n+1} \circ S$ neboli $T_{n+1} \circ T_n \circ \cdots \circ T_1$.

5a.11: Indukce, pro $n = 1$ evidentně platí.

Indukční předpoklad: tvrzení platí pro n množin. Mějme konečné množiny A_1, \dots, A_N, A_{n+1} . Podle indukčního předpokladu je konečná množina $B = \bigcup_{i=1}^n A_i$ a platí $|B| \leq \sum_{i=1}^n |A_i|$. Pak podle Věty 2c.7 je konečná i množina

$$B \cup A_{n+1} = \bigcup_{i=1}^{n+1} A_i \text{ a platí } \left| \bigcup_{i=1}^{n+1} A_i \right| = |B \cup A_{n+1}| \leq |B| + |A_{n+1}| \leq \sum_{i=1}^{n+1} |A_i|.$$

Pokud jsou navíc disjunktní, pak podle indukčního předpokladu $|B| = \sum_{i=1}^n |A_i|$, množiny B a A_{n+1} jsou disjunktní

$$\text{a proto obdobně } \left| \bigcup_{i=1}^{n+1} A_i \right| = \sum_{i=1}^{n+1} |A_i|.$$

5a.12: Indukce: (0) Pro $n = 0$ nerovnost říká $(1+h)^0 \geq 1+0$, což platí.

(1) Předpoklad: Dáno $n \in \mathbb{N}_0$, pro každé $h > -1$ platí $(1+h)^n \geq 1+hn$. Pak

$$(1+h)^{n+1} = (1+h)(1+h)^n \geq (1+h)(1+hn) = 1+h(n+1)+h^2 \geq 1+h(n+1).$$

5a.13: $a^n - b^n = a^n - a^{n-1}b + a^{n-1}b - b^n = a^{n-1}(a-b) - b(a^{n-1} - b^{n-1})$.

5a.14: (1) Předpokládejme, že $V(n)$ platí. Pak $A^{n+1} = A \cdot A^k = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \cdot \begin{pmatrix} a^k & 0 \\ 0 & b^k \end{pmatrix} = \begin{pmatrix} a^{k+1} & 0 \\ 0 & b^{k+1} \end{pmatrix}$.

5a.15: $V(n)$: Jestliže je $4n+1$ zápalek, tak má druhý hráč výherní strategii.

(0) $n = 0$: Je jedna zápalka, zbyla na prvního hráče, ten prohrál.

(1) Předpoklad: Druhý hráč umí vyhrát hru s $4n+1$ zápalkami. Předpokládejme hru s $4(n+1)+1 = 4n+5$ zápalkami. První hráč vezme zápalky, povolený počet je 1, 2, 3. Cokoliv vezme, druhý hráč může vždy vzít tak, aby zbylo $4n+1$ zápalek a je na tahu první, tedy hra znova začíná s $4n+1$ zápalkami a druhý má vyherní strategii. Důkaz, že jinak má výherní strategii první hráč: Jestliže je počet zápalek $4n+2$, $4n+3$ nebo $4n+4$, tak první hráč odebere tak, aby zbylo $4n+1$ a je na tahu druhý hráč. Začná hra s $4n+1$ zápalkami, ve které původně první hráč hraje roli druhého a má výherní strategii.

5a.16: 1) Důkaz indukcí: (1) Mějme $n \in \mathbb{N}$, předpokládejme $G(n) \leq 2n-4$. Teď uvažujme $n+1$ osob. Jedna z nich někomu ze skupiny řekne svou informaci. Pak se odpojí, zbývající skupině o n lidech stačí $G(n)$ hovorů k tomu, aby v ní už všechni věděli všechno, včetně informace první osoby. Podle indukčního předpokladu na to stačí $2n-4$ hovorů. Pak někdo ze skupiny ještě řekne všechny informace té první osobě. Celkem stačí $1+(2n-4)+1 = 2n-2$ hovorů. Nevíme ale, jestli zrovna tato strategie je optimální, takže je to horní odhad, máme tedy $G(n+1) \leq 2n-2 = 2(n+1)-4$. (1) je dokázáno.

2) Důkaz přímý: Pro 1, 2, 3, 4 to platí. Nechť $n \geq 5$. Označíme si nějaké čtyři osoby jako 1, 2, 3, 4. Všechni ostatní si promluví s někým z této čtveřice. Pak tedy tato čtveřice jako skupina ví všechno.

Pak si informace navzájem vymění, na to stačí čtyři rozhovory 1–2, 3–4, 1–3, 2–4, tím každý ze skupiny ví všechno.

Nakonec třeba 1 promluví s těmi $n-4$ mimo skupinu. Celkem tedy stačí $(n-4)+4+(n-4) = 2n-4$ hovorů.

Poznámka: Kolik by bylo třeba hovorů, kdyby všech $n-1$ řeklo svou informaci člověku 1, ten pak zná vše, tak to řekne ostatním?

Kolik by bylo třeba hovorů, kdyby v té speciální skupině byli 2, popřípadě 3 lidi? A pět lidí?

5b. Rekurze a strukturální indukce

Indukce a rekurze (jak jsme již viděli, jde vlastně o dva pohledy na jednu věc) se často používá i k definování různých objektů. S definicemi indukcí jsme se již setkali v kapitole o množinách, například v definici mocniny T^n pro zobrazení, v předchozí kapitole jsme viděli indukci třeba v definici faktoriálu. Při důkazu korektnosti takové definice hraje obvykle zásadní roli indukce.

! Příklad 5b.a: Definujme funkci $f: \mathbb{N}_0 \mapsto \mathbb{Z}$ takto:

- (0) $f(0) = 1$.
- (1) $f(n+1) = -f(n)$ pro $n \geq 0$.

Tvrdíme, že f je takto dobré definována pro všechna $n \in \mathbb{N}_0$. Pro ilustraci si to vyzkoušíme. Máme $f(0) = 1$ přímo z definice. Pomocí indukčního kroku použitého s $n = 0$ vidíme, že $f(1) = f(0+1) = -f(0)$ a $f(0)$ již známe, tedy $f(1) = -1$. Teď lze použít (1) s $n = 1$ a dostáváme $f(2) = f(1+1) = -f(1) = 1$, podobně $f(3) = f(2+1) = -f(2) = -1$, $f(4) = f(3+1) = -f(3) = 1$ atd. Zdá se, že opravdu dokážeme najít hodnoty f na \mathbb{N}_0 , zároveň si můžeme tipnout, že platí $f(n) = (-1)^n$ pro $n \in \mathbb{N}_0$. Jak se dá čekat, takové tvrzení se dá nejlépe dokázat matematickou indukcí.

Pro $n_0 \in \mathbb{N}_0$ dokážeme indukcí vlastnost $V(n)$: Funkce f je dobré definována v n a $f(n) = (-1)^n$.

- (0) Pro $n = 0$ to evidentně platí dle (0) v definici.
- (1) Nechť $n \in \mathbb{N}_0$. Předpokládejme, že $f(n)$ je definováno a $f(n) = (-1)^n$. Pak podle (1) v definici je definováno i $f(n+1)$ a platí $f(n+1) = -f(n) = -(-1)^n = (-1)^{n+1}$.

Důkaz je hotov.

△

! Příklad 5b.b: Definujme funkci $f: \mathbb{N} \mapsto \mathbb{Z}$ takto:

- (0) $f(1) = 3, f(2) = 5$.
- (1) $f(n+1) = 2f(n) - f(n-1)$ pro $n \geq 2$.

Spočítáme si nějaké hodnoty. Pokud použijeme (1) s $n = 2$, dostaneme $f(3) = 2f(2) - f(1) = 2 \cdot 5 - 3 = 7$, pak už můžeme použít (1) s $n = 3$ a dostaneme $f(4) = 9$, podobně $f(5) = 11$ a tak dále. Z těchto hodnot odhadneme, že $f(n) = 2n + 1$ pro $n \in \mathbb{N}$ a aby tvrzení $V(n)$ to dokážeme indukcí. Protože k získání $f(n+1)$ potřebujeme znalost dvou předchozích kroků, musíme použít modifikovaný princip s dvěma výchozími hodnotami.

- (0) Pro $n = 1$ a $n = 2$ je funkce definovaná a vzorce $f(1) = 2 \cdot 1 + 1, f(2) = 2 \cdot 2 + 1$ platí dle (0) v definici.
- (1) Nechť $n \in \mathbb{N}, n \geq 2$. Předpokládejme, že $f(n)$ a $f(n-1)$ jsou definovány a splňují vzorce z $V(n)$, tedy $f(n) = 2n + 1$ a $f(n-1) = 2(n-1) + 1 = 2n - 1$. Pak je podle (1) v definici dobré definováno i $f(n+1)$ a platí (zase dle (1) z definice a podle indukčního předpokladu), že

$$f(n+1) = 2f(n) - f(n-1) = 2[2n+1] - [2n-1] = 2n+3 = 2(n+1)+1,$$

což je přesně vzorec, který jsme potřebovali dokázat.

△

Induktivně lze definovat i množiny. Začneme s přirozenými čísly, které se také musí nějak zavést a Peano je kdysi zadefinoval induktivně takto:

- (0) $1 \in \mathbb{N}$.
- (1) Jestliže $n \in \mathbb{N}$, pak $n+1 \in \mathbb{N}$.

Existuje zajímavá souhra mezi induktivní definicí množiny a možností na ni dále budovat nové objekty indukcí, popřípadě na ni indukcí něco dokazovat. Například pokud přijmeme Peanovu definici \mathbb{N} jako platnou (tedy jako axiom), pak z toho dostaneme platnost principu matematické indukce. V klasické teorii množin se ale přirozená čísla dělají jinak, takže princip indukce je třeba přijmout jako samostatný axiom, jak už jsme o tom psali v předchozí kapitole.

Souhru mezi definicí množiny a indukcí si ukážeme na příkladě.

! Příklad 5b.c: Množinu A všech kladných sudých čísel lze definovat takto:

- (0) $2 \in A$.
- (1) $n \in A \implies n+2 \in A$.

Podrobnejší se tímto způsobem definice budeme zabývat za chvíli, zatím nám bude stačit intuitivní představa, že by to mělo fungovat, opakováním kroku (1) do a postupně dodáme všechna kladná sudá čísla.

Když tedy máme induktivně definovanou množinu A , můžeme na ní definovat nový objekt, třeba funkci, přičemž bude třeba zachovávat stejnou indukční strukturu jako v definici A , tedy základní krok musí definovat hodnotu v 2 a indukční krok musí jít ob dva. Vypadá to třeba takto:

- (0) $f(2) = 1$.
- (1) $f(n+2) = f(n) + 1$.

Dostáváme pak například $f(4) = f(2+2) = f(2) + 1 = 2$, $f(6) = f(4+2) = f(4) + 1 = 3$, $f(8) = f(6+2) = f(6) + 1 = 4$ atd.

Tvrdíme, že tato induktivní/rekurzivní definice (vyberte si, které s těch slov se vám víc líbí) definuje f na množině všech kladných sudých čísel a tato funkce splňuje $f(n) = \frac{n}{2}$ pro všechna $n \in A$.

Důkaz provedeme indukcí, kterou ale zase musíme přizpůsobit tomu, jak byla množina A definována.

- (0) $n = 0$: funguje to, dle definice $f(2) = \frac{2}{2}$.
- (1) Nechť $n \in A$. Předpokládejme, že $f(n)$ je definováno a že $f(n) = \frac{n}{2}$. Pak je podle (1) v definici definováno i $f(n+2)$ a

$$f(n+2) = f(n) + 1 = \frac{n}{2} + 1 = \frac{n+2}{2}.$$

Důkaz je hotov.

Proč by toto mělo stačit? Všechny prvky množiny A (kromě dvojky) se do ní dostaly přičtením dvojky k něčemu, co už v A bylo. Představíme si tedy, jak se množina A postupně rozrůstá. My jsme udělali to, že jsme hned v základním kroku zároveň s přidanou dvojkou pro tu dvojku něco udělali, jmenovitě definovali funkci a pak ukázali určitou vlastnost (funkce je dána pěkným vzorečkem). Zároveň jsme nastavili mechanismus, že se při každém přidání dalšího prvku do množiny A na tento nový prvek rozšíří definice funkce, a ještě jsme rovnou dokázali, že když už máme dotyčnou vlastnost (pěkný vzoreček) pro stávající prvky z A , tak se tato vlastnost přenese i na ten nový, který tam právě přidáváme. Je tedy dobré si to představit jako souběžný jev, množinu A zvětšujeme a zároveň si hlídáme, že nové prvky jsou stejně „dobré“ jako ty předchozí. Selský rozum naznačuje, že by pak v A mělo být uvdobré všechno, platnost takového principu indukce si budeme muset potvrdit řádně matematicky.

Než se do toho dáme, tak poznámejme, že platnost vzorce $f(n) = \frac{n}{2}$ lze dokázat i jinak. Opět půjde o důkaz blízký tomu, jak množinu A vnímáme. My totiž víme, že lze napsat $A = \{2k; k \in \mathbb{N}\}$. Můžeme tedy dokazovat vlastnost $V(k)$: $f(2k) = k$ pro $k \in \mathbb{N}$, což se snadno udělá slabým principem indukce.

△

Teď si přesně zformulujeme induktivní způsob vytváření množin.

! 5b.1. Induktivní definice množin.

Při definici konkrétní množiny M uvažujme následující dva druhy specifikací:

- (0) **Základní pravidla** definují přímo, které prvky jsou v množině M .
- (1) **Induktivní pravidla** určují, jak lze pomocí prvků, které již v množině jsou (tzv. **předpoklady** pravidla), vytvářet další prvky z M (tzv. **závěr** pravidla).

Množina M se pak skládá ze všech prvků, které lze obdržet konečným počtem použití pravidel (0) a (1) (tedy prvky, které lze takto získat, leží v M , a ty, které takto získat nelze, pak v M neleží, čímž je množina M jednoznačně určena).

! Příklad 5b.d: Vymyslíme pravidla, která definují množinu všech neprázdných (a konečných) binárních řetězců, tedy objektů, které vypadají jako třeba 010010110. Myšlenka je jednoduchá, začne se jedním znakem a k němu budeme přilepovat další, například zprava.

- (0) $0 \in M$, $1 \in M$.

- (1) Jestliže je r binární řetězec, pak řetězce $r0$ a $r1$ jsou také binární řetězce.

Zdá se zřejmé, že množina M neprázdných binárních řetězců je právě množina dána induktivními pravidly (0) a (1). Například k řetězci „00101“ dojdeme takto:

$$\xrightarrow{(0)} 0 \xrightarrow{(1)} 00 \xrightarrow{(1)} 001 \xrightarrow{(1)} 0010 \xrightarrow{(1)} 00101.$$

Zde je trochu nepříjemné, že ze zápisu vzniku dotyčného řetězce vlastně nevíme, které ze dvou pravidel v (1) jsme použili. Bývá tedy dobré to rovnou v definici udělat pořádně.

- (0a) $0 \in M$.

- (0b) $1 \in M$.

- (1a) $r \in M \implies r0 \in M$.

- (1b) $r \in M \implies r1 \in M$.

Pak už máme $\xrightarrow{(0a)} 0 \xrightarrow{(1a)} 00 \xrightarrow{(1b)} 001 \xrightarrow{(1a)} 0010 \xrightarrow{(1b)} 00101$.

Aby byla definice úplná, měli bychom teď správně dokázat, že opravdu množina vytvořená pomocí pravidel (0) a (1) je právě množina všech binárních řetězců. To se obvykle dělá ve dvou krocích, jednak se ukáže, že každý objekt vytvořený pravidly dává neprázdný binární řetězec, a naopak se ukáže, že každý neprázdný binární řetězec lze získat pomocí dotyčných pravidel. Obvykle se využívá klasická indukce, ukážeme si to na nějakém komplikovanějším příkladě.

△

Induktivní definice objektů pomocí pravidel se používá ve více oborech computer science a mnohé z nich si zavádějí své vlastní formální zápisy. My se zde spokojíme se zápisem předvedeným výše, ale pro zajímavost si ukážeme několik jiných formalismů.

1) Při práci s datovými typy se definice zapisují pomocí tzv. *Bakchus-Neurovy formy*, jejíž matematická verze by pro naši množinu vypadala takto:

$$R ::= 0 \mid 1 \mid R0 \mid R1$$

2) Další zajímavý zápis je pomocí *odvozovacích pravidel*. Definice by se u našeho příkladu značily takto:

$$\overline{0}^{(0)} \mid \overline{1}^{(1)} \mid \frac{s}{s0}^{(-0)} \mid \frac{s}{s1}^{(-1)}.$$

Odvození řetězce „110“ by se pak napsalo takto:

$$\begin{array}{c} \overline{1}^{(1)} \\ \hline \overline{11}^{(-1)} \\ \hline 110^{(0)}. \end{array}$$

Pokud se používá tento jazyk, tak se základním pravidlům říká axiomy.

Pomocí pravidel se dá vybudovat například vstupní filtr, který přijme jen správně utvořené výrazy určitého typu, často nám je zároveň předpřipraví pro další použití tím, že ukáže vnitřní strukturu vstupních dat, například vytvořením stromového schématu. Ukážeme to na nečem, co všichni dobře známe.

Příklad 5b.e: Množinu M korektních algebraických výrazů skládajících se z čísel a malých písmen anglické abecedy lze definovat například takto:

$$(0a) a \in M, b \in M, \dots, z \in M.$$

$$(0b) \alpha \in \mathbb{R} \implies \alpha \in M.$$

$$(1) \text{ Jestliže } \alpha, \beta \in M, \text{ pak } (\alpha) + (\beta) \in M, (\alpha) - (\beta) \in M, (\alpha) \cdot (\beta) \in M, \frac{\alpha}{\beta} \in M, (\alpha)^{(\beta)} \in M, \sqrt{\alpha} \in M.$$

Podle této definice je třeba $(1) + \left(\frac{(a)+(b)}{5}\right)$ správně utvořený zápis, ale $((1 + \frac{a}{3}) - 5)$ či $3x + -7 + \sqrt{3}$ nejsou z M .

Čtenáře asi napadlo, že naše pravidla vyžadují zbytečně mnoho závorek, například výraz $2 \cdot a + z$ je správný, ale podle naší definice vytvořit nejde, ta umí jen $((2) \cdot (a)) + (z)$. Abychom se zbavili zbytečných závorek, museli bychom použít složitější definice. Jedna možnost je již na vstupu zjišťovat strukturu vstupů, třeba pravidly

$$(1) a + b, c + d \in M \implies (a + b) \cdot (c + d) \in M \text{ (zde jsou závorky potřebné)}$$

a

$$(1) a + b, c + d \in M \implies a + b + c + d \in M \text{ (zde závorky nejsou třeba).}$$

Je zřejmé, že pokud bychom chtěli vytvářet výhradně „pěkné“ výrazy, tak by se takových pravidel musela vytvořit spousta, jsou i další problémy. Například nelze přímo zadefinovat pravidlo $a, b \implies a + b$, protože kdyby bylo b záporné, dostali bychom věci jako $23 + -13$. Vytváření pravidel je občas náročné.

Takovéto „gramatiky“ se používají například v teorii jazyků (matematických, programovacích). Ta má zajímavé předchůdce. Již cca 500 př.n.l. se hindský učenec jménem Pānini rozhodl sepsat gramatiku sanskritu. Použil na to strukturální indukci a vyšla mu z toho báseň o 3959 verších. Byl to první formální popis přirozeného jazyka v historii.

△

! Příklad 5b.f: Vymyslíme pravidla, která by definovala množinu M všech přirozených čísel, my se ale na ně teď nebudeme dívat jako na čísla, místo toho je budeme vnímat jako obrázky, tedy řetězce určitých značek. Abychom nemuseli psát dvacet pravidel, zavedeme si množinu znaků $C = \{0, 1, 2, 3, \dots, 8, 9\}$. Čísla pak budeme vytvářet podobně jako v příkladě s binárními řetězci, tedy přilepováním znaků z C , ale bude tu jedna výjimka: Nebudeme ochotni přijmout všechny možné řetězce, nebývá totiž zvykem začínat čísla nulami. Pravidla tedy musíme upravit.

$$(0) c \in C - \{0\} \implies c \in M.$$

$$(1) r \in M \wedge c \in C \implies rc \in M.$$

Protože přidáváme další cifry zprava a první cifra je nenulová, vznikají tak zásadně řetězce nezačínající nulou. V tom je jedna z velkých výhod induktivních definic, dají se (někdy) relativně snadno upravit, aby se výsledné množině vnitila nějaká struktura.

Zajímavé je porovnání s oním příkladem 5b.d. Tam bylo v zásadě jedno, jestli se nové znaky přilepovaly zprava nebo zleva, my jsme zvolili zprava čistě proto, že jsme tak zvyklí psát rukou, ale šlo by klidně do (1) dát $0r \in M$ a $1r \in M$. Naopak v tomto příkladě bylo přidávání zprava rozhodující výhodou, protože jsme hned v základním kroku zajistili, že nebudeme mít na levém konci nuly. Pokud bychom se rozhodli přidávat nové znaky zleva, tak bychom museli zajistit, že po každém přidání nul se následně objeví něco nenulového, což se ale induktivními pravidly dle našeho vzoru nedá rozumně zajistit. Jak by to vypadalo?

V kroku (1) by určitě bylo pravidlo $r \in M \ \& \ c \in C - \{0\} \implies cr \in M$. Pak bychom mohli zkousit pravidlo $r \in M \ \& \ c \in C - \{0\} \implies c0r \in M$, jenž tak bychom neuměli vyrobit dvě po sobě jdoucí nuly. Bylo by proto třeba přidat i pravidlo s $c00r \in M$, ale co tři nuly? Protože je možné vyrobit čísla s libovolným počtem po sobě jdoucích nul uprostřed, potřebovali bychom nekonečně mnoho takových indukčních pravidel, což je zjevně slepá ulička.

△

Příklad 5b.g: Předchozí příklady nás přivádí k zajímavé oblasti zvané teorie jazyků. Ta pracuje obecně s nějakou **abecedou** Σ , což je množina používaných znaků, třeba $\Sigma = \{0, 1\}$ jako v prvním příkladě, C v druhém nebo třeba 26 písmen anglické abecedy. Z těchto znaků pak vytváříme **slova** nad tuto abecedou, což jsou řetězce znaků z abecedy Σ . V teorii jazyků se uvažuje i slovo prázdné, označované λ .

Množina všech slov nad danou abecedou Σ se značí Σ^* a definuje takto:

- (0) $\lambda \in \Sigma^*$.
- (1) Jestliže $s \in \Sigma^*$ a $x \in \Sigma$, pak $sx \in \Sigma^*$.

V teorii jazyků je základní operací tzv. konkatenace neboli spojování, což už jsme vlastně použili v definici, kdy se dva řetězce spojí za sebe. Například konkatenací slov „auto“ a „drom“ vznikne slovo „autodrom“.

Většinou nepotřebujeme množinu Σ^* všech slov, ale zajímají nás jen slova některá, takže se zavádějí komplikovanější pravidla a gramatiky a teorie jazyků pak začne být velice zajímavá.

Již jsme se zmíňovali, že na množině definované indukcí lze definovat další struktury pomocí indukce stejné formy. Ukážeme si dva zajímavé objekty na množině všech slov.

1) V teorii jazyků se definuje délka slova, a to následovně:

- (0) $d(\lambda) = 0$.
- (1) Jestliže $s \in \Sigma^*$ a $x \in \Sigma$, pak $d(sx) = d(s) + 1$.

Tím je délka slova definovaná pro všechna slova a snadno si rozmyslíme, že funguje přesně tak, jak bychom čekali. Vlastně jsme tak definovali funkci na Σ^* .

2) Teď zadefinujeme operaci na Σ^* . Operace na množině je procedura, která vezme jeden či více objektů z dané množiny a vrátí nějaký objekt z téže množiny. Nás bude zajímat operace, která dané slovo převrátí, takže bude pozpátku. Pro dané slovo s se výsledek takové operace značí s^R , říkáme tomu *obrácené slovo* a definujeme to takto:

- (0) $\lambda^R = \lambda$.
- (1) Jestliže $s \in \Sigma^*$ a $x \in \Sigma$, pak $(sx)^R = xs^R$.

Zkusme si obě definice na nějakém příkladě. Vezměme $\Sigma = \{a, b, c, t\}$ a slovo $s = bat \in \Sigma^*$. Jakou má délku?

Neplatí $s \in \Sigma$, proto musíme použít (1) a napsat si s jako spojení $s_1 = ba$ a $x = t$. Podle definice $d(bat) = d(ba) + 1$. Voláme rekurzivně (1) a napíšeme si s_1 jako spojení $s_2 = b$ a $x = a$, tedy $d(ba) = d(b) + 1$, nakonec si vyjádříme s_2 coby spojení $s_3 = \lambda$ a b , konečně se dostáváme k něčemu, co umíme. Podle (0) je $d(\lambda) = 0$, zpětným chodem pak $d(b) = 0 + 1 = 1$, $d(ba) = 1 + 1 = 2$ a $d(bat) = 2 + 1 = 3$.

Jak vypadá s^R ? Struktura rekurze je obdobná, nejprve podle (1) najdeme $(bat)^R$ jako $t(ba)^R$, pak zase podle (1) je $(ba)^R = ab^R$, nakonec se dostaneme k $b^R = (\lambda b)^R = b\lambda^R = b\lambda = b$. Zpětným chodem pak $(ba)^R = ab$ a $(bat)^R = t(ab) = tab$. Zdá se, že to funguje.

△

Na induktivně definovaných množinách často potřebujeme něco dokázat, například že objekty, které jsme na nich definovaly, splňují nějaké podmínky. Nikterak překvapivě se to dělá indukcí, ale upravenou tak, aby odpovídala definici dotyčné množiny.

!

5b.2. Princip strukturální indukce (structural induction).

Uvažujme množinu M definovanou induktivně pomocí nějakých základních pravidel (0) a induktivních pravidel (1). Uvažujme vlastnost $V(m)$, která má smysl pro všechna $m \in M$.

Předpokládejme, že jsou splněny následující podmínky:

(0) V je splněna pro všechny prvky, které jsou do M dodány základními pravidly.

(1) Pro každé induktivní pravidlo platí: Jestliže je V splněna pro prvky z jeho předpokladů, pak je splněna i pro prvek z jeho závěru.

Pak je vlastnost V splněna pro všechny prvky $m \in M$.

To je náš poslední indukční princip a ve skutečnosti zase nejde o nic nového, Princip strukturální indukce je ekvivalentní principům z kapitolky 5a. Ještě ale nejsme připraveni to dokázat, necháme si to na konec této kapitoly. Zatím se podíváme na různé příklady a také si představíme nějaké nové myšlenky.

! **Příklad 5b.h:** V tomto příkladě budeme definovat množinu všech řetězců ze symbolů $C = \{1, 2, 3\}$, které neobsahují číslo 11. Hlavní myšlenka je, že postupně přidáváme číslice zprava, ale beztrestně můžeme přidávat jen 2 a 3, protože pak nehrozí nebezpečí vzniku 11. S jedničkou musíme být opatrnejší, tu můžeme přidat jen za 2 nebo 3, takže je rovnou přidáme jako dvojice. Tím je jasné, jak bude vypadat (1). Máme ale problém, řetězec „1“ vyhovuje definici, ale nedostaneme jej přidáváním 21 ani 31. Tak tento řetězec zahrneme jako speciální případ do základního pravidla. Dostáváme tedy následující definici množiny M :

(0a) $\lambda \in M$.

(0b) $1 \in M$.

(1a) $w \in M \implies w2 \in M$.

(1b) $w \in M \implies w3 \in M$.

(1c) $w \in M \implies w21 \in M$.

(1d) $w \in M \implies w31 \in M$.

Je tato definice správná? Přesněji řečeno, pokud budeme uvažovat množinu M definovanou induktivně našimi pravidly, bude rovna množině ze zadání? To se obvykle dokazuje dvěma kroky.

1) Dokážeme, že žádný prvek w ze vzniklé množiny M nemůže obsahovat „11“, označíme tuto vlastnost $W(n)$ a podle očekávání použijme strukturální indukci. Při ní musíme sledovat stejnou strukturu jako při definici množiny M , viz Princip strukturální indukce.

(0) Žádný z prvků λ či 1 ze základních pravidel neobsahuje 11, proto je splněno $W(\lambda)$ a $W(1)$.

(1) V tomto kroku potřebujeme ověřit u všech induktivních pravidel, že se platnost W pro prvky z předpokladu přenáší i na prvek ze závěru.

Nejprve uvažujme induktivní pravidlo (1a). Předpokládejme, že řetězec w z jeho předpokladu splňuje $W(w)$, takže neobsahuje 11. Přidáním znaku 2 na konec se to nezmění, tudíž i $W(w2)$ je splněno.

Podobně ukážeme pravdivost implikací $W(w) \implies W(w3)$, $W(w) \implies W(w21)$ a $W(w) \implies W(w31)$ pro zbyvající pravidla v (1).

Podle principu strukturální indukce tedy W platí pro všechny prvky množiny M .

2) Teď potřebujeme naopak ukázat, že každý řetězec nad C neobsahující 11 je i v množině M . Uděláme to indukcí na délku řetězce. Dokážeme tedy pro $n \in \mathbb{N}_0$ silnou indukcí tvrzení $V(n)$, že řetězce délky n nad C neobsahující 11 lze vytvořit pomocí (0) a (1). Budeme muset udělat speciální krok pro délku $n = 1$, protože jeden z těchto řetězců, jmenovitě 1, nelze vytvořit z řetězce kratší délky, címž se vymyká indukci.

(0) Evidentně dokážeme získat řetězce délky 0 a 1, protože řetězec λ je v kroku (0a), řetězec „1“ je v kroku (0b) a ostatní o délce 1, tedy „2“ a „3“ dostaneme kombinací (0a) a (1a) popř. (1b).

(1) Mějme teď $n \in \mathbb{N}$, $n \geq 2$ a předpokládejme, že umíme pomocí pravidel (0) a (1) vyrobit libovolný řetězec délky 0 až n . Uvažujme libovolný řetězec r délky $n + 1$.

Jestliže je jeho poslední znak 3, pak $r = r'3$ pro nějaký řetězec r' délky n , který už podle indukčního předpokladu umíme získat pomocí našich axiomů, z něj pak řetězec r získáme axiomem (1b). Takže r končící na 3 lze vyrobit pomocí pravidel (0) a (1).

Podobně se vyrovnané s případem, že r končí na 2.

Jestliže r končí na 1, pak předchozí znak (který existuje, $n \geq 2$) musí být buď 2 nebo 3, protože r neobsahuje 11. Předpokládejme první případ, pak $r = r'21$ pro nějaký řetězec r' délky $n - 1$. Odvoláme na indukční předpoklad (tady jsme potřebovali indukci silnou) a pravidlo (1c) a máme r odvozeno, případ $r'31$ se dělá obdobně.

Vyčerpali jsme všechny možnosti, libovolný prvek délky $n + 1$ lze odvodit, tedy $V(n + 1)$ platí. Důkaz je hotov.

Někdy je nám povoleno do pravidel v definici vkládat podmínky. Není to ale tak jednoduché, protože pokud si vezmeme slovo w , tak nemáme nástroj, jak se zeptat na jeho poslední znak. Dá se to elegantně vyřešit tak, že bereme v úvahu jen řetězce jistého typu, například $1w3$ je řetězec, který začíná jedničkou a končí trojkou, pod

jménem w pak máme k dispozici jeho prostřední část. Díky tomu je pak možné vyjádřit induktivní podmínky (1) následovně:

- (1a) $w \in M \implies w2 \in M$.
- (1b) $w \in M \implies w3 \in M$.
- (1c) $w2 \in M \implies w21 \in M$.
- (1d) $w3 \in M \implies w31 \in M$.

V některých případech je to výrazně výhodnější, ale ne vždy je nám toto povoleno aplikací, kde induktivní definici používáme.

△

Příklad 5b.i: Mějme abecedu $C = \{0, 1, 2, 3\}$. Zajímá nás množina všech řetězců ze znaků z C , které neobsahují víc nul než jedniček. Abychom ji vybudovali, zkusíme vyjít z klasického postupu, tedy budujeme řetězce postupně, třeba přidáváním zprava. Ze zadání je jasné, že znaky 1, 2 a 3 lze přilepovat bez omezení, ale s přilepováním 0 je třeba být opatrný, musíme zajistit, že za každé přidání se přidá i jednička. Nelze to ovšem udělat přidáním dvojznaku 01, protože odpovídající jednička může být v konečném slově klidně někde jinde, za i před nulou. Nabízí se možnost, že budeme přilepovat 0 na jeden konec a 1 na druhý konec dosavadního řetězce, tedy dvě různá pravidla.

Tím ovšem narázíme na další problém, tímto způsobem nepůjde vyrobit třeba 3120, protože po přilepení 1 zleva a 0 zprava ke dvojce už neumíme přilepovat zleva. Takže se to budeme muset naučit a přidat i přilepování znaků 1, 2, 3 zleva.

Jak bude vypadat základní krok? Začneme řetězci 1, 2 a 3 (řetězec 0 nevyhovuje pravidlům). Docházíme tak k následujícím definicím.

- (0a) $1 \in M$.
- (0b) $2 \in M$.
- (0c) $3 \in M$.
- (1a) $r \in M \implies r1 \in M$.
- (1b) $r \in M \implies r2 \in M$.
- (1c) $r \in M \implies r3 \in M$.
- (1d) $r \in M \implies 1r0 \in M$.
- (1e) $r \in M \implies 1r \in M$.
- (1f) $r \in M \implies 2r \in M$.
- (1g) $r \in M \implies 3r \in M$.
- (1h) $r \in M \implies 0r1 \in M$.

Je to už dobrá definice? Čtenář teď má příležitost si trochu pohrát a přemýšlet, zda všechny řetězce daného typu, které jej napadnou, dokáže vytvořit pomocí pravidel výše.

Já jsem si docela dlouho myslel, že ano, ale pak mě napadl řetězec 0110, který evidentně pomocí daných pravidel neuklohníme (zádné z nich nemá na krajích nuly).

Jediné rozumné východisko je nějak zařídit, abychom mohli přidávat nulu a jedničku někam doprostřed řetězce. To zní jako dobrý nápad, ale má problém, že když induktivní pravidlo začneme „nechť $w \in M$ “, tak to je jeden objekt a my neumíme sáhnout do jeho středu. Budeme tedy rovnou muset začít s jednotlivými úseky, mezi které chceme vsazovat, tedy třemi řetězci. Jenže co když někdy budeme chtít pracovat jen se dvěma segmenty (či jedním)? To hravě vyřešíme zahrnutím prázdného řetězce do naší množiny, čímž si také elegantně zkrátíme základní krok.

- (0) $\lambda \in M$.
- (1a) $r \in M \implies r1 \in M$.
- (1b) $r \in M \implies r2 \in M$.
- (1c) $r \in M \implies r3 \in M$.
- (1d) $r, s, t \in M \implies r1s0t \in M$.
- (1e) $r, s, t \in M \implies r0s1t \in M$.

Máme novou definici a starou otázkou: Dá nám všechny řetězce zkoumaného typu? Tentokrát se to opravdu povedlo a ukážeme si alespoň částečně, jak by se to dokazovalo.

1) Vzhledem k tomu, že vždy přidáváme alespoň tolik jedniček co nul, zdá se jasné, že vytvořené řetězce splňují zadání. Formální důkaz se dělá strukturální indukcí a pomůže při něm, když si zavedeme dvě pomocné funkce: $f_0(m)$ udává, kolik je v řetězci m nul, a $f_1(m)$ udává, kolik je v řetězci m jedniček. Uvažujme vlastnost $W(m)$, která říká, že $f_0(m) \leq f_1(m)$. Chceme ukázat, že W platí pro všechny prvky $m \in M$.

- (0) Evidentně $f_0(\lambda) = 0 = f_1(\lambda)$, tedy W platí pro všechny prvky ze základních pravidel.

(1) Teď si vezměme nějaké induktivní pravidlo z definice M a předpokládejme, že W platí pro všechny prvky z jeho předpokladu. Rozebereme si případy:

a) Pokud je to pravidlo (1a), tak předpokládáme platnost $W(r)$, tedy že $f_0(r) \leq f_1(r)$. Protože $f_0(r1) = f_0(r)$ a $f_1(r1) = f_1(r) + 1$, dostáváme $f_0(r1) = f_0(r) \leq f_1(r) < f_1(r) + 1 = f_1(r1)$. Takže W platí i pro závěr $r1$ pravidla (1a).

b) Pokud je to pravidlo (1b), tak předpokládáme, že $f_0(r) \leq f_1(r)$. Protože $f_0(r2) = f_0(r)$ a $f_1(r2) = f_1(r)$, dostáváme $f_0(r2) = f_0(r) \leq f_1(r) = f_1(r2)$. Takže W platí i pro závěr pravidla (1b). Podobně se to dokáže případ (1c).

c) Pokud je to pravidlo (1d), tak předpokládáme platnost W pro r, s, t , tedy že $f_0(r) \leq f_1(r)$, $f_0(s) \leq f_1(s)$ a $f_0(t) \leq f_1(t)$. Dostáváme pak $f_0(r1s0t) = f_0(r) + f_0(s) + f_0(t) + 1 \leq f_1(r) + f_1(s) + f_1(t) + 1 = f_1(r1s0t)$. Takže W platí i pro závěr pravidla (1d). Podobně se to dokáže pro (1e).

Podle principu strukturální indukce platí W pro všechna $m \in M$.

2) Teď bychom měli ukázat opačnou inkluzi, tedy že každý řetězec nad C obsahující alespoň tolik jedniček, kolik má nul, leží v naší množině M vytvořené pravidly (0) a (1). Jinými slovy, pro daný objekt musíme vytvořit způsob, kterým jej lze odvodit pomocí pravidel (0) a (1), přičemž ale nevíme, který konkrétní objekt máme, címž se situace evidentně dosti dramatizuje. Takže tento směr bývá obvykle dosti náročný, zde to proto nedokážeme, spíš zkusíme naznačit, kde je problém.

Obvykle se podobná tvrzení dokazují indukcí, zde bychom použili silnou indukcí na počet nul ve výsledném řetězci. Chtěli bychom dokazovat pro $n \in \mathbb{N}_0$ tvrzení $V(n)$, že každý řetězec nad C , který obsahuje n nul a alespoň n jedniček, lze dostať pomocí pravidel (0) a (1).

(0) $n = 0$. Mějme řetězec w , který neobsahuje nuly. Je tedy složen čistě ze znaků 1,2,3. Protože nám pravidla (1) dovolují právě tyto znaky zcela libovolně spojovat za sebe, bude možné vytvořit i řetězec w . To je třeba dokázat a není to příliš obtížné, snadno se to udolá indukcí na délku w .

(1) Mějme $n \in \mathbb{N}_0$. Předpokládejme, že umíme pomocí pravidel sestavit všechny řetězce nad C , které obsahují 0 až n nul a alespoň stejně jedniček. Musíme ukázat, že umíme sestavit i řetězce s $n+1$ nulami a alespoň $n+1$ jedničkami. Vezměme si tedy nějaký takový řetězec w .

Základní myšlenka je jasná, chceme využít pravidlo (1) a přejít k podřetězcům, které už budou mít méně nul, pro ně pak využít indukční předpoklad. Takže si v našem řetězci najdeme nějakou nulu a nějakou jedničku, třeba napravo od té nuly. Naše slovo pak lze napsat jako $w = r0s1t$. Pokud bychom byli schopni aplikovat indukční předpoklad na části r, s, t , pak bychom věděli, že je lze pomocí pravidel (0) a (1) vytvořit, další pravidlo (1) už nám dává zkoumané slovo w .

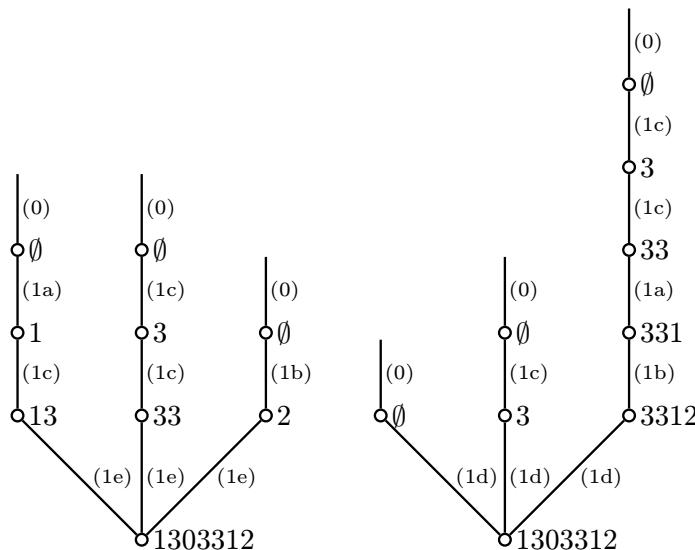
Máme ale velký problém. Protože w mělo $n+1$ nul a my jsme při vytváření r, s, t jednu nulu z w odebrali, je jasné, že všechny tyto tři části obsahují nejvýše n nul. Není ale žádný důvod, proč by každá z těchto částí měla mít alespoň tolik jedniček, kolik má nul. Jedničky totiž mohly být ve w rozloženy v zásadě libovolně, takže se klidně mohlo stát, že po rozdělení w na části se všechny dostali do jedné z nich, zatímco jiná část má nuly, ale jedničky už na ni nezbyly, na takovou část pak ale nelze aplikovat indukční předpoklad.

Klíčem k úspěchu je tedy vyřešit tento problém dělení w na části. Je třeba ukázat, že lze w rozdělit tak, aby žádná výsledná část neměla méně jedniček než nul. To je kombinatorický problém, který není triviální, takže jím nebudeme tuto kapitolu natahovat.

△

! Když máme množinu M definovanou indukcí, tak se každý prvek z M vytvořil nějakým konečným počtem použití pravidel z (0) a (1), říká se tomu derivační postup a lze jej pěkně znázornit **derivačním stromem** či **odvozovacím stromem** (anglicky **parsing tree**). Prvek je kořenem dole, podle indukčních pravidel (značí se zkratkou vedle hran) se postupně dojde k listům, tj. prvkům ze základních pravidel (0). Počet úrovní stromu (přesněji řečeno největší počet kroků, který je ve stromu možné jedním směrem udělat) je **výška** stromu, takže prvky ze základních pravidel mají výšku 1 (nejprve je nic, jedním krokem dle pravidla (0) se pak dojde k dotyčnému prvku). Má to ale malý zádrhel, může se totiž snadno stát, že se ke zkoumanému prvku dokážeme dostat pomocí pravidel z (0) a (1) více způsoby.

Pro příklad se vrátíme k předchozí definici řetězců s alespoň tolik jedničkami co nulami, najdeme si dva odvozovací stromy pro řetězec 1303312:



Stát se to tedy snadno může. Protože jsou ale výšky všech stromů alespoň jedna, dá se hledat nejnižší možný z nich. Definujeme **výšku prvku** jako nejmenší možnou výšku odvozovacího stromu pro dotyčný prvek. Tím se ale zase poněkud zkomplicuje praktické určování výšky prvku. Nahoře máme dva odvozovací stromy pro 1303312, menší z nich má výšku 4, ale to ještě neznamená, že prvek 1303312 má výšku 4. Museli bychom dokázat, že neexistuje žádný nižší strom pro 1303312 (což je mimochodem pravda).

Poslední příklad ukázal, že strukturální definování množin nemusí být zase tak snadné. Někdy dokonce stačí jen malá modifikace zadání a máme problém, pro ukázku se vrátíme k námětu příkladu 5b.h.

Příklad 5b.j: V tomto příkladě se pokusíme definovat množinu všech řetězců ze symbolů $C = \{1, 2, 3\}$, které neobsahují číslo 13. To vypadá obdobně jako příklad 5b.h, proto použijeme stejnou myšlenku a budeme postupně přidávát číslice ze zadu (zprava) s tím, že beztrestně můžeme přidávat jen 1 a 2, protože pak nehrází nebezpečí vzniku 13. S trojkou musíme být opatrnejší, tu můžeme přidat jen za 2 nebo 3. Pro začátek si ukážeme, jak může vypadat definice, pokud je nám povoleno použít podmínky.

- (0a) $\lambda \in M$.
- (0b) $3 \in M$.
- (1a) $w \in M \implies w1 \in M$.
- (1b) $w \in M \implies w2 \in M$.
- (1c) $w2 \in M \implies w23 \in M$.
- (1d) $w3 \in M \implies w33 \in M$.

Dá se dokázat, že tato definice dělá to, co po ní chceme.

Ne vždy je ovšem toto možné, takže mnohem zajímavější je vytvářet definice bez podmínek. Problémem je přidávání trojky. V příkladě 5b.h jsme toto řešili tak, že jsme přidávali dvojznaky, což by naznačovalo následující definici.

- (0a) $\lambda \in M$.
- (0b) $3 \in M$.
- (1a) $w \in M \implies w1 \in M$.
- (1b) $w \in M \implies w2 \in M$.
- (1c) $w \in M \implies w23 \in M$.

Proč jsme nezahrnuli pravidlo $w \in M \implies w33 \in M$? Protože pokud by řetězec W končil jedničkou (což může), tak by vzniklo 133 a máme třináctku. Tím ovšem vznikl zásadní problém, protože řetězec 233 je dozajista korektní, ale našimi pravidly jej vytvořit nejde. Proč tedy nepřidat pravidlo?

- (1d) $w \in M \implies w233 \in M$.

To je sice pěkné, ale zase neumíme vytvořit 2333. Přidáním

- (1e) $w \in M \implies w2332 \in M$

zase nevyřešíme problém řetězce 2333 atd., potřebovali bychom nekonečně mnoho pravidel.

Tudy tedy cestička nevede. Pomohlo by, pokud bychom povolili přidávání i zleva?

- (0a) $\lambda \in M$.
- (0b) $3 \in M$.
- (1a) $w \in M \implies w1 \in M$.
- (1b) $w \in M \implies w2 \in M$.
- (1c) $w \in M \implies w23 \in M$.

- (1d) $w \in M \implies 2w \in M.$
- (1e) $w \in M \implies 3w \in M.$
- (1f) $w \in M \implies 12w \in M.$

Ted' už libovolný řetězec typu 23333 vytvoříme, nejprve opakováním (1e) dodáme ty trojky a pak zakončíme pomocí (1f). Ale pořád neumíme řetězec 11233.

Přiznám se, že se mi nepodařilo vymyslet rozumně krátký soubor axiomů bez podmínek pro řetězce bez 13. Takže strukturální definice je nástroj mocný, ale občas překvapivě obtížný na použití.

△

Příklad 5b.k: Dokonce ani u některých „pěkných“ množin není jasné, jak je efektivně definovat indukcí. Ty pickým příkladem jsou celá čísla. Když se necháme inspirovat Peanovým pohledem na čísla přirozená, dostaneme toto:

- (0) $0 \in \mathbb{Z}.$
- (1a) $n \in \mathbb{Z} \implies n + 1 \in \mathbb{Z}.$
- (1b) $n \in \mathbb{Z} \implies n - 1 \in \mathbb{Z}.$

To samozřejmě funguje, ale má to jednu nevýhodu, každé číslo je definované mnohemkrát, dokonce nekonečně mnohemkrát. Například abychom odvodili číslo 13 z nuly, tak nejprve zopakujeme pravidlo (1a) třeba milionkrát a pak dáme pravidlo (1b) milion minus 13 krát. Je to tedy velice plýtvavá definice. Mimochodem, jde docela zajímavě smrsknout:

- (0a) $0 \in \mathbb{Z}.$
- (0b) $1 \in \mathbb{Z}.$
- (1) $m, n \in \mathbb{Z} \implies m - n \in \mathbb{Z}.$

Pořád ale strašlivě plýtváme. Jde to lépe? Ano, pokud si dovolíme logiku.

- (0) $0 \in \mathbb{Z}.$
- (1a) $n \in \mathbb{Z} \wedge n \geq 0 \implies n + 1 \in \mathbb{Z}.$
- (1a) $n \in \mathbb{Z} \wedge n > 0 \implies -n \in \mathbb{Z}.$

Ted' už vůbec neplýtváme, ale takovéto podmínky se zase obtížně vyjadřují některými z používaných formalismů, takže bychom se jim měli spíš vyhýbat. Někdy to prostě ideálně nejde.

△

Ukážeme si ještě dva příklady, kde budeme mít několik příležitostí si indukci vyzkoušet.

Příklad 5b.l (delší pro pokročilé, ale poučný): Dokážeme, že každé symetrické číslo se sudým počtem číslic je dělitelné 11.

Možných přístupů je více, tady to zkusíme přes strukturální indukci. Nejprve taková čísla nadefinujeme. Aby se nám to lépe dělalo, označme si $C = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Zajímají nás čísla jako 340043, 10377301 či 1331. Taková čísla budeme jistě vytvářet souběžným přidáváním zleva i zprava.

Aby se nám pravidla lépe pilovala, budeme se zatím dívat na čísla jako na řetězce cifer. Po troše přemýšlení čtenáře jistě napadlo, že přidávání je třeba dělat opatrně, problémem je nula. Když třeba k 33 přidáme nulu na obě strany, dostaneme 330, což už není symetrické číslo. Čísla s nulami ale jsou, problém vyřešíme tak, že když přidáváme nulu, tak povinně přidáme i něco jiného.

Čím začneme? Nemůžeme začít prázdným řetězcem (není to číslo), takže v základním kroku budeme muset nastrkat do M všechny možné dvojíčky cifer z C . To ale nestačí, jak uděláme číslo 2002? Nelze dát jako základní číslo 00 a pak případně přidat 2 na oba konce, protože to 00 je vlastně 0. Musíme tedy v základním kroku dodat i čísla typu 2002.

- (0a) $c \in C \implies cc \in M.$
- (0b) $c \in C \implies c0c \in M.$
- (1a) $r \in M \wedge c \in C \implies crc \in M.$
- (1b) $r \in M \wedge c \in C \implies c0r0c \in M.$

Tato definice je správná a zároveň efektivní, když dostaneme symetrické číslo se sudým počtem číslic, tak jej lze coby řetězec vyrobit jediným způsobem pomocí našich pravidel.

Abychom s tím teď mohli pracovat matematicky (chce se po nás dělitelnost), musíme ty operace s řetězci (spojovalní neboli konkatenace) přepsat do řeči čísel, což znamená, že si budeme hrát se zápisem čísla v desítkové soustavě. Například chceme-li k číslu r „přilepit“ na konec „38“, tak si nejprve musíme na konci udělat na ty znaky místo a „r00“ dostaneme v řeči čísel jako $100 \cdot r$. Znaky pak dolepíme přičtením: $100 \cdot r + 3 \cdot 10 + 8$. Podobné triky teď použijeme k přepisu pravidel výše, čímž dostaneme opravdu množinu čísel, ne řetězců. Označíme ji proto jinak, třeba S , a prvky z C a 0 teď bereme jako čísla.

- (0a) $c \in C \implies 10c + c \in S.$

(0b) $c \in C \implies 1000c + c \in S.$

(1a) Jestliže $s = \sum_{i=0}^m s_i 10^i \in S$, kde $s_m \neq 0$, a $c \in C$, pak $c \cdot 10^{m+2} + \sum_{i=0}^m s_i 10^{i+1} + c = 10^{m+2}c + 10s + c \in S.$

(1b) Jestliže $s = \sum_{i=0}^m s_i 10^i \in S$, kde $s_m \neq 0$, a $c \in C$, pak $10^{m+4}c + 100s + c \in S.$

Nejprve si dokážeme že opravdu výsledná čísla mají sudý počet cifer. Použijeme strukturální indukci.

(0) To je jasné, pro $c \in C$ jsou $10c + c$ dvouciferná a $1000c + c$ čtyřciferná.

(1a) Má-li s sudý počet cifer, pak $s = \sum_{i=0}^m s_i 10^i$, $s_m \neq 0$ a m je liché (rozmyslete si to). Pak je ale i $m+2$ liché a proto má $10^{m+2}c + 10s + c$ sudý počet cifer.

Důkaz pro (1b) je obdobný.

Víme tedy, že když si vezmeme nějaké $s = \sum_{i=0}^m s_i 10^i \in S$, kde $s_m \neq 0$, tak je m liché.

Teď už strukturální indukcí dokážeme to hlavní: Pro každé $s \in S$ platí, že je dělitelné 11.

(0) Nejprve to dokážeme pro prvky ze základních kroků.

(0a) Nechť $c \in C$. Pak $10c + c = 11c$, tedy číslo dělitelné 11.

(0b) Nechť $c \in C$. Pak $1000c + c = 1001c = 11 \cdot 91c$, tedy číslo dělitelné 11.

(1) Teď probereme induktivní pravidla.

(1a) Nechť $s \in S$ a předpokládejme, že je dělitelné 11, tedy $s = 11k$. Víme, že když si zapíšeme $s = \sum_{i=0}^m s_i 10^i \in S$, kde $s_m \neq 0$, tak je m liché, $m = 2n-1$. Pak pro $c \in C$ je $10^{m+2}c + 10s + c = 10 \cdot (11k) + (10^{2n+1} + 1)c$. V příkladě 5a.b jsme dokázali, že čísla typu $10^{2n+1} + 1$ jsou také dělitelná 11, tedy $10^{m+2}c + 10s + c = 11 \cdot (10k) + 11l \cdot c = 11a$, i nové číslo je dělitelné 11.

(1b) Nechť $s \in S$ a předpokládejme, že $s = 11k$ pro $k \in \mathbb{Z}$. Podobně jako v (1a) zapíšeme $s = \sum_{i=0}^m s_i 10^i \in S$, kde $m = 2n-1$, a pro $c \in C$ máme $10^{m+4}c + 100s + c = 11 \cdot (10k) + (10^{2(n+1)+1} + 1)c = 11b$, zase je i nové číslo dělitelné 11.

Důkaz je hotov.

tento důkaz nebyl nejfektivnější, ale pěkně ukázal různé aspekty práce s řetězci i číslami a strukturální indukci v akci.

△

Příklad 5b.m: Uvažujme šachovnici „nekonečnou směrem nahoru a doprava“, jejíž políčka jsme si zakódovali pomocí dvojic (i, j) pro $i, j \in \mathbb{N}$, kde i ukazuje vodorovně doprava (tedy udává číslo sloupce) a j ukazuje nahoru. Na levé dolní rohové políčko, tedy na souřadnici $(1, 1)$, položíme šachového koně. To je figurka s nejjednodušším pohybem, jsou jí totiž povoleny jen tahy ve tvaru L. Přesně řečeno, tah koně se skládá z poskoku o dvě pole v nějakém vodorovném či svislém směru, následovaného poskokem o jedno pole do pravého úhlu.

Tvrdíme, že se kůň pomocí těchto tahů dostane na zcela libovolné pole na naší šachovnici.

Evidentně je to úloha, kterou bychom rádi řešili matematickou indukcí. Problémem ale je, že šachovnice je dvourozměrná. S tím se dá vyrovnat několika způsoby, ukážeme si čtyři.

1) Jedna možnost je si situaci zjednodušit tím, že si pohyb v rámci šachovnice rozložíme do dvou směrů, takže když chceme někam dojít, tak nejdřív jdeme pořád napravo, dokud nebudeme ve správném sloupci, a pak půjdeme nahoru na cílové pole. Pohyb napravo už má jen jeden parametr, stejně tak jako pohyb nahoru, takže tam by měla indukce pomoci.

Zkusíme nejprve indukci slabou, která se odvolává jen na předchozí situaci. Abychom uspěli, musíme vymyslet, jestli se dá někam dostat za předpokladu, že už jsme o políčko vedle (to je ten rekurzivní způsob myšlení). Umíte se nějak koněm dostat o jedno pole doprava, popřípadě nahoru? Rozmyslete si to, já už to umím, a tak vím, že indukce projde.

1a) Nejprve dokážeme, že se z počátku $(1, 1)$ dokážeme dostat libovolně daleko doprava. Formálně:

Pro $i \in \mathbb{N}$ dokazujeme $V(i)$: kůň se umí dostat z pole $(1, 1)$ na pole $(i, 1)$.

Uděláme to matematickou indukcí.

(0) $i = 1$: Na poli $(1, 1)$ už kůň je, takže se tam určitě umí dostat.

(1) Nechť $i \in \mathbb{N}$ a předpokládejme, že se kůň umí dostat z $(1, 1)$ na pole $(i, 1)$.

Chceme ukázat, že se dostane i na pole $(i + 1, 1)$. To se provede následujícími třemi tahy:

tah 1: $(i, 1) \mapsto (i + 1, 3)$,

tah 2: $(i + 1, 3) \mapsto (i + 3, 2)$,

tah 3: $(i + 3, 2) \mapsto (i + 1, 1)$. Takže $V(1)$ platí.

Důkaz $V(i)$ je hotov.

1b) Teď bychom se potřebovali zase posunout nahoru a čtenář by měl mít pocit, že je to vlastně stejný problém, cílem důkazu by měl být stejný. Je tomu tak, až na jednu maličkost, teď už totiž nestačí dokazovat cestu vzhůru jen pro první sloupec, ale důkaz musí cestování nahoru potvrdit pro sloupec libovolný, z místa $(m, 1)$, kam jsme došli v první fázi.

Nechť $m \in \mathbb{N}$ je pevně zvolené číslo (parametr). Pro $j \in \mathbb{N}$ dokazujeme $W_m(j)$: kůň se umí dostat z pole $(m, 1)$ na pole (m, j) .

Uděláme to matematickou indukcí.

(0) $j = 1$: Na poli $(m, 1)$ už kůň je, takže se tam umí dostat.

(1) Nechť $j \in \mathbb{N}$ a předpokládejme, že se kůň umí dostat z $(m, 1)$ na pole (m, j) . Chceme ukázat, že se dostane i na pole $(m, j + 1)$. To se provede následujícími třemi tahy:

tah 1: $(m, j) \mapsto (m + 2, j + 1)$,

tah 2: $(m + 2, j + 1) \mapsto (m + 1, j + 3)$,

tah 3: $(m + 1, j + 3) \mapsto (m, j + 1)$. Takže $V(1)$ platí.

Důkaz $W_m(j)$ je hotov.

Teď to dáme dohromady. Má-li se kůň dostat na pole (m, n) , pak se pomocí vlastnosti $V(m)$ nejprve dostane na pole $(m, 1)$ a pak pomocí vlastnosti $W_m(n)$ na pole (m, n) . Hotovo.

Všimněte si jedné podstatné věci. Aby byl důkaz úplný, tak je třeba dát pozor na to, že na ony tři tahy máme vždy na šachovnici místo. To je pravda, protože k přesunu využíváme směry doprava a nahoru, kde je šachovnice neomezená. Ještě se k tomu vrátíme v poznámce na konci příkladu.

2) Někdy je možné se na vícerozměrnou situaci podívat z jiného úhlu pohledu tak, že už se dá popsat jedním parametrem. V tomto případě k tomu dojde, pokud se díváme na pohyb po šachovnici vzhledem k diagonálám. Každá diagonála má své číslo. Dokážeme se dostat na libovolnou danou diagonálu, pokud předpokládáme, že už jsme na diagonále, která je o jedno blíž k počátku? Ano, stačí se posunout o jedno pole doprava či nahoru a to už jsme zvládli. Udělejme to pořádně.

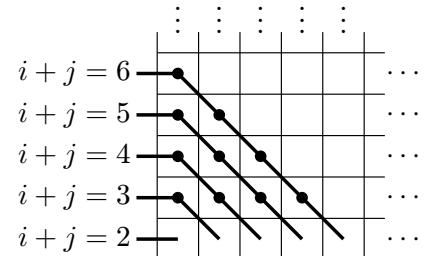
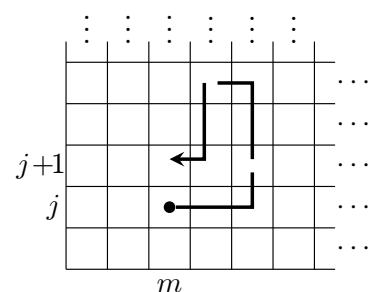
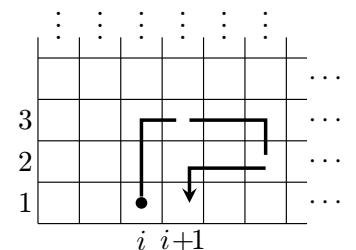
To, na jaké diagonále stojíme, když jsme na poli (i, j) , se pozná podle hodnoty součtu $i + j$ (viz obrázek). Tento parametr $n = i + j$ použijeme v naší indukci. Nejmenší možná hodnota je $1 + 1 = 2$. Budeme tedy pro $n \geq 2$ dokazovat tvrzení $V(n)$: Dokážeme se koněm dostat z pole $(1, 1)$ na libovolné pole (i, j) splňující $i + j = n$.

(0) Nechť $n = 2$. Jsme v bodě $(1, 1)$, máme se dostat na bod (i, j) splňující $i + j = 2$, takový bod je jediný, $(1, 1)$, a tam už jsme.

(1) Nechť $n \geq 2$. Předpokládáme platnost $V(n)$, tedy že se umíme dostat z bodu $(1, 1)$ do libovolného bodu (u, v) splňujícího $u + v = n$. Potřebujeme dokázat $V(n + 1)$, tedy že se z bodu $(1, 1)$ dostaneme i do libovolného bodu (i, j) splňujícího $i + j = n + 1$.

Vezměme si tedy nějaký takový bod (i, j) . Jestliže $i > 1$, tak nejsme na levém kraji a můžeme se podívat na bod $(i - 1, j)$ o jedno doleva. Ten splňuje $(i - 1) + j = (i + j) - 1 = n + 1 - 1 = n$ (je na předchozí diagonále), tudíž se tam podle indukčního předpokladu umíme dostat. Zbývá posun o jedno políčko doprava, což už třemi tahy dokážeme, viz předchozí důkaz. Tento případ je tedy hotov.

Jestliže $j > 1$, tak nejsme na dolním okraji a můžeme se podívat na bod $(i, j - 1)$ o jedno níže. Ten splňuje $i + (j - 1) = (i + j) - 1 = n + 1 - 1 = n$, tudíž se tam podle indukčního předpokladu umíme dostat. Zbývá tedy posun o jedno políčko nahoru, což už také třemi tahy dokážeme, viz předchozí důkaz. Tento případ je tedy také hotov. Všimněte si, že pokud náhodou nejsme na nějakém kraji, tak jsou splněny obě podmínky, tudíž máme na výběr, který způsob si vybereme. To není nikterak na závadu, hlavní je, aby alespoň jeden způsob fungoval.



Vyčerpali jsme tím všechny možnosti? Máme $n \geq 2$, takže vlastně uvažujeme body (i, j) splňující $i+j = n+1 \geq 3$, takže alespoň jedno z i, j musí být větší než 1. Jinými slovy, těmito dvěma možnostmi jsme pokryli možnosti všechny a důkaz je hotov.

3) Předchozí důkaz lze v jednom bodě zjednodušit. Všimneme si, že pokud jde kůň tahem od dvě pole dolů a doprava, tak se ocitnul na diagonále s o jedno menším číslem. K přechodu mezi diagonálami v předchozím důkazu tudíž vůbec nemusíme používat trojskok, ale stačí jeden základní tah koněm. Má to ale drobný zádrhel, ne vždy na něj musí být dost místa. Může se stát, že se nedokážeme vrátit o diagonálu zpět? Ano, pokud nemáme ani dvě řady směrem dolů, ani dvě řady směrem doleva. Jinými slovy, problémová jsou pole o souřadnicích $(2, 1)$, $(2, 2)$ a $(1, 2)$. Pokud bychom tedy chtěli důkaz 2) upravit na základní tah koně, musel by vypadat takto:

(0) V základním kroku bychom ukázali, že se lze z $(1, 1)$ dostat na pole (i, j) se součtem $i + j$ rovným 2, 3 a 4, to se prostě jen najdou konkrétní tahy pro tato pole.

(1) V indukčním kroku bychom uvažovali $n \geq 4$. Máme koně na výchozím poli (i, j) splňujícím $i + j = n + 1$, tedy $i + j$ je nejméně 5. Proto určitě $i \geq 3$ nebo $j \geq 3$. V prvním případě uvažujeme pole $(i - 2, j + 1)$. To je na diagonále číslo $(i - 2) + (j + 1) = n$, proto se na něj dle indukčního předpokladu dostaneme, jeden tah koně nás pak doveze na (i, j) . V druhém případě použijeme pole $(i + 1, j - 2)$.

4) Další zajímavou možností je použít strukturální indukci. Nejprve si musíme zadefinovat množinu $M = \mathbb{N} \times \mathbb{N}$ pomocí indukčních pravidel, pak se indukcí se stejnou strukturou dokáže vlastnost $V(i, j)$: lze dojet koněm z $(1, 1)$ na (i, j) .

Jednou z možností, jak definovat $\mathbb{N} \times \mathbb{N}$, je tato:

(0) $(1, 1) \in M$.

(1a) Jestliže $(i, 1) \in M$, tak $(i + 1, 1) \in M$.

(1b) Jestliže $(i, j) \in M$, tak $(i, j + 1) \in M$.

Rozmyslete si, že těmito pravidly dostanete libovolnou dvojici $(i, j) \in \mathbb{N} \times \mathbb{N}$. Princip strukturální indukce pak dovoluje důkaz vlastnosti V provést následujícími kroky:

(0) Platí $V(1, 1)$.

(1a) Jestliže platí $V(i, 1)$, pak platí $V(i + 1, 1)$.

(1b) Jestliže platí $V(i, j)$, pak platí $V(i, j + 1)$.

Zajímavou shodou okolností jsme přesně tato tvrzení dokázali v řešení 1), takže i onen důkaz šlo prezentovat jako strukturální indukci.

Jsou i jiné možnosti, jak zavést \mathbb{N}^2 . Jak příklad uvedeme následující definici.

(0) $(1, 1) \in M$.

(1) Jestliže $(m, n) \in M$, tak $(m + 1, n) \in M$ a $(m, n + 1) \in M$.

Strukturální indukce pak vede na řešení, které odpovídá důkazům z 2) a 3).

△

Poznámka: Máme několik důkazů pro šachovnici nekonečnou, zajímavá otázka je, zda se s koněm dokážeme dostat na libovolné pole šachovnice konečné. Kritickým faktorem bude, zda máme dost místa na tahy nutné k provedení kroků indukce. Podíváme se na první důkaz. Pokud se chceme posunout o jedno pole doprava, pak naše tři tahy vyžadují alespoň dvě řady navíc směrem nahoru a jeden sloupec navíc směrem doprava. Pokud není místo vpravo, tak máme možnost jít v druhém tuhu doleva, zase budeme potřebovat jeden sloupec navíc. Podobně k posunu nahoru potřebujeme alespoň dva sloupce nalevo či napravo a jednu řadu navíc nad či pod.

Po kratší úvaze lze dojít k závěru, že důkaz indukcí bude fungovat vždy, když má šachovnice alespoň 4 řady a sloupce.

Zbývají případy šachovnic 1×1 , 2×2 a 3×3 , které již snadno prozkoumáme kratší prací s tužkou a papírem. Zjistíme, že případ $n = 1$ je triviální a v případech $n = 2$ a $n = 3$ se nedokážeme dostat na pole $(2, 2)$. Nakonec tedy máme úplnou informaci, víme, že u všech velikostí šachovnic (včetně nekonečné) s výjimkou 2×2 a 3×3 se kůň dokáže dostat kamkoliv.

△

Výklad indukce teď zakončíme slíbeným důkazem ekivalence principů.

Věta 5b.3.

Platnost principu strukturální indukce je ekvivalentní platnosti principu matematické indukce.

Důkaz (drsný): 1) Nejprve ukážeme, že slabý princip indukce plyne ze strukturální indukce.

Uvažujme tedy nějakou vlastnost $V(n)$ celých čísel, která má smysl pro $n \geq n_0$ a splňuje tyto podmínky:

(s0) $V(n_0)$ platí.

(s1) Pro všechna $n \geq n_0$ platí implikace: $V(n)$ platí $\implies V(n + 1)$ platí.

Ukážeme, že když předpokládáme platnost strukturální indukce, tak V platí pro všechna $n \geq n_0$. Uvažujme množinu M definovanou předpisy

- (0) $n_0 \in M$.
- (1) $n \in M \implies n + 1 \in M$.

Pak $M \subseteq \mathbb{Z}$, proto má V smysl pro prvky M . Předpoklad (s0) ukazuje, že V je splněna pro všechny prvky ze základního pravidla (0). Předpoklad (s1) ukazuje, že když je V splněna pro nějaký prvek z předpokladu induktivního pravidla (1), pak je splněna i pro prvek z jeho závěru. V tedy splňuje předpoklady principu strukturální indukce, proto podle něj V platí pro všechny prvky M , přičemž evidentně $M = \{n \in \mathbb{Z}; n \geq n_0\}$. Proto $V(n)$ platí pro všechna $n \geq n_0$.

2) Teď ukážeme, že princip strukturální indukce plyne ze silného principu matematické indukce. Uvažujme tedy nějakou množinu M danou základními pravidly (0i) a induktivními pravidly (1j). Uvažujme také vlastnost V definovanou na M a splňující předpoklady strukturální indukce:

- (s0) V platí pro všechny prvky základních kroků.

(s1j) Jestliže je V splněna pro všechny prvky z předpokladu j -tého pravidla, pak platí i pro prvek z jeho závěru.

Ukážeme pomocí silného principu indukce, že V pak musí platit pro všechny prvky z M . Definujme proto novou vlastnost $W(n)$ na \mathbb{N} takto: $W(n)$ platí, jestliže je V splněno pro všechny prvky M s výškou n .

Tvrdíme, že tato vlastnost W splňuje předpoklady silného principu matematické indukce.

(S0): Nechť $n = 1$. $W(1)$ platí, pokud je V splněno pro všechny prvky M výšky jedna, tedy prvky ze základních pravidel. To ale platí dle (s0).

(S1): Předpokládejme, že platí $W(1)$ až $W(n)$. To znamená, že V platí pro všechny prvky množiny M , jejichž výška je nejvýše n .

Platí $W(n+1)$? Máme ukázat, že V platí pro všechny prvky množiny M výšky $n+1$. Vezměme tedy jeden takový prvek m . Protože je to prvek z M a má výšku $n+1 > 1$, tak se v M ocitnul na základě nějakého induktivního pravidla. Vezměme si tedy jeho derivační strom, který dává výšku $n+1$, a vidíme, že m vzniklo použitím nějakého induktivního pravidla (1j). Toto pravidlo má ve svém předpokladu nějaké prvky $m_i \in M$, které se v našem derivačním stromě pro m objeví o úroveň výš. Mají proto derivační strom, jehož výška je určitě menší než výška pro m , tedy všechny m_i mají výšku nejvýše n . Podle indukčního předpokladu pro ně V platí, a proto podle předpokladu (s1j) strukturální indukce musí V platit i pro prvek m , přesně jak jsme potřebovali.

$W(n+1)$ tedy platí. Ukázali jsme, že W splňuje (S0) i (S1), proto podle silného principu matematické indukce $W(n)$ platí pro všechna n , tedy V platí pro všechny prvky M . □

Cvičení

Cvičení 5b.1 (rutinní): Najděte $f(1)$, $f(2)$, $f(3)$, $f(4)$ pro f definované indukcí jako

- (i) (0) $f(0) = 1$, (1) $f(n+1) = f(n) + 2$ pro $n \in \mathbb{N}_0$;
- (ii) (0) $f(0) = 1$, (1) $f(n+1) = 3f(n)$ pro $n \in \mathbb{N}_0$;
- (iii) (0) $f(0) = 1$, (1) $f(n+1) = f(n)^2 + f(n) + 1$ pro $n \in \mathbb{N}_0$;
- (iv) (0) $f(0) = 1$, (1) $f(n+1) = 2^{f(n)}$ pro $n \in \mathbb{N}_0$.

Cvičení 5b.2 (rutinní): Najděte $f(2)$, $f(3)$, $f(4)$ pro f definované indukcí jako

- (i) (0) $f(0) = 1$, $f(1) = -2$, (1) $f(n+1) = f(n-1)^2 f(n)$ pro $n \in \mathbb{N}$;
- (ii) (0) $f(0) = 1$, $f(1) = -2$, (1) $f(n+1) = f(n-1) - 2f(n)$ pro $n \in \mathbb{N}$;
- (iii) (0) $f(0) = 1$, $f(1) = -2$, (1) $f(n+1) = \frac{f(n-1)}{f(n)}$ pro $n \in \mathbb{N}$.

Cvičení 5b.3 (rutinní, zkouškové): Uvažujte funkce definované induktivně následujícími vzorci. Pro každou z nich spočítejte několik hodnot a zkuste odhadnout, jakým vzorcem je $f(n)$ dáno. Pak dokažte, že je to správně.

- (i) (0) $f(0) = 0$, (1) $f(n+1) = 2f(n)$ pro $n \in \mathbb{N}_0$;
- (ii) (0) $f(1) = 0$, (1) $f(n+1) = f(n) + 1$ pro $n \in \mathbb{N}$;
- (iii) (0) $f(1) = 1$, (1) $f(n+1) = f(n) \cdot \frac{n}{n+1}$ pro $n \in \mathbb{N}$;
- (iv) (0) $f(1) = 1$, $f(2) = 2$, (1) $f(n+1) = 2f(n) - f(n-1)$ pro $n \in \mathbb{N}$, $n \geq 2$;
- (v) (0) $f(1) = 1$, $f(2) = 1$, $f(3) = 1$, (1) $f(n+1) = f(n) + f(n-1) - f(n-2)$ pro $n \in \mathbb{N}$, $n \geq 3$;
- (vi) (0) $f(1) = 1$, $f(2) = 0$, $f(3) = 1$, (1) $f(n+1) = f(n) + f(n-1) - f(n-2)$ pro $n \in \mathbb{N}$, $n \geq 3$;
- (vii) (0) $f(0) = 1$, $f(1) = 3$, (1) $f(n+1) = \begin{cases} 3f(n), & n \in \mathbb{N} \text{ liché;} \\ 9f(n-1), & n \in \mathbb{N} \text{ sudé;} \end{cases}$;
- (viii) (0) $f(1) = 1$, $f(2) = 2$, (1) $f(n+1) = 2f(n-1)$ pro $n \in \mathbb{N}$, $n \geq 2$;
- (ix) (0) $f(0) = 1$, $f(1) = 0$, $f(2) = 2$, (1) $f(n) = 2f(n-3)$ pro $n \in \mathbb{N}$, $n \geq 3$.

Cvičení 5b.4 (rutinní, zkouškové): Uvažujte funkce definované induktivně následujícími vzorci. Pro každou z nich dokažte zadanou (ne)rovnost.

- (i) (0) $f(1) = 1$, $f(2) = 2$, (1) $f(n+1) = f(n) + n f(n-1)$ pro $n \geq 2$; nerovnost $f(n) \leq n!$;
- (ii) (0) $f(1) = 1$, $f(2) = 2$, (1) $f(n+1) = \frac{1}{n} f(n) + f(n-1)$ pro $n \geq 2$; nerovnost $f(n) \leq n^2$;
- (iii) (0) $f(1) = 1$, $f(2) = 2$, (1) $f(n+1) = n f(n) + n f(n-1)$ pro $n \geq 2$; rovnost $f(n) = n!$;
- (iv) (0) $f(1) = 2$, $f(2) = 3$, (1) $f(n+1) = n f(n) + n^2 f(n-1)$ pro $n \geq 2$; nerovnost $f(n) \geq n!$.

Cvičení 5b.5 (rutinní, poučné): Uvažujme posloupnost danou předpisem

- (0) $F_1 = F_2 = 1$.
- (1) $F_{n+1} = F_n + F_{n-1}$ pro $n \geq 2$.

(Je to tzv. Fibonaciho posloupnost, viz příklad 9a.c.)

a) Odhadněte, která F_n jsou lichá, a dokažte to.

b) Dokažte následující vztahy:

- (i) $F_1^2 + F_2^2 + \cdots + F_n^2 = F_n F_{n+1}$ pro $n \in \mathbb{N}$;
- (ii) $F_1 + F_3 + \cdots + F_{2n-1} = F_{2n}$ pro $n \in \mathbb{N}$;
- (iii) $F_{n+1} F_{n-1} - F_n^2 = (-1)^n$ pro $n \in \mathbb{N}$;
- (iv) $F_1 F_2 + \cdots + F_{2n-1} F_{2n} = F_{2n}^2$ pro $n \in \mathbb{N}$;
- (v) $F_1 - F_2 + \cdots + F_{2n-1} - F_{2n} = 1 - F_{2n-1}$ pro $n \in \mathbb{N}$;
- (vi) $F_k F_n + F_{k+1} F_{n+1} = F_{n+k+1}$ pro $n, k \in \mathbb{N}$;
- (vii) $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} F_{n+1} & F_n \\ F_n & 0 \end{pmatrix}$ pro $n \in \mathbb{N}$.

Cvičení 5b.6 (poučné, zkouškové): Definujte množinu všech binárních slov, která:

- (i) neobsahuje více nul jdoucích po sobě;
- (ii) končí nulou;
- (iii) nekončí nulou;
- (iv) obsahuje někde v sobě kombinaci 101.

Cvičení 5b.7 (poučné, zkouškové): Definujte množinu všech slov nad abecedou $C = \{1, 2, 3, 4\}$, která:

- (i) neobsahuje více trojek jdoucích po sobě;
- (ii) začínají dvojkou;
- (iii) nekončí jedničkou;
- (iv) obsahuje stejný počet sudých a lichých číslic.

Cvičení 5b.8 (poučné): Napište nějakou rekurzivní definici správných množinových výrazů složených z velkých písmen, \cap , \cup , $-$, $\overline{}$ a závorek.

Cvičení 5b.9 (poučné, zkouškové): Napište nějakou rekurzivní definici množiny M všech slov nad anglickou abecedou C (26 malých písmen), které jsou palindromy, tj. čtou se stejně zleva doprava a zprava doleva.

Cvičení 5b.10 (poučné): Napište nějakou rekurzivní definici množiny všech polynomů s reálnými koeficienty. Návod: Indukce může zvyšovat stupně.

Cvičení 5b.11 (poučné, dobré): Napište rekurzivní definice těchto množin:

- (i) $M = \{(a, b) \in \mathbb{N} \times \mathbb{N}; a + b$ liché};
- (ii) $M = \{(a, b) \in \mathbb{N} \times \mathbb{N}; a | b\}$;
- (iii) $M = \{(a, b) \in \mathbb{N} \times \mathbb{N}; a$ nebo b liché}.

Dokažte, že vaše definice jsou správné.

Cvičení 5b.12 (poučné, dobré): Uvažujte množinu M neprázdných řetězců nad $\{a, b, c\}$ zadanou pravidly

- (0) $aa \in M$.
- (1a) $r \in M \implies raa \in M$.
- (1A) $r \in M \implies ara \in M$.
- (1b) $r \in M \implies rb \in M$.
- (1c) $r \in M \implies rc \in M$.
- (1B) $r \in M \implies br \in M$.
- (1C) $r \in M \implies cr \in M$.

Dokažte, že každý řetězec z M obsahuje sudý počet znaků a .

Návod: Uvažujte funkci $f(r)$ na M udávající počet znaků a v řetězci r .

Cvičení 5b.13 (poučné, zkouškové): Uvažujte množinu čísel M definovanou induktivně takto:

- (s0) $23 \in M$.
- (s1) $m \in M \implies 13 \cdot m \in M$.

Dokažte, že $M = \{n \in \mathbb{N}; \exists k \in \mathbb{N}_0: n = 23 \cdot 13^k\} = \{23 \cdot 13^k; k \in \mathbb{N}_0\}$.

Cvičení 5b.14 (poučné): Použijte strukturální indukci k důkazu, že čísla zadaná

- (0) $a(0, 0) = 0$;
- (1a) $a(m + 1, 0) = a(m, 0) + 1$ pro $m \in \mathbb{N}_0$;
- (1b) $a(m, n + 1) = a(m, n) + 1$ pro $m, n \in \mathbb{N}_0$

splňují $a(m, n) = m + n$ pro $m, n \in \mathbb{N}_0$.

Cvičení 5b.15 (poučné): Použijte strukturální indukci k důkazu, že čísla zadaná

- (0) $a(1, 1) = 5$;
- (1a) $a(m + 1, 1) = a(m, 1) + 2$ pro $m \in \mathbb{N}$;
- (1b) $a(m, n + 1) = a(m, n) + 2$ pro $m, n \in \mathbb{N}$

splňují $a(m, n) = 2(m + n) + 1$ pro $m, n \in \mathbb{N}$.

Řešení:

5b.1: (i): $f(1) = f(0) + 2 = 3$, $f(2) = f(1) + 2 = 5$, $f(3) = 7$, $f(4) = 9$; (ii): $f(1) = 3f(0) = 3$, $f(2) = 3f(1) = 9$, $f(3) = 27$, $f(4) = 81$; (iii): $f(1) = f(0)^2 + f(0) + 1 = 3$, $f(2) = f(1)^2 + f(1) + 1 = 13$, $f(3) = 183$, $f(4) = 33673$; (iv): $f(1) = 2^{f(0)} = 2$, $f(2) = 2^{f(1)} = 4$, $f(3) = 16$, $f(4) = 65536$.

5b.2: (i): $f(2) = f(0)^2 f(1) = -2$, $f(3) = f(1)^2 f(2) = -8$, $f(4) = -32$; (ii): $f(2) = f(0) - 2f(1) = 5$, $f(3) = f(1) - 2f(2) = -12$, $f(4) = 29$; (iii): $f(2) = \frac{f(0)}{f(1)} = -\frac{1}{2}$, $f(3) = \frac{f(1)}{f(2)} = 4$, $f(4) = -\frac{1}{8}$.

5b.3: (i): $f(n) = 0$. Slabý princip. (0) $n = 0$ funguje. (1) Nechť $n \in \mathbb{N}_0$. Předpokládejme, že $f(n) = 0$. Pak $f(n + 1) = 2f(n) = 0$, souhlasí pro $n + 1$.

(ii): $f(n) = n - 1$. Slabý princip. (0) $n = 1$ funguje. (1) Nechť $n \in \mathbb{N}$. Předpokládejme, že $f(n) = n - 1$. Pak $f(n + 1) = f(n) + 1 = n - 1 + 1 = (n + 1) - 1$, souhlasí pro $n + 1$.

(iii): $f(n) = \frac{1}{n}$. Slabý princip. (0) $n = 1$ funguje. (1) Nechť $n \in \mathbb{N}$. Předpokládejme, že $f(n) = \frac{1}{n}$. Pak $f(n + 1) = f(n) \frac{n}{n+1} = \frac{1}{n} \frac{n}{n+1} = \frac{1}{n+1}$, souhlasí pro $n + 1$.

(iv): $f(n) = n$. Modifikovaný princip. (0) $n = 1$ a $n = 2$ funguje. (1) Nechť $n \in \mathbb{N}$. Předpokládejme, že $f(k) = k$ pro $k = n - 1, n$. Pak $f(n + 1) = 2f(n) - f(n - 1) = 2n - (n - 1) = n + 1$, souhlasí pro $n + 1$.

(v): $f(n) = 1$. Modifikovaný princip. (0) $n = 1, n = 2$ a $n = 3$ funguje. (1) Nechť $n \in \mathbb{N}$. Předpokládejme, že $f(k) = 1$ pro $k = n - 2, n - 1, n$. Pak $f(n + 1) = f(n) + f(n - 1) - f(n - 2) = 1 + 1 - 1 = 1$, souhlasí pro $n + 1$.

(vi): $f(n) = \begin{cases} 1, & n \text{ liché;} \\ 0, & n \text{ sudé.} \end{cases}$ Modifikovaný princip: (0) Pro $n = 1, 2, 3$ to funguje.

(1) Nechť $n \in \mathbb{N}$, $n \geq 3$. Předpokládejme, že $f(k) = \begin{cases} 1, & k \text{ liché;} \\ 0, & k \text{ sudé} \end{cases}$ pro $k = n - 2, n - 1, n$.

a) Je-li n sudé, pak $n - 2$ je sudé, zato $n - 1$ a $n + 1$ jsou liché. Pak $f(n + 1) = f(n) + f(n - 1) - f(n - 2) = 0 + 1 - 0 = 1$, souhlasí pro liché $n + 1$.

b) Je-li n liché, pak $n - 2$ je liché, zato $n - 1$ a $n + 1$ jsou sudé. Pak $f(n + 1) = f(n) + f(n - 1) - f(n - 2) = 1 + 0 - 1 = 0$, souhlasí pro sudé $n + 1$.

Alternativa: $f(n) = \frac{1}{2}(1 - (-1)^n)$, pak lze přímo, z indukčního předpokladu vyjde

$$f(n + 1) = f(n) + f(n - 1) - f(n - 2) = \frac{1}{2}(1 - (-1)^n) + \frac{1}{2}(1 - (-1)^{n-1}) - \frac{1}{2}(1 - (-1)^{n-2}) = \frac{1}{2}(1 - (-1)^n + 1 + (-1)^n - 1 + (-1)^n) = \frac{1}{2}(1 + (-1)^n) = \frac{1}{2}(1 - (-1)^{n+1}).$$

(vii): $f(n) = 3^n$. Modifikovaný princip. (0) $n = 1$ a $n = 2$ funguje. (1) Nechť $n \in \mathbb{N}$. Předpokládejme, že $f(k) = 3^k$ pro $k = n - 1, n$. Uvažujme $n + 1$.

Je-li $n + 1$ sudé, pak $f(n + 1) = 9f(n - 1) = 9 \cdot 3^{n-1} = 3^{n+1}$.

Je-li $n + 1$ liché, pak $f(n + 1) = 3f(n) = 3 \cdot 3^n = 3^{n+1}$.

(viii): $f(n) = 2^{\lfloor n/2 \rfloor}$. Modifikovaný princip. (0) $n = 1$ a $n = 2$ funguje. (1) Nechť $n \in \mathbb{N}$. Předpokládejme, že $f(k) = \begin{cases} 2^{k/2}, & k \text{ sudé;} \\ 2^{(k-1)/2}, & k \text{ liché} \end{cases}$ pro $k = n - 1, n$. Uvažujme $n + 1$.

Je-li $n + 1$ sudé, pak je i $n - 1$ sudé a $f(n + 1) = 2f(n - 1) = 2 \cdot 2^{(n-1)/2} = 2^{(n+1)/2}$.

Je-li $n + 1$ liché, pak je i $n - 1$ liché a $f(n + 1) = 2f(n - 1) = 2 \cdot 2^{(n-1-1)/2} = 2^{n/2} = 2^{((n+1)-1)/2}$.

(ix): Jak zapsat $1, 0, 2, 2, 0, 4, 4, 0, 8, 8, 0, 16, 16, 0, \dots$? Nápad: $f(n) = \begin{cases} 0, & n = 3k + 1; \\ 2^{\lfloor (n+1)/3 \rfloor}, & n \neq 3k + 1. \end{cases}$

Důkaz modifikovanou indukcí podobný předchozímu, ale teď se musí řešit tři případy.

5b.4: (i): Modifikovaná indukce (0) Pro $n = 1, 2$ to platí. (1) Nechť $n \geq 2$, předpoklad platnosti $V(k)$: $f(k) \leq k!$ pro $k = n - 1, n$. Pak $f(n + 1) = f(n) + n f(n - 1) \leq n! + n \cdot (n - 1)! = 2n! \leq (n + 1)n! = (n + 1)!$.

(ii): Modifikovaná indukce (0) Pro $n = 1, 2$ to platí. (1) Nechť $n \geq 2$, předpoklad platnosti $V(k)$: $f(k) \leq k^2$ pro $k = n - 1, n$. Pak $f(n + 1) = \frac{1}{k}f(n) + f(n - 1) \leq \frac{1}{n}n^2 + (n - 1)^2 = n + n^2 - 2n + 1 = n^2 - n + 1 \leq n^2 + 2n + 1 = (n + 1)^2$.

(iii): Modifikovaná indukce (0) Pro $n = 1, 2$ to platí. (1) Nechť $n \geq 2$, předpoklad platnosti $V(k)$: $f(k) = k!$ pro $k = n - 1, n$. Pak $f(n + 1) = n f(n) + n f(n - 1) = n \cdot n! + n \cdot (n - 1)! = n \cdot n! + n! = (n + 1) \cdot n! = (n + 1)!$.

(iv): Modifikovaná indukce (0) Pro $n = 1, 2$ to platí. (1) Nechť $n \geq 2$, předpoklad platnosti $V(k)$: $f(k) \geq k!$ pro $k = n-1, n$. Pak $f(n+1) = n f(n) + n^2 f(n-1) = n \cdot n! + n^2 \cdot (n-1)! = n \cdot n! + n \cdot n! = 2n \cdot n! \geq (n+1) \cdot n! = (n+1)!$

5b.5: a): F_{3n} sudé, F_{3n+1} a F_{3n+2} liché. Důkaz najednou indukcí: (0) $n = 0$ funguje, pokud doplníme $F_0 = 0$. (1) Předpoklad: $n \in \mathbb{N}_0$ a F_{3n} sudé, F_{3n+1} a F_{3n+2} liché.

Pak $F_{3(n+1)} = F_{3n+3} = F_{3n+2} + F_{3n+1}$. Protože jsou dle indukčního předpokladu F_{3n+2} a F_{3n+1} liché, je $F_{3(n+1)}$ sudé.

Dále $F_{3(n+1)+1} = F_{3n+4} = F_{3n+3} + F_{3n+2}$. Protože je dle indukčního předpokladu F_{3n+2} liché a již jsme dokázali, že F_{3n+3} je sudé, je $F_{3(n+1)+1}$ liché.

Dále $F_{3(n+1)+2} = F_{3n+5} = F_{3n+4} + F_{3n+3}$. Protože jsme dokázali, že F_{3n+4} je liché a F_{3n+3} je sudé, je $F_{3(n+1)+2}$ liché.

b) (i): Slabá indukce: (0) $n = 1$ funguje, $1^2 = 1 \cdot 1$.

(1) Předpoklad: Platí to pro jisté $n \in \mathbb{N}$. Pak

$$F_1^2 + F_2^2 + \cdots + F_n^2 = [F_1^2 + F_2^2 + \cdots + F_n^2] + F_{n+1}^2 = F_n F_{n+1} + F_{n+1}^2 = F_{n+1}(F_n + F_{n+1}) = F_{n+1} F_{n+2}.$$

Zbytek podobně.

5b.6: (i): (0a) $0 \in M$. (0b) $1 \in M$. (0c) $10 \in M$.

(1a) $w \in M \implies w1 \in M$. (1b) $w \in M \implies w10 \in M$.

Poznámka: Bez (0c) nelze získat 101.

(ii): (0) $0 \in M$.

(1a) $w \in M \implies 0w \in M$. (1b) $w \in M \implies 1w \in M$.

Poznámka: Nutno přidávat nalevo, zprava nejde zaručit správné ukončení.

(iii): (0) $1 \in M$.

(1a) $w \in M \implies 0w \in M$. (1b) $w \in M \implies 1w \in M$.

Poznámka: Nutno přidávat nalevo, zprava nejde zaručit správné ukončení.

(iv): (0) $101 \in M$.

(1a) $w \in M \implies 0w \in M$. (1b) $w \in M \implies 1w \in M$. (1c) $w \in M \implies w0 \in M$.

(1d) $w \in M \implies w1 \in M$.

5b.7: (i): (0a) $c \in C \implies c \in M$. (0b) $c \in C - \{3\} \implies c3 \in M$.

(1a) $[w \in M \wedge c \in C - \{3\}] \implies wc \in M$. (1b) $[w \in M \wedge c \in C - \{3\}] \implies wc3 \in M$.

Poznámka: Bez (0b) nelze získat 13.

(ii): (0) $2 \in M$. (1) $[w \in M \wedge c \in C] \implies wc \in M$.

(iii): (0) $c \in C - \{1\} \implies c \in M$. (1) $[w \in M \wedge c \in C] \implies cw \in M$.

Poznámka: Nutno přidávat nalevo, zprava nejde zaručit správné ukončení.

(iv): (0a) $\lambda \in M$. (0b) $[c \in \{1, 3\} \wedge d \in \{2, 4\}] \implies cd \in M$. (0c) $[c \in \{1, 3\} \wedge d \in \{2, 4\}] \implies dc \in M$.

(1a) $[r, s \in M \wedge c \in \{1, 3\} \wedge d \in \{2, 4\}] \implies rcsd \in M$. (1b) $[r, s \in M \wedge c \in \{1, 3\} \wedge d \in \{2, 4\}] \implies rdsc \in M$.

Poznámka: Přidáváme napravo, vždy někam doprostřed vsuneme číslo opačné parity. Je to dobré? Úvaha přes rekurentní postup, od daného řetězce vždy odebereme pravý krajiní znak a s ním i nějaký znak opačné parity ze zbytku řetězce. Je pak nutné umožnit odebírání zevnitř, viz řetězec 11222211, ale lze se obejít bez odebírání z druhého kraje. Má-li řetězec alespoň 4 znaky, pak v něm musí existovat alespoň dva znaky parity opačné než je ta na pravém konci, tudíž alespoň jeden znak správné parity je někde uprostřed a dá se odebrat.

Podmínku (0a) jsme museli přidat, jinak bychom nedokázali vytvořit třeba 1122.

5b.8: (0) $A, B, \dots, Z \in \mathcal{M}$.

(1a) $v \in \mathcal{M} \implies \bar{v} \in \mathcal{M}$. (1c) $v_1, v_2 \in \mathcal{M} \implies (v_1 \cup v_2) \in \mathcal{M}$.

(1b) $v_1, v_2 \in \mathcal{M} \implies (v_1 \cap v_2) \in \mathcal{M}$. (1d) $v_1, v_2 \in \mathcal{M} \implies (v_1 - v_2) \in \mathcal{M}$.

5b.9: (0a) $c \in C \implies c \in M$. (0b) $c \in C \implies cc \in M$.

(1) $[w \in M \wedge c \in C] \implies cwc \in M$.

5b.10: (0) $a \in \mathbb{R} \implies a \in P$. (1) $[p \in P \wedge a \in \mathbb{R}] \implies x \cdot p + a \in P$.

Alternativa (méně elegantní): (1) $[p \in P \wedge a \in \mathbb{R} \wedge n \in \mathbb{N}] \implies p + ax^n \in P$.

Poznámka: Proč by nefungovala definice (1) $[p \in P \wedge a \in \mathbb{R}] \implies (x - a) \cdot p \in P$? Což takhle $x^2 + 1$?

5b.11: (i): (0) $(2, 0) \in S$, $(2, 1) \in S$.

(1a) $(a, b) \in S \implies (a + 2, b) \in S \in S$. (1b) $(a, b) \in S \implies (a, b + 2) \in S$.

a) $S \subseteq M$ strukturální indukcí: (0) $1 + 2 = 3$ liché, proto $(1, 2), (2, 1) \in M$.

(1) Předpoklad: $(a, b) \in S$ splňuje $(a, b) \in M$. Pak $a + b$ je liché a tudíž je i $a + b + 2$ liché, proto prvky ze závěru

(1a) a (1b) splňují $(a + 2, b) \in M$ a $(a, b + 2) \in M$.

b) $M \subseteq S$ nejlépe silnou indukcí na $a + b$. Vlastnost $V(n)$: Každá dvojice $(a, b) \in \mathbb{N}^2$ s vlastností $a + b = 2n + 1$ leží v S . Protože $a, b \geq 1$, je nejmenší možný lichý součet 3, proto bereme $n \geq 1$.

(0) $n = 1$: Jestliže $a + b = 1$, pak z $a, b \in \mathbb{N}_0$ plyne, že $(a, b) = (1, 0)$ nebo $(a, b) = (0, 1)$, každopádně dle (0) v definici $(a, b) \in S$.

(1) Předpokládáme platnost pro jisté $n \in \mathbb{N}$. Nechť $(a, b) \in \mathbb{N}_0^2$ splňuje $a + b = 2(n+1) + 1 = 2n + 3$. Pak $a + b \geq 5$ a proto je alespoň jedno z čísel a, b větší než 2. Možnost $a \geq 3$: Pak $(a-2, b) \in \mathbb{N}^2$ a $(a-2) + b = 2n + 1$, proto dle indukčního předpokladu $(a-2, b) \in S$. Pak ale dle (1a) také $(a, b) = ((a-2) + 2, b) \in S$. Možnost $b \geq 3$ obdobně.

(ii): $(0, 1, 1) \in S$.

(1a) $[(a, b) \in S \wedge c \in \mathbb{N}] \implies (ac, bc) \in S$. (1b) $[(a, b) \in S \wedge c \in \mathbb{N}] \implies (a, bc) \in S$.

a) $S \subseteq M$ lehce strukturální indukcí. a b) $M \subseteq S$ nejlépe ve dvou krocích. V prvním kroku indukci na a dokázat, že $(a, a) \in S$. V druhém kroku silnou indukcí na $\frac{b}{a}$ dokázat $M \subseteq S$.

(iii): $(0, 1, 1) \in S, (1, 2) \in S, (2, 1) \in S$.

(1a) $(a, b) \in S \implies (a+2, b) \in S$. (1b) $(a, b) \in S \implies (a, b+2) \in S$.

a) $S \subseteq M$ lehce strukturální indukcí. b) $M \subseteq S$ nejlépe silnou indukcí na $a+b$, protože $a+b-2$ znamená, že $a-2$ či $b-2$ má stejnou paritu jako a či b , tedy z lichého bude liché.

5b.12: Uvažujme $f(r)$ na M udávající počet znaků a v řetězci r . Dokážeme, že f má v M sudé hodnoty.

Silná indukce. (0) V pravidle (d0) vznikají prvky aa , pro něž $f(aa) = 2$.

(1) Nechť r je prvek z M , indukční předpoklad je, že $f(r)$ je sudé.

Pak pro prvek vzniklý z (1a) platí $f(raa) = f(r) + 2$, což je také sudé. Pro prvek vzniklý z (1A) platí $f(ara) = f(r) + 2$, což je také sudé. Pro prvek vzniklý z (1b) platí $f(rb) = f(r)$, což je také sudé. Podobně pro ostatní pravidla.

5b.13: Dvě inkluze

1) $W(k)$: $23 \cdot 13^k \in M$ indukcí (slabým principem):

(0) $k=0$: $23 \cdot 13^0 = 23 \in M$ dle (s0).

(1) $k \in \mathbb{N}_0$, nechť $W(k)$ platí, tedy $23 \cdot 13^k \in M$. Pak podle (s1) je v M i $13 \cdot 23 \cdot 13^k = 23 \cdot 13^{k+1}$, tedy $W(k+1)$ platí.

Proto W platí pro všechna $k \in \mathbb{N}_0$ a $\{23 \cdot 13^k; k \in \mathbb{N}_0\} \subseteq M$.

2) $V(m)$ vlastnost, že pro $m \in M$ existuje $k \in \mathbb{N}_0$: $m = 23 \cdot 13^k$. Důkaz strukturální indukci, že $V(m)$ platí pro všechna $m \in M$.

(0) Základní pravidlo obsahuje jen 23 , a $23 = 23 \cdot 13^0$. V platí pro prvky ze základního kroku.

(1) Vezměme prvek $m \in M$ z předpokladu induktivního pravidla. Předpoklad: V pro něj platí, tj. existuje $k \in \mathbb{N}_0$ splňující $m = 23 \cdot 13^k$. Závěr pravidla do M dává prvek $13m$. Pro něj máme $13m = 13 \cdot 23 \cdot 13^k = 23 \cdot 13^{k+1}$, tedy i pro něj platí V .

Podle (0), (1) a strukturální indukce V platí pro všechna $m \in M$, tedy $M \subseteq \{23 \cdot 13^k; k \in \mathbb{N}_0\}$.

5b.14: Definujme množinu $M = \mathbb{N}_0^2$ strukturální indukci takto:

(d0): $(0, 0) \in M$.

(d1a): $(m, 0) \in M \implies (m+1, 0) \in M$.

(d1b): $(m, n) \in M \implies (m, n+1) \in M$.

Vlastnost $V(m, n)$ na množině M : pro $(m, n) \in M$ platí $a(m, n) = m+n$. Platnost dokážeme strukturální indukci dle definice M :

(0) Pro prvek $a(0, 0) = 0$ to platí.

(1) Pravidlo (1a): Předpokládejme, že $V(m, n)$ platí pro prvek $(m, 0) \in M$, tedy $a(m, 0) = m+0 = m$. Pak dle (1a) platí $a(m+1, 0) = a(m, 0) + 1 = (m+1) + 0$, tedy $V(m, n)$ platí také pro prvek $(m+1, 0)$.

Pravidlo (1b): Předpokládejme, že $V(m, n)$ platí pro prvek $(m, n) \in M$, tedy $a(m, n) = m+n$. Pak dle (1b) platí $a(m, n+1) = a(m, n) + 1 = m+(n+1)$, tedy $V(m, n)$ platí také pro prvek $(m, n+1)$.

5b.15: (1b) $a(m, n+1) = a(m, n) + 2$ pro $m, n \in \mathbb{N}$

splňují $a(m, n) = 2(m+n)+1$ pro $m, n \in \mathbb{N}$.

Definujme množinu $M = \mathbb{N}^2$ strukturální indukci takto:

(d0): $(1, 1) \in M$.

(d1a): $(m, 1) \in M \implies (m+1, 1) \in M$.

(d1b): $(m, n) \in M \implies (m, n+1) \in M$.

Vlastnost $V(m, n)$ na množině M : pro $(m, n) \in M$ platí $a(m, n) = 2(m+n)+1$. Platnost dokážeme strukturální indukci dle definice M :

(0) Pro prvek $a(1, 1) = 5$ to platí.

(1) Pravidlo (1a): Předpokládejme, že $V(m, n)$ platí pro prvek $(m, 1) \in M$, tedy $a(m, 1) = 2(m+1)+1 = 2m+3$. Pak dle (1a) platí $a(m+1, 1) = a(m, 1) + 2 = 2m+5 = 2((m+1)+1)+1$, tedy $V(m, n)$ platí také pro prvek $(m+1, 1)$.

Pravidlo (1b): Předpokládejme, že $V(m, n)$ platí pro prvek $(m, n) \in M$, tedy $a(m, n) = 2(m+n)+2 = 2m+2n+2$. Pak dle (1b) platí $a(m, n+1) = a(m, n) + 2 = 2m+2n+4 = 2(m+(n+1))+2$, tedy $V(m, n)$ platí také pro prvek $(m, n+1)$.

6. Dělitelnost a prvočísla

Tato kapitola má jednak význam vysoce praktický pro computer science, zároveň je to také jemný úvod do oblasti zvané teorie čísel, která se zabývá mimo jiné prvočíslami a má za sebou úctyhodnou historii a zajímavé a hluboké výsledky, o některých se zde občas lehce zmíníme. Velice úzce souvisí s kapitolou následující, dokonce to původně byla jedna kapitola, než narostla tak, že už to na jednu bylo moc.

6a. Dělitelnost

Zde se pořádně matematicky podíváme na situace, kdy jedno číslo dělí druhé, mnohé z toho čtenář jistě zná.

! Definice.

Nechť $a, b \in \mathbb{Z}$. Řekneme, že a **dělí** b , značeno $a|b$, jestliže existuje $k \in \mathbb{Z}$ takové, že $b = k \cdot a$.

V takovém případě říkáme, že a je **faktor** b a že b je **násobek** a . Také říkáme, že b je **dělitelné číslem** a .

Let $a, b \in \mathbb{Z}$. We say that a **divides** b , denoted $a|b$, if there is $k \in \mathbb{Z}$ such that $b = k \cdot a$.

Then we say that a is a **factor** of b and that b is a **multiple** of a .

Příklad 6a.a: Evidentně $3|12$, protože $12 = 4 \cdot 3$, na druhou stranu neplatí $5|13$.

△

Dělitelnost je velice důležitý pojem v teorii čísel a pro některá a existují algoritmy, jak snadno poznat, zda dělí určité b , čtenář jistě zná kritérium pro dělitelost dvěma (číslo je sudé), pro další viz cvičení 6a.11 a 7a.4 a poznámka 7a.14. Teď se podíváme, jaké vlastnosti pojmu dělitelnosti splňuje, některé věci jsou jasné hned.

! Fakt 6a.1.

Pro každé $a \in \mathbb{Z}$ platí:

- (i) $1|a$;
- (ii) $a|a$;
- (iii) $a|0$.

Důkaz je tak snadný, že jej s důvěrou necháme jako cvičení 6a.2. Začátečníka může zarazit, že Fakt lze aplikovat i na $a = 0$. Opravdu dělí nula nulu? Podle definice by mělo existovat $k \in \mathbb{Z}$ tak, aby $0 = k \cdot 0$, což se nám hravě podaří splnit, třeba $k = 13$ bude fungovat. Takže ano, $0|0$.

Toto ukazuje, že pojmu dělitelnosti se zásadně liší od dělení jako matematické operace, protože samozřejmě nulou dělit nelze nic, ani nulu. Pojem dělitelnosti $a|b$ se ptá, zda lze jedno číslo nějak získat z druhého, zkoumáme tedy jistý vzájemný vztah. Na druhou stranu když napíšeme $\frac{a}{b}$, tak nás zajímá výsledek jakési operace, což nás (až na pár výjimek) nebude v této kapitole zajímat. Silně čtenáři doporučujeme se zlomkům vyhýbat, zejména v důkazech, protože svádí na zcestí. I my je zde použijeme výjimečně.

Příklad 6a.b: Zvolme si nějaké $d \in \mathbb{N}$. Kolik existuje násobků d , které jsou menší než nějaké $n \in \mathbb{N}$? Jinými slovy, kolik z přirozených čísel z $\{1, 2, \dots, n\}$ je dělitelných d ?

Pomůže postřeh, že ještě v množině $\{1, 2, \dots, d-1\}$ není žádné, v množinách $\{1, 2, \dots, d\}$ až $\{1, 2, \dots, 2d-1\}$ je jedno, v množinách $\{1, 2, \dots, 2d\}$ až $\{1, 2, \dots, 3d-1\}$ jsou dvě a tak dále. Rozmyslete si, že se to dá vzorcem vyjádřit snadno takto: Těchto čísel je $\left\lfloor \frac{n}{d} \right\rfloor$.

△

Práci s dělitelností nám usnadní užitečná pravidla.

! Věta 6a.2.

Nechť $a, b, c \in \mathbb{Z}$.

- (i) Jestliže $a|b$ a $a|c$, pak $a|(b+c)$.
- (ii) Jestliže $a|b$, pak $a|(nb)$ pro všechna $n \in \mathbb{Z}$.
- (iii) Jestliže $a|b$ a $b|c$, pak $a|c$.
- (iv) $a|b$ právě tehdy, když $|a||b|$.
- (v) Jestliže $a|b$ a $b \neq 0$, tak $|a| \leq |b|$.

Důkaz (rutinní, poučný): (i): Z předpokladu $a|b$ máme $b = ka$ pro nějaké konkrétní $k \in \mathbb{Z}$, podobně $c = la$ pro nějaké $l \in \mathbb{Z}$. Pak $b + c = (k + l)a$ a $(k + l) \in \mathbb{Z}$, proto podle definice $a|(b + c)$.

(ii) a (iii) viz cvičení 6a.3.

(iv): Viz cvičení 6a.4, ukazuje se tam, že na znaménku opravdu nezáleží.

(v): $a|b$ dává $b = ka$ pro nějaké $k \in \mathbb{Z}$. Jestliže $b \neq 0$, tak také $k \neq 0$, tudíž $|k| \in \mathbb{N}$. To znamená, že $|k| \geq 1$ a proto $|b| = |k| \cdot |a| \geq |a|$. □

S 6a.3 Poznámka o důkazech: Důkazy byly velice snadné a vysoce poučné. Měly stejné schéma: Začalo se danými předpoklady, z nich se získala nějaká informace podle definice dělitelnosti. Pak se tato informace použila k bližšímu nahlédnutí na objekt, který jsme měli zkoumat (třeba na $b + c$ v prvním tvrzení). Toto schéma bude čtenáři dobré sloužit v důkazech jednodušších tvrzení ze všech oborů matematiky.

Využijeme důkazu tvrzení (i) k upozornění na dva faktory, které někdy působí začátečníkům potíže.

1. Aby se ze studenta stal zkušený důkazovník, musí se mimo jiné naučit, kdy má svobodu volby písmenka a kdy ne. Občas je u studenta možné vidět takovýto začátek důkazu: Máme dánou, že $b = k \cdot a$ a $c = k \cdot a$. To první je dobře, druhé ne. Ve chvíli, kdy píšeme ten první vztah, můžeme na místě k dát libovolné písmeno (kromě a, b), zde máme svobodu, proč tedy nevzít tradiční k .

Jakmile ale napíšeme, že $b = k \cdot a$, tak už k získalo nějakou pevnou hodnotu, kterou sice neznáme, ale ona existuje (závisí na a, b), takže k už není k dispozici. Proto když začneme pracovat s číslem c , musíme nutně zvolit v definici dělitelnosti jiné písmenko. Jinak bychom totiž dostali $b = k \cdot a$, $c = k \cdot a$ a tedy vynutili $b = c$, což rozhodně nemusí být pravda.

2. V důkazech tohoto typu se často objevuje klasická začátečnická chyba, kdy důkaz začne slovy „nechť $b+c = k \cdot a$ “. Pokud ještě student následně napíše, že z předpokladu získá $b = k \cdot a$, tak najel do slepé uličky, protože jedno k hraje dvě rozdílné role a to už nerozechodí.

Představme si tedy studenta trochu chytřejšího, který k předpokladu $b + c = ka$ ještě přidá, že $b = xa$ a $c = ya$ pro nějaká $x, y \in \mathbb{Z}$. Jednoduchou úpravou pak získá rovnost $k = x + y$, tu dvakrát podtrhne a je spokojen. Bohužel však nemá platný důkaz. Začal totiž něčím, co nemá k dispozici, co naopak chce získat na konci.

Samozřejmě že když důkaz vymýslíme, tak si položíme otázku, kam chceme dojít, a odpověď $b + c = ka$ nás pak navádí správným směrem, Po rozmyšlení situace je pak ale třeba začít psát pořádně důkaz, tedy začít tím, co je dánou, a skončit tím, co chceme dokázat, jak jsme to udělali v dukazu Faktu výše.

△

Vraťme se k tvrzením z Faktu. Bod (iv) ukazuje, že u dělitelnosti stačí dobře rozumět, jak funguje na \mathbb{N}_0 , a už jí rozumíme všude. Někteří autoři by tuto kapitolu dělali výhradně pro čísla z \mathbb{N} , čímž by se také elegantně vyhnuli problémům s nulou, která se občas musí dělat jako zvláštní případ. Bylo by to pak snažší, ale my budeme potřebovat některé z poznatků pro všechna čísla ze \mathbb{Z} , tak to tu probereme v plné obecnosti.

Zkušenost naznačuje, že směr v (i) nejde obecně obrátit, například $3|(2+4)$, ale neplatí $3|2$ ani $3|4$. Trocha experimentování ukáže, že když platí $a|(b+c)$, ale neplatí výrok „ $a|b$ a $a|c$ “, tak se musí pokazit obě jeho části, nelze to zkazit jen u jedné. To se občas hodí, tak si to vyjádříme. Přidáme i zajímavou kombinaci tvrzení z (i) a (ii), mohlo by vám to připomenout lineární kombinace z lineární algebry.

Důsledek 6a.4.

Nechť $a, b, c \in \mathbb{Z}$.

- (i) Jestliže $a|b$ a $a|c$, pak $a|(mb + nc)$ pro všechna $m, n \in \mathbb{Z}$.
- (ii) Jestliže $a|(b+c)$ a $a|b$, pak $a|c$.

Druhé tvrzení plyne hned, protože z $a|b$ díky Větě 6a.2 (ii) dostaneme $a|(-b)$ a tudíž podle Věty 6a.2 (i) musí a dělit číslo $(b+c)+(-b)=c$. Nemusíme tedy jít do hloubky, do detailů, stačilo chytře poskládat již dokázané výsledky. Tomuto se někdy říká „měkký důkaz“ a máme je rádi, je samozřejmě efektivní nedělat věci znova, když už můžeme využít plodů naší předchozí práce. Zde by mimochodem důkaz přes definici také nebyl těžký, zkuste si to jako cvičení.

Teď si připomeneme něco z kapitoly 4b.

! Věta 6a.5.

Relace $a|b$ je částečné uspořádání na \mathbb{N} a na \mathbb{N}_0 .

Důkaz (rutinní): Důkaz provedeme společně pro obě množiny. Reflexivita plyne z Faktu 6a.1 (ii), tranzitivita zase z Věty 6a.2 (iii). Zbývá antisimetrie.

Nechť $a, b \in \mathbb{N}_0$ splňují $a | b$ a $b | a$. Pak $b = ka$ a $a = lb$ pro nějaké $k, l \in \mathbb{Z}$. Pokud by alespoň jedno z čísel a, b bylo nula, pak z našich dvou rovností okamžitě dostáváme, že i druhé je nula. Dostáváme tedy $a = b$ a antisimetrie pro tento případ funguje.

Zbývá prozkoumat situaci, kdy jsou obě a, b kladné, tudíž musí být kladné i k, l , tedy $k, l \in \mathbb{N}$. Když z první rovnosti dosadíme do druhé, dostaneme $a = kla$, také $a \neq 0$, tedy $kl = 1$. Protože $k \geq 1$, máme $1 = kl \geq l \geq 1$, tedy $l = 1$, odtud $k = 1$. Proto $a = b$. Antisimetrie dokázána. \square

Rozhodně neplatí, že by relace $a | b$ byla uspořádáním na \mathbb{Z} , tam ztrácíme antisimetrii. Například $13 | (-13)$ a $(-13) | 13$, ale neplatí $-13 = 13$. To je jeden z důvodů, proč se s dělitelností pracuje lépe jen na \mathbb{N} . Připomeňme také, že v kapitole jsme viděli, že uspořádání dělitelností není lineární ani dobré.

Při práci s dělitelností se vyplatí umět dělit se zbytkem, tedy pro daná a, d umět spočítat $\frac{a}{d} = q + \frac{r}{d}$. Protože zde se dělení a zlomkům vyhýbáme, přepíšeme tento vztah do jazyka násobení.

!

Věta 6a.6. (o dělení se zbytkem) (division theorem) (division algorithm)

Nechť $a, d \in \mathbb{Z}$, $d \neq 0$. Pak existují $q \in \mathbb{Z}$ a $r \in \mathbb{N}_0$ takové, že $a = qd + r$ a $0 \leq r < |d|$.

Čísla q a r jsou jednoznačně určena.

Důkaz (dobrý, poučný): Neprve dokážeme existenci q a r pro $d > 0$, a to hned dvakrát za cenu jednoho důkazu, no neberte to.

1) První verze začíná důkazem, že to umíme pro $a, d > 0$, a to indukcí. Vhodná formulace vyžaduje trochu experimentování, autorovi přijde jako dobré východisko tato vlastnost:

$V(n)$: Pro každé $a \in \{1, 2, \dots, n\}$ a pro každé $d \in \mathbb{N}$ existují q, r dle předpisu.

Dokážeme ji pro $n \in \mathbb{N}$.

(0) $n = 1$: Platí tvrzení pro $a = 1$? Buď $d = 1$, pak $q = 1$ a $r = 0$ splňují zadání ($1 = 1 \cdot 1 + 0$ a $r < d$), nebo $d > 1$, pak $q = 0$ a $r = a = 1$ (ano, $1 = 0 \cdot d + 1$ a $r < d$). Platí.

(1) Pro $n \in \mathbb{N}$ předpokládejme platnost $V(n)$, tedy že umíme dělit se zbytkem čísla $a = 1, 2, \dots, n$. Dokážeme teď $V(n+1)$. Vezměme tedy libovolné $a \in \{1, 2, \dots, n, n+1\}$. Jestliže $a \leq n$, pak dělit umíme dle indukčního předpokladu. Co když $a = n+1$?

Jestliže $d > a$, pak dáme $q = 0$, $r = a$ a je hotovo, $a = 0 \cdot d + a$ a $a < d$.

Jestliže $d = a$, pak dáme $q = 1$, $r = 0$ a je hotovo.

Jestliže $d < a$, pak uvažujme $b = a - d$. Máme $1 \leq b \leq n$, proto dle indukčního předpokladu $b = q'd + r$, kde $0 \leq r < d$. Pak ovšem $a = b + d = (q' + 1)d + r$ a je to hotovo, pořád $0 \leq r < d$.

Tím je dokázána možnost dělení se zbytkem pro $a, d > 0$. Pro $a = 0$ je to triviální ($q = r = 0$), zbývá případ $a < 0, d > 0$.

Uvažujme $(-a) > 0$. Podle první části důkazu je $-a = q'|d| + r'$. Jestliže $-a = q'd$, pak $a = (-q')d$ a je to hotovo, $r = 0$. Druhá alternativa je, že $0 < r' < d$. Pak $a = (-q')d - r' = (-q')d - d + d - r' = (-q' - 1)d + (d - r')$ a $r = d - r'$ splňuje $0 < r < d$, hotovo.

2) Teď to zkusíme jinak: Budeme postupně odečítat $d > 0$ od a a čekat, až to dopadne dobré. To se nejlépe udělá najednou chytrou definicí množiny. Uvažujme tedy množinu $M = \{a - qd; q \in \mathbb{Z}, a - qd \geq 0\}$. Tato množina je neprázdná: Jestliže je $a \geq 0$, tak volba $q = 0$ ukazuje $a \in M$. Jestliže $a < 0$, pak stačí zvolit $q = -a$ a máme $a - qd = a(1 - d) \geq 0$, neboť díky $d \geq 1$ máme $(1 - d) \leq 0$.

Už z definice jsou všechny prvky M nezáporné, jsou to samozřejmě celá čísla. Máme tedy neprázdnou podmnožinu \mathbb{N}_0 , vezmeme její nejmenší prvek r . Evidentně $r \geq 0$ a $a = q_0d + r$ pro nějaké $q_0 \in \mathbb{Z}$. Platí $r < d$? Kdyby ne, pak $a - q_0d \geq d$, proto $r_1 = a - (q_0 + 1)d \geq 0$, tedy $r_1 \in M$ a $r_1 < r$, spor s tím, že r je nejmenší prvek M . Proto $r \geq d$ nemůže nastat.

3) Máme tedy dokázánu existenci pro $d > 0$. Pokud $d < 0$, pak $-d > 0$ a dle první části 1) nebo 2) najdeme q, r tak, aby $a = q(-d) + r$, pak $a = (-q)d + r$.

4) Jednoznačnost: Předpokládejme, že $a = qd + r = q'd + r'$, kde $0 \leq r, r' < |d|$. Pak $qd + r = q'd + r'$, proto $(q - q')d = r - r'$. Jelikož $|r - r'| < |d|$, musí být i $|q - q'| \cdot |d| < |d|$, což znamená $|q - q'| < 1$. Ale $(q - q') \in \mathbb{Z}$, proto $q - q' = 0$ a tedy i $q = q'$. Pak také $r = r'$. \square

Všimněte si, že první důkaz potřeboval k platnosti princip matematické indukce a druhý existenci nejmenšího prvku pro $M \subseteq \mathbb{N}$. My už víme, že tyto dvě věci jsou ekvivalentní, viz Věta 5a.8.

!

Definice.

Nechť $a, d \in \mathbb{Z}$, $d \neq 0$. Číslu r z Věty 6a.6 říkáme **zbytek (remainder)** při dělení a číslem d a značíme jej $r = a \text{ mod } d$, čteno „ r je a modulo d “.

Číslu q říkáme **částečný podíl (quotient)**.

Pro q značení zavádět nemusíme, protože máme následující.

!

Fakt 6a.7.

Nechť $a, d \in \mathbb{Z}$, $d \neq 0$, nechť je q částečný podíl a a d . Pak $q = \lfloor \frac{a}{d} \rfloor$ pro $d > 0$ a $q = \lceil \frac{a}{d} \rceil$ pro $d < 0$.

Důkaz (rutinní, poučný): Máme $a = qd + r$ a $0 \leq r < |d|$. Pak $\frac{a}{d} = q + \frac{r}{d}$, přičemž $q \in \mathbb{Z}$ a $0 \leq \left| \frac{r}{d} \right| < 1$. Čísla q a $\varepsilon = \left| \frac{r}{d} \right|$ pak spolu s Faktem 2b.15 (i) a (ii) dávají žádaný výsledek. □

Dále se budeme pro jednoduchost soustředit na případ $d > 0$. Fakt nabízí možnost, jak se k číslům z Věty 6a.6 dostat, nejprve spočítáme (pro $d > 0$) $q = \lfloor \frac{a}{d} \rfloor$, pak $r = a - qd$. Z hlediska rychlosti výpočtu je klíčové to dělení, kdy je tradiční školní algoritmus zoufale pomalý pro větší čísla. Dá se mu zcela vyhnout metodou z důkazu Věty, kdy prostě necháme od čísla $a > 0$ odečítat číslo d tak dlouho, dokud nedostaneme $0 \leq a - d - \dots - d < d$. U $a < 0$ zase přiřítáme d , dokud nebude $d > a + d + \dots + d \geq 0$. Počet přičtení/odečtení dá q . To ale znamená, že tento algoritmus funguje rychle jen pro a, d podobná, v případě velkého q by těch cyklů bylo příliš mnoho.

Obecně se tedy dělení vyhnout nelze a zajímavý přístup je najít nejprve nějak chytře $\frac{1}{d}$, protože násobení $a \cdot \frac{1}{d}$ už rychle umíme. Populární je třeba tzv. Newton-Raphsonovo dělení, které hledá approximaci $\frac{1}{d}$ použitím Newtonovy metody na nalezení kořene funkce $f(x) = \frac{1}{x} - d$. Vychází iterační schéma $x_{n+1} = x_n(2 - dx_n)$, které je kvadratického rádu, přibližně řečeno s každou iterací zdvojnásobí počet správně nalezených desetinných míst. Jako u mnohých numerických receptů, i zde čihají nebezpečné zákruty a je dobré si to nastudovat, například se ukáže, že dobrá iniciační hodnota je $x_0 = \frac{48}{17} - \frac{32}{17}d$.

To už se ale dostáváme někam daleko, zde nebudeme technické nalezení q a r řešit, prostě oznámíme „nechť $r = a \text{ mod } d$ “ a volbu implementace necháme na uživateli.

! **Příklad 6a.c:** Zkusíme si dělení se zbytkem pro $a = 13$ a $d = 3$. Uhodneme, že $13 = 4 \cdot 3 + 1$ a $1 < 4$, proto je 1 zbytek po dělení čísla 13 číslem 3 čili $1 = 13 \text{ mod } 3$, částečný podíl je $q = 4$. Nebo můžeme počítat $\lfloor \frac{13}{3} \rfloor = 4$ a $13 - 4 \cdot 3 = 1$.

Dělit kladná čísla se zbytkem umíme již ze základní školy, například povědomý výpočet napravo ukazuje, že $4147 \text{ mod } 37 = 3$. U menších čísel to vidíme na první pohled, třeba $48 \text{ mod } 16 = 0$ nebo $48 \text{ mod } 15 = 3$. Jak to bude fungovat, když $a < 0$?

Pro $a = -13$ a $d = 3$ hravě uhodneme, že $-13 = (-5) \cdot 3 + 2$, tedy $-13 \text{ mod } 3 = 2$. Neplatí úvaha $-13 = (-4) \cdot 4 - 1$, tedy zbytek -1 , konec končů $q = \lfloor \frac{-13}{3} \rfloor = \lfloor -4.33\dots \rfloor = -5$.

Obecně se dá dokázat, že jestliže pro $a > 0$ máme $a \text{ mod } d = r > 0$, pak $(-a) \text{ mod } d = d - r$ (viz Cvičení 6a.6). Je tedy možné použít běžné dělení se zbytkem pro kladná čísla a pak modifikovat zbytek.

△

Příklad 6a.d: Dělení se zbytkem nabízí efektivní algoritmus, jak najít zápis čísla n vzhledem k základu b . Ukážeme dvě verze, jednu s koeficienty pro matematickou práci a druhou efektivní pro programování. Nebudeme preferovat nějaký existující jazyk, raději použijeme pseudokód srozumitelný snad každému.

Verze 1.

Iniciace: $q_0 := n$, $k := -1$.

Krok: $k := k + 1$, $q_k = q_{k+1}b + a_k$.

Opakovat dokud nenastane $q_k = 0$.

Pak $n = (a_k \dots a_1 a_0)_b$

nebo Verze 2.

procedure conversion (n, b : positive integer)

$q := n$; $k := -1$;

repeat

$k := k + 1$;

$a_k := q \text{ mod } b$;

$q := \lfloor \frac{q}{b} \rfloor$;

until $q = 0$;

output: $n = (a_k \dots a_1 a_0)_b$;

Z praktického hlediska je lepší přidat registr a ušetřit operace, tedy

$$\begin{aligned} k &:= k + 1; \\ x &:= \lfloor \frac{q}{b} \rfloor; \\ a_k &:= q - xb; \\ q &:= x; \end{aligned}$$

Jako příklad si převedeme 86 do trojkové soustavy.

Iniciace: $q = 86$, $b = 3$, $k = -1$.

Krok 1: $k = 0$, $x = \lfloor \frac{86}{3} \rfloor = 28$, $a_0 = 86 \bmod 3 = 86 - 28 \cdot 3 = 2$, $q = 28$.

Krok 2: $k = 1$, $x = \lfloor \frac{28}{3} \rfloor = 9$, $a_1 = 28 \bmod 3 = 28 - 9 \cdot 3 = 1$, $q = 9$.

Krok 3: $k = 2$, $x = \lfloor \frac{9}{3} \rfloor = 3$, $a_2 = 9 \bmod 3 = 9 - 3 \cdot 3 = 0$, $q = 3$.

Krok 4: $k = 3$, $x = \lfloor \frac{3}{3} \rfloor = 1$, $a_3 = 3 \bmod 3 = 3 - 1 \cdot 3 = 0$, $q = 1$.

Krok 5: $k = 4$, $x = \lfloor \frac{1}{3} \rfloor = 0$, $a_4 = 1 \bmod 3 = 3 - 0 \cdot 3 = 1$, $q = 0$.

Konec, $86 = (10012)_3$.

Zkouška: $1 \cdot 3^4 + 0 \cdot 3^3 + 0 \cdot 3^2 + 1 \cdot 3^1 + 2 \cdot 3^0 = 81 + 3 + 2 = 86$.

Zajímavou otázkou je, jak převádět (kladná) desetinná čísla. Jejich celou část převedeme algoritmem výše, zbývá vymyslet, jak převést část za desetinnou čárkou. K tomu je dobré nejprve pochopit, jak vlastně dotyčný algoritmus pro celá čísla funguje. Představme si číslo $n \in \mathbb{N}$ vyjádřené vzhledem k základu b , tedy $n = (a_k \dots a_1 a_0)_b$. To znamená, že $n = \sum_{k=0}^n a_i b^i$. Chytře si to rozepíšeme:

$$n = a_k b^k + \dots + a_2 b^2 + a_1 b + a_0 = (a_k b^{k-1} + \dots + a_2 b + a_1) \cdot b + a_0.$$

Potřebujeme, aby se nám cifra a_0 z toho davu nějak vydělila, aby získala jiný charakter. To se stane, pokud n vydělíme číslem b . Dostaneme totiž číslo $\frac{n}{b} = a_k b^{k-1} + \dots + a_2 b + a_1 + \frac{a_0}{b}$, přičemž všechny části až na tu poslední jsou celá čísla, zatímco díky $a_0 < b$ je to poslední desetinné, menší než 1. Jinými slovy, a_0 je přesně zbytek po dělení n číslem b .

Ted' totéž zopakujeme s $a_k b^{k-1} + \dots + a_2 b + a_1 = (a_k b^{k-2} + \dots + a_2) b + a_1$ a dostaneme a_1 atd.

Nyní se podívejme na kladné číslo c menší než 1. Takové číslo se v b -soustavě napíše jako

$$c = \sum_{k=1}^{\infty} a_{-k} b^{-k} = a_{-1} \frac{1}{b} + a_{-2} \frac{1}{b^2} + a_{-3} \frac{1}{b^3} + \dots$$

(rozvoj může a nemusí být nekonečný). Potřebujeme vypreparovat cifry a_{-k} pomocí počítání s celými čísly. Dokážeme nějak zařídit, aby se ta první část s a_{-1} nějak vydělila svou povahou od ostatních? Ano, stačí c vynásobit číslem b . Dostaneme pak $cb = a_{-1} + a_{-2} \frac{1}{b} + a_{-3} \frac{1}{b^2} + \dots$, přičemž první číslo a_{-1} je celé a ostatní části jsou menší než 1, dokonce i po sečtení (to je třeba trochu prozkoumat matematicky, není to tak těžké). Cifru a_{-1} tedy vidíme jako celou část čísla cb , zatímco desetinná část nám dá ten zbytek.

Ten lze zapsat jako $cb - a_{-1} = a_{-2} \frac{1}{b} + a_{-3} \frac{1}{b^2} + \dots$ a máme ted' šanci vyseparovat cifru a_{-2} tím, že toto číslo zase vynásobíme číslem b . Tak pokračujeme buď donekonečna, nebo dokud nedostaneme jako zbytek nulu.

procedure conversion (c : real $\in (0, 1)$, b : positive integer)

$x = c$, $k := 0$;

repeat

$k := k - 1$;

$a_k := \lfloor xb \rfloor$;

$x := xb - a_k$;

until $x = 0$;

output: $c = (0.a_{-1}a_{-2}a_{-3} \dots)_b$;

△

! Příklad 6a.e: Zde si ukážeme několik praktických aplikací modula.

1) Knižní kód ISBN je navržen tak, aby částečně fungoval jako opravný kód, přesněji řečeno tak, abychom snadno a s vysokou pravděpodobností poznali, že nám při jeho předávání vznikla chyba. Jeho starší verze měla 10 cifer. Prvních 9 cifer identifikuje jazyk, nakladatele a číslo knihy dle katalogu nakladatele. Jako poslední číslo se vždy dává zbytek po dělení počátečního devítimístného čísla jedenácti, je samozřejmě třeba vyřešit problém zbytku 10, to se pak dává znak X . Tvrdíme, že výsledné číslo je pak již vždy dělitelné jedenácti.

Označme $n = 10a + r$, přičemž a je to počáteční devítimístné číslo a $r = a \bmod 11$. Pak $a = 11k + r$, proto $n = 110k + 10r + r = 11(10k + r)$, tedy číslo dělitelné jedenácti. Jiný důkaz, možná rychlejší (viz příští kapitola): Podle definice máme $a \equiv r \pmod{11}$, také $10 \equiv (-1) \pmod{11}$, proto $10a + r \equiv (-1)r + r = 0 \pmod{11}$.

To znamená, že když nám někdo dá ISBN číslo, my jej zkusíme vydělit 11 a nevyjde to, tak už víme, že se někde stala chyba. Pokud by to vyšlo, tak je buď číslo dobré, nebo se pokazila víc než jedna cifra a zrovna tak šikovně, že to dělitelnost nezkazilo.

Mimochodem, proč jsme použili zrovna jedenáctku? Pokud použijeme menší číslo, pak chybu v jedné cifře nemusí odhalit. Například pokud bychom použili dělitelnost desíti, tak nepoznáme správné číslo 20 od chybného čísla 30. Podobně když se rozhodneme testovat dělitelnost čtyřmi a správné číslo je 36, pak chyba v cifře v čísle 32 není dělitelností poznatelná.

Číslo 11 je tedy nejmenší (tudíž nejpraktičtější), které v testu dělitelnosti umí odhalit chybu v jedné číslici (rozmyslete si, že záměna jedné číslice v čísle nutně vede ke změně zbytku po dělení jedenácti).

2) Hashovací funkce. Představte si, že chceme ukládat data o lidech, kteří jsou kódováni rodnými čísly, ale máme jen n paměťových adres. Hledáme funkci h , která nám řekne, že data člověka s rodným číslem a se mají dát na adresu $h(a)$. Jedním z možných řešení je použít funkci $h(a) = a \bmod n$.

Výhody: h je na, rychle se počítá.

Nevýhoda: h není prostá, vznikají tzv. kolize. Jsou nutné strategie, co pak (např. metoda „první následující volné místo“).

3) Když už mluvíme o rodných číslech: Rodná čísla se dělají následovně: První dvoučíslí je rok narození, druhé měsíc narození zvýšený u žen o 50, třetí dvoučíslí je den narození, další tři pak identifikují oblast a pořadové číslo dítěte v rámci této oblasti. Jako poslední cifra rodného čísla se dá buď zbytek po dělení počátečního devítimístného čísla jedenácti, pokud vyjde menší než 10, nebo 0, pokud ten zbytek vyjde 10.

Co to znamená? že každé rodné číslo nekončící nulou musí být dělitelné jedenácti, u čísel nulou končících už to ale nemusí být pravda. Moc jich nebývá: statisticky každé jedenácté, ale zejména v posledních desetiletích se takovým číslům při přidělování snaží vyhýbat, takže jich je výrazně méně než jedenáctina. Díky tomu přezívá fáma, že se dělitelností jedenáctkou dají kontrolovat správná rodná čísla.

4) Od rodných čísel přejdeme k náhodným. Pro různé simulace a samozřejmě také hry je potřeba mít zdroj náhodných čísel. To ale není tak snadné zařídit, protože tento zdroj musí být algoritmický (počítač má naprogramovanou metodu, jak to dělat). Nevznikají tak čísla náhodná, ale pseudonáhodná, jejich zdroji se říká generátor.

Když už se tedy smíříme s tím, že máme generátor jen pseudonáhodných čísel, tak bychom alespoň chtěli, aby ten algoritmus z dlouhodobého hlediska nezvýhodňoval žádná čísla ani nevykazoval pravidelnosti. To je velice náročný úkol, u méně náročných aplikací (třeba her) se dá od striktních nároků částečně ustoupit a pak přichází vhod tzv. **lineární kongruentní generátor**.

Funguje to následovně. Zvolíme modulus $n \in \mathbb{N}$. Pak zvolíme multiplikátor $a \in \mathbb{N}$ splňující $2 \leq a < n$ a posun $c \in \mathbb{N}$ splňující $0 \leq c < n$. Jako náhodná čísla používáme posloupnost $x_{k+1} = (a \cdot x_k + c) \bmod n$. Je nutno ji nastartovat pomocí zdrojové hodnoty $x_0 \in \mathbb{N}$. Vychází pak z toho čísla z rozmezí 0 až $n - 1$, která se tváří náhodně (ale nejsou, protože se opakují, nejdéle možný řetězec má délku n , ale může se zacyklit dříve, zabráníme tomu tak, že zvolíme jako n prvočíslo).

Například pokud zvolíme $n = 6$, $a = 4$, $c = 1$, dostáváme vzorec $x_{k+1} = (4x_k + 1) \bmod 6$. Když se rozhodneme začít dvojkou, dostaneme posloupnost $2, 3, 1, 5, 3, 1, 5, 3, \dots$, délka cyklu je 3.

Když si zvolíme $n = 9$, $a = 7$, $c = 4$, pak ze vzorce $x_{k+1} = (7x_k + 4) \bmod 9$ už vyjde řetězec délky 9.

Často chceme čísla z intervalu $(0, 1)$, pak bereme x_k/n . Při volbě hodně velkého n a a to vychází docela zajímavě.

Často se volí $c = 0$, tzv. čistě multiplikativní generátor, pak nechceme $x_k = 0$ a je snaha volit n, a tak, aby vznikl opravdu řetězec délky $n - 1$. Typická volba je třeba $n = 2^{31} - 1$ a $a = 7^5 = 16807$, kdy pak opravdu dostaneme $2^{31} - 2 = 4294967294$ hodnot. To už je pro praktické účely docela dost.

△

Následující fakt by měl být čtenáři po menším zamýšlení jasný.

Fakt 6a.8.

Nechť $a, b \in \mathbb{Z}$, $a \neq 0$. Pak $a | b$ právě tehdy, když $b \bmod |a| = 0$, tedy zbytek po dělení b číslem $|a|$ je 0.

Důkaz necháme jako cvičení 6a.5. Mimochodem, absolutní hodnota u a v textu tvrzení není nutná, protože jsme ve větě o dělení se zbytkem pracovali i se zápornými děliteli a není s nimi problém, ale v praxi stejně znaménko ignorujeme, takže jsme zvolili praktický pohled na věc.

Ted' zavedeme nové užitečné pojmy.

!

Definice.

Nechť $a, b \in \mathbb{Z}$.

Číslo $d \in \mathbb{N}$ je **společný dělitel (common divisor)** čísel a, b , jestliže $d | a$ a $d | b$.

Číslo $d \in \mathbb{N}$ je **společný násobek (common multiple)** čísel a, b , jestliže $a | d$ a $b | d$.

Například číslo 1 je určitě společným dělitelem čísel 40 a 60, zatímco $40 \cdot 60 = 2400$ je určitě jejich společným násobkem. Čtenář ale asi tuší, že nás budou zajímat trochu méně triviální odpovědi, třeba 20 jako společný dělitel a 120 jako společný násobek. Jak vlastně množiny všech společných násobků a společných dělitelů vypadají?

Pro každá dvě čísla $a, b \in \mathbb{Z}$ je 1 společným dělitelem, tudíž množina společných dělitelů je neprázdná. Pokud je jedno z čísel a, b nenulové, například a , pak každý společný dělitel d musí splňovat $d \leq |a|$, viz Věta 6a.2 (v). Proto je pak množina společných dělitelů konečná a tudíž má v \mathbb{N} největší a nejmenší prvek.

Kdyby ale bylo $a = b = 0$, tak je společným dělitelem libovolné $d \in \mathbb{N}$, takže vlastně množina společných dělitelů je \mathbb{N} . To je nepříjemné a budeme se s tímto případem opakovaně nuceni vypořádávat zvláště.

U společných násobků je rovněž třeba hlídat nuly, ale trochu jinak. Pokud jsou obě a, b nenulové, tak je $a \cdot b$ společným násobkem. Množina společných násobků je shora neomezená, protože máme-li společný násobek d , pak je i kd společným násobkem pro libovolné $k \in \mathbb{N}$. Množina společných násobků tedy nemá největší prvek, ale jako neprázdná podmnožina \mathbb{N} má určitě prvek nejmenší (všimněte si, že i pro záporná a, b bereme v definici jen kladná čísla jako jejich společné dělíteli či násobky).

Pokud je alespoň jedno z čísel a, b nulové, tak máme problém, protože jediným násobkem nuly je zase nula, což definice společného násobku nepřipouští. Množina společných násobků je v tomto případě prázdná.

! Definice.

Nechť $a, b \in \mathbb{Z}$.

Definujeme jejich **největší společný dělitel (greatest common divisor)**, značeno $\gcd(a, b)$, jako největší prvek množiny jejich společných dělitelů, pokud je alespoň jedno z a, b nenulové.

Jinak definujeme $\gcd(0, 0) = 0$.

Řekneme, že čísla $a, b \in \mathbb{Z}$ jsou **nesoudělná**, jestliže $\gcd(a, b) = 1$.

Definujeme jejich **nejmenší společný násobek (least common multiple)**, značeno $\text{lcm}(a, b)$, jako nejmenší prvek množiny jejich společných násobků, pokud jsou obě a, b nenulové.

Jinak definujeme $\text{lcm}(a, 0) = \text{lcm}(0, b) = 0$.

Tyto pojmy se dají zobecnit na více čísel, viz cvičení 6a.19.

Volby gcd a lcm pro výjimečné případy jsou vedeny snahou zachovat užitečné vlastnosti. V obou případech je zachován smysl, číslo 0 opravdu dělí 0 a je násobkem 0. U největšího společného dělitele je takových kandidátů nekonečně mnoho (to už jsme viděli), tam byla naše volba rozumná například proto, že teď máme obecně $\gcd(a, b) \leq |a|$ a $\gcd(a, b) \leq |b|$ pro všechna a, b včetně nulových, což je jistě příjemné. U společného násobku to bohužel nevyšlo, jediný násobek čísla 0 je zase 0, tudíž jsme v definici neměli na výběr a musíme se smířit s tím, že $|a| \leq \text{lcm}(a, b)$ a $|b| \leq \text{lcm}(a, b)$ platí jen pro nenulová a, b (musíme být proto opatrní).

Je to další ponouknutí, abychom pracovali jen s čísly $a, b \in \mathbb{N}$, často to bude možné a máme po problémech, ale ne vždy to jde.

Protože $\gcd(a, b)$ dělí a i b , tak pro nenulová a, b máme $\frac{a}{\gcd(a, b)} \in \mathbb{Z}$ a $\frac{b}{\gcd(a, b)} \in \mathbb{Z}$. To je zjevné, ale budeme to opakovaně používat, tak na to upozorňujeme. Vtělíme to do tvrzení s ještě jedním pozorováním.

! Fakt 6a.9.

Nechť $a, b \in \mathbb{Z}$, $a \neq b$. Pak jsou $\frac{a}{\gcd(a, b)}$ a $\frac{b}{\gcd(a, b)}$ nesoudělná celá čísla.

Důkaz (rutinní): Předpokládejme, že číslo $d \in \mathbb{N}$ je společný dělitel $\frac{a}{\gcd(a, b)}$ a $\frac{b}{\gcd(a, b)}$. To znamená, že pro nějaká $k, l \in \mathbb{Z}$ máme $\frac{a}{\gcd(a, b)} = kd$ a $\frac{b}{\gcd(a, b)} = ld$. Pak $a = k[d\gcd(a, b)]$ a $b = l[d\gcd(a, b)]$, čili $d\gcd(a, b)$ je společný dělitel a, b . Protože $\gcd(a, b)$ je mezi společnými děliteli největší, musí být $d \leq 1$, což pro $d \in \mathbb{N}$ znamená nutně $d = 1$.

Takže jediný společný dělitel těch dvou čísel je 1, jsou tedy nesoudělná. □

Intuitivně je to jasné, když vydělíme obě čísla tím, co mají společné, tak už v nich nic společného zbýt nemůže.

Další zkoumání těchto pojmu začneme pomocným tvrzením, které čtenář jistě zná z vlastní zkušenosti. Pokud je nějaké číslo d dělitelné číslu a, b , tak ještě to d nemusí být dělitelné číslem $a \cdot b$, ale určitě půjde vydělit číslem $\text{lcm}(a, b)$.

Lemma 6a.10.

Nechť $a, b \in \mathbb{Z}$. Jestliže je d společný násobek a, b , pak $\text{lcm}(a, b)$ dělí d .

Důkaz (poučný): Jestliže $a = 0$ nebo $b = 0$, pak společné násobky neexistují a tvrzení platí automaticky. Předpokládejme tedy dále, že $a, b \neq 0$.

Protože je $\text{lcm}(a, b)$ nejmenší společný násobek, musí platit $d \geq \text{lcm}(a, b)$. Nechť podle věty o dělení $d = q \text{lcm}(a, b) + r$. Protože a dělí $\text{lcm}(a, b)$, tak dělí i $q \text{lcm}(a, b)$, dělí rovněž d , proto podle Důsledku 6a.4 (ii) musí a také dělit r . Stejně ukážeme, že b dělí r , navíc $r \in \mathbb{N}_0$, takže buď $r = 0$ nebo je r společný násobek a, b . To druhé je ale nemožné, protože $\text{lcm}(a, b)$ je nejmenší společný násobek a $r < \text{lcm}(a, b)$. Takže $r = 0$ a $d = q \text{lcm}(a, b)$ pro nějaké $q \in \mathbb{Z}$. \square

Dá se to číst i jinak, ukazuje to, že množina společných násobků dvou čísel má zajímavou strukturu, všechny pocházejí od jednoho základního, nemůže se objevit nějaký úplně jiný. Lemma to dokázalo pro čísla kladná, platí to i obecně.

Fakt 6a.11.

Nechť $a, b, n \in \mathbb{Z}$. Jestliže $a | n$ a $b | n$, pak $\text{lcm}(a, b) | n$.

Důkaz (rutinní, možná poučný): Jestliže je $a = 0$ či $b = 0$, tak nutně i $n = 0$ a $\text{lcm}(a, b) = 0$, tvrzení pak platí. Zbývá probrat případ, kdy $a, b \neq 0$.

Jestliže $a | n$ a $b | n$, pak také $a | |n|$ a $b | |n|$ (viz cvičení 6a.4), takže $|n|$ je společným násobkem a, b . Podle Lemma 6a.10 tedy musí platit $\text{lcm}(a, b) | |n|$, proto i $\text{lcm}(a, b) | n$. \square

Ukážeme ještě jedno zajímavé čtení tohoto faktu. Pro $a, b \neq 0$ je $\text{lcm}(a, b)$ definován jako nejmenší prvek množiny společných násobků, přičemž „nejmenší“ je ve smyslu nerovnosti. Právě jsme se dozvěděli, že $\text{lcm}(a, b)$ je také nejmenším prvkem množiny společných násobků vzhledem k relaci dělitelnosti. Obdobné tvrzení platí i pro $\text{gcd}(a, b)$, ale tam je důkaz těžší a postupně se k němu propracujeme.

Naším cílem teď bude odvodit praktické postupy k nalezení $\text{gcd}(a, b)$ a $\text{lcm}(a, b)$. Nejprve si různými pozorováními zredukujeme obecnou situaci na případ $0 < b < a$. Začneme tím, že stačí umět hledat oba pojmy pro čísla z \mathbb{N}_0 .

Fakt 6a.12.

Nechť $a, b \in \mathbb{Z}$. Pak $\text{gcd}(a, b) = \text{gcd}(|a|, |b|)$ a $\text{lcm}(a, b) = \text{lcm}(|a|, |b|)$.

Důkaz (rutinní): Případy s $a = 0$ či $b = 0$ se snadno rozmyslí, zaměříme se na případ $a, b \neq 0$.

Společní dělitelé a, b jsou kladná čísla d splňující $d | a$ a $d | b$, což jsou podle Věty 6a.2 (iv) přesně čísla splňující $d | |a|$ a $d | |b|$. Množina společných dělitelů a, b je tedy stejná jako množina společných dělitelů $|a|, |b|$, tudíž se musejí rovnat i jejich největší prvky.

Důkaz pro $\text{lcm}(a, b)$ je obdobný. \square

Hodnoty $\text{gcd}(0, 0) = 1$ a $\text{lcm}(0, 0) = 0$ známe z definice, snadno nalezneme i hodnoty pro další specifické případy.

Fakt 6a.13.

Nechť $a \in \mathbb{N}$. Pak $\text{gcd}(a, 0) = a$, $\text{lcm}(a, 0) = 0$ a $\text{gcd}(a, a) = \text{lcm}(a, a) = a$.

Druhé tvrzení je přímo definice, důkaz ostatních je snadný a necháme jej jako cvičení 6a.10. Teď si dále zjednodušíme situaci.

Věta 6a.14.

Nechť $a, b \in \mathbb{Z}$. Pak $\text{lcm}(a, b) \cdot \text{gcd}(a, b) = |a| \cdot |b|$.

Důkaz (poučný): 1) Nejprve budeme předpokládat, že $a, b \in \mathbb{N}$.

Označme $z = \frac{ab}{\text{gcd}(a, b)}$, chceme ukázat, že je to $\text{lcm}(a, b)$. Máme $z = \frac{a}{\text{gcd}(a, b)}b$ a $\frac{a}{\text{gcd}(a, b)} \in \mathbb{Z}$, tedy z je násobek b , a symetricky také $z = \frac{b}{\text{gcd}(a, b)}a$ a z je násobek a . Podle Lemma 6a.10 tedy musí platit $z = q \text{lcm}(a, b)$ pro nějaké $q \in \mathbb{N}$. Potřebujeme ukázat, že $q = 1$.

Všimněme si, že $\text{lcm}(a, b) = \frac{z}{q} = \frac{ab}{q \gcd(a, b)}$. Když využijeme toho, že a i b dělí $\text{lcm}(a, b)$, dostáváme $\frac{a}{q \gcd(a, b)} = \frac{\text{lcm}(a, b)}{b} \in \mathbb{Z}$, tedy $q \gcd(a, b)$ dělí a , a symetricky $\frac{b}{q \gcd(a, b)} = \frac{\text{lcm}(a, b)}{a} \in \mathbb{Z}$, tedy $q \gcd(a, b)$ dělí b . To znamená, že $q \gcd(a, b)$ je společný dělitel a, b , tudíž z definice \gcd musí platit $q \gcd(a, b) \leq \gcd(a, b)$. Zároveň ale $q \in \mathbb{N}$, což znamená, že nutně $q = 1$.

2) Jsou-li a, b nenulové, ale některé z nich je záporné (či obě), pak výsledek vyplývá z Faktu 6a.12.

3) Případy s $a = 0$ nebo $b = 0$ se snadno ověří. □

! Tato věta nám dává vzorec pro výpočet nejmenšího společného násobku, v počítání je ale samozřejmě lepší namísto $\text{lcm}(a, b) = \frac{ab}{\gcd(a, b)}$ používat $\text{lcm}(a, b) = \frac{a}{\gcd(a, b)} b$ či $\text{lcm}(a, b) = \frac{b}{\gcd(a, b)} a$, protože tento postup nevyžaduje ukládání velkého čísla ab .

Zůstává otázka, jak najít $\gcd(a, b)$, podle Faktů 6a.12 a 6a.13 už víme, že to musíme hlavně vyřešit pro případ $a, b > 0, a \neq b$. Pro malá čísla se to často dělá rozkladem na prvočísla, jenže ten jsme ještě neprobrali a hlavně to je ukrutně výpočetně náročná úloha, tudíž neperspektivní. Existuje lepší způsob, populární a mocný Euklidův algoritmus. Je založen na následujícím pozorování.

Lemma 6a.15.

Nechť $a, b \in \mathbb{N}$, nechť $r = a \bmod b$. Pak platí následující:

- (i) $d \in \mathbb{N}$ je společný dělitel a, b právě tehdy, když je to společný dělitel b, r .
- (ii) $\gcd(a, b) = \gcd(b, r)$.

Důkaz (poučný): Nechť $r = a \bmod b$. (i): \implies : Je-li d společný dělitel a a b , pak dělí a i qb , tedy podle Důsledku 6a.4 (ii) musí dělit také r a je to společný dělitel b, r . \implies : Důkaz je obdobný.

(ii): Podle (i) se množina společných dělitelů a, b rovná množině společných dělitelů b, r , proto se musí rovnat i jejich největší prvky. □

Toto lemma je klíčem k algoritmu. Namísto hledání \gcd pro dvojici $a > b$ nám stačí hledat \gcd pro odpovídající čísla $b > r$, nic nám ale nebrání aplikovat naše lemma znova, najít $r' = b \bmod r$ a namísto dvojice $b > r$ hledat \gcd pro $r > r'$. Takto můžeme pokračovat.

Protože se v každém kroku to menší číslo dále zmenší, ale nikdy není záporné, tak tento postup nemůže trvat do nekonečna, ale musí se zarazit. Kdy se tak stane? Kdy vlastně nemůžeme Lemma použít? Když některé z čísel (či obě) opustí \mathbb{N} , což se nejdříve stane tomu menšímu, tedy zbytku po dělení. Zbytek po dělení opustí \mathbb{N} tehdy, je-li nulový, takže opakujeme Lemma tak dlouho, dokud nenařazíme na nulový zbytek, $\gcd(x, 0)$ pak hravě určíme podle Faktu 6a.13. Tím dostáváme algoritmus k nalezení $\gcd(a, b)$ pro $a > b$. Ukážeme jej na příkladě.

Příklad 6a.f: Chceme najít $\gcd(408, 108)$.

Máme $408 = 3 \cdot 108 + 84$, proto $\gcd(408, 108) = \gcd(108, 84)$.

Máme $108 = 1 \cdot 84 + 24$, proto $\gcd(408, 108) = \gcd(108, 84) = \gcd(84, 24)$.

Máme $84 = 3 \cdot 24 + 12$, proto $\gcd(408, 108) = \gcd(108, 84) = \gcd(84, 24) = \gcd(24, 12)$.

Máme $24 = 2 \cdot 12 + 0$, proto $\gcd(408, 108) = \gcd(108, 84) = \gcd(84, 24) = \gcd(24, 12) = \gcd(12, 0) = 12$.

Jako obvykle jsme neřešili, kde se ty rozklady berou, já jsem si je dělal tužkou přes $q = \lfloor \frac{x}{y} \rfloor$. △

Algoritmus ukážeme ve dvou podobách. Jedna si pamatuje, co se kdy dělo. Ta bude výhodná při důkazu, že algoritmus dělá to, co má. Druhá verze se nestará o minulost, což je samozřejmě verze, kterou bychom použili v praxi. Jak už jsme zmínili po Faktu 6a.7, budeme předpokládat nějakou implementaci procesu nalezení q, r .

Algoritmus 6a.16. Euklidův algoritmus pro nalezení $\gcd(a, b)$ pro $a > b \in \mathbb{N}$.

Verze 1.

Iniciace: $r_0 := a, r_1 := b, k := 0$.

Krok: $k := k + 1, r_{k-1} = q_k \cdot r_k + r_{k+1}$

Opakovat dokud nenastane $r_{k+1} = 0$.

Pak $\gcd(a, b) = r_k$.

nebo Verze 2.

procedure $\gcd(a, b: \text{integer}, a > b > 0)$

repeat

$r := a \bmod b;$

$a := b; b := r;$

until $b = 0$;

output: a ;

△

Teď dokážeme, že algoritmus dělá, co má.

1) Algoritmus skončí: Variantem je r_k , to splňuje $r_1 > r_2 > r_3 > \dots$ a zároveň $r_k \in \mathbb{N}_0$, proto musí dojít k terminační podmínce $r_k = 0$.

2) Algoritmus má na výstupu $\gcd(a, b)$: Pomocí indukce dokážeme následující $V(k)$: Jestliže $r_k > 0$, pak $\gcd(r_k, r_{k+1}) = \gcd(a, b)$.

(0) Pro $k = 0$ to platí, neboť $r_0 = a$ a $r_1 = b$.

(1) Nechť $k \in \mathbb{N}_0$ a předpokládejme, že $\gcd(r_k, r_{k+1}) = \gcd(a, b)$ a $r_k > 0$.

Nechť také $r_{k+1} > 0$. Pak $r_k = qr_{k+1} + r_{k+2}$ podle věty o dělení a podle Lemma 6a.15 máme $\gcd(r_{k+1}, r_{k+2}) = \gcd(r_k, r_{k+1})$, spolu s indukčním předpokladem pak $\gcd(r_{k+1}, r_{k+2}) = \gcd(a, b)$, tedy $V(k+1)$ platí.

$V(k)$ je dokázáno. V okamžiku terminace máme $r_k > 0$ a $r_{k+1} = 0$, tedy podle $V(k)$ a Faktu 6a.13 je $\gcd(a, b) = \gcd(r_k, r_{k+1}) = \gcd(r_k, 0) = r_k$.

! 6a.17 Ruční výpočet.

Při ručním počítání zachycujeme stavy registrů tabulkou, existuje několik verzí. Začneme verzí podrobnou, která přesně ilustruje běh algoritmu. Každý jeho krok dělá dvě věci, nejprve spočítá r a pak si připraví půdu pro nové kolo posunem dat v registrech. Totéž uděláme v tabulce, nejprve si spočítáme r a zapíšeme jej do nového řádku jako novou hodnotu b , pak doplníme na nový řádek starou hodnotou b jako nové a a tím se připravíme na další kolo. Ukážeme si to pro nás předchozí příklad.

Najdeme $\gcd(408, 108)$ pomocí Euklidova algoritmu.

$\begin{array}{ c c } \hline a & b \\ \hline 408 & 108 \\ \hline \end{array} \quad 408 \text{ mod } 108 = 84 \implies$	$\begin{array}{ c c } \hline a & b \\ \hline 408 & 108 \\ \hline 84 & \\ \hline \end{array} \quad \begin{array}{ c c } \hline a & b \\ \hline 408 & 108 \\ \hline 108 & 84 \\ \hline \end{array}$
$108 \text{ mod } 84 = 24 \implies$	$\begin{array}{ c c } \hline a & b \\ \hline 408 & 108 \\ \hline 108 & 84 \\ \hline \end{array} \quad \begin{array}{ c c } \hline a & b \\ \hline 408 & 108 \\ \hline 108 & 84 \\ \hline 84 & 24 \\ \hline \end{array} \quad 84 \text{ mod } 24 = 12 \implies$
$24 \text{ mod } 12 = 0 \implies$	$\begin{array}{ c c } \hline a & b \\ \hline 408 & 108 \\ \hline 108 & 84 \\ \hline 84 & 24 \\ \hline 24 & 12 \\ \hline \end{array} \quad \begin{array}{ c c } \hline a & b \\ \hline 408 & 108 \\ \hline 108 & 84 \\ \hline 84 & 24 \\ \hline 24 & 12 \\ \hline 12 & 0 \\ \hline \end{array} \quad \begin{array}{ c c } \hline a & b \\ \hline 408 & 108 \\ \hline 108 & 84 \\ \hline 84 & 24 \\ \hline 24 & 12 \\ \hline 12 & 0 \\ \hline \end{array}$

Když počítáme takto ručně, tak se nemusíme přesně držet algoritmu a lze vynechat tu poslední tabulku. Jakmile se nám objeví na řádku nula, tak si přečteme hodnotu $\gcd(a, b)$ v políčku nad tou nulou.

Zásadní nedostatek tohoto algoritmu je, že se v něm každé číslo objevuje dvakrát, což je zbytečné. Mnohem jednodušší je použít jen jeden sloupec a pracovat vždy se dvěma posledními čísly. Opět si najdeme $\gcd(408, 108)$.

$\begin{array}{ c c } \hline a, b \\ \hline 408 \\ \hline 108 \\ \hline \end{array} \implies$	$\begin{array}{ c c } \hline a, b \\ \hline 408 \\ \hline 108 \\ \hline 84 \\ \hline \end{array} \implies$	$\begin{array}{ c c } \hline a, b \\ \hline 408 \\ \hline 108 \\ \hline 84 \\ \hline 24 \\ \hline \end{array} \implies$	$\begin{array}{ c c } \hline a, b \\ \hline 408 \\ \hline 108 \\ \hline 84 \\ \hline 24 \\ \hline 12 \\ \hline \end{array} \implies$	$\begin{array}{ c c } \hline a, b \\ \hline 408 \\ \hline 108 \\ \hline 84 \\ \hline 24 \\ \hline 12 \\ \hline 0 \\ \hline \end{array}$
---	--	---	--	---

Samozřejmě není důvod psát pod sebe, nejjednodušší je rovnou psát čísla za sebe, začneme s čísly 408 a 108, z nich odvodíme další.

$$\begin{array}{ccccccc}
 & 408 & & 108 & & 84 & \\
 & \downarrow & & \downarrow & & \downarrow & \\
 408 \mod 108 & = & 84 & & 24 & & 0
 \end{array}$$

Takto pokračujeme:

$$408 \quad 108 \quad 84 \quad 24 \quad 12 \quad 0$$

Tato metoda je evidentně nejpohodlnější, ale my si tento algoritmus brzy dále rozšíříme a pak mnozí dávají přednost tomu předchozímu svislému zápisu.

Pro úplnost si ještě dopočítáme $\text{lcm}(408, 108) = \frac{408 \cdot 108}{\text{gcd}(408, 108)} = 408 \frac{108}{12} = 408 \cdot 9 = 3672$.

△

Jak je tento algoritmus rychlý?

Věta 6a.18. (Lamého věta)

Nechť $a > b \in \mathbb{N}$. Pak Euklidův algoritmus pro $\text{gcd}(a, b)$ vyžaduje nanejvýš tolik kroků, kolik je pětkrát počet cifer v b .

Důkaz (drsný, poučný): Nejprve si všimněme následujícího. V každém kroku počítáme $r_{k-1} = q_k \cdot r_k + r$, kde $r < r_k$. Kdyby bylo $q_k = 0$, tak dostáváme $r = r_{k-1}$, což by znamenalo $r_{k-1} < r_k$, jenže v našem algoritmu je to přesně naopak, my máme $r_{k-1} > r_k$. Proto musí být vždy $q_k \geq 1$.

Je to vidět i jinak. Protože je $r_{k-1} > r_k$, tak při hledání zbytku po dělení musíme r_k alespoň jednou odečíst od r_{k-1} .

Každopádně máme $r_{k-1} = q_k \cdot r_k + r_{k+1} \geq r_k + r_{k+1}$.

Jak to vypadá pro případ, kdy $r_{k+1} = 0$, tedy poslední krok algoritmu? Pak $r_{k-1} = q_k \cdot r_k$. Kdyby bylo $q_k = 1$, pak dostáváme $r_{k-1} = r_k$, což zase není možné. Je tedy $q_k \geq 2$ a máme pro poslední nenulové r_k nerovnost $r_{k-1} \geq 2r_k$.

Uvažujme teď běh algoritmu pro konkrétní $a > b$ a předpokládejme, že skončil pro jisté $n \in \mathbb{N}$ s $r_{n+1} = 0$ a $r_n = \text{gcd}(a, b)$. Tvrdíme následující:

Pro $k = 1, \dots, n$ platí $V(k)$: $r_{n-k+1} \geq F_{k+1}$, kde F_m jsou Fibonacciho čísla, viz příklad 9a.c. Dokážeme to indukcí na k (modifikovaným principem s návratem o dva kroky).

(0) Pro $k = 1$ rovnost říká $r_n \geq F_2$, tedy $r_n \geq 1$, což je pravda, r_n je poslední nenulový zbytek.

Pro $k = 2$ tvrzení říká, že $r_{n-1} \geq F_3$, tedy $r_{n-1} \geq 2$. To je pravda, protože pro poslední krok jsme odvodili $r_{n-1} \geq 2r_n \geq 2$.

(1) Předpokládejme, že tvrzení platí pro $1, \dots, k$, kde $k \geq 2$. Chceme odhad pro $r_{n-(k+1)+1} = r_{n-k}$. Použijeme nerovnici odvozenou na začátku důkazu, dává $r_{n-k} \geq r_{n-k+1} + r_{n-k+2} = r_{n-k+1} + r_{n-(k-1)+1}$. Pomocí indukčního předpokladu a vlastnosti Fibonacciho čísel pak dostaneme

$$r_{n-(k+1)+1} = r_{n-k} \geq F_{k+1} + F_{(k-1)+1} = F_{k+1} + F_k = F_{k+2} = F_{(k+1)+1}.$$

$V(k+1)$ a tím i krok (1) jsou dokázány. Indukce potvrdila, že $V(k)$ platí pro $k = 1, \dots, n$.

Z $V(n)$ máme $r_1 \geq F_{n+1}$, tedy počet kroků programu n splňuje nerovnost $F_{n+1} \leq b$. Co z toho plyne pro n ?

Z výsledků příkladu 9a.c máme následující. Když označíme $\alpha = \frac{1+\sqrt{5}}{2}$, pak $F_n \geq \alpha^{n-1}$. Pro nás to znamená, že $\alpha^n \leq b$, proto $n \log_{10}(\alpha) \leq \log_{10}(b)$. Kalkuliho přístroj potvrdí, že $\log_{10}(\alpha) > \frac{1}{5}$, z čehož dostaneme $n \leq 5 \log_{10}(b)$. Důkaz je hotov. □

Říkáme tím vlastně, že výpočetní náročnost algoritmu je řádově $\log(b)$, jenže jsme nezapočítali nároky na výpočet rozkladu $a = qb + r$, zvolená implementace výslednou náročnost dost ovlivní. Používané algoritmy mají obdobnou náročnost (závisí na délce dělitele), takže obecně lze říci, že celková výpočetní náročnost Euklidova algoritmu je řádově $[\log(b)]^2$. Z hlediska computer science je jednodušší hovořit o bitech zápisu čísla, můžeme například říct, že při práci s n -bitovými čísly je potřeba zhruba n^2 základních operací. V kapitole 9b si o náročnosti povíme více, pak budeme říkat, že náročnost Euklidova algoritmu je $O(n^2)$.

Mimochodem, zajímavé je, že efektivní implementace dělení celých čísel dokážou průměrnou náročnost hledání zbytku, takže postup přes $q = \lfloor \frac{a}{b} \rfloor$ zase není tak špatný. Dá se ale ukázat, že při výpočtu Euklidova algoritmu vznikají s vysokou pravděpodobností malá q (pravděpodobnosti pro $q = 1, 2, 3, 4$ jsou po řadě 41.5%, 17.0%, 9.3%, 5.9%), jinými slovy v polovině případů lze čekat $q = 1$ či $q = 2$ a to už se hledání q a r odečítáním opravdu vyplatí.

Zajímavou alternativou je použití různých fint, například je možné použít tyto rovnosti:

- $\text{gcd}(a, b) = 2 \text{gcd}(a/2, b/2)$, jsou-li obě sudá,
- $\text{gcd}(a, b) = \text{gcd}(a/2, b)$, je-li a sudé a b liché,
- $\text{gcd}(a, b) = \text{gcd}(a - b, b)$, jsou-li obě lichá.

Takže nejprve opakováním dělením dvěma (což je relativně levná operace) dosáhneme situace s jedním či dvěma lichými čísly, pak aplikujeme opakováně druhý či třetí vzorec, dokud nedostaneme dvě sudá čísla, to zase redukujeme dělením dvěma a pořád dokola. V zásadě lze ale říct, že lépe než k n^2 se stejně nedostaneme.

Tak jsme se naučili hledat efektivně největší společný dělitel (tedy i nejmenší společný násobek) a vrátíme se zase k teorii.

!

Věta 6a.19. (Bezoutova věta/rovnost) (Bezout's identity)Nechť $a, b \in \mathbb{Z}$. Pak existují $A, B \in \mathbb{Z}$ takové, že $\gcd(a, b) = Aa + Bb$.

Důkaz (drsný, poučný): Nejprve poznamenejme, že pokud $a = 0$, pak $\gcd(0, b) = b = 0 \cdot a + 1 \cdot b$, podobně identita platí pro $b = 0$. Dále tedy budeme předpokládat, že $a, b \neq 0$.

Uvažujme množinu $M = \{Aa + Bb; A, B \in \mathbb{Z}, Aa + Bb > 0\}$, tedy všechna kladná čísla, která lze dostat jako lineární kombinace a, b . Pak evidentně $M \neq \emptyset$, třeba $|a| + |b| \in M$, protože toto číslo dostaneme sečtením $s_a a + s_b b$ pro vhodně zvolená $s_a, s_b = \pm 1$. Je to neprázdná podmnožina \mathbb{N} , proto dle principu dobrého uspořádání (4c.14) existuje její nejmenší prvek c . Tvrdíme, že $c = \gcd(a, b)$.

1) Podle Důsledku 6a.4 (i) každý společný dělitel a a b dělí všechny prvky M , mimo jiné také c .

I $\gcd(a, b)$ je společný dělitel a, b , proto $\gcd(a, b) | c$, tedy $\gcd(a, b) \leq c$.

2) Ukážeme, že c je společný dělitel a a b . Nechť $a = qc + r$, kde $0 \leq r < c$. Máme $r = a - qc = a - (Aa + Bb) = (1 - A)a + Bb$, takže kdyby platilo $r > 0$, tak už $r \in M$ a zároveň $r < c$, což je spor s tím, že c je nejmenší v M . Proto $r = 0$ a $c | a$. Obdobně také ukážeme, že $c | b$. Takže c je společný dělitel, proto musí splňovat $c \leq \gcd(a, b)$.

Spojením 1) a 2) dostáváme, že opravdu $c = \gcd(a, b)$. Ale $c \in M$, proto se dá $\gcd(a, b)$ coby prvek M napsat jako $\gcd(a, b) = Aa + Bb$. □

Když víme, že $c = \gcd(a, b)$, tak lze pozorování z 1) formulovat takto:

Důsledek 6a.20.Nechť $a, b \in \mathbb{Z}$. Jestliže je d společný dělitel a, b , pak d dělí $\gcd(a, b)$.

Takže opravdu je $\gcd(a, b)$ největším společným dělitelem i vůči uspořádání dělitelností. Tím jsme splnili jeden dloužek z počátku této kapitoly.

Jak se taková kombinace najde? Někdy se to dá uhádnout. Víme například, že $\gcd(24, 60) = 12$, a uhádneme $12 = 3 \cdot 24 + (-1) \cdot 60$. Je ovšem také možné zkousit třeba $12 = (-2) \cdot 24 + 1 \cdot 60$. Bezoutova věta netvrzila, že je jediná možnost, ve skutečnosti je jich nekonečně mnoho, viz Diofantické rovnice 6c. Z hlediska aplikací mezi těmito možnostmi není rozdíl, cílem je najít nějaké takové vyjádření pro $\gcd(a, b)$. Existuje na to algoritmus vycházející ze zpětného chodu Euklidovým algoritmem. Vraťme se k předchozímu příkladu.

Příklad 6a.g: Zjistili jsme, že $\gcd(408, 108) = 12$. Jak vyjádříme 12 jako lineární kombinaci 408 a 108? Přečteme si běh Euklidova algoritmu od konce.

Máme $d = 12$ a poslední rozklad říká, že $84 = 3 \cdot 24 + d$, tedy $d = 84 - 3 \cdot 24$. Řádek předtím dá $108 = 84 + 24$, tedy $24 = 108 - 84$, a proto $d = 84 - 3 \cdot (108 - 84) = (-3) \cdot 108 + 4 \cdot 84$. První řádek dá $408 = 3 \cdot 108 + 84$, tedy $84 = 408 - 3 \cdot 108$, a proto $d = (-3) \cdot 108 + 4 \cdot (408 - 3 \cdot 108) = 4 \cdot 408 + (-15) \cdot 108$.

Máme $\gcd(408, 108) = 4 \cdot 408 + (-15) \cdot 108$. △

Takto to samozřejmě dělat nechceme, raději hledání A, B zabudujeme přímo do Euklidova algoritmu, dostaneme tak jeho rozšířenou verzi. V tomto algoritmu již nebude moct používat $r = a \bmod b$, protože i částečný podíl q_k bude hrát roli. Opět uvedeme jednu verzi indexovanou pro důkazy a jednu verzi počítací.

S Algoritmus 6a.21. Rozšířený Euklidův algoritmus pro nalezení $\gcd(a, b) = Aa + Bb$ pro $a > b \in \mathbb{N}$.

Verze 1.

Inicializace: $r_0 := a, r_1 := b, k := 0$, $A_0 := 1, A_1 := 0, B_0 := 0, B_1 := 1$.Krok: $k := k + 1, r_{k+1} = q_k \cdot r_k + r_{k+1}$, $A_{k+1} := A_{k-1} - q_k A_k, B_{k+1} := B_{k-1} - q_k B_k$.Opakovat dokud nenastane $r_{k+1} = 0$.Pak $\gcd(a, b) = r_k = A_k a + B_k b$.

nebo

Verze 2.

procedure gcd-Bezout (a, b : integer, $a > b > 0$) $A_0 := 1; A_1 := 0; B_0 := 0; B_1 := 1;$ **repeat** $a = q \cdot b + r$; $a := b; b := r$; $r_a := A_0 - q A_1$; $r_b := B_0 - q B_1$; $A_0 := A_1; A_1 := r_a$; $B_0 := B_1; B_1 := r_b$;**until** $b = 0$;**output:** a, A_0, B_0 ;

△

Všimněte si, že nový zbytek počítáme jako $r_{k+1} = r_{k-1} - q_k \cdot r_k$, což je přesně stejný vzorec jako pro A_{k+1} a B_{k+1} , dokonce používají stejné číslo q_k . To znamená, že jakmile si zjistíme $q_k = \lfloor \frac{r_{k-1}}{r_k} \rfloor$, tak už všechna tři čísla z nové generace počítáme stejným způsobem.

! procedure gcd-Bezout(a, b: integer, a > b > 0)

$A_0 := 1; A_1 := 0; B_0 := 0; B_1 := 1;$

repeat

$q := \lfloor \frac{a}{b} \rfloor;$

$r := a - q \cdot b;$

$r_a := A_0 - qA_1;$

$r_b := B_0 - qB_1;$

$a := b; b := r;$

$A_0 := A_1; A_1 := r_a;$

$B_0 := B_1; B_1 := r_b;$

until $b = 0;$

output: $a, A_0, B_0;$

Tento postup je ideálním vychodiskem pro ruční výpočet. Znamená to totiž, že jakmile doplníme políčko s q , tak do dalšího řádku dopočítáváme „nové b “, „nové A_1 “ a „nové B_1 “ pomocí stejného cestování tabulkou.

S 6a.22 Ruční výpočet. Najdeme $\gcd(408, 108)$ a vyjádříme jej jako lineární kombinaci 408 a 108. Začneme verzí se svislým zápisem (do řádků).

Nejprve je třeba vytvořit dva iniciační řádky. Hodnoty 408 a 108 jsou jasné, hodnoty pro A, B si je třeba pamatovat, například tak, že vypadají jako jednotková matice. Pak spočítáme $q = 3$ jako částečný podíl a zapíšeme do tabulkou.

a, b	q	A	B
408		1	0
108		0	1

a, b	q	A	B
408		1	0
108		0	1

V normálním Euklidově algoritmu následně najdeme zbytek $408 - 3 \cdot 108 = 84$, který zapíšeme pod 108. To znamená, že když se podíváme na první sloupec, tak děláme postup „číslo v předposledním mínus q krát číslo v posledním, výsledek zapíšeme na nový řádek“. Přesně tento vzorec pak aplikujeme na dvojici sloupců pro A a B , dostáváme $1 - 3 \cdot 0 = 1$ jako „nové A “ a $0 - 3 \cdot 1 = -3$ jako „nové B “. Tím jsme zkompletovali nový řádek a jsme připraveni celý algoritmus opakovat. To děláme tak dlouho, dokud v levém sloupci nebude nula.

a, b	q	A	B
408		1	0
108		0	1
408		1	0
108	3	0	1
84		1	-3
408		1	0
108	3	0	1
84	1	1	-3
24		-1	4

a, b	q	A	B
408		1	0
108	3	0	1
84	1	1	-3
24	3	-1	4
12		4	-15
408		1	0
108	3	0	1
84	1	1	-3
24	3	-1	4
12	2	4	-15
0			

Výsledek najdeme jako obvykle v řádku nad nulou, opravdu $12 = 4 \cdot 408 + (-15) \cdot 108$.

Všimněte si, že stejný vztah platí i pro řádek před tím: $24 = (-1) \cdot 408 + 4 \cdot 108$. A také ten před tím atd., až se dostaneme na první dva řádky: $108 = 0 \cdot 408 + 1 \cdot 108$ a $408 = 1 \cdot 408 + 0 \cdot 108$. To pro někoho může být další mnemotechnická pomůcka, jak si pamatovat výchozí hodnoty pro A, B . Zároveň to bude níže východiskem pro důkaz správnosti rozšířeného Euklidova algoritmu.

Tento výpočet svislou tabulkou se zdá být vhodným kompromisem mezi praktičností a zároveň názorností, budeme jej proto v této knize používat. Pro úplnost ukážeme ještě jeden algoritmus, který je graficky úspornější, vychází z vodorovného zápisu Euklidova algoritmu. Probíhá ve dvou fázích, začneme tak, že prostě provedeme Euklidův algoritmus s tím, že do prvního řádku píšeme hodnoty a, b a pod to odpovídající hodnoty q .

a, b	408	108	84	24	12	0
q			3	1	3	2

Teď vytvoříme třetí řádek, ten ale děláme zprava doleva. Nejprve pod poslední hodnoty q napíšeme (zprava) 0 a 1.

a, b	408	108	84	24	12	0
q		3	1	3	2	
				1	0	

Jsme připraveni k druhé fázi algoritmu, která funguje následovně: Vezmeme poslední číslo v třetím řádku (tedž 0) a odečteme od něj předposlední číslo (tedž 1) vynásobené koeficientem q nad předposledním číslem (tedž 3), dostaneme $0 - 1 \cdot 3 = -3$ a napišeme to doleva od předposledního čísla, tedy pod jedničku. Proces opakujeme, jen se vzorec přesune v tabulce o pole doleva. Vezmeme druhé číslo zprava (jedničku), odečteme číslo nalevo (tedž -3) vynásobené číslem nad ním (jedničkou), výsledek $1 - (-3) \cdot 1 = 4$ zapíšeme ještě o jedno doleva. V posledním kroku počítáme $-3 - 4 \cdot 3 = -15$ a zapíšeme zcela doleva.

a, b	408	108	84	24	12	0
q		3	1	3	2	
	-15	4	-3	1	0	

Co teď s tím? Začněme zprava, v horním a dolním řádku vidíme napravo dvojice $12 \leftrightarrow 0$ a $24 \leftrightarrow 1$. Když hodnoty z dolního řádku prohodíme, dostaneme $12 \leftrightarrow 1$ a $24 \leftrightarrow 0$ a $12 \cdot 1 + 24 \cdot 0 = 12 = \gcd(408, 108)$. Posuňme se o jedno doleva. Porovnáním horního a dolního řádku máme dvojice $24 \leftrightarrow 1$ a $84 \leftrightarrow -3$, prohodíme $24 \leftrightarrow -3$ a $84 \leftrightarrow 1$ a dostaneme $24 \cdot (-3) + 84 \cdot 1 = 12$. Další posun si zkuste sami. Happy end: V levých dvou sloupcích máme $408 \leftrightarrow -15$ a $108 \leftrightarrow 4$, prohodíme $408 \leftrightarrow 4$ a $108 \leftrightarrow -15$ a dostaneme $408 \cdot 4 + 108 \cdot (-15) = 12$, což je přesně hledaná Bezoutova identita. Náhoda? Nikoliv, takto to funguje vždy.

Tento způsob je bezesporu nejúspornější, ale je to už docela černá magie, je tedy více náchylný na chybu. Je například snadné zapomenout na to, že se mají na konci výsledná dvě čísla prohodit. Pokud má čtenář pocit, že mu chyby nehrozí (ani na písemce, kde přijde nervozita, nevyspání atd.), pak tento postup samozřejmě může s úspěchem používat. V této knize nicméně budeme po mnoha zkušenostech se studenty u zkoušek raději používat ten bezpečnejší předchozí algoritmus.

Poznámka důkaz správnosti algoritmu:

Potvrďme správnost našeho pozorování o tabulce. Přesně řečeno, dokážeme modifikovaným principem indukce (zpětný krok o dva) následující tvrzení pro $k \in \mathbb{N}_0$:

$V(k)$: Jestliže $r_k > 0$, pak $r_k = A_k a + B_k b$.

(0) $k = 0$: $r_0 = a$, $A_0 a + B_0 b = a$, v pořádku.

$k = 1$: $r_1 = b$, $A_1 a + B_1 b = b$, v pořádku.

(1) Nechť $k \geq 2$, předpokládáme, že V platí pro k a $k - 1$. Pak

$$\begin{aligned} r_{k+1} &= r_{k-1} - q_k \cdot r_k = (A_{k-1} a + B_{k-1} b) - q_k (A_k a + B_k b) \\ &= (A_{k-1} - q_k A_k) a + (B_{k-1} - q_k B_k) b = A_{k+1} a + B_{k+1} b. \end{aligned}$$

Důkaz správnosti algoritmu je hotov.

Pro úplnost ještě dokážeme, že ten kratší a adrenalinovější algoritmus opravdu funguje. Nejprve se vrátíme ke značení z důkazu správnosti Euklidova algoritmu, kde jsme zvolili $r_0 = a$, $r_1 = b$ a poté definovali rekurzí $q_k = \lfloor \frac{r_{k-1}}{r_k} \rfloor$, $r_{k+1} = r_{k-1} - q_k r_k$. Toto se opakuje, dokud nenastane $r_{n+1} = 0$, pak $r_n = \gcd(a, b)$.

Nyní definujeme $s_0 = 0$, $s_1 = 1$ a $s_{k+1} = s_{k-1} - q_{n-k} s_k$ pro $k = 1, \dots, n-1$. Tvrdíme, že pro každé $k = 1, \dots, n$ platí $r_{n-k+1} s_k + r_{n-k} s_{k-1} = r_n$, dokážeme to indukcí.

(0) Pro $k = 1$ začneme levou stranou vzorce: $r_{n-1+1} s_1 + r_{n-1} s_0 = r_n \cdot 1 + r_{n-1} \cdot 0 = r_n$, souhlasí.

(1) Nechť $k \in \{1, \dots, n-1\}$, předpokládejme, že $r_{n-k+1} s_k + r_{n-k} s_{k-1} = r_n$. Dosadíme za r_{n-k+1} z rekurentní definice a po úpravě dostaneme vzorec, který potřebujeme pro $k+1$:

$$\begin{aligned} r_n &= r_{n-k+1} s_k + r_{n-k} s_{k-1} = (r_{n-k-1} - q_{n-k} r_{n-k}) s_k + r_{n-k} s_{k-1} = r_{n-k-1} s_k - q_{n-k} r_{n-k} s_k + r_{n-k} s_{k-1} \\ &= r_{n-k-1} s_k + r_{n-k} (s_{k-1} - q_{n-k} s_k) = r_{n-k-1} s_k + r_{n-k} s_{k+1} = r_{n-k} s_{k+1} + r_{n-k-1} s_k \\ &= r_{n-(k+1)+1} s_{k+1} + r_{n-(k+1)} s_{(k+1)-1}. \end{aligned}$$

Tím je naše tvrzení dokázáno. Teď jej použijeme pro $k = n$:

$$\gcd(a, b) = r_n = r_1 s_n + r_0 s_{n-1} = b \cdot s_n + a \cdot s_{n-1}.$$

Opravdu tedy Bezoutovu identitu získáme pomocí posledních dvou čísel zpětného chodu, když je prohodíme.

△

! Bezoutova identita byla dokázána pro všechna celá čísla a, b , ale náš algoritmus funguje jen pro $a > b > 0$. Jak se najde vyjádření $\gcd(a, b) = Aa + Bb$ pro jiné možnosti? Pokud je některé z čísel nulové, tak je to triviální, viz důkaz Bezoutovy identity. Pokud $|a| = |b|$, pak taky není co řešit, viz Fakta 6a.12 a 6a.13. Zbývá případ, kdy je $|a| \neq |b| > 0$, ale to už se snadno udělá s trochou selského rozumu.

Příklad 6a.h: Chceme najít $\gcd(-108, 408)$ a vyjádřit jej Bezoutovým způsobem.

Víme, že $\gcd(-108, 408) = \gcd(|-108|, |408|) = \gcd(408, 108)$, tou druhou rovností jsme si to upravili, aby $a = 408$ bylo větší než $b = 108$. Nyní aplikujeme běžný Euklidův rozšířený algoritmus (viz výše) a dostaneme $\gcd(408, 108) = 12 = 4 \cdot 408 + (-15) \cdot 108$. Tedž už jenom vyrobíme správná znaménka u čísel:

$$\gcd(-108, 408) = \gcd(408, 108) = 12 = 15 \cdot (-108) + 4 \cdot 408.$$

△

Dal by se napsat obecný algoritmus, ale snad to není nutné.

Bezoutova identita je velice mocný nástroj pro praktické výpočty, jak ještě uvidíme zde níže i dalších kapitolách, ale také se nám bude hodit v důkazech. Jako ukázku si pomocí této identity ukážeme tvrzení, že když umíme číslem d vydělit součin ab , ale číslo d nemá nic společného s a , pak už musíme to d najít celé v b . To zní jako něco, co je evidentně pravdivé, ale bez pomoci Bezouta by se to dokazovalo překvapivě obtížně.

!

Lemma 6a.23.

Nechť $a, b \in \mathbb{Z}$ a $d \in \mathbb{N}$. Jestliže $d \mid ab$ a čísla d, a jsou nesoudělná, pak $d \mid b$.

Důkaz (rutinní): Podle předpokladu existuje $k \in \mathbb{Z}$ takové, že $ab = kd$. Protože jsou d, a nesoudělná, musí existovat čísla $A, B \in \mathbb{Z}$ taková, že $1 = \gcd(d, a) = Dd + Aa$. Pak $b = Ddb + Aab$, tedy $b = Ddb + Akd$, tedy $b = (Db + Ak)d$. To ukazuje, že d dělí b .

□

Cvičení

Cvičení 6a.1 (rutinní): Najděte částečný podíl a zbytek pro $2002/87$, $-1030/13$, $0/7$, $2/17$, $-3/7$, $8/2$.

Cvičení 6a.2 (rutinní): Dokažte, že pro každé $a \in \mathbb{Z}$ platí $1 \mid a$, $a \mid a$ a $a \mid 0$ (viz Fakt 6a.1).

Cvičení 6a.3 (rutinní, zkouškové): (i) Nechť $a, b \in \mathbb{Z}$. Dokažte, že jestliže $a \mid b$, pak $a \mid (nb)$ pro všechna $n \in \mathbb{Z}$.
(ii) Nechť $a, b, c \in \mathbb{Z}$. Dokažte, že jestliže $a \mid b$ a $b \mid c$, pak $a \mid c$ (viz Věta 6a.2).

Cvičení 6a.4 (rutinní, poučné): Nechť $a, b \in \mathbb{Z}$. Dokažte, že následující podmínky jsou ekvivalentní:

- (i) $a \mid b$,
- (ii) $(-a) \mid b$,
- (iii) $a \mid (-b)$,
- (iv) $(-a) \mid (-b)$,
- (v) $|a| \mid |b|$.

Nápověda: Vytvořte z implikací nějaký uzavřený cyklus zahrnující (i) až (iv), třeba (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (i) a dokažte jej. Rozmyslete si, že pak už z toho plyne libovolná implikace mezi nějakými dvěma podmínkami z těchto čtyř, jsou tedy všechny ekvivalentní. Pak toho využijte k důkazu ekvivalence (i) a (v).

Cvičení 6a.5 (rutinní): Nechť $a, b \in \mathbb{Z}$, $a > 0$. Dokažte, že $a \mid b$ právě tehdy, když $b \bmod a = 0$.

Cvičení 6a.6 (poučné, zkouškové): Dokažte, že pro každé $a, d \in \mathbb{N}$ takové, že $r = a \bmod d \neq 0$, platí $(-a) \bmod d = d - r$.

Jak to funguje pro případ $a \bmod d = 0$?

Cvičení 6a.7 (rutinní, zkouškové): Nechť $a, b, c, d \in \mathbb{Z}$. Dokažte, že jestliže $a \mid c$ a $b \mid d$, pak $ab \mid cd$.

Cvičení 6a.8 (rutinní, zkouškové): Nechť $a, b, c \in \mathbb{Z}$. Dokažte, že jestliže $ac \mid bc$ a $c \neq 0$, pak $a \mid b$.

Cvičení 6a.9 (poučné): Nechť $a, b, c \in \mathbb{Z}$. Dokažte/vyvraťte, že jestliže $a \mid bc$, pak $a \mid b$ nebo $a \mid c$.

Cvičení 6a.10 (rutinní, poučné): Dokažte, že pro $a \in \mathbb{N}$ platí $\gcd(a, 0) = a$ a $\gcd(a, a) = \text{lcm}(a, a) = a$ (viz Fakt 6a.13).

Cvičení 6a.11 (poučné): Nechť $n \in \mathbb{N}_0$.

(i) Dokažte, že n je dělitelné pěti právě tehdy, je-li jeho poslední cifra rovna 0 nebo 5.

(ii) Dokažte, že n je dělitelné čtyřmi právě tehdy, je-li jeho poslední dvoučíslí dělitelné 4.

Cvičení 6a.12 (poučné): Dokažte, že součin libovolných tří po sobě následujících celých čísel je vždy dělitelný 6.

Cvičení 6a.13 (rutinní, poučné, zkouškové): Dokažte, že pro každé $n \in \mathbb{N}_0$ platí následující dělitelnosti:

- | | | |
|------------------------|--------------------------|---|
| (i) $2 (n^2 - n)$; | (iii) $3 (n^3 + 2n)$; | (v) $6 (n^3 - n)$; |
| (ii) $2 (n^2 + n)$; | (iv) $5 (n^5 - n)$; | (vi) $21 (4^{n+1} + 5^{2n-1})$ (zde $n \geq 1$). |

Nápověda: Indukce to jistí.

Cvičení 6a.14 (rutinní, zkouškové): Pro následující dvojice $a, b \in \mathbb{Z}$ najděte $\gcd(a, b)$ faktORIZACÍ, pak rozšířeným Euklidovým algoritmem a napište příslušnou Bezoutovu identitu. Pak najděte $\text{lcm}(a, b)$.

- (i) $a = 420, b = 231$; (ii) $a = -60, b = -156$; (iii) $a = 118, b = -131$.

Cvičení 6a.15 (poučné): Nechť $a, b \in \mathbb{Z}$. Dokažte, že jestliže $\gcd(a, b) = 1$, pak $\text{lcm}(a, b) = |a| \cdot |b|$.

Cvičení 6a.16 (rutinní): Dokažte, že pro každé $a, b, k \in \mathbb{N}$ platí:

- (i) $\gcd(ka, kb) = k \gcd(a, b)$.
(ii) Jestliže k dělí a i b , pak $\gcd\left(\frac{a}{k}, \frac{b}{k}\right) = \frac{\gcd(a, b)}{k}$.

Cvičení 6a.17 (dobré): Dokažte, že pro každé $a, b, c \in \mathbb{N}$ platí, že $\gcd(a, bc)$ dělí $\gcd(a, b) \cdot \gcd(a, c)$

Cvičení 6a.18 (drsné): Dokažte: Vybereme-li $n + 1$ různých čísel z množiny $\{1, 2, 3, \dots, 2n\}$, pak mezi nimi musí být dvě takové, že jedno dělí druhé.

Viz také příklad 11b.k.

Cvičení 6a.19 (drsné): Jak zobecníme pojem gcd a lcm pro více čísel? Jsou dva obecné přístupy.

Jedna možnost je přes indukci. Pro tři čísla $a, b, c \in \mathbb{N}$ můžeme definovat $\gcd(a, b, c) = \gcd(\gcd(a, b), c)$ a $\text{lcm}(a, b, c) = \text{lcm}(\text{lcm}(a, b), c)$, gcd i lcm se pokaždé aplikují jen na dvě čísla.

Když to umíme pro tři, můžeme pro čtyři čísla $a, b, c, d \in \mathbb{N}$ definovat $\gcd(a, b, c, d) = \gcd(\gcd(a, b, c), d)$ a $\text{lcm}(a, b, c, d) = \text{lcm}(\text{lcm}(a, b, c), d)$, atd., indukcí tak dokážeme zadefinovat oba pojmy pro libovolný (konečný) počet čísel.

Druhý přístup je významový, $\gcd(a_1, a_2, \dots, a_n)$ definujeme jako největší přirozené číslo d splňující vlastnost, že $d | a_i$ pro všechna i , obdobně $\text{lcm}(a_1, a_2, \dots, a_n)$ definujeme jako nejmenší přirozené číslo d splňující vlastnost, že $a_i | d$ pro všechna i .

Dá se ukázat, že oba přístupy dávají stejné výsledky. To by bylo na jedno cvičení trochu moc, dokažte tedy následující:

Nechť $a, b, c \in \mathbb{N}$ jsou libovolná. Nechť d je největší přirozené číslo takové, že $d | a$, $d | b$ a $d | c$. Pak $d = \gcd(\gcd(a, b), c)$.

Cvičení 6a.20 (poučné): Nechť $a_1, a_2, \dots, a_n \in \mathbb{N}$.

Rozmyslete si, zda platí $\text{lcm}(a_1, a_2, \dots, a_n) = \frac{a_1 \cdot a_2 \cdots a_n}{\gcd(a_1, a_2, \dots, a_n)}$.

Rozmyslete si, zda platí, že když jsou a_i po dvou nesoudělná, pak $\text{lcm}(a_1, a_2, \dots, a_n) = a_1 \cdot a_2 \cdots a_n$.

Pro případný důkaz platnosti viz cvičení 6b.4

Řešení:

6a.1: $q = 23, r = 17; q = -80, r = 10; q = 0, r = 0; q = 0, r = 2; q = -1, r = 4; q = 4, r = 0$.

6a.2: $a = a \cdot 1, a = 1 \cdot a, 0 = 0 \cdot a$.

6a.3: (i): $b = ka \implies bn = \dots$ (ii) podobně.

6a.4: (i) \implies (ii): $a | b \implies b = ka, k \in \mathbb{Z} \implies b = -k \cdot (-a) \wedge -k \in \mathbb{Z} \implies -a | b$. (ii) \implies (iii), (iii) \implies (iv), (iv) \implies (i) obdobně.

(i) \implies (v): $a | b \implies b = ka, k \in \mathbb{Z} \implies |b| = |k| \cdot |a| \wedge |k| \in \mathbb{Z} \implies |a| | |b|$.

(v) \implies ?: $|a| | |b| \implies |b| = k|a|$. Zbavíme se absolutních hodnot, podle známének a a b se ve vztahu objeví plusy či mínusy $\pm b = k(\pm a)$, tedy důkaz se rozpadne na čtyři případy, pokaždé se skončí nějakou konkrétní situací $(\pm a) | (\pm b)$ neboli jedním z tvrzení (i) až (iv).

6a.5: Jestliže $a | b$, pak $b = ka = ka + 0$, tedy $r = 0$.

Jestliže $b \bmod a = 0$, pak $b = qa + 0 = qa$ a $q \in \mathbb{Z}$.

Jestliže $r = a \bmod d$, pak $a = qd + r$ pro nějaké $q \in \mathbb{N}_0$ a $0 \leq r < d$. Předpoklad dále dává $r > 0$. Pak také $(-a) = (-q)d - r$ a $-d < -r < 0$, proto $(-a) = (-q - 1)d + (d - r)$. Teď $(-q - 1) \in \mathbb{Z}$ a $0 < d - r < d$, číslo $d - r$ proto splňuje podmínu z definice $(-a) \bmod d$.

Jestliže $r = 0$ neboli $d | a$, pak $(-a) \bmod d = 0 = r$.

Cvičení 6a.6 (poučné, zkouškové): Dokažte, že pro každé $a, d \in \mathbb{N}$ platí: $(-a) \bmod d = d - a \bmod d$.

Nápověda: Označte si $r = a \bmod d$.

6a.7: $c = ka, d = lb \wedge k, l \in \mathbb{Z} \implies cd = (kl)ab \wedge (kl) \in \mathbb{Z} \implies ab | cd$.

6a.8: $ac | bc \implies bc = kac \wedge k \in \mathbb{Z} \implies b = ka \wedge k \in \mathbb{Z} \implies a | b$.

6a.9: Neplatí, $6 \mid (4 \cdot 9)$, ale není $6 \mid 4$ ani $6 \mid 9$.

6a.10: Libovolné $d \in \mathbb{N}$ dělí 0, proto je množina společných dělitelů $a, 0$ totožná s množinou dělitelů $d \in \mathbb{N}$ čísla a . Takoví dělitelé nutně splňují $d \leq a$ a víme, že také $a \mid a$, tudíž a náleží do množiny společných dělitelů a je tam největší.

$\gcd(a, a)$ má být největší dělitel a , což je samozřejmě a . Důkaz pro $\text{lcm}(a, a)$ je obdobný.

6a.11: (i) Označme $n = 10a + b$, tedy b je poslední cifra. Protože $10a = (2a) \cdot 5$, tak (podle Důsledku 6a.4) $5 \mid n$ právě tehdy, když $5 \mid b$. Pro jednociferné číslo b ale $5 \mid b$ jen pro $b = 0$ a $b = 5$.

(ii): Označte $n = 100a + b$, kde b je dvouciferné.

6a.12: Jedno z nich musí být sudé, jedno z nich musí být dělitelné třemi, viz Fakt 6a.11.

6a.13: Vše indukcí.

(i): (0) $n = 0 \implies n^2 - n = 0, 2 \mid 0$.

(1) Nechť $n \in \mathbb{N}_0$ libovolné, předpoklad: $2 \mid (n^2 - n) \implies (n^2 - n) = 2k, k \in \mathbb{Z}$. Pak

$$(n+1)^2 - (n+1) = n^2 + n = (n^2 - n) + 2n = 2k + 2n = 2(k+n) \text{ a } n+k \in \mathbb{Z}, \text{ tedy } 2 \mid [(n+1)^2 - (n+1)].$$

(vi): (0) $n = 1 \implies 21 \mid (16+5)$ platí.

(1) Nechť $n \in \mathbb{N}$ libovolné, předpoklad: $21 \mid (4^{n+1} + 5^{2n-1}) \implies (4^{n+1} + 5^{2n-1}) = 21k$ pro $k \in \mathbb{Z}$. Pak

$$4^{(n+1)+1} + 5^{2(n+1)-1} = 4 \cdot 4^{n+1} + 25 \cdot 5^{2n-1} = 4 \cdot 4^{n+1} + 4 \cdot 5^{2n-1} + 21 \cdot 5^{2n-1} = 4(4^{n+1} + 5^{2n-1}) + 21 \cdot 5^{2n-1} = 4 \cdot 21k + 21 \cdot 5^{2n-1} = 21(4k + 5^{2n-1}), \text{ kde } (4k + 5^{2n-1}) \in \mathbb{Z}.$$

6a.14: (i):

420		1	0	$\gcd(2^2 \cdot 3 \cdot 5 \cdot 7, 3 \cdot 7 \cdot 11) = 3 \cdot 7 = 21$
231	1	0	1	$\text{lcm}(2^2 \cdot 3 \cdot 5 \cdot 7, 3 \cdot 7 \cdot 11) = 2^2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 = 4620$
189	1	1	-1	
42	4	-1	2	
21•	2	5•	-9•	
0				

$$\gcd(420, 231) = 21 = 5 \cdot 420 + (-9) \cdot 231.$$

(ii): Hledat $\gcd(|-60|, |-156|) = \gcd(156, 60)$.

156		1	0	$\gcd(-2^2 \cdot 3 \cdot 5, -2^2 \cdot 3 \cdot 13) = 2^2 \cdot 3 = 12$
60	2	0	1	$\text{lcm}(-2^2 \cdot 3 \cdot 5, -2^2 \cdot 3 \cdot 13) = 2^2 \cdot 3 \cdot 5 \cdot 13 = 780$
36	1	1	-2	
24	1	-1	3	
12•	2	2•	-5•	
0				

$$\gcd(-60, -156) = 12 = 2 \cdot 156 + (-5) \cdot 60;$$

(iii): Hledat $\gcd(|-131|, 118) = \gcd(131, 118)$.

131		1	0	$\gcd(-131, 2 \cdot 59) = 1$
18	1	0	1	$\text{lcm}(-131, 2 \cdot 59) = 131 \cdot 2 \cdot 59 = 15458$
13	9	1	-1	
1•	13	-9•	10•	
0				

$$\gcd(118, -131) = 1 = (-9) \cdot 131 + 10 \cdot 118 = 10 \cdot 118 + 9 \cdot (-131).$$

6a.15: Stačí použít $\text{lcm}(a, b) = \frac{|a| \cdot |b|}{\gcd(a, b)}$.

Správný matematický přístup je využívat již odvedené práce.

6a.16: (i): Označme $d = \gcd(a, b)$ a $e = \gcd(ka, kb)$. Pak d dělí a, b , proto kd dělí ka, kb , platí tedy $kd \leq e$. Naopak: Podle Bezouta $d = Aa + Bb$, proto $kd = Aka + Bkb$, e dělí obě napravo, proto dělí i kd a tedy $e \leq kd$.

(ii) Podobně jako v (i) nebo přímo, aplikujte (i) na $a' = \frac{a}{k}$ a $b' = \frac{b}{k}$.

6a.17: Podle Bezouta $\gcd(a, b) = A_b a + B_b b$ a $\gcd(a, c) = A_c a + C_b c$.

Pak $\gcd(a, b) \gcd(a, c) = a(A_b A_c a + A_b C_b b + A_c B_b b) + bcBC$. $\gcd(a, bc)$ dělí a i bc , tudíž musí dělit i ten součin.

6a.18: Indukcí na n .

(0) $n = 1$: Vybereme dvě různá čísla z $\{1, 2\}$, pak to jsou 1 a 2 a $1 \mid 2$.

Ze zvědavosti $n = 2$: Vybereme tři různá čísla z $\{1, 2, 3, 4\}$, pokud je mezi nimi 1, tak ta určitě dělí nějaké další. Pokud mezi nimi 1 není, tak jsme nutně vybrali 2, 3, 4 a $2 \mid 4$.

(1) Předpokládejme platnost pro nějaké $n \in \mathbb{N}$. Teď vybereme $(n+1)+1 = n+2$ čísel z množiny $\{1, 2, \dots, 2n, 2n+1, 2n+2 = 2(n+1)\}$. Dvě možnosti:

a) Jestliže je alespoň $n+1$ z těchto čísel vybráno z množiny $\{1, 2, \dots, 2n\}$, tak aplikujeme indukční předpoklad a najdeme mezi nimi dvě, že jedno dělí druhé.

b) Z těch $n + 2$ čísel je v množině $\{1, 2, \dots, 2n\}$ jen n . To znamená, že jsme nutně vybrali čísla $2n + 1$ a $2n + 2$. Pokud jsme vybrali i číslo $n + 1$, tak jsme hotovi.

Zbývá následující situace: Vybrali jsme čísla $2n + 1$, $2n + 2$ a ještě množinu M obsahující n různých čísel z množiny $\{1, 2, \dots, 2n\}$, přičemž M neobsahuje $n + 1$. Pak je $M \cup \{n + 1\}$ množina obsahující $n + 1$ různých čísel z množiny $\{1, 2, \dots, 2n\}$, proto podle indukčního předpokladu v této množině existují a, b taková, že $a | b$. Z toho plyne $a < b$, proto a nemůže být $n + 1$. Pokud ani b není $n + 1$, pak $a, b \in M$, jsou tedy i v původním výběru a jedno dělí druhé, hotovo.

Zbývá možnost, že $b = n + 1$, pak je v M číslo, které dělí $n + 1$, což zase dělí $2n + 2$, které taktéž je v našem výběru, takže v našem původním výběru je číslo dělící $2n + 2$, hotovo.

6a.19: Označme $D = \gcd(\gcd(a, b), c)$. 1) Z definice $D | c$ a $D | \gcd(a, b)$, odtud pak zase $D | a$ a $D | b$. Takže D je společný dělitel všech čísel a, b, c , tudíž $D \leq d$, neboť d je největší takový.

2) Pokud je d největší společný dělitel a, b, c , pak je to i společný dělitel a, b , tudíž musí platit $d | \gcd(a, b)$. Také $d | c$, takže d je společný dělitel čísel $\gcd(a, b)$ a c , tudíž $d \leq D$.

6b. Prvočísla

Teď se budeme soustředit na čísla z \mathbb{N} . Podíváme se na speciální druh čísel, která jsou z hlediska dělitelnosti základními stavebními bloky všech čísel.

! Definice.

Nechť $a \in \mathbb{N}$, $a \neq 1$.

Řekneme, že je to **prvočíslo (prime)**, jestliže jediná přirozená čísla, která jej dělí, jsou 1 a a .

Řekneme, že a je **složené číslo (composite number)**, jestliže to není prvočíslo.

Prvočísla tradičně značíme jako p , popř. q . Vidíme, že 1 není prvočíslo ani složené číslo, je to nějaké jiné číslo. Občas si na to budeme muset dát pozor.

Všimneme si, že číslo $a \in \mathbb{N}$ je složené, jestliže existuje $k \in \mathbb{N}$ takové, že $k | a$ a $1 < k < a$. Ještě jinak: Číslo a je složené, jestliže existují $x, y \in \mathbb{N}$ takové, že $a = xy$ a $x < a$, $y < a$.

! Příklad 6b.a: Prvočísla byla zkoumána už od nepaměti a každý určitě nějaká zná. Mezi první stovkou přirozených čísel jsou tato prvočísla (je jich 25):

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.

Další je pak 101.

△

Prvočísla najdeme ve všech číslech (skoro, kromě jedničky).

Fakt 6b.1.

Nechť $n \in \mathbb{N}$, $n \neq 1$. Pak existuje prvočíslo p takové, že $p | n$.

Důkaz (poučný): Dokážeme to silnou indukcí na n pro $n \geq 2$.

(0) $n = 2$: Ano, existuje prvočíslo $p = 2$, které dělí $n = 2$.

(1) Nechť $n \in \mathbb{N}$, $n \geq 2$. Předpokládejme, že tvrzení platí pro všechna čísla $2, 3, \dots, n$. Uvažujme číslo $n + 1$. Pokud je to prvočíslo, tak dáme $p = n + 1$ a jsme hotovi.

Pokud to prvočíslo není, pak musí existovat $k \in \mathbb{N}$ takové, že $k | (n + 1)$ a $1 < k < n + 1$. Pak je ale k z množiny $2, 3, \dots, n$, tudíž existuje prvočíslo p takové, že $p | k$. Spolu s $k | (n + 1)$ a tranzitivitou této relace dostáváme $p | (n + 1)$ a jsme hotovi.

□

Při manipulací s dělitelností máme občas problém, že nám číslo d dělí součin $a \cdot b$, ale my z toho nejsme schopni nic říct. Třeba $6 | (4 \cdot 9)$, ale neplatí ani $6 | 4$, ani $6 | 9$, intuitivně se kousek šestky schoval do 4 a druhý kousek do 9. S prvočísly takové problémy nenastanou.

! Lemma 6b.2.

Nechť $a_1, \dots, a_m \in \mathbb{N}$ a p je prvočíslo. Jestliže $p | (a_1 a_2 \cdots a_m)$, pak existuje i takové, že $p | a_i$.

Důkaz (poučný): Dokážeme to indukcí na m .

(0) Jestliže $m = 1$, tak předpokládáme $p \mid a_1$, z čehož plyne $i = 1$ a je to.

(1) Předpokládejme, že pro nějaké $m \geq 1$ tvrzení Lemma platí pro libovolné $a_1, \dots, a_m \in \mathbb{Z}$. Mějme teď $a_1, \dots, a_m, a_{m+1} \in \mathbb{Z}$ takové, že $p \mid (a_1 \cdots a_m a_{m+1})$. Protože jediní dělitelé p jsou p a 1, tak je i $\gcd(p, a_{m+1})$ buď p nebo 1. V prvním případě je p dělitelem a_{m+1} , tedy $p \mid a_{m+1}$ a jsme hotovi.

V opačném případě $\gcd(p, a_{m+1}) = 1$. Označme $b = a_1 \cdots a_m$, máme tedy situaci, kdy $p \mid (a_{m+1}b)$ a také $\gcd(p, a_{m+1}) = 1$, tudíž Lemma 6a.23 říká, že $p \mid b$, tedy $p \mid (a_1 \cdots a_m)$. Pak ale podle indukčního předpokladu musí existovat i takové, že $p \mid a_i$. □

Jako rychlý důsledek si dokažme jednu užitečnou ekvivalenci.

Lemma 6b.3.

Nechť $d, a, b \in \mathbb{N}$. Pak $\gcd(d, ab) = 1$ právě tehdy, když $\gcd(d, a) = 1$ a $\gcd(d, b) = 1$.

Důkaz (poučný): 1) \implies : Předpokládejme, že $\gcd(d, ab) = 1$. Nechť $x \in \mathbb{N}$ je společný dělitel d a a . Pak podle Věty 6a.2 (ii) x dělí také d a ab , tedy platí $x \leq \gcd(d, ab) = 1$. Jediný společný dělitel a a d je tedy 1, proto jsou nesoudělná. Důkaz pro d, b je obdobný.

2) Druhý směr dokážeme obměnou. Předpokládejme, že $\gcd(d, ab) > 1$. Pak existuje číslo $k > 1$, které dělí d i ab . Podle Faktu 6b.1 tudíž musí existovat i provočíslo p , které je společným dělitelem d a ab . Lemma 6b.2 ovšem tvrdí, že si p musí vybrat, řekněme, že $p \mid a$. Pak ale p dělí d i a , proto $\gcd(d, a) \geq p > 1$. Neplatí tedy výrok „ $\gcd(d, a) = 1$ a $\gcd(d, b) = 1$ “. □

Lemma 6b.2 bude hrát hlavní roli při důkazu následující věty.

!

Věta 6b.4. (Fundamentální věta aritmetiky) (Fundamental theorem of arithmetics)

Nechť $n \in \mathbb{N}$. Pak existují prvočísla p_1, p_2, \dots, p_m a exponenty $k_1, k_2, \dots, k_m \in \mathbb{N}_0$ takové, že

$$n = p_1^{k_1} \cdot p_2^{k_2} \cdots p_m^{k_m} = \prod_{i=1}^m p_i^{k_i}.$$

Jestliže pro $n \geq 2$ přidáme podmínky $p_1 < p_2 < \dots < p_m$ a $k_i > 0$, tak je tato dekompozice jednoznačně určena.

Důkaz (poučný, drsný): 1) K důkazu existence dekompozice použijeme silný princip indukce na $n \in \mathbb{N}$.

(0) Jestliže $n = 1$, tak zvolíme libovolné prvočíslo, třeba $p = 13$, a $k_1 = 0$, dostáváme $1 = p^0$.

(1) Předpokládejme, že rozklad existuje pro všechna čísla $1, 2, \dots, n$. Uvažujme nyní číslo $n + 1$.

Jestliže je to prvočíslo, pak zvolíme $p_1 = n + 1$ a $k_1 = 1$, hotovo.

Jinak je to číslo složené, tedy existují $a, b \in \mathbb{N}$ takové, že $n + 1 = a \cdot b$ a $a, b < n + 1$. Podle indukčního předpokladu máme $a = \prod_{i=1}^M q_i^{k_i}$ a $b = \prod_{i=1}^N r_i^{k_i}$ pro nějaká prvočísla q_i, r_j . Pak $n + 1 = \prod_{i=1}^M q_i^{K_i} \prod_{i=1}^N r_i^{L_i}$, takže je to opravdu součin mocnin prvočísel. Aby to odpovídalo formálně, provedeme přejmenování: Označíme $m = M + N$, $p_i = q_i$ a $k_i = K_i$ pro $i = 1, \dots, M$, dále $p_i = r_{i-M}$ a $k_i = L_{i-M}$ pro $i = M + 1, \dots, m$ a dostáváme $n + 1 = \prod_{i=1}^m p_i^{k_i}$. Důkaz je hotov.

2) Teď jednoznačnost: Nechť $n \in \mathbb{N}$, $n \geq 2$, a předpokládejme, že $n = \prod_{i=1}^m p_i^{k_i}$ a $n = \prod_{i=1}^m q_i^{l_i}$, kde $p_1 < \dots < p_m$, $q_1 < \dots < q_M$ a $k_i, l_j > 0$.

a) Nejprve ukážeme, že $p_i = q_i$ pro každé i takové, že p_i či q_i existuje. Dokážeme to silnou indukcí na číslo i .

(0) Protože je p_1 prvočíslo a dělí n , musí podle Lemma 6b.2 dělit i jedno z čísel q_j . Jenže q_j má jen dělitele 1 a q_j a p_1 coby prvočíslo není 1, proto $p_1 = q_j$. Protože jsou prvočísla srovnána dle velikosti, vyplývá z toho také, že $q_1 \leq q_j = p_1$.

Naopak prvočíslo q_1 dělí n , proto symetricky musí existovat p_j takové, že $q_1 = p_j$, a tedy $p_1 \leq p_j = q_1$.

Tyto dvě nerovnosti dávají $p_1 = q_1$.

(1) Předpokládejme, že už máme $p_1 = q_1, \dots, p_i = q_i$. Pokud p_{i+1} existuje, pak dělí n a musí existovat q_j takové, že $p_{i+1} = q_j$. Jenže q_1, \dots, q_i jsou už obsazena jinými p , musí platit $j \geq i + 1$ (takže existuje i q_{i+1} a možná nějaká další). Máme tedy $q_{i+1} \leq q_j = p_{i+1}$. Symetrickým argumentem z existence q_{i+1} dostaneme $p_{i+1} \leq q_{i+1}$, tedy i $p_{i+1} = q_{i+1}$.

Tím je dokončen indukční krok, zároveň z toho vyplývá nemožnost situace $m < M$ či $M < m$.

b) Teď již víme, že oba rozklady zahrnují stejná prvočísla, tedy máme $n = \prod_{i=1}^m p_i^{k_i}$ a $n = \prod_{i=1}^m p_i^{l_i}$. Potřebujeme ukázat, že $k_i = l_i$ pro všechna $i = 1, \dots, m$. Vezměme nějaké takové i a ze symetrie situace předpokládejme, že $k_i \leq l_i$. Vydělíme oba rozklady číslem $p_i^{k_i}$ a dostaneme dva rozklady pro číslo $\frac{n}{p_i^{k_i}}$. V tom prvním se p_i vůbec nevyskytuje, v tom druhém je s exponentem $l_i - k_i \geq 0$. Ale podle části a) aplikované na tyto dva vydelené rozklady musí mít oba stejná prvočísla, což nastane jedině v případě, že $l_i - k_i = 0$, tedy $k_i = l_i$.

Důkaz pro $n \geq 2$ je hotov. \square

Jednoznačnost vlastně platí i pro $n = 1$, ale je to trikem, proto jsme to do věty nezahrnuli. Jak se vůbec vyjádří 1 pomocí prvočísel, když máme podmínu $k_1 > 0$? Zvolíme $m = 0$ (prvočísla žádná nevybíráme), pak se v součinu $\prod_{i=1}^0$ násobí přes prázdnou množinu, což je podle definice právě 1. Jiná možnost není.

Vyjádření čísla n jako ve větě říkáme **prvočíselný rozklad**. V mnoha situacích máme rádi jednoznačnost z části b), ale někdy je pro změnu výhodné si dovolit přidat do rozkladu prvočísla s mocninou 0 (třeba $13 = 13 \cdot 3^0 = 13 \cdot 23^0 \cdot 33^0$), což například umožní sjednotit použitá prvočísla pro více čísel. Dobrým příkladem je následující aplikace prvočíselného rozkladu na dělitelnost.

Lemma 6b.5.

Nechť $a \in \mathbb{N}$ je číslo s prvočíselným rozkladem $\prod_{i=1}^m p_i^{k_i}$, nechť $d \in \mathbb{N}$. Číslo d dělí a právě tehdy, když existují čísla $K_i \in \mathbb{N}_0$ splňující pro všechna i podmínu $0 \leq K_i \leq k_i$ taková, že $d = \prod_{i=1}^m p_i^{K_i}$.

Přeloženo do lidštiny, aby číslo d dělilo číslo a , nemůže mít v rozkladu prvočísla jiná, než jsou v a , a prvočíslo, které v d je, tam nemůže být vicekrát, než je v a .

Důkaz je variací na důkaz předchozí Věty. Pokud d dělí a , tak se pro každé p z rozkladu d ukáže, že musí být v rozkladu a , načež se vydelením p^k ukáže, že v číslu a musí být exponent alespoň tak velký jako u d . Pokud naopak nějaké prvočíslo p z rozkladu a chybí v daném rozkladu d , tak jej tam prostě přidáme s mocninou 0.

Odtud hned dostaneme následující tvrzení, které matematicky potvrdí oblíbený způsob hledání gcd a lcm pro menší čísla. Obě čísla napíšeme pomocí prvočíselného rozkladu, gcd se pak získá pomocí nejmenších mocnin a lcm pomocí největších mocnin u prvočísel (pokud nějaké prvočíslo z jednoho rozkladu chybí v druhém, dodáme jej tam s mocninou 0). Formálně řečeno:

Fakt 6b.6.

Nechť $a, b \in \mathbb{N}$. Předpokládejme, že máme prvočísla $p_1 < \dots < p_m$ a čísla $k_i, l_i \in \mathbb{N}_0$ taková, že $a = \prod_{i=1}^m p_i^{k_i}$ a $b = \prod_{i=1}^m p_i^{l_i}$. Pak $\gcd(a, b) = \prod_{i=1}^m p_i^{\min(k_i, l_i)}$ a $\text{lcm}(a, b) = \prod_{i=1}^m p_i^{\max(k_i, l_i)}$.

Důkaz (z povinnosti): 1) Označme $n = \prod_{i=1}^m p_i^{\min(k_i, l_i)}$. Protože má n ve svém rozkladu stejná prvočísla jako a i b a jejich exponenty splňují $\min(k_i, l_i) \leq k_i$ a $\min(k_i, l_i) \leq l_i$, podle předchozího Lemmatu $n | a$ a $n | b$. Je to tedy společný dělitel. Zbývá ukázat, že je největší.

Nechť d je nějaký společný dělitel a, b . Pak podle předchozího Lemmatu musí existovat čísla K_i taková, že $d = \prod_{i=1}^m p_i^{K_i}$ a přitom $K_i \leq k_i$ a $K_i \leq l_i$ pro všechna i . To pak ale znamená, že $K_i \leq \min(k_i, l_i)$ pro všechna i , tudíž zase podle Lemmatu $d | n$.

Takže n je opravdu největší společný dělitel a, b .

2) Důkaz vzorce pro $\text{lcm}(a, b)$ je obdobný. \square

Z toho hned dostaneme zajímavý důkaz Věty 6a.14. Pro všechna $k, l \in \mathbb{N}_0$ platí $\max(k, l) + \min(k, l) = k + l$ (zkuste si to rozmyslet, stačí rozebrat varianty podle toho, které z těch čísel je větší), což dává

$$\begin{aligned}\gcd(a, b) \cdot \text{lcm}(a, b) &= \prod_{i=1}^m p_i^{\min(k_i, l_i)} \cdot \prod_{i=1}^m p_i^{\max(k_i, l_i)} = \prod_{i=1}^m p_i^{\min(k_i, l_i) + \max(k_i, l_i)} \\ &= \prod_{i=1}^m p_i^{k_i + l_i} = \prod_{i=1}^m p_i^{k_i} \cdot \prod_{i=1}^m p_i^{l_i} = a \cdot b.\end{aligned}$$

Pro úplnost ukážeme několik příkladů na hledání gcd a lcm pomocí rozkladu.

Příklad 6b.b: Uvažujme čísla 24 a 60. Jejich prvočíselné rozklady jsou $24 = 2^3 \cdot 3$ a $60 = 2^2 \cdot 3 \cdot 5$, takže máme $\gcd(24, 60) = \gcd(2^3 \cdot 3 \cdot 5^0, 2^2 \cdot 3 \cdot 5) = 2^2 \cdot 3 = 12$ a $\text{lcm}(24, 60) = 2^3 \cdot 3 \cdot 5 = 120$.

Podobně $\gcd(2 \cdot 3^2 \cdot 5 \cdot 7^2, 3 \cdot 7^4 \cdot 13) = 3 \cdot 7^2$ a $\text{lcm}(2 \cdot 3^2 \cdot 5 \cdot 7^2, 3 \cdot 7^4 \cdot 13) = 2 \cdot 3^2 \cdot 5 \cdot 7^4 \cdot 13$.

△

Tento postup je ovšem jako obecný přístup neperspektivní, protože prvočíselný rozklad daného čísla je jedním z nejnáročnějších problémů.

6b.7 Rozklad na prvočísla. Zásadním problém je najít k danému číslu n nějaké prvočíslo p , které jej dělí. Pokud toto umíme, tak aplikací téhož postupu na číslo n/p atd. získáme nakonec rozklad. Budeme se tedy soustředit na problém nalezení prvočíselného dělitele.

Jako první metoda se nabízí prostě zkoušet dělit rozkládané číslo n číslami 2, 3, 4, To asi čtenář zná. Vezme 45, zkusí vydělit dvojkou, nic, zkusí trojkou, zásah, máme $45 \div 3 = 15$. Pokračujeme s patnáctkou, zkusíme zase trojku, bingo, $15 \div 3 = 5$, rozklad hotov, $45 = 3^2 \cdot 5$.

U malých čísel toto může fungovat efektivně, zejména když si vzpomeneme na rozličná kritéria dělitelnosti, viz například cvičení 6a.11 a 7a.4.

Pro větší čísla je tento přístup totální katastrofa, protože pokud máme smůlu (n je prvočíslo), tak musíme projít všechna čísla 1, 2, ..., n , náročnost algoritmu je tedy n , a to jsme ještě ani nevzali v úvahu, že samotné rozhodování, zda nějaké číslo d dělí n , také něco stojí. Jisté zlepšení se nabízí.

! Fakt 6b.8.

Jestliže je n složené číslo, pak existuje jeho prvočíselný dělitel menší či roven číslu \sqrt{n} .

Důkaz: Předpokládejme, že $n = ab$ a $a, b > 1$. Tvrdíme, že buď $a \leq \sqrt{n}$ nebo $b \leq \sqrt{n}$. V opačném případě bychom totiž měli $ab > \sqrt{n} \cdot \sqrt{n} = n$.

Takže předpokládejme, že třeba $a \leq \sqrt{n}$. Vezměme libovolné prvočíslo p z prvočíselného rozkladu a , to pak splňuje $p \leq a \leq \sqrt{n}$ a dělí n . □

To znamená, že pokud dané číslo n nedělí nic až po \sqrt{n} , tak už víme, že je to prvočíslo. Podobně lze ukázat, že pokud nějaké číslo n vzniklo jako součin tří prvočísel, pak to nejmenší z nich nesmí být větší než $\sqrt[3]{n}$, a tak dále.

Náročnost našeho naivního algoritmu (kterému se také říká „direct search algorithm“ či „trial division“) je tedy \sqrt{n} kroků, když do toho započítáme náročnost dělení, budeme na tom ještě hůř. V praxi se často velikost čísla soudí podle počtu cifer $m = \log_2(n)$, pak $n = 2^m$ a lze říci, že v nejhorším případě potřebujeme pro m -ciferné číslo použít $2^{m/2}$ kroků, což je hodně.

Při postupném dělení čísla 2, 3, 4, ... by pomohlo, kdybychom měli tabulkou prvočísel, protože pak bychom nemuseli dělit n všemi čísly až po \sqrt{n} , stačilo by brát jen prvočísla. Těch je relativně málo, například jsme viděli, že je jen 25 prvočísel menších než 100. To znamená, že kdybychom chtěli udělat rozklad čísla 10003, tak bychom nemuseli zkoušet dělit všemi čísly až po $\sqrt{10001} \sim 100$, ale jen oněmi 25 prvočísly.

Takovou tabulku naštěstí nemusíme tvorit postupným testováním čísel, existuje metoda, která to zvládá relativně efektivně.

! Příklad 6b.c:

Pokud potřebujeme identifikovat všechna prvočísla v rozmezí 1 až n , pak můžeme s úspěchem použít metodu zvanou **Eratosthenovo síto** (sieve of Eratosthenes). Funguje to následovně.

Nejprve si všechna čísla od 2 do n napíšeme na papír, třeba do tabulky nebo za sebe, to je jedno:

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, ...

Začneme s $a = 2$. Je to prvočíslo, tak jej označíme a pak ze seznamu vyškrtneme všechny jeho násobky:

[2], 3, , 5, , 7, , 9, , 11, , 13, , 15, , 17, , ...

Podíváme se do seznamu a najdeme první další nevyškrtnuté a neoznačené číslo, je to $a = 3$. Musí to být prvočíslo, tak jej označíme a pak ze seznamu vyškrtneme všechny jeho násobky, které tam ještě zbyly:

[2], [3], , 5, , 7, , , , 11, , 13, , , , 17, , ...

Podíváme se do seznamu a najdeme první další neoznačené číslo, tedy $a = 5$. Ooznačíme jej coby prvočíslo a pak ze seznamu vyškrtneme atd.: [2], [3], [5], , 7, , , , 11, , 13, , , , 17, , ...

Takto pokračujeme, dokud nedojdeme k \sqrt{n} . Všimněte si, že nemusíme dělit, jen sčítáme (3, 3 + 3, 6 + 3, 9 + 3, ...), což je mnohem lepší, a ještě hezčí je, že k vyřazení každého neprvočísla jsme potřebovali jen jednu operaci. Je to tedy velmi efektivní metoda.

△

Metoda je to sice pěkná, ale my většinou potřebujeme faktorizovat či testovat na prvočíselnost čísla řádu třeba 10^{200} a vyrábět si kvůli tomu tak velké monstrózní síto by bylo pořád nepředstavitelně drahé. Tak či onak, pokud zkoušíme najít rozklad n -ciferného (n -bitového) čísla pomocí podobných přímočarých metod, tak se v zásadě díváme na náročnost okolo $e^{n/2}$ operací.

Faktorizaci už teď opustíme, poznamenejme jen, že nejpoužívanější veřejné šifrování na Internetu je založeno na tom, že faktorizovat velké číslo by i dnešním superpočítacům zabralo desítky až stovky let. Vrátíme se k tomu v kapitole 7a, v příkladě 7a.m si popovídáme i o dokazování, že nějaké číslo je či není prvočíslem.

Ve zbytku této sekce se podíváme blíže na prvočísla. Začneme jednoduchou odpovědí na otázku, kolik jich je.

Věta 6b.9.

Pročísel je nekonečně mnoho.

Důkaz (poučný): Ukážeme si klasický Euklidův důkaz, takže jdeme až ke starým Řekům, cca 300 př.n.l. Dělá se sporem, předpokládáme, že existuje jen konečně mnoho prvočísel p_1, p_2, \dots, p_m .

Vezměme číslo $a = p_1 \cdot p_2 \cdots p_m + 1$. Podle Faktu 6b.1 existuje prvočíslo p takové, že dělí a . Podle předpokladu to ale musí být jedno z těch p_i , proto také p dělí ten součin $p_1 \cdot p_2 \cdots p_m$. Máme $p | a$, $p | (p_1 \cdot p_2 \cdots p_m)$, proto nutně p dělí jejich rozdíl, viz Důsledek 6a.4 (ii). Takže $p | 1$ a to je spor, neboť $p \geq 2$. \square

Zajímavé je, že ve skutečnosti jsou čísla typu $a = p_1 \cdot p_2 \cdots p_m + 1$ docela často prvočísla, třeba $2 \cdot 3 + 1 = 7$, $2 \cdot 3 \cdot 5 + 1 = 31$ nebo $2 \cdot 3 \cdot 5 \cdot 7 + 1 = 211$. Známá čísla tohoto typu se používají pro testování kvality nových superpočítaců.

Víme, že prvočísel je nekonečně mnoho, což samozřejmě už po staletí nedá lidem spát a snaží se najít co největší. Je to věc prestiže, výsledky jsou kombinací brutální výpočetní síly a vysoce sofistikovaných matematických metod (viz třeba příklad 7a.m). Má to ale i praktické důsledky, třeba pro bezpečnost šifrování. Posledních 300 let byla největší nalezená prvočísla ve tvaru $2^p - 1$ pro prvočíslo p . Samozřejmě ne každé takové číslo je prvočíslem, to by bylo moc snadné, například $2^2 - 1 = 3$, $2^3 - 1 = 7$, $2^5 - 1 = 31$ jsou prvočísla, ale $2^{11} - 1 = 2047 = 23 \cdot 89$. Říká se jim Mersennova prvočísla a jsou populární díky tomu, že výrazy typu $2^p - 1$ se relativně dobře testují na prvočíselnost.

Sice jsme odpověděli na otázku, kolik je prvočísel, ale to byla snadná odpověď. Lepší otázka je, kolik jich je relativně, třeba takto: Označme pro $n \in \mathbb{N}$ jako $\pi(n)$ počet prvočísel p splňujících $p \leq n$ (seznámek 2, 3, 5, 7 ukazuje, že $\pi(10) = 4$, už třeba víme, že $\pi(100) = 25$). Jak rychle tato funkce roste? Hluboké výsledky říkají, že pro velká n je $\pi(n)$ přibližně rovno $\frac{n}{\ln(n)}$. (Má to dost komplikované důkazy, jako hypotéza se to objevilo v 19. století, první důkaz 1896.) Znamená to tedy, že relativní hustota prvočísel mezi prvními n čísly je přibližně $\frac{1}{\ln(n)}$, takže se zvětšujícím se n klesá, tedy prvočísel je čím dál relativně méně.

Tento výsledek se dá interpretovat různými způsoby, třeba takto: Jestliže si mezi čísla 1 až n zvolíte náhodně jedno, tak je pravděpodobnost $\frac{1}{\ln(n)}$, že to bude prvočíslo. Ještě jinak: Tato pravděpodobnost je nepřímo úměrná počtu cifer čísla n (čím více cifer, tím menší pravděpodobnost). Dá se také říct, že průměrná vzdálenost mezi prvočíslily okolo čísla n je zhruba $\ln(n)$.

Přesto je prvočísel v jistém smyslu dost. Připomeňme si známou divergentní harmonickou řadu $\sum_{k=1}^{\infty} \frac{1}{k} = \infty$. Když z té řady vynecháme relativně dost členů (neboli necháme si jich relativně málo), tak začne konvergovat. Pokud například označíme jako M množinu všech druhých mocnin přirozených čísel, tak již $\sum_{k \in M} \frac{1}{k}$ konverguje. Když si ale vezmeme množinu P všech prvočísel, pak $\sum_{p \in P} \frac{1}{p} = \infty$ (Eulerův výsledek). Takže jich zase tak málo není.

Víme, že prvočísla nám dají všechna přirozená čísla prostřednictvím násobení. Zajímavá otázka je, jestli je jich dost na to, aby nám dala přirozená čísla prostřednictvím sčítání. K tomu se váže Goldbachova hypotéza (1742), která říká: Každé liché $n \in \mathbb{N}$ větší než 5 je součtem tří prvočísel. Tato hypotéza má i ekvivalentní vyjádření: Každé sudé $n \in \mathbb{N}$ větší než 2 je součtem dvou prvočísel. Neví se, zda toto platí, zatím je dokázáno, že každé sudé $n \in \mathbb{N}$ větší než 2 je součtem nejvýše 6 prvočísel, což je od cíle dost daleko.

Hodně úsilí šlo do odhalování různých pravidelností ve výskytu prvočísel. Hned na začátku je třeba říct, že celkově v jejich výskytu žádná pravidelnost není. Mohou se ale vyskytovat zajímavé pravidelnosti dočasné, lokální. Několik výsledků:

• Kdykoliv zvolíme a, b nesoudělné, pak v aritmetické posloupnosti $\{an+b\}$ najdeme nekonečně mnoho prvočísel (Dirichletova věta).

• V opačném směru je tu hypotéza: Pro libovolné m existuje $a, d \in \mathbb{N}$ takové, že $a, a+d, a+2d, \dots, a+md$ jsou prvočísla. Krátce řečeno, lze vytvořit konečné aritmetické posloupnosti libovolných délek, které se skládají z prvočísel. Zdá se, že je to pravda, v roce 2004 byl prezentován důkaz, ale byl tak hrozný, že ještě v době psaní tohoto skripta nebyl pořádně prověřen.

Určitě ale víme, že nejde najít nekonečnou aritmetickou posloupnost z prvočísel.

• Pro libovolné $n \in \mathbb{N}$ existuje mezi prvočísly někde mezera délky alespoň n . Ekvivalentně, existuje n po sobě jdoucích čísel takových, že jsou složená. Toto je vlastně snadné, začne se číslem $(n+1)! + 2$ a skončíme $(n+1)! + (n+1)$. Pro každé $2 \leq k \leq n+1$ totiž platí, že k dělí $(n+1)!$, proto jsou pak čísla $(n+1)! + k$ dělitelná k a tedy složená.

• Hodně by pomohlo najít funkci takovou, aby $f(n)$ bylo prvočíslo pro všechna $n \in \mathbb{N}$. Zatím ji nikdo nenašel, i když se zajímavé věci našly, například funkce $f(n) = n^2 - n + 41$, která dává prvočísla pro $n = 1, \dots, 40$, ale pak už ne, $f(41) = 41^2$. Jenže takovéto polynomy jsou stejně slepá ulička, pro každý polynom p s celočíselnými koeficienty existuje $y \in \mathbb{N}$ takové, že $p(y)$ je složené.

I to je vlastně snadné. Nechť $p(x) = a_nx^n + \dots + a_0$. Jestliže $a_0 \neq 0$, pak $f(|a_0|) = a_0(\pm a_n a_0^{n-1} \pm \dots \pm 1)$ a máme číslo složené, v případě $a_0 = 0$ pak $p(x) = x(a_nx^{n-1} + \dots + 1)$ a stačí dosadit jakékoli $a \in \mathbb{N}$.

• Není známo, zda existuje nekonečně mnoho prvočísel typu $n^2 + 1$. Zatím je známo, že pro nekonečně mnoho n je $n^2 + 1$ buď prvočíslo, nebo součin dvou prvočísel.

• Není známo, zda existuje nekonečně mnoho dvojic $p, p+2$, kde obě jsou prvočísla (známe třeba dvojice 3 a 5, 11 a 13, 17 a 19 atd.).

Takto by se dalo pokračovat ještě dlouho, ale jako úvod do teorie čísel to stačí.

Cvičení

Cvičení 6b.1 (rutinní): Najděte faktorizaci následujících čísel: (i) 156; (ii) 165; (iii) 504.

Cvičení 6b.2: Dokažte/vyvraťte: Existují tři po sobě jdoucí lichá čísla, která jsou prvočísla, tj. $p, p+2, p+4$.

Cvičení 6b.3 (drsné): Najděte nějaký předpis používající prvočísla a prvočíselné rozklady pro následující posloupnosti:

(i) 0, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, ...

(ii) 1, 2, 3, 2, 5, 2, 7, 2, 3, 2, 11, 2, 13, 2, ...

(iii) 1, 2, 2, 3, 2, 4, 2, 4, 3, 4, 2, 6, 2, 5, ...

(iv) 1, 2, 3, 3, 5, 5, 7, 7, 7, 7, 11, 11, 13, 13, ...

(v) 1, 2, 6, 30, 210, 2310, 30030, 510510, 9699690, 223092870, ...

Cvičení 6b.4 (poučné): Nechť $a_1, a_2, \dots, a_n \in \mathbb{N}$. Dokažte, že když jsou a_i po dvou nesoudělná, pak $\text{lcm}(a_1, a_2, \dots, a_n) = a_1 \cdot a_2 \cdots a_n$.

Řešení:

6b.1: (i): $156 = 2^2 \cdot 3 \cdot 13$;

(ii): $165 = 3 \cdot 5 \cdot 11$;

(iii): $504 = 2^3 \cdot 3^2 \cdot 7$.

6b.2: 3, 5, 7.

6b.3: (i): prvočíslo ano/ne; (ii): a_n je nejmenší prvočíselný dělitel n ; (iii): počet kladných dělitelů; (iv): a_n je největší prvočíslo $\leq n$; (v): součin prvních $n-1$ prvočísel.

6b.4: Označme $a = a_1 \cdot a_2 \cdots a_n$. Protože $a_i | a$, pak také $\text{lcm}(a_1, a_2, \dots, a_n) | a$. Potřebujeme opačný směr.

Nechť $a = \prod_j p_j^{k_j}$ je prvočíselný rozklad. Zvolme nějaké j . Pak podle Lemma 6b.2 musí existovat i takové, že $p_j | a_i$. Protože jsou všechna a_k po dvou nesoudělná, tak už žádné jiné a_k nemůže mít p_j jako dělitele, tudíž dokonce $p_j^{k_j} | a_i$, tedy i $p_j^{k_j} | \text{lcm}(a_1, a_2, \dots, a_n)$. Ukázali jsme, že všechny prvočíslé faktory $p_j^{k_j}$ z rozkladu a dělí $\text{lcm}(a_1, a_2, \dots, a_n)$, tedy a dělí $\text{lcm}(a_1, a_2, \dots, a_n)$.

6c. Diofantické rovnice

Teď si ukážeme jedno praktické použití toho, co jsme se zatím naučili. V mnoha aplikacích pracujeme zcela přirozeně ve světě celých čísel, což je ale problém, protože naše obvyklé metody řešení rovnic pracují v množině reálných čísel. Ilustruje to následující příklad:

Příklad 6c.a: Představme si letadlo, kde sedadlo v turistické třídě zabírá 6 decimetrů délky trupu, zato sedadlo v business class zabírá 8 dm. Celková délka trupu je 305 dm. My potřebujeme rozdělit letadlo na turistickou a lepší třídu, čili vybrat, kolik řad bude turistických (označme to t) a kolik bude pro business class b . Samozřejmě nechceme plýtvat prostorem, takže dostáváme rovnici $6t + 8b = 305$.

Takovéto rovnice umíme řešit. Víme, že má nekonečně mnoho řešení, snadno se zjistí, že pro libovolnou hodnotu parametru t je dvojice $(t, \frac{305}{8} - \frac{3}{4}t)$ řešením.

My ale těžko budeme do letadla cpát třeba tři čtvrtiny sedadla. Zajímají nás tedy výhradně celočíselná řešení. Nabízí se nápad začít zkoušet, jestli pro nějakou hodnotu t nevyjdou t a b celé, ale zrovna u tohoto příkladu by to trvalo docela dlouho, protože to nikdy nevyjde.

Je tedy jasné, že omezit se na celá čísla může znamenat, že běžné postupy začnou selhávat. Je potřeba vyvinout specializované nástroje.

△

Rovnice, kde máme koeficienty i očekávaná řešení celá, se nazývají diofantické rovnice (Diophantine equations). Jmenují se podle člověka jménem Diofanus z Alexandrie, který je zkoumal ve 3. století, ale najdeme je třeba i ve starých textech indických (už od 800 př.n.l s důležitými aplikacemi v astronomii, okolo roku 800 zase jistý Brahmagupta zkoumal rovnici $x^2 + y^2 = z^2$ neboli pravoúhlé trojúhelníky s celočíselnými stranami). Většinou byly zkoumány izolovaně, ucelená teorie přišla teprve ve 20. století.

Mohou se zkoumat buď rovnice velice obecné, pak se toho nevyzkoumá moc, nebo se omezujeme na pěknější poddruhy. Jednou zkoumanou třídou jsou rovnice polynomální, tedy rovnice ve tvaru $p(x_1, x_2, \dots, x_n) = 0$, kde p je polynom. Dobrým příkladem je rovnice o třech proměnných $x^n + y^n = z^n$, o jejíž celočíselná řešení se zajímal v 17. století Fermat a tvrdil, že pro $n \geq 3$ žádná nejsou, což je ona slavná Fermatova věta. My se zde zaměříme na rovnice typu, který jsme potkali v leteckém příkladě.

! Definice.

Pojmem **lineární diofantická rovnice** označujeme libovolnou rovnici typu $ax + by = c$ s neznámými x, y , kde $a, b, c \in \mathbb{Z}$ a vyžadujeme také řešení $x, y \in \mathbb{Z}$.

Takovéto rovnice nám mohou odpovědět třeba na tyto otázky:

- Lze vyplnit c korun pomocí mincí o hodnotách a a b ?
- Lze vyměřit c litrů pomocí nádob o objemu a a b ?

O řešitelnosti má konečné slovo následující tvrzení.

! Věta 6c.1.

Lineární diofantická rovnice $ax + by = c$ má alespoň jedno řešení právě tehdy, když c je násobkem $\gcd(a, b)$.

Důkaz (poučný): 1) \implies : Předpokládejme, že existují $x, y \in \mathbb{Z}$ takové, že $ax + by = c$. Protože $\gcd(a, b)$ dělí a i b , musí podle Důsledku 6a.4 (i) dělit i c .

2) \iff : Předpokládejme, že $c = k\gcd(a, b)$ pro nějaké $k \in \mathbb{Z}$. Podle Bezoutovy rovnosti existují $A, B \in \mathbb{Z}$ takové, že $Aa + Bb = \gcd(a, b)$. Pak $kAa + kBb = k\gcd(a, b)$ neboli $a(kA) + b(kB) = c$, tedy celá čísla $x = kA$, $y = kB$ řeší $ax + by = c$.

□

Víme tedy, kdy má taková úloha řešení, důkaz nám dokonce dává návod, jak jedno najít pomocí Bezouta.

S Algoritmus 6c.2. pro nalezení nějakého celočíselného řešení x, y rovnice $ax + by = c$.

0. Jestliže c není násobkem $\gcd(a, b)$, tak řešení neexistuje.

1. Jestliže c je násobkem $\gcd(a, b)$, tak například rozšířeným Euklidovým algoritmem najděte $A, B \in \mathbb{Z}$ takové, že $\gcd(a, b) = Aa + Bb$. Když tuto rovnici vynásobíme (celým) číslem $\frac{c}{\gcd(a, b)}$, tak hned vidíme, že čísla $x = A\frac{c}{\gcd(a, b)}$, $y = B\frac{c}{\gcd(a, b)}$ jsou řešením dané rovnice.

△

Příklad 6c.b: Lze vyplnit 1250 korun pomocí mincí o hodnotách 6 a 15? Zajímá nás tedy řešení rovnice $6x + 15y = 1250$. Víme, že $\gcd(6, 15) = 3$, a číslo 1250 není dělitelné třemi, proto to podle Věty nejde.

△

Příklad 6c.c: Lze odměřit 1251 litrů pomocí nádob s objemem 6 a 15 litrů?

Doplňující otázka: Ve kterém filmu se řešila podobná úloha?

Hledáme řešení rovnice $15x + 6y = 1251$, dali jsme si větší číslo jako a , abychom to měli připraveno na rozšířený Euklidův algoritmus. Hned vidíme, že $\gcd(15, 6) = 3$, a protože $\frac{1251}{3} = 417 \in \mathbb{Z}$, je tato úloha řešitelná. Potřebujeme koeficienty Bezoutovy identity.

a, b	q	A	B
15		1	0
6		0	1

a, b	q	A	B
15		1	0
6	2	0	1
3		1	-2

a, b	q	A	B
15		1	0
6	2	0	1
3•	2	1•	-2•
0			

Máme $3 = 1 \cdot 15 + (-2) \cdot 6$. Tuto rovnost vynásobíme číslem 417, u těch součinů opatrně, potřebujeme, aby tam zůstaly koeficienty rovnice 6 a 15. Dostáváme $15 \cdot 417 + 6 \cdot (-834) = 1251$. Porovnáním se zadanou rovnicí vidíme, že $x = 417$ a $y = -834$ je hledané řešení. Takže nejprve do nádrže přidáme 417 krát obsah patnáctilitrové nádoby, pak odebereme 834 krát obsah šestilitrové a zůstane nám 1251 litrů. Z praktického pohledu asi bude lepší nalévání a vybíráni střídat, abychom nepotřebovali nádrž o objemu $417 \cdot 15 = 6225$.

Poznámka: V případě, že vidíme $\gcd(a, b)$ hned a dělí c , tak se může vyplatit celou rovnici tímto číslem pokrátit, čímž samozřejmě vznikne ekvivalentní úloha, která se řeší stejně, ale při výpočtech používáme menší čísla. V našem příkladě bychom z rovnice $15x + 6y = 1251$ dělením trojkou dostali rovnici $5x + 2y = 417$, na kterou pak aplikujeme stejný postup. Rozšířený Euklidův algoritmus aplikovaný na čísla 5, 2 (s \gcd rovným 1) dá rovnost $1 = 1 \cdot 5 + (-2) \cdot 2$, vynásobíme ji číslem 417 a dostaneme $417 = 5 \cdot 417 + 2 \cdot (-834)$, odtud pak máme hledaná řešení $x = 417$ a $y = -834$. Dokonce ani nemusíme používat Euklidův algoritmus, protože rovnost $1 = 1 \cdot 5 + (-2) \cdot 2$ lze snadno uhodnout.

Doplňující odpověď: Die Hard 3 (with a Vengeance). Bruce Willis neznal diofantické rovnice a málem na to doplatil.

△

Toto řešení není úplně uspokojivé, nabízí se otázka, zda náhodou neexistuje i jiné a úspornější, v ideálním případě takové, které by vodu jen nalévalo. U Bezoutovy věty výše jsme zmínili, že možností, jak \gcd vyjádřit pomocí a, b , je obecně hodně, z nich dostáváme více řešení diofantické rovnice. Takže odpověď zní, že ano, máme si z čeho vybírat. Posouváme se tedy k další otázce: Máme-li řešitelnou lineární diofantickou rovnici, chceme určit množinu všech jejích řešení.

Na to se budeme muset hlouběji zamyslet nad tím, jakou má tato množina strukturu. V této chvíli bude lepší dívat se na určité řešení jako na vektor $(x, y) \in \mathbb{Z}^2$, například v případě s vodou bychom mohli napsat, že $(417, -834)$ je řešením dané rovnice. Množina všech řešení je pak určitou podmnožinou postoru \mathbb{Z}^2 .

Čtenáře obeznámeného s lineární algebrou následující pasáže jistě nepřekvapí, protože už něco podobného viděl u soustav lineárních rovnic. Linearita je mocná vlastnost, a jakmile ji máme, spousta věcí už vyplýne.

! Definice.

Je-li dána lineární diofantická rovnice $ax + by = c$, pak definujeme její **přidruženou homogenní rovnici** jako $ax + by = 0$.

Následující tvrzení je klasikou pro lineární rovnice a umožní nám redukovat problematiku hledání všech řešení jen na případ homogenních rovnic.

! Věta 6c.3.

Uvažujme lineární diofantickou rovnici $ax + by = c$. Nechť $(x_p, y_p) \in \mathbb{Z}^2$ je nějaké její řešení.

$(x, y) \in \mathbb{Z}^2$ je řešení této rovnice právě tehdy, když existuje $(x_h, y_h) \in \mathbb{Z}^2$ takové, že $(x, y) = (x_p, y_p) + (x_h, y_h)$ a (x_h, y_h) řeší přidruženou homogenní rovnici.

Důkaz (poučný): Mějme nějaké řešení (x_p, y_p) dané rovnice.

2) \iff : Předpokládejme, že $(x_h, y_h) \in \mathbb{Z}^2$ řeší přidruženou homogenní rovnici. Dosaďme tedy $x = x_p + x_h$ a $y = y_p + y_h$ do dané rovnice, začneme levou stranou a uvidíme, co se z ní vyvrší:

$ax + by = a(x_p + x_h) + b(y_p + y_h) = (ax_p + by_p) + (ax_h + by_h) = c + 0 = c$. Ano, (x, y) je opravdu řešením dané rovnice.

1) \implies : Předpokládejme, že (x, y) řeší danou rovnici. Definujme $x_h = x_p - x$ a $y_h = y_p - y$. Pak $(x, y) = (x_p, y_p) + (x_h, y_h)$ a $(x_h, y_h) \in \mathbb{Z}^2$, zbývá ukázat, že (x_h, y_h) řeší přidruženou homogenní rovnici. Zkusíme dosadit do její levé strany, využijeme pak toho, že (x_p, y_p) i (x, y) jsou řešení:

$ax_h + by_h = a(x_p - x) + b(y_p - y) = (ax_p + by_p) - (ax + by) = c - c = 0$. Ano, (x_h, y_h) řeší přidruženou homogenní rovnici.

□

! Takže jakmile už jedno konkrétní řešení (x_p, y_p) dané diofantické rovnice máme (tomu pak říkáme **partikulární řešení** a umíme ho najít tím algoritmem výše), tak lze množinu všech (celočíselných) řešení získat takto:

$$\{(x_p, y_p) + (x_h, y_h); (x_h, y_h) \in \mathbb{Z}^2 \wedge ax_h + by_h = 0\}.$$

Zbývá vymyslet, jak zcela řešit homogenní rovnice, tedy jak najít všechna celočíselná řešení takových rovnic. To nám prozradí následující věta, ale nejprve si to zkuse rozmyslet. Rovnici $ax + by = 0$ lze přepsat jako $ax = -by$. Snadno pak uhádneme jedno řešení, stačí dát třeba $x = -b$ a $y = a$. Všimněme si také, že jakmile máme jedno řešení (x_0, y_0) , tak získáme další vynásobením čísel x_0, y_0 libovolným celým číslem k , protože toto k lze po dosazení do rovnice $ax = -by$ na obou stranách vykrátit. Řešení tedy bude nekonečně mnoho, například všechna ve tvaru $x = -bk, y = ak$ pro $k \in \mathbb{Z}$. Kritická otázka ovšem je, jestli jsou i jiná.

Představme si na chvíli, že a, b v rovnici $ax = -by$ jsou nesoudělná. Koeficient b dělí levou stranu ax , ale je nesoudělný s a , takže (viz Lemma 6a.23) musí dělit x . Jinými slovy, řešení x hledáme jen mezi násobky b . Obdobně budeme hledat y jen mezi násobky čísla a . Oma řešení z předchozího odstavce tedy budou (pro a, b nesoudělná) jediná možná.

Co když a, b nesoudělná nejsou? Pak jsou i řešení jiná než $(-kb, ka)$. Například u rovnice $4x + 6y = 0$ nám předchozí postup dává řešení ve tvaru $x = -6k, y = 4k$ neboli $(-6k, 4k)$ pro $k \in \mathbb{Z}$, ale vidíme také řešení $x = 3, y = -2$, které nelze volbou $k \in \mathbb{Z}$ získat z toho obecného vzorce. Situace je tedy obecně složitější.

Naštěstí existuje jednoduchý trik: Když rovnici $ax = -by$ vydělíme číslem $\gcd(a, b)$, vznikne ekivalentní rovnice $\frac{a}{\gcd(a, b)}x = -\frac{b}{\gcd(a, b)}y$, která už má (celé) nesoudělné kořeny a tudíž dokážeme najít všechna řešení postupem předvedeným výše. Dostaneme se tak k následujícím tvrzení.

! Věta 6c.4.

Uvažujme rovnici $ax + by = 0$ pro $a, b \in \mathbb{Z}$. Pak množina všech jejích celočíselných řešení je

$$\left\{ \left(k \frac{b}{\gcd(a, b)}, -k \frac{a}{\gcd(a, b)} \right); k \in \mathbb{Z} \right\}.$$

Důkaz (poučný): 1) Nejprve ověříme, že dvojice $x = k \frac{b}{\gcd(a, b)}, y = -k \frac{a}{\gcd(a, b)}$ jsou opravdu řešení ze \mathbb{Z} .

Celočíselnost plyne z toho, že $\frac{b}{\gcd(a, b)}$ a $\frac{a}{\gcd(a, b)}$ jsou celá čísla, po dosazení x, y do rovnice pak okamžitě dostáváme $ax + by = ak \frac{b}{\gcd(a, b)} - bk \frac{a}{\gcd(a, b)} = 0$. Takže to souhlasí.

2) Zbývá ukázat, že řešení daná tímto předpisem jsou všechna, tj. že žádné jiné neexistuje. Nechť je tedy x, y nějaké řešení rovnice $ax + by = 0$. Vydělíme ji číslem $\gcd(a, b)$ a převedeme jeden člen na druhou stranu:

$\frac{b}{\gcd(a, b)}y = -\frac{a}{\gcd(a, b)}x$. Vidíme, že celé číslo $\frac{b}{\gcd(a, b)}$ musí dělit $\frac{a}{\gcd(a, b)}x$, jenž podle Faktu 6a.9 jsou $\frac{b}{\gcd(a, b)}$ a $\frac{a}{\gcd(a, b)}$ nesoudělná čísla, tudíž musí podle Lemma 6a.23 číslo $\frac{b}{\gcd(a, b)}$ dělit x . Existuje tedy $k \in \mathbb{Z}$ takové, že $x = k \frac{b}{\gcd(a, b)}$, z rovnice $by = -ax$ pak snadno dostaneme příslušný vzorec pro y .

□

Pozorný čtenář si jistě všimnul, že ve znění věty je znaménko mínus u y , zatímco v úvahách před větou bylo u x . Vysvětlení je snadné, nás zajímá množina všech řešení a tam na umístění znaménka nesejde. Pokud bychom například při řešení rovnice $3x + 4y = 0$ použili intuitivní přístup předvedený před Větou, dostaneme množinu $x = -4k, y = 3k$ pro $k \in \mathbb{Z}$. Volbou $k = 3$ pak dostaneme konkrétní řešení $x = -12, y = 9$ neboli $(-12, 9)$. Pokud bychom použili vzorec $(4k, -3k)$ z věty, pak totéž řešení dostaneme volbou $k = -3$. To ukazuje, proč v případě, že necháme k proběhnout všemi celými čísly, dostáváme nakonec v obou případech shodné množiny dvojic. V praxi většinou volíme tu variantu, která nám přijde příjemnější, například u rovnice $10x - 15y$ dostáváme buď řešení ve tvaru $(3k, 2k)$, $k \in \mathbb{Z}$ nebo ve tvaru $(-3k, -2k)$, $k \in \mathbb{Z}$, první se mi líbí více.

Podobně podle situace volíme i zápis. V teoretických úvahách je lepší pracovat s vektory, u praktických úloh bývá často příjemnější používat zápis $x = \dots, y = \dots$

Ted' už umíme řešit homogenní lineární diofantické rovnice. Když shrneme naše dosavadní výsledky, dostáváme následující.

!

Důsledek 6c.5.

Uvažujme lineární diofantickou rovnici $ax + by = c$. Předpokládejme, že c je násobkem $\gcd(a, b)$. Nechť $A, B \in \mathbb{Z}$ splňují $\gcd(a, b) = Aa + Bb$. Pak množina všech řešení dané rovnice je

$$\left\{ \left(A \frac{c}{\gcd(a, b)} + k \frac{b}{\gcd(a, b)}, B \frac{c}{\gcd(a, b)} - k \frac{a}{\gcd(a, b)} \right); k \in \mathbb{Z} \right\}.$$

Příklad 6c.d (pokračování 6c.c): Řešili jsme rovnici $15x + 6y = 1251$ a zjistili jsme, že $\gcd(15, 6) = 3 = 1 \cdot 15 + (-2) \cdot 6$, což dělí pravou stranu $c = 1251$. Podle Důsledku máme množinu řešení

$$\left\{ (1 \cdot 417 - k \frac{6}{3}, (-2) \cdot 417 + k \frac{15}{3}); k \in \mathbb{Z} \right\} = \left\{ (417 - 2k, 5k - 834); k \in \mathbb{Z} \right\}$$

neboli $x = 417 - 2k$, $y = 5k - 834$ pro $k \in \mathbb{Z}$ (znaménko u k jsme umístili tak, aby u každé neznámé byl alespoň jeden kladný člen, protože nám to tak přišlo hezčí, ale klidně si to udělejte jinak).

Můžeme udělat zkoušku, jako obvykle dosazením do rovnice:

$$15 \cdot (417 - 2k) + 6 \cdot (5k - 834) = 6255 - 30k + 30k - 5004 = 1251.$$

Vyšla.

Jestliže nás zajímají řešení z oboru \mathbb{N}_0 , tak potřebujeme, aby $5k - 834 \geq 0$ a $417 - 2k \geq 0$ neboli $k \geq \frac{834}{5}$ a $k \leq \frac{417}{2}$. Taková k existují, jmenovitě jde o všechna $k \in \mathbb{N}$ splňující $167 \leq k \leq 208$. Mohli bychom si tedy vypsat všechna řešení $(x, y) \in \mathbb{N}_0^2$, ale je jich docela dost, tak se spokojíme s jedním. Zvolíme třeba $k = 200$ a vidíme, že 1251 litrů získáme například tak, že do nádrže nalejeme 17 patnáctilitrových nádob a 166 šestilitrových.

△

S Praktický výpočet. Při praktickém výpočtu je možné postupovat více způsoby, jedna možnost je použít právě předvedený postup, kdy použijeme výsledný vzorec pro množinu všech řešení. Na tom není nic špatného, má ale jednu nevýhodu: Je zcela závislý na zapamatování vzorce, který je v látce poněkud izolovaný a tudíž se snadno zapomene. Mnoho lidí dává přednost postupu vícekrokovému, jehož hlavní struktura vychází z obecné teorie lineárních rovnic, je to tedy postup, který se zde ještě několikrát objeví a čtenář jej již může znát z lineární algebry. Jmenovitě, nejprve se najde partikulární řešení dané rovnice a pak se vyřeší homogenní rovnice. Tento postup vychází z pochopení fungování rovnic, takže pokud čtenář látku zná, v zásadě si už skoro nic navíc nemusí pamatovat. Oblíbeným trikem, který se může a nemusí použít, je také vykrácení rovnice co nejdříve, zejména pokud dokážeme $\gcd(a, b)$ hned uhodnout.

Možných přístupů je tedy několik, čtenář si vybere dle toho, v jakém bodě na škále mezi pamatováním a porozuměním se nejlépe cítí. My zde uvedeme (a budeme používat) algoritmus přemýšlecí a strukturovaný, který autorovi přijde nejlepší.

S Algoritmus 6c.6. pro nalezení všech celočíselných řešení rovnice $ax + by = c$.

Verze 1 bez hádání.

0. Pomocí například rozšířeného Euklidova algoritmu najděte $\gcd(a, b) = Aa + Bb$.

1. Jestliže c není násobkem $\gcd(a, b)$, pak řešení rovnice neexistuje.

2. Případ $\gcd(a, b)$ dělí c :

a) Získanou rovnici $aA + bB = \gcd(a, b)$ vynásobte číslem $c' = \frac{c}{\gcd(a, b)}$ tak, aby se zachovaly koeficienty a, b ,

a dostanete $a(Ac') + b(Bc') = c$, tudíž i jedno partikulární řešení $x_p = Ac'$, $y_p = Bc'$ neboli vektor (Ac', Bc') .

b) Přidruženou homogenní rovnici $ax + by = 0$ zkraťte číslem $\gcd(a, b)$ na tvar $a'x + b'y = 0$ neboli $a'x = -b'y$, což dává řešení $x_h = -b'k$, $y_h = a'k$ neboli dvojice $(-b'k, a'k)$ pro $k \in \mathbb{Z}$.

c) Sečtením partikulárního a obecného homogenního řešení získáte množinu všech celočíselných řešení

$$\{(Ac' - kb', Bc' + ka'); k \in \mathbb{Z}\} \text{ neboli } x = Ac' - kb', y = Bc' + ka' \text{ pro } k \in \mathbb{Z}.$$

Verze 2 s hádáním.

1. Uhodněte $\gcd(a, b)$ a danou rovnici zkraťte tímto číslem. Pokud to nejde, tedy jestliže c není násobkem $\gcd(a, b)$, pak řešení rovnice neexistuje.

2. Případ $\gcd(a, b)$ dělí c :

Vydělte danou rovnici číslem $\gcd(a, b)$, dostanete novou diofantickou rovnici $a'x + b'y = c'$, kde teď a', b' jsou nesoudělné.

a) Pomocí například rozšířeného Euklidova algoritmu najděte $1 = \gcd(a', b') = Aa' + Bb'$. Získanou rovnici $a'A + b'B = 1$ vynásobte číslem c' tak, aby se zachovaly koeficienty, dostanete $a(Ac') + b(Bc') = c'$ a tudíž i jedno partikulární řešení $x_p = Ac'$, $y_p = Bc'$ neboli vektor (Ac', Bc') .

b) Přidruženou homogenní rovnici $a'x + b'y = 0$ si přepište jako $a'x = -b'y$, což napoví řešení $x_h = -b'k$, $y_h = a'k$ neboli dvojice $(-b'k, a'k)$ pro $k \in \mathbb{Z}$.

c) Sečtením partikulárního a obecného homogenního řešení získáte množinu všech celočíselných řešení

$$\{(Ac' - kb', Bc' + ka'); k \in \mathbb{Z}\} \text{ neboli } x = Ac' - kb', y = Bc' + ka' \text{ pro } k \in \mathbb{Z}.$$

U obou verzí je možné dát míinus k y_h namísto k x_h .

△

Někteří studenti u homogenní rovnice $a'x + b'y = 0$ už žádné úpravy nedělají a prostě si pamatují, že prohozené koeficienty (s jedním mínusem) dávají řešení.

Pokud chceme získat řešení z \mathbb{N}_0^2 a máme $a', b' > 0$, pak k musí splňovat podmínky $-\frac{Ac'}{b'} \leq k \leq \frac{Bc'}{a'}$.

I tento algoritmus si ukážeme na úloze s nádobami.

! Příklad 6c.e (pokračování 6c.c): Rovnici $15x + 6y = 1251$ vyřešíme strukturovaně. Číslo $\gcd(15, 6) = 3$ umíme uhádnout, třeba tak, že 6 má netriviální dělitele 2 a 3, jen jeden z nich dělí i 15. Rovnici tedy vydělíme číslem 3 a dostaváme rovnici $5x + 2y = 417$, dělení proběhlo bez problémů a tudíž je rovnice řešitelná v celočíselném oboru.

Snadno pomocí rozšířeného Euklidova algoritmu najdeme popřípadě uhodneme, že $\gcd(5, 2) = 1 = 1 \cdot 5 + (-2) \cdot 2$. Rovnici $1 = 5 \cdot 1 + 2 \cdot (-2)$ vynásobíme číslem 417 a dostaneme $5 \cdot 417 + 2 \cdot (-834) = 417$, máme tedy partikulární řešení $x_p = 417$, $y_p = -834$.

Přidružená homogenní rovnice $5x + 2y = 0$ neboli $5x = -2y$ má obecné řešení $x_h = -2k$, $y_h = 5k$ pro $k \in \mathbb{Z}$.

Sečtením dostaváme obecné celočíselné řešení $x = 417 - 2k$, $y = 5k - 834$ pro $k \in \mathbb{Z}$.

Již jsme odvodili, že pokud bychom chtěli řešení jen z \mathbb{N}_0 , tak použijeme $167 \leq k \leq 208$. Je možné zkusit i další optimalizaci.

Představme si například, že pro nás není váhový rozdíl mezi 15 a 6 litry až tak velký, ale vadí nám běhání pro vodu. Ocenili bychom řešení, u kterého běháme nejméně, což znamená řešení s co nejmenším počtem použitých nádob. Matematicky to znamená, že chceme minimalizovat $x + y = (417 - 2k) + (5k - 834) = 3k - 417$, ale zajímají nás jen hodnoty k mezi 167 a 208. Řešením je evidentně volba co nejmenšího možného k , tedy $k = 167$. Nejméně se naběháme, pokud použijeme $x = 83$ patnáctilitrové a jednu šestilitrovku.

△

Zdá se, že je to takto příjemnější než řešení předchozí, ale rozhodnutí, který algoritmus používat, je samozřejmě na čtenáři. Na závěr dva příklady, první příjemný a druhý standardní písemkový.

! Příklad 6c.f: Dostali jste stokorunu s tím, že za ni máte nakoupit lízátka a bonbóny na dětský den. Lízátka stojí šest korun a bonbón dvě koruny. Jaké se nabízí možnosti, jestliže si nechcete nechat nic od cesty ani nákup dotovat ze svého?

Máme hledat řešení rovnice $6l + 2b = 100$ z oboru \mathbb{N}_0 . Protože hned vidíme, že $\gcd(6, 2) = 2$, vydělíme rovnici, ono to jde, budou tedy řešení.

Řešíme rovnici $3l + b = 50$. K nalezení partikulárního řešení potřebujeme najít vyjádření 1 pomocí 3 a 1, tedy $1 = 3A + B$. To snadno uhádneme, $3 \cdot 1 + 1 \cdot (-2) = 1$. Tuto rovnici vynásobíme padesáti, ať máme správnou pravou stranu. Dáme si přitom pozor, abychom na levé straně dali padesát na správná místa, potřebujeme zachovat koeficienty 3 a 1: $3 \cdot 50 + 1 \cdot (-100) = 50$, tedy vidíme řešení $l_p = 50$, $b_p = -100$.

Přidružená homogenní rovnice $3l + b = 0$ neboli $3l = -b$ má obecné řešení $l_h = -k$, $b_h = 3k$, dostaváme tak obecné celočíselné řešení pro danou úlohu jako $l = 50 - k$, $b = 3k - 100$ pro $k \in \mathbb{Z}$.

Která řešení splňuje $l, b \geq 0$? Dostaváme rovnice $50 \geq k$ a $3k \geq 100$, což dává rozmezí $34 \leq k \leq 50$. Máme tedy řešení $(16, 2), (15, 5), (14, 8), (13, 11), (12, 14), (11, 17), (10, 20), (9, 23), (8, 26), (7, 29), (6, 32), (5, 35), (4, 38), (3, 41), (2, 44), (1, 47), (0, 50)$.

Na řešeních je možné dělat další analýzy, například pokud bychom chtěli, aby bylo lízátek a bonbónů stejně, řešili bychom rovnici $3k - 100 = 50 - k$ neboli $4k = 150$, což dává $k = 37.5$, takže hodnoty $k = 37$ a $k = 38$ dávají výsledky nejblíže rovnováze, dostaváme $l = 13$, $b = 11$, případně $l = 12$, $b = 14$. To druhé asi bude lepší, protože bude víc sladkostí.

Mohl bychom také chtít koupit co nejvíce věcí, takže bychom maximalizovali funkci $s(k) = x + y = 2k - 50$ na intervalu $(34, 50)$. Funkce s rostoucím k roste, její největší hodnotu tedy dostaneme v co největším možném k . Nejvíce sladkostí dává volba $k = 50$, tedy žádná lízátka a 50 bonbónů.

Poznámka: Víme, že Bezoutův rozklad není jednoznačný, někdo třeba uhodne $1 = 3 \cdot (-1) + 1 \cdot 4$. Pak bychom dostali $3 \cdot (-50) + 1 \cdot 200 = 50$ a řešení $x = -50 - k$, $y = 200 + 3k$ pro $k \in \mathbb{Z}$. Zde by asi bylo lepší volit v homogenním řešení znaménka naopak, řešení je pak elegantnější, $x = k - 50$, $y = 200 - 3k$ pro $k \in \mathbb{Z}$. Jsou to sice jiné vzorečky, ale jako množinu všech řešení dostaváme totéž. Například vyrovnanou variantu 12 lízátek a 14 bonbónů jsme předtím dostali volbou $k = 38$, u nových vzorečků ji dostaneme volbou $k = 62$.

To se samořejmě dalo čekat, věta o struktuře řešení (viz 6c.3 a poznámka za ní) zaručuje, že při vytváření množiny všech řešení lze to partikulární volit libovolně.

△

! Příklad 6c.g:

Vyřešíme rovnici $154x - 259y = 105$.

Asi by se $\gcd(154, -259)$ dal i uhádnout, ale zopakujeme si Euklidův algoritmus na číslech 259 a 154.

a, b	q	A	B
259		1	0
154	1	0	1
105	1	1	-1
49	2	-1	2
7•	7	3•	-5•
0			

Dostáváme $\gcd(154, -259) = \gcd(259, 154) = 7$, takže podělíme rovnici: $22x - 37y = 15$. Šlo to, rovnice je řešitelná.

Dostali jsme také vyjádření $7 = 3 \cdot 259 + (-5) \cdot 154$. Abychom měli na levé straně 105, vynásobíme tuto identitu patnácti, na pravé straně si chceme zachovat čísla 154 a 259: $105 = 45 \cdot 259 + (-75) \cdot 154$. Ještě si tuto rovnost přeorganizujeme, aby souhlasila znaménka a pořadí: $154 \cdot (-75) - 259 \cdot (-45) = 105$. Porovnáním s danou rovnicí vidíme řešení $x_p = -75$, $y_p = -45$.

Vykrácenou rovnici upravíme na homogenní: $22x - 37y = 0$. Dostáváme řešení $x_h = 37k$, $y_h = 22k$ pro $k \in \mathbb{Z}$.

Sečtením partikulárního a homogenních řešení dostáváme obecné řešení dané rovnice $x = 37k - 75$, $y = 22k - 45$ pro $k \in \mathbb{Z}$.

Kdyby to někdo chtěl množinově, tak množina všech řešení dané rovnice je

$$\{(37k - 75, 22k - 45); k \in \mathbb{Z}\}.$$

Zkouška: Dosadíme do dané rovnice: $154 \cdot (37k - 75) - 259 \cdot (22k - 45) = 5698k - 11550 - 5698k + 11655 = 105$.

Pokud bychom chtěli řešení z \mathbb{N}_0 , dostali bychom jich nekonečně mnoho, $x = 37k - 75$, $y = 22k - 45$ pro $k \in \mathbb{Z}$, $k \geq 3$.

Poznámka: Pokud někdo postupu dobře rozumí, může se od něj někdy výhodně odchýlit. Podívejte se na třetí řádek v tabulce algoritmu. Říká se tam, že $105 = 1 \cdot 259 + (-1) \cdot 154$. Takže přímo v tabulce vidíme kombinaci koeficientů, která dává původní pravou stranu. Po úpravě $154 \cdot (-1) - 259 \cdot (-1) = 105$ a hned vidíme alternativní partikulární řešení $x_p = -1$, $y_p = -1$.

△

Cvičení

Cvičení 6c.1 (rutinní, zkouškové): Najděte všechna řešení $(x, y) \in \mathbb{Z}^2$ a $(x, y) \in \mathbb{N}_0^2$ pro následující diofantické rovnice:

- | | | |
|--------------------------|-------------------------|---------------------------|
| (i) $6x + 9y = 204$; | (iii) $10x - 4y = 26$; | (v) $819x + 315y = 126$; |
| (ii) $10x - 15y = 131$; | (iv) $105x - 75y = 0$; | (vi) $65x + 273y = 157$. |

Cvičení 6c.2 (dobré): Máte k dispozici klasické váhy s dvěma miskami a libovolný počet závaží o váze 15 nebo 55 gramů. Jakou nejmenší hmotnost jste schopni odvážit?

Cvičení 6c.3 (dobré): Máte dvě tyče, jedna má délku 60 dm a druhá má délku 25 dm. Jaká je nejmenší délka látky, kterou pomocí nich dokážete odměřit, pokud si odměřujete podél okraje a děláte čárky?

Řešení:

6c.1: (i): $\gcd(6, 9) = 3$ uhodneme, řešíme $2x + 3y = 68$: $\gcd(3, 2) = 1 = 1 \cdot 3 + (-1) \cdot 2$, proto $\gcd(2, 3) = 1 = (-1) \cdot 2 + 1 \cdot 3$, po vynásobení $2 \cdot (-68) + 3 \cdot 68 = 68$. Řešení $(x, y) = (-68 + 3k, 68 - 2k)$ neboli $x = 3k - 68$, $y = 68 - 2k$ pro $k \in \mathbb{Z}$. Řešení v \mathbb{N}_0 : $23 \leq k \leq 34$.

(ii): $\gcd(10, -15) = 5$ nedělí 131. Nemá řešení.

(iii): $\gcd(10, -4) = 2$ uhodneme, řešíme $5x - 2y = 13$: $\gcd(5, 2) = 1 = 1 \cdot 5 + (-2) \cdot 2$, proto $\gcd(5, -2) = 1 = 1 \cdot 5 + 2 \cdot (-2)$, po vynásobení $5 \cdot 13 - 2 \cdot 26 = 13$. Řešení $(x, y) = (13 - (-2)k, 26 + 5k)$ neboli $x = 2k + 13$, $y = 5k + 26$ pro $k \in \mathbb{Z}$. Řešení v \mathbb{N}_0 : $k \geq -5$.

(iv): $\gcd(105, 75) = 15$, $7x - 5y = 0$ homogenní rovnice. Řešení $(x, y) = (5k, 7k)$ neboli $x = 5k$, $y = 7k$ pro $k \in \mathbb{Z}$. Řešení v \mathbb{N}_0 : $k \geq 0$.

(v): $\gcd(819, 315) = 63$, $13x + 5y = 2$. $\gcd(819, 315) = 63 = 2 \cdot 819 + (-5) \cdot 315$, proto $819 \cdot 4 + 315 \cdot (-10) = 126$. Řešení $(x, y) = (4 - 5k, -10 + 13k)$ neboli $x = 4 - 5k$, $y = 13k - 10$ pro $k \in \mathbb{Z}$. Řešení v \mathbb{N}_0 : nelze.

(vi): $\gcd(65, 273) = 13$ nedělí 157. Nemá řešení.

6c.2: Váhu c odměříme, pokud lze napsat $c = 15x + 55y$, kde $x, y \in \mathbb{Z}$ a záporné hodnoty znamenají, že takováto závaží dáváme na stejnou misku jako dotyčný předmět. Rovnice má řešení, pokud $\gcd(15, 55)$ dělí c , tedy nejmenší váha je 5 gramů.

6c.3: Délku c odměříme, pokud lze napsat $c = 60x + 25y$, kde $x, y \in \mathbb{Z}$ a záporné hodnoty znamenají, že nanášíme na opačnou stranu. Rovnice má řešení, pokud $\gcd(60, 25)$ dělí c , tedy nejmenší délka je 5 dm.

7. Počítání modulo

V této kapitole se podíváme na téma, bez kterého se neobejde žádná diskuse o fungování počítačů, nakonec skončíme u Internetu. Tato látka je přirozené pokračování kapitoly 6.

7a. Kongruence, počítání modulo

V mnoha aplikacích se omezujeme na malou množinu čísel a při vyskočení se do ní vracíme cyklicky, tak jak to děláme běžně u hodin. Zde se na to podíváme pořádně a matematicky.

! Definice.

Nechť $n \in \mathbb{N}$. Řekneme, že čísla $a, b \in \mathbb{Z}$ jsou **kongruentní modulo n** , značeno $a \equiv b \pmod{n}$, jestliže $n | (a - b)$.

Let $n \in \mathbb{N}$. We say that numbers $a, b \in \mathbb{Z}$ are **congruent modulo n** , denoted $a \equiv b \pmod{n}$, if $n | (a - b)$.

Příklad 7a.a: Z hodin víme, že $21 \equiv 9 \pmod{12}$. Zkouška podle definice: $21 - 9 = 12$, což je dělitelné dvanácti. Jiný příklad: $(-2) \equiv 13 \pmod{5}$, protože $(-2) - 13 = -15$, což je dělitelné pěti.

△

Někdy se hodí poznávat kongruenci jinak než podle definice.

! Věta 7a.1.

Nechť $n \in \mathbb{N}$. Pro čísla $a, b \in \mathbb{Z}$ jsou následující podmínky ekvivalentní:

- (i) $a \equiv b \pmod{n}$,
- (ii) existuje $k \in \mathbb{Z}$ takové, že $a = b + kn$,
- (iii) $a \text{ mod } n = b \text{ mod } n$, tj. jsou si rovny zbytky po dělení číslem n .

Důkaz (rutinní, poučný): (i) \Rightarrow (ii): Jestliže $a \equiv b \pmod{n}$, pak $n | (a - b)$. Proto existuje $k \in \mathbb{Z}$: $(a - b) = kn$, tedy $a = b + kn$.

(ii) \Rightarrow (iii): Předpokládejme, že $a = b + kn$ pro nějaké $k \in \mathbb{Z}$. Nechť $r = b \text{ mod } n$ (zbytek po dělení), tedy máme rozklad $b = qn + r$ splňující $q \in \mathbb{Z}$ a $0 \leq r < n$. Pak $a = b + kn = (q + k)n + r$, kde $(q + k) \in \mathbb{Z}$ a $0 \leq r < n$, proto jde o rozklad z věty o dělení a $r = a \text{ mod } n$.

(iii) \Rightarrow (i): Nechť $a \text{ mod } n = b \text{ mod } n = r$. Pak existují $p, q \in \mathbb{Z}$ takové, že $a = pn + r$ a $b = qn + r$. Odtud $b - a = (q - p)n$ a $q - p \in \mathbb{Z}$, tedy $n | (b - a)$, což podle definice znamená $a \equiv b \pmod{n}$.

Uzavřeli jsme kruh, proto je libovolné z tvrzení (i) až (iii) ekvivalentní s libovolným jiným.

□

Zejména podmínka (ii) je příjemná pro rychlé počítání s malými čísly. Říká, že $a \equiv b \pmod{n}$, jestliže se od a k b (či naopak) dokážeme dostat opakováním přičítáním/odčítáním čísla n .

! Příklad 7a.b:

Tvrdíme, že $21 \equiv 9 \pmod{6}$. Podle definice máme ověřit, že 6 dělí $21 - 9 = 12$, což tedy platí.

Podle podmínky (ii) to vidíme také, dvojím přičtením 6 k 9 dostaneme 21. I zbytky po dělení hravě spočítáme, $21 \text{ mod } 6 = 3$ a $9 \text{ mod } 6 = 3$ se rovnají a podmínka (iii) dává $21 \equiv 9 \pmod{12}$.

Podmínka (ii) bývá pro mnohé pohodlná, když dojde na záporná čísla, nemusí si totikdávat pozor na znaménka. Například dvojím odečtením trojky od -68 dostaneme -74 , proto určitě $-68 \equiv -74 \pmod{3}$, mnoha lidem to přijde pohodlnější než odečítat $(-68) - (-74)$.

Podmínka (iii) se hodí v případech, kdy zbytky po dělení n vidíme hned, což je zejména případ $n = 5$. Například $37 \text{ mod } 5 = 2$ a $12 \text{ mod } 5 = 2$, proto určitě $37 \equiv 12 \pmod{5}$.

Jakmile si na tohle čtenář zvykne, tak hned vidí, že ve světě počítání modulo 5 je $3 \equiv 8 \equiv 13 \equiv 18 \equiv 23 \equiv \dots$ a také $3 \equiv -2 \equiv -7 \equiv -12 \equiv -17 \equiv \dots$

△

Ze třetí podmínky okamžitě dostáváme následující.

! Fakt 7a.2.

Nechť $n \in \mathbb{N}$. Pak platí:

- (i) Pro každé $c \in \mathbb{Z}$ je $c \equiv c \pmod{n}$.
- (ii) Nechť $a \in \mathbb{Z}$. $a \equiv 0 \pmod{n}$ právě tehdy, když n dělí a .

Důkaz necháváme jako cvičení.

Kongruence splňuje mnoho vlastností, které nám usnadňují práci. Přenesme se tedy do světa, kde vše funguje modulo nějaké konkrétní n . Všechna čísla se tam rozpadnou do skupin podle toho, které je s kterým kongruentní. Například ve světě modulo 2 to bude skupina zahrnující čísla $0, 2, -2, 4, -4, \dots$ (všechna jsou navzájem kongruentní, lze se mezi nimi přesouvat přičítáním/odčítáním dvojky) a druhá skupina zahrnující čísla $1, -1, 3, -3, 5, -5, \dots$ (i mezi těmi se lze přesunovat přičítáním/odčítáním dvojky). Formálně si to zavedeme později, nejprve si o takových skupinách něco ukážeme.

Důležité na těch skupinách je, že se v rámci jedné skupiny čísla mohou navzájem zastupovat. Jinak řečeno, pokud máme nějaké číslo a ono se nám nelibí (třeba je moc velké a nám se s ním nechce počítat), tak si jej ve světě modulo můžeme v mnoha situacích nahradit libovolným jiným číslem z jeho skupiny a nic tím neovlivníme. To je klíčové tvrzení, které si zaslouží přesnější vyjádření a také důkaz.

Věta 7a.3.

Nechť $n \in \mathbb{N}$, uvažujme $a, b, u, v \in \mathbb{Z}$ takové, že $a \equiv u \pmod{n}$ a $b \equiv v \pmod{n}$. Pak platí následující:

- (i) $a + b \equiv u + v \pmod{n}$;
- (ii) $a - b \equiv u - v \pmod{n}$;
- (iii) $ab \equiv uv \pmod{n}$.

Důkaz (poučný): (iii): Podle předpokladu a Věty platí $a = u + kn$ a $b = v + ln$ pro nějaká $k, l \in \mathbb{Z}$. Pak máme také $ab = uv + uln + vkn + kln^2 = (uv) + (ul + vk + kln)n$ a $(ul + vk + kln) \in \mathbb{Z}$, závér zase plyne z dotyčné Věty.

Důkazy (i) a (ii) necháme jako cvičení, jsou obdobné. □

! Díky této větě například ve světě modula 5 můžeme místo výpočtu $195376 \cdot 16239 + 32532675$ počítat $1 \cdot 4 + 0 = 4$ a výsledky se budou (modulo 5) rovnat. Zástupce jsme přitom našli velice snadno, již totiž víme, že se čísla ve skupinách modulo poznají podle zbytků a zbytek po dělení pěti určíme hravě podle poslední cifry v čísle.

Samozřejmě jsme mohli použít i jiné zástupce, například počítat $(-4) \cdot 114 + 290$, ale proč pracovat, když nemusíme. Potvrďme si tedy obecně, že ve světě modulo n můžeme každé číslo nahradit tím nejjednodušším kandidátem, tedy zbytkem po dělení.

Fakt 7a.4.

Nechť $n \in \mathbb{N}$, uvažujme $a \in \mathbb{Z}$. Jestliže $r = a \bmod n$, tedy r je zbytek po dělení a číslem n , pak $a \equiv r \pmod{n}$.

Důkaz (rutinní): Zbytek splňuje $a = qn + r$ pro jisté $q \in \mathbb{Z}$, proto $a - r = qn$, tedy n dělí $a - r$. □

Poznamenejme nicméně, že ne vždy je zrovna zbytek po dělení ten nejlepší zástupce. Pokud potřebujeme spočítat $398 \cdot 1243$ modulo 100, pak přechod ke zbytkům dá $98 \cdot 43$, ale my určitě dáme přednost výpočtu $(-2) \cdot 43 = -86 \equiv 14 \pmod{100}$.

Čtenář jistě ví, že odčítání je vlastně přičítání opačného čísla, takže jsme vlastně ani nemuseli dokazovat speciální pravidlo (ii), stačilo by (i) a (iii). Ještě se k tomuto tématu vrátíme. Věta naopak neřešila komplikovanější algebraické výpočty, ale k těm se snadno dostaneme, protože je stejně vždy děláme postupně podle priorit. Můžeme tak princip zastupování rozšířit i na složitější výrazy standardním způsobem, například případ sčítání více čísel se jistě bude dělat indukcí.

Důsledek 7a.5.

Nechť $n \in \mathbb{Z}$.

- (i) Uvažujme $a_1, u_1, \dots, a_m, u_m \in \mathbb{Z}$ takové, že $a_i \equiv u_i \pmod{n}$ pro všechna $i = 1, \dots, m$.

Pak $\sum_{i=1}^m a_i \equiv \sum_{i=1}^m u_i \pmod{n}$ a $\prod_{i=1}^m a_i \equiv \prod_{i=1}^m u_i \pmod{n}$.

- (ii) Uvažujme $a_1, b_1, u_1, v_1, \dots, a_m, b_m, u_m, v_m \in \mathbb{Z}$ takové, že $a_i \equiv u_i \pmod{n}$ a $b_i \equiv v_i \pmod{n}$ pro všechna $i = 1, \dots, m$. Pak $\sum_{i=1}^m a_i b_i \equiv \sum_{i=1}^m u_i v_i \pmod{n}$.

Důkaz (rutinní): (i): Dokážeme to indukcí na m pro sčítání, násobení necháme jako cvičení.

(0) $m = 1$: Předpoklad $a_1 \equiv b_1 \pmod{n}$ je zároveň závěrem, tedy platí.

(1) Předpokládejme, že sčítací vzorec platí pro nějaké $m \in \mathbb{N}$ a všechna a_i, u_i . Mějme čísla $a_1, u_1, \dots, a_{m+1}, u_{m+1}$ splňující $a_i \equiv u_i \pmod{n}$ pro všechna i . Podle indukčního předpokladu pak máme $\sum_{i=1}^m a_i \equiv \sum_{i=1}^m u_i \pmod{n}$, proto podle Věty (i) také $\left(\sum_{i=1}^m a_i\right) + a_{m+1} \equiv \left(\sum_{i=1}^m u_i\right) + u_{m+1} \pmod{n}$ neboli $\sum_{i=1}^{m+1} a_i \equiv \sum_{i=1}^{m+1} u_i \pmod{n}$, důkaz je hotov.

(ii): Podle Věty (iii) platí $a_i b_i \equiv u_i v_i \pmod{n}$ pro všechna i , na tyto čísla pak aplikujeme část (i) a sečteme je. \square

Stručně řečeno, v jakémkoliv algebraickém výrazu poskládaném ze sčítání (odčítání) a násobení, případně závorek lze zúčastněná čísla nahrazovat. Doplňme ještě jedno užitečné výpočetní pravidlo.

Fakt 7a.6.

Nechť $n \in \mathbb{N}$, uvažujme $a, u \in \mathbb{Z}$ takové, že $a \equiv u \pmod{n}$. Pak pro všechna $k \in \mathbb{N}$ platí $a^k \equiv u^k \pmod{n}$.

Důkaz (poučný): Protože $a^k = a \cdot a \cdots a$, plyne to hned z (i) v Důsledku výše. Pro zajímavost ukážeme ještě jeden důkaz pro případ $k \geq 2$.

Předpoklad dává $m \in \mathbb{Z}$ takové, že $a - u = mn$. Pak

$$a^k - u^k = (a - u)(a^{k-1} + a^{k-2}u + \cdots + au^{k-2} + u^{k-1}) = m(a^{k-1} + a^{k-2}u + \cdots + au^{k-2} + u^{k-1})n,$$

kde $m(u^{k-1} + a^{k-2}u + \cdots + au^{k-2} + u^{k-1}) \in \mathbb{Z}$. Proto $n | (a^k - u^k)$ a závěr následuje.

Další alternativní důkaz (indukcí) najdete jako cvičení . \square

Příklad 7a.c: Vypočítáme, čemu je kongruentní výraz $(3 \cdot 5 \cdot 17 + 6 \cdot 3 + 9)^8 \cdot (2 + 35 - 4 \cdot 5)$ modulo 6. Podle vět víme, že můžeme prakticky všechna čísla (kromě exponentu 8, pro ten jsme zatím pravidlo neměli) nahradit čísla příjemnějšími, čím menší tím určitě lépe. Proto podle Faktu zkusíme dávat rovnou zbytky po dělení šesti, ke kterým se často nejsnáze dostaneme odečítáním šestky.

$$(3 \cdot 5 \cdot 17 + 6 \cdot 3 + 10)^8 \cdot (2 + 35 - 4 \cdot 5) \equiv (3 \cdot 5 \cdot 5 + 0 \cdot 3 + 4)^8 \cdot (2 + 5 - 4 \cdot 5) \pmod{6}.$$

Tedž si započítáme obyčejným způsobem:

$$(3 \cdot 5 \cdot 5 + 0 \cdot 3 + 4)^8 \cdot (2 + 5 - 4 \cdot 5) = (15 \cdot 5 + 3)^8 \cdot (7 - 20) \pmod{6}.$$

A zase můžeme nahradit, pak zase počítat, napíšeme celý výpočet.

$$\begin{aligned} (3 \cdot 5 \cdot 17 + 6 \cdot 3 + 10)^8 \cdot (2 + 35 - 4 \cdot 5) &\equiv (3 \cdot 5 \cdot 5 + 0 \cdot 3 + 4)^8 \cdot (2 + 5 - 4 \cdot 5) = (15 \cdot 5 + 4)^8 \cdot (7 - 20) \\ &\equiv (3 \cdot 5 + 4)^8 \cdot (1 - 2) = (15 + 4)^8 \cdot (-1) \\ &= (19)^8 \cdot (-1) \equiv 1^8 \cdot (-1) = 1 \cdot (-1) = -1 \pmod{6}. \end{aligned}$$

Všimněte si, jak v postupu pečlivě rozlišujeme mezi běžnou algebrou s čísly (značenou rovníkem) a místy, kde nahrazujeme pomocí rozličných pravidel pro kongruenci (značeno \equiv).

\triangle

V aplikacích, kde se pracuje výhradně modulo nějaké konkrétní n , se takové pečlivé rozlišování už nedělá a zkušení lidé prostě píšou rovnítko. Například pokud bychom pracovali čistě s hodinami, tak namísto správného $3 \cdot 16 - 21 \equiv 3 \cdot 4 - 9 = 3 \pmod{12}$ prostě napíšeme $3 \cdot 16 - 21 = 3 \cdot 4 - 9 = 3$. Je to příjemné, ale tady zatím ještě tak zkušení nejsme a navíc se tu snažíme pochopit i teorii, takže budeme pečlivě psát kongruenze, ať v tom není zmatek.

! Jistě jste si všimli, že jsme zatím nezmínili některé operace. Asi to nebude náhoda.

• Nemáme pravidlo pro nahrazování v exponentu. Konkrétně, není obecně pravda, že když $k \equiv l \pmod{n}$, tak $a^k \equiv a^l \pmod{n}$. Například počítáme-li 2^4 modulo 3, pak by nahrazení v exponentu dalo $2^1 = 2$, což ale není správně, $2^4 = 16 \equiv 1 \pmod{3}$.

Je to dáno tím, že mocnění ve skutečnosti není algebraická operace, ale zápis (zkratka) pro opakování násobení. Třeba $a^4 = a \cdot a \cdot a \cdot a$, zde jsou čísla a z určitého světa (ve kterém se počítá modulo n nebo i jinak, mocnina se zavádí mnohem obecněji), zatímco to 4 (obecně exponent) do toho světa nepatří, je vždy ze světa \mathbb{N} a říká, kolik objektů násobíme. Nedá se proto čekat, že by pravidla ze světa, odkud bereme a , platila i pro exponent (pro exponenty máme jiná pravidla). Blíže se o tom dočteme v kapitole .

Přesto by se nám často hodilo umět zmenšovat exponenty, na to jsou triky jednoduché (viz následující příklad níže) i sofistikované, viz sekce o Eulerově větě.

• Nemáme pravidla pro dělení. My totiž dokonce ani nemáme dělení. Je to podobné jako s odčítáním, dělení je jen pohodlný převlek pro něco jiného. Tak jako je $5 - 2$ jen zkratka pro $5 + (-2)$, je i $6/2$ jen příjemný způsob, jak napsat $6 \cdot \frac{1}{2}$. Takže by nám vlastně ani pravidla pro dělení neměla chybět, ale chybí, ona je to totiž moc příjemná zkratka. Budeme ale na to muset jít jinak, dáme se do toho za chvíli, až si k tomu připravíme podmínky. Jmenovitě si proces počítání modulo zachytíme pomocí speciální matematické struktury.

! Viděli jsme, že si ve světě modula n při počítání zahrnujícím operace sčítání a násobení (a taky to odčítání) vystačíme jen s malou množinou čísel, což je ohromně užitečné třeba pro počítače, které umí ukládat jen konečně mnoho dat. Nejtradičnější je pracovat čistě se zbytky, tedy s čísly z rozmezí 0 až $n-1$. Vzniká pak nový matematický svět.

! Definice.

Nechť $n \in \mathbb{N}$. Symbolem \mathbb{Z}_n značíme množinu $\mathbb{Z}_n = \{0, 1, 2, \dots, n-1\}$.

Pro všechna $a, b \in \mathbb{Z}_n$ definujme operace

$$a \oplus b = (a + b) \text{ mod } n,$$

$$a \odot b = (a \cdot b) \text{ mod } n.$$

Takže chceme-li sečít/vynásobit dvě čísla ze \mathbb{Z}_n , tak začneme tím, že to uděláme normálně, čímž se ale můžeme dostat mimo tuto množinu. Vrátíme se do ní, když výsledek nahradíme oblíbeným zástupcem neboli zbytkem po dělení n .

! **Příklad 7a.d:** Nechť $n = 5$. Pak $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$ a máme třeba $2 \oplus 1 = 3$, neboť $2 + 1 = 3$ a $3 \text{ mod } 5 = 3$. Zajímavější je $3 \oplus 4 = 2$, neboť $3 + 4 = 7$ a $7 \text{ mod } 5 = 2$. Máme také $1 \oplus 4 = 0$ (rozmyslete si) nebo třeba $3 \odot 4 = 2$, neboť $3 \cdot 4 = 12 \text{ mod } 5 = 2$.

△

Čtenář se setkal s tím, že se třeba operace sčítání, což je určitý nápad, jak kombinovat čísla, dala používat v rozličných světech: Sčítali jsme ve světě přirozených čísel \mathbb{N} , reálných čísel \mathbb{R} nebo třeba ve světě komplexních čísel. Nejde tedy o nic nového, teď používáme sčítání v dalším světě čísel \mathbb{Z}_n (museli jsme jej na to trochu upravit). Posléze ukážeme, že i v tomto světě platí pro „sčítání“ stejná pravidla, na jaká jsme zvyklí, obdobně pro násobení.

! **Příklad 7a.e:** Chování operací u konečných množin se dá dobře zachytit tabulkou. Ukažme si tabulky pro operace \oplus a \odot v \mathbb{Z}_6 .

\oplus	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

\odot	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Ověřte si, že se v tabulkách vyznáte, takže například umíte v levé najít, že $3 \oplus 4 = 1$, a v pravé $2 \odot 4 = 2$.

Vyzkoušejte si také, že takto umíte počítat, dokázali byste tu tabulku sami vyrobit?

△

Možná vás teď napadlo, jaký je vlastně rozdíl mezi počítáním modulo n a počítáním v prostoru \mathbb{Z}_n ? V podstatě žádný, v obou případech lze zúčastněná čísla nahradit příjemnějšími a počítáme stejně, rozdíl je až na konci, kde si v případě počítání modulo můžeme pro výsledek vybrat libovolného zástupce, zatímco v případě počítání v \mathbb{Z}_n si musíme vybrat zástupce z této množiny. Dá se říct, že při počítání v \mathbb{Z}_n jsme vlastně počítání modulo pěkně zabalili.

Další drobný rozdíl je v tom, že některé malé triky bychom při práci v \mathbb{Z}_n vlastně neměli používat, ale ony se tam dají udělat neoficiálně, jak hned uvidíme.

Příklad 7a.f: Hledáme výsledek výrazu $(7 + 3 \cdot 5)^{18}$ modulo 8. Počítáme

$$(7 + 3 \cdot 5)^{18} = (7 + 15)^{18} \equiv (7 + 7)^{18} = 14^{18} \equiv 6^{18} \pmod{8}.$$

Již jsme diskutovali, že mocninu 18 nahradit lepším zástupcem nelze, a přímý výpočet také není nejlepší strategie. O zbytku po dělení totiž rozhoduje i poslední cifra, jinými slovy není možné pracovat se zaokrouhlenými čísly, musíme znát všechny cifry. Tím se u vyšších mocnin vylučuje použití kalkulačky, protože tradičně vycházejí čísla delší, než kalkulačka dokáže udržet. Nezbývá než použít nějaký trik, velice populární je postupné odebrání malých mocnin z exponentu, které pak snadno vypočítáme. Běžná pravidla pro práci s exponenty totiž platí i obecně.

$$(7 + 3 \cdot 5)^{18} \equiv 6^{18} = 6^{2 \cdot 9} = (6^2)^9 = 36^9 \equiv 4^9 \pmod{8}.$$

Ted' už dvojku oddělit neumíme. Mohli bychom oddělit trojku a počítat $4^9 = (4^3)^3 = \dots$, ale druhá mocnina je přece jen hezčí, tak použijeme jiný způsob odebírání z exponentu.

$$(7 + 3 \cdot 5)^{18} \equiv 4^9 = 4^1 4^8 = 4 \cdot (4^2)^4 = 4 \cdot 16^4 \equiv 4 \cdot 0^4 = 0 \pmod{8}.$$

Pomocí těchto dvou druhů odebírání dokážeme i velice vysokou mocninu zredukovat bez větších problémů, jen to někdy chvíli trvá.

Tento výpočet se dá ovšem také interpretovat jako počítání v prostoru \mathbb{Z}_8 , kde operace už v sobě zahrnují jak běžný výpočet, tak okamžitý přechod k ideálním kongruentním zástupcům:

$$(7 \oplus 3 \odot 5)^{18} = (7 \oplus 7)^{18} = 6^{18} = (6^2)^9 = (6 \odot 6)^9 = 4^9 = 4 \odot 4^8 = 4 \odot (4^2)^4 = 4 \odot (4 \odot 4)^4 = 4 \odot 0^4 = 0.$$

Vidíme, že se to od výpočtu modulo opravdu liší jen zápisem.

Zkušení výpočetníci by při práci v \mathbb{Z}_n psali obyčejné plus a krát, ale my si v této kapitole musíme dávat pozor na teorii, tak si budeme speciálním značením připomínat, že používáme speciální operace z prostoru \mathbb{Z}_n . Je to o to důležitější, že se nám vlastně budou míchat tři druhy operací: obyčejné, výpočty modulo a výpočty v \mathbb{Z}_n , rozličné značení nám pomůže, ať v tom není zmatek.

Ted' si ukážeme příklad, kde již rozdíl mezi počítáním modulo a počítáním v \mathbb{Z}_n bude. Pokud budeme chtít spočítat $147 \cdot 148$ modulo 150, můžeme to provést takto:

$$147 \cdot 148 = 21756 \equiv 6 \pmod{150}.$$

Výpočet v \mathbb{Z}_{150} by se dělal stejně, $147 \odot 148 = 21756 \text{ mod } 150 = 6$.

Jenže při výpočtu modulo 150 můžeme zkousit ještě něco jiného, co celý výpočet výrazně zjednoduší:

$$147 \cdot 148 \equiv (-3) \cdot (-2) = 6.$$

A tento trik v \mathbb{Z}_{150} udělat nelze, protože tam žádná záporná čísla neexistují. Ale to není problém, v takových případech si prostě na chvíli odskočíme do počítání modulo n a pak výsledek nahradíme kongruentním číslem ze \mathbb{Z}_n .

△

7a.7 Další operace v \mathbb{Z}_n : Odčítání.

Na první pohled by se zdálo, že s odčítáním v \mathbb{Z}_n nebude problém. Nabízí se definice $a \ominus b = (a - b) \text{ mod } n$ a ani prakticky to nevypadá špatně. Například bychom si tipli, že v prostoru \mathbb{Z}_6 máme $5 \ominus 3 = 2$, což potvrší i zkouška: $2 \oplus 3 = 5$. Aby to bylo zajímavější, zkusíme si $2 \ominus 5 = -3 \text{ mod } 6 = 3$ a zkouška nám opět vyjde: $3 \oplus 5 = 2$ v \mathbb{Z}_6 .

Nicméně se to takto nedělá. Již jsme se zmínili, že odčítání se nebere jako jedna ze základních algebraických operací, mimo jiné proto, že nesplňuje některé věci, které považujeme za zásadní, například asociativní zákon. Ve skutečnosti se odečítání bere jako příjemná zkratka pro přičítání opačného prvku. Například opačný prvek k 5 je -5 , namísto $3 + (-5)$ píšeme $3 - 5$.

Proto se odčítání nezavádí ani ve světě modulo jako samostatná operace, místo toho se vytvoří pojem opačného čísla. Protože v \mathbb{Z}_n záporná čísla nejsou, musíme na to trochu jinak. Začneme otázkou, co je to vlastně opačné číslo, jak jej známe. Co mají společná čísla 5 a -5 ? To, že se navzájem vynulují, tedy $5 + (-5) = 0$. Přesně toto nám může fungovat i v \mathbb{Z}_n .

Podívejme se třeba do světa \mathbb{Z}_6 . Dokážeme najít číslo opačné k 5 neboli číslo takové, že po přičtení k 5 dostaneme nulu? Podíváme-li se do tabulky v příkladě, zjistíme, že $5 \oplus 1 = 0$. Zkusmo tedy prohlašme, že opačné číslo k 5 je 1 v \mathbb{Z}_6 , psáno $(-5) = 1$ modulo 6. Letmý pohled na tabulkou sčítání v \mathbb{Z}_6 odhalí, že opačné číslo lze najít ke všem prvkům \mathbb{Z}_6 , což je slibný začátek, čtenář již dokonce asi tuší, jak se ta čísla hledají a že se najdou obdobnou metodou v každém \mathbb{Z}_n .

Dobrá otázka je, zda nám tato opačná čísla opravdu dokážou nahradit odčítání modulo. Před chvílí jsme zkousili spočítat $2 - 5$ modulo 6, vyšlo $-3 \equiv 3 \pmod{6}$. Pokud budeme počítat v prostoru \mathbb{Z}_6 , musíme přejít k přičtení opačného čísla: $2 \ominus 5 = 2 \oplus (-5) = 2 \oplus 1 = 3$. Funguje to.

To samozřejmě mohla být náhoda, ale není, jsme na správné cestě. Další povzbuzení nám dodá, když se na opačný prvek podíváme trochu jinou cestou. Říkali jsme si, že někdy se vyplatí ze \mathbb{Z}_n vyskočit k počítání modulo, tam dostat výsledek a nahradit jej správným zástupcem ze \mathbb{Z}_n . Opačné číslo k 5 je -5 , to platí i modulo 6, ted' najdeme správného zástupce čísla -5 modulo 6, což je 1, souhlasí.

Než se dáme do formálních definic, poznamenejme, že modulo je zde stále zásadní. Pokud bychom chtěli počítat $2 - 5$ v \mathbb{Z}_7 , tak musíme použít opačný prvek k 5 vzhledem k modulu 7, což je evidentně jiné číslo než ta 1 ze světa modulo 6.

Definice.

Nechť $n \in \mathbb{N}$, nechť $a \in \mathbb{Z}_n$. Řekneme, že $b \in \mathbb{Z}_n$ je **opačný prvek** k a v \mathbb{Z}_n , jestliže $a \oplus b = 0$ v \mathbb{Z}_n . Pak značíme $b = (-a)$.

Diskuse před definicí naznačila, jak opačné prvky v prostorech \mathbb{Z}_n hledat v praxi, pro dané $a \in \mathbb{Z}_n$ začneme s $-a$ a najdeme si jeho zástupce ze \mathbb{Z}_n , což je $(-a) + n = n - a$. Možná. Rozmyslete si, že to neplatí úplně vždy, pak čtěte dál.

! Fakt 7a.8.

Nechť $n \in \mathbb{N}$.

(i) $(-0) = 0$.

(ii) Jestliže $a \in \mathbb{Z}_n$ a $a \neq 0$, pak $(-a) = n - a$.

Důkaz je snadný, necháme jej jako cvičení.

S opačnými prvky jsme se setkali i při práci s běžnými čísly například při řešení rovnic.

Příklad 7a.g: Uvažujme rovnici $x \oplus 5 = 2$ v \mathbb{Z}_6 . Kdyby to byla rovnice „normální“, tak bychom prostě od obou stran odečetli 5 neboli přičetli -5 . V \mathbb{Z}_6 je to stejně, jen musíme hledat opačný prvek jinak. Víme už, že v \mathbb{Z}_6 je $(-5) = 6 - 5 = 1$, takže upravujeme:

$$x \oplus 5 = 2 \implies (x \oplus 5) \oplus 1 = 2 \oplus 1 \implies x \oplus (5 \oplus 1) = 3 \implies x \oplus 0 = 3 \implies x = 3.$$

Postup je delší, protože jsme schválňě zdůraznili klíčové kroky při řešení. Viděli jsme tak, že toto řešení závisí na možnosti změnit pozici závorek na levé straně neboli na platnosti tzv. asociativního zákona. Jak víme, že v prostoru \mathbb{Z}_6 platí? To je dobrá otázka a brzy se k ní vrátíme.

△

7a.9 Další operace v \mathbb{Z}_n : Dělení.

Zde je obdobná situace jako u odčítání. U reálných čísel bychom výpočet $4/2$ přepsali jako $4 \cdot \frac{1}{2}$, kde souvislost mezi 2 a $\frac{1}{2}$ je zjevná, $2 \cdot \frac{1}{2} = 1$. Vztah $x \cdot \frac{1}{x} = 1$ lze zkoumat i ve světě \mathbb{Z}_n , třeba si můžeme všimnout, že $3 \cdot 7 = 1 \pmod{10}$, tudíž bychom ve světě \mathbb{Z}_{10} mohli psát, že $\frac{1}{3} = 7$.

To se zase hodí třeba při řešení rovnic. V prostoru \mathbb{Z}_{10} lze rovnici $3x = 4$ řešit tímto postupem:

$$3 \odot x = 4 \implies 3^{-1} \odot (3 \odot x) = 3^{-1} \odot 4 \implies (7 \odot 3) \odot x = 7 \odot 4 \implies 1 \odot x = 8 \implies x = 8.$$

Zkuste si ale rozmyslet, že k číslu 4 nenajdete žádné x , aby platilo $4 \cdot x = 1 \pmod{10}$, stačí projít všechna čísla 0 až 9 a vyloučit je. To ukazuje, že tentokrát bude situace poněkud jiná než u opačných čísel.

Definice.

Uvažujme $n \in \mathbb{N}$.

Nechť $a \in \mathbb{Z}$. Řekneme, že $b \in \mathbb{Z}$ je **inverzní prvek (inverse element)** k a **modulo n** , jestliže $a \cdot b \equiv 1 \pmod{n}$.

Nechť $a \in \mathbb{Z}_n$. Řekneme, že $b \in \mathbb{Z}_n$ je **inverzní prvek** k a v \mathbb{Z}_n , jestliže $a \odot b = 1$ v \mathbb{Z}_n .

Pokud takovýto prvek b existuje, pak jej značíme $b = a^{-1}$ a řekneme, že a je **invertibilní (invertible)** modulo n , resp. v \mathbb{Z}_n .

Jako obvykle platí, že nalezený inverzní prvek v \mathbb{Z}_n již dává inverzní prvek modulo n dle první definice, naopak libovolný inverzní prvek modulo n nám okamžitě dá i inverzní prvek z prostoru \mathbb{Z}_n , když jej nahradíme vhodným kongruentním zástupcem. Jde tedy opět o stejnou myšlenku, jen trochu jinak oblečenou.

Podobně jako u opačného prvku, i zde na modulu záleží, například v \mathbb{Z}_{10} jsme uholí $3^{-1} = 7$, ale v \mathbb{Z}_8 již je $3^{-1} = 3$ (ověřte). A stejně jako u opačného prvku se modulo ze značení nepozná, což je ze striktně matematického pohledu nešťastné, ale z praktického pohledu to zase takový problém není, protože používané modulo je vždy jasné z kontextu, prakticky nikdy nepracujeme s více moduly najednou.

! V této chvíli čtenáři doporučíme, aby si přečetl kapitolu , přinejmenším část . Pochopí, proč byla u definice opačného prvku zrovna nula a u inverzního zrovna jednička, a vůbec celou tuhoto partií uvidí v trochu jiném světle.

Invertibilní a inverzní prvky jsou velice užitečné a musíme se je naučit hledat, příklad výše ovšem ukazuje, že někdy to je nemožné. To vlastně není nic nového, v prostoru \mathbb{R} také neumíme najít inverzní prvek k nule, ale v \mathbb{Z}_n těchto nepříjemných situací může být víc.

Příklad 7a.h: Podíváme se na několik tabulek násobení v zajímavých \mathbb{Z}_n . Prvky, které mají inverzi (jsou invertibilní), poznáme podle toho, že se v jejich řádku vyskytne výsledek jedna, v příslušném sloupci pak nahoře dohledáme onen inverzní prvek.

Nejprve si připomeneme případ (\mathbb{Z}_6, \odot) .

\odot	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Zde vidíme, že jediné invertibilní prvky jsou 1 a 5, platí $1^{-1} = 1$ a $5^{-1} = 5$.

Zato tu máme „dělitele nuly“, třeba $2 \odot 3 = 0$ či $3 \odot 4 = 0$.

Na to nejsme zvyklí a má to zase dopady na řešení rovnic. Zatímco v \mathbb{Z} (a v \mathbb{R} atd.) má rovnice $2x = 0$ automaticky jediné řešení $x = 0$, v \mathbb{Z}_6 už je i řešení $x = 3$. Z toho někdy plynou zajímavé komplikace.

Toto byl asi extrémně pesimistický případ. Teď si ukážeme naopak nejlepší možný případ, zastoupený příkladem (\mathbb{Z}_5, \odot) .

\odot	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Vidíme, že v \mathbb{Z}_5 jsou všechny prvky kromě nuly invertibilní, máme $1^{-1} = 1$, $2^{-1} = 3$ a $3^{-1} = 2$ neboť $2 \odot 3 = 1$, $4^{-1} = 4$ neboť $4 \odot 4 = 1$.

Zde by tedy mnohé věci měly fungovat dost podobně jako ve světě reálných čísel a dokonce někdy lépe než ve světě \mathbb{Z} . Například rovnici $3x = 4$ nedokážeme ve světě \mathbb{Z} vyřešit, zatímco v \mathbb{Z}_5 stačí rovnici vynásobit číslem $3^{-1} = 2$ a dostáváme $x = 3$, což je opravdu řešení dané rovnice.

Jen pro úplnost, $(-0) = 0$, $(-1) = 4$, $(-2) = 3$, $(-3) = 2$ a $(-4) = 1$.

Na takovéto výrazně příjemné případy se brzy podíváme blíže. Typický prostor \mathbb{Z}_n je nicméně někde mezi právě předvedenými extrémy, pěkně to ukazuje třeba \mathbb{Z}_{14} .

\odot	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
2	0	2	4	6	8	10	12	0	2	4	6	8	10	12
3	0	3	6	9	12	1	4	7	10	13	2	5	8	11
4	0	4	8	12	2	6	10	0	4	8	12	2	6	10
5	0	5	10	1	6	11	2	7	12	3	8	13	4	9
6	0	6	12	4	10	2	8	0	6	12	4	10	2	8
7	0	7	0	7	0	7	0	7	0	7	0	7	0	7
8	0	8	2	10	4	12	6	0	8	2	10	4	12	6
9	0	9	4	13	8	3	12	7	2	11	6	1	10	5
10	0	10	6	2	12	8	4	0	10	6	2	12	8	4
11	0	11	8	5	2	13	10	7	4	1	12	9	6	3
12	0	12	10	8	6	4	2	0	12	10	8	6	4	2
13	0	13	12	11	10	9	8	7	6	5	4	3	2	1

Vidíme, že máme invertibilní prvky 1, kde $1^{-1} = 1$, dále $3^{-1} = 5$ (kontrola: $3 \cdot 5 = 15$, modulo 14 to dává opravdu $15 - 14 = 1$), dále $5^{-1} = 3$, $9^{-1} = 11$ (kontrola: $9 \cdot 11 = 99$, modulo 14 to dává opravdu $99 - 7 \cdot 14 = 1$), dále $11^{-1} = 9$ a nakonec $13^{-1} = 13$.

△

Jak se pozná, které prvky v \mathbb{Z}_n jsou invertibilní? Pokud vám příklady nenapovídely, tady je odpověď. Opět poskytneme dvě verze, jednu pro počítání modulo a jednu pro počítání v prostoru \mathbb{Z}_n , jako obvykle půjde o stejnou věc, jen jinak vyjádřenou.

!

Věta 7a.10.

Nechť $n \in \mathbb{N}$.

(i) Nechť $a \in \mathbb{Z}$. Existuje $x \in \mathbb{Z}$ takové, že $ax \equiv 1 \pmod{n}$, právě tehdy, když $\gcd(a, n) = 1$.

Pak je toto x určeno jednoznačně až na modulo n a všechna y s x kongruentní také splňují $ay \equiv 1 \pmod{n}$.

(ii) Nechť $a \in \mathbb{Z}_n$. Inverzní prvek a^{-1} v \mathbb{Z}_n existuje právě tehdy, když $\gcd(a, n) = 1$. Pak je tento prvek jediný.

Důkaz (poučný): (i): 1) Takové x existuje právě tehdy, když existuje $k \in \mathbb{Z}$ takové, že $1 - ax = kn$. Jinými slovy, nalezení prveku x je ekvivalentní tomu, že umíme najít řešení $x, k \in \mathbb{Z}$ rovnice $1 = ax + kn$. To je ale diofantická rovnice, kterou jsme se už naučili řešit, a podle Věty 6c.1 to lze právě tehdy, když je 1 násobkem $\gcd(a, n)$. Takže řešení existuje právě tehdy, když $\gcd(a, n) = 1$. Důkaz ekvivalence je hotov.

2) Jednoznačnost: Nechť $x, y \in \mathbb{Z}$ obě splňují dotyčnou podmítku. Pak existují $k, l \in \mathbb{Z}$ takové, že $1 = ax + kn$ a $1 = ay + ln$. Odečtením získáme $ax - ay = kn - ln$, tedy $a(x - y) = (k - l)n$. To znamená, že n dělí $a(x - y)$, a protože je n nesoudělné s a , musí podle Lemma 6a.23 n dělit $x - y$, tedy $x \equiv y \pmod{n}$.

3) $y \equiv x \pmod{n}$ dává $k \in \mathbb{Z}$ tak, aby $y = x + kn$. Pak $ay = a(x + kn) = ax + (ak)n \equiv 1 + 0 = 1 \pmod{n}$, použili jsme Fakt (ii) na násobek $(ak)n$.

(ii): Číslo $x \in \mathbb{Z}_n$ splňuje $a \odot x = 1$ právě tehdy, když $ax \equiv 1 \pmod{n}$, což je právě tehdy, když $ax \equiv 1 \pmod{n}$, což je podle (i) právě tehdy, když $\gcd(a, n) = 1$.

Jednoznačnost plyne buď z (i), nebo také z Faktu . □

Vlastně jsme v bodě (i) dokázali, že jakmile je nějaký prvek a invertibilní modulo n , tak množina všech jeho inverzních čísel je přesně množina všech čísel kongruentních s nějakým konkrétním inverzním číslem x .

Už umíme poznat, kdy a^{-1} existuje, ale jak jej najdeme? Úplně snadné to není, žádný vzoreček totiž neexistuje. Jediný rozumný postup vychází z důkazu Věty , přes řešení příslušné diofantické rovnice. Na to máme Algoritmus 6c.6, postup lze ale zjednodušit, protože teď nás vlastně jedna neznámá nezajímá.

S Algoritmus 7a.11. pro hledání inverzního prvku k a vzhledem k násobení modulo n , popřípadě pro hledání inverzního prvku k a v \mathbb{Z}_n .

0. Například pomocí rozšířeného Euklidova algoritmu najděte $\gcd(a, n) = Aa + Bn$.

1. Jestliže $\gcd(a, n) > 1$, pak inverzní prvek k a v \mathbb{Z}_n neexistuje.

Pokud umíte $\gcd(a, n)$ získat snadněji než Euklidovým algoritmem (třeba pohledem) a vyjde číslo větší než 1, je možné krok 0 přeskočit.

2. Jestliže $\gcd(a, n) = 1$, pak Bezoutova identita dává $1 = a \cdot A + B \cdot n$. To znamená, že $a \cdot A \equiv 1 \pmod{n}$ a $x = A$ je hledaný inverzní prvek.

Pokud hledáte inverzní prvek v \mathbb{Z}_n , pak najděte vhodného zástupce x z rozmezí $1, 2, \dots, n - 1$ buď přičtením/odečtením vhodného násobku n , nebo dělením se zbytkem.

△

! Příklad 7a.i: Najdeme inverzní prvek k $a = 36$ modulo 175.

Nejprve si to přeložíme: Hledáme x splňující $36x \equiv 1 \pmod{175}$ neboli x takové, aby pro nějaké $m \in \mathbb{Z}$ bylo $36x + 175m = 1$.

Není jasné, zda vidíme bez větší práce, kolik je $\gcd(36, 175)$, tak rovnou zkusíme rozšířený Euklidův algoritmus pro jeho nalezení.

175		1	0
36	4	0	1
31	1	1	-4
5	6	-1	5
1•	5	7•	-34•
0			

Dostáváme $\gcd(175, 36) = 1 = 7 \cdot 175 + (-34) \cdot 36$. Když se na obě strany Bezoutovy rovnosti podíváme modulo 175, dostáváme $36 \cdot (-34) + 0 \cdot 7 \equiv 1$, tedy $36 \cdot (-34) \equiv 1 \pmod{175}$. Proto $36^{-1} = -34$ vzhledem k počítání modulo 175.

Čtenář si samořejmě může vybrat i jiného zástupce dle osobní preference, například $-34 \equiv 316 \pmod{175}$.

Pokud bychom hledali inverzní prvek k 36 v \mathbb{Z}_{175} , pak bychom také nejprve přeložili zadání: Hledáme x splňující $36 \odot x = 1$ neboli $36x \equiv 1 \pmod{175}$ neboli $36x + 175m = 1$ pro nějaké $m \in \mathbb{Z}$.

Pak bychom postupovali stejně jako výše, ale na konci bychom pro -34 museli najít zástupce ze \mathbb{Z}_{175} , nejsnáze přičtením čísla 175.

Závěr: 36 je v \mathbb{Z}_{175} invertibilní a $36^{-1} = 141$.

Zkouška: $36 \cdot 141 = 5076 \equiv 1 \pmod{175}$, neboť $5076 = 29 \cdot 175 + 1$.

△

Všimněte si, že pokud je n prvočíslo, tak vlastně všechny nenulové prvky \mathbb{Z}_n jsou nesoudělné s n a tudíž mají inverzi. To už je jako u reálných čísel. V mnoha aplikacích skutečně úspěšně nahrazujeme svět \mathbb{R} světem \mathbb{Z}_p pro p prvočíslo.

Teď si ukážeme jednu zajímavou aplikaci počítání modulo.

! Příklad 7a.j: Úzký vztah mezi kryptografií a počítáním modulo jde zpět minimálně ke starým Římanům. Takzvanou Césarovu šifru si nejlépe představíme takto: Máme dva soustředné kruhy, jeden menší než druhý, a po obvodu napišeme na oba písmena, vždy stejná proti sobě. Pak jeden kruh otočíme o tři pozice a vzniká tím šifra, namísto A píšeme D , namísto B píšeme E a tak dále, třeba Y přejde na B .

Matematicky se to simuluje jednoduše, nahradíme písmena číslami $1, \dots, 26$ a pak používáme jako šifru bijekci $T(a) = (a + 3) \pmod{26}$.

Obecně lze posouvat i o jiné číslo než o tu Césarovu trojku. Zvolíme si nějaké k mezi 1 a 25 a dostáváme „šifrování posunem“: $T(a) = (a + k) \pmod{26}$. Toto se snadno dešifruje, $T^{-1}(b) = (b - k) \pmod{26}$, a pokud nechceme odečítat, tak si najdeme opačný prvek $(-k)$ a máme $T^{-1}(b) = (b + (-k)) \pmod{26}$.

Například pokud zvolíme číslo $k = 8$, tak písmeno 20 zašifrujeme jako $T(20) = (20 + 8) \pmod{26} = 28 \pmod{26} = 2$. Opačný prvek ke $k = 8$ modulo 26 je 18 (zkouška: $8 + 18 = 26 \equiv 0 \pmod{26}$), proto bychom se posunem o 18 měli zase dostat zpět: $T^{-1}(2) = (2 + 18) \pmod{26} = 20 \pmod{26} = 20$. Zajímavá volba je $k = 13$, pak $T^{-1} = T$.

Tato šifra není příliš bezpečná. Protože se dané písmeno vždy kóduje stejně, je vysoce náchylná na frekvenční analýzu, kdy si prostě spočítáme, které písmeno se v zašifrované zprávě vyskytuje nejčastěji, a je vysoce pravděpodobné, že odpovídá nejčastějšímu písmenu daného jazyka. Velice pěkně toto popsal E.A. Poe v povídce *Zlatý skarabeus*.

Lépe vypadá šifrování dané předpisem $T(a) = (ea + k) \text{ mod } 26$, kde e je zvoleno tak, aby T bylo prosté (tedy je třeba zvolit něco nesoudělného s 26). Jak se takový vzkaz dekóduje? Zvolili jsme e nesoudělné s 26, pak už víme (Věta), že k němu existuje inverzní prvek d modulo 26, tedy prvek splňující $ed \text{ mod } 26 = 1$. Ukážeme, že $T^{-1}(b) = d(b + (-k))$ dekóduje zprávu:

$$T^{-1}(T(a)) = d(T(a) + (-k)) \equiv d((ea + k) + (-k)) \equiv d(ea + 0) = (de)a \equiv 1 \cdot a = a \pmod{26}.$$

Zvolme třeba $e = 7$ a $k = 3$. Pak $-k = 23$ a ještě potřebujeme vyřešit rovnici $7x + 26m = 1$, buď algoritmem nebo to zkuseme uhádnout. Vyjde například $x = -11$ a $m = 3$, nás zajímá x , ale z prostoru \mathbb{Z}_{26} , dostáváme tedy $d = 15$.

Tedž zakódujeme třeba písmeno B odpovídající hodnotě $a = 2$, takže vyšleme zprávu $T(2) = 7 \cdot 2 + 3 \pmod{26} = 17$ neboli písmeno Q . Příjemce na zprávu aplikuje T^{-1} :

$$T^{-1}(17) = 15(17 + 23) = 15 \cdot 40 \equiv 15 \cdot 14 = 210 \equiv 2 \pmod{26}.$$

Vyšlo to.

Ale lépe tato šifra jen vypadá, pořád je to dětská šifra zranitelná přes frekvenční analýzu. Zajímavé zobecnění je použít modulární aritmetiku na bloky číslic, nikoliv jednotlivé číslice, tam už frekvence nepomohou. Přesto jsou ale šifry tohoto typu pořád zranitelné díky své pravidelnosti. K lepším šífrám se dostaneme brzy.

△

Problém inverzních prvků jsme vyřešili tak dobře, jak jen to jde. Vrátíme se ještě k problematice umocňování. Sice jsme vyvinuli metodu postupné redukce exponentu, ale počítat tak třeba 13^{1946} modulo 17 nezní moc lákavě. Pokud je n prvočíslo (a už víme, že to je velice příjemný případ), naskýtá se ještě jiná zajímavá metoda snižování mocniny.

! Věta 7a.12. (malá Fermatova věta)

Nechť $n \in \mathbb{N}$ je prvočíslo.

- (i) Je-li $a \in \mathbb{Z}$ nesoudělné s n , pak platí $a^{n-1} \equiv 1 \pmod{n}$.
- (ii) Pro každé $a \in \mathbb{Z}$ pak platí $a^n \equiv a \pmod{n}$.

Důkaz (poučný): (i) Nejprve ukážeme, že čísla $a, 2a, \dots, (n-1)a$ nejsou navzájem kongruentní modulo n . Když totiž $ia \equiv ja \pmod{n}$, pak n dělí $a(i-j)$, ale n je nesoudělné s a , proto (Lemma 6a.23) n dělí $i-j$. Nicméně $|i-j| < n$, proto $i-j = 0$, tedy $ia = ja$.

Když tedy vezmeme $a, 2a, \dots, (n-1)a$ modulo n , dostaneme v nějakém pořadí všechna čísla $1, 2, \dots, n-1$.

Když je všechny spolu vynásobíme, což lze psát jako $\prod_{i=1}^{n-1} (ia) \pmod{n}$, dostaneme $(n-1)!$. Upravením toho součinu máme $(n-1)! \equiv a^{n-1}(n-1)! \pmod{n}$. Protože $\gcd((n-1)!, n) = 1$, je prvek $(n-1)!$ invertibilní v \mathbb{Z}_n , proto vynásobením obou stran kongruence jeho inverzí dostaneme $1 \equiv a^{n-1} \pmod{n}$.

(ii) Nechť $a \in \mathbb{Z}$, rozebereme dva případy. Jestliže $\gcd(a, n) = 1$, tak lze aplikovat a) a dostaneme $a^{n-1} \equiv 1 \pmod{n}$, takže podle Věty (iii) je $a^n = a \cdot a^{n-1} \equiv a \cdot 1 = a \pmod{n}$.

Jestliže $\gcd(a, n) > 1$, pak existuje společný dělitel $d > 1$. Jenže n je prvočíslo, takže jediný jeho dělitel (kromě 1) je $d = n$. Takže vlastně $n | a$, pak $a \equiv 0 \pmod{n}$, tedy podle Faktu $a^n \equiv 0^n = 0 = a \pmod{n}$.

□

Alternativní důkaz se najde jako Poznámka .

Čtenáře možná napadne, proč jsme vlastně uváděli (i), když je verze (ii) obecnější a možná i elegantnější. Důvod je jednoduchý, verze (i) je tradiční a rovněž praktických výpočtech užitečnější. Proto všichni za „malou Fermatovu větu“ považují tvrzení (i), budeme to tak dělat i my.

! Příklad 7a.k: Spočítáme 136^{182} modulo 13. Nejprve použijeme nahrazení v základu: $136^{182} \equiv 6^{182} \pmod{13}$. Poznamenejme, že číslo 6^{182} má 142 cifer, takže by nám v této fázi kalkulačka rozhodně nepomohla. Musíme redukovat exponent, a protože je 13 prvočíslo, můžeme na to použít malého Fermata. Na to si tam ale musíme vyrobit mocninu $13 - 1 = 12$, což je snadné, $6^{182} = 6^{12 \cdot 15+2}$. Podle malé Fermatovy věty pak máme $6^{12} \equiv 1 \pmod{13}$, tedy

$$136^{182} \equiv 6^{182} = (6^{12})^{15} \cdot 6^2 \equiv 1^{15} \cdot 36 = 36 \equiv 10 \pmod{13}.$$

Je to rozhodně lepší, než snižování mocniny po dvou, což by vypadalo nějak takto:

$$6^{182} = (6^2)^{91} \equiv 10 \cdot 10^{90} = 10 \cdot (10^2)^{45} \equiv 10 \cdot 9 \cdot 9^{44} = \dots$$

Chytřejší by bylo použít $10 \equiv -3 \pmod{13}$ a při umocňování na sudý exponent se dá znaménko ignorovat, tedy

$$6^{182} = (6^2)^{91} \equiv 10 \cdot (-3)^{90} = 10 \cdot (3^2)^{45} = 10 \cdot 9 \cdot 9^{44} \equiv 10 \cdot 9 \cdot (-4)^{44} = \dots$$

I tak by to bylo na dlouhé zimní večery, ten Fermat byl výrazně kratší.

Všimněte si, že pokud bychom chtěli použít tvrzení (ii) výše, dostali bychom

$$6^{182} = 6^{13 \cdot 14} = (6^{13})^{14} \equiv 6^{14} \pmod{13}.$$

Čekala nás další práce, ta jednička coby výsledek u verze (i) je příjemnější.

△

Pro další podobný příklad a poznámku s úderným trikem viz příklad .

Ve výpočtu jsme použili postup, který lze vyjádřit obecně a je užitečný při práci s velkými čísly.

Fakt 7a.13.

Nechť $n \in \mathbb{N}$ je prvočíslo a $a \in \mathbb{Z}$ není dělitelné n . Pak pro každé $k \in \mathbb{N}_0$ platí $a^k \equiv a^r \pmod{n}$, kde $r = k \bmod (n-1)$ neboli zbytek po dělení k číslem $n-1$.

V zásadě se dá říct, že už ve světě \mathbb{Z}_n umíme počítat. Teď se podíváme na další aplikaci.

! Příklad 7a.1 (pokračování): Vráťme se k problému šifrování. Pro zjednodušení každou zprávu převedeme na jedno číslo v zásadě libovolným způsobem, třeba se rozhodneme, že každé písmeno ze zprávy nahradíme dvoučíslím od 00 do 26 a dáme je za sebe. Chceme tedy vytvořit metodu, která umí kódovat celá čísla.

Jako inspiraci si představme následující kód. Zvolíme $e \in \mathbb{N}$. Zprávu $M \in \mathbb{N}$ zašifrujeme jako $T(M) = M^e$. Jak se dostaneme k původnímu textu? Zobrazením $T^{-1}(C) = \sqrt[e]{C}$. Tato šifra je již výrazně lepší než predchozí pokusy, protože se maskují frekvence a má obecně méně vnitřních pravidelností, čímž se protivníkovi ztěžuje protiútok.

Její nevýhodou je, že výpočet mocniny i odmocniny je velice náročná operace, zejména výpočet odmocniny znamená, že metoda je v praxi nepoužitelná. Proto si teď nápad vylepšíme.

Zvolíme nějaké prvočíslo n strašlivě velké, aby byly zprávy vždy o hodně menší (dlouhé zprávy můžeme sekat). Zvolme libovolné číslo $e \in \mathbb{N}$ nesoudělné s $n-1$, pak podle Věty existuje také $d \in \mathbb{N}$ takové, že $de \equiv 1 \pmod{n-1}$, tedy $ed = 1 + k(n-1)$ pro nějaké $k \in \mathbb{Z}$. Máme pak k dispozici následující způsob šifrování.

Předpokládejme, že M je zpráva splňující $M < n$. Zakódujeme ji zobrazením $T(M) = M^e \pmod{n}$ (proto jsme volili zkratku e jako „encode“). Jak se dělá dešifrování? Tvrdíme, že to dělá zobrazení $T^{-1}(C) = C^d \pmod{n}$ (proto d jako „decode“). Důkaz plyne z malé Fermatovy věty, zde je dobré si uvědomit, že n je prvočíslo, proto jej číslo $1 < M < n$ nemůže dělit a jsou tedy nesoudělná. Máme pak (počítáme modulo n)

$$T^{-1}(T(M)) \equiv (M^e)^d = M^{ed} = M^{1+k(n-1)} = M \cdot (M^{n-1})^k \equiv M \cdot 1^k = M.$$

A je to. Nemusíme odmocňovat, navíc na mocnění modulo máme pěkné triky. Největší slabina této metody je v praktickém provedení, což je mimochodem velkou slabinou většiny šifrovacích schémat. Odesílatel musí nějak dopravit dekódovací klíč d příjemci zprávy, jakákoli cesta je zranitelná a hrozí tak nebezpečí, že si naši zprávu někdo po zachycení klíče přečeťte.

Šifra je zranitelná i opačným směrem. Řekněme, že chceme, aby nám někdo poslal tajnou zprávu. Pošleme mu šifrovací klíč e a číslo n nutné k operaci modulo, sami si schováme dešifrovací klíč d . Jenže pokud někdo naši zprávu odesílateli zachytí, tak si z hodnot e a n hravě náš dešifrovací klíč d spočítá, protože najít inverzi k e modulo $n-1$ je pro rozšířeného Euklida relativně snadný úkol.

Výrazné zvýšení bezpečnosti se dá dosáhnout, pokud nějakou fintou znemožníme, aby odposlouchávač dokázal z hodnot e a n odvodit náš dešifrovací klíč d . Tím se dostáváme ke zlatému hřebu naší procházky kódováním.

Jedním z nejrozšířenějších veřejných šifrovacích schémat na Internetu je v současnosti takzvané **RSA šifrování** (nazvané podle autorů jménem Rivest, Shamir a Adleman, nápad publikovali v roce 1978, i když v tajných službách byl znám i dříve, ale patrně nebyl použit). Na začátku zvolíme dvě prvočísla p, q (typicky o 200 cifrách). Nechť $n = pq$. Zvolíme $e \in \mathbb{N}$ tak, aby bylo nesoudělné s $(p-1)(q-1)$, pak najdeme (rozšířeným Euklidovým algoritmem) $d \in \mathbb{N}$ tak, aby $de \equiv 1 \pmod{(p-1)(q-1)}$, tj. d je inverzní prvek k e vzhledem k násobení modulo $(p-1)(q-1)$. Dvojici (n, e) sdělíme tomu, kdo nám má zprávy posílat, je to tzv. „veřejný klíč“. Sami si schováme „soukromý klíč“ (n, d) .

Kódování: Zprávu $M \in \mathbb{N}$ splňující $M < p, q$ zašifrujeme pomocí zobrazení $T(M) = M^e \pmod{n}$. Tvrdíme, že ji lze dešifrovat pomocí zobrazení $T^{-1}(C) = C^d \pmod{n}$.

Opravdu? Protože je p prvočíslo a díky $M < p$ je s ním M nesoudělné, podle malé Fermatovy věty platí

$$(M^e)^d = M^{1+k(p-1)(q-1)} = M \cdot (M^{p-1})^{k(q-1)} \equiv M \cdot 1^{k(q-1)} = M \pmod{p}.$$

Číslo q má ovšem stejné vlastnosti, proto stejně ukážeme $(M^e)^d \equiv M \pmod{q}$. Podle Cvičení tudíž musí platit $(M^e)^d \equiv M \pmod{\text{lcm}(p, q)}$, a protože jsou p, q coby různá prvočísla nesoudělná, máme dle Cvičení 6a.15 také $(M^e)^d \equiv M \pmod{n}$. Tím jsme dokázali, že ze zprávy M^e dalším umocněním na d a přechodem ke zbytku po dělení modulo n dostáváme původní text M .

Jak bezpečná je tato metoda? Aby zprávu někdo rozšifroval, musel by najít d , k tomu ale potřebuje znát $(p-1)(q-1)$. Takže RSA kód je tak bezpečný, jako je číslo $(p-1)(q-1)$. To se dá získat jedině nalezením příslušné faktorizace n na $p \cdot q$, což už jsme zde několikrát zmiňovali jako pořádný problém. Existují efektivní metody pro určité kombinace, například když jsou p, q dosti blízké nebo když je d relativně malé číslo, ale pro dobře vybrané p, q se odhadovaný čas faktorizace blíží tomu nejhoršímu scénáři, náročnost faktorizačních algoritmů je horší než mocniny, patří do skupiny a^n , což už je hodně (viz kapitola).

Mimochodem, pokud bychom prozradili zároveň n a $m = (p-1)(q-1)$, tak už nejen může kdokoliv zjistit e řešením $ed \equiv 1 \pmod{m}$, ale dokonce snadno zjistí naši faktorizaci: Máme $m = pq - p - q + 1 = n - p - q + 1$, čísla p, q tedy řeší rovnice $pq = n$, $p + q = n - m + 1$, což je snadná algebraická úloha. Například z druhé rovnice vyjádříme q , dáme do první a dostáváme $p^2 - (n - m + 1)p + n = 0$.

△

Máme tedy kvalitní kódování, ale také nový problém: Kde vezmeme prvočísla o 200 cifrách?

Příklad 7a.m: Zde se vrátíme k problému, jak poznat, zda je nějaké $n \in \mathbb{N}$ prvočíslo. Již jsme diskutovali, že zkoušet dělit čísla mezi 1 a \sqrt{n} je extrémně časově náročné, je načase zapřemýšlet nad lepšími alternativami.

Malá Fermatova věta nabízí následující zajímavou obměnu:

Jestliže jsou čísla $a \in \mathbb{Z}$ a n nesoudělná a neplatí $a^{n-1} \equiv 1 \pmod{n}$, pak už n nemůže být prvočíslo.

Toto může sloužit jako test prvočíselnosti. Vezměme libovolné liché $n > 2$, chceme vědět, zda je to prvočíslo. použijeme $a = 2$, protože pro liché číslo určitě $\gcd(2, n) = 1$. Jestliže $2^{n-1} \not\equiv 1 \pmod{n}$, pak podle malé Fermatovy věty n určitě není prvočíslo.

Bohužel, malá Fermatova věta je jen implikace. Takže pokud by platilo $2^{n-1} \equiv 1 \pmod{n}$, pak n prvočíslo být může, ale nemusí. Jsou zvláštní čísla, která $2^{n-1} \equiv 1 \pmod{n}$ splňují, ale jsou složená, říkáme jim **pseudoprvočísla**. Dobrá zpráva je, že pseudoprvočísel je velice málo, například mezi prvními 10^{10} přirozenými čísla je cca 450,000,000 prvočísel, ale jen cca 15,000 pseudoprvočísel. To znamená, že tento test je vysoce účinný při vyřazování čísel, která prvočísky nejsou, a pokud už nějaké n tímto testem projde, tak je vysoká pravděpodobnost, že jsme opravdu ulovili prvočíslo, a vyplatí se investovat další námahu na skutečné potvrzení této skutečnosti.

Tato myšlenka se dá samozřejmě rozvést dále. Řekneme, že n je pseudoprvočíslo vzhledem k základu a , jestliže $a^{n-1} \equiv 1 \pmod{n}$. Takže pokud nějaké pseudoprvočíslo přežije první test, zvolíme nějaké nesoudělné a , například další prvočíslo $a = 3$, a zkusíme, zda neplatí $a^{n-1} \equiv 1 \pmod{n}$. To při troše štěstí zase vyřadí neprvočíslo.

Řekneme, že n je Carmichaelovské číslo, jestliže $a^{n-1} \equiv 1 \pmod{n}$ pro všechna $a \in \mathbb{N}$ s $\gcd(a, n) = 1$. Např. 561 je takové číslo. Těchto je sice nekonečně mnoho, ale zase hrozně málo, čili když začneme s n a protestujeme jej pro hodně a , tak v případě úspěchu už je skoro jisté, že n je prvočíslo.

Je to dobrá strategie pro hledání velkých prvočísel, například pro kódování RSA. Dělá se to tak, že si prostě zvolíme nějaké vhodně dlouhé číslo (liché a nekončící pětkou, samozřejmě). Testovat přímo dělením, zda je to prvočíslo, by trvalo strašně dlouho (mluvíme zde o desítkách let na těch nejvýkonnějších počítačích). Místo toho jej rychle proženeme testy popsanými výše a ono asi vypadne jako složené číslo. Tak zkusíme jiné velké číslo (třeba přičteme 2 k tomu neúspěšnému) a jedeme znova. Nakonec nějaké číslo těmi testy projde, pak je téměř jisté, že je to prvočíslo. Tak prostě z takového čísla zkusíme RSA kódování udělat, zkusmo něco zakódujeme a rozkódujeme a pokud to vyjde, tak jsme našli, co jsme potřebovali.

△

7a.14 Poznámka (kritéria dělitelnosti): V kapitole 6 jsme se zmínili o existenci kritérií dělitelnosti, ale pořádně se na ně podíváme až zde, protože počítání modulo občas nabídne pohodlný zápis. Využijeme pak toho, že n dělí a právě tehdy, pokud $a \equiv 0 \pmod{n}$. Známá kritéria se dají rozdělit do skupin podle toho, z jaké myšlenky vycházejí. Ukážeme si několik populárních myšlenek, obvykle nejprve na kritériu známém a pak se pokusíme zjistit něco o dělitelnosti sedmičkou.

Jedna skupina kritérií vychází z toho, že si dané číslo a napíšeme jako $a = 100A + r$, kde $r = a \bmod 100$. Při pohledu vzhledem k počítání modulo d občas objevíme zajímavé věci. Jako ukázkou dokážeme kritérium dělitelnosti čtyřkou, viz 6a.11. Modulo 4 totiž máme

$$a = 100A + r \equiv 0A + r = r \pmod{4}.$$

Vidíme, že $a \equiv 0 \pmod{4}$ právě tehdy, když $r \equiv 0 \pmod{4}$, jinak řečeno, číslo a je dělitelné čtyřkou právě tehdy, když je čtyřkou dělitelné r neboli poslední dvojcíslí čísla a . Podobně se dokazuje kritérium pro $d = 25$, pomocí rozpisu $a = 10A + r$ takto snadno dokážeme kritérium pro dělitelnost dvojkou nebo pětkou.

Co dostaneme, když počítáme modulo 7? $a = 10A + r \equiv 3A + r$. Chceme-li tedy vědět, zda je číslo dělitelné sedmičkou, oddělíme poslední cifru a přičteme ji k trojnásobku „začátku“, pak otestujeme nové. To zní překně, ale je to pracné. Pokud bychom potřebovali znát dělitelnost čísla $a = 87654$, toto kritérium nabízí testovat místo toho číslo $3 \cdot 8765 + 4$, to se mi ani nechce počítat. Je to ale východisko k zajímavému algoritmu. Číslo se dá totiž probírat principem $10x + y \mapsto 3x + y$ postupně zleva (detaily raději vynecháme), címž vznikne tento postup:

- Vezmi levou cifru, vynásob třemi a přičti druhou cifru zleva. Výsledné číslo vynásob třemi a přičti třetí cifru zleva, to vynásob třemi a přičti čtvrtou cifru zleva atd., dokud se nedojde k poslední (pravé) cifře. Výsledné číslo je dělitelné sedmi přesně tehdy, když to původní. Vždy po ukončení kroku (přičtení, před násobením třemi) je možné přejít ke zbytku modulo 7.

Ukážeme pro $a = 87654$. Nejprve $3 \cdot 8 + 7 = 31$, zbytek je 3. Pak $3 \cdot 3 + 6 = 15$, zbytek je 1. Pak $3 \cdot 1 + 5 = 8$, pak $3 \cdot 8 + 4 = 28$. Toto je výsledné číslo. Je dělitelné sedmi, proto je i $a = 87654$ dělitelné sedmi.

Další populární rodinka kritérií vychází z dekadického rozvoje čísla. Jako inspiraci si ukážeme, proč funguje kritérium dělitelnosti trojkou. Když se na dané číslo v dekadickém tvaru $a = \sum_k a_k 10^k$ podíváme modulo 3, můžeme podle věty o kongruenci a operacích nahrazovat jednotlivé části.

$$a = \sum_k a_k 10^k \equiv \sum_k a_k \cdot (10 \pmod{3})^k = \sum_k a_k \cdot 1^k = \sum_k a_k \pmod{3}.$$

Vidíme, že číslo a je dělitelné třemi právě tehdy, pokud je dělitelný ciferný součet. Podobný důkaz ukáže i známá kritéria pro dělitelnost devíti a jedenácti. Dokonce bychom mohli aplikovat modulo i na cifry samotné, tedy $a = \sum_k (a_k \pmod{3})$. Je tedy možné rovnou sčítat namísto cifer jejich zbytky po dělení třemi.

Pomohlo by to se sedmičkou? Modulo 7 dostáváme $a = \sum_k a_k \cdot (10 \pmod{7})^k = \sum_k a_k \cdot 3^k$. Namísto čísla $a = 87654$ bychom mohli testovat číslo $8 \cdot 3^4 + 7 \cdot 3^3 + 6 \cdot 3^2 + 5 \cdot 3^1 + 4$, ani to se mi nechce počítat. Přesto to není zcela slepá ulička. Pokud se podíváme, jaké jsou zbytky čísel 10^k po dělení sedmi, dostáváme cyklickou posloupnost $1, 3, 2, 6, 4, 5, 1, 3, 2, \dots$. Můžeme tedy sčítat cifry daného a (bráno zprava) násobené těmito váhami. Takže namísto $a = 87654$ lze testovat číslo $4 \cdot 1 + 5 \cdot 3 + 6 \cdot 2 + 7 \cdot 6 + 8 \cdot 4 = 105$. To dělitelné sedmi je, což nám potvrzuje, že opravdu $7|87654$. Toto kritérium je asi méně příjemné než předchozí algoritmus, ale také se používá.

Lepší trik dostaneme, když dané číslo nebudeme dělit na cifry, ale na větší skupiny cifer. Dvojice ještě moc nepomohou, vedou na $a \equiv \sum (a_k \pmod{7}) \cdot 2^k \pmod{7}$. Když dané číslo rozložíme na trojcíslí, $a = \sum_k a_k 1000^k$, dostáváme modulo 7 rovnost

$$a \equiv \sum (a_k \pmod{7}) \cdot (1000 \pmod{7})^k = \sum (a_k \pmod{7}) \cdot (-1)^k \pmod{7}.$$

Chceme-li tedy vědět, zda je číslo a dělitelné sedmi, tak místo toho můžeme testovat číslo vytvořené takto: první trojcíslí zprava nahradíme zbytkem po dělení sedmi a vezmeme se znaménkem plus. Od toho odečteme zbytek po dělení druhého trojcíslí zprava sedmi. K tomu přičteme zbytek po dělení třetího trojcíslí zprava sedmi atd. V případě $a = 87654$ bychom místo toho testovali číslo $(654 \pmod{7}) - (87 \pmod{7}) = 3 - 3 = 0$, to je dělitelné sedmi a potvrzujeme nezávisle, že 87654 je dělitelné sedmičkou.

Asi nejpoužívanější kritérium pro dělitelnost sedmi vypadá ještě jinak. Zapišme zase $a = 10A + r$. Tvrdíme, že je dělitelné sedmi právě tehdy, pokud je sedmi dělitelné číslo $A - 2r$. Důkaz by vypadal třeba takto. Protože je 2 nesoudělná se sedmi, pak má číslo $2a$ stejnou dělitelnost sedmi jako a . Když od čísla a odečteme násobek sedmičky, také jeho dělitelnost sedmi nezměníme, takže číslo $2a - 21A = 2r - A$ má zase stejnou dělitelnost jako a . Z praktického důvodu je pak lepší ještě změnit znaménko.

Obvyklá ukázka: Chceme znát dělitelnost čísla $a = 87654$, místo toho koukneme na $8765 - 2 \cdot 4 = 8757$, pak na $875 - 2 \cdot 7 = 861$, pak na $86 - 2 \cdot 1 = 84$ a zde již vidíme, že jde o číslo dělitelné sedmičkou.

Toto kritérium má obdobu i pro dělitelnost třinácti, sedmnácti a podobně, tak si ukážeme obecnýustr.

Odvodíme kritérium pro dělitelnost $a = 10A + r$ číslem $d \in \mathbb{N}$. Začneme tím, že najdeme číslo c tak, aby bylo nesoudělné s d , ale aby d dělilo $10c + 1$. Díky nesoudělnosti víme, že $d|a$ právě tehdy, když $d|(ca)$. Pak si šikovně napíšeme

$$ca = 10Ac + cr = (10c + 1)A - (A - cr).$$

Výraz nalevo je násobkem d právě tehdy, pokud je jím výraz napravo. Protože ale podle předpokladu d dělí $(10c + 1)A$, tak o všem rozhodne výraz $A - cr$.

Volba $c = 2$ dá již výše zmíněné kritérium pro sedmičku, $c = -4$ zase vede na kritérium pro třináctku a podobně.

Poněkud jednodušší a známá kritéria připomínáme ve cvičení.

△

7a.15 Strukturální teoretický pohled na počítání modulo

Existují dva pohledy na svět zvaný \mathbb{Z}_n , každý má své výhody a nevýhody a autoři si volí ten, který se jim lépe hodí do koncepce knihy. Zatím jsme představili definici, která je praktičtější, dá se říct, že pokud člověk počítání modulo používá při práci, tak je tento pohled na \mathbb{Z}_n ten pravý.

Existuje ještě pohled jiný, který je teoretičtější (pro mnohé zbytečně), ale zase toho o množině \mathbb{Z}_n více řekne, občas ušetří trochu práce a přednost mu dávají autoři, kteří se na \mathbb{Z}_n dívají jako na zajímavou matematickou strukturu. Teď si jej představíme, myslím, že pro pokročilejšího čtenáře to může být zajímavá alternativa, ale pro ty, kteří mají méně vyvinutý smysl pro abstrakci, to může být spíš matoucí. Pokud by čtenář při četbě začínal mít pocit, že se topí (pokud jej už tedy nemá), nic se nestane, když skočí buď k Eulerově větě, nebo dokonce rovnou na cvičení. Přesto doporučuji, aby se alespoň zastavil u Věty .

Jdeme na to.

!

Fakt 7a.16.

Pro každé $n \in \mathbb{N}$ je relace „být kongruentní modulo n “ ekvivalence na \mathbb{Z} .

Důkaz (rutinní): Reflexivitu a symetrii necháme jako cvičení, zde ukážeme tranzitivitu.

Jestliže $a \equiv b \pmod{n}$ a $b \equiv c \pmod{n}$, pak $a = b + kn$ pro nějaké $k \in \mathbb{Z}$ a $b = c + ln$ pro nějaké $l \in \mathbb{Z}$. Odtud $a = c + (k + l)n$ a $(k + l) \in \mathbb{Z}$, tedy $a \equiv c \pmod{n}$. □

Můžeme teď aplikovat teorii z kapitoly a vidíme, že se nám množina \mathbb{Z} rozpadne na třídy ekvivalence, kterým říkáme **zbytkové třídy**. Zbytkovou třídu čísla a vzhledem k relaci kongruence modulo n budeme značit $[a]_n$.

Například při volbě $n = 4$ máme třeba $[1]_4 = [5]_4 = [-3]_4 = \dots = \{\dots, -7, -3, 1, 5, 9, \dots\}$. Obecně samozřejmě $[a]_n = \{a + kn; k \in \mathbb{Z}\}$, mimo jiné vždy $a \in [a]_n$. Výsledky z kapitoly říkají, že je úplně jedno, kterého zástupce si vybereme, také víme, že $[a]_n = [b]_n$ právě tehdy, když $a \equiv b \pmod{n}$ neboli když dají stejný zbytek při dělení n neboli když je $a - b$ dělitelné n neboli když se od a k b dokážeme dostat opakováním přičítáním/odčítáním n .

Množina \mathbb{Z} se nám rozpadla na přesně n zbytkových tříd, čímž nám vznikl nový objekt, množina těchto zbytkových tříd. Naším cílem je s touto množinou normálně pracovat, tedy brát zbytkové třídy jako objekty, se kterými normálně manipuluje, jako jsme to zvyklí dělat s čísly. To je oblíbená matematická věc, vezme se nějaký třeba i dosti komplikovaný objekt (v našem případě nekonečná množina navzájem kongruentních čísel) a ten se vhodně zabalí a označí písmenkem, takže se navenek tváří jako nějaká obyčejná věc, se kterou pak v pohodě manipuluje pomocí odvozených pravidel.

V tomto případě bychom se zbytkovými třídami rádi prováděli běžné algebraické operace. K tomu využijeme, že na každou zbytkovou třídu se lze odvolat nějakým zástupcem. Je čas zadefinovat hlavní pojmy.

!

Definice.

Prostor \mathbb{Z}_n definujeme jako množinu všech tříd ekvivalence v \mathbb{Z} vzhledem k relaci být kongruentní modulo n , tedy $\mathbb{Z}_n = \{[a]_n; a \in \mathbb{Z}\}$.

Pro $[a]_n, [b]_n \in \mathbb{Z}_n$ definujeme

$$\begin{aligned}[a]_n \oplus [b]_n &= [a + b]_n, \\ [a]_n \odot [b]_n &= [a \cdot b]_n.\end{aligned}$$

Další časté značení pro tuto množinu tříd ekvivalence je $\mathbb{Z}/n\mathbb{Z}$. Je lepší z hlediska formálního, protože je to obecné značení pro faktorovou množinu (což zde vlastně děláme), na druhou stranu je delší. Nakonec jsme zvolili pohodlnější verzi \mathbb{Z}_n , zejména proto, že záhy ukážeme, že je to vlastně naše stará dobrá \mathbb{Z}_n v jiném převleku.

Než začneme s touto definicí pracovat, je třeba hněd vyjasnit jeden potencionální problém. Není totiž vůbec jasné, jestli jsme operace na zbytkových třídách definovali tak, aby to mělo smysl. Například umíme vypočítat $[5]_8 \odot [6]_8 = [30]_8$, ale víme, že třeba $[5]_8 = [13]_8$, je to tentýž objekt, tatáž třída. Co kdybychom tedy ve výpočtu použili 13 namísto 5? Bude $[13]_8 \odot [6]_8$ dávat stejný výsledek, tedy stejnou třídu jako původní výpočet? Ano, protože $[13]_8 \odot [6]_8 = [78]_8 = [30]_8$, neboť $8 | (78 - 30)$. Takže to vyšlo, ale to byl jen jeden příklad. Aby byla definice korektní, musíme ukázat obecně, že konkrétní volba zástupce nemá vliv na výsledek operace.

Věta 7a.17.

Nechť $n \in \mathbb{N}$. Uvažujme $a, b, u, v \in \mathbb{Z}$ takové, že $[a]_n = [u]_n$ a $[b]_n = [v]_n$. Pak $[a+b]_n = [u+v]_n$ a $[a \cdot b]_n = [u \cdot v]_n$.

Důkaz (rutinní): $[a]_n = [u]_n$ znamená $a \equiv u \pmod{n}$, podobně pro b a v , pak nám Věta dává

$$a + b \equiv u + v \pmod{n} \text{ neboli } [a + b]_n = [u + v]_n.$$

Důkaz pro součin je obdobný.

□

Takže už víme, že definice není sporná, zkusíme si to.

Příklad 7a.n: Například v \mathbb{Z}_5 máme $[3]_5 \odot [4]_5 = [3 \cdot 4]_5 = [12]_5$. Komu se nelibí tento zástupce, může udělat třeba $[3]_5 \odot [4]_5 = [12]_5 = [2]_5 = [-3]_5 = [127]_5 = \dots$

Nebo třeba v \mathbb{Z}_{13} je $[8]_{13} \oplus [5]_{13} = [13]_{13} = [0]_{13}$, tohle mimochodem ukazuje, že $-[8]_{13} = [5]_{13}$.

△

! Máme teď zajímavou situaci, množina \mathbb{Z}_n je definována dvakrát. Samozřejmě se ukáže, že vlastně jde o totéž. Začneme tím, že každá zbytková třída je dána nějakým svým zástupcem. Můžeme se rozhodnout, že budeme jako zástupce zásadně vybírat právě zbytky po dělení číslem n , pak dostáváme $\mathbb{Z}_n = \{[0]_n, [1]_n, \dots, [n-1]_n\}$. Kdybychom zavedli „ k “ jako symbol pro $[k]_n$, tak máme přesně množinu $\mathbb{Z}_n = \{0, 1, 2, \dots, n-1\}$. Je to ale zatím jen podobnost vizuální. Abychom ukázali, že jde opravdu o zcela stejné struktury, musíme ještě ukázat, že si odpovídají i operace.

Vezměme $a, b \in \mathbb{Z}_n$ (dle staré definice) a uvažujme $[a]_n, [b]_n$ jako prvky z nového \mathbb{Z}_n . Ve starém i novém \mathbb{Z}_n provedeme operace sčítání, popř. násobení, a ukážeme, že to v obou světech dopadne stejně.

Podle první definice dostáváme $a \oplus b = (a+b) \pmod{n}$ a $a \odot b = (ab) \pmod{n}$. Když totéž provedeme v nové definici přes třídy, dostáváme výsledky $[a]_n \oplus [b]_n = [a+b]_n$ a $[a]_n \odot [b]_n = [ab]_n$. My jsme se ale rozhodli třídy zastupovat vždy zbytky po dělení n , tudíž $[a]_n \oplus [b]_n = [(a+b) \pmod{n}]_n$ a dostáváme stejný výsledek jako u první definice, podobně $[a]_n \odot [b]_n = [(ab) \pmod{n}]_n$ a zase máme stejný výsledek.

Takže se dá říct, že nová definice vlastně znamená, že kolem prvků z první definice \mathbb{Z}_n nakreslíme ohrádky, operace děláme v zásadě stejně. Původní definice nás ještě nutí po provedení výpočtu přejít ke zbytkům coby zástupcům ze \mathbb{Z}_n , zatímco nová definice ne, to je jedna z jejích výhod. Nechává nám svobodu vybírat si zástupce tak, aby se nám co nejlépe počítalo. Například opačný prvek k 5 v \mathbb{Z}_{137} najdeme dle nové definice jako $[-5]_{137}$, nemusíme používat 132 podle první definice \mathbb{Z}_{137} .

Všechny výpočty modulo z předchozích příkladů o \mathbb{Z}_n by se proto daly přepsat jako výsledky o třídách, třeba v příkladě jsme našli inverzní prvek k $a = 36$ v monoidu \mathbb{Z}_{175} , $36^{-1} = 141$. V novém znění můžeme říct, že $[36]_{175}^{-1} = [141]_{175}$. Opravdu, $[36]_{175} \cdot [141]_{175} = [36 \cdot 141]_{175} = [5076]_{175} = [1]_{175}$, protože $5076 \equiv 1 \pmod{175}$. Krásně to souhlasí.

Všechny dosavadní výsledky o \mathbb{Z}_n bychom teď mohli přepsat do nového jazyka, ale nestojí to za to, uvedeme jen jedno klíčové tvrzení.

! Důsledek 7a.18.

Nechť $n \in \mathbb{N}$, uvažujme $[a]_n \in \mathbb{Z}_n$.

- (i) Vždy existuje prvek opačný $-[a]_n = [n-a]_n$.
- (ii) $[a]_n$ je invertibilní vůči \odot právě tehdy, když jsou a a n nesoudělné.

Přímý důkaz vzorce pro opačný prvek: $[a]_n \oplus [n-a]_n = [a + (n-a)]_n = [n]_n = [0]_n$, neboť $n \equiv 0 \pmod{n}$.

Takže například $-[3]_7 = [4]_7$ a $-[3]_{13} = [10]_{13}$. Všimněte si, že na rozdíl od počítání v \mathbb{Z}_n podle té první definice si teď nemusíme dělat speciální případ pro inverzní prvek k nule. Vzorec dává výsledek $[n]_n$, což je správně, neboť $[n]_n = [0]_n$.

Invezní prvky hledáme stejně jako předtím. Od inverzního prvku $[x]_n$ k prvku $[a]_n$ očekáváme, že $[a]_n \cdot [x]_n = [1]_n$ neboli $[ax]_n = [1]_n$ neboli $ax \equiv 1 \pmod{n}$ neboli $ax + kn = 1$ pro nějaké $k \in \mathbb{Z}$. Zase tedy přijde ke slovu rozšířený Euklidův algoritmus, který rovnou dá správnou třídu $[x]_n$, protože nemusíme hledat zbytek po dělení n coby vzorového zástupce. Pokud to ale uděláme, dopadne to přesně jako u \mathbb{Z}_n podle první definice. Dostáváme se zpět k tomu, že jde vlastně o tutéž věc.

Zatím to vypadá, že vlastně jen dokreslujeme symboly kolem čísel, což vypadá jako zbytečná práce. Má to nějaké výhody? Jednu už jsme viděli, nenutí nás to hledat správné zástupce. Hlavní výhoda je ale ještě jinde. Protože nová definice používá pokročilejší matematické struktury, můžeme využívat rozličné nástroje, které nám byly u původního přístupu inspirovány praktickým počítáním nepřistupné.

První ukázkou je splacení dluhu, který máme. Již jsme ve výpočtech v \mathbb{Z}_n použili běžná pravidla, na která jsme zvyklí od reálných čísel, jenže jsme ještě nepotvrzili, že opravdu platí. V původní definici by to dalo trochu práce, v té nové je „zdědíme“ téměř bez práce.

!

Věta 7a.19.

Nechť $n \in \mathbb{N}$. Pak platí následující:

- (i) pro všechna $a, b \in \mathbb{Z}$ platí $[a]_n \oplus [b]_n = [b]_n \oplus [a]_n$,
- (ii) pro všechna $a, b, c \in \mathbb{Z}$ platí $[a]_n \oplus ([b]_n \oplus [c]_n) = ([a]_n \oplus [b]_n) \oplus [c]_n$,
- (iii) pro všechna $a \in \mathbb{Z}$ platí $[a]_n \oplus [0]_n = [a]_n$,
- (iv) pro každé $a \in \mathbb{Z}$ platí $[a]_n \oplus [-a]_n = [0]_n$,
- (v) pro všechna $a, b \in \mathbb{Z}$ platí $[a]_n \odot [b]_n = [b]_n \odot [a]_n$,
- (vi) pro všechna $a, b, c \in \mathbb{Z}$ platí $[a]_n \odot ([b]_n \odot [c]_n) = ([a]_n \odot [b]_n) \odot [c]_n$,
- (vii) pro všechna $a \in \mathbb{Z}$ platí $[a]_n \odot [1]_n = [a]_n$,
- (viii) pro všechna $a, b, c \in \mathbb{Z}$ platí $[a]_n \odot ([b]_n \oplus [c]_n) = ([a]_n \odot [b]_n) \oplus ([a]_n \odot [c]_n)$.

Důkaz (poučný): Všechny vlastnosti platí pro počítání na \mathbb{Z} a díky chytré definice operací v \mathbb{Z}_n se přenesou na operace \oplus a \odot . Ukážeme to pro komutativitu neboli (i): $[a]_n \oplus [b]_n = [a + b]_n = [b + a]_n = [b]_n \oplus [a]_n$. Jen mírně komplikovanější je třeba (viii):

$$[a]_n \odot ([b]_n \oplus [c]_n) = [a]_n \odot [b + c]_n = [a(b + c)]_n = [ab + ac]_n = [ab]_n \oplus [ac]_n = ([a]_n \odot [b]_n) \oplus ([a]_n \odot [c]_n).$$

□

To znamená, že množina \mathbb{Z}_n je velice příjemná z abstraktního pohledu an operace, viz kapitola , kterou by už bylo opravdu dobré si teď přečíst. Konkrétně tato Věta ukazuje, že $(\mathbb{Z}_n, \oplus, [0]_n)$ je komutativní grupa a $(\mathbb{Z}_n, \odot, [1]_n)$ je komutativní monoid, $(\mathbb{Z}_n, \oplus, \odot)$ je pak komutativní okruh (viz kapitola). Díky tomu automaticky dostáváme spoustu vlastností, které jsme předtím museli dokazovat, například jednoznačnost inverzního prvku (pokud existuje).

Přeložíme do jazyka algebry ještě další poznatek, který plyne z Důsledku .

!

Věta 7a.20.

Nechť $n \in \mathbb{N}$, uvažujme \mathbb{Z}_n . Jestliže je n prvočíslo, tak je každý prvek $[a]_n \in \mathbb{Z}_n$ splňující $[0]_n \neq [a]_n$ invertibilní. To znamená, že pro prvočíslo n je $(\mathbb{Z}_n, \oplus, \odot)$ těleso.

Jinak řečeno, pro prvočísla n se v \mathbb{Z}_n pracuje prakticky stejně jako v reálných číslech, všechno funguje báječně, dokonce můžeme dělit nenulovými prvky.

7a.21 Poznámka: Kombinací chytré nové definice \mathbb{Z}_n a obecných poznatků o množinách s operacemi z kapitoly dostáváme také zajímavý důkaz malé Fermatovy věty .

Nejprve si ale musíme rozmyslet, co je to $[a]^k_n$. To je zkratka pro $[a]_n \odot [a]_n \odot \dots \odot [a]_n$, což je podle definice násobení rovno $[a \cdot a \cdots a]_n = [a^k]_n$. Rovnost $a^{n-1} \equiv 1 \pmod{n}$ tedy znamená, že v řeči zbytkových tříd máme $[a^{n-1}]_n = [1]_n$ neboli $[a]_n^{n-1} = [1]_n$. Jsme připraveni.

Dokážeme, že když je n prvočíslo a a je nesoudělné s n , pak $[a]_n^{n-1} = [1]_n$.

Podle Věty je množina všech invertibilních prvků \mathbb{Z}_n rovna $G = \{[1]_n, \dots, [n-1]_n\}$. Podle Věty je (G, \odot) grupa. Její mohutnost je $n - 1$, tudíž podle Důsledku platí pro každý prvek $[a]_n \in G$, že $[a]_n^{n-1} = [1]_n$.

Hotovo.

Důkaz byl tedy kratší a zdánlivě snažší než ten první, což je způsobeno tím, že využívá spousty práce, která byla udělána v kapitole .

△

Eulerova věta.

Tím končí hlavní blok této kapitoly, pro pokročilé přidáme ještě část další, která se pokusí pomoci s následujícím problémem: Umíme už efektivně počítat mocniny pro případ, že je n prvočíslo. Co dělat v případě, kdy n prvočíslo není? Jinými slovy, co se pak stane s malou Fermatovou větou?

Na to budeme muset hlouběji zabrousit do teorie čísel, jmenovitě se pořádněji podívat na otázku, kolik je v množině $\{1, 2, \dots, n\}$ čísel nesoudělných s n . Nejprve si to nazveme.

Definice.

Eulerova funkce (Euler function or totient) φ je definována takto: Pro $n \in \mathbb{N}$ je $\varphi(n)$ rovno počtu přirozených čísel, která jsou menší než n a nesoudělná s n .

Kolik je třeba $\varphi(6)$? Snadno nahlédneme, že z množiny $\{1, 2, 3, 4, 5\}$ jsou jen čísla 1, 5 nesoudělná s šestkou, proto $\varphi(6) = 2$. Z množiny $\{1, 2, 3, 4, 5, 6, 7, 8\}$ jsou s devítkou nesoudělná čísla 1, 2, 4, 5, 7, 8, proto $\varphi(9) = 6$. Snadno si

rozmyslíme, že pro prvočíslo p je $\varphi(p) = p - 1$, protože nesoudělná s prvočíslem p jsou všechna čísla $1, 2, \dots, p - 1$. K čemu tato Eulerova funkce je? Díky ní hravě dokážeme zobecnění malé Fermatovy věty.

Věta 7a.22. (Eulerova věta)

Nechť $n \in \mathbb{N}$. Jestliže je $a \in \mathbb{N}$ nesoudělné s n , pak $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Důkaz (poučný): Uvažujme $G = \{[a]_n; [a]_n \text{ invertibilní v } \mathbb{Z}_n\}$. Podle Důsledku (ii) víme, že $[a]_n \in G$ právě tehdy, když je a nesoudělné s n . To znamená, že G je dána těmi $a \in \{1, 2, \dots, n - 1\}$, které jsou nesoudělné s n , proto $|G| = \varphi(n)$.

Podle Věty je (G, \odot) grupa, tudíž podle Důsledku platí pro každý prvek $[a]_n \in G$, že $[a]_n^{\varphi(n)} = [1]_n$. Jinými slovy, pro každé a nesoudělné s n platí $a^{\varphi(n)} \equiv 1 \pmod{n}$. □

Tato věta nám tedy umožňuje efektivně umocňovat i při práci v \mathbb{Z}_n pro n neprvočíselné, jen je třeba umět nacházet $\varphi(n)$. Už víme, že pro prvočíslo n máme $\varphi(n) = n - 1$, takže z této věty dostáváme jako důsledek malou Fermanovu větu.

Mimochodem, někdy se dá najít i menší číslo než $\varphi(n)$ tak, aby umocnilo a na jedničku. Například pro prvočíslo $n = 31$ nám malý Fermat i Euler zaručí, že $2^{30} \equiv 1 \pmod{31}$, ale máme i $2^5 \equiv 1 \pmod{31}$. To není v rozporu s Eulerovou větou, ta jen zaručí existenci vhodné mocniny a nikde netvrdí, že našla tu nejlepší.

Hledání hodnot Eulerovy funkce také není zrovna snadné a odložíme to na konec kapitoly, abychom čtenáře předčasně nevyčerpali. Teď si ukážeme pár aplikací.

Příklad 7a.o: Vypočítáme $6^{1040} \pmod{91}$. Již tradičně nejde použít kalkulačku, toto číslo má totiž 810 cifer, takže by udolalo i vše, co má typický počítač k dispozici. Nadšenci si napíšou vlastní rutiny na operace s takto dlouhými čísly, my raději zkusíme redukovat zlomek. Začneme pokročilými nástroji.

Protože 91 není prvočíslo, nelze použít malou Fermatovu větu, ale je třeba použít Eulerovu větu, což možné je, protože 6 je nesoudělné s 91. Podle Věty a definice spočítáme $\varphi(91) = \varphi(7 \cdot 13) = \varphi(7) \cdot \varphi(13) = 6 \cdot 12 = 72$. Ještě si spočítáme $q = \lfloor \frac{1040}{72} \rfloor = \lfloor 14.44.. \rfloor = 14$ a můžeme počítat modulo 91.

$$6^{1040} = 6^{14 \cdot 72 + 32} = (6^{72})^{14} \cdot 6^{32} \equiv 1 \cdot 6^{32}.$$

To pořád není žádná sranda (25 cifer, moje kalkulačka nemá se svými 13 pamatovanými ciframi šanci), Euler už nepomůže, takže je čas na jednoduché metody, jmenovitě postupné snižování mocniny. Umíme relativně snadno mocnit $6^4 = 36^2 = 1296 \equiv 22 \pmod{91}$, díky čemuž

$$6^{1040} \equiv \dots \equiv 6^{32} = (6^4)^8 \equiv 22^8 \pmod{91}.$$

Toto je desetimístné číslo, to už kalkulačka zvládne, vydelením, zaokrouhlením podílu dolů a odečtením příslušného násobku konečně nacházíme zbytek 27.3. To nevypadá moc dobře, je to tím, že jsme použili funkci mocninu, tedy zmáčknuli jsme 22, pak „ x^y “ a pak 8, takové obecné mocniny počítá kalkulačka přes logaritmy a chybky se projevují. Zkusíme to znova pomocí klávesy „ x^2 “, která opravdu násobí čísla, pak $((22^2)^2)^2 - 603031577 \cdot 91 = 29$. Dostáváme celé číslo, ale raději ještě kalkulačce pomůžeme.

$22^2 = 484 \equiv 29 \pmod{91}$, proto $6^{1040} \equiv \dots \equiv 22^8 = (22^2)^4 \equiv 29^4 \pmod{91}$. Tohle už na kalkulačce bezpečně utlučeme, $6^{1040} \equiv 29^4 = 707281 \equiv 29 \pmod{91}$. Takže téměř 29 věříme.

Pro alternativní způsob výpočtu se podívejte na příklad .

△

Poznámka: Díky rozličným trikům (redukce mocniny oddělováním dvojký, malým Fermatem, Eulerovinami) je umocňování modulo mnohem příjemnější než umocňování běžné. Navíc se dá jakékoli umocňování velice zrychlit následující fintou pro nalezení a^b :

Díky opakovanému násobení umíme rychle najít mocniny $a^1 = a$, a^2 , $(a^2)^2 = a^4$, $(a^4)^2 = a^8$, $(a^8)^2 = a^{16}$ atd., tedy mocniny typu a^{2^i} . Stačí si tedy vyjádřit b v dvojkové soustavě, $b = \sum_{i=0}^m b_i 2^i$, viz příklad 6a.d, a dostáváme $a^b = \prod (a^{2^i})^{b_i}$. To vypadá komplikovaně, ale b_i nabývá pouze hodnot 0 či 1, takže $a^b = \prod_{b_i=1} a^{2^i}$.

Například $a^{21} = a^{16+4+1} = a^{16}a^4a^1 = (((a^2)^2)^2)(a^2)a^1$. Takovéto mocnění je velice rychlé, viz cvičení . Při počítání modulo se navíc při všech krocích přechází k příjemnějším zástupcům, takže díky kombinaci všech triků se z mocniny stává relativně nenáročná operace.

△

Na začátku této kapitoly jsme ukázali, že při výpočtech modulo n lze v běžných algebraických operacích a v základu mocniny čísla nahrazovat jejich kongruentními bratříčky. Pomocí Eulerovy věty se ukáže, že to lze dělat i s exponentem, ale musí se použít jiný modulus.

Důsledek 7a.23.

Nechť $n \in \mathbb{N}$, uvažujme $a \in \mathbb{N}$ nesoudělné s n . Pak pro všechna $x, y \in \mathbb{N}$ platí: Jestliže $x \equiv y \pmod{\varphi(n)}$, pak $a^x \equiv a^y \pmod{n}$.

Důkaz (rutinní): $x \equiv y \pmod{\varphi(n)}$ znamená, že $x = y + k\varphi(n)$ pro nějaké $k \in \mathbb{Z}$.

Pak $a^x = a^{y+k\varphi(n)} = a^y(a^{\varphi(n)})^k \equiv a^y 1^k = a^y \pmod{n}$. □

Ve zbytku kapitoly se podíváme na výpočet $\varphi(n)$, což je mimochodem dost náročná úloha. Dělat si seznam čísel a hledat v něm ta nesoudělná s n rozhodně není pro větší n vhodnou metodou, raději bychom nějaké vzorečky. Jeden už jsme odvodili pro prvočísla, $\varphi(p) = p - 1$, teď se podíváme na další.

Fakt 7a.24.

Nechť p je prvočíslo. Pak pro všechna $k \in \mathbb{N}$ platí $\varphi(p^k) = p^k - p^{k-1} = (p - 1)p^{k-1}$.

Důkaz (poučný): Která čísla menší než p^k mají netriviálního společného dělitele s p^k ? Jestliže je d dělitel p^k , tak podle Lemma 6b.5 musí být ve tvaru $d = p^i$, tudíž čísla, která mají netriviálního společného dělitele s p^k , jsou přesně čísla dělitelná nějakou mocninou p^k . Rozmyslete si, že to jsou všechny násobky p . Kolik takových čísel je? Jsou to čísla $p, 2p, 3p, \dots, (p^{k-1} - 1)p$, protože další násobek je $p^{k-1}p = p^k$, ten už je moc velký. Je tedy $p^{k-1} - 1$ čísel menších než p^k a soudělných s p^k . Celkem je $p^k - 1$ přirozených čísel menších než p^k , proto těch nesoudělných je $(p^k - 1) - (p^{k-1} - 1) = p^k - p^{k-1}$. □

Abychom našli vzorec pro všechna čísla, musíme si nejprve připravit půdu.

Lemma 7a.25.

Nechť $n \in \mathbb{N}$. Pak pro každé $a \in \mathbb{Z}$ platí $\gcd(a, n) = \gcd(a \bmod n, n)$.

Toto lemma je vlastně jen přepsané Lemma 6a.15 s volbou $b = n$, neboť pak $r = a \bmod n$.

Lemma 7a.26.

Nechť $m, n \in \mathbb{N}$ a $a \in \mathbb{N}$. Pak $\gcd(a, mn) = 1$ právě tehdy, když $\gcd(a \bmod m, m) = 1$ a $\gcd(a \bmod n, n) = 1$.

Důkaz: Jde jen o Lemma 6b.3 přepsané pomocí předchozího lemmatu. □

Potřebujeme ještě jedno lemma, tentokráté poněkud vydatné. Abychom usnadnili jeho strávení, uvedeme nejprve inspirační podobenství.

7a.27 Souřadnice. Čtenář jistě zná pojem souřadnicového systému v \mathbb{R}^2 . Je-li dán vektor $\vec{u} = (x, y)$, tak jej dokážeme rozložit do základních směrů $\vec{e} = (1, 0)$, $\vec{f} = (0, 1)$ a naopak jej z těch složek zase poskládat zpět, $\vec{u} = x \cdot \vec{e} + y \cdot \vec{f}$. Čísla x a y nám říkají, jak velká část z vektoru \vec{u} působí ve směrech \vec{e} a \vec{f} .

Toto funguje zcela obecně, můžeme si v rovině zvolit i jiné dva vektory \vec{e}, \vec{f} (nesmí být rovnoběžné) a stejný postup bude fungovat, jen už velikost působení \vec{u} v těchto dvou směrech nebude rovna souřadnicím x, y , ale nějakým jiným číslům, kterým zcela přirozeně říkáme souřadnice \vec{u} vůči těmto novým vektorům \vec{e}, \vec{f} . Podstatné je, že pak libovolný vektor \vec{u} dokážeme kódovat pomocí souřadnic, přičemž toto kódování je jednoznačné. Ještě užitečnější je, že takovéto souřadnice nám umožňují provádět snadno vektorové operace. Máme-li najít součet dvou vektorů, je možné začít rýsovat a odměřovat, nebo prostě sečteme jejich souřadnice a dostaneme tak hledaný vektor.

Teď si zavedeme něco obdobného, ale v úplně jiné situaci. Uvažujme dvě čísla $m, n \in \mathbb{N}$ větší než 1. To budou jakoby základní směrové vektory. Pro jiná čísla a se pak můžeme ptát, jak veliká je jakási „složka vzhledem k m “ a „složka vzhledem k n “. Není jasné, co to vůbec znamená, ale tato kapitola je o modulu, tak můžeme zkoušit $x = a \bmod m$ a $y = a \bmod n$.

Příklad to vysvětlí nejlépe. Zvolíme základní směry $m = 3, n = 5$, a vyzkoušíme si vytváření „souřadnic“. Například pro číslo $a = 10$ dostáváme „souřadnice“ $(10 \bmod 3, 10 \bmod 5) = (1, 0)$, pro $a = 7$ je to $(1, 2)$, pro

$a = 13$ je to $(1, 3)$. Vidíme, že pro různá čísla dostáváme různé souřadnice, což vypadá vysoce nadějně. Je nicméně jasné, že to je jen díky tomu, že jsme si hráli s malými čísly, například pro $a = 13 + 3 \cdot 5 = 28$ dostáváme zase $(1, 3)$. To by byl probém, pokud bychom si nevšimli, že se mu jednoduše vyhneme, když se omezíme na relativně malá čísla.

Snadno si rozmyslíme, že pokud se podíváme na číslo větší než $m \cdot n$, tak dostaneme stejné „souřadnice“ jako u nějakého menšího (nezáporného) čísla. V následujícím tvrzení ukážeme, že to funguje i naopak, pokud zůstaneme u čísel menších než mn , tak již ke každému číslu naleží unikátní „souřadnice“.

Lemma 7a.28.

Nechť $m, n \in \mathbb{N}$ jsou nesoudělná. Definujme zobrazení $T: \{0, 1, \dots, mn-1\} \mapsto \{0, 1, \dots, m-1\} \times \{0, 1, \dots, n-1\}$ předpisem $T(a) = (a \bmod m, a \bmod n)$. Toto zobrazení je bijekce.

Důkaz (poučný): Definice je evidentně korektní, pro libovolné $a \in \mathbb{Z}$ platí

$$(a \bmod m, a \bmod n) \in \{0, 1, \dots, m-1\} \times \{0, 1, \dots, n-1\}.$$

Máme $|\{0, 1, \dots, mn-1\}| = mn = |\{0, 1, \dots, m-1\}| \cdot |\{0, 1, \dots, n-1\}| = |\{0, 1, \dots, m-1\} \times \{0, 1, \dots, n-1\}|$, jinými slovy mohutnosti definičního oboru a cílové množiny se shodují (a jsou konečné). Podle Faktu proto stačí ukázat, že je T prosté. Mějme tedy $a, b \in \mathbb{Z}$ takové, že $T(a) = T(b)$. To znamená, že $a \bmod m = b \bmod m$ a $a \bmod n = b \bmod n$, takže dle Věty platí $m | (a - b)$ a $n | (a - b)$. Podle Faktu 6a.11 pak $\text{lcm}(m, n) | (b - a)$, a jelikož jsou m, n nesoudělné, pak už nutně $(mn) | (a - b)$. Proto existuje $k \in \mathbb{Z}$ takové, že $(a - b) = kmn$. Jenže $a, b \in \{0, 1, \dots, mn-1\}$, tedy $|a - b| < mn$, což je možné jen pro $k = 0$, tedy $a = b$. Prostota a tudíž i bijektivita T jsou dokázány. \square

Toto nám ukazuje dvě věci. První je, že opravdu pro čísla z množiny $\{0, 1, \dots, mn-1\}$ jsou ony „souřadnice“ jednoznačně dány. Druhá věc je, že se takto dají dostat všechny možné souřadnice. Jinými slovy, kdykoliv si zvolíme nějaké $x \in \{0, 1, \dots, m-1\}$ a $y \in \{0, 1, \dots, n-1\}$, tak už musí existovat $a \in \{0, 1, \dots, mn-1\}$ takové, že $a \bmod m = x$ a $a \bmod n = y$.

Jak se ale takové a najde? Bohužel důkaz věty není konstruktivní, tedy nedává návod, jak takové a identifikovat. Čtenář si může zkusit uhodnout třeba takové a , aby $a \bmod 101 = 13$ a $a \bmod 103 = 17$, aby docenil, jak obtížný problém to je. Naštěstí na to existuje metoda, viz Soustavy lineárních kongruencí.

Je zajímavé, že se tam vrátíme k našim „souřadnicím“, dokonce tam dokážeme, že se operace mezi čísly dají převést na operace s jejich „souřadnicemi“, přesně jako v případě vektorů. To je ale budoucnost, teď si ukážeme, co vyplýne z našeho posledního lemmatu.

! Věta 7a.29.

Nechť $m, n \in \mathbb{N}$ jsou nesoudělná. Pak $\varphi(m \cdot n) = \varphi(m) \cdot \varphi(n)$.

Důkaz (poučný): Zavedeme si množiny, které je třeba znát při výpočtu Eulerových čísel. Nechť B je množina přirozených čísel menších než m a nesoudělných s m , C je množina přirozených čísel menších než n a nesoudělných s n a nechť A je množina přirozených čísel menších než mn , která jsou s mn nesoudělná. Pak podle definice Eulerovy funkce je $\varphi(mn) = |A|$, $\varphi(m) = |B|$ a $\varphi(n) = |C|$.

Z definice množin máme $A \subseteq \{0, 1, \dots, mn-1\}$, $B \subseteq \{0, 1, \dots, m-1\}$ a $C \subseteq \{0, 1, \dots, n-1\}$. Uvažujme zobrazení S dané jako restrikce T z Lemmatu na množinu A . Tvrdíme, že je to bijekce $A \mapsto B \times C$.

Jestliže $a \in A$, pak $\gcd(a, mn) = 1$, proto je podle Lemmatu první složka $T(a)$ nesoudělná s m a druhá nesoudělná s n , tedy platí $T(a) \in B \times C$. S je opravdu zobrazení do $B \times C$. Protože je to restrikce prostého zobrazení T , musí být prosté.

Zbývá ukázat, že je to zobrazení na. Nechť $(x, y) \in B \times C$. Z předchozího Lemmatu plyne, že existuje číslo $a \in \{0, 1, \dots, mn-1\}$ takové, že $T(a) = (x, y)$. Ovšem z definice B a C máme $\gcd(x, m) = 1$ a $\gcd(y, n) = 1$, tudíž podle definice T platí i $\gcd(a \bmod m, m) = 1$ a $\gcd(a \bmod n, n) = 1$, proto podle Lemmatu máme $\gcd(a, mn) = 1$ a $a \in A$.

Ukázali jsme, že restrikce T je bijekce z A na $B \times C$, proto $|A| = |B| \cdot |C|$. \square

Důsledek 7a.30.

Nechť $n \in \mathbb{N}$ má prvočíselný rozklad $n = p_1^{k_1} \cdots p_m^{k_m}$, kde $p_1 < \cdots < p_m$ jsou prvočísla. Pak

$$\varphi(n) = p_1^{k_1} \left(1 - \frac{1}{p_1}\right) \cdots p_m^{k_m} \left(1 - \frac{1}{p_m}\right) = n \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

Důkaz (rutinní): Podle předchozí věty a Faktu máme

$$\varphi(n) = \prod_{i=1}^m \varphi(p_i^{k_i}) = \prod_{i=1}^m (p_i^{k_i} - p_i^{k_i-1}) = \prod_{i=1}^m p_i^{k_i} \left(1 - \frac{1}{p_i}\right) = \prod_{i=1}^m p_i^{k_i} \cdot \prod_{i=1}^m \left(1 - \frac{1}{p_i}\right) = n \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

□

Příklad 7a.p: Platí $\varphi(36) = \varphi(2^2 3^2) = 2^2 \left(1 - \frac{1}{2}\right) 3^2 \left(1 - \frac{1}{3}\right) = 36 \left(1 - \frac{1}{2}\right) \left(1 - \frac{1}{3}\right) = 12$.

△

Získaný vzorec lze pro jeden speciální případ zajímavě přeorganizovat.

Důsledek 7a.31.

Pro všechna $n, N \in \mathbb{N}$ platí $\varphi(n^N) = n^{N-1} \varphi(n)$.

Důkaz (rutinní, poučný): Rozložme si $n = p_1^{k_1} \cdots p_m^{k_m}$, kde $p_1 < \cdots < p_m$ jsou prvočísla. Pak také máme $n^N = p_1^{Nk_1} \cdots p_m^{Nk_m}$, tudíž

$$\varphi(n^N) = \prod_{i=1}^m p_i^{Nk_i} \cdot \prod_{i=1}^m \left(1 - \frac{1}{p_i}\right) = n^N \prod_{p|n} \left(1 - \frac{1}{p}\right) = n^{N-1} n \prod_{p|n} \left(1 - \frac{1}{p}\right) = n^{N-1} \varphi(n).$$

□

Naším cílem bylo vyhnout se pracnému procházení množiny a zjišťování, kdo je soudělný a kdo ne, abychom zjistili φ . Teď máme vzorec pro obecné n , což zní nadějně, ale ve skutečnosti to také není žádná výhra. Vyžaduje totiž najít prvočíselný rozklad čísla, což je výpočetně vysoce náročná úloha. Proto je vhodné hledat alternativy ke zjišťování $\varphi(n)$. Tato funkce byla a je v teorii čísel pilně studována a už se toho o ní spoustu ví, ukažme si pro zajímavost bez důkazu některé věci:

- Pro všechna $n \in \mathbb{N}$ platí $n = \sum_{d|n} \varphi(d)$.

- Schrammův vzorec: $\varphi(n) = \sum_{k=1}^n \gcd(k, n) e^{-2\pi i k/n}$.

- $\varphi(n)$ roste skoro tak rychle jako n . Přesně:

Pro každé $n \in \mathbb{N}$ platí $\sqrt{\frac{n}{2}} \leq \varphi(n) \leq n - 1$.

Pro každé $\varepsilon > 0$ existuje $N(\varepsilon) \in \mathbb{N}$ takové, že $n^{1-\varepsilon} < \varphi(n) < n$ pro všechna $n > N(\varepsilon)$.

Cvičení

Cvičení 7a.1 (rutinní): Spočítejte následující výrazy (zbytky po dělení), tedy ideální zástupce v kongruenci modulo dané číslo:

- | | | | |
|----------------------------|------------------------------|-----------------------------|-------------------------------|
| (i) $81 \text{ mod } 11$; | (iii) $3 \text{ mod } 11$; | (v) $48 \text{ mod } 8$; | (vii) $-8 \text{ mod } 4$; |
| (ii) $-1 \text{ mod } 7$; | (iv) $-14 \text{ mod } 13$; | (vi) $-37 \text{ mod } 5$; | (viii) $-15 \text{ mod } 6$. |

Cvičení 7a.2 (rutinní): Rozhodněte, které dvojice čísel z následujícího seznamu jsou kongruentní modulo 7: $-13, -4, 0, 1, 3, 7, 9, 17, 28$.

Cvičení 7a.3 (rutinní, zkouškové): Pro daná n spočítejte dané výrazy modulo n tak, aby výsledkem bylo číslo z rozmezí $0, 1, \dots, n-1$:

- $n = 6$, $(3 \cdot 13 + 11)^4 \cdot (37 + 14 \cdot 5)$;
- $n = 5$, $(13 - 39) \cdot 37 \cdot (-14)^2$;
- $n = 8$, $(24 \cdot 135 + 9)^7 \cdot 15 \cdot 18$.

Cvičení 7a.4 (rutinní, poučné): Nechť $a = \sum_{i=0}^m a_i 10^i$. Dokažte následující:

- a je dělitelné třemi právě tehdy, když je ciferný součet a (daný $\sum a_i$) dělitelný třemi.

- a je dělitelné devíti právě tehdy, když je ciferný součet a dělitelný devíti.

- a je dělitelné jedenácti právě tehdy, když je jedenácti dělitelné číslo, které získáme sečtením sudých cifer a a odečtením lichých cifer a .

Ná pověda: $n | a$ právě tehdy, když $a \equiv 0 \pmod{n}$.

Viz poznámka .

Cvičení 7a.5 (rutinní): Pro daná n a a najděte opačný prvek $(-a)$ a inverzní prvek a^{-1} v prostoru \mathbb{Z}_n , tedy prvky z množiny $\{0, 1, \dots, n-1\}$ takové, že $a + (-a) \equiv 0 \pmod{n}$ a $a^{-1} \cdot a \equiv 1 \pmod{n}$:

- (i) $n = 35, a = 12$; (iii) $n = 42, a = 25$;
- (ii) $n = 36, a = 15$; (iv) $n = 146, a = 75$.

Cvičení 7a.6 (rutinní, zkouškové): Použijte malou Fermatovu větu k výpočtu následujících výrazů modulo zadané n . Očekávají se výsledky z $\{0, 1, \dots, n-1\}$.

- (i) 3^{33} modulo $n = 11$; (ii) 4^{44} modulo $n = 13$; (iii) 5^{55} modulo $n = 23$.

Cvičení 7a.7 (rutinní, zkouškové): Spočítejte následující výrazy v daném \mathbb{Z}_n . Nejprve převeďte odčítání na sčítání s opačnými prvky.

- (i) $(7+8)^{146} - 21$ modulo $n = 13$; (ii) $(31 \cdot 4 - 1)^{192}$ modulo $n = 20$.

Cvičení 7a.8 (dobré): Dokažte, že jestliže je $n \in \mathbb{N}$ liché, pak $n^2 \equiv 1 \pmod{8}$.

Cvičení 7a.9 (rutinní, poučné): Která pseudonáhodná posloupnost je generována pomocí $x_{k+1} = (4x_k + 1) \pmod{7}$ při $x_0 = 3$?

Cvičení 7a.10 (rutinní, poučné, zkouškové): Nechť $n \in \mathbb{N}$, uvažujme $a, b, u, v \in \mathbb{Z}$ takové, že $a \equiv u \pmod{n}$ a $b \equiv v \pmod{n}$. Dokažte, že pak $a + b \equiv u + v \pmod{n}$ a $a - b \equiv u - v \pmod{n}$ (viz Věta).

Cvičení 7a.11 (poučné): Nechť $n \in \mathbb{Z}$, uvažujme $a_1, u_1, \dots, a_m, u_m \in \mathbb{Z}$ takové, že $a_i \equiv u_i \pmod{n}$ pro všechna $i = 1, \dots, m$. Dokažte, že pak $\prod_{i=1}^m a_i \equiv \prod_{i=1}^m u_i \pmod{n}$, viz Důsledek .

Cvičení 7a.12 (rutinní, poučné, zkouškové): Nechť $n \in \mathbb{N}$. Dokažte matematickou indukcí na k , že když $a, u \in \mathbb{Z}$ splňují $a \equiv u \pmod{n}$, pak pro libovolné $k \in \mathbb{N}$ platí $a^k \equiv u^k \pmod{n}$ (viz Fakt).

Cvičení 7a.13 (rutinní, zkouškové): Nechť $n \in \mathbb{N}$. Dokažte, že

- (i) pro každé $a \in \mathbb{Z}$ platí $a \equiv a \pmod{n}$;
- (ii) pro každé $a \in \mathbb{Z}$ platí: $a \equiv 0 \pmod{n}$ právě tehdy, když $n | a$.

(Viz Fakt .)

Cvičení 7a.14 (rutinní, zkouškové): Nechť $n \in \mathbb{N}$. Dokažte, že pro $a \in \mathbb{Z}_n$, $a \neq 0$ platí $(-a) = n - a$.
(Viz Fakt .)

Cvičení 7a.15 (rutinní, zkouškové): Nechť $n \in \mathbb{N}$. Dokažte, že pro každé $a, b \in \mathbb{Z}$ platí: $a \equiv b \pmod{n}$ právě tehdy, když $b \equiv a \pmod{n}$.

(Viz Věta .)

Cvičení 7a.16 (poučné): Nechť $m, n \in \mathbb{N}$. Dokažte, že pro každé $a, b \in \mathbb{Z}$ platí: Jestliže $a \equiv b \pmod{m}$ a $a \equiv b \pmod{n}$, pak $a \equiv b \pmod{\text{lcm}(m, n)}$.

Řešení:

7a.1: (i): $\lfloor \frac{81}{11} \rfloor = \lfloor 7.4\dots \rfloor = 7$, proto $81 \pmod{11} = 81 - 7 \cdot 11 = 4$; (ii): $\lfloor \frac{-1}{7} \rfloor = \lfloor -0.14\dots \rfloor = -1$, proto $-1 \pmod{7} = -1 - (-1) \cdot 7 = 6$, nebo prostě $-1 + 7 = 6$; (iii): 3 hotovo; (iv): $-14 + 13 + 13 = 12$; (v): 0 neboť $8 | 48$; (vi): $\lfloor \frac{-37}{5} \rfloor = \lfloor -7.4 \rfloor = -8$, proto $-37 \pmod{5} = -37 - (-8) \cdot 5 = 3$; (vii): 0 neboť $4 | (-8)$; (viii): třeba $-15 + 6 + 6 + 6 = 3$.

7a.2: $0 \equiv 7 \equiv 28 \pmod{7}$, $-4 \equiv 3 \equiv 17 \pmod{7}$, $-13 \equiv 1 \pmod{7}$, číslo 9 není kongruentní s nikým v seznamu. Mimochodem, právě jsme viděli rozklad dané množiny na zbytkové třídy.

7a.3: (i): $(3 \cdot 13 + 11)^4 \cdot (37 + 14 \cdot 5) \equiv (3 \cdot 1 + 5)^4 \cdot (1 + 2 \cdot 5) = 8^4 \cdot 11 \equiv 2^4 \cdot 5 = 16 \cdot 5 \equiv 4 \cdot 5 = 20 \equiv 2 \pmod{6}$.

(ii): $(13 - 39) \cdot 37 \cdot (-14)^2 \equiv (3 - 4) \cdot 2 \cdot 1^2 = (-1) \cdot 2 = -2 \equiv 3 \pmod{5}$.

(iii): $(24 \cdot 135 + 9)^7 \cdot 15 \cdot 18 \equiv (0 \cdot 135 + 1)^7 \cdot 7 \cdot 2 = 1^7 \cdot 14 \equiv 1 \cdot 6 = 6 \pmod{8}$.

Mimochodem, kdyby v tom prvním součinu nevyšla nula, museli bychom nahradit i 135. To odečítáním trvá dlohu, zde je asi lepší přístup přes zbytek po dělení. $q = \lfloor \frac{135}{8} \rfloor = \lfloor 16.87\dots \rfloor = 16$, $135 - 16 \cdot 8 = 135 - 128 = 7$. Proto $135 \equiv 7 \pmod{8}$.

7a.4: (i): $10 \equiv 1 \pmod{3}$, proto $a = \sum a_i 10^i \equiv \sum a_i \cdot 1^i = \sum a_i \pmod{3}$.

(iii): $10 \equiv (-1) \pmod{11}$, proto $a = \sum a_i 10^i \equiv \sum a_i \cdot (-1)^i = \sum a_{2i} - \sum a_{2i+1} \pmod{11}$.

7a.5:

- (i): $(-a) = n - a = 35 - 12 = 23$,
hledáme $x \in \mathbb{Z}$ aby $12x + 35k = 1$ pro nějaké $k \in \mathbb{Z}$,
toto děláme Euklidem.
Dostali jsme $3 \cdot 12 + (-1) \cdot 35 = 1$,
modulo 35 to dává $3 \cdot 12 \equiv 1$.
Takže $12^{-1} = 3$.

35		1	0
12	2	0	1
11	1	1	-2
1•	11	-1•	3•
0			

(ii): $(-a) = 36 - 15 = 21$,
 hledáme $x \in \mathbb{Z}$ aby $15x + 36k = 1$ pro nějaké $k \in \mathbb{Z}$,
 toto děláme Euklidem.
 Dostali jsme $\gcd(15, 36) > 1$,
 proto 15^{-1} v \mathbb{Z}_{36} neexistuje.

(iii): $(-a) = 42 - 25 = 17$,
 hledáme $x \in \mathbb{Z}$ aby $25x + 42k = 1$ pro nějaké $k \in \mathbb{Z}$,
 toto děláme Euklidem.
 Dostali jsme $(-5) \cdot 25 + 3 \cdot 42 = 1$,
 modulo 42 to dává $(-5) \cdot 25 \equiv 1$.
 Přičteme $-5 + 42 = 37$, takže $25^{-1} = 37$.

(iv): $(-a) = 146 - 75 = 71$,
 hledáme $x \in \mathbb{Z}$ aby $75x + 146k = 1$ pro nějaké $k \in \mathbb{Z}$,
 toto děláme Euklidem.
 Dostali jsme $(-19) \cdot 146 + 37 \cdot 75 = 1$,
 modulo 146 to dává $37 \cdot 75 \equiv 1$.
 Takže $75^{-1} = 37$.

36		1	0
15	2	0	1
6	2	1	-2
3•	2	-2•	5•
0			

42		1	0
25	1	0	1
17	1	1	-1
8	2	-1	2
1•	8	3•	-5•
0			

146		1	0
75	1	0	1
71	1	1	-1
4	17	-1	2
3	1	18	-35
1•	3	-19•	37•
0			

7a.6: (i): $= 3^{3 \cdot 10+3} = (3^{10})^3 \cdot 3^3 \equiv 1^3 \cdot 3^3 = 27 \equiv 5 \pmod{11}$. Výpočet je platný, protože $\gcd(3, 11) = 1$ a 11 je prvočíslo.

(ii): $= 4^{3 \cdot 12+8} = (4^{12})^3 \cdot 4^8 \equiv 1^3 \cdot (4^2)^4 = 16^4 \equiv 3^4 = 81 \equiv 3 \pmod{13}$. Výpočet je platný, protože $\gcd(4, 13) = 1$ a 13 je prvočíslo.

(iii): $= 5^{2 \cdot 22+11} = (5^{22})^2 \cdot 5^{11} \equiv 1^2 \cdot 5 \cdot (5^2)^5 = 5 \cdot 25^5 \equiv 5 \cdot 2^5 = 5 \cdot 32 \equiv 5 \cdot 9 = 45 \equiv 22 \pmod{23}$. Výpočet je platný, protože $\gcd(5, 23) = 1$ a 23 je prvočíslo.

7a.7: (i): $\equiv (7+8)^{146} + 5 = 15^{146} + 5 \equiv 2^{146} + 5 = 2^{12 \cdot 12+2} + 5 = (2^{12})^{12} \cdot 2^2 + 5 = 1^{12} \cdot 4 + 5 = 9 \pmod{13}$. Výpočet je platný, protože $\gcd(2, 13) = 1$ a 13 je prvočíslo.

(ii): $\equiv (31 \cdot 4 + 19)^{192} \equiv (11 \cdot 4 + 19)^{192} = (44+19)^{192} \equiv (4+19)^{192} = 23^{192} \equiv 3^{192} \pmod{20}$. Nelze použít malého fermata (20 není prvočíslo).

Redukce mocnin: $3^{192} = 3^{3 \cdot 64} = (3^3)^{64} = 27^{64} \equiv 7^6 4 = (7^2)^3 2 \equiv 9^{32} = (9^2)^{16} \equiv 1^{16} = 1 \pmod{20}$.

Euler: $\varphi(20) = \varphi(2^2 \cdot 5) = 20(1 - \frac{1}{2})(1 - \frac{1}{5}) = 8$, dále $\gcd(3, 20) = 1$, proto $3^{192} = 3^{8 \cdot 24} = (3^8)^{24} \equiv 1^{24} = 1 \pmod{20}$.

7a.8: $n = 2k + 1 \implies n^2 = 4k^2 + 4k + 1 = 4k(k+1) + 1$ a 2 dělí $k(k+1)$.

7a.9: 3, 6, 4, 3, 6, 4, 3, 6, 4, 3...

7a.10: $a = u + kn$, $b = v + ln$ pak $a+b = (u+v) + (k+l)n$.

7a.11: Indukcí. $m = 1$ dává $a_1 \equiv u_1 \pmod{n}$, což platí dle předpokladu.

(1) Nechť $m \in \mathbb{N}$. Předpoklad $\prod_{i=1}^m a_i \equiv \prod_{i=1}^m u_i \pmod{n}$.

Mějme a_1, \dots, a_{m+1} a u_1, \dots, u_{m+1} po dvou kongruentní viz předpoklad tvrzení. Předpoklad indukce dává $\prod_{i=1}^m a_i \equiv$

$\prod_{i=1}^m u_i \pmod{n}$, také $a_{m+1} \equiv u_{m+1} \pmod{n}$, proto dle Věty (iii) platí $\left(\prod_{i=1}^m a_i\right) \cdot a_{m+1} \equiv \left(\prod_{i=1}^m u_i\right) \cdot u_{m+1} \pmod{n}$

neboli $\prod_{i=1}^{m+1} a_i \equiv \prod_{i=1}^{m+1} u_i \pmod{n}$.

7a.12: Indukcí, $k = 1$ jasné.

(1) Nechť $k \in \mathbb{N}$. Indukční předpoklad: $a^k \equiv u^k \pmod{n}$. Také $a \equiv u$, proto dle Věty (iii) platí $a^k \cdot a \equiv u^k \cdot u \pmod{n}$ neboli $a^{k+1} \equiv u^{k+1} \pmod{n}$.

7a.13: (i): $a - a = 0$, proto $n | (a - a)$.

(ii): $a \equiv 0 \pmod{n} \iff n | (a - 0) \iff n | a$.

7a.14: Evidentně $0 \leq n - a \leq n - 1$, proto $n - a \in \mathbb{Z}_n$. Platí $a \oplus (-a) = a \oplus (n - a) = (a + (n - a)) \pmod{n} = n \pmod{n} = 0$.

7a.15: $a \equiv b \pmod{n} \implies n | (a - b) \implies n | (b - a) \implies b \equiv a \pmod{n}$.

7a.16: Předpoklad dává $m | (x-y)$ a $n | (x-y)$, takže číslo $x-y$ je společný násobek m, n , tudíž podle Lemma 6a.10 také $\text{lcm}(m, n) | (x-y)$.

7b. Řešení rovnic modulo

Jednou ze základních aritmetických úloh je řešení rovnic. Když řešíme rovnice v \mathbb{R} či \mathbb{Q} , tak na to máme tradiční nástroje, jmenovitě ekvivalentní (a neekvivalentní) úpravy rovnic, kterými si je přetváříme na příjemnější verze,

a vzorečky, které nám pomáhají s určitými typy rovnic. V této kapitole se zaměříme na rovnice ve světě modula, začneme zamýšlením, jak jsou vlastně takové rovnice formulovány a jaká řešení budeme očekávat.

Jako příklad uvažujme rovnici $x^2 = 6$. Podle toho, jakou verzi počítání modulo zrovna používáme, dostaváme různé úlohy.

- Pracujeme v \mathbb{Z} modulo n .

Pak bychom řešili úlohu „Najdi $x \in \mathbb{Z}$ takové, že $x^2 \equiv 6 \pmod{n}$ “.

Tato formulace je nejjednodušší (nepoužívá složitější konstrukce) a záhy uvidíme, že z praktického pohledu je také základem, i úlohy z dalších formulací převádíme při výpočtech na tuto.

- Pracujeme v \mathbb{Z}_n dle novější definice s třídami.

Pak bychom řešili úlohu „Najdi $[x]_n \in \mathbb{Z}_n$ takové, že $[x]_n^2 = [6]_n$ “. Zde samozřejmě $[x]_n^2 = [x]_n \odot [x]_n$.

- Pracujeme v \mathbb{Z}_n dle první definice.

Pak bychom řešili úlohu „Najdi $x \in \mathbb{Z}_n = \{0, 1, \dots, n-1\}$ takové, že $x^2 = 6$ “. Tentokrát je x^2 zkratka pro $x \odot x$.

Jaká řešení očekáváme?

• V první formulaci chceme najít množinu všech čísel $x \in \mathbb{Z}$, které splňují danou rovnici. Protože víme, že při počítání modulo je možné čísla ve výrazech nahrazovat kongruenčními bratříčky, je hned jasné, že jakmile najdeme jedno řešení, bude jich už nekonečně mnoho.

Například při volbě $n = 10$ můžeme zkusmo zjistit, že $x = 4$ splňuje rovnici $x^2 \equiv 6 \pmod{10}$, máme tedy jedno řešení, a proto budou řešeními i všechna čísla typu $x = 4 + 10k$ pro $k \in \mathbb{Z}$.

Zajímavější otázka je, jestli už tak dostaváme všechna možná řešení, nebo existují ještě i jiná. V našem případě je možné najít i číslo $y = 6$, které rozhodně není kongruentní s $x = 4$ a také řeší danou rovnici. Je tedy druhá „rodina“ řešení ve tvaru $6 + 10k$. Je ještě nějaká další? To je jedna z otázek, kterými se zde budeme zabývat.

• Formulace v jazyce $[]_n$: Pokud v rovnosti $[x]_n \odot [x]_n = [6]_n$ přepíšeme násobení podle definice, dostaváme $[x^2]_n = [6]_n$ neboli $x^2 \equiv 6 \pmod{n}$. To znamená, že každé řešení první formulace úlohy rovnou dá i řešení druhé formulace (po přikreslení ohrádky), naopak pokud nějaké $[x]_n$ splňuje druhou formulaci, pak x musí splňovat rovnici v jazyce modula. Zjistili jsme, že první dvě verze rovnice jsou v praxi totéž, stačí umět řešit rovnici dle první formulace (což je příjemnější) a po připsání ohrádek dostaváme řešení úlohy dle druhé formulace. Například rovnice $[x]_{10}^2 = [6]_{10}$ má určitě řešení $[4]_{10}$ a $[6]_{10}$.

Kolik bude řešení? Je snadné si rozmyslet, že všechna řešení typu $4 + 10k$ dají po přechodu k třídám kongruence totéž, třídu $[4]_{10}$, podobně to platí pro $[6 + 10k]_{10} = [6]_{10}$. Obecně, každá kongruenční rodinka řešení pro první formulaci úlohy daná jedním konkrétním řešením x dá ve světě \mathbb{Z}_n jedno řešení $[x]_n$. Počet řešení úlohy dle druhé formulace je tedy stejný jako počet různých kongruenčních skupin nalezených při řešení úlohy dle první formulace.

• Formulace v jazyce \mathbb{Z}_n : I zde existuje blízký vztah s řešeními první formulace. Pokud nějaké $x \in \mathbb{Z}_n$ řeší $x^2 = 6$, pak to znamená, že $x \in \{0, 1, \dots, n-1\}$ a $x^2 \equiv 6 \pmod{n}$. Máme-li naopak nějaké řešení $z \in \mathbb{Z}$, pak k němu najdeme zbytek modulo n neboli x , které je se z kongruentní (tudíž také řeší danou rovnici) a je z množiny $\{0, 1, \dots, n-1\}$, dostaneme tak řešení dle třetí formulace.

Jinak řešeno, pokud vezmeme všechna řešení $[x]_n$ dle druhé formulace a z každého vezmeme ideálního zástupce, dostaneme množinu všech řešení dle třetí formulace. Vidíme, že zase stačí umět řešit rovnice v modulární formulaci, rozdělit je do skupin podle kongruence a z každé pak vhodnou volbou dostaváme řešení pro svět \mathbb{Z}_n .

Například jsme již zjistili, že rovnice $x^2 = 6$ má v \mathbb{Z}_{10} řešení 4 a 6 (ale nevíme, zda nejsou nějaká další).

Jaký je závěr? Základem je rovnice vzhledem k počítání modulo n řešená v prostoru \mathbb{Z} , řešení v jiných variantách pak dostaváme tak, že si nalezená řešení vhodně uspořádáme a vybereme vhodné zástupce.

Teď bychom rádi odvodili na řešení postupy, jmenovitě chceme algoritmy, které pro danou rovnici umí rozhodnout, zda je řešitelná, a v případě, že ano, tak nalézt množinu všech řešení včetně kongruenční struktury. To je ovšem úloha vysoce náročná a obecně neřešitelná. Již tradičně se proto omezíme na speciální typy rovnic, jmenovitě na typ nejpříjemnější.

! 7b.1 Lineární kongruenze

Rozumíme tím rovnice typu $ax \equiv b \pmod{n}$ pro daná $a, b \in \mathbb{Z}$, popřípadě tutéž rovnici v řeči \mathbb{Z}_n .

Shrňme poznatky z úvodu:

1) Rovnice formulované v \mathbb{Z}_n převedeme do jazyka modula, například namísto úlohy „najděte (všechna) řešení $5x = 7$ v \mathbb{Z}_8 “ bychom řešili úlohu „najděte (všechna) řešení $x \in \mathbb{Z}$ rovnice $5x = 7 \pmod{8}$ “.

Nalezená řešení pak rozdělíme do skupin dle kongruence, z každé vybereme zástupce a máme řešení v \mathbb{Z}_n .

2) Základem je proto umět řešit rovnici $ax = b \pmod{n}$. Množina řešení je buď prázdná, nebo nekonečná, pro každé nalezené řešení x již bude řešením i celá množina $x + kn$ pro $k \in \mathbb{Z}$.

Dokažme si to: Jestliže x splňuje $ax = b \pmod{n}$, pak platí $a(x + kn) = ax + (ka)n \equiv b + 0 = b \pmod{n}$.

Než se dáme do práce, trochu si zjednodušíme značení.

Úmluva. V této sekci přestaneme pro kongruenci používat značení \equiv a coby zkušení modulární veteráni budeme prostě psát třeba $3x = 5 \pmod{13}$. Rovněž budeme psát obvyklé značky pro operace namísto \oplus a \odot . Je to (zejména u rovnic) tradiční a příjemnější. Pozorný čtenář by měl být schopen při čtení odvodit, kdy se hovoří o běžné rovnosti (a operacích) nebo o práci vzhledem k modulo n .

Jak budeme modulární rovnice řešit? Základem je následující pozorování, které vychází přímo z definice.

- $x \in \mathbb{Z}$ splňuje $ax \equiv b \pmod{n}$ právě tehdy, když existuje $y \in \mathbb{Z}$ takové, že $ax + yn = b$.

To je ovšem diofantická rovnice, u které nás navíc zajímá jen jedna souřadnice řešení x . Diofantické rovnice dokážeme zcela vyřešit, máme na to algoritmy a věty, když si z toho vytáhneme informace o první proměnné, okamžitě dostáváme toto.

Věta 7b.2.

Nechť $n \in \mathbb{N}$, nechť $a, b \in \mathbb{Z}$. Předpokládejme, že $\gcd(a, n) = Aa + Bn$.

(i) Jestliže b není násobkem $\gcd(a, n)$, tak řešení rovnice $ax \equiv b \pmod{n}$ neexistuje.

(ii) Jestliže $\gcd(a, n) | b$, tak rovnice $ax \equiv b \pmod{n}$ má řešení a množina všech jejích řešení je

$$\left\{ A \frac{b}{\gcd(a, n)} + k \frac{n}{\gcd(a, n)}; k \in \mathbb{Z} \right\}.$$

Tím je v zásadě vše hotovo, dostáváme z toho první možný postup na řešení rovnice $ax \equiv b \pmod{n}$.

Příklad 7b.a:

Uvažujme rovnici $14x = 38 \pmod{40}$.

Přeložíme si ji jako $14x + 40y = 38$ pro $x, y \in \mathbb{Z}$. Dle algoritmu pro diofantické rovnice potřebujeme Bezoutovu identitu, použijeme rozšířený Euklidův algoritmus.

40		1	0
14	2	0	1
12	1	1	-2
2	6	-1	3
0			

Máme $\gcd(a, n) = 2$ a pravá strana $b = 38$ je násobkem dvou, takže daná rovnice má řešení. Bezoutova identita říká $14 \cdot 3 + 40 \cdot (-1) = 2$, tedy $A = 3$ a $B = -1$. Podle vzorce dostáváme množinu řešení $x = 3 \cdot \frac{38}{2} + k \frac{40}{2} = 57 + 20k$. Protože pracujeme modulo 40, je možné vybrat pěknějšího zástupce:

Množina všech řešení je $x = 17 + 20k$, $k \in \mathbb{Z}$.

Tím je vyřešena úloha, jak je zadána. Nyní se podíváme na další varianty.

- Vyřešte rovnici $14x = 38$ v \mathbb{Z}_{40} .

Nejprve bychom ji převedli na tvar $14x = 38 \pmod{40}$ a vyřešili. Abychom nalezená řešení převedli do \mathbb{Z}_{40} , je třeba zjistit, jaké kongruenční třídy se tam vyskytují. Víme, že čísla $14 + 40k$ dávají jedno řešení z prostoru \mathbb{Z}_{40} , vycerpali jsme tím všechna řešení?

První následující řešení po 17 je $17 + 20 = 37$ a to nevznikne jako $17 + 40k$ pro $k \in \mathbb{Z}$. To znamená, že existuje i další kongruenční skupina $37 + 40k$. Je ještě nějaká další?

Snadno si rozmyslíme, že všechna řešení typu $17 + 20k$ jsou již v obou nalezených skupinách obsaženy.

Závěr: Rovnici $14x = 38$ má v \mathbb{Z}_{40} řešení $x = 17, 37$.

- Vyřešte rovnici $[14]_{40}[x]_{40} = [38]_{40}$ v \mathbb{Z}_{40} .

Analýza výše ukazuje, že řešením jsou $x = [17]_{40}$ a $x = [37]_{40}$.

△

Protože rovnice řešíme v řeči \mathbb{Z}_n často, bylo by dobré si vnést nějaký řád do hledání skupin. Nejprve to zkuseme intuitivně.

Podle výše máme řešení ve tvaru $x = x_p + kn'$, kde $n' = \frac{n}{\gcd(a, n)}$. Kolik skupin (tříd ekvivalence) dle modula n dostáváme?

Určitě máme třídu danou $x_p + kn$ neboli $[x_p]_n$. Dostáváme tak všechna řešení? Hned vidíme následující:

• Pokud $\gcd(a, n) = 1$, pak jsou všechna řešení dané rovnice kongruentní modulo n , v jazyce tříd ekvivalence (či v prostoru $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$) bude mít jediné řešení $[x_p]_n$.

Mimochedem, řešení vždy existuje, protože podmínka existence je $\gcd(a, n) | b$ a to je pro $\gcd(a, n) = 1$ vždy splněno.

Co se stane, když $n' < n$? Pak zbyla nějaká řešení, která nedokážeme najít jako $x_p + kn$. První vynechané řešení po x_p je $x_p + n'$. Protože n nemůže dělit n' , nejsou čísla x_p a $x_p + n'$ kongruentní, dostáváme tak proto jinou kongruenční třídu, jmenovitě řešení daná $\{(x_p + n') + kn\} = [x_p + n']_n$.

Zbyla nějaká řešení? Další řešení po $x_p + n'$ je $x_p + 2n'$. To by bylo obsaženo ve třídě $[x_p]_n$ jedině tehdy, pokud $2n' = n$, jinak dostáváme další třídu a tak dále.

Kolik tříd dostaneme? Tolik, kolikrát dokážeme přičíst n' k x_p , než se dostaneme k $x_p + n$. To znamená, že je celkem $\frac{n}{n'} = \gcd(a, n)$ různých tříd řešení.

Toto pozorování si potvrďme, zvolíme jazyk tříd ekvivalence, který je zde asi nejpříjemnější.

!

Věta 7b.3.

Nechť $n \in \mathbb{N}$, nechť $[a]_n, [b]_n \in \mathbb{Z}_n$. Předpokládejme, že $\gcd(a, n)$ dělí b . Nechť $[x_p]_n$ je nějaké řešení rovnice $[a]_n \cdot [x]_n = [b]_n$.

Rovnice $[a]_n \cdot [x]_n = [b]_n$ má v \mathbb{Z}_n celkem $\gcd(a, n)$ různých řešení, jmenovitě

$$\begin{aligned} & \left\{ [x_p]_n, \left[x_p + \frac{n}{\gcd(a, n)} \right]_n, \left[x_p + 2 \frac{n}{\gcd(a, n)} \right]_n, \dots, \left[x_p + (\gcd(a, n) - 1) \frac{n}{\gcd(a, n)} \right]_n \right\} \\ &= \left\{ \left[x_p + k \frac{n}{\gcd(a, n)} \right]_n ; k = 0, 1, \dots, \gcd(a, n) - 1 \right\}. \end{aligned}$$

Důkaz (poučný): 1) Nejprve ukážeme, že všechny vyspané třídy jsou opravdu řešením uvažované rovnice.

Jestliže $[x_p]_n$ řeší tuto rovnici, pak musí x_p řešit rovnici $ax = b \pmod{n}$, proto podle Věty existuje $k' \in \mathbb{Z}$ takové, že $x_p = A \frac{b}{\gcd(a, n)} + k' \frac{n}{\gcd(a, n)}$.

Uvažujme nějakou třídu $\left[x_p + k \frac{n}{\gcd(a, n)} \right]_n$. Pak $x_p + k \frac{n}{\gcd(a, n)} = A \frac{b}{\gcd(a, n)} + (k' + k) \frac{n}{\gcd(a, n)}$, proto podle Věty toto číslo také řeší rovnici $ax = b \pmod{n}$ a tedy dotyčná třída opravdu řeší $[a]_n \cdot [x]_n = [b]_n$.

2) Dále ukážeme, že každé řešení je obsaženo v některé z tříd. Uvažujme nějaké řešení $[x]_n$, to x pak musí řešit rovnici $ax = b \pmod{n}$. Podle Věty je určitě tvaru $x = A \frac{b}{\gcd(a, n)} + k \frac{n}{\gcd(a, n)}$ pro nějaké $k \in \mathbb{N}$. Proto $x = x_p + (k - k') \frac{n}{\gcd(a, n)}$.

Číslo $k - k'$ lze vyjádřit jako $k - k' = \gcd(a, n)y + r$, kde $r \in \{0, 1, \dots, \gcd(a, n) - 1\}$. Pak

$$x = x_p + (r + \gcd(a, n)y) \frac{n}{\gcd(a, n)} = x_p + r \frac{n}{\gcd(a, n)} + yn,$$

proto

$$[x]_n = \left[x_p + r \frac{n}{\gcd(a, n)} + yn \right]_n = \left[x_p + r \frac{n}{\gcd(a, n)} \right]_n.$$

Vidíme, že tato třída je mezi těmi vyspanými výše.

3) Již víme, že seznam tříd ve větě opravdu udává všechna řešení. Zbývá dokázat, že jde opravdu o různé třídy.

Uvažujme $[x]_n = \left[x_p + k \frac{n}{\gcd(a, n)} \right]_n$ a $y = \left[x_p + l \frac{n}{\gcd(a, n)} \right]_n$ pro nějaká $k, l \in \{0, 1, \dots, \gcd(a, n) - 1\}$. Pak $|k - l| < \gcd(a, n)$, proto $|x - y| = |k - l| \frac{n}{\gcd(a, n)} < \gcd(a, n) \frac{n}{\gcd(a, n)} = n$. Protože $|x - y| < n$, nemůže n dělit $x - y$ a tudíž tato čísla nejsou kongruentní modulo n . Třídy $[x]_n$ a $[y]_n$ jsou opravdu různé. □

Souhlasí to s naším příkladem výše? Ano, tam jsme našli řešení x_p a k němu třídy $[x_p]_{40}$ a $[x_p + 20]_{40}$, přesně jak je třeba pro případ $\gcd(14, 40) = 2$.

Shrňme to: Umíme řešit rovnice v řeči modula, umíme také nalézt množinu všech řešení pro formulaci v jazyce tříd kongruence. Pokud by úloha byla zadána v jazyce \mathbb{Z}_n dle první definice, pak stačí z každé kongruenční třídy vybrat vhodné řešení a dostaneme množinu všech řešení. Umíme tedy řešit lineární kongruence.

Mnozí uživatelé volí poněkud jiný postup řešení, takový, který využívá znalosti struktury množiny řešení. Výhodou je, že se pak postup podobá běžné práci s lineárními rovnicemi. I já mu dávám přednost, proto si tu odvodíme příslušný algoritmus. Začneme tradiční větou, viz lineární algebra či kapitola 6c.

Věta 7b.4.

Nechť $n \in \mathbb{N}$. Uvažujme rovnici $ax \equiv b \pmod{n}$ pro nějaká $a, b \in \mathbb{Z}$, nechť x_p je nějaké její řešení. Pak množina všech jejích řešení je

$$\{x_p + x_h; x_h \in \mathbb{Z}\text{ řeší přidruženou homogenní rovnici } ax \equiv 0 \pmod{n}\}.$$

Důkaz je natolik podobný důkazu Věty 6c.3, že jej s klidným svědomím necháme jako cvičení. Co to pro nás znamená prakticky? Je třeba najít jedno partiklární řešení, a to pomocí Bezoutovy identity, tam to jinak nepřejde.

Protože teď ale vyvíjíme algoritmus založený na pochopení procesu, nebudeme si pamatovat vzorec, ale postup, ukážeme to níže.

Dále je třeba znát, jak vypadá množina všech řešení homogenní lineární kongruence. To jsme již vlastně viděli, stačí si ve Větě dosadit $b = 0$. Z cvičných důvodů si výsledek dokážeme pro tento speciální případ, čtenář to zkusí samostatně.

Fakt 7b.5.

Nechť $n \in \mathbb{N}$, $a \in \mathbb{Z}$. Množina všech řešení rovnice $ax \equiv 0 \pmod{n}$ je $\left\{ k \frac{n}{\gcd(a, n)} ; k \in \mathbb{Z} \right\}$.

Důkaz (rutinní): 1) Ukážeme, že každé $x = k \frac{n}{\gcd(a, n)}$ je řešením dané rovnice. Dosadíme do levé strany:

$$ax = k \frac{an}{\gcd(a, n)} = nk \frac{a}{\gcd(a, n)}, \text{ což je násobek } n \text{ neboť } \frac{a}{\gcd(a, n)} \in \mathbb{Z} \text{ a tudíž rovno } 0 \pmod{n}.$$

2) Naopak nechť x je nějaké řešení. Pak $ax + yn = 0$ pro nějaké $y \in \mathbb{Z}$, tedy $ax = -yn$. Vydělíme tím \gcd , dostaneme $\frac{a}{\gcd(a, n)}x = -y \frac{n}{\gcd(a, n)}$. Celé číslo $\frac{n}{\gcd(a, n)}$ tedy dělí $\frac{a}{\gcd(a, n)}x$, jenž podle Faktu 6a.9 jsou $\frac{n}{\gcd(a, n)}$ a $\frac{a}{\gcd(a, n)}$ nesoudělná čísla, tudíž musí podle Lemma 6a.23 číslo $\frac{n}{\gcd(a, n)}$ dělit x . \square

Protože se snažíme redukovat množství vzorců, které si pamatujeme, podíváme se na to jinak. Práci si ulehčíme, když rovnici hned na začátku co nejvíce zkrátíme. Zde je ovšem třeba pracovat opatrně, protože aby se zachovala množina řešení, je třeba krátit i modulo.

Lemma 7b.6.

Nechť $n \in \mathbb{N}$, uvažujme $a, b \in \mathbb{Z}$. Předpokládejme, že $d \in \mathbb{N}$ dělí čísla a, b, n .

Pak číslo $x \in \mathbb{Z}$ řeší rovnici $ax \equiv b \pmod{n}$ právě tehdy, když řeší rovnici $\frac{a}{d}x \equiv \frac{b}{d} \pmod{\frac{n}{d}}$.

Důkaz (rutinní): $ax \equiv b \pmod{n}$ právě tehdy, když existuje nějaké $k \in \mathbb{Z}$, aby $ax = b + kn$, což je právě tehdy, když existuje nějaké $k \in \mathbb{Z}$, aby $\frac{a}{d}x = \frac{b}{d} + k \frac{n}{d}$, což je právě tehdy, když $\frac{a}{d}x \equiv \frac{b}{d} \pmod{\frac{n}{d}}$. \square

V praxi ale většinou zkrátíme až odvozenou diofantickou rovnici $ax + yn = b$. Zajímavou shodou okolností je největší číslo, kterým lze zkrátit na levé straně, právě $\gcd(a, n)$. Vzniká tak docela zajímavý sled kroků, které vypadají docela přirozeně:

Pamatujeme si, že chceme rovnici zkrátit, logicky tedy začneme nalezením čísla $\gcd(a, n)$, uhodnutím nebo rozšířeným Euklidem. Pak rovnici zkusíme zkrátit. Pokud se to na pravé straně nepovede, tak nemá řešení. Pokud se to povede, řešíme jednodušší zkrácenou rovnici.

Pak si musíme zapamatovat, že k nalezení partikulárního řešení vyjdeme z Bezoutovy identity, tu pak upravíme tak, abychom v ní rozpoznali danou rovnici (původní či zkrácenou, vyjde to nastejno).

Zbývá vyřešit přidruženou homogenní rovnici, ale u té je to ve zkráceném stavu jednoduché, máme rovnici ve tvaru $a'x + yn' = 0$ a pro a', n' nesoudělné má množina všech řešení tvar $x = kn'$.

Tím jsme dospěli k algoritmu. Ten má ještě dvě možné varianty, liší se tím, zda zkrátit danou rovnici hned na začátku, nebo až ve chvíli, kdy řešíme homogenní verzi. Souvisí to s tím, jak jsme našli $\gcd(a, n)$. Pokud jsme to hned viděli, pak má smysl rovnou rovnici vykrátit. Pokud jsme na to museli použít rozšířený Euklidův algoritmus, pak je asi jednodušší rovnou z výsledné Bezoutovy identity najít x_p pomocí původní verze rovnice a ke krácení přistoupit u homogenní rovnice. Tato druhá verze je obecnější, proto ji použijeme v algoritmu a student se jí může bez problémů držet.

Obecně se dá říct, že pokud student postupu dobře rozumí, může si dovolit se od něj občas jemně odchýlit, například tím, že použije první verzi nastíněnou výše čijinou zkratku.

S Algoritmus 7b.7. pro řešení rovnice $ax \equiv b \pmod{n}$ v \mathbb{Z} , popřípadě rovnice $[a]_n[x]_n = [b]_n$ v \mathbb{Z}_n , popřípadě rovnice $ax = b$ v \mathbb{Z}_n pro $a, b \in \mathbb{Z}_n$.

0. Přepište si rovnici do tvaru $ax + ny = b$. Rozšířeným Euklidovým algoritmem najděte $\gcd(a, n) = Aa + Bn$ (či to uhodněte).

1. Jestliže $\gcd(a, n)$ nedělí b , pak řešení neexistuje.
2. Jestliže $\gcd(a, n)$ dělí b , rovnice má řešení.

a) Vynásobte identitu $\gcd(a, n) = Aa + Bn$ číslem $\frac{b}{\gcd(a, n)}$, čímž se změní na tvar $a \frac{Ab}{\gcd(a, n)} + n \frac{Bb}{\gcd(a, n)} = b$, přesně jako rovnice v kroku 0. Vidíme partikulární řešení $x_p = \frac{Ab}{\gcd(a, n)}$.

b) Přidruženou homogenní rovnici $ax + ny = 0$ zkraťte číslem $\gcd(a, n)$ na tvar $a'x + n'y = 0$, ta má obecné řešení $x_h = kn'$, $k \in \mathbb{Z}$.

c) Obecné řešení dané rovnice je pak $x_p + x_h$.

V závislosti na formě zadání tak dostáváte následující:

- Množina všech celočíselných řešení kongruence $ax \equiv b \pmod{n}$ je $\{x_p + kn'; k \in \mathbb{Z}\}$ neboli $x = x_p + kn'$, $k \in \mathbb{Z}$.
- Množina všech řešení v \mathbb{Z}_n rovnice $[a]_n[x]_n = [b]_n$ je $\{[x_p + kn']_n; k = 0, 1, 2, \dots, \gcd(a, n) - 1\}$.
- Množinu řešení v \mathbb{Z}_n rovnice $ax = b$ dostaneme tak, že pro každé $k = 0, 1, \dots, \gcd(a, n) - 1$ vybereme vhodného zástupce třídy $[x_p + kn']_n$.

Formálně, nechť k_0 je nejmenší celé číslo takové, že $x_p + kn' \geq 0$. Pak množina všech řešení v \mathbb{Z}_n rovnice $ax = b$ je $\{x_p + kn'; k = k_0, k_0 + 1, \dots, k_0 + \gcd(a, n) - 1\}$.

△

! Příklad 7b.b: Vyřešíme rovnici $66x = 18$ ve světě modulo $n = 150$ v různých podobách.

Ať už je zadání jakékoliv, vždy si nejprve rovnici přepíšeme jako $66x + 150y = 18$. Podle algoritmu máme nejprve najít $\gcd(66, 150)$, takže na koeficienty poštěveme rozšířený Euklidův algoritmus.

150		1	0
66	2	0	1
18	3	1	-2
12	1	-3	7
6•	2	4•	-9•
0			

Dostáváme $\gcd(150, 66) = 6 = 4 \cdot 150 + (-9) \cdot 66$. Protože $\gcd(66, 150) = 6$ dělí pravou stranu, rovnice má řešení. Abychom našli partikulární řešení, je třeba upravit rovnost $66 \cdot (-9) + 150 \cdot 4 = 6$ do tvaru odpovídajícího zadанé rovnici. Koeficienty 66 a 150 levé strany již souhlasí, zbyvá upravit pravou stranu. Číslo 18 dostaneme, když celou rovnici vynásobíme trojkou, na levé straně tu trojku rozhodně nedáme ke koeficientům. Dostáváme $66 \cdot (-27) + 150 \cdot 12 = 18$, což mimochodem modulo 150 dává $66 \cdot (-27) \equiv 18$, prostě vidíme partikulární řešení $x_p = -27$.

Ted' vyřešíme homogenní rovnici $66x + 150y = 0$. Zkrátíme nalezeným $\gcd(66, 150) = 6$, dostáváme rovnici $11x + 25y = 0$ a z ní $x_h = 25k$, $k \in \mathbb{Z}$.

Sečteme: Obecné řešení je $x = x_p + x_h = 25k - 27$.

Odpovědi dle formy zadání:

- Pokud máme řešit rovnici $66x = 18 \pmod{150}$, odpověď zní:

Množina řešení je $x = 25k - 27$, $k \in \mathbb{Z}$, popřípadě $\{25k - 27, k \in \mathbb{Z}\}$.

Je také možné použít lepšího zástupce, třeba $x = 25k - 2$ nebo $x = 23 + 25k$.

- Pokud máme řešit rovnici $[66]_{150}[x]_{150} = [18]_{150}$, pak víme, že existuje $\gcd(66, 150) = 6$ různých řešení, jmenovitě $[-27]_{150}, [-27 + 25]_{150}, [-27 + 50]_{150}, [-27 + 75]_{150}, [-27 + 100]_{150}, [-27 + 125]_{150}$. Množina všech řešení je $\{[-27]_{150}, [-2]_{150}, [23]_{150}, [48]_{150}, [73]_{150}, [98]_{150}\}$.

- Pokud máme řešit rovnici $66x = 18$ v \mathbb{Z}_{150} , pak množinu všech řešení, kterých je $\gcd(66, 150) = 6$, dostaneme výběrem vhodných kongruentních zástupců z šesti tříd výše. V praxi se to ale často dělá hned z prvního výsledku takto: nejprve najdeme nejmenší nezáporné číslo typu $x = -27 + 25k$, jmenovitě 23, z něj pak vyrobíme 6 řešení obvyklým způsobem. Množina všech řešení je $\{23, 48, 73, 98, 123, 148\}$.

Zkouška: Namátkou třeba $66 \cdot 48 = 3168 \pmod{150} = 18$, použili jsme $3168 - 21 \cdot 150 = 18$. Ověření ostatních řešení necháme pilnému čtenáři.

△

Poznámka: Řešení, které jsme právě viděli, je vzorové, ale nabízí se pár odboček.

1) Jedna možnost je zkrátit přímo danou rovnici číslem 6, což je rozumné dělat jen v situaci, kdy $\gcd(66, 150)$ umíme odhadnout. Dostaneme rovnici $11x + 25y = 3$, kterou pak dále řešíme obvyklým způsobem. Je třeba najít Bezoutovo vyjádření pro novou zkrácenou rovnici, pomocí rozšířeného Euklidova algoritmu odhalíme $\gcd(11, 25) = 1 = (-9) \cdot 11 + 4 \cdot 25$, tuto rovnost pak vynásobením trojkou upravíme na tvar $11 \cdot (-27) + 25 \cdot 12 = 3$, což odpovídá řešené (zkrácené rovnici) a máme $x_p = -27$.

Dál už pokračujeme běžným způsobem (x_h , následně $x_p + x_h$).

Proč takovéto brzké krácení rovnice doporučujeme jen pro případ, že $\gcd(66, 150)$ uhodneme? Pokud jej uhodnout neumíme a použijeme Euklidův algoritmus, tak po jeho běhu dostaneme $\gcd(66, 150) = 6 = 66 \cdot (-9) + 150 \cdot 4$. Z této rovnosti se většinou snáze dostaneme úpravou k rovnici původní než k té zkrácené, je tedy lepší odložit krácení až na homogenní případ.

2) Občas se naskytne jiný trik na zkrácení výpočtu. Pokud čtenář chápe, že smyslem prvního kroku je najít něco, co vypadá jako daná rovnice, pak je pro něj zajímavý třetí (nenulový) řádek zdola v tabulce Euklidova algoritmu

výše. Ten totiž říká, že $18 = 1 \cdot 150 + (-2) \cdot 66$ neboli $66 \cdot (-2) + 150 \cdot 1 = 18$. Rovnou vidíme řešení $x_p = 18$, protože dostáváme verzi dané rovnice, souhlasí koeficienty i pravá strana.

Je jasné, že člověk musí mít štěstí, aby se mu v tabulce objevila pravá strana rovnice, také musí dobře rozumět algoritmu i smyslu Euklidovy tabulky, ale když už se to zadaří, tak to potěší.

△

! 7b.8 Soustavy lineárních kongruencí

Zde budeme uvažovat následující typ soustav. Jsou dány moduly $n_1, \dots, n_m \in \mathbb{N}$ a pravé strany $b_1, \dots, b_m \in \mathbb{Z}$. Hledáme celá čísla x taková, že

$$\begin{aligned} x &\equiv b_1 \pmod{n_1}, \\ x &\equiv b_2 \pmod{n_2}, \\ &\vdots \\ x &\equiv b_m \pmod{n_m}. \end{aligned}$$

Jde o speciální případ soustav lineárních kongruencí $a_i x \equiv b_i \pmod{n_i}$, ale ty by byly nad naše síly. Dokonce i naše speciální volba $a_i = 1$ pořád vede na zajímavé věci. Začneme klasikou.

Věta 7b.9.

Uvažujme moduly $n_1, n_2, \dots, n_m \in \mathbb{N}$ a čísla $b_1, b_2, \dots, b_m \in \mathbb{Z}$.

Nechť x_p je nějaké řešení soustavy kongruencí

$$\begin{aligned} x &\equiv b_1 \pmod{n_1}, \\ x &\equiv b_2 \pmod{n_2}, \\ &\vdots \\ x &\equiv b_m \pmod{n_m}. \end{aligned}$$

Číslo x je také řešením této soustavy právě tehdy, pokud existuje číslo x_h takové, že $x = x_p + x_h$ a x_h je řešením přidružené homogenní soustavy kongruencí

$$\begin{aligned} x &\equiv 0 \pmod{n_1}, \\ x &\equiv 0 \pmod{n_2}, \\ &\vdots \\ x &\equiv 0 \pmod{n_m}. \end{aligned}$$

Důkaz se od dvou obdobných vět dokázaných již dříve liší jen v detailech a necháme jej čtenáři.

Jako obvykle tedy stačí umět najít jedno partikulární řešení a pak pořádně prozkoumat homogenní rovnice. Těmi začneme, jsou snadné.

Uvažujme tedy soustavu rovnic

$$\begin{aligned} x &\equiv 0 \pmod{n_1}, \\ x &\equiv 0 \pmod{n_2}, \\ &\vdots \\ x &\equiv 0 \pmod{n_m}. \end{aligned}$$

Každá z těchto rovnic vyžaduje, aby x bylo násobkem příslušného modulu, takže vlastně hledáme taková x , která jsou společnými násobky všech modulů n_i .

Fakt 7b.10.

Uvažujme moduly $n_1, n_2, \dots, n_m \in \mathbb{N}$. Číslo $x \in \mathbb{Z}$ splňuje kongruence $x \equiv 0 \pmod{n_i}$ pro všechna $i = 1, \dots, m$ právě tehdy, když je x násobkem čísla $\text{lcm}(n_1, n_2, \dots, n_m)$.

Důkaz: Každé číslo ve tvaru $k \text{lcm}(n_1, \dots, n_m)$ pro $k \in \mathbb{Z}$ je dělitelné všemi n_i , tudíž vždy dává modulo n_i nulu a řeší uvažované kongruence.

Naopak každé řešení x oněch kongruencí musí být násobkem jednotlivých n_i , je to tedy společný násobek n_1, \dots, n_m . Protože je $\text{lcm}(n_1, \dots, n_m)$ nejmenším společným násobkem, a to i ve smyslu dělitelnosti, musí platit $\text{lcm}(n_1, \dots, n_m) | x$.

□

Pokud máme moduly n_1, \dots, n_m takové, že pro $i \neq j$ jsou n_i, n_j nesoudělná čísla, pak je množina všech řešení příslušných kongruencí $x \equiv 0 \pmod{n_i}$ rovna $\{kn_1 n_2 \cdots n_m; k \in \mathbb{Z}\}$. Právě tento případ budeme dále uvažovat.

Zbývá vymyslet, jak nějak najít jedno partikulární řešení, což bude znatelně komplikovanější než v případě diofantických rovnic a lineárních kongruencí probraných výše. Existenci takového řešení nám za určitých podmínek potvrdí věta, dokonce dodá návod, jak takové řešení najít. Abychom ale důkazu lépe porozuměli, bude dobré si nejprve rozmyslet, odkud se výsledný vzorec bere.

Začneme první rovnicí. Pokud ji x řeší, tak jistě musí mít tvar $x = b_1 + kn_1$. Podobně snadno najdeme obecná řešení i pro další rovnice, problém je v tom, že potřebujeme jedno řešení společné.

Vezměme tedy všechna možná řešení první rovnice $x = b_1 + kn_1$, měli bychom zařídit, aby mezi nimi bylo i partikulární řešení druhé rovnice, jinými slovy, měli bychom zařídit, aby se při pohledu modulo n_2 objevilo b_2 . Klíčová myšlenka je následující: Protože budeme mít více rovnic, tak nechceme, aby se nám v x vlivy míchaly. Přesněji řečeno, máme x jako součet dvou částí a zatím to funguje tak, že se při pohledu modulo n_1 druhá vynuluje a první dá žádané b_1 . Rádi bychom, aby to obdobně (jen obráceně) fungovalo modulo n_2 .

Nápad: Použijeme $x = b_1 + b_2 kn_1$. Zatím jsme měli k jako libovolný parametr, čehož teď využijeme. Zvolíme takové k , aby se z výrazu $b_2 kn_1$ při pohledu modulo n_2 stalo b_2 . To vyžaduje, aby $kn_1 \equiv 1 \pmod{n_2}$, tedy stačí za k zvolit n_1^{-1} vzhledem k modulu n_2 . Pokud to uděláme, tak $(b_2 n_1^{-1} n_1) \pmod{n_2} = b_2 \cdot 1 = b_2$, zatímco $(b_2 n_1^{-1} n_1) \pmod{n_1} = b_2 n_1^{-1} \cdot 0 = 0$. Vidíme, že druhý člen teď funguje tak, jak chceme, vůči jednomu modulu se vynuluje a vůči druhému dá potřebnou pravou stranu.

První člen to zatím ale neumí, modulo n_1 dává žádané b_1 , ale modulo n_2 se nevynuluje. Máme už ale nápad, jak to zařídit. Dostáváme lepší verzi $x = b_1 n_2^{-1} n_2 + b_2 n_1^{-1} n_1$, kde se inverze n_2^{-1} bere vzhledem k modulu n_1 .

Funguje to pěkně, rozmyslíme si případ tří rovnic. Pak x sestavíme ze tří členů, u každého potřebujeme, aby se vynuloval vzhledem ke dvěma modulům, což se snadno udělá zahrnutím těchto modulů. Napíšeme si kandidáty a přehledně si napíšeme, co od nich očekáváme vzhledem k různým modulům.

	$b_1 x_1 n_2 n_3$	$b_2 x_2 n_1 n_3$	$b_3 x_3 n_1 n_2$
n_1	b_1	0	0
n_2	0	b_2	0
n_3	0	0	b_3

Ty nuly již opravdu fungují, bez ohledu na to, co zvolíme za x_i , takže máme svobodu si zvolit x_i tak, aby dobře dopadly i zbývající políčka v tabulce. Jestliže například má být $(b_1 x_1 n_2 n_3) \pmod{n_1} = b_1$, tak potřebujeme $(x_1 n_2 n_3) \pmod{n_1} = 1$. To znamená, že by x_1 měl být inverzní prvek k $n_2 n_3$ vzhledem k modulu n_1 , podobně by x_2 měl být inverzní prvek k $n_1 n_3$ vzhledem k modulu n_2 a x_3 by měl být inverzní prvek k $n_1 n_2$ vzhledem k modulu n_3 .

Aby šly tyto prvky najít, musí být vždy n_i nesoudělné se součinem ostatních modulů. Máme nápad, který zdá se funguje. Abychom vše rádně dokázali, uděláme si nejprve lematko.

Lemma 7b.11.

Nechť $n_1, n_2, \dots, n_m \in \mathbb{N}$ jsou po dvou nesoudělná. Označme $n = n_1 \cdot n_2 \cdots n_m$.
Jestliže $a, b \in \mathbb{Z}$ splňují $a \equiv b \pmod{n_i}$ pro všechna $i = 1, \dots, m$, pak $a \equiv b \pmod{n}$.

Důkaz (poučný): Předpoklad říká, že $n_i | (a - b)$ pro všechna i . Podle Faktu výše je pak nutně $a - b$ násobkem čísla $\text{lcm}(n_1, n_2, \dots, n_m)$. Protože jsou n_i navzájem nesoudělná, podle cvičení 6b.4 je $\text{lcm}(n_1, n_2, \dots, n_m) = n$, tedy n dělí $a - b$. □

Jako cvičení nabízíme alternativní důkaz indukcí, který může čtenáři přijít stravitelnější (nebo také ne). Jsme připraveni.

!

Věta 7b.12. (Čínská věta o zbytcích)

Nechť $n_1, n_2, \dots, n_m \in \mathbb{N}$, $b_1, b_2, \dots, b_m \in \mathbb{Z}$. Uvažujme soustavu rovnic

$$x \equiv b_1 \pmod{n_1},$$

$$x \equiv b_2 \pmod{n_2},$$

⋮

$$x \equiv b_m \pmod{n_m}.$$

Jestliže jsou všechna čísla n_i po dvou nesoudělná, pak má tato soustava řešení.

Toto řešení je jediné modulo $n = n_1 n_2 \cdots n_m$, množina všech řešení je $\{x + kn; k \in \mathbb{Z}\}$.

Důkaz (poučný): 1) Nejprve ukážeme, že řešení existuje. Pro $i = 1, \dots, m$ definujeme $N_i = \frac{n}{n_i}$, tedy je to součin všech n_j s výjimkou n_i . Tvrdíme, že $\text{gcd}(N_i, n_i) = 1$.

Kdyby tomu tak nebylo, pak by existovalo číslo, tudíž podle Faktu 6b.1 i prvočíslo p , které by dělilo n_i a $N_i = \prod_{j \neq i} n_j$. Podle Lemmatu 6b.2 by tedy p dělilo některé n_j a máme spor s $\gcd(n_i, n_j) = 1$.

Čísla N_i, n_i jsou tedy nesoudělná, proto existuje inverzní prvek k N_i vzhledem k násobení modulo n_i , tedy x_i takové, že $x_i N_i \equiv 1 \pmod{n_i}$. Nechť $x = \sum_{i=1}^m b_i N_i x_i$. Tvrdíme, že je to řešení dané soustavy.

Zvolme i . Pro $j \neq i$ pak $n_i | N_j$, proto $N_j \equiv 0 \pmod{n_i}$, tedy i $(b_j N_j x_j) \equiv 0 \pmod{n_i}$. Následně modulo n_i dostaneme $x \equiv b_i N_i x_i \equiv b_i \cdot 1 = b_i \pmod{n_i}$.

2) Jednoznačnost: Nechť je y nějaké řešení. Pak $(x - y) \equiv (b_i - b_i) = 0 \pmod{n_i}$, tedy $x \equiv y \pmod{n_i}$ pro všechna n_i . Podle Lemmatu pak $x \equiv y \pmod{n}$.

3) Tvar množiny všech řešení vyplývá okamžitě z našich předchozích úvah. □

Při důkazu jsme opakovaně používali vzájemnou nesoudělnost jednotlivých dvojic. Poznamenejme, že by nestačilo jen chtít, aby největší společný dělitel všech n_i byl 1, protože to je jiná, mnohem slabší podmínka. Například největší společný dělitel čísel 3, 4, 8 je 1, ale 4 a 8 rozhodně nejsou nesoudělné, tudíž by naše triky nefungovaly. Požadavek „po dvou nesoudělná“ tuto trojici správně vyřadí.

Důkaz věty a předchozí pozorování o homogenní rovnici dávají algoritmus.

S Algoritmus 7b.13. pro řešení soustavy kongruencí $x \equiv b_1 \pmod{n_1}, x \equiv b_2 \pmod{n_2}, \dots, x \equiv b_m \pmod{n_m}$ pro případ, že jsou všechna čísla n_i po dvou nesoudělná.

1. Označte $n = n_1 n_2 \cdots n_m$ a $N_i = \frac{n}{n_i}$ pro všechna i .
2. Pro každé i najděte inverzní prvek k N_i vzhledem k násobení modulo n_i , viz algoritmus .
3. Nechť $x = \sum_{i=1}^m b_i N_i x_i$. Množina všech řešení soustavy je $\{x + kn; k \in \mathbb{Z}\}$.

△

! Příklad 7b.c: Větě se říká čínská, protože soustavy kongruencí jdou zpět ke starým Číňanům někam do 3. století. Asi nejznámější je následující úloha z klasické knihy *Matematický manuál* mistra Sun-Tzu (to byl matematik, neplést se stejnojmenným autorem klasické knihy o vojenské strategii známe jako *The Art of War*).

Mějme určitý neznámý počet věcí. Když je uspořádáme po třech, zbydou dvě. Když je uspořádáme po pěti, zbydou tři. Když je uspořádáme po sedmi, zbydou dvě. Kolik je věcí?

Přeloženo do moderního jazyka, hledáme řešení soustavy rovnic $x \equiv 2 \pmod{3}$, $x \equiv 3 \pmod{5}$ a $x \equiv 2 \pmod{7}$. Použijeme příslušný algoritmus.

Máme $n_1 = 3, n_2 = 5, n_3 = 7$, proto $n = 3 \cdot 5 \cdot 7 = 105$. Uděláme si doplňkové součiny $N_1 = \frac{n}{n_1} = n_2 \cdot n_3 = 35, N_2 = \frac{n}{n_2} = n_1 \cdot n_3 = 21, N_3 = \frac{n}{n_3} = n_1 \cdot n_2 = 15$.

Ted' pro každé i potřebujeme inverzní prvek k N_i vzhledem k násobení modulo n_i . Budeme tedy řešit diofantické rovnice $35x + 3k = 1, 21x + 5k = 1$ a $15x + 7k = 1$.

35	1	0	
3	11	0	1
2	1	1	-11
1•	2	-1•	12•
0			

21	1	0	
5	4	0	1
1•	5	1•	-4•
0			

15	1	0	
7	2	0	1
1•	7	1•	-2•
0			

Dostáváme následující:

$$\gcd(35, 3) = 1 = (-1) \cdot 35 + 12 \cdot 3, \text{ tedy } 2 \cdot 35 \equiv 1 \pmod{3} \text{ a proto } x_1 = 35^{-1} = -1;$$

$$\gcd(21, 5) = 1 = 1 \cdot 21 + (-4) \cdot 5, \text{ tedy } 1 \cdot 21 \equiv 1 \pmod{5} \text{ a proto } x_2 = 21^{-1} = 1;$$

$$\gcd(15, 7) = 1 = 1 \cdot 15 + (-2) \cdot 7, \text{ tedy } 1 \cdot 15 \equiv 1 \pmod{7} \text{ a proto } x_3 = 15^{-1} = 1.$$

Ty poslední dva se daly odhadnout i bez výpočtu.

Dosadíme do vzorce a dostáváme $x = 2 \cdot (-1) \cdot 35 + 3 \cdot 1 \cdot 21 + 2 \cdot 1 \cdot 15 = 23$.

Řešení je $23 + 105k$ pro $k \in \mathbb{N}_0$ (vzhledem k tomu, že jde o počty věcí, jsme nezahrnuli záporná k).

△

Poznámka: Pokud nesoudělnost těch n_i, n_j nemáme, nastává vážný problém, a to nejen s nalezením řešení (Čínskou větu nelze použít), ale dokonce s jeho existencí: Například soustava $x \equiv 2 \pmod{6}$ a $x \equiv 3 \pmod{9}$ vůbec nemá řešení. Existuje přesná podmínka: Máme-li dánou soustavu a n_i jsou obecná (ne nutně nesoudělná) přirozená čísla, pak tato soustava má řešení právě tehdy, když pro každé $i \neq j$ platí $a_i \equiv a_j \pmod{\gcd(n_i, n_j)}$. Pak je řešení jednoznačné modulo $\text{lcm}(n_1, \dots, n_m)$. To už ale zase zabíháme mimo rozsah tohoto skripta.

△

Příklad 7b.d: Vyřešíme soustavu $x \equiv 8 \pmod{5}$, $x \equiv -1 \pmod{6}$ a $x \equiv 14 \pmod{7}$.

Tento příklad má připomenout, že se ve větě ani algoritmu nikde nepožadovalo, aby n_i byla prvočísla, jen nesoučinnost po dvojicích, a na pravé strany b_i nebyly už vůbec žádné požadavky. Ukážeme si také pár zjednodušujících triků.

První věc je, že u každé rovnice si můžeme pravé strany modifikovat dle příslušného modula. Budeme tedy namísto té zadané řešit soustavu $x \equiv 3 \pmod{5}$, $x \equiv -1 \pmod{6}$ a $x \equiv 0 \pmod{7}$.

Teď také vidíme další zjednodušení, třetí člen v řešení se násobí nulou, tedy vůbec jej nemusíme vytvářet. Ale z cvičných důvodů si to také uděláme. Pro vytváření jednotlivých členů řešení použijeme systematický zápis, který některým (třeba mi) vyhovuje.

$$\begin{array}{c|c|c} \begin{array}{l} x \equiv 3 \pmod{5} \\ 3 \cdot 6 \cdot 7 \cdot ? \\ x_1 = 42^{-1} \pmod{5} \\ 42x_1 + 5y = 1 \\ x_1 = -2 \\ x = 3 \cdot 42 \cdot (-2) \end{array} & \begin{array}{l} x \equiv -1 \pmod{6} \\ -1 \cdot 5 \cdot 7 \cdot ? \\ x_2 = 35^{-1} \pmod{6} \\ 35x_2 + 6y = 1 \\ x_2 = -1 \\ x = +(-1) \cdot 35 \cdot (-1) \end{array} & \begin{array}{l} x \equiv 0 \pmod{7} \\ 0 \cdot 5 \cdot 6 \cdot ? \\ x_3 = 30^{-1} \pmod{7} \\ 30x_3 + 7y = 1 \\ x_3 = -3 \\ x = +0 \end{array} \\ \hline & & = -252 + 35 = -217 \end{array}$$

Inverze x_i jsme uhádli, to je často možné, v případě nouze si bokem uděláme tabulky pro rozšířený Euklidův algoritmus. Máme také $n = 5 \cdot 6 \cdot 7 = 210$.

Dostáváme množinu řešení $x = 210k - 217$ pro $k \in \mathbb{Z}$. Kdyby chtěl někdo lepšího reprezentanta, nabízí se $-217 + 2 \cdot 210 = 203$, tedy množina všech řešení je $x = 203 + 210k$, $k \in \mathbb{Z}$.

△

Poznámka: Hledání inverzních čísel lze podstatně zjednodušit, když si uvědomíme, že pracujeme ve světě modula. Například u prvního člena jsme měli řešit rovnici $42x \equiv 1 \pmod{5}$, což je ale ekvivalentní rovnici $2x_1 \equiv 1$, takže stačí hledat 2^{-1} pro modulo 5. Řešení příslušné rovnice $2x + 5y = 1$ lze snadno uhodnout. Podobně v dalších dvou případech stačí hledat $x_2 = 5^{-1} \pmod{6}$ a $x_3 = 2^{-1} \pmod{7}$.

Další prostor pro zjednodušení nám nabízí fáze formování členů. My jsme si do prvního přidávali $6 \cdot 7$, abychom zajistili vynulování vůči modulům 6 a 7. Jenže pravá strana první rovnice už dodala trojku, lze tedy pracovat se členem $3 \cdot 2 \cdot 7x_1$, kde $x_1 = 14^{-1} \pmod{5}$. Z tohoto pohledu se může vyplatit přepis druhé rovnice do tvaru $x \equiv 5 \pmod{6}$, protože pak druhý člen nemusí být $5 \cdot 5 \cdot 7x_2$, ale stačí $5 \cdot 7x_2$ a hledat $x_2 = 7^{-1} \pmod{6}$.

Pokud vidíme, že nám x bobtná, máme jej někdy možnost redukovat vhodnou volbou čísel x_i . Každé z nich je totiž určeno jen modulo n_i , takže například u druhého člena jsme coby řešení rovnice $35x_2 \equiv 1 \pmod{6}$ namísto $x_2 = -1$ mohli vzít $x_2 = -1 + 6 = 5$. Pak bychom měli $x = -252 - 175 = -427$. Dalo by se také použít $x_2 = -1 - 2 \cdot 6 = -13$ a máme $x = -252 + 455 = 203$, našeho nejlepšího zástupce.

Algoritmus je tedy (při ručním provádění) docela flexibilní, zejména pokud víme, oč v něm jde.

△

Čínská věta o zbytcích má mnoho praktických aplikací. Naštěstí se řešení dá najít algoritmicky, takže to nedělá počítacům problém a nemusíme se bát postupů, které Čínskou větu používají. Tímto optimistickým prohlášením končí hlavní část této sekce o soustavách, pokročilejší či odvážnější čtenáři si ale jistě její zbytek užijí. V něm si ukážeme jednu aplikaci Čínské věty, která dokáže významným způsobem urychlit násobení velkých čísel. Nejprve se trochu připravíme.

Souřadnice podruhé. Připomeneme si pojem kartézského součinu, nás bude konkrétně zajímat součin typu $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \dots \times \mathbb{Z}_{n_m}$, je to tedy množina vektorů o m souřadnicích, přičemž každá souřadnice používá čísel z příslušného \mathbb{Z}_{n_i} .

Na takovéto množině si zavedeme operace. Sčítání bude obdobné jako běžné sčítání vektorů, prostě sčítáme po souřadnicích, jen si teď musíme dát pozor, abychom v každé souřadnici používali správné sčítání z prostoru \mathbb{Z}_{n_i} . Podobně zavedeme i násobení vektorů, dva vektory násobíme tak, že vždy spolu vynásobíme odpovídající souřadnice dle odpovídající operace ze \mathbb{Z}_{n_i} a z výsledků vytvoříme nový vektor.

Příklad: Množina $\mathbb{Z}_4 \times \mathbb{Z}_7$ je množina všech vektorů (x, y) , přičemž x může nabývat hodnot 0, 1, 2, 3 a y může nabývat hodnot 0, 1, ..., 6. Když chceme sčítat dva takové vektory, tak v první souřadnici sčítáme modulo 4 a ve druhé modulo 7, podobně násobíme. Takže třeba $(3, 2) \oplus (1, 4) = (3 + 1 \pmod{4}, 2 + 4 \pmod{7}) = (0, 6)$ nebo $(3, 2) \odot (1, 4) = (3 \cdot 1 \pmod{4}, 2 \cdot 4 \pmod{7}) = (3, 1)$.

V části jsme ukázali, nač taková věc může být, pomocí souřadnic ze $\mathbb{Z}_4 \times \mathbb{Z}_7$ si dokážeme kódovat čísla z množiny $\mathbb{Z}_{4 \cdot 7} = \mathbb{Z}_{28}$ předpisem, že souřadnice čísla $a \in \mathbb{Z}_{28}$ jsou dána jako $(a \pmod{4}, a \pmod{7})$. Lemma nám zaručilo, že opravdu každé číslo ze \mathbb{Z}_{28} má jedinečně přiřazené souřadnice a naopak ke každé dvojici souřadnic $(x, y) \in \mathbb{Z}_4 \times \mathbb{Z}_7$ dokážeme najít odpovídající číslo a . Mělo to malý zádrhel, v té chvíli jsme ještě nevěděli, jak to a najít.

Tento problém je nyní vyřešen. Hledané a totiž musí splňovat $a \equiv x \pmod{4}$ a $a \equiv y \pmod{7}$, což je soustava lineárních kongruencí, kterou už umíme řešit. Vidíme tedy, že Čínská věta nám doplnila chybějící cihličku při

výstavbě souřadnicového nápadu. Čínská věta to také umí pro více souřadnic než jen dvě, což ukazuje, že to budeme chtít takto zobecnit.

Než se k tomu dostaneme, připomeneme si, proč jsou pro nás souřadnice zajímavé. Každý čtenář zná souřadnice vektoru. Vektory v rovině jsou šipky, se kterými sice manipulovat umíme, ale dělat to graficky dá práci. Mnohem jednodušší je vyjádřit si vektory pomocí souřadnic, operace pak provádime mnohem snáze. Podobně zde jsme v situaci, kdy si umíme čísla z rozmezí $0, 1, \dots, n_1 \cdot n_2$ kódovat pomocí souřadnic, které neprevýší větší čísel z n_1, n_2 , třeba v našem příkladě nemusíme pracovat s ohromnými číslami do 27, ale stačí umět pracovat s číslami do 6, na což stačí prsty na rukou. Máme tu tedy analogii s těmi vektory, ale aby byla analogie úplná (a užitečná), musíme také ukázat, že operace, které bychom chtěli provádět s těmi velkými číslami, můžeme namísto toho provádět s jejich souřadnicemi a vyjde to nestejně.

Tím se konečně dostaváme k jádru problému. Máme čísla n_1, n_2, \dots, n_m a v prostorech \mathbb{Z}_{n_i} odpovídající operace. Pomocí nich teď už také umíme sčítat a násobit vektory z prostoru $\mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_m}$. Na druhé straně máme čísla z množiny $\mathbb{Z}_{n_1 \cdot n_2 \cdots n_m}$, která bychom také rádi sčítali a násobili. Klíčová otázka zní: Když dvě čísla ze $\mathbb{Z}_{n_1 \cdots n_m}$ sečteme, popř. vynásobíme, dokážeme tento výsledek získat tak, že si dotyčná čísla nahradíme jejich souřadnicemi z $\mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_m}$, provedeme tam žádanou operaci a k výslednému vektoru souřadnic zase najdeme (Čínskou větou) odpovídající číslo ze $\mathbb{Z}_{n_1 \cdots n_m}$?

Dokážeme, že to funguje, nejprve si ale ujasníme, jak se kongruence chová při změně modula.

Lemma 7b.14.

Nechť $m, n \in \mathbb{N}$ a $a, b \in \mathbb{Z}$. Jestliže $a \equiv b \pmod{n}$ a m dělí n , pak $a \equiv b \pmod{m}$.

Důkaz necháváme jako cvičení. Mělo by to být jasné, například když $9 \equiv 21 \pmod{12}$, tedy od 9 se dá k 21 doskákat pomocí 12, tak už se dá určitě doskákat i pomocí 3 neboli $9 \equiv 21 \pmod{3}$. Je také zjevné, že toto tvrzení nebude platit naopak, pokud se dá od a k b doskákat pomocí menšího čísla, pak to ještě nemusí znamenat, že to půjde i pomocí většího.

Teď jsme připraveni dokázat, že operace lze provádět pomocí souřadnic, což ovšem musíme vyjádřit ve správném matematickém jazyce. Přechod od čísla k jeho souřadnicím vlastně vytváří jisté zobrazení, jeho inverze pak reprezentuje přechod od souřadnic zpět. Fungování operací, jak jsme jej výše naznačili, se pak musí převést na vlastnosti tohoto zobrazení.

Protože teď budeme muset opatrně pracovat s různými operacemi, v následujícím lemmatu a jeho důkazu raději zase budeme používat \oplus a \odot pro operace v \mathbb{Z}_n a zavedeme ještě (dočasně) speciální značení pro operace s vektory.

Lemma 7b.15.

Nechť $n_1, n_2, \dots, n_m \in \mathbb{N}$ jsou po dvou nesoudělná, $n_i \geq 2$. Označme $n = n_1 n_2 \cdots n_m$.

Tvrdíme, že zobrazení $T: \mathbb{Z}_n \mapsto \mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_m}$ definované jako $T(a) = (a \bmod n_1, a \bmod n_2, \dots, a \bmod n_m)$ je bijekce.

Pro $(x_1, x_2, \dots, x_m), (y_1, y_2, \dots, y_m) \in \mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_m}$ definujme operace

$$(x_1, x_2, \dots, x_m) \boxplus (y_1, y_2, \dots, y_m) = (x_1 \oplus y_1, x_2 \oplus y_2, \dots, x_m \oplus y_m),$$

$$(x_1, x_2, \dots, x_m) \boxdot (y_1, y_2, \dots, y_m) = (x_1 \odot y_1, x_2 \odot y_2, \dots, x_m \odot y_m),$$

kde $x_i \oplus y_i$ a $x_i \odot y_i$ jsou operace v příslušném \mathbb{Z}_{n_i} , tedy operace modulo n_i , a $x \oplus y$ a $x \odot y$ jsou operace v \mathbb{Z}_n . Pak pro všechna $a, b \in \mathbb{Z}_n$ platí $T(a \oplus b) = T(a) \boxplus T(b)$ a $T(a \odot b) = T(a) \boxdot T(b)$.

Důkaz (náznak): T je evidentně dobře definováno, neboť $a \bmod n_i \in \mathbb{Z}_{n_i}$ a tedy obrazy $T(a)$ neboli vektory $(a \bmod n_1, a \bmod n_2, \dots, a \bmod n_m)$ opravdu leží ve specifikované cílové množině.

Prostota: Jestliže $T(x) = T(y) = (b_1, b_2, \dots, b_m)$, pak obě čísla řeší soustavu rovnic $x \equiv b_i \pmod{n_i}$, tudíž podle Čínské větě o zbytcích $x \equiv y \pmod{n}$, pro prvky ze \mathbb{Z}_n pak nutně $x = y$.

T je na díky Čínské větě o zbytcích, pro dané $b_i \in \mathbb{Z}_{n_i}$ hledáme x takové, že $x \equiv b_i \pmod{n_i}$. Máme tedy bijekci.

Zbývají ověřit pravidla pro operace. Nejprve ověříme, že pro $a, b \in \mathbb{Z}_n$ platí $(a \oplus b) \bmod n_i = (a + b) \bmod n_i$: Podle definice $a \oplus b \equiv a + b \pmod{n}$, a protože $n_i | n$, pak podle Lemma také $a \oplus b \equiv a + b \pmod{n_i}$. Proto podle Věty dávají obě čísla stejný zbytek po dělení n_i .

Dále si všimneme, že $(a + b) \bmod n_i = ([a \bmod n_i] + [b \bmod n_i]) \bmod n_i$. To plyne z Faktu, při počítání modulo n_i lze vstupní čísla nahradit kongruentními alternativami.

Ted' už můžeme počítat

$$\begin{aligned} T(a \oplus b) &= ((a \oplus b) \bmod n_1, \dots, (a \oplus b) \bmod n_m) \\ &= ((a + b) \bmod n_1, \dots, (a + b) \bmod n_m) \\ &= (([a \bmod n_1] + [b \bmod n_1]) \bmod n_1, \dots, ([a \bmod n_m] + [b \bmod n_m]) \bmod n_m) \\ &= (a \bmod n_1, \dots, a \bmod n_m) \boxplus (b \bmod n_1, \dots, b \bmod n_m) = T(a) \boxplus T(b). \end{aligned}$$

Obdobný důkaz potvrdí pravidlo pro součin.

□

Toto Lemma vlastně říká, že prostory \mathbb{Z}_n a $\mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_m}$ jsou z hlediska algebraického totožné. Když chci něco vypočítat, tak se pomocí T a T^{-1} mohu volně pohybovat mezi těmito dvěma prostory a je jedno, kde výpočet provedu, nakonec to vyjde nestejno. Ukažme si to pro sčítání. Pokud na vzoreček z Lemmatu aplikujeme inverzní zobrazení T^{-1} , dostaneme $T^{-1}(T(a \oplus b)) = T^{-1}(T(a) \oplus T(b))$ neboli $a \oplus b = T^{-1}(T(a) \oplus T(b))$. Přeloženo do lidštiny: Pokud chci spočítat $a \oplus b$, tak namísto přímého výpočtu mohu nejprve obě čísla nahradit souřadnicemi pomocí T , tyto vektory sečist a od výsledného vektoru pak pomocí T^{-1} přejít zpět do původní množiny. Vidíme, že náš výsledek o zobrazení T opravdu říká to, o čem jsme neformálně rozmýšleli dříve.

Jaký je praktický dopad? Pokud umíme dobré násobit „malá“ čísla velikosti zhruba n_i , pak pomocí souřadnic už vlastně umíme i násobit čísla velikosti $n_1 \cdots n_m$. To je velice důležité, protože každý procesor má určitou mezní velikost, po kterou umí násobit opravdu hbitě, větší čísla pak násobí podstatně líněji. Náš trik nám umožní využít počítání s malými čísly k výpočtům s velkými čísly, časově se to i přes tu Čínskou větu vyplatí.

Například pokud procesor umí rychle počítat do 100, tak lze za n_i zvolit 95, 97, 98 a 99 a pak můžeme počítat rychle až do 89403930, navíc ty souřadnicové výpočty lze dělat paralelně. Populární volba jsou čísla tvaru $2^a - 1$, protože $\gcd(2^a - 1, 2^b - 1) = 2^{\gcd(a,b)} - 1$ (viz cvičení).

Příklad 7b.e: Nechť $n_1 = 3$, $n_2 = 4$. Pak $n = 12$ a množina \mathbb{Z}_{12} je reprezentována množinou $\mathbb{Z}_3 \times \mathbb{Z}_4$ takto:

$$T(0) = (0, 0), T(1) = (1, 1), T(2) = (2, 2), T(3) = (0, 3), T(4) = (1, 0), T(5) = (2, 1), T(6) = (0, 2), T(7) = (1, 3), \\ T(8) = (2, 0), T(9) = (0, 1), T(10) = (1, 2), T(11) = (2, 3).$$

Fungují opravdu operace? Pro zjednodušení už zase budeme psát normální značky.

Začneme sčítáním, v \mathbb{Z}_{12} máme například $7 + 10 = 5$. Ted' to zkusíme přes souřadnice. Nejprve čísla nahradíme souřadnicemi, tedy použijeme T , pak provedeme operaci na vektorech.

$$7 + 10 \mapsto T(7) + T(10) = (1, 3) + (1, 2) = ([1 + 1] \bmod 3, [3 + 2] \bmod 4) = (2, 1).$$

Ted' se vrátíme do \mathbb{Z}_{12} , $T^{-1}(2, 1) = 5$, opravdu to souhlasí.

Zkusíme násobení, kolik je $2 \cdot 3$? Přeneseme pomocí T : $(2, 2) \cdot (0, 3) = ([2 \cdot 0] \bmod 3, [2 \cdot 3] \bmod 4) = (0, 2)$ a $T^{-1}(0, 2) = 6$. Vyšlo to.

△

Kromě násobení dokáže Čínská věta pomoci i s umocňováním.

Příklad 7b.f: V příkladě jsme počítali 6^{1040} modulo 91 a dalo to docela práci. Zde si zkusíme jiný postup.

Máme $91 = 7 \cdot 13$, tak se podíváme na výpočet modulo tato dvě čísla. Jsou to prvočísla, takže lze použít malou Fermatovu větu, navíc budeme mít ve výpočtu znatelně menší čísla než v.

$$\text{modulo } 7 : 6^{1040} = 6^{173 \cdot 6 + 2} = (6^6)^{173} \cdot 6^2 \equiv 1^{173} \cdot 6^2 = 1 \cdot 36 \equiv 1 \pmod{7},$$

$$\text{modulo } 13 : 6^{1040} = 6^{86 \cdot 12 + 8} = (6^{12})^{86} \cdot 6^8 \equiv 1^{86} \cdot 36^4 \equiv 1 \cdot 10^4 = 100^2 \equiv 9^2 = 81 \equiv 3 \pmod{13}.$$

Číslo $x = 6^{1040}$ tedy splňuje kongruenze $x \equiv 1 \pmod{7}$ a $x \equiv 3 \pmod{13}$, my už víme, jak takovou soustavu řešit (algoritmus).

$$\begin{array}{l|l} \begin{array}{l} x \equiv 1 \pmod{7} \\ 1 \cdot 13 \cdot ? \\ x_1 = 13^{-1} \pmod{7} \\ 13x_1 + 7y = 1 \\ x_1 = -1 \\ x = 1 \cdot 13 \cdot (-1) \end{array} & \begin{array}{l} x \equiv 3 \pmod{13} \\ 3 \cdot 7 \cdot ? \\ x_2 = 7^{-1} \pmod{13} \\ 7x_2 + 13y = 1 \\ x_2 = 2 \\ +3 \cdot 7 \cdot 2 \end{array} \\ \hline & = -13 + 42 = 29 \end{array}$$

Dostali jsme 29, což je stejný výsledek jako v příkladě a zdá se, že pohodlněji, ještě lepší by bylo, kdybychom měli na řešení soustav kongruencí hotový programek.

△

Pro lidi zabývající se rychlými výpočty modulo jsou podobné metody velice důležité.

7b.16 Bonus: Úpravy rovnic

Samozřejmě ne všechny rovnice jsou lineární, pak to začne být zajímavé. My se zde doplňkově zaměříme na otázku, které z našich obvyklých triků pro práci s rovnicemi fungují i ve světě modulo n , kde řešíme kongruence typu $x \equiv y \pmod{n}$. Mnohé z výsledků již vlastně máme dokázány, jen se na ně podíváme z jiného úhlu.

Ze symetrie relace kongruence (Fakt) třeba vyplývá, že můžeme u rovnic modulo prohazovat strany, to je dobrý začátek. Měli jsme tam i tranzitivitu, což nám zase říká, že pokud k rovnici $x \equiv y \pmod{n}$ dostaneme $y \equiv z \pmod{n}$, tak už platí $x \equiv z \pmod{n}$. Jinými slovy, jednu stranu rovnice můžeme nahradit něčím jiným, co je jí rovno ve smyslu modula.

Zkusme něco méně triviálního. Jak se změní množina řešení, pokud k rovnicím něco přičteme či je vynásobíme číslem?

! Fakt 7b.17.

Nechť $n \in \mathbb{N}$. Uvažujme rovnici $x \equiv y \pmod{n}$ pro $x, y \in \mathbb{Z}$. Pak pro každé $c \in \mathbb{Z}$ platí:

- (i) Rovnice $x + c \equiv y + c \pmod{n}$ a $x - c \equiv y - c \pmod{n}$ mají stejnou množinu řešení jako rovnice původní.
- (ii) Jestliže x, y řeší rovnici původní, tak řeší i rovnice $cx \equiv cy \pmod{n}$ a $x^c \equiv y^c \pmod{n}$.

Důkaz (rutinní, poučný): (i) 1) Nejprve ukážeme, že každé řešení původní rovnice je i řešením upravené rovnice $x + c \equiv y + c \pmod{n}$. Mějme tedy čísla $x, y \in \mathbb{Z}$ splňující $x \equiv y \pmod{n}$. Máme určitě i $c \equiv c \pmod{n}$ a podle Věty po sečtení zase dostaneme platnou kongruenci.

2) Teď potřebujeme ukázat, že naopak pokud máme nějaká řešení x, y rovnice $x + c \equiv y + c \pmod{n}$, pak už nutně musí splňovat $x \equiv y \pmod{n}$. Ale to je snadné, aplikujeme 1) s opačným prvkem $-c$ a dostaneme $(x + c) + (-c) \equiv (y + c) + (-c) \pmod{n}$, asociativní zákon nám umožňuje to přepsat jako $x + [c + (-c)] \equiv y + [c + (-c)] \pmod{n}$, tedy $x + 0 \equiv y + 0 \pmod{n}$ a je to.

Tvrzení o odčítání se snadno dostane tak, že se ekvivalence přičítání aplikuje na opačný prvek $-c$.

(ii) Nechť čísla x, y řeší rovnost $x \equiv y \pmod{n}$. Pak máme kongruenci $x \equiv y \pmod{n}$, tudíž podle Faktu platí $x^c \equiv y^c \pmod{n}$, a navíc $c \equiv c \pmod{n}$, proto podle Věty (iii) platí $cx \equiv cy \pmod{n}$. □

Vidíme, že přičítat k rovnicím je bez problémů, ale u násobení a umocňování se dají čekat problémy, protože jsme v našem tvrzení napsali implikaci, ne ekvivalenci.

U umocňování to není překvapovák, dokonce i v oboru reálných čísel se snadno stane, že rovnice $x = y$ má méně řešení než rovnice $x^2 = y^2$, tedy umocněním rovnice mohou přibýt další řešení. Je to známý problém a nebudeme se v tom dále vrtat.

Mnohem zajímavější je násobení rovnic. Ani tam nejsme v \mathbb{R} zcela v bezpečí, my totiž víme, že násobení rovnice je ekvivalentní operací jen v případě, že použijeme nenulové číslo, jinými slovy číslo, ke kterému existuje inverze. Tím máme možnost ono dodané číslo zase zkrátit a vrátit se k původní rovnici, množiny řešení tedy souhlasí. A přesně tady je problém při počítání modulo. Máme tedy následující výstrahu:

- Z platnosti rovnice $cx \equiv cy \pmod{n}$ obecně nedostáváme $x \equiv y \pmod{n}$. Jinými slovy, **nelze krátit**. Příklad: Platí $3 \cdot 4 \equiv 3 \cdot 2 \pmod{6}$, neboť $12 - 6 = 6$ je dělitelné šesti. Když ale zkuste zkrátit trojkou, dostaneme nepravdivou rovnost $4 \equiv 2 \pmod{6}$.

Zajímavý případ je, když na pravé straně použijeme $y = 0$.

- Z platnosti rovnice $cx \equiv 0 \pmod{n}$ obecně nemůžeme odvodit, že alespoň jedno z čísel nalevo je nulové modulo n . Například $3 \cdot 2 \equiv 0 \pmod{6}$. I s tímto jsme se již setkali, když jsme disktovali dělitele nuly.

Protože jádrem problému je možnost najít inverzní prvek, poznatky z kapitoly nám okamžitě řeknou, jak se věci mají.

! Fakt 7b.18.

Nechť $n \in \mathbb{N}$, uvažujme $y, x, c \in \mathbb{Z}$.

- (i) Jestliže $cx \equiv cy \pmod{n}$ a $\gcd(c, n) = 1$, pak $x \equiv y \pmod{n}$.
- (ii) Jestliže $cx \equiv 0 \pmod{n}$ a $\gcd(c, n) = 1$, pak $x \equiv 0 \pmod{n}$.

Důkaz (rutinní): (i): Podle Věty má c inverzní prvek c^{-1} modulo n , díky kterému lze počítat

$$x = 1x \equiv (c^{-1}c)x \equiv c^{-1}(cx) \equiv c^{-1}(cy) \equiv 1y = y \pmod{n}.$$

Alternativa: Inverzním prvkem c^{-1} vynásobíme podle Faktu obě strany dané rovnice, použijeme asociativitu k přezávorkování a máme $(c^{-1}c)x \equiv (c^{-1}c)y \pmod{n}$ neboli $1x \equiv 1y \pmod{n}$, a je to.

(ii): Vynásobíme obě strany rovnice prvkem c^{-1} . □

Tím máme jasno. Pro doplnění se podíváme na situaci, kdy máme $cx \equiv cy \pmod{n}$, ale c a n nesoudělné nejsou. Pak lze pořád něco odvodit, ale za cenu změny referenčního základu modula.

Věta 7b.19.

Nechť $n \in \mathbb{N}$, uvažujme $x, y, c \in \mathbb{Z}$. Jestliže $cx \equiv cy \pmod{n}$, pak $x \equiv y \pmod{\frac{n}{\gcd(c, n)}}$.

Důkaz (poučný): Podle předpokladu máme $k \in \mathbb{Z}$ takové, že $c(x - y) = kn$. To znamená, že $\frac{c}{\gcd(c, n)}(x - y) = k \frac{n}{\gcd(c, n)}$, takže $\frac{n}{\gcd(c, n)}$ dělí $\frac{c}{\gcd(c, n)}(x - y)$. Podle Lemma 6a.9 jsou ovšem $\frac{n}{\gcd(c, n)}$ a $\frac{c}{\gcd(c, n)}$ čísla nesoudělná, tudíž podle Lemma 6a.23 musí $\frac{n}{\gcd(c, n)}$ dělit $x - y$. □

Vyzkoušíme si to. Před chvílí jsme v poznámce o nemožnosti krátit uvažovali rovnici $3 \cdot 4 \equiv 3 \cdot 2 \pmod{6}$. Podle právě dokázané věty by mělo platit $4 \equiv 2 \pmod{\frac{6}{\gcd(3, 6)}}$ neboli $4 \equiv 2 \pmod{2}$, což opravdu funguje.

To je sice pěkné, ale my přece žijeme ve světě modula n ! Informace používající nějaké jiné modulo se občas dá částečně použít i ve světě n , ale je to komplikovanější a záleží na typu rovnice, který řešíme, takže se tomu nebudeme dále věnovat. Dá se říct, že se změně modula uprostřed výpočtu snažíme vyhýbat a dochází k tomu málokdy.

Cvičení

Cvičení 7b.1 (rutinní, zkouškové): Vyřešte následující kongruence:

- | | | |
|---------------------------------|-------------------------------------|---------------------------------|
| (i) $3x \equiv 7 \pmod{10}$; | (iii) $84x \equiv -56 \pmod{308}$; | (v) $6x \equiv 10 \pmod{8}$; |
| (ii) $12x \equiv 0 \pmod{20}$; | (iv) $3x \equiv 7 \pmod{9}$; | (vi) $11x \equiv 0 \pmod{40}$. |

Cvičení 7b.2 (rutinní, zkouškové): Vyřešte následující rovnice v daném \mathbb{Z}_n :

- | | | |
|--------------------------------------|---------------------------------------|--|
| (i) $12x = 18$ v \mathbb{Z}_{42} ; | (iii) $10x = 0$ v \mathbb{Z}_{35} ; | (v) $84x = 126$ v \mathbb{Z}_{210} ; |
| (ii) $9x = 7$ v \mathbb{Z}_{20} ; | (iv) $8x = 10$ v \mathbb{Z}_{12} ; | (vi) $8x = 0$ v \mathbb{Z}_{12} ; |

Cvičení 7b.3 (rutinní, zkouškové): Vyřešte následující soustavy kongruencí:

- | | | | |
|---------------------------|----------------------------|-----------------------------|----------------------------|
| (i) $x \equiv 0 \pmod{3}$ | (ii) $x \equiv 4 \pmod{2}$ | (iii) $x \equiv 1 \pmod{7}$ | (iv) $x \equiv 3 \pmod{5}$ |
| $x \equiv 1 \pmod{4}$ | $x \equiv -4 \pmod{3}$ | $x \equiv 0 \pmod{9}$ | $x \equiv 4 \pmod{4}$ |
| $x \equiv 2 \pmod{5}$; | $x \equiv 4 \pmod{5}$; | $x \equiv -1 \pmod{11}$; | $x \equiv 5 \pmod{3}$. |

Cvičení 7b.4 (rutinní, zkouškové): Které z následujících rovnic jsou řešitelné v \mathbb{Z}_{168} ?

- | | | | |
|-----------------|-----------------|-----------------|-----------------|
| a) $25x = 13$; | b) $30x = 12$; | c) $30x = 15$; | d) $16x = 24$. |
|-----------------|-----------------|-----------------|-----------------|

Cvičení 7b.5 (dobré, zkouškové): Uvažujme rovnici $(6 - t)x = 24$ v \mathbb{Z}_{40} . Pro které hodnoty t z rozmezí $0, \dots, 5$ má tato rovnice

- | | | | |
|-------------------------|-----------------------|-----------------------|------------------|
| a) přesně čtyři řešení? | b) přesně tři řešení? | c) přesně pět řešení? | d) žádné řešení? |
|-------------------------|-----------------------|-----------------------|------------------|

Cvičení 7b.6 (rutinní, poučné): Nechť $n \in \mathbb{N}$ a $c \in \mathbb{Z}$, předpokládejme, že $\gcd(c, n) = 1$. Dokažte, že jestliže $x, y \in \mathbb{Z}$ splňují $cx \equiv 1 \pmod{n}$ a $cy \equiv 1 \pmod{n}$, pak $x \equiv y \pmod{n}$.

Cvičení 7b.7 (rutinní, poučné): Nechť $m, n \in \mathbb{N}$ a $a, b \in \mathbb{Z}$. Dokažte, že jestliže $a \equiv b \pmod{n}$ a m dělí n , pak $a \equiv b \pmod{m}$.

Cvičení 7b.8 (poučné): (i) Nechť m, n jsou nesoudělná. Předpokládejte, že $a \equiv b \pmod{m}$ a $a \equiv b \pmod{n}$. Pak existuje $k \in \mathbb{Z}$ takové, že $a - b = km$. Z druhého předpokladu zase víme, že $n|(a - b)$, tedy $n|(km)$. Použijte Lemma 6a.23 k důkazu, že $a \equiv b \pmod{mn}$.

(ii) Nechť $n_1, n_2, \dots, n_m \in \mathbb{N}$ jsou po dvou nesoudělná. Označme $n = n_1 \cdot n_2 \cdots n_m$. Dokažte matematickou indukcí na m , že jestliže $a, b \in \mathbb{Z}$ splňují $a \equiv b \pmod{n_i}$ pro všechna $i = 1, \dots, m$, pak $a \equiv b \pmod{n}$.

Bude se hodit (i).

Cvičení 7b.9 (rutinní, poučné): Nechť $n \in \mathbb{N}$, nechť $a, b \in \mathbb{Z}$. Uvažujme nějaké řešení x_p kongruence $ax \equiv b$.

- | |
|--|
| (i) Dokažte, že když je i $x \in \mathbb{Z}$ řešením této kongruence, tak číslo $x_h = x - x_p$ řeší kongruenci $ax \equiv 0 \pmod{n}$. |
| (ii) Dokažte, že když $x_h \in \mathbb{Z}$ je řešením kongruence $ax \equiv 0 \pmod{n}$, tak $x = x_p + x_h$ řeší kongruenci $ax \equiv b \pmod{n}$. |

Cvičení 7b.10 (poučné): Nechť $c \in \mathbb{N}$. Zde dokážeme, že pro $a, b \in \mathbb{N}$ platí $\gcd(c^a - 1, c^b - 1) = c^{\gcd(a, b)} - 1$.

Vyplyne to z následujících kroků:

- | |
|--|
| (i) Pomocí cvičení dokažte, že když $d a$, pak také $(c^d - 1) (c^a - 1)$. |
| (ii) Odvodte, že $(c^{\gcd(a, b)} - 1) (c^a - 1)$ a $(c^{\gcd(a, b)} - 1) (c^b - 1)$. |

(iii) Ukažte, že když číslo d dělí $c^a - 1$ a $c^b - 1$, pak $d | (c^{\gcd(a,b)} - 1)$.

Návod pro (iii): Přepište dělitelnost jako kongruenci, použijte Bezouta.

(ii) a (iii) říkají, že $c^{\gcd(a,b)} - 1$ je společný dělitel a je největší takový.

Řešení:

7b.1: (i): $7 = 3x + 10k$, evidentně $\gcd(3, 10) = 1 = (-3) \cdot 3 + 1 \cdot 10$ (lze uhádnout), vynásobíme Bezouta sedmi, $7 = 3 \cdot (-21) + 7 \cdot 10$, tedy $x = -21$ je řešení.

Rovnice $3x \equiv 0 \pmod{10}$ má řešení $x_h = 10k$, proto řešení dané rovnice je $x = -21 + 10k$, $k \in \mathbb{Z}$. Kdo chce, použije $x = 9 + 10k$, $k \in \mathbb{Z}$.

(ii): Evidentně $\gcd(12, 20) = 4$, zkrátíme, $3x \equiv 0 \pmod{5}$ má řešení $x = 5k$, $k \in \mathbb{Z}$.

(iii): $-56 = 84x + 308k$, Euklidem $\gcd(308, 84) = 28 = (-1) \cdot 308 + 4 \cdot 84$, protože $\frac{-56}{28} = -2 \in \mathbb{Z}$ má rovnice řešení, vynásobíme Bezouta tou mínus dvojkou, $-56 = 84 \cdot (-8) + 2 \cdot 308$, tedy $x = -8$ je řešení.

Rovnice $84x \equiv 0 \pmod{308}$ se vydělí 28 na $3x \equiv 0 \pmod{11}$, má řešení $x_h = 11k$, proto řešení dané rovnice je $x = -8 + 11k$, $k \in \mathbb{Z}$. Kdo chce, použije $x = 3 + 11k$, $k \in \mathbb{Z}$.

(iv): protože $\gcd(3, 9) = 3$ a 7 není násobkem 3, rovnice nemá řešení.

(v): $10 = 6x + 8k$, evidentně $\gcd(6, 8) = 2 = (-1) \cdot 6 + 1 \cdot 8$ (lze uhádnout), rovnice má řešení, neboť $\frac{10}{2} = 5 \in \mathbb{Z}$, vynásobíme Bezouta tou pětkou, $10 = 6 \cdot (-5) + 5 \cdot 8$, tedy $x = -5$ je řešení.

Rovnice $6x \equiv 0 \pmod{8}$ se vydělí 2 na $3x \equiv 0 \pmod{4}$, má řešení $x_h = 4k$, proto řešení dané rovnice je $x = -5 + 4k$, $k \in \mathbb{Z}$. Kdo chce, použije $x = 3 + 4k$, $k \in \mathbb{Z}$.

(vi): Protože $\gcd(11, 40) = 1$, je množina řešení $x = 40k$, $k \in \mathbb{Z}$.

7b.2: (i): $18 = 12x + 42k$, Euklidem nebo odhadem $\gcd(42, 12) = 6 = 1 \cdot 42 + (-3) \cdot 12$, protože $\frac{18}{6} = 3 \in \mathbb{Z}$ má rovnice řešení, vynásobíme Bezouta tou trojkou, $18 = (-9) \cdot 12 + 3 \cdot 42$, tedy $x = -9$ je řešení.

Rovnice $12x \equiv 0 \pmod{42}$ se vydělí 6 na $2x \equiv 0 \pmod{7}$, má řešení $x_h = 7k$, proto řešení kongruenční rovnice je $x = -9 + 7k$, přepíšeme na kongruentní $x = -9 + 7 \cdot 2 + 7k = 5 + 7k$. Je $\gcd(42, 12) = 6$ řešení v \mathbb{Z}_{42} , proto řešení dané úlohy je $x = 5 + 7k$ pro $k = 0, 1, 2, 3, 4, 5$ neboli $\{5, 12, 19, 26, 33, 40\}$.

(ii): $9x = 7$ v \mathbb{Z}_{20} ; $7 = 9x + 20k$, Euklidem nebo odhadem $\gcd(20, 9) = 1 = (-4) \cdot 20 + 9 \cdot 9$, protože $\frac{7}{1} = 7 \in \mathbb{Z}$ má rovnice řešení, vynásobíme Bezouta tou sedmičkou, $7 = 9 \cdot 63 + (-28) \cdot 20$, tedy $x = 63$ je řešení.

Rovnice $9x \equiv 0 \pmod{20}$ má řešení $x_h = 20k$, proto řešení kongruenční rovnice je $x = 63 + 20k$, přepíšeme na kongruentní $x = 3 + 20k$. Je $\gcd(20, 9) = 1$ řešení v \mathbb{Z}_{20} , proto řešení dané úlohy je $x = 3$.

(iii): Řešíme $10x \equiv 0 \pmod{35}$, uhodneme $\gcd(35, 10) = 5$, vydělíme rovnici na $2x \equiv 0 \pmod{7}$, takže řešení kongruenční rovnice jsou $x = 7k$. Je $\gcd(35, 10) = 5$ řešení v \mathbb{Z}_{35} , proto řešení dané úlohy je $x = 7k$ pro $k = 0, 1, 2, 3, 4$ neboli $\{0, 7, 14, 21, 28\}$.

(iv): $10 = 8x + 12k$, Euklidem nebo odhadem $\gcd(12, 8) = 4 = 1 \cdot 12 + (-1) \cdot 8$, protože 4 nedělí 10, rovnice nemá řešení.

(v): $126 = 84x + 210k$, Euklidem $\gcd(210, 84) = 42 = 1 \cdot 210 + (-2) \cdot 84$, protože $\frac{126}{42} = 3 \in \mathbb{Z}$ má rovnice řešení, vynásobíme Bezouta tou trojkou, $126 = (-6) \cdot 84 + 3 \cdot 210$, tedy $x = -6$ je řešení.

Rovnice $84x \equiv 0 \pmod{210}$ se vydělí 42 na $2x \equiv 0 \pmod{5}$, má řešení $x_h = 5k$, proto řešení kongruenční rovnice je $x = -6 + 5k$, přepíšeme na kongruentní $x = 4 + 5k$. Je $\gcd(210, 84) = 42$ řešení v \mathbb{Z}_{210} , proto řešení dané úlohy je $x = 4 + 5k$ pro $k = 0, 1, \dots, 41$ neboli $\{4, 9, 14, 19, \dots, 204, 209\}$.

(vi): Řešíme $8x \equiv 0 \pmod{12}$, uhodneme $\gcd(12, 8) = 4$, vydělíme rovnici na $2x \equiv 0 \pmod{3}$, takže řešení kongruenční rovnice jsou $x = 3k$. Je $\gcd(12, 8) = 4$ řešení v \mathbb{Z}_{12} , proto řešení dané úlohy je $x = 3k$ pro $k = 0, 1, 2, 3$ neboli $\{0, 3, 6, 9\}$.

7b.3: (i): $n = 60$, $N_1 = 20$, inverze v \mathbb{Z}_3 je $x_1 = -1$; $N_2 = 15$, inverze v \mathbb{Z}_4 je $x_2 = -1$; $N_3 = 12$, inverze v \mathbb{Z}_5 je $x_3 = -2$. $x = 0 \cdot 20 \cdot (-1) + 1 \cdot 15 \cdot (-1) + 2 \cdot 12 \cdot (-2) = -63 \equiv 57 \pmod{60}$. Řešení jsou $x = 60k - 63$ nebo třeba $57 + 60k$ pro $k \in \mathbb{Z}$.

(ii): $n = 30$, $N_1 = 15$, inverze v \mathbb{Z}_2 je $x_1 = 1$; $N_2 = 10$, inverze v \mathbb{Z}_3 je $x_2 = 1$; $N_3 = 6$, inverze v \mathbb{Z}_5 je $x_3 = 1$. $x = 4 \cdot 15 \cdot 1 + (-4) \cdot 10 \cdot 1 + 4 \cdot 6 \cdot 1 = 44 \equiv 14 \pmod{30}$. Řešení jsou $x = 44 + 30k$ nebo třeba $14 + 30k$ pro $k \in \mathbb{Z}$.

(iii): $n = 693$, $N_1 = 99$, inverze v \mathbb{Z}_7 je $x_1 = 1$; $N_2 = 77$, inverze v \mathbb{Z}_9 je $x_2 = 2$; $N_3 = 63$, inverze v \mathbb{Z}_{11} je $x_3 = -4$. $x = 1 \cdot 99 \cdot 1 + 0 \cdot 77 \cdot 2 + (-1) \cdot 63 \cdot (-4) = 351$. Řešení jsou $x = 351 + 693k$ pro $k \in \mathbb{Z}$.

(iv): Přepis na $x \equiv 3 \pmod{5}$, $x \equiv 0 \pmod{4}$, $x \equiv 2 \pmod{3}$. $n = 60$, $N_1 = 12$, inverze v \mathbb{Z}_5 je $x_1 = 3$; N_2 netřeba řešit; $N_3 = 20$, inverze v \mathbb{Z}_3 je $x_3 = 2$. $x = 3 \cdot 12 \cdot 3 + 0 + 2 \cdot 20 \cdot 2 = 188$. Řešení jsou $x = 188 + 60k$ nebo třeba $x = 8 + 60k$ pro $k \in \mathbb{Z}$.

7b.4: Podmínka je $\gcd(a, n) | b$. a): $\gcd(25, 168) = 1$, $1 \mid 13$, ano. b): $\gcd(30, 168) = 6$, $6 \mid 12$, ano. c): $\gcd(30, 168) = 6$, neplatí $6 \mid 15$, ne. d): $\gcd(16, 168) = 8$, $8 \mid 24$, ano.

7b.5: Počet řešení je roven $\gcd(a, n)$, ale musí platit $\gcd(a, n) | b$. a): Potřebujeme $\gcd(6 - t, 40) = 4$, pak také $4 \mid 24$, to platí pro $t = 2$.

b): Potřebujeme $\gcd(6 - t, 40) = 3$, to není možné.

c): Potřebujeme $\gcd(6 - t, 40) = 5$, nastane pro $t = 1$, ale neplatí $5 \mid 24$, takže žádné řešení.

d): Žádné řešení nastane, když $\gcd(6 - t, 40)$ nedělí 24. Protože $40 = 8 \cdot 5$ a $6 - t \leq 6$, možná \gcd jsou 2, 4, 5. Z nich jen 5 nedělí 24, u ostatních budou řešení. Závěr: Žádné řešení nebude pro $t = 1$.

7b.6: Použijte tranzitivitu, $cx \equiv cy \pmod{n}$, pak krácení.

Nebo: Z rovnice plyne, že pro nějaká $k, l \in \mathbb{Z}$ plací $cx = 1 + kn$ a $cy = 1 + ln$. Pak $cy - cx = (k - l)n$, tedy $c(x - y) = (k - l)n$, a protože $\gcd(c, n) = 1$, musí n dělit $x - y$ (Lemma 6a.23).

7b.7: $n \mid (a - b)$, $m \mid n$ a tranzitivita, nebo $a - b = kn$, $n = lm$ a dosadit.

7b.8: (i) Víme, že $n \mid (km)$. Ale $\gcd(m, n) = 1$, proto podle Lemma 6a.23 platí $n \mid k$ neboli $k = ln$ pro nějaké $l \in \mathbb{Z}$. Pak $a - b = lnm$, tedy $(mn) \mid (a - b)$, což dává závěr.
(ii) (0) $m = 1$ evidentně platí.

(1) Předpokládejme platnost pro m . Mějme n_1, \dots, n_m, n_{m+1} po dvou nesoudělná a a, b dle předpokladu. Protože $a \equiv b \pmod{n_1}, \dots, a \equiv b \pmod{n_m}$, musí podle indukčního předpokladu platit $a \equiv b \pmod{n'}$, kde $n' = n_1 \cdot n_2 \cdots n_m$. Také $a \equiv b \pmod{n_{m+1}}$, proto podle (i) platí $a \equiv b \pmod{n' n_{m+1}}$, ale $n' n_{m+1} = n_1 \cdot n_2 \cdots n_m \cdot n_{m+1}$, přesně jak jsme potřebovali.

7b.9: (i): $a(x - x_p) = ax - ax_p \equiv b - b = 0 \pmod{n}$.

(ii) je podobné.

7b.10: (i): $d \mid a \implies a = dm$, podle cvičení s dosazením c za x je $c^a - 1 = (c^d - 1)(c^{(m-1)d} + \cdots + c^d + 1)$.

(ii): Protože $\gcd(a, b) \mid a$ a $\gcd(a, b) \mid b$, plyne to hned z (i).

(iii): $d \mid (c^a - 1)$, proto $c^a \equiv 1 \pmod{d}$, podobně $c^b \equiv 1 \pmod{d}$. Podle Bezouta $\gcd(a, b) = ax + by$, tedy $c^{\gcd(a, b)} = (c^a)^x \cdot (c^b)^y \equiv 1^x \cdot 1^y = 1 \pmod{d}$, tedy $d \mid (c^{\gcd(a, b)} - 1)$.

7c. Matice a polynomy modulo

S maticemi a polynomy se pracuje také v jiných světech než ve světě reálných čísel. Věci pak mohou fungovat trochu jinak, než jsme zvyklí. Pro příklad není třeba chodit daleko, například polynom $p = 2x - 1$ má kořen v oboru reálných čísel (jmenovitě $x = \frac{1}{2}$ splňuje $p(x) = 0$), ale nemá kořen v oboru celých čísel. Nebo třeba matice $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ má inverzní matici $A^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}$ v oboru reálných čísel, ale v oboru celých čísel už invertibilní není (stává se singulární). Ještě zajímavější to je, když začneme pracovat v \mathbb{Z}_n . Podíváme se, na které věci se stále můžeme spolehnout (to bude krátký seznam) a kde naopak může číhat překvapení.

7c.1 Matice nad \mathbb{Z}_n

Začneme dobrou zprávou: Sčítání a násobení matic i přechod k transponované matici fungují v \mathbb{Z} i v \mathbb{Z}_n přesně tak, jak jsme zvyklí. Rovněž determinanty mají stále dobrý smysl a můžeme je počítat podle definice (přes všechny permutace, což známe dobře pro matice 2×2 a 3×3) a také rozvojem podle sloupce či řádku. Platí také, že podle determinantu poznáme, zda je matice regulární neboli invertibilní, tedy zda k ní existuje inverzní matice. Zde si ale musíme kritérium trochu upravit.

Věta 7c.2.

Ke čtvercové matici A nad \mathbb{Z}_n existuje matice inverzní právě tehdy, když je její determinant $|A|$ invertibilní v \mathbb{Z}_n .

Tato inverzní matice je pak dána vzorcem $A^{-1} = |A|^{-1} D^T$, kde D je matice kofaktorů.

Připomeňme, že prvek d_{ij} matice D získáme tak, že z matice A vyškrtneme řádek i a sloupec j , spočítáme determinant výsledné matice a vynásobíme jej číslem $(-1)^{i+j}$.

Tato věta je zobecněním situace z reálného oboru, kde jsou regulární ty matice, které mají nenulový determinant, což souhlasí, právě nenulová čísla jsou v \mathbb{R} invertibilní. Podobně matice nad \mathbb{Z} jsou regulární právě tehdy, když mají determinant roven ± 1 .

Příklad 7c.a: Najdeme A^{-1} pro $A = \begin{pmatrix} 2 & 3 \\ 4 & 13 \end{pmatrix}$ v prostoru \mathbb{Z}_{45} .

Nejprve spočítáme $|A| = 2 \cdot 13 - 4 \cdot 3 = 14$. Protože $\gcd(14, 45) = 1$, je 14 invertibilní v \mathbb{Z}_{45} a tudíž je i A regulární.

Rovnou najdeme $|A|^{-1} = 14^{-1}$ v \mathbb{Z}_{45} . Pomocí rozšířeného Euklidova algoritmu získáme $1 = 5 \cdot 45 + (-16) \cdot 14$, proto $(-16) \cdot 14 \equiv 1 \pmod{45}$, tedy $14^{-1} = 29$ v \mathbb{Z}_{45} .

Ted' sestavíme matici D : d_{11} dostaneme vyškrtnutím prvního řádku a sloupce z A a nalezením determinantu výsledné matice (13), výsledek je $d_{11} = (-1)^{1+1} \cdot 13 = 13$, podobně je $d_{12} = (-1)^{1+2} \cdot 4 = -4 \equiv 41 \pmod{45}$, $d_{21} = (-1)^{2+1} \cdot 3 = -3 \equiv 42 \pmod{45}$, $d_{22} = (-1)^{2+2} \cdot 2 = 2$. Proto $D = \begin{pmatrix} 13 & 41 \\ 42 & 2 \end{pmatrix}$ a tedy

$$A^{-1} = 29 \begin{pmatrix} 13 & 42 \\ 41 & 2 \end{pmatrix} = \begin{pmatrix} 17 & 3 \\ 19 & 13 \end{pmatrix}.$$

Ověřte, že opravdu $\begin{pmatrix} 2 & 3 \\ 4 & 13 \end{pmatrix} \begin{pmatrix} 17 & 3 \\ 19 & 13 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ při počítání v \mathbb{Z}_{45} .

△

Čtenáře možná napadne, proč jsme u metod výpočtu determinantu, popřípadě u metody výpočtu inverzní matice nevedli Gaussovou eliminaci. Důvod je jednoduchý, tato metoda totiž při práci v \mathbb{Z}_n narází na nečekané potíže.

Příklad 7c.b: Uvažujme matici $A = \begin{pmatrix} 2 & 3 \\ 2 & 2 \end{pmatrix}$ nad \mathbb{Z}_6 . Výpočtem modulo 6 hravě zjistíme, že tato matice má determinant $|A| = 2 \cdot 2 - 3 \cdot 2 = 2 \cdot 2 + 3 \cdot 2 = 4$.

Jak bychom na to útočili Gaussovou? Nejprve bychom si vytvořili vlevo nahore jedničku dělením prvního řádku dvěma, což ovšem v \mathbb{Z}_6 nejde, tam umíme jen násobit. Protože $\gcd(2, 6) > 1$, nemáme inverzní prvek k 2 a tudíž pomocí násobení z dvojkou jedničku neuděláme. Z podobného důvodu si násobením nevyrobíme jedničku ani na začátku druhého řádku.

To ale zase takový problém není, pro nás je ted' hlavním cílem trojúhelníkový tvar, což zde uděláme relativně snadno, stačí přičíst první řádek vynásobený dvěma k řádku druhému a dostaneme $\begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}$, což vypadá nadějně.

Otázka ovšem zní, zda je tato operace korektní i nad \mathbb{Z}_n . Odpověď zní, že ano, přičítání násobku jednoho řádku k řádku jinému je i nad \mathbb{Z} či \mathbb{Z}_n zcela korektní úprava. Nová matice proto musí mít a evidentně má determinant 4. Co bychom ale dělali, kdyby v druhém řádku nebylo číslo, které lze takto vynulovat (tedy pokud by tam nebyl násobek dvou)? Začalo by to být zajímavé.

Máme ovšem ještě další oblíbenou operaci, tedy násobení řádku číslem. Co dostaneme, když v nové matici vynásobíme druhý řádek trojkou? Matici $\begin{pmatrix} 2 & 3 \\ 0 & 0 \end{pmatrix}$, která má determinant nulový, což jasné ukazuje, že to není korektní úprava.

Přemýšlivějšího čtenáře možná napadne, že problém trojky je v tom, že je dělitelem nuly, viz kapitola . Pokud násobíme řádky výhradně čísla, která jsou v daném \mathbb{Z}_n invertibilní, pak se problémům tohoto typu vyhneme, na druhou stranu si takovým pravidlem silně omezujeme možnosti výpočtu. Praktickým důsledkem je, že Gaussova eliminační metoda přestává být spolehlivým nástrojem pro výpočet determinantu, výpočet inverzní matice či řešení soustav rovnic.

△

Protože na Gaussově eliminaci závisí důkaz toho, že transponováním reálné matice se nezmění její hodnost, naskýtá se otázka, zda i nad \mathbb{Z}_n platí důležitá rovnost $\text{hod}(A^T) = \text{hod}(A)$. Ukáže se, že ne, což je docela problém.

Příklad 7c.c: Uvažujme matici $A = \begin{pmatrix} 1 & 1 & 29 \\ 0 & 2 & 3 \end{pmatrix}$ nad \mathbb{Z}_{30} .

Vzhledem k jejímu tvaru by se zdálo, že řádky jsou lineárně nezávislé, tudíž má hodnost 2. Pro jistotu to zkusíme pořádně podle definice: Jaká řešení má rovnost $\alpha(1, 1, 29) + \beta(0, 2, 3) = (0, 0, 0)$?

Vzniká tak soustava $\alpha = 0$, $\alpha + 2\beta = 0$ a $29\alpha + 3\beta = 0$ řešená v \mathbb{Z}_{30} . Z první rovnice dosadíme do druhých dvou a máme systém $2\beta = 0$ a $3\beta = 0$ v \mathbb{Z}_{30} . Rovnice $2\beta = 0$ má v \mathbb{Z}_{30} množinu řešení $\{0, 15\}$, rovnice $3\beta = 0$ má v \mathbb{Z}_{30} množinu řešení $\{0, 10, 20\}$ a celý systém má tedy jediné řešení $\beta = 0$. Řešená vektorová rovnice má tedy jen triviální řešení a vektory jsou proto lineárně nezávislé.

Ted' se podíváme na transponovanou matici $A^T = \begin{pmatrix} 1 & 0 \\ 1 & 2 \\ 29 & 3 \end{pmatrix}$. Podle definice je její hodnost rovna maximálnímu počtu lineárně nezávislých řádků. Ukážeme, že žádné dva nejsou lineárně nezávislé, protože z každé dvojice umíme vyrobit pomocí netriviální lineární kombinace nulový řádek:

$$15 \cdot (1, 0) + 15 \cdot (1, 2) = (0, 0), \quad 10 \cdot (1, 0) + 10 \cdot (29, 3) = (0, 0), \quad 6 \cdot (1, 2) + 6 \cdot (29, 3) = (0, 0).$$

To dokazuje, že maximální počet lineárně nezávislých řádků této transponované matice je 1, tedy $\text{hod}(A^T) = 1$.

△

Těmto nepříjemnostem se vyhneme, pokud se nám podaří pracovat v \mathbb{Z}_n , kde n je prvočíslo. Pak je totiž \mathbb{Z}_n těleso (viz kapitola), hlavně je každý nenulový prvek v \mathbb{Z}_n invertibilní a tudíž se zase můžeme odvolávat na své zkušenosti z práce nad \mathbb{R} , například nám bude fungovat Gaussovka. Inverzní matici tedy bude možné hledat převodem $(A|E_n)$ pomocí řádkových úprav na $(E_n|A^{-1})$, což je samozřejmě ta nejpraktičtější metoda.

Příklad 7c.d: Najdeme matici inverzní k $A = \begin{pmatrix} 2 & 3 \\ 4 & 0 \end{pmatrix}$ nad \mathbb{Z}_5 .

Je vůbec regulární? V \mathbb{Z}_5 počítáme $|A| = 2 \cdot 0 - 4 \cdot 3 = -12 = 3$, což je v \mathbb{Z}_5 invertibilní. Proto s důvěrou upravujeme, pamatujeme si ovšem, že můžeme jen násobit a přičítat kladné násobky řádků. Nejprve přičteme první řádek k druhému, pak řádky prohodíme a přičteme trojnásobek nového prvního řádku k řádku druhému.

$$\left(\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ 4 & 0 & 0 & 1 \end{array} \right) \sim \left(\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ 1 & 3 & 1 & 1 \end{array} \right) \sim \left(\begin{array}{cc|cc} 1 & 3 & 1 & 1 \\ 2 & 3 & 1 & 0 \end{array} \right) \sim \left(\begin{array}{cc|cc} 1 & 3 & 1 & 1 \\ 0 & 2 & 4 & 3 \end{array} \right).$$

Nakonec přičteme druhý řádek k prvnímu a poté jej vynásobíme číslem $3 = 2^{-1}$.

$$\left(\begin{array}{cc|cc} 1 & 3 & 1 & 1 \\ 0 & 2 & 4 & 3 \end{array} \right) \sim \left(\begin{array}{cc|cc} 1 & 0 & 0 & 4 \\ 0 & 2 & 4 & 3 \end{array} \right) \sim \left(\begin{array}{cc|cc} 1 & 0 & 0 & 4 \\ 0 & 1 & 2 & 4 \end{array} \right).$$

Zkouška:

$$\left(\begin{array}{cc} 2 & 3 \\ 4 & 0 \end{array} \right) \cdot \left(\begin{array}{cc} 0 & 4 \\ 2 & 4 \end{array} \right) = \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right).$$

Takže opravdu $A^{-1} = \begin{pmatrix} 0 & 4 \\ 2 & 4 \end{pmatrix}$.

△

Jednou ze zajímavých aplikací lineární algebry nad \mathbb{Z}_n jsou samoopravné kódy, o kterých se lze dočíst například ve skriptu kolegy Velebila zmíněném v úvodu.

7c.3 Polynomy nad \mathbb{R} a \mathbb{Z}

Toto je opakovací kapitola, která jen připomene, co všechno nám bude chybět, když se pak přesuneme k \mathbb{Z}_n .

Polynomy nad \mathbb{R} jsou výrazy typu $a_nx^n + \dots + a_1x + a_0$, jejichž koeficienty a_i jsou z \mathbb{R} , třeba $p = \sqrt{2}x^5 - \pi x + e$. Množina všech takovýchto polynomů se značí $\mathbb{R}[x]$. Pro polynomy máme pravidla pro sčítání a násobení číslem (třeba umíme spočítat $3p$ pro ten polynom výše) i pro násobení polynomů mezi sebou, například pro $p = x + 1$ a $q = x - 1$ dostáváme $p \cdot q = x^2 - 1$.

Každý polynom zároveň dává vzniknout funkci, tedy zobrazení $x \mapsto p(x)$. Čistě formálně jde o dvě různé věci, polynom a funkce z něj vznikající, ale zrovna u polynomů nad \mathbb{R} se to nějak neřeší, protože polynomy a z nich vznikající funkce jsou úzce svázány. Konkrétně, jedna ze základních vlastností polynomů nad reálnými čísly je, že formální polynomy (tedy výrazy typu $a_nx^n + \dots + a_0$) a funkce jimi definované si jednoznačně odpovídají: Pokud se dva polynomy rovnají svými hodnotami, pak musí mít i stejné koeficienty, tedy jde o stejný polynom. To je užitečné v mnoha aplikacích, například pokud víme, že pro jisté parametry a, b, c platí rovnice $ax^2 + bx + c = x^2 + 13x + 14$ pro všechna $x \in \mathbb{R}$, pak nutně musí být $a = 1$, $b = 13$ a $c = 14$.

U polynomů $a_nx^n + \dots + a_0$ umíme obecně zadefinovat stupeň polynomu jako největší koeficient i takový, aby $a_i \neq 0$, a pro polynomy reálné se stupeň chová velice rozumně. Reálné polynomy dokonce umíme i navzájem dělit se zbytkem a zbytek po dělení i částečný podíl jsou jednoznačné (viz Věta o dělení pro čísla 6a.6). Shrňeme si to základní ve větě.

Věta 7c.4.

Uvažujme polynomy p, q nad \mathbb{R} . Pak platí následující:

- (i) $\text{st}(pq) = \text{st}(p) + \text{st}(q)$.
- (ii) Existují jediné polynomy d a r takové, že $p = d \cdot q + r$ a $\text{st}(r) < \text{st}(q)$.
- (iii) Polynom p má nejvýše $\text{st}(p)$ kořenů v \mathbb{R} .
- (iv) $a \in \mathbb{R}$ je kořenem p právě tehdy, když polynom $x - a$ dělí p .

Pro polynomy můžeme definovat i dělitelnost naprostě stejným způsobem jako pro celá čísla, tedy $q | p$ pokud existuje polynom r tak, aby $p = qr$. Například polynom $q = x - 1$ dělí polynom $p = x^2 - x$, protože $p = q \cdot r$ pro volbu $r = x$.

Když máme dělitelnost, můžeme zkousit vymyslet pojem největší společný dělitel. Tady je ale problém v nejednoznačnosti rozkladu, kdy můžeme násobící konstanty dle libosti přesouvat mezi faktory, například takto:

$$4x^2 - 16 = (4x - 8)(x + 2) = (2x - 4)(2x + 4) = (x - 2)(4x + 8) = \left(\frac{1}{2}x - 1\right)(8x + 16) = \dots$$

Naštěstí to ale není až tak hrozné, protože v zásadě jsou jen faktory dva, $x - 2$ a $x + 2$, které chytře vynásobíme konstantami tak, aby to celkově vyšlo. Když tedy mluvíme u reálných polynomů o faktorech, tak tím ani tak nemyslíme konkrétní polynomy, jako spíš množiny, jeden „faktor“ je polynom $x - 2$ a všechny jeho nenulové násobky, druhý „faktor“ je $x + 2$ a všechny jeho nenulové násobky. Není to až zas tak velký problém, například $x - 2$ i všechny jeho nenulové násobky mají stejný stupeň, stejně kořeny a podobně, takže je to z hlediska vlastností množina stejných věcí. Samozřejmě pokud bychom chtěli tuto definici vybudovat pořádně a matematicky korektně, museli bychom se s tímto vypořádat, ale zde si jen tak povídáme.

Když pak u reálných polynomů hledáme největšího společného dělitele, tak přirozeně dostaneme množinu polynomů, které jsou zase jeden vzorový polynom a všechny jeho násobky nenulovými konstantami, což se dá vnímat jako rozumná odpověď. Existuje zajímavý způsob, jak vyjádřit, že jde vlastně o stejný polynom až na násobení konstantou. Platí, že všechny polynomy z této množiny se navzájem dělí.

Příklad: Pro polynomy $p = 8x^2 + 4x$ a $q = 8x^2 - 16x$ snadno najdeme společné dělitele stupně 1, jmenovitě x , $2x$, $4x$, ale třeba i $20x$ nebo $\sqrt{13}x$. Opravdu je $20x$ společným dělitelem? Ano, $p = 20x \cdot (\frac{2}{5}x + \frac{1}{5})$ a $q = 20x \cdot (\frac{2}{5}x - \frac{4}{5})$. Vidíme, že největší společný dělitel je množina polynomů ve tvaru ax pro $a \neq 0$ a opravdu libovolné dva z nich se navzájem dělí ve smyslu polynomů, třeba $4x$ dělí $6x$, protože $6x = \frac{3}{2}(4x)$ a $r = \frac{3}{2}$ je polynom a naopak $4x = \frac{2}{3}(6x)$ a $r = \frac{2}{3}$ je polynom.

Něco podobného už ostatně známe: Při hledání největšího společného dělitele čísel $a, b \in \mathbb{Z}$ jsme dostali jednoznačný výsledek jedině díky tomu, že jsme se omezili jen na kladná čísla. Kdybychom se takto neomezili, pak bychom vlastně měli dva největší společné dělitele, číslo $\gcd(a, b)$ a číslo $-\gcd(a, b)$, přičemž hned vidíme, že se navzájem dělí, takže jde o situaci podobnou jako u těch polynomů.

Není těžké ukázat, že rozšířený Euklidův algoritmus (který využívá jen dělitelnost se zbytkem) funguje i pro polynomy, jeho výsledkem je jeden z polynomů, které lze považovat za největší společný dělitel, a jeho vyjádření pomocí vstupních polynomů. Prostor $\mathbb{R}[x]$ se tedy chová velice civilizovaně.

Obdobně můžeme definovat $\mathbb{Z}[x]$ jako množinu všech polynomů $a_n x^n + \dots + a_0$, jejichž koeficienty jsou ze \mathbb{Z} , kupodivu pak věci fungují úplně stejně. Zase si odpovídají polynomy jako výrazy s funkcemi, přesně řečeno koeficienty takového výrazu jsou jednoznačně určeny hodnotami z něj vzniklé funkce. To se občas velice hodí a my to využijeme v kapitole v tzv. metodě neučitých koeficientů. Také bychom teď mohli opsat Větu výše, jen se změnou \mathbb{R} v \mathbb{Z} , a platila by, stejně jako je pro $\mathbb{Z}[x]$ pravdivá i poznámka za ní o dělitelnosti, Euklidovi atd. a můžeme se spolehnout na rozšířený Euklidův algoritmus. Máme i obdobný problém s jednoznačností rozkladu a \gcd .

Rozdíly mezi polynomy nad \mathbb{R} a \mathbb{Z} přesto jsou. Čtenáře asi hned napadne, že polynomy nad \mathbb{Z} nemívají tolik kořenů jako polynomy nad \mathbb{R} . Třeba polynom $p(x) = 2x - 1$ má nad \mathbb{R} kořen $x = \frac{1}{2}$, zatímco nad \mathbb{Z} žádný kořen nemá. To je ale spíš podružné.

Podstatnější jsou komplikace, které nastanou okolo dělitelnosti. Například pro polynomy $p = 8x^2 + 4x$ a $q = 8x^2 - 16x$ brány nad \mathbb{Z} dostaneme jako největší společný dělitel jen množinu $\{x, 2x, 4x\}$. Na rozdíl od reálného případu již neplatí, že by se všechny navzájem dělily, takže toto kritérium, jak poznat, že jde v zásadě o tentýž objekt, nelze použít pro práci nad \mathbb{Z} . Nicméně to hlavní funguje podobně, ani u polynomů nad \mathbb{Z} nečíhají nějaké zásadní závludnosti.

7c.5 Polynomy nad \mathbb{Z}_n

Obsah této části se dá shrnout velice snadno. Jakmile začneme pracovat s polynomy nad \mathbb{Z}_n , tak už se nedá spoléhat na nic, co jsme připomněli v předchozí části.

Začneme tím, že hodnoty polynomu už jej nemusíme jednoznačně určovat.

Příklad 7c.e: Polynom $p = x^3$ nad \mathbb{Z}_6 definuje tuto funkci: $p(0) = 0$, $p(1) = 1^3 = 1$, $p(2) = 2^3 \equiv 2 \pmod{6}$, $p(3) = 3^3 \equiv 3 \pmod{6}$, $p(4) = 4^3 \equiv 4 \pmod{6}$ a $p(5) = 5^3 \equiv 5 \pmod{6}$, což je přesně stejná funkce, jakou definuje polynom $q = x$. To tedy znamená, že dva zcela různé polynomy dávají stejnou funkci.

Praktický dopad je, že přestávají fungovat mnohé oblíbené metody na určování koeficientů. Například když nám někdo dodá informaci, že jistý celostátně hledaný polynom $ax^5 + bx^4 + cx^3 + dx^2 + ex + f$ splňuje rovnici $2x^3 + 2x = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$ pro všechna $x \in \mathbb{Z}_6$ (tedy jde o rovnost funkcí), tak už nemusí platit, že je to zrovna ten $2x^3 + 2x$, klidně to může být i polynom $4x$. Ověřte si, že ani ten není sám, vyhovují třeba i polynomy $4x$, $4x^3$, $4x^5$, $2x^5 + x^3 + x$ a mnoho dalších.

To je zásadní změna. Mnoho úloh, u kterých oprávněně čekáme jednoznačné řešení při práci nad \mathbb{R} , může mít po přechodu do světa $\mathbb{Z}_n[x]$ třeba i nekonečně mnoho řešení.

△

To je ovšem teprve začátek. Rozpadnou se zcela i naše představy o kořenech polynomů a jejich rozkladech.

Příklad 7c.f: Polynom $p = 2x^2 + 1$ nad \mathbb{Z}_6 dává funkci $p(0) = p(3) = 1$ a $p(1) = p(2) = p(4) = p(5) = 3$, takže tento polynom nemá kořeny. To zase není nic divného (nemá je ani nad \mathbb{R}), ale málokterý čtenář by asi čekal, že tento polynom lze přesto rozložit na lineární faktory! Ověřte si pro sebe roznásobením, že

$$2x^2 + 1 = (2x + 1)(4x + 1).$$

Rozkládat lze evidentně všelicos, čtenář by asi také nečekal třeba toto:

$$x = (3x + 4)(4x + 3) \text{ nad } \mathbb{Z}_6.$$

Zde je zajímavé, že $x = 0$ je evidentně kořenem polynomu nalevo, ale není to kořenem ani jednoho z faktorů napravo. Také to ukazuje, že rovnost $\text{st}(pq) = \text{st}(p) + \text{st}(q)$ evidentně neplatí. Extrémní příklad: $3x(2x + 4) = 0$, součinem dvou polynomů stupně 1 dostaneme polynom stupně $-\infty$.

Poslední podivnosti: Uvažujme polynom $p = x^2 + x$ nad \mathbb{Z}_6 . Přesvědčte se, že čísla $x = 0, 2, 3, 5$ jsou kořeny tohoto polynomu, je tedy více kořenů, než je stupeň polynomu. Extrémní příklad v tomto směru: polynom $p = 2x^3 + 4x$ nad \mathbb{Z}_6 je jako funkce roven identicky nule, tedy všechna čísla ze \mathbb{Z}_6 jsou jeho kořeny!

△

Podíváme-li se do Věty o vlastnostech reálných polynomů, tak nám tento příklad ukázal, že při práci nad \mathbb{Z}_n přestávají obecně platit tvrzení (i) a (iii). Teď si pokazíme i (ii).

Příklad 7c.g: Zkusíme vydelit se zbytkem polynom $p = x$ polynomem $q = 3x + 4$ nad \mathbb{Z}_6 . V předchozím příkladě jsme už jeden rozklad viděli:

$$x = (4x + 3)(3x + 4) + 0.$$

Dá se ovšem zkusit třeba toto:

$$x = 2 \cdot (3x + 4) + 4.$$

Nemáme tedy jednoznačnost, není proto definován ani zbytek po dělení p polynomem q . Mimochodem pokus o zbytek jednou vyšel 0 a podruhé nenulový, je tedy p dělitelné polynomem q ? Zde naštěstí zmatek nevzniká, v definici se říká, že q dělí p , pokud lze vyjádřit p jako násobek q , což zde lze a víc už definice neřeší. Takže q dělí p .

△

Tento příklad tedy ukazuje, že pojem dělitelnosti se vybudovat dá, ale rozumné dělení se zbytkem nemáme. Pak už se ani nedá čekat rozumné fungování při hledání největšího společného dělíteli.

Příklad 7c.h: Jak vypadá největší společný dělitel polynomů $p = x^2 + x$ a $q = x^2 + 5$ nad \mathbb{Z}_6 ? Podívejme se na rozklady:

$$\begin{aligned} x^2 + x &= x(x + 1) = (x + 3)(x + 4), \\ x^2 + 5x &= x(x + 5) = (x + 3)(x + 2). \end{aligned}$$

Podle prvních rozkladů je společným dělitelem polynom x , podle druhých rozkladů je to zase $x + 3$. Oba jsou to ale polynomy prvního stupně, o kterých rozhodně neplatí, že by jeden byl násobkem druhého, což je nepříjemná komplikace.

△

Pracovat s polynomy nad \mathbb{Z}_n je tedy dobrodružné.

Teď dobrá zpráva. Pokud je n prvočíslo, pak nám zase většina věcí bude fungovat, jak jsme zvyklí u $\mathbb{R}[x]$ (viz sekce).

Věta 7c.6.

Nechť n je prvočíslo, uvažujme polynomy p, q nad \mathbb{Z}_n . Pak platí následující:

- (i) $\text{st}(pq) = \text{st}(p) + \text{st}(q)$.
- (ii) Existují jediné polynomy d a r takové, že $p = d \cdot q + r$ a $\text{st}(r) < \text{st}(q)$.
- (iii) Polynom p má nejvýše $\text{st}(p)$ kořenů.
- (iv) $a \in \mathbb{Z}_n$ je kořenem p právě tehdy, když polynom $x + (-a)$ dělí p .

Díky (ii) nám bude fungovat Euklidův algoritmus, zkusíme si to.

Příklad 7c.i: Najdeme $\gcd(x^3 + 3, x^2 + 1)$ pro polynomy nad \mathbb{Z}_5 pomocí rozšířeného Euklidova algoritmu.

Jeho nedílnou součástí je dělení se zbytkem, proto si zde připomeneme, jak to funguje pro polynomy. Když budeme počítat $\gcd(x^3 + 3, x^2 + 1)$, prvním krokem bude vydelení $(x^3 + 3) : (x^2 + 1)$. Algoritmus:

1. Vydelíme nejvyšší mocniny, výsledek je x .

2. Najdeme zbytek jako $(x^3 + 3) - x \cdot (x^2 + 1) = 3 - x \equiv 3 + 4x \pmod{5}$.

3. Pokud má zbytek nižší stupeň než dělitel, algoritmus končí, viz zde $\text{st}(4x+3) < \text{st}(x^2+1)$. V opačném případě se vrátíme do kroku 1 s tím, že dělíme zbytek původním dělitelem.

Příklad neukázal, co bychom dělali, kdyby u vedoucích mocnin byly koeficienty. Pak bychom v kroku 1 čelili třeba tomuto: $(2x^5) : (3x^2)$. V takovém případě nejprve vydělíme mocniny, vyjde x^3 , a pak se postaráme o koeficienty, přičemž dělení převádíme na násobení inverzním prvkem modulo n . V tomto případě namísto $2 : 3$ počítáme $2 \cdot 3^{-1} = 2 \cdot 2 = 4$, použili jsme, že $3^{-1} = 2 \pmod{5}$, což jsme odhadli zkusemo. Výsledek je $(2x^5) : (3x^2) = 4x^3$.

Ted' už by nás nemělo v následujícím běhu Euklidova algoritmu nic překvapit.

a, b	q	A	B	Použito
$x^3 + 3$		1	0	
$x^2 + 1$	x	0	1	$(x^3 + 3) : (x^2 + 1) = x$, zbytek $4x + 3$
$4x + 3 \bullet$	$4x + 2$	$1 \bullet$	$-x = 4x \bullet$	$(x^2 + 1) : (4x + 3) = 4x + 2$, zb. 0
0				

Vychází nám $\gcd(x^3 + 3, x^2 + 1) = 4x + 3$, což snadno ověříme:

$$x^3 + 3 = (4x + 3)(4x^2 + 2x + 1), \quad x^2 + 1 = (4x + 3)(4x + 2).$$

Máme také Bezoutovu identitu $4x + 3 = 1 \cdot (x^3 + 3) + 4x \cdot (x^2 + 1)$, což souhlasí.

Zajímavá věc:

$$x^3 + 3 = (x + 2)(x^2 + 3x + 4), \quad x^2 + 1 = (x + 2)(x + 3).$$

Takže také $\gcd(x^3 + 3, x^2 + 1) = x + 2$. Potvrzuje se tedy naše očekávání, u polynomů dostáváme jako největší společný dělitel ne jeden, ale celou množinu polynomů. Ovšem tvrdili jsme, že pro prvočíslo $n = 5$ bychom měli mít podobnou situaci jako u \mathbb{R} , tedy všechny tyto polynomy by měly být násobky (konstantou) jednoho vzorového polynomu. Je opravdu náš nový kandidát $x + 2$ násobkem toho, který vysel z Euklida? Ano, $x + 2 = 4 \cdot (4x + 3)$.

Snadno se ověří, že libovolný nenulový násobek polynomu $x + 2$ je také společným dělitelem, protože například rozklad u prvního polynomu lze zapsat jako $x^3 + 3 = [a(x + 2)] \cdot [a^{-1}(x^2 + 3x + 4)]$. Dá se ukázat, že jiní společní dělitelé nejsou.

△

Psali jsme, že „většina bude fungovat“, takže je dobré ještě přidat varování, kde i pro n prvočíselné může být problém. Asi nejpříjemější je, že různé polynomy mohou pořád dávat stejné funkce, například nad \mathbb{Z}_2 oba polynomy 1 a $x^2 + x + 1$ dávají konstantní funkci $x \mapsto 1$. Konec konců, nemusíme ani složitě shánět příklady: Pro prvočíslo n platí malá Fermatova věta a tu lze číst i tak, že polynomy x a x^n dávají stejné funkce.

Musíme si také dávat pozor, abychom do práce v \mathbb{Z}_n nepřenášeli návyky z reálných polynomů, například automaticky považujeme polynomy typu $x^2 + 1$ za nerozložitelné a bez kořenů, ale nad \mathbb{Z}_5 má tento polynom kořeny $x = 2, 3$ a lze jej napsat jako $x^2 + 1 = (x + 2)(x + 3)$.

Tímto končíme stručnou exkurzi do podivuhodného světa polynomů nad \mathbb{Z}_n , který má mimochodem zásadní aplikace například při kódování.

8. Binární operace

Binární operace je jeden z nejobecnějších pojmu, které v této knize potkáme. Úkolem zkoumat abstraktní struktury z hlediska operací se zabývá obor algebra, tuto kapitolu lze tedy brát jako jemný úvod do algebry.

Podstata binárních operací je jednoduchá, jde o proces, kdy vezmeme dva objekty, zamícháme a dostaneme objekt jiný. Prohláklé příklady binárních operací jsou třeba sčítání a násobení čísel, dále jsme potkali skládání zobrazení, binární operací je také násobení matic nebo třeba procedura, kdy vezmeme dva řetězce písmen a vytvoříme z nich třetí tak, že bereme postupně písmena z obou střídavě a řadíme za sebe. Nemusíme se omezovat na matematiku, jako operace můžeme interpretovat i mnoho interakcí v reálném světě, třeba mísení barev.

Jak vyjádříme takovou operaci matematicky? Vždycky nám ze dvou objektů dá třetí, takže je to vlastně přiřazení či posílání. Třeba sčítání dvou čísel v podstatě funguje takto: $(x, y) \mapsto (x + y)$. Vhodným modelem tedy bude zobrazení.

! Definice.

Nechť M je množina. Pod pojmem **binární operace** na M rozumíme libovolné zobrazení $\circ: M \times M \mapsto M$.

Let M be a set. By a **binary operation** we mean any mapping $\circ: M \times M \mapsto M$.

Aby se nám věci lépe psaly, zavedeme speciální značení, takže pro prvky $x, y \in M$ budeme namísto $\circ(x, y)$ psát $x \circ y$. To ostatně známe, sčítání dvou čísel nepíšeme $+(1, 13) = 14$, ale $1 + 13 = 14$. Někdy používáme i jiné značky, třeba $x \diamond y$, $x * y$ a podobně.

Při práci s operací nám pomáhá, pokud známe rozličná pravidla. Ne každá operace ovšem splňuje věci, které máme rádi. S tím je spojena zásadní otázka, co je třeba k tomu, aby určité pravidlo platilo. Někdy to vyplývá ze speciální podstaty objektů, se kterými pracujeme, ale jindy za tím stojí obecnější principy. Dobrým příkladem je pravidlo, které při přechodu k inverzi převrací pořadí v operaci. Vzoreček $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$ jsme viděli pro skládání zobrazení, později se objevil u skládání relací a čtenář jej potkal v lineární algebře u násobení matic: $(AB)^{-1} = B^{-1}A^{-1}$. Dokonce i důkazy byly vlastně stejné ve všech třech případech. To silně naznačuje, že za platnost tohoto pravidla nevděčíme jakési maticovosti či zobrazeňovosti, ale že vyplývá z nějakého obecného principu. Pokud bychom tento princip dokázali identifikovat a zjistili, co přesně je k jeho platnosti potřebné, tak už platnost tohoto pravidla dostaneme automaticky ve všech situacích, která splňují dotyčné požadavky.

Právě zkoumání prapůvodních přičin různých chování je jednou z náplní matematiky. Vnese se tím pořádek do mnohdy nepřehledných situací, navíc je ekonomické dělat věci jednou, tam, kde na tom opravdu záleží, než stejnou věc dělat stejným způsobem znova a znova pro různé inkarnace téhož. Nevýhoda z hlediska začátečníka je, že se tak dostáváme ke zkoumání velice abstraktních struktur, u nichž se často nemůžeme odvolat na intuici. Jde tedy obvykle o dosti pokročilou látku.

Když matematici takto obecně přemýšleli o operacích a vlastnostech, které je porůznu zajímají, tak postupně zjistili, že mnoho případů „pěkného“ chování ve skutečnosti vyplývá jen z několika málo základních principů. Právě to je téma této kapitoly. Jako inspiraci si ukažme, jak se poznají ty v zásadě nejlepší operace.

! Definice.

Uvažujme binární operaci \circ na množině M . Řekneme, že dvojice (M, \circ) je **grupa**, jestliže splňuje následující axiomy:

(A1) asociativita:

pro všechna $x, y, z \in M$ platí $x \circ (y \circ z) = (x \circ y) \circ z$.

(A2) existence jednotkového prvku:

existuje $e \in M$ takové, že $x \circ e = e \circ x = x$ pro všechna $x \in M$.

(A3) existence inverzních prvků:

pro každé $x \in M$ existuje $y \in M$ takové, že $x \circ y = y \circ x = e$.

Consider a set A with a binary operation \circ on it. We say that the couple (A, \circ) is a **group** if it satisfies the following conditions:

(A1) associative law: $\forall x, y, z \in M: x \circ (y \circ z) = (x \circ y) \circ z$.

(A2) existence of an identity element: $\exists e \in M \forall x \in M: x \circ e = e \circ x = x$.

(A3) existence of inverse elements: $\forall x \in M \exists y \in M: x \circ y = y \circ x = e$.

Někteří autoři začínají ještě jedním axiomem:

(A0) pro všechna $x, y \in M$ platí, že $x \circ y \in M$. $(M$ je **uzavřená** na \circ)

Toto je ale už obsaženo v slovech „binární operace \circ “, takže je to zbytečné tam dávat. Na druhou stranu to nevadí, má to pedagogický efekt, protože to čtenáři připomene tuto důležitou vlastnost binárních operací. Když vymýslíme operaci, tak si musíme pohlídat, že její výsledky nemohou vyskočit z množiny, na které pracujeme. To se týká zejména případů, kdy už operaci máme, ale rozhodněme se s ní pracovat na nějaké menší množině.

Grupy jsou jedna z nejpopulárnějších obecných struktur v matematice. Pomáhají v mnoha aplikacích, například ve fyzikálních teoriích mikrosvěta i vesmíru, při lámání určitých typů šifer (viz prolomení německého kódování Enigma polským algebraikem před druhou světovou) nebo třeba při analýze Rubikovy kostky.

My se jim ale nebude věnovat hned, nejprve se pořádně podíváme na to, jaký dopad vlastně jednotlivé axiomy mají. Ne všechny operace jsou totiž natolik pěkné, aby vytvořily grupy, a pak je dobré vědět, co můžeme čekat, když operaci nějaký axiom chybí. Je navíc zajímavé vidět, jak postupným přibíráním axiomů dokážeme o dané operaci říct stále více, pochopíme tak lépe souvislosti.

Toto zkoumání také vyjasní některé nejasnosti v definici grupy, kterou jsme zde uvedli. Například v této chvíli nevíme, kolik identit a inverzních prvků vlastně máme, což způsobuje problémy v axiomu (A3). Jak se to vlastně s inverzním prvkem myslí, kdyby těch identit existovalo více?

Čtenář, který potřebuje jen povšechnou znalost a praktické aplikace grup, může první část přeskočit a jít rovnou na kapitolu 8b, konec konců to tak dělají i někteří autoři. Přesto bych kapitolu 8a doporučil, hlavně část o monoidech, která souvisí s kapitolou 7a.

Než se do toho dáme, poznamenejme, že existují i operace jiné než binární, třeba unární (vezmou objekt a něco s ním udělají, čtenář asi zná konjugaci pro komplexní čísla, vlastně každé zobrazení $A \rightarrow A$ je unární operace) nebo naopak obecně n -ární, například v geometrii se používá smíšený součin $\vec{u} \bullet (\vec{v} \times \vec{w})$, který pracuje se třemi objekty. Teď už se ale podívejme na operace binární.

8a. Pologrupy a monoidy

Začneme tím nejjednodušším.

Definice.

Pojmem **grupoid** označujeme každou dvojici (M, \circ) , kde M je množina a \circ je binární operace na M .

Tento pojem se příliš nepoužívá, už proto, že zde nepředpokládáme splnění žádných dalších podmínek, tudíž ani nemůžeme čekat, že bychom o takovéto struktuře mohli něco dokázat. I v anglické literatuře vyskytuje zřídka, říkají tomu **groupoid** či častěji **magma**.

Jedna podmínka tam vlastně je: že dotyčná operace je zobrazení do M . Tomu říkáme, že množina M je **uzavřená** na (vzhledem k) operaci \circ , čímž se některé případy přece jen vyřadí.

Grupoidy potkáváme pořád, například $(\mathbb{N}, +)$, $(\mathbb{Z}, +)$, $(\mathbb{Z}, -)$, (\mathbb{R}, \cdot) .

Grupoidem nebude $(\mathbb{N}, -)$, protože například nemůžeme v množině \mathbb{N} odečítat 13 – 23. Množina tedy není na odečítání uzavřená a nejde o binární operaci. Podobně není grupoidem (\mathbb{R}, \div) , protože neumíme dělit nulou. Ale můžeme si rozmyslet, že $(\mathbb{R} - \{0\}, \div)$ už grupoid je, vydělením dvou nenulových čísel zase získáme nenulové číslo.

Teď už je čas zkusit něco zajímavějšího.

Definice.

Nechť (M, \circ) je grupoid.

Řekneme, že operace \circ je **asociativní**, jestliže pro všechna $x, y, z \in M$ platí $x \circ (y \circ z) = (x \circ y) \circ z$.

Tím jsme zavedli pojem, teď vytvoříme strukturu.

Definice.

Nechť (M, \circ) je grupoid. Řekneme, že je to **pologrupa (semigroup)**, jestliže \circ je asociativní.

Mnoho autorů začíná touto definicí a grupoid prostě přeskočí. Vidíme, že jsme se omezili pouze na axiom (A1) z definice grupy. Než se podíváme, co se dá s pologrupami dělat, ukážeme si nějaké příklady.

Příklad 8a.a:

1) Klasickými pologrupami jsou $(\mathbb{N}, +)$, $(\mathbb{R}, +)$, vlastně můžeme se sčítáním použít libovolnou známou množinu čísel $(\mathbb{Z}, \mathbb{Q}, \mathbb{C})$, protože víme snad už od základní školy, že sčítání je asociativní.

Dalšími dobrými příklady je násobení s různými množinami jako (\mathbb{N}, \cdot) , (\mathbb{R}, \cdot) a podobně.

Ale $(\mathbb{Z}, -)$ pologrupa není. Sice je splněna uzavřenosť množiny vzhledem k odčítání, ale neplatí asociativní zákon: $(5 - 2) - 1 = 2$, zatímco $5 - (2 - 1) = 4$.

Ze stejného důvodu není pologrupou $(\mathbb{R} - \{0\}, \div)$. Operace odečítání a dělení se tedy do této teorie nevejdou. To ukazuje, že nejsou stejného druhu jako sčítání a odčítání. Proto se s nimi v teorii operací vůbec nepracuje, stejně brzy uvidíme, že je vlastně vůbec nepotřebujeme, protože se k nim dostaneme oklikou přes (dobře vychovaná) sčítání a násobení.

Zajímavá poznámka: V reálném světě paradoxně nepracujeme s reálnými čísly, ale jen určitou podmnožinou v závislosti na použité výpočetní technice. Například moje kalkulačka si pamatuje jen 12 míst z každého reálného čísla. To má nečekané dopady, abychom je lépe viděli, trochu zmenšíme rozsah. Nechť M je množina všech reálných čísel s 4 platnými číslíci. Obsahuje tedy čísla jako 145, -13.76 , 1492000 a podobně. Zato neobsahuje číslo 74367, to se v našem světě M automaticky převede na 74360.

Uvažujme prostor $(M, +)$, kde $+$ je teď sčítání „počítačové“, kdy se výsledky automaticky převádějí do M . Tvrdíme, takovéto sčítání není asociativní, tudíž $(M, +)$ není pologrupa. Dokážeme to snadno, porovnejte

$$(11110 + 7) + 7 = 11110 + 7 = 11110,$$

$$11110 + (7 + 7) = 11110 + 14 = 11120.$$

Pokud bychom při převodu výsledku operace do M neorezávali, ale zaokrouhlovali, stejně by to dopadlo jinak:

$$(11110 + 7) + 7 = 11120 + 7 = 11130,$$

$$11110 + (7 + 7) = 11110 + 14 = 11120.$$

Takže sčítání, teoreticky krásná asociativní operace, není v počítačích asociativní (a také není komutativní), což může při výpočtech způsobovat fatální chyby. Lidé, kteří se počítačovými výpočty zaobírají, na toto musí myslet.

2) Ve vícedimenzionálních prostorech umíme sčítat vektory (děláme to po souřadnicích), máme pologrupu $(\mathbb{R}^n, +)$ nebo třeba $(\mathbb{N}^n, +)$, protože sčítáním vektorů se souřadnicemi z \mathbb{N} zase dostaneme vektory se souřadnicemi z \mathbb{N} . Dokážeme to pro dvouozměrný případ.

Nechť $(u, v), (x, y) \in \mathbb{N}^2$. Pak $u, v, x, y \in \mathbb{N}$, proto $(u, v) + (x, y) = (u + x, v + y)$ má souřadnice $u + x \in \mathbb{N}$ a $v + y \in \mathbb{N}$. Množina \mathbb{N}^2 je tedy opravdu uzavřená na sčítání vektorů.

Teď si ještě přiberme $(s, t) \in \mathbb{N}^2$. Pak pomocí asociativity pro běžné sčítání dostaneme

$$\begin{aligned} [(s, t) + (u, v)] + (x, y) &= (s + u, t + v) + (x, y) = ([s + u] + x, [t + v] + y) \\ &= (s + [u + x], t + [v + y]) = (s, t) + (u + x, v + y) = (s, t) + [(u, v) + (x, y)]. \end{aligned}$$

Asociativita je potvrzena, množiny \mathbb{R}^n či \mathbb{Z}^n tvoří se sčítáním vektorů pologrupu. Čtenář si hravě rozmyslí, že totéž lze dokázati o $(\mathbb{Q}^n, +)$, $(\mathbb{Z}^n, +)$ atd.

Standardní násobení na vektorech neexistuje, tak si operaci spojenou s násobením vymyslíme. Necháme se inspirovat klasickým sčítáním a zavedeme násobení po souřadnicích, pro $(u_1, \dots, u_n), (x_1, \dots, x_n) \in \mathbb{R}^n$ definujeme $(u_1, \dots, u_n) \cdot (x_1, \dots, x_n) = (u_1 \cdot x_1, \dots, u_n \cdot x_n)$. Podobně jako u sčítání se snadno dokáže, že se vlastnosti obvyklého násobení čísel přenášejí i na tuto operaci, dostáváme tak pologrupu (\mathbb{R}^n, \cdot) či třeba (\mathbb{N}^n, \cdot) .

Operací lze na vektorech vymyslet spoustu, viz třeba cvičení 8a.1.

3) Další známá pologrupa je $(M_{m \times n}, +)$, sčítání reálných matic velikosti $m \times n$, podobně je pologrupou (M_n, \cdot) neboli násobení reálných čtvercových matic rádu n . Přinejmenším u sčítání by to nemělo překvapit, protože to funguje po jednotlivých prvcích matic (podobně jako u vektorů jsme sčítali po souřadnicích), takže není problém si matici představit přerovnanou do vektoru, vše funguje stejně.

4) Snadno také dokážeme, že z množiny P všech reálných polynomů dostaneme pologrupu operacemi sčítání i násobení polynomů. Ani to není překvapením, protože polynomy i operace s nimi lze reprezentovat pomocí vektorů čísel (koeficientů polynomu).

△

Příklad 8a.b:

Díky kapitole 7 máme další zajímavý případ, prostory $(\mathbb{Z}_n, +)$ a (\mathbb{Z}_n, \cdot) , kde $\mathbb{Z}_n = \{0, 1, \dots, n-1\}$. Ve Větě 7a.20 jsme dokázali, že sčítání i násobení na \mathbb{Z}_n je asociativní.

△

Příklad 8a.c: Nechť U je libovolná množina, uvažujme $M = P(U)$ (všechny podmnožiny U). Pak (M, \cap) i (M, \cup) jsou pologrupy. Pokud sjednotíme či pronikneme dvě podmnožiny U , pak je výsledná množina zase podmnožinou U , máme tedy uzavřenosť M vzhledem k těmto dvěma operacím. Asociativitu průniku i sjednocení dává Věta 2a.7 (vi). Pologrupu ale nedostaneme s operací množinového rozdílu, viz cvičení 2a.2 (iv).

△

Příklad 8a.d:

1) Zvolme nějakou množinu A . Nechť M je množina všech zobrazení z A do A . Pak je M spolu s operací skládání pologrupa, viz Věta 2b.1.

2) Nechť M je množina všech relací na A . Pak je M s operací skládání pologrupa, viz Fakt 3a.3.

3) Vezmeme-li interval I reálných čísel a uvažujeme množinu $F(I)$ všech reálných funkcí na I , pak jsou $(F(I), +)$ a $(F(I), \cdot)$ pologrupy.

Trochu víc práce dá dokázat, že pologrupa je i $(F(\mathbb{R}), \circ)$, kde \circ je operace skládání funkcí.

4) Ukážeme si ještě jednu trochu komplikovanější množinu funkcí. Označíme písmenem F množinu všech funkcí, které mají jako definiční obor nějakou podmnožinu reálných čísel a jdou do reálných čísel. Taková množina už nebude uzavřená ani na sčítání funkcí, protože co dostaneme při sčítání dvou funkcí, které mají disjunktní definiční obory? Vyřešíme to tak, že si zavedeme jako speciální funkci \emptyset , to je funkce, která má definiční obor \emptyset a neposílá nic nikam. Ta nám pak pomůže, kdykoliv se pokusíme o něco srádovního, třeba výsledkem $\sqrt{x} + \ln(-x)$ je právě funkce \emptyset .

Ted' už můžeme funkce z této množiny libovolně kombinovat. Protože se algebraické operace mezi funkcemi definují pomocí běžného sčítání a násobení, které jsou asociativní, tak snadno ukážeme, že $(F, +)$ i (F, \cdot) jsou pologrupy. U skládání se to dokáže také, takže i (F, \circ) je pologrupa.

△

Příklad 8a.e: Uvažujme $A = \mathbb{N}$ a operace $x \circ_l y = \text{lcm}(x, y)$, tj. nejmenší společný násobek, a $x \circ_g y = \text{gcd}(x, y)$, tj. největší společný dělitel.

Například $3 \circ_l 7 = 21$, $3 \circ_g 7 = 1$, $4 \circ_l 6 = 12$, $4 \circ_g 6 = 2$.

Uzavřenosť množiny \mathbb{N} na tuto operaci je zjevná, ale ještě potřebujeme dokázat asociativitu obou operací. To dá trochu práce, asi nejjednodušší je ukázat, že $(x \circ_l y) \circ_l z$ odpovídá nejmenšímu společnému násobku x, y, z stejně jako $x \circ_l (y \circ_l z)$, tudíž se musí rovnat. Podobně se dá vyšetřit druhá operace (viz cvičení 6a.19).

Máme tedy pologrupy (\mathbb{N}, \circ_l) a (\mathbb{N}, \circ_g) .

△

Příklad 8a.f: Ted' něco trochu exotičtějšího.

1) Uvažujme $A = \mathbb{R}$ a zadefinujeme operaci \circ_{13} takto: Pro reálná čísla x, y je $x \circ_{13} y$ to z čísel x, y , které je blíže číslu 13; pokud jsou obě stejně daleko, dá se přednost tomu levému.

Takže $14 \circ_{13} 27 = 27 \circ_{13} 14 = 14$, $10 \circ_{13} 17 = 17 \circ_{13} 10 = 10$, $12 \circ_{13} 14 = 12$ ale $14 \circ_{13} 12 = 14$ (takže tato operace není komutativní), $\sqrt{8} \circ_{13} (-\pi) = \sqrt{8}$.

Vidíme také, že $13 \circ_{13} x = x \circ_{13} 13 = 13$ pro všechna $x \in \mathbb{R}$. U této operace tedy hraje číslo 13 stejnou roli jako číslo 0 u násobení, ale nějaké závěry z toho dělat nebudeme, je to jen zajímavá náhoda.

Asociativita platí, ale dá práci ji dokázat, protože výsledek operace není dán vzorcem. Asi nejjednodušší cesta je si rozmyslet, že výsledkem $(x \circ_{13} y) \circ_{13} z$ je to z čísel x, y, z , které je nejblíže k číslu 13, v případě rovnosti vyhrává to nejvíce vlevo ve výrazu. Výraz $x \circ_{13} (y \circ_{13} z)$ dává totéž, musí se tedy rovnat. Vytvořili jsme tedy pologrupu.

2) Barvu očí určují mimo jiné dva geny, z nichž každý přichází ve dvou formách, H a b . Gen H je dominantní a jakmile se v té dvojici objeví, budou oči hnědé. Gen b je recessivní a převládne jen tehdy, když jsou takové ve dvojici oba, pak jsou oči modré. Můžeme to vyjádřit takto: $H \circ H = H$, $H \circ b = H$, $b \circ H = H$, $b \circ b = b$. Tím vznikla binární operace na množině $M = \{b, H\}$. Rozmyslíme si, že pro tuto operaci platí asociativní zákon.

Máme-li výrazy $(x \circ y) \circ z$ a $x \circ (y \circ z)$, tak jsou dvě možnosti. Jestliže je některý z x, y, z roven H , pak dominantní gen ovládne všechny operace a výsledky obou vzorců jsou H . V opačném případě jsou ve vzorcích samá b a výsledek bude u obou výrazů b .

Takže tento objekt (M, \circ) je pologrupa.

△

Operace na konečných množinách, které nelze definovat vzorcem, se spíš než výčtem možností definují tzv. **Cayleyho tabulkou**. Tabulka pro geny očí z příkladu 8a.f vypadá takto:

\circ	H	b
H	H	H
b	H	b

Operace zadané tabulkou jsou nepřijemné, protože chceme-li zjistit, zda platí nějaký zákon, tak musíme vyzkoušet všechny možnosti dosazení, což je docela dost. V případě asociativního volíme na tři pozice, takže u množiny s n prvky to znamená ověřit n^3 možností.

Příklad 8a.g: Uvažujte množinu s dvěma prvky, $M = \{a, b\}$. Vytvoříme na ní binární operaci následující tabulkou.

\circ	a	b
a	a	a
b	b	a

Dává to pologrupu? Musíme prozkoumat osm možností. Ukáže se, že tato operace asociativní není, protože $(b \circ b) \circ b = a \circ b = a$, zatímco $b \circ (b \circ b) = b \circ a = b$.

Nedostáváme tedy pologrupu.

△

Příklad 8a.h: Na množině \mathbb{N} definujme následující operaci: Pro $m, n \in \mathbb{N}$ je $m \circ n$ dáno jako největší prvočíslo, které dělí $m + n$. Například $23 + 37 = 60$, to dělí prvočísla 2, 3, 5, proto $23 \circ 37 = 5$. Evidentně je \mathbb{N} uzavřená na tuto operaci, ale (\mathbb{N}, \circ) není pologrupa, protože selhává asociativita. Například $(3 \circ 5) \circ 10 = 2 \circ 10 = 3$, zatímco $3 \circ (5 \circ 10) = 3 \circ 5 = 2$. Takže tuto operaci už tady znova nepotkáme.

Mimochodem, všimněte si, že je komutativní, $m \circ n = n \circ m$, ale nikterak jí to nepomůže. V algebře hraje komutativita spíš pomocnou roli.

△

K většině těchto příkladů se budeme opakováně vracet.

V příkladech jsme se setkali s pologrupami $(\mathbb{Z}, +)$ a $(\mathbb{N}, +)$. Pokud si to čtenář zkusil dokázat, tak zjistil, že důkaz asociativity pro případ \mathbb{N} je zcela stejný jako pro \mathbb{Z} , protože jde o stejnou operaci. Patrně jsme tedy dělali zbytečnou práci, selský rozum říká, že když operaci, pro kterou funguje určité pravidlo na množině, používáme na části té množiny, tak se dotyčné pravidlo nemůže pokazit.

Formálně, máme pologrupu a rádi bychom se omezili na nějakou její část, aniž bychom ztratili vlastnost být pologrupou. To je v matematice častý požadavek, který jsme už ostatně potkali u více pojmu (restrikcí funkce vznikla nová funkce, u uspořádaných množin jsme často ocenili možnost přejít k podmnožinám a zase to byla uspořádaná množina atd.).

Definice.

Nechť (M, \circ) je pologrupa a $N \neq \emptyset$ je podmnožinou M . Řekneme, že N je **podpologrupa** (M, \circ) , jestliže (N, \circ_N) je pologrupa, kde \circ_N je restrikce binární operace \circ na množinu $N \times N$.

V našem inspiračním příkladě bychom tedy řekli, že \mathbb{N} je podpologrupa $(\mathbb{Z}, +)$. Operaci u \mathbb{N} není třeba uvádět, protože se automaticky dědí ta původní z větší množiny, v tom je princip pojmu podpologrupy. My jsme si v definici zavedli speciální značení \circ_N , ale to je jen dočasně, abychom v důkazu níže jasně ukázali, kde se na operaci \circ díváme v rámci původní množiny a kde v rámci N .

Teď bychom se měli dostat k té úspore. Představme si, že uvažujeme podmnožinu N pologrupy (M, \circ) . Co by se mohlo pokazit, aby to nebyla podpologrupa? Zkoumáme podle definice, zda je (N, \circ_N) pologrupa, a již jsme zmínili intuitivní pocit, že výpočty by měly probíhat stejně v N jako v M , tudíž by asociativita měla platit. Takže jediná podmínka z definice pologrupy, která není zaručena, je uzavřenosť operace. Potvrďme to důkazem.

Věta 8a.1.

Nechť (M, \circ) je pologrupa a $N \neq \emptyset$ je podmnožinou M . Pak N je podpologrupa (M, \circ) právě tehdy, když je N uzavřená na \circ .

Důkaz (rutinní): 1) \implies : Jestliže je N podpologrupa, pak je (N, \circ_N) pologrupa, tudíž již z definice musí platit $x \circ_N y \in N$ pro všechna $x, y \in N$. Protože je \circ_N restrikcí \circ na N , tak pak máme i $x \circ y = x \circ_N y \in N$ pro všechna $x, y \in N$, tedy N je uzavřená vzhledem k \circ .

2) \impliedby : Předpokládejme, že $x \circ y \in N$ pro všechna $x, y \in N$. Pak má zobrazení $\circ: M \times M \mapsto M$ po restrikci na $N \times N$ hodnoty v N , tedy $\circ_N: N \times N \mapsto N$. Dvojice (N, \circ_N) je tedy grupoid. Zbývá ověřit asociativní zákon. Použijeme asociativitu \circ na M a faktu, že operace \circ_N coby restrikce splňuje $x \circ_N y = x \circ y$ pro $x, y \in N$.

Vezměme tedy libovolné $x, y, z \in N$. Pak

$$(x \circ_N y) \circ_N z = (x \circ y) \circ z = x \circ (y \circ z) = x \circ_N (y \circ_N z).$$

□

Takže při přechodu k podmnožinám stačí hlídat uzavřenosť, což je obvykle snadné. Často tak ze známého příkladu pologrupy vyrábíme další.

S podpologrupami už si rozumíme, tak přestaneme používat \circ_N , je to pořád stejná operace \circ .

Příklad 8a.i (pokračování 8a.a):

1) Vyjdeme-li z pozorování, že $(\mathbb{C}, +)$ a (\mathbb{C}, \cdot) jsou pologrupy, dostáváme prakticky okamžitě také pologrupy $(\mathbb{R}, +)$, $(\mathbb{N}, +)$, (\mathbb{Z}, \cdot) a další.

Zkusíme si zajímavější modifikace. Připomeňme, že když je X nějaká známá číselná množina, tak z ní značením X^+ vybereme jen kladná čísla a značením X^- jen záporná čísla, značením X_0 pak přidáme nulu.

Snadno si rozmyslíme, že třeba $(\mathbb{Z}^-, +)$, $(\mathbb{R}^+, +)$, $(\mathbb{Q}_0^+, +)$, (\mathbb{Q}^+, \cdot) , $(\mathbb{R} - \{0\}, \cdot)$ jsou pologrupy. Stačí například ukázat, že součet dvou záporných celých čísel je záporné celé číslo, a podle Věty to již dokazuje \mathbb{Z}^+ coby podpologrupu $(\mathbb{C}, +)$, ale také podpologrupu $(\mathbb{Z}, +)$, což je zase podpologrupa $(\mathbb{R}, +)$. Jak vidíte, můžeme dělat celé řetízky podpologrups.

Naopak (\mathbb{Z}^-, \cdot) pologrupou není, protože to sice je podmnožina pologrupy (\mathbb{Z}, \cdot) , ale není uzavřená na násobení, vynásobením dvou záporných čísel nedostaneme záporné číslo.

Jiný zajímavý způsob: Zvolme přirozené číslo n a definujme $n\mathbb{Z} = \{nk; k \in \mathbb{Z}\}$, například $2\mathbb{Z}$ je množina všech celých násobků 2 čili množina sudých čísel. Ověříme uzavřenosť takové množiny na obě známé operace:

Jestliže $x, y \in n\mathbb{Z}$, pak existují $k, l \in \mathbb{Z}$ takové, že $x = nk$ a $y = nl$. Proto $x + y = n(k + l)$ a $k + l \in \mathbb{Z}$, tudíž podle definice $n\mathbb{Z}$ platí $x + y \in n\mathbb{Z}$. Podobně $xy = n(nkl)$ a $nkl \in \mathbb{Z}$, tudíž $x \cdot y \in n\mathbb{Z}$.

Proto jsou $(n\mathbb{Z}, +)$ a $(n\mathbb{Z}, \cdot)$ pologrupy.

Ukážeme si ještě jednu pěknou pologrupu. Čtenář je jistě obeznámen s pravidly pro interakci znamének v součinech/podílech (dva míny jsou plus atd.). Můžeme zavést množinu $M = \{+, -\}$ a pravidla zachytíme tabulkou.

○	-	+
-	+	-
+	-	+

Vyhodnocením osmi možností se ukáže, že tato (M, \circ) je pologrupa.

Elegantnější přístup: namísto znamének pracujeme s číslami $+1$ a -1 a násobením. Množina $\{1, -1\}$ je uzavřená na násobení, proto je to podpologrupa (\mathbb{Z}, \cdot) a tudíž pologrupa.

2) I ve vícedimenzionálních prostorech teď můžeme utíkat do podmnožin s tím, že si musíme hlídat uzavřenosť. Třeba $(\mathbb{Z}^n, +)$, $(\mathbb{Q}^n, +)$, $((\mathbb{R}^+)^n, \cdot)$ nebo $(\mathbb{N}^n, +)$ jsou pologrupy.

Toto je ale docela nudné, zkuste se zamyslet hlouběji nad běžným vektorovým prostorem se sčítáním. Jak vlastně množiny uzavřené na sčítání vypadají? V tom nám pomůže lineární algebra, jsou to například všechny lineární podprostory. Takže v rovině budou takovými podmnožinami všechny přímky procházející počátkem, v třírozměrném prostoru zase všechny přímky či roviny procházející počátkem. Takto je tedy možné vytvářet zajímavé pologrupy, třeba $N = \{(s, t, 2s - t); s, t \in \mathbb{R}\}$ je rovina procházející počátkem v \mathbb{R}^3 , proto je $(N, +)$ pologrupa.

3) Uvažujme množinu M_n^R všech regulárních reálných matic $n \times n$, tedy čtvercových matic s nenulovým determinantem. Z lineární algebry víme, že součin dvou regulárních matic je regulární, tudíž je tato množina uzavřená na násobení a (M_n^R, \cdot) je pologrupa.

Tato množina už ale není uzavřená na sčítání. Se sčítáním můžeme zkousit jiné věci, třeba množinu všech matic, jejichž členy jsou kladné.

4) I u polynomů se dají vyrábět zajímavé podmnožiny uzavřené na běžné operace, například množina M všech polynomů, které mají 13 jako kořen, je uzavřená na násobení.

△

Příklad 8a.j (pokračování 8a.c): Máme množinu M všech podmnožin nějakého universa U . Tady už dá trochu práce vybírat množiny z M tak, aby byly uzavřené na sjednocení, popřípadě na průnik.

Pokud si například jako N zkuseme vzít všechny podmnožiny U se sudým počtem prvků, tak to nebude fungovat, je snadné si představit, že průnikem dvou chytře vybraných dvouprvkových možin vznikne jednoprvková a nemáme uzavřenosť vzhledem k průniku, sjednocením dvouprvkových množin zase může vzniknout i tříprvková.

Chytřejší nápad: Zvolme si pevně nějaký prvek $u \in U$, nechť N je množina všech podmnožin U , které obsahují u . Pak už jsou (N, \cap) i (N, \cup) pologrupy (rozmyslete si uzavřenosť, viz cvičení 8a.2).

△

Příklad 8a.k (pokračování 8a.d):

1) Použijeme-li Fakt 2b.10, tak zjistíme, že množina všech prostých zobrazení z A do A je s operací skládání pologrupa, podobně je uzavřená na skládání i množina všech zobrazení na a množina všech bijekcí.

2) U relací na A a operace skládání se zase z původní pologrupy můžeme omezit například na relace reflexivní (Věta 3c.4) nebo třeba na uspořádání či na ekvivalence.

3) U funkcí můžeme zkoušit vzít třeba jen množinu všech funkcí na I , které mají vždy kladné hodnoty, a bude již uzavřená na sčítání a násobení funkcí, dostaneme tak pologrupu.

Zajímavější trik: Vezměme množinu $C(I)$ všech spojitých funkcí na I a množinu $C^1(I)$ všech funkcí na I , které jsou tam spojité, mají první derivaci a i ta je spojitá na I . Věty z matematické analýzy zaručují, že tyto dvě množiny jsou uzavřené jak na sčítání funkcí, tak na jejich násobení, dostáváme tak zajímavé pologrupy $(C(I), +)$, $(C(I), \cdot)$, $(C^1(I), +)$ a $(C^1(I), \cdot)$.

Podobně se ukáže, že pologrupami jsou také $(C(\mathbb{R}), \circ)$ a $(C^1(\mathbb{R}), \circ)$, ani skládání funkcí nezakazí spojitost či diferencovatelnost.

△

Díky asociativitě umíme zavést operaci i pro více prvků a také mocninu.

Definice.

Nechť (M, \circ) je pologrupa.

(i) Pro libovolné $n \in \mathbb{N}$ a $x_1, \dots, x_n \in M$ definujeme $x_1 \circ \dots \circ x_n$ rekurzivním vzorcem

$$(0) x_1 = x_1;$$

$$(1) x_1 \circ \dots \circ x_n \circ x_{n+1} = (x_1 \circ \dots \circ x_n) \circ x_{n+1} \text{ pro } n \in \mathbb{N}.$$

(ii) Pro libovolné $n \in \mathbb{N}$ a $x \in M$ definujeme mocninu x jako

$$(0) x^1 = x;$$

$$(1) x^{n+1} = (x^n) \circ x \text{ pro } n \in \mathbb{N}.$$

Úmluva: Mocnina má prioritu před operací \circ , není-li závorkou řečeno jinak.

Díky asociativitě má tato definice smysl a můžeme vynechávat závorky v opakování operacích. Například rekurzivním použitím definice získáme

$$x_1 \circ x_2 \circ x_3 \circ x_4 = (x_1 \circ x_2 \circ x_3) \circ x_4 = ((x_1 \circ x_2) \circ x_3) \circ x_4$$

a teď už je operace \circ aplikována vždy jen na dvojici prvků.

Pro začátečníka je mocnina zrádná, protože se podobá běžné mocnině, ale v skutečnosti je to jen zkratka pro opakování operace z uvažované pologrupy. Pokud například pracujeme v pologrupě $(\mathbb{N}, +)$, pak výsledek výrazu 2^3 rozhodně není 8, ale je to $2 + 2 + 2 = 6$.

Mocnina je někdy velice užitečná, ale občas se chová zajímavě. V příkladě 8a.e jsme měli operaci \gcd , pro celé číslo x pak máme

$$x^3 = x \circ_g x \circ_g x = (x \circ_g x) \circ_g x = \gcd(x, x) \circ_g x = x \circ_g x = \gcd(x, x) = x.$$

Indukcí se snadno dokáže, že $x^n = x$ pro všechna $x \in \mathbb{N}$, takže vlastně mocnina nic nedělá.

Teď si dokážeme pravidlo, které čtenář dobře zná z násobení čísel. Je to tedy typický příklad toho, o čem jsme mluvili v úvodu kapitoly, toto pravidlo platí obecně pro všechny asociativní operace.

Fakt 8a.2.

Nechť (M, \circ) je pologrupa a $x \in M$. Pak pro všechna $m, n \in \mathbb{N}$ platí:

$$(i) x^m \circ x^n = x^{m+n},$$

$$(ii) (x^m)^n = x^{mn}.$$

Důkaz (poučný): (i): Důkaz povedeme indukcí na n . Zvolme tedy pevné $m \in \mathbb{N}$ a dokážeme pro všechna $n \in \mathbb{N}$ následující $V(n)$: $(x^m) \circ (x^n) = x^{m+n}$.

$$(0) (x^m) \circ (x^1) = (x^m) \circ x = x^{m+1} \text{ podle definice mocniny.}$$

(1) Pro nějaké (libovolné) $n \in \mathbb{N}$ předpokládejme, že $(x^m) \circ (x^n) = x^{m+n}$. Pak pomocí vzorců z definice mocniny, asociativního zákona a indukčního předpokladu dostaneme

$$(x^m) \circ (x^{n+1}) = (x^m) \circ (x^n \circ x) = (x^m \circ x^n) \circ x = (x^{m+n}) \circ x = x^{(m+n)+1} = x^{m+(n+1)}.$$

Vzhledem k prioritě mocniny byly některé závorky zbytečné, ale při práci v abstraktních prostorzech je lepší být opatrný. Všichni jsme totiž zvyklí pracovat s reálnými čísly a běžnými operacemi a mnohé věci děláme skoro bez

přemýšlení, ale u pologrup vůbec nemusí platit. Je proto důležité si každý krok pečlivě rozmyslet, proč vlastně platí.

(ii): Důkaz se nejlépe dělá indukcí na n , použije se i výsledek z (i). Necháme to jako cvičení 8a.3. \square

Abychom se dostali k dalším zajímavým vlastnostem, potřebujeme přidat další z axiomů.

!

Definice.

Nechť (M, \circ) je grupoid. Řekneme, že prvek $e \in M$ se nazývá **jednotkový prvek** či **identita (identity element)**, jestliže pro všechna $x \in M$ platí $e \circ x = x \circ e = x$.

Je docela zajímavé, že nemusíme o grupoidu M nic předpokládat a stejně už bude platit, že pokud nějaký jednotkový prvek v M existuje, tak je jediný.

!

Fakt 8a.3.

Nechť (M, \circ) je grupoid. Jestliže v M existují jednotkové prvky e, f , pak $e = f$.

Důkaz (poučný): Použijeme nejprve to, že f je jednotkový prvek, a pak to, že i e je jednotkový prvek. Dostaneme $e = e \circ f = f$. \square

Podle definice můžeme takový prvek hledat u jakékoliv operace, ale opravdu zajímavé to začne být u pologrup.

!

Definice.

Nechť (M, \circ) je grupoid. Řekneme, že je to **monoid**, jestliže je to pologrupa a má jednotkový prvek e . Značíme jej pak (M, \circ, e) .

Takže přidáváme, teď už se budeme dívat na struktury, které splňují axiomy (A1) a (A2) z definice grupy. Jestliže je dotyčná operace silně podobná sčítání a označená + či podobně, pak se často jednotkový prvek značí také n či 0 a nazývá se **neutrální prvek**. Tady to ale raději dělat nebude, aby nebyl zmatek.

Zase si projdeme naše příklady, dá se čekat, že prořídnou.

Příklad 8a.1 (pokračování 8a.a):

1) Asi nepřekvapí, že $(\mathbb{R}, +, 0)$ a $(\mathbb{R}, \cdot, 1)$ jsou monoidy, protože vždy platí $x + 0 = 0 + x = x$ a také $x \cdot 1 = 1 \cdot x = x$.

Teď už ale musíme být opatrní při přechodu k podpologrupám. Například $(\mathbb{N}, \cdot, 1)$ monoid je (obsahuje jednotkový prvek 1), zato $(\mathbb{N}, +)$ už monoid není, protože ztratil jednotkový prvek sčítání 0. Napravíme to například takto: $(\mathbb{N}_0, +, 0)$ je monoid.

Podobně vidíme, že zatímco $(n\mathbb{Z}, +, 0)$ monoidy jsou, protože obsahují jednotkový prvek pro sčítání, s násobením už to fungovat nebude, protože jestliže $n > 1$, tak už jednotkový prvek 1 pro násobení nebude v množině $n\mathbb{Z}$. Pak je struktura $(n\mathbb{Z}, \cdot)$ pologrupa, ale ne monoid.

2) Je snadné si rozmyslet, že $(\mathbb{R}^n, +, 0_n)$ je monoid, podobně je i $(\mathbb{R}^n, \circ, 1_n)$ monoid, kde 1_n používáme pro vektor složený ze samých jedniček. Ukážeme to pro $n = 2$: Pro libovolné $(u, v) \in \mathbb{R}^2$ platí

$$(u, v) \circ (1, 1) = (u \cdot 1, v \cdot 1) = (u, v), \quad (1, 1) \circ (u, v) = (1 \cdot u, 1 \cdot v) = (u, v).$$

3) Připomeňme množinu $M_{m \times n}$ reálných matic. Z lineární algebry víme, že nulová matice $m \times n$ (označíme ji $0_{m \times n}$) je jednotkový prvek pro $(M_{m \times n}, +)$, zatímco jednotková matice E_n je jednotkový prvek pro (M_n, \cdot) .

4) Konstantní polynom $p_0(x) = 0$ je jednotkovým prvkem pro $(P, +)$, zatímco konstantní polynom $p_1(x) = 1$ je jednotkovým prvkem pro (P, \cdot) .

△

K problematice ztráty jednotkového prvku se po příkladech vrátíme blíže.

Příklad 8a.m (pokračování 8a.b):

V kapitole 7 jsme ukázali (a používali), že $(\mathbb{Z}_n, +, 0)$ a $(\mathbb{Z}_n, \cdot, 1)$ jsou monoidy. Budou to naše oblíbené příklady. △

Příklad 8a.n (pokračování 8a.c): Nechť U je libovolná množina, víme, že $(P(U), \cap)$ a $(P(U), \cup)$ jsou pologrupy. Ukážeme, že oba jsou i monoidy. Pro první je jednotkovým prvkem množina U , neboť pro libovolnou podmnožinu $A \subseteq U$ platí $A \cap U = U \cap A = A$. Pro tu druhou je jednotkovým prvkem množina \emptyset .

△

Příklad 8a.o (pokračování 8a.d):

1) Zde uvažujeme množinu M všech zobrazení z A do A . Fakt 2b.2 říká, že (M, \circ, i_A) je monoid.

Protože je i_A prosté, na a bijekce, budou i množina všech prostých zobrazení, množina všech zobrazení na a množina všech bijekcí s operací skládání monoidy.

2) Podobně díky Faktu 3c.1 víme, že relace $\Delta(A)$ poslouží jako jednotkový prvek pro množinu M všech relací na A a skládání.

3) Konstantní funkce $f_0(x) = 0$ je jednotkovým prvkem pro funkce vzhledem ke sčítání, zatímco konstantní funkce $f_1(x) = 1$ je jednotkovým prvkem pro násobení funkcí. Tyto dvě funkce jsou samozřejmě i spojité a spojitě diferencovatelné, takže vidíme, že i struktury $(F(I), +, f_0)$, $(F(I), \cdot, f_1)$, $(C(I), +, f_0)$, $(C(I), \cdot, f_1)$, $(C^1(I), +, f_0)$ a $(C^1(I), \cdot, f_1)$ jsou monoidy.

Podobně je monoidem $(F, +, f_0)$ a (F, \cdot, f_1) .

Jak je tomu se skládáním funkcí? Snadno se rozmyslí, že lineární funkce $i_{\mathbb{R}}(x) = x$ vyhovuje požadavkům (viz Fakt 2b.2), je také spojita a spojitě diferencovatelná, takže $(F(\mathbb{R}), \circ, i_{\mathbb{R}})$, $(C(\mathbb{R}), \circ, i_{\mathbb{R}})$, $(C^1(\mathbb{R}), \circ, i_{\mathbb{R}})$ a $(F, \circ, i_{\mathbb{R}})$ jsou rovněž monoidy.

△

Příklad 8a.p (pokračování 8a.e): Zde jsme pracovali s operacemi $x \circ_l y = \text{lcm}(x, y)$ a $x \circ_g y = \text{gcd}(x, y)$ na \mathbb{N} .

Je snadné si rozmyslet, že $(\mathbb{N}, \circ_l, 1)$ je monoid, ale pro operaci největšího společného dělitele jednotkový prvek nenajdeme. Proč? Zkusíme-li libovolného kandidáta n , pak největším společným dělitelem čísel n a $x = 2n$ je n , tedy $n \circ_g x = n$ a neplatí $n \circ_g x = x$. Toto číslo n tedy jako jednotkový prvek sloužit nemůže.

△

Příklad 8a.q (pokračování 8a.f):

1) Zkusme najít jednotkový prvek e pro operaci „blíž ke 13“. Musí splňovat $e \circ_{13} x = x$, ale výsledek této operace je bližší z čísel e, x ke třináctce, což znamená, že aby to fungovalo, muselo by číslo e být dál od 13 než všechna čísla z $M = \mathbb{R}$. Takové číslo ale neexistuje, tento příklad tedy není monoid.

Zajímavá finta: Uvažujme nějaký konečný interval okolo 13, třeba interval $I = \langle 2, 23 \rangle$. Protože je výsledkem operace $x \circ_{13} y$ vždy jedno z čísel x, y , bude libovolná podmnožina \mathbb{R} (i ta naše) uzavřená na tuto operaci. To znamená, že pro libovolnou (neprázdnou) podmnožinu $N \subseteq \mathbb{R}$ je (N, \circ_{13}) pologrupa. Zkusme se teď podívat na tu naši I .

Nejvzdálenější bod této množiny od 13 je 2. To znamená, že libovolné jiné číslo x z I je blíže, proto podle definice $x \circ_{13} 2 = 2 \circ_{13} x = x$. No a pro $x = 2$ máme $2 \circ_{13} 2 = 2$. Vidíme tedy, že $(I, \circ_{13}, 2)$ je monoid.

Tento příklad ukazuje, že opravdu děláme velice abstraktní teorii, do které se vejde ledasjaká šílenost.

Zkuste si rozmyslet, že monoidem bude libovolná (N, \circ_{13}) taková, že N je nějaká omezená podmnožina \mathbb{R} , jejíž supremum a infimum nejsou stejně vzdáleny od 13, přičemž ten vzdálenější z nich musí být v N (bude dělat jednotkový prvek).

2) Interakce genů pro barvu očí obsahuje také rovnosti $b \circ H = H$, $H \circ b = H$ a $b \circ b = b$, což ukazuje, že $(\{b, H\}, \circ, b)$ je monoid.

△

Monoidy jsou docela populární v computer science, například v teorii konečných automatů nebo v teorii jazyků.

Příklad 8a.r: Připomeňme si příklad 5b.g, kde jsme definovali slova Σ^* nad abecedou Σ . Slova se spojují tak, že se prostě dají za sebe, operace se jmenuje konkatenace a značí $x * y$, popřípadě prostě xy . Například konkatenací slov „mate“ a „matika“ je „matematika“. Snadno se ukáže, že jde o asociativní operaci, prázdný řetězec λ pak slouží jako jednotkový prvek. Dostáváme tedy monoid.

△

Vraťme se k problému s přechodem k podmnožině. U pologrup stačilo vyžadovat uzavřenosť podmnožiny na operaci a již vznikla pologrupa. Jak to funguje u monoidů? Tam máme o problém víc, musíme se postarat o jednotkový prvek. Co když přejdeme na podmnožinu a jednotkový prvek vynecháme? Kupodivu to může být pořád monoid, pokud se mu podaří najít si nějaký vlastní jednotkový prvek. Jak je to možné? Přechodem na menší množinu se sníží nároky na jednotkový prvek v definici (musí obhospodařit méně prvků), tudíž je naděje, že bude fungovat i prvek, který by v původní množině někde selhal.

Příklad 8a.s (pokračování 8a.a 2): Uvažujme klasickou množinu $\mathbb{R}^2 = \{(x, y); x, y \in \mathbb{R}\}$ dvojrozměrných vektorů s naší operací násobení $(u, v) \circ (x, y) = (ux, vy)$. Víme, že $(\mathbb{R}^2, \circ, (1, 1))$ je monoid.

Ted' uvažujme podmnožinu $M = \{(x, 0); x \in \mathbb{R}\}$ (osa x). Hned vidíme, že je na operaci \circ uzavřená, protože

$$(u, 0) \circ (x, 0) = (u \cdot x, 0 \cdot 0) = (ux, 0) \in M.$$

Je to tedy pologrupa, ale původní jednotkový prvek $(1, 1)$ v ní není. Ted' se ovšem podívejme na toto:

$$(u, 0) \circ (1, 0) = (u \cdot 1, 0 \cdot 0) = (u, 0), \quad (1, 0) \circ (u, 0) = (1 \cdot u, 0 \cdot 0) = (u, 0).$$

Vidíme, že $(M, \circ, (1, 0))$ je monoid, s jiným jednotkovým prvkem, který ovšem v původním \mathbb{R}^2 nefungoval vždy správně a proto jím tam nebyl.

△

Je zajímavé si rozmyslet, že tím nedochází ke sporu s Faktem 8a.3, spíš ho to zajímavě ilustruje. Představme si, že máme monoid (M, \circ, e) a uvažujeme jeho podmnožinu N . Kdybychom v podmnožině N původní jednotkový prvek ponechali, tak už si to N svůj vlastní najít nemůže, protože tím by v N byly dva různé a byl by spor s dotyčným Faktem. Teprve tím, že původní e z N odebereme, mu otevříme dvírka k tomu, aby si našlo své vlastní f , to ale zase podle Faktu nemůže být jednotkovým prvkem v M .

V matematice je ovšem taková změna jednotkového prvku nepříjemná, protože vlastně jde o jinou strukturu, zpřetrhá se tím spojení s původním světem. Pak už se ani nedá čekat, že by se dědily původní vlastnosti, čímž nás to přestane zajímat. Proto budeme trvat v definici podmonoidu na tom, že i menší množina používá původní e .

Definice.

Nechť (M, \circ, e) je monoid a $N \neq \emptyset$ je podmnožina M . Řekneme, že N je **podmonoid** (M, \circ, e) , jestliže $e \in N$ a (N, \circ, e) je monoid.

Pomocí Věty 8a.1 dokážeme následující:

Věta 8a.4.

Nechť (M, \circ, e) je monoid a $N \neq \emptyset$ je podmnožina M . Pak je N podmonoid (M, \circ, e) právě tehdy, když $e \in N$ a N je uzavřená na \circ .

Důkaz (rutinní): 1) \implies : Jestliže je (N, \circ, e) podmonoid, pak $e \in N$ už z definice, navíc je (N, \circ) pologrupa a tudíž podle Věty 8a.3 musí být N uzavřená na \circ .

2) \impliedby : Protože je N uzavřená na \circ , je podle Věty 8a.3 (N, \circ) pologrupa. Zbývá ukázat, že prvek $e \in N$ je v ní jednotkový. To je ale jasné, protože prvky z N jsou i prvky z M a operace je stejná, tedy pro $x \in N$ platí $e \circ x = x \circ e = x$. □

Takže můžeme snadno vyrábět spousty monoidů z těch základních známých, když si dáme pozor, jaké podmnožiny bereme. Například \mathbb{N} je podpologrupa $(\mathbb{Z}, +, 0)$, ale není to podmonoid, protože jsme vyneschali nulový prvek. \mathbb{N}_0 už podmonoid je, tedy $(\mathbb{N}_0, +, 0)$ je monoid.

Ted' si uvedeme další pojem. Inspiraci najdeme u násobení, pro které je 1 jednotkový prvek. Ve výpočtech se často hodí, že k číslu x dokážeme (většinou) najít číslo $\frac{1}{x}$, jehož klíčovou vlastností je $x \cdot \frac{1}{x} = 1$.

!

Definice.

Nechť (M, \circ, e) je monoid. Řekneme, že prvek $x \in M$ je **invertibilní** (**invertible**), jestliže existuje $y \in M$ takové, že $x \circ y = y \circ x = e$.

Matematika hned zajímá, kolik takových y k invertibilnímu x najdeme.

!

Fakt 8a.5.

Nechť (M, \circ, e) je monoid, nechť $x \in M$ je invertibilní. Předpokládejme, že $y_1, y_2 \in M$ splňují $x \circ y_1 = y_1 \circ x = e$ a $x \circ y_2 = y_2 \circ x = e$. Pak $y_1 = y_2$.

Důkaz (poučný): Využijeme obě vlastnosti a také asociativitu.

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

Důkaz je hotov. □

Víme tedy, že takové prvky jsou jedinečné, proto pro ně můžeme zavést rozumné značení.

!

Definice.

Nechť (M, \circ, e) je monoid a $x \in M$ je invertibilní prvek. Pak se prvku $y \in M$ splňujícímu $x \circ y = y \circ x = e$ říká **inverzní prvek** k x a značí se x^{-1}

Abychom ušetřili závorky, dohodneme se, že inverze prvku má vždy **prioritu** před operací \circ , není-li závorkou řečeno jinak.

Pokud nějaký autor pracuje s operací, která se chová jako sčítání, značí ji $+$ či podobně a používá názvu neutrální prvek, pak většinou také inverzní prvek značí $(-x)$ a říká mu prvek opačný. Aby v tom nebyl zmatek, zde to dělat nebudejme (s výjimkou kapitoly 8c).

Než se podíváme na příklady, stojí za to se vrátit k předchozímu Faktu. Pozorné čtení důkazu ukáže, že vlastně stačí, aby jeden prvek fungoval zleva (byl „levou inverzí“) a druhý zprava (byl „pravou inverzí“), a už dostáváme, co potřebujeme.

Věta 8a.6.

Nechť (M, \circ, e) je monoid. Jestliže pro $x \in M$ existují prvky $y_1, y_2 \in M$ takové, že $y_1 \circ x = e$ a $x \circ y_2 = e$, pak je x invertibilní a $y_1 = y_2 = x^{-1}$.

Důkaz (poučný): Předpoklady této věty jsou přesně to, co je třeba, aby platil výpočet v předchozím důkazu. Máme tedy $y_1 = y_2$, nazvěme tento prvek y . Když jej pak ale dosadíme do předpokladů místo y_1 a y_2 , dostáváme $y \circ x = e = x \circ y$, tedy x je invertibilní a $x^{-1} = y$. □

Je důležité si všimnout, že „inverze z jedné strany“, třeba levá, ještě nic neznamená. Klidně se může stát, že pro $x \in M$ najdeme prvek y takový, že $y \circ x = e$, ale x je přitom neinvertibilní. Jeden takový příklad ukážeme níže v příkladu 8a.v.

Když ale již dopředu víme, že je prvek invertibilní, pak stačí ověřit jen jednu z rovností, abychom věděli, že jsme našli prvek k němu inverzní.

Fakt 8a.7.

Nechť (M, \circ, e) je monoid. Jestliže je $x \in M$ invertibilní a $y \in M$ splňuje $y \circ x = e$ nebo $x \circ y = e$, pak $y = x^{-1}$.

Důkaz (rutinní): Předpokládejme, že y splňuje $y \circ x = e$. Protože je x invertibilní, pak máme také prvek splňující $x \circ x^{-1} = e$, tudíž podle Věty 8a.6 platí $y = x^{-1}$. Druhá možnost se dokáže stejně. □

Kolik máme v monoidu invertibilních prvků? Vždy alespoň jeden.

Fakt 8a.8.

Nechť (M, \circ, e) je monoid. Pak e je invertibilní a $e^{-1} = e$.

Důkaz hned plyne z definice e a definice inverzního prvku. Hledáme x tak, aby $x \circ e = e \circ x = e$, prvek $x = e$ vyhovuje.

Ted' se již podívejme na příklady. Uvidíme, že některé monoidy v počtu invertibilních prvků s bídou přeskočí laťku nasazenou již tak nízko tímto faktem, zatímco v jiných jsou invertibilní prvky všechny.

! Příklad 8a.t (pokračování 8a.a):

1) V monoidu $(\mathbb{R}, +, 0)$ má každý prvek x svou inverzi $-x$, protože pro libovolné x platí $x + (-x) = (-x) + x = 0 = e$.

Naopak v monoidu $(\mathbb{R}_0^+, +, 0)$ je jediným invertibilním prvkem 0 s $-0 = 0$, protože jsme vyneschali záporná čísla.

V monoidu $(\mathbb{R}, \cdot, 1)$ jsou invertibilní všechny prvky $x \neq 0$ a jejich inverze je $x^{-1} = \frac{1}{x}$, neboť pro libovolné nenulové x platí $x \cdot \frac{1}{x} = \frac{1}{x} \cdot x = 1 = e$.

V monoidu $(\mathbb{N}, \cdot, 1)$ je jediný invertibilní prvek, a to povinný $x = 1$. Podobně v monoidu $(\mathbb{N}_0, +, 0)$ je jediný invertibilní prvek $x = 0$.

Zajímavější je monoid $(\mathbb{Z}, \cdot, 1)$, kde jsou invertibilní prvky dva, $x = 1$ a $x = -1$, pokaždé $x^{-1} = x$. Jiné tam nenajdeme, pro $x = 0$ máme pro libovolného kandidáta y výsledek $xy = 0$, takže inverze nejde. Pokud $x \neq 0, \pm 1$, pak $|x| \geq 2$, proto pro libovolné $y \in \mathbb{Z}$ máme buď $xy = 0$ pro $y = 0$, nebo máme $|xy| = |x| \cdot |y| \geq |x| \geq 2$, takže k jedničce se neostaneme.

2) I v monoidu $(\mathbb{R}^n, +, 0)$ má každý prvek x svou inverzi $-x$, protože pro libovolné x platí $x + (-x) = (-x) + x = 0_n = e$, zde se sčítají vektory po souřadnicích.

U násobení vektorů po souřadnicích dle očekávání vadí nuly, tentokrát stačí byt jediná nulová souřadnice, aby již vektor nebyl invertibilní. Pokud jsou ale všechny x_i různé od nuly, tak $(x_1, x_2, \dots, x_n)^{-1} = (\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_n})$.

3) V monoidu $(M_{m \times n}, +, 0_{m \times n})$ má každá matice A inverzní prvek $-A$.

V monoidu (M_n, \cdot, E_n) jsou invertibilní regulární matice, neboť pro ně $A \cdot A^{-1} = A^{-1} \cdot A = E_n$.

4) V prostoru polynomů $(P, +, p_0)$ má každý polynom p svou inverzi $-p$, zatímco v prostoru polynomů (P, \cdot, p_1) mají svou inverzi jen nenulové konstantní polynomy $p(x) = a$, pro ně $p^{-1}(x) = \frac{1}{a}$.

△

! Příklad 8a.u (pokračování 8a.b):

Z kapitoly 7 již víme, že v $(\mathbb{Z}_n, +)$ je to s invertibilními prvky jednoduché (jsou to všechny), zatímco v (\mathbb{Z}_N, \cdot) je to zajímavé (jsou to jen taková a , kde $\gcd(a, n) = 1$, a hledání a^{-1} dá trochu práce).

Jako příklad si ukážeme (\mathbb{Z}_{14}, \cdot) , na který se budeme později odkazovat.

\odot	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
2	0	2	4	6	8	10	12	0	2	4	6	8	10	12
3	0	3	6	9	12	1	4	7	10	13	2	5	8	11
4	0	4	8	12	2	6	10	0	4	8	12	2	6	10
5	0	5	10	1	6	11	2	7	12	3	8	13	4	9
6	0	6	12	4	10	2	8	0	6	12	4	10	2	8
7	0	7	0	7	0	7	0	7	0	7	0	7	0	7
8	0	8	2	10	4	12	6	0	8	2	10	4	12	6
9	0	9	4	13	8	3	12	7	2	11	6	1	10	5
10	0	10	6	2	12	8	4	0	10	6	2	12	8	4
11	0	11	8	5	2	13	10	7	4	1	12	9	6	3
12	0	12	10	8	6	4	2	0	12	10	8	6	4	2
13	0	13	12	11	10	9	8	7	6	5	4	3	2	1

Jednotkový prvek se pozná tak, že se v jeho řádku i sloupci zopakují po řadě všechny prvky (jako v záhlaví), což 1 splňuje.

Invertibilní prvky poznáme podle toho, že mají v řádku i sloupci někde jednotkový prvek, tedy jedničku. Pokud se podíváme na řádek takového prvku x , najdeme v něm jedničku a od ní se podíváme nahoru na hlavičku sloupce, najdeme tam x^{-1} .

Vidíme, že máme invertibilní prvky 1, kde $1^{-1} = 1$ (to ostatně plyne z Faktu 8a.8), dále 3, kde $3^{-1} = 5$ (kontrola: $3 \cdot 5 = 15 \equiv 1$ modulo 14), dále 5, kde je $5^{-1} = 3$, dále 9, kde $9^{-1} = 11$, také 11, kde $11^{-1} = 9$ a nakonec 13, kde $13^{-1} = 13$.

Všimněte si zajímavé věci. Začneme třeba s $x = 9$, inverzní prvek je 11, a když se podíváme na inverzi k němu, dostaneme zase 9, situace je pěkně symetrická. Funguje to tak u všech dvojic v tomto příkladě a také obecně, za chvíli to dokážeme.

Všimněte si také, že třeba v třetím řádku vidíme zajímavou věc: $2 \odot 7 = 0$. Na to nejsme zvyklí, aby se nám nenulové prvky násobily na nulu, ale jak vidíme, u monoidů (alespoň některých) si na něco takového zvyknout musíme. Existuje na to dokonce teorie, viz kapitola 8c.

△

! Příklad 8a.v (pokračování 8a.d):

1) V monoidu všech zobrazení z množiny A do A se skládáním jsou invertibilní bijekce, inverzní prvky odpovídají inverzním zobrazením.

Zde se na chvíli zastavíme, abychom ilustrovali nebezpečí testování inverze pouze z jedné strany, viz diskuse po Větě 8a.6. Uvažujme zobrazení T a S z \mathbb{N} do \mathbb{N} dané těmito vzorci: $T(n) = n + 1$ pro všechna $n \in \mathbb{N}$, zatímco $S(1) = 1$ a $S(n) = n - 1$ pro $n \geq 2$.

Pak pro každé $n \in \mathbb{N}$ platí $T(n) \geq 2$, proto $(S \circ T)(n) = S(T(n)) = T(n) - 1 = (n + 1) - 1 = n$. Máme tedy $S \circ T = i_{\mathbb{N}}$, ale přesto není S inverzním zobrazením k T , protože neplatí $T \circ S = i_{\mathbb{N}}$. Máme totiž $(T \circ S)(1) = T(S(1)) = T(1) = 2$.

2) U relací na A to není hned vidět, ale dá se rozmyslet, že i v monoidu všech relací na množině A jsou invertibilní právě ty relace, které jsou zároveň bijektivními zobrazeními.

3) V prostorech funkcí na I (všechny, spojité, spojitě diferencovatelné) jsou všechny funkce invertibilní vůči sčítání, přičemž $f^{-1} = -f$.

Vůči násobení jsou invertibilní jen funkce, které nejsou nikde na I nulové, pro ně $f^{-1} = \frac{1}{f}$.

V prostoru funkcí $F(\mathbb{R})$ se skládáním jsou invertibilní jen bijekce \mathbb{R} na \mathbb{R} , pro ně je inverzním prvkem inverzní funkce.

Méně náročné je to v prostoru funkcí F , kde jsou invertibilní všechny prosté funkce. To je jeden z důvodů, proč jsme tuto komplikovanější množinu zaváděli, jsme totiž z matematické analýzy zvyklí invertovat prosté funkce. Například funkce arctg je prostá, má tedy v množině F inverzi, ale v množině $F(\mathbb{R})$ ji nemá, protože to není bijekce \mathbb{R} na \mathbb{R} .

△

Příklad 8a.w (pokračování 8a.e): Rozmysleli jsme si, že \mathbb{N} s operací \circ_l danou jako $\text{lcm}(x, y)$ je monoid s $e = 1$. To je také jediný invertibilní prvek, protože pro libovolné $n > 1$ a libovolné $x \in \mathbb{N}$ platí $\text{lcm}(n, x) \geq n > 1$, tedy nedokážeme z n vyrobit operaci číslo 1.

△

Příklad 8a.x (pokračování 8a.f):

1) Připomeňme si operaci \circ_{13} určující číslo nejbližší k 13, kde jsme si rozmysleli, že na množině $I = \langle 2, 23 \rangle$ spolu $e = 2$ je to monoid.

Jediným invertibilním prvkem je zde 2. Proč? Vezměme si libovolné jiné číslo $x \in I$ a hledejme k němu inverzní prvek. Ať už zkusíme jakékoli y , tak je vždycky $x \circ_{13} y$ nejvýše tak daleko od 13, jako je x , tudíž to nikdy nemůže být 2, které je ze všech nejdále.

2) Připomeňme si tabulkou monoidu genů pro barvy očí, kde je jednotkovým prvkem recesivní gen b . Pak je b automaticky invertibilní, zatímco H díky své dominanci invertibilní není, nikdy z něj nedostaneme b .

△

Jaké jsou základní vlastnosti platící pro inverzní prvky?

!

Věta 8a.9.

Nechť (M, \circ, e) je monoid.

(i) Jestliže je $x \in M$ invertibilní, pak je také jeho inverze x^{-1} invertibilní a $(x^{-1})^{-1} = x$.

(ii) Jestliže jsou $x, y \in M$ invertibilní, pak je také $x \circ y$ invertibilní a $(x \circ y)^{-1} = y^{-1} \circ x^{-1}$.

Důkaz (rutinní): (i) necháme jako cvičení 8a.4.

(ii): Hledáme prvek z takový, že $(x \circ y) \circ z = e$ a $z \circ (x \circ y) = e$. Dosazením a pomocí asociativního zákona ukážeme, že prvek $y^{-1} \circ x^{-1}$ toto splňuje:

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

Druhá rovnost se dokáže stejně.

□

Kdybychom začali skriptum touto kapitolou, tak jsme Větu 2b.6 ani Větu 3a.4 nemuseli dokazovat, plynuly by okamžitě z tvrzení (ii).

Invertibilní prvky jsou velice užitečné při řešení rovnic.

!

Fakt 8a.10. (o krácení)

Nechť (M, \circ, e) je monoid. Nechť prvky $c, x, y \in M$ splňují rovnici $c \circ x = c \circ y$ nebo rovnici $x \circ c = y \circ c$. Jestliže je c invertibilní, pak $x = y$.

Důkaz (rutinní): Předpokládejme, že $c \circ x = c \circ y$. Pak

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

Druhý případ se dělá obdobně, viz cvičení 8a.5.

□

Prvky, které nejsou invertibilní, pak v rovnicích krátit nemůžeme, což přináší komplikace, jak jsme to již viděli třeba v kapitole 7b.

Pomocí krácení se snadno dokáže jeden užitečný způsob, jak poznat jednotkový prvek:

Lemma 8a.11.

Nechť (M, \circ, e) je monoid a $g \in M$ je invertibilní prvek. Jestliže $g^2 = g$, pak $g = e$.

Důkaz necháváme jako snadné cvičení 8a.6. Poznamenejme, že obecně z $g^2 = g$ nic neplyne, například v (\mathbb{Z}_6, \cdot) máme $3^2 = 3$ (neboť $3 \cdot 3 = 9 \equiv 3 \pmod{6}$), ale není to jednotkový prvek. Také tam ovšem 3 není invertibilní, proto se na něj naše Lemma nevztahuje.

Mocniny prvků jsme definovali pro pologrupy. V monoidech se tato definice dá rozšířit.

Definice.

Nechť (M, \circ, e) je monoid a $x \in M$. Definujeme **mocniny** následovně:

- pro $n \in \mathbb{N}$ definujeme x^n dle původní definice;
- pro $n = 0$ definujeme $x^0 = e$;
- jestliže je x invertibilní, pak pro $n \in \mathbb{N}$ definujeme $x^{-n} = (x^{-1})^n$.

Pořád platí, že je třeba být ostražitý. Například v monoidu $(\mathbb{Z}, +, 0)$ je operací sčítání, proto 2^{-1} je -2 a tudíž $2^{-3} = (2^{-1})^3 = (-2)^3 = (-2) + (-2) + (-2) = -6$. Stojí za zopakování, že x^n není běžná mocnina, ale symbol pro opakování operace \circ .

Je snadné ale pracné dokázat (musí se rozebrat všechny kombinace kladných, záporných a nulových exponentů), že vzorce pro mocninu stále platí:

Věta 8a.12.

Nechť (M, \circ, e) je monoid, nechť $x \in M$. Pak vzorce $(x^m) \circ (x^n) = x^{m+n}$ a $(x^m)^n = x^{mn}$ platí pro všechna $m, n \in \mathbb{N}_0$.

Jestliže je x invertibilní, pak tyto vzorce platí pro všechna $m, n \in \mathbb{Z}$.

Platí i další rozumné pravidlo.

Fakt 8a.13.

Nechť (M, \circ, e) je monoid, nechť $x \in M$ je invertibilní. Pak je invertibilní i x^n pro každé $n \in \mathbb{Z}$ a $(x^n)^{-1} = x^{-n}$.

Důkaz (rutinní): 1) Nejprve to dokážeme pro $n \in \mathbb{N}$. Důkaz povedeme indukcí.

(0) Evidentně $(x^1)^{-1} = x^{-1}$.

(1) Předpokládejme, že pro nějaké $n \in \mathbb{N}$ je x^n invertibilní a $(x^n)^{-1} = x^{-n}$. Pak je podle Věty 8a.9 (ii) invertibilní i $x^n \circ x = x^{n+1}$ a podle téže věty je

$$(x^{n+1})^{-1} = (x^n \circ x)^{-1} = x^{-1} \circ (x^n)^{-1} = x^{-1} \circ (x^{-n}) = x^{-n-1} = x^{-(n+1)}.$$

2) Pro $n = 0$ tvrzení evidentně platí, $(x^0)^{-1} = e^{-1} = e = e^{-0}$.

3) Jestliže je $n \in \mathbb{Z}$ a $n < 0$, pak $m = -n \in \mathbb{N}$. Připomeňme, že pro invertibilní prvek y se záporná mocnina definuje jako $(y^{-1})^{-n} = (y^{-1})^m$, budeme to dále aplikovat na x i x^{-1} .

Protože je x invertibilní, je podle Věty 8a.9 (i) invertibilní i x^{-1} , tudíž je podle části 1) tohoto důkazu invertibilní i $(x^{-1})^m = x^n$. Opět podle části 1) máme

$$(x^n)^{-1} = ((x^{-1})^m)^{-1} = (x^{-1})^{-m} = ((x^{-1})^{-1})^m = x^{-n}.$$

Důkaz je hotov. □

8a.14 Mocniny

Mocniny prvku jsou zajímavý objekt. Zvolme si pevně prvek g z monoidu (M, \circ, e) a začněme jej kombinovat samý se sebou, tedy uvažujme $g, g \circ g = g^2, g \circ g \circ g = g^3, g^4, g^5, \dots$. Co se může stát? Jsou dvě možnosti.

a) U některých prvků tak dostaneme nekonečnou posloupnost $\{g^n\}_{n=1}^{\infty}$ různých prvků množiny M .

Například prvek 2 v monoidu $(\mathbb{Z}, +, 0)$ dává $2^1 = 2, 2^2 = 2 + 2 = 4, 2^3 = 2 + 2 + 2 = 6$, snadno se nahlédne, že 2^n (ve smyslu mocniny monoidu) je číslo $2n$.

b) U jiných prvků se stane, že se nějaký prvek po čase zopakuje, tedy pro jisté n se najde menší k takové, že $g^n = g^k$. Pak se již nic nového nestane a pořád dokola se budou opakovat prvky $g^k, g^{k+1}, \dots, g^{n-1}$, protože $g^{n+1} = g^n \circ g = g^k \circ g = g^{k+1}, g^{n+2} = g^n \circ g^2 = g^k \circ g^2 = g^{k+2}$ a tak dále.

Příklad 8a.y: Zajímavé případy nabízí $(\mathbb{Z}_{12}, \cdot, 1)$.

Mocniny prvku $g = 3$ se zacyklí na začátek: $3^1 = 3, 3^2 = 3 \cdot 3 = 9, 3^3 = 9 \cdot 3 = 27 \bmod 12 = 3$, zacyklili jsme se, pak $3^4 = 3^3 \cdot 3 = 3 \cdot 3 = 9$, atd. Dostáváme řetízek $3 \mapsto 9 \mapsto 3 \mapsto 9 \mapsto \dots$

Mocniny prvku $g = 2$ se na začátek nevrátí: $2^1 = 2, 2^2 = 4, 2^3 = 8, 2^4 = 4, 2^5 = 8$ atd., tedy máme $2 \mapsto 4 \mapsto 8 \mapsto 4 \mapsto 8 \mapsto \dots$

To zacyklení může být extrémně krátké: $g = 4$ má mocniny $4 \mapsto 4 \mapsto 4 \mapsto \dots$. Lze k tomu dorazit i později: $g = 6$ nabízí mocniny $6 \mapsto 0 \mapsto 0 \mapsto 0 \mapsto \dots$

Pro trochu delší řetězec se podíváme do monoidu $(\mathbb{Z}_{14}, \cdot, 1)$. Volba $g = 3$ dává $3^1 = 3, 3^2 = 9, 3^3 = 13, 3^4 = 11, 3^5 = 5, 3^6 = 1, 3^7 = 3$, vidíme cyklus $3 \mapsto 9 \mapsto 13 \mapsto 11 \mapsto 5 \mapsto 1 \mapsto 3 \mapsto \dots$

△

Případy, kdy cyklení začne později, nejsou až tak přínosné, mnohem raději máme prvky, u kterých se mocnina vrátí na začátek ke g . Ukážeme, že pro invertibilní prvky je to zaručeno, příklad výše nicméně ukazuje, že to není podmínka nutná. Měli jsme $g = 3 = g^3$, ale 3 není invertibilní v \mathbb{Z}_{12} . Následující tvrzení tedy bude mít formu implikace.

Fakt 8a.15.

Nechť (M, \circ, e) je monoid, nechť $g \neq e$ je invertibilní prvek z M . Předpokládejme, že množina $A = \{n \in \mathbb{N}; \exists k \in \{1, 2, \dots, n-1\}: g^n = g^k\}$ je neprázdná, a označme její nejmenší prvek jako m . Pak $g^m = g$.

Důkaz (poučný): Protože $m \in A$, existuje $k \in \{1, 2, \dots, m-1\}$ takové, že $g^m = g^k$. Prvek g je invertibilní, tudíž můžeme podle Faktu 8a.13 najít také inverzní prvek g^{-k} k g^k . Pomocí tohoto prvku a rovnosti $g^m = g^k$ pak odvodíme, že

$$g^{m-k+1} = [g^m \circ g^{-k}] \circ g = [g^k \circ ((g^k)^{-1})] \circ g = e \circ g = g.$$

To znamená, že se mocnina g^{m-k+1} zacyklila k předchozí množině, a tedy $m - k + 1$ leží v množině A . Zároveň je m její nejmenší prvek, což je možné jen tehdy, když $k = 1$, tedy $g^m = g^k$ je vlastně $g^m = g$, přesně jak jsme potřebovali dokázat.

□

Všimněte si, že pak $g^{m-1} = g^m \circ g^{-1} = g \circ g^{-1} = e$. Invertibilní prvky, které se cykly, se tedy o mocninu dříve dostanou k e , jak ostatně ukázal případ $g = 3$ v $(\mathbb{Z}_{14}, \cdot, 1)$. To nám umožňuje udělat následující definici:

Definice.

Nechť (M, \circ, e) je monoid a $g \in M$ je invertibilní prvek. Definujeme **řád (order)** prvku g , značeno $\text{ord}(g)$, jako nejmenší prvek množiny $\{n \in \mathbb{N}; g^n = e\}$, jestliže je neprázdná, jinak definujeme $\text{ord}(g) = \infty$.

Řád se také někdy značí $|g|$, ale raději to nebudeme používat, ať se to neplete s absolutní hodnotou. Snadno se nahlédne, že v každém monoidu je $\text{ord}(e) = 1$. Je to jediný takový prvek, je to vidět přímo z definice, $\text{ord}(g) = 1$ znamená, že $g^1 = e$ neboli $g = e$. Z našich příkladů vidíme, že v monoidu $(\mathbb{Z}, +, 0)$ je $\text{ord}(2) = \infty$ a v $(\mathbb{Z}_{14}, \cdot, 1)$ je $\text{ord}(3) = 6$.

Tedě jsme připraveni na jednu užitečnou větu.

Věta 8a.16.

Nechť (M, \circ, e) je monoid a $g \in M$ je invertibilní prvek takový, že $\text{ord}(g) < \infty$. Pak platí následující:

- (i) Pro $n \in \mathbb{Z}$ platí $g^n = e$ právě tehdy, když n je násobek $\text{ord}(g)$.
- (ii) Pro $m, n \in \mathbb{Z}$ platí $g^m = g^n$ právě tehdy, když $m - n$ je násobek $\text{ord}(g)$.

Důkaz (poučný): (i): Označme $o = \text{ord}(g)$. Podle Věty 6a.6 existují čísla $r \in \mathbb{N}_0$ a $k \in \mathbb{Z}$ taková, že $n = ok + r$ a $r < o$, tedy $r = n \bmod o$ je zbytek po dělení n číslem o . Pak máme $e = g^n = g^{ok+r} = (g^o)^k \circ g^r = e^k \circ g^r = g^r$. Takže $g^r = e$, ale zároveň je o nejmenší přirozené číslo s touto vlastností a $r < o$. Jediná možnost, jak se vyhnout sporu, je $r = 0$, tedy $n = ko$.

(ii): Jestliže $g^m = g^n$, pak $e = g^m \circ (g^m)^{-1} = g^m \circ (g^n)^{-1} = g^m \circ g^{-n} = g^{m-n}$, podle (i) je tedy $m - n$ násobkem řádu.

Nechť naopak $m - n = ko$ pro nějaké $k \in \mathbb{Z}$. Pak podle (i) $g^{m-n} = e$ a $g^m = g^{n+ko} = g^n \circ g^{ko} = g^n \circ e = g^n$.

□

Tedě uděláme další zajímavou věc, všechny mocniny včetně záporných sesypeme do jedné množiny.

!

Definice.

Nechť g je nějaký invertibilní prvek v monoidu (M, \circ, e) . Definujeme **monoid generovaný prvkem g** jako množinu

$$\langle g \rangle = \{g^n; n \in \mathbb{Z}\}.$$

Díky našim předchozím zkoumáním už dobře víme, jak takový generovaný monoid vypadá. Pro prvky nekonečného řádu je to množina $\{\dots, g^{-2}, g^{-1}, e, g, g^2, \dots\}$, kde jsou všechny mocniny různé. Pro prvky konečného řádu je to množina $\{e, g, g^2, \dots, g^{\text{ord}(g)-1}\}$.

Sice tomu říkáme „monoid generovaný“, ale ještě vlastně nevíme, že je to monoid. Hned to napravíme.

!

Věta 8a.17.

Nechť (M, \circ, e) je monoid a $g \in M$ nějaký invertibilní prvek. Pak $\langle g \rangle$ je podmonoid (M, \circ, e) .

Jestliže $\text{ord}(g) < \infty$, pak $\langle g \rangle = \{e, g, g^2, \dots, g^{\text{ord}(g)-1}\}$ a $|\langle g \rangle| = \text{ord}(g)$.

Důkaz (rutinní): Podle Věty 8a.4 musíme ukázat dvě věci. Jednak že $e \in \langle g \rangle$, ale to je snadné, $e = g^0$. Pak musíme ukázat uzavřenosť na \circ , ale to je také snadné. Nechť $x, y \in \langle g \rangle$. Pak existují $k, l \in \mathbb{Z}$ takové, že $x = g^k$ a $y = g^l$, proto $x \circ y = g^k \circ g^l = g^{k+l} \in \langle g \rangle$.

Druhá část: Označme $N = \{e, g, g^2, \dots, g^{\text{ord}(g)-1}\}$. Již z této definice platí $N \subseteq \{g^n; n \in \mathbb{N}\} = \langle g \rangle$. Teď ukážeme, že $\langle g \rangle \subseteq N$. Pro libovolné $n \in \mathbb{Z}$ vyjádříme $n = k \text{ord}(g) + r$, kde $k \in \mathbb{Z}$ a $r \in \mathbb{N}_0$ je zbytek splňující $r < \text{ord}(g)$. Pak je $n - r$ násobkem $\text{ord}(g)$, proto dle Věty 8a.16 (ii) je $g^n = g^r \in N$. Rovnost množin dokázána.

Nakonec ukážeme, že vypsané prvky N jsou navzájem různé. I to plyne z Věty 8a.16 (ii) a z toho, že čísla $0 < m, n < \text{ord}(g)$ splňují podmínu $m - n = k \text{ord}(g)$ jen tehdy, když $m = n$. Tato množina má tedy $\text{ord}(g)$ různých prvků. \square

Množinu $\langle g \rangle$ lze definovat i pro prvky, které nejsou invertibilní, ale pak už obecně nedostáváme podmonoid, takže to přestává být zajímavé. V příkladě výše jsme třeba viděli, že volba $g = 2$ v monoidu $(\mathbb{Z}_{12}, \cdot, 1)$ dává $\langle 2 \rangle = \{2, 4, 8\}$.

V poslední části této kapitoly se budeme věnovat zajímavému úkazu, když máme prvek g v monoidu (M, \circ, e) a necháme jej působit prostřednictvím operace \circ nikoliv na různé jednotlivé prvky, ale rovnou na celý kus množiny M . Inspirace pochází například z geometrie. Představme si nějaký útvar v rovině, třeba jistý kruh. Matematicky to znamená, že máme určitou množinu K vektorů. Teď si vezmeme ještě jiný vektor \vec{v} a přičteme jej ke všem vektorům množiny K . Geometricky to znamená, že jsme celý kruh posunuli, přičemž směr a velikost posunu jsou dány vektorem \vec{v} .

Matematicky bychom tento posunutý kruh nazvali $\vec{v} + K$ a je dán jako $\vec{v} + K = \{\vec{v} + \vec{u}; \vec{u} \in K\}$.

Tato operace je v geometrii často používána, například z přímek procházejících počátkem (které jsou docela výjimečné, například tvoří podprostory) dostáváme ostatní přímky právě posunem.

Stejnou fintu lze provést i obecně, přičemž ale představa zůstává stejná, máme množinu objektů N , a když na všechny z nich provedeme operaci \circ s účastí nějakého prvku g , tak jakoby množinu N posouváme.

Definice.

Nechť (M, \circ, e) je monoid. Pro libovolnou podmnožinu $N \subseteq M$ a prvek $c \in M$ definujeme cN jako $\{c \circ x; x \in N\}$.

Klíčová otázka je, co se s takovou množinou po posunu stane. Geometrické posouvání z našeho inspiračního příkladu zachovává tvar i velikost, ale obecně nemá smysl o tvaru mluvit, zato otázka velikosti (u množin tedy mohutnosti) může mít velký význam. Zajímat nás někdy mohou i další vlastnosti.

Příklad 8a.z: Uvažujme $(\mathbb{R}^2, \cdot, 1_2)$, tedy pracujeme s dvourozměrnými vektory, které mezi sebou násobíme po souřadnicích, jak jsme vymysleli v příkladě 8a.a 2).

Zvolme nějakou množinu N v \mathbb{R}^2 . Jak si představujeme $(2, -1)N$? Je to množina $\{(2, -1) \cdot (u, v); (u, v) \in N\} = \{(2u, -v); (u, v) \in N\}$. Geometricky to znamená, že jsme vzali útvar daný množinou N , přetočili okolo osy x , takže vznikl zrcadlový obraz, a pak jej roztahli dvakrát ve směru osy x (oběma směry od počátku). V zásadě se dá říct, že základní tvar zůstal do velké míry zachován (protažení udělá ze čtverce kosodélník či z kružnice elipsu, ale hlavní prvky se nezmění), zůstala i dimenze tohoto útvaru (pokud byla N dvourozměrná, třeba bramboroid, tak i $(2, -1)N$ je pořád dvourozměrná) a rovněž mohutnost množiny je stejná.

Teď se podíváme na $(0, 1)N = \{(0, y); (x, y) \in N\}$. Toto je takzvaná projekce na osu y , všechny body z množiny N jsme nahradili jejich nejbližšími sousedy z osy y . Pokud byla N trochu tučnější (třeba kruh), tak už výsledný útvar rozhodně nevypadá podobně, protože se celý nachází na ose y , z kruhu či čtverce by vznikla úsečka. Ztratili jsme tedy jednu dimenzi, útvar N se nám „posunem“ $(0, 1)N$ podstatně změnil.

Ke stejnemu jevu může docházet i v dalších monoidech. Uvažujme množinu $N = \{0, 1, 2, 3, 4, 5\}$ v našem oblíbeném $(\mathbb{Z}_{12}, \cdot, 1)$. Pak $5N = \{0, 5, 10, 3, 8, 1\}$, zde zůstala velikost zachována, zatímco $4N = \{0, 4, 8, 0, 4, 8\} = \{0, 4, 8\}$, množina se smrskla na polovinu.

△

Mizení části množin je při posunech velice nepříjemné, tak se podíváme, kdy se toho nemusíme bát.

Věta 8a.18.

Nechť (M, \circ, e) je monoid a $N \subseteq M$. Pro $c \in M$ definujme zobrazení $T: N \mapsto cN$ předpisem $T(x) = c \circ x$ pro $x \in N$. Jestliže je c invertibilní, pak je T bijekce.

Důkaz (poučný): T je na: Vezměme libovolné $y \in cN$. Pak pro nějaké $x \in N$ platí $y = c \circ x$. Máme $x \in N$ a $T(x) = c \circ x = y$, tedy T je na.

T je prosté: Nechť pro $x, y \in N$ platí $T(x) = T(y)$. Pak $c \circ x = c \circ y$, c je invertibilní a podle Faktu 8a.10 tedy $x = y$.

□

Tato věta nám říká dvě věci. Jedna je, že pro invertibilní c mají množiny N a cN stejnou mohutnost, to ještě použijeme. Stejně užitečný bude i druhý poznatek: Pokud máme v monoidu dva různé prvky a vynásobíme je invertibilním prvkem, pak musí být výsledky rovněž různé (to nám říká prostota toho T).

Je zajímavé aplikovat závěr Věty přímo na samotnou množinu M . Pokud vezmeme invertibilní prvek $c \in M$, pak je cM zase množina M . Podívejme se do příkladu 8a.b na tabulkou. Vidíme, že v řádcích u invertibilních prvků 1, 3, 5, 9, 11, 13 se vždy opravdu vyskytnou všechny prvky \mathbb{Z}_{14} , bez opakování, jen někdy v jiném pořadí.

V závěru této kapitoly si spojíme dva posledně zkoumané pojmy, tedy $\langle g \rangle$ a cN . Opět něco pro inspiraci: Budeme pracovat v rovině, což je teď pro nás množina vektorů s operací sčítání. Typickým podmonoidem je tam přímka procházející počátkem. Jednu takovou si zvolme, nazveme ji p . Už jsme si rozmysleli, že pro vektor \vec{v} dává $\vec{v}p$ přímku p posunutou o vektor \vec{v} , ta je s tou původní rovnoběžná.

Všimněme si dvou věcí. Jedna je, že pokud provedeme dva takové posuny, neboli pokud uvažujeme $\vec{u}p$ a $\vec{v}p$, tak budou výsledné přímky totožné, nebo naopak disjunktní. To by mělo čtenáři připomenout situaci s třídami ekvivalence. Druhá věc je, že celý prostor dokážeme získat tak, že sjednotíme různé posuny této konkrétní přímky. To už opravdu zavádí rozkladem a taky tam je, pod tímto povídáním je ve skutečnosti schovaná relace. Nás v této chvíli ale mnohem více zajímá, že to funguje úplně stejně i v obecných monoidech.

Věta 8a.19.

Nechť (M, \circ, e) je monoid a $g \in M$ invertibilní prvek. Pak platí následující:

- (i) Pro libovolné invertibilní $x \in M$ je zobrazení $T(g^k) = x \circ g^k$ bijekce z $\langle g \rangle$ na $x\langle g \rangle$. Proto také $|\langle g \rangle| = |x\langle g \rangle|$.
- (ii) Pro každé $x \in M$ platí: $x\langle g \rangle = \langle g \rangle$ právě tehdy, když $x \in \langle g \rangle$.
- (iii) Pro všechna $x, y \in M$ platí: $x\langle g \rangle = y\langle g \rangle$ právě tehdy, když $y = x \circ g^k$ pro nějaké $k \in \mathbb{Z}$.
- (iv) Množina $\{x\langle g \rangle\}_{x \in M}$ je rozklad množiny M .

Důkaz (poučný): (i) je vlastně Věta 8a.18.

(ii): Předpokládejme, že $x\langle g \rangle = \langle g \rangle$. Protože $e \in \langle g \rangle$, pak i $x = x \circ e \in x\langle g \rangle$ neboli $x \in \langle g \rangle$.

Nechť naopak $x \in \langle g \rangle$. Pak $x = g^n$ pro nějaké $n \in \mathbb{Z}$. Proto pro každé $x \circ g^k \in x\langle g \rangle$ máme $x \circ g^k = g^{n+k} \in \langle g \rangle$ a naopak pro každé $g^k \in \langle g \rangle$ máme $g^k = g^{n+k-n} = g^n \circ g^{k-n} = x \circ g^{k-n} \in x\langle g \rangle$.

(iii): Jestliže $y\langle g \rangle = x\langle g \rangle$, pak $y \in x\langle g \rangle$ a tedy $y = x \circ g^k$. Důkaz opačným směrem je podobný důkazu (ii).

(iv): Protože $x \in x\langle g \rangle$, evidentně platí $M = \bigcup_{x \in M} x\langle g \rangle$. Teď potřebujeme druhou podmínku.

Předpokládejme, že $x\langle g \rangle \cap y\langle g \rangle \neq \emptyset$. Pak najdeme prvek z takový, že $z = x \circ g^m$ a $z = y \circ g^n$ pro nějaká $m, n \in \mathbb{Z}$. Pak dostáváme

$$x = x \circ e = x \circ (g^m \circ g^{-m}) = (x \circ g^m) \circ g^{-m} = z \circ g^{-m} = (y \circ g^n) \circ g^{-m} = y \circ g^{n-m},$$

a podle (iii) je $x\langle g \rangle = y\langle g \rangle$.

□

Dostáváme docela zajímavý obrázek. Máme monoid (M, \circ, e) a vybereme si invertibilní prvek g . Pak $\langle g \rangle$ dává podmnožinu M , která si s operací \circ rozumí tak dobře, že sama tvoří monoid, je to tedy velice pěkný kousek množiny M . Množinu M pak dokážeme poskládat z posunů onoho pěkného kousku $\langle g \rangle$. Pokud se k posouvání používaly invertibilní prvky, tak dokonce všechny ty kousky mají stejný „tvar“ jako původní $\langle g \rangle$. Tato představa je tak žádoucí, že se omezíme jen na tyto situace, což nás přinutí jít do další kapitoly.

Cvičení

Cvičení 8a.1 (rutinní, poučné, zkouškové): Uvažujme následující grupoidy. Pro každý z nich rozhodněte, zda je \circ asociativní (tedy zda je M pologrupa). Pokud ano, zjistěte, zda existuje jednotkový prvek. Pokud půjde o monoid, vyšetřete, které jeho prvky jsou invertibilní, a najděte předpis pro inverzní prvky.

- (i) $M = \mathbb{Z}$, $x \circ y = x + y + 13$;
- (ii) $M = \mathbb{R}$, $x \circ y = \sqrt{x^2 + y^2}$;
- (iii) $M = \mathbb{R}$, $x \circ y = x^2 y$;
- (iv) $M = \mathbb{Q}$, $x \circ y = |xy|$;
- (v) $M = \mathbb{R} - \{0\}$, $x \circ y = \frac{1}{\frac{1}{x} + \frac{1}{y}}$ (takto se sčítají paralelní rezistory v obvodech);
- (vi) $M = \mathbb{R}^2$, $(u, v) \circ (x, y) = (u + x, v \cdot y)$;
- (vii) $M = \mathbb{Z}^2$, $(u, v) \circ (x, y) = (u + x, v \cdot y)$;
- (viii) $M = \mathbb{Z}^2$, $(u, v) \circ (x, y) = (u^x, v + y)$;
- (ix) $M = \mathbb{Q}^2$, $(u, v) \circ (x, y) = (uy, vx)$.

Cvičení 8a.2 (rutinní, poučné): Nechť U je množina, zvolme pevně prvek $u \in U$. Ukažte, že následující množiny N jsou uzavřené na operace \cap a \cup , viz příklad 8a.j.

- (i) $N = \{X \subseteq U; u \in X\}$;
- (ii) $N = \{X \subseteq U; u \notin X\}$.

Cvičení 8a.3 (dobré, poučné): Dokažte, že když je x prvek v pologrupě (M, \circ) , pak pro všechna $m, n \in \mathbb{N}$ platí $(x^m)^n = x^{mn}$ (viz Fakt 8a.2).

Cvičení 8a.4 (rutinní, poučné): Dokažte, že když je x invertibilní prvek v monoidu (M, \circ, e) , pak je invertibilní i x^{-1} a $(x^{-1})^{-1} = x$ (viz Věta 8a.9).

Cvičení 8a.5 (rutinní, poučné): Nechť (M, \circ, e) je monoid a prvky $c, x, y \in M$ splňují rovnici $x \circ c = y \circ c$. Dokažte, že jestliže je c invertibilní, pak $x = y$ (viz Fakt 8a.10).

Cvičení 8a.6 (rutinní, poučné): Nechť (M, \circ, e) je monoid a prvek $g \in M$ je invertibilní. Dokažte, že když $g^2 = g$, tak $g = e$.

Cvičení 8a.7 (rutinní): Dokažte, že když je g prvek v monoidu (M, \circ, e) , pak pro $m, n \in \mathbb{N}$ platí $g^m \circ g^n = g^n \circ g^m$.

Řešení:

8a.1: (i): $(x \circ y) \circ z = (x + y + 13) \circ z = (x + y + 13) + z + 13 = x + y + z + 26$ a $x \circ (y \circ z) = x \circ (y + z + 13) = x + (y + z + 13) + 13 = x + y + z + 26$. Rovno, je asociativní.

Jednotka: rovnice $x \circ e = x$ dává $x + e + 13 = x$, tedy $e = -13$. Zkouška: $x \circ e = x \circ (-13) = x + (-13) + 13 = x$, $e \circ x = (-13) \circ x = (-13) + x + 13 = x$, ano, máme monoid.

Inverzní prvky: Dáno $x \in M = \mathbb{Z}$, hledáme $x \circ y = e$, tedy $x + y + 13 = -13$. Řešení: $y = -26 - x$. Závěr: Všechny prvky jsou invertibilní, $x^{-1} = -x - 26$.

(ii): $(x \circ y) \circ z = \sqrt{x^2 + y^2} \circ z = \sqrt{x^2 + y^2 + z^2}$, $x \circ (y \circ z) = x \circ \sqrt{y^2 + z^2} = \sqrt{x^2 + y^2 + z^2}$, rovno. Rovnice $x \circ e = x$ dá po umocnění $e = 0$, ale to bohužel není řešením, protože pro $x < 0$ to neřeší původní rovnici. Je totiž jasné, že pro x záporné se k němu nemůžeme dostat pomocí odmocniny. Jednotkový prvek tedy není.

Kdybychom pozměnili zadání úlohy a uvažovali dotyčnou operaci na \mathbb{R}_0^+ , pak už by $e = 0$ bylo identitou. Jak by pak tomu bylo s inverzními prvky? Žádné $x > 0$ není invertibilní, protože pro libovolné y platí $x \circ y = \sqrt{x^2 + y^2} \geq |x| > 0$.

(iii): Není asociativní, $(x \circ y) \circ z = x^4 y^2 z$, zatímco $x \circ (y \circ z) = x^2 y^2 z$. Protipříklad: třeba $x = 2$, $y = 1$, $z = 1$, aby byl vidět rozdíl mezi těmito dvěma vzorci, pak $(2 \circ 1) \circ 1 = 4 \circ 1 = 16$, zatímco $2 \circ (1 \circ 1) = 2 \circ 1 = 4$.

(iv): Je asociativní, $(x \circ y) \circ z = |xy| \circ z = ||xy|z| = |xyz| = x \circ (y \circ z)$.

Nemá jednotkový prvek, pro $x = -1$ hledáme e takové, aby $x \circ e = x$, tedy $| - e | = -1$, což nelze. Není to monoid.

(v): Kupodivu asociativní, $(x \circ y) \circ z = \frac{1}{\frac{1}{x} + \frac{1}{y}} \circ z = \frac{1}{\frac{1}{x} + \frac{1}{y} + \frac{1}{z}} = x \circ (y \circ z)$.

Jednotkový prvek e by musel splňovat $x = \frac{1}{\frac{1}{x} + \frac{1}{e}}$, tedy $\frac{1}{e} = 0$, takový není. Neexistuje tedy rezistor s odporem, který by při paralelním zapojení nic nezměnil. Pokud bychom ale povolili $e = \infty$, už by to fungovalo, takže dát jako paralelní rezistor s nekonečným odporem (třeba rozstříhnutý drát) dá původní odpor.

(vi): $[(s, t) \circ (u, v)] \circ (x, y) = (s + u, tv) \circ (x, y) = (s + u + x, tvy)$ a $(s, t) \circ [(u, v) \circ (x, y)] = (s, t) \circ (u + x, vy) = (s + u + x, tvy)$. Rovno, je asociativní.

Identita: $(x, y) \circ (e_1, e_2) = (x, y)$ dává $x + e_1 = x$ a $y e_2 = y$ pro všechna x, y , vyhovuje $e_1 = 0$ a $e_2 = 1$. Kontrola: $(x, y) \circ (0, 1) = (0, 1) \circ (x, y) = (x, y)$.

Inverze: dáno (x, y) , chceme $(x + u, yv) = (0, 1)$. Pak $u = -x$ a $v = \frac{1}{y}$, pokud to jde. Závěr: pro $y \neq 0$ je (x, y) invertibilní a $(x, y)^{-1} = (-x, \frac{1}{y})$.

(vii): Asociativita a jednotkový prvek viz (vi).

Inverze: dáno (x, y) , chceme $(x + u, yv) = (0, 1)$. Pak $u = -x$ a $v = \frac{1}{y}$, pokud to jde. Závěr: pro $y = \pm 1$ je (x, y) invertibilní a $(x, y)^{-1} = (-x, y)$.

(viii): $[(s, t) \circ (u, v)] \circ (x, y) = (s^u, t + v) \circ (x, y) = ((s^u)^x, t + v + y)$ a $(s, t) \circ [(u, v) \circ (x, y)] = (s, t) \circ (u^x, v + y) = (s^{u^x}, t + v + y)$. Není to stejné, protože $(s^u)^x = s^{ux}$ se liší od $s^{(u^x)}$, takže není asociativní, protipříklad vyrobíme snadno pomocí té mocniny. Třeba $[(2, 0) \circ (2, 0)] \circ (3, 0) = (4, 0) \circ (3, 0) = (64, 0)$, zatímco $(2, 0) \circ [(2, 0) \circ (3, 0)] = (2, 0) \circ (8, 0) = (256, 0)$.

(ix): $[(s, t) \circ (u, v)] \circ (x, y) = (sv, tu) \circ (x, y) = (svy, tux)$ a $(s, t) \circ [(u, v) \circ (x, y)] = (s, t) \circ (uy, vx) = (svx, tuy)$. Není to stejné, takže není asociativní, protipříklad vyrobíme snadno, musíme odlišit x a y , třeba $[(1, 1) \circ (1, 1)] \circ (1, 2) = (1, 1) \circ (1, 2) = (2, 1)$, zatímco $(1, 1) \circ [(1, 1) \circ (1, 2)] = (1, 1) \circ (2, 1) = (1, 2)$.

8a.2: (i): $A, B \in N \implies A, B \subseteq U \wedge u \in A \wedge u \in B$. Pak $A \cap B \subseteq U \wedge u \in A \cap B$, tedy $A \cap B \in N$.

Pro cup a (ii) obdobně.

8a.3: Indukce na n , (1) $(x^m)^{n+1} = (x^m)^n \circ (x^m) = x^{mn} \circ x^m = x^{mn+m} = x^{m(n+1)}$.

8a.4: Ověřte, že $y = x$ splňuje podmínu $x^{-1} \circ y = y \circ x^{-1} = e$.

8a.5: $x = x \circ e = x \circ (c \circ c^{-1})$ tak dále.

8a.6: Napište si to jako $g \circ g = g \circ e$ a použijte vhodný Fakt.

8a.7: Využijte Fakt 8a.2.

8b. Grupy

Definice.

Monoid (M, \circ, e) se nazývá **grupa**, jestliže je každý jeho prvek invertibilní.

Tím jsme se vrátili k původní definici s jejími axiomy (A1) až (A3). Abychom si to oddělili i graficky, budeme grupy značit G . Nejprve se podívejme na příklady. Projdeme si známé monoidy z předchozí kapitoly a zamyslíme se, jak to v nich chodí s invertibilními prvky.

Příklad 8b.a (pokračování 8a.a):

1) Opačný prvek k číslu a vzhledem ke sčítání je $-a$. Je tedy jasné, že reálná čísla se sčítáním $(\mathbb{R}, +, 0)$ tvoří grupu, stejně jako vhodné podmnožiny, třeba $(\mathbb{Q}, +, 0)$ či $(\mathbb{Z}, +, 0)$. Naopak $(\mathbb{N}_0, +)$ grupu netvoří.

Grupami jsou také $(n\mathbb{Z}, +, 0)$ pro všechna $n \in \mathbb{N}$.

Žádné monoidy postavené na standardních číselních množinách a násobení nebudou grupy, máme problém s inverzí pro nulu, u celých čísel s inverzemi vůbec. Je tedy potřeba trochu množiny změnit. Podle Věty 8b.4 či selským rozumem dovodíme, že $(\mathbb{R} - \{0\}, \cdot, 1)$ je grupa, $(\mathbb{Q} - \{0\}, \cdot, 1)$ je grupa. To byl nejpřímočařejší způsob, zbavili jsme se problémového prvku. Jde to i jinak, například $(\mathbb{R}^+, \cdot, 1)$ je grupa či $(\{-1, 1\}, \cdot, 1)$ je grupa.

2) U vektorů je to podobné, $(\mathbb{R}^n, +, 0_n)$ je grupa, zato u „našeho násobení“ \cdot je problém s vektory s nulami. Vyhnut se jim umíme, například $((\mathbb{R} - \{0\})^n, \cdot, 1_n)$ či $((\mathbb{R}^+)^n, \cdot, 1_n)$ jsou grupy.

Rozmyslete si, že neprojde pokus $(\mathbb{R}^n - \{0_n\}, \cdot, 1_n)$; které vektory tam leží a nejsou invertibilní?

3) Matice se sčítáním tvoří grupu $(M_{m \times n}, +, 0_{m \times n})$, ale u násobení matic máme problém se singulárními maticemi. Vyhne se jim a dopadne to dobře, (M_n^R, \cdot, E_n) je grupa.

4) Reálné polynomy jsou s operací sčítání grupy. U násobení je problémových polynomů hodně. Jakmile $st(p) \geq 1$, tak už nutně $st(pq) \geq 1$ pro libovolný nenulový polynom q , takže takový p nemůže být invertibilní. Jediné invertibilní polynomy jsou tedy nenulové konstanty, $p(x) = a$, čímž ale vlastně dostáváme $(\mathbb{R} - \{0\}, \cdot)$, takže rozumnou grupu z reálných polynomů udělat nelze.

△

Příklad 8b.b (pokračování 8a.c): Vůči průniku a sjednocení na množinách inverzi nenajdeme s výjimkou (povinného) jednotkového prvku, tedy nejde o grupy.

△

! Příklad 8b.c (pokračování 8a.d):

V kapitole 7 jsme vlastně ukázali, že $(\mathbb{Z}_n, +, 0)$ jsou grupy pro všechna $n \in \mathbb{N}$, zatímco $(\mathbb{Z}_n, \cdot, 1)$ jsou grupy jen v případě, že n je prvočíslo.

△

! Příklad 8b.d (pokračování 8a.d):

1) Nechť A je množina. Množina všech zobrazení $A \rightarrow A$ s operací skládání obecně není grupa (výjimkou jsou extrémní případy, kdy je A jednoprvková).

Zato množina všech bijekcí na množině A je s operací skládání grupa.

2) Nechť A je množina. Množina všech relací na A s operací skládání obecně není grupa, vyjde to zase jen u jednoprvkové A .

3) Funkce na intervalu I jsou s operací sčítání grupy. U násobení je zase problém s nulami, u skládání s funkcemi, které nejsou bijekce či prosté.

△

Příklad 8b.e (pokračování 8a.e): Ani operace $\text{lcm}(x, y)$ nedá na \mathbb{N} grupu.

△

Příklad 8b.f (pokračování 8a.f):

1) Už jsme zjistili dříve, že v prostoru $(\langle 2, 23 \rangle, \circ_{13}, 2)$ je inverzní jen jednotkový prvek 2. Takže grupa to není ani náhodou.

2) Monoid genů pro barvu očí není grupa.

△

Docela nám to prořídlo, to se dalo čekat. Zkusíme něco nového.

Příklad 8b.g: Uvažujme uspořádanou množinu X , nechť M je množina všech permutací této množiny. Jako operaci bereme skládání, tedy jsou-li π_1, π_2 permutace, pak $\pi_2 \circ \pi_1$ je permutace, jejíž akce na množinu X funguje tak, že nejprve X zpermutujeme pomocí π_1 , pak výslednou množinu zpermutujeme pomocí π_2 .

Dá se ukázat, že tímto vznikne pologrupa (ověříme asociativní zákon). Permutace, která nechá množinu na místě, je jednotkovým prvkem množiny všech permutací na X , a každou permutaci množiny X je možné zase zamíchat na původní stav, takže máme i inverze. Množina permutací množiny X je tedy grupa. Právě věty o grupách permutací výrazně pomohly při zlomení německé šifry Enigma, což podle některých odhadů uspíšilo porážku Německa v druhé světové až od dva roky.

Formálně se permutace definuje jako bijekce množiny, pak už nám to vyplýne z příkladu 8a.d.

△

Pro grupy samozřejmě platí vše, co jsme si řekli o monoidech, ale s tím, že už nemusíme dělat předpoklady o inverzních prvcích, to už je zabudováno v grupě. Shrňme si to nejdůležitější, začneme základními vlastnostmi.

! Věta 8b.1.

Nechť (G, \circ, e) je grupa. Pak platí následující:

- (i) Jednotkový prvek e je jediný, stejně jako jsou jednoznačně určeny všechny inverzní prvky x^{-1} .
- (ii) (o krácení) Nechť $c \in G$. Jestliže pro prvky $x, y \in G$ platí $c \circ x = c \circ y$ nebo $x \circ c = y \circ c$, pak už $x = y$.
- (iii) Pro každé $x \in G$ existují mocniny x^n pro všechna $n \in \mathbb{Z}$, platí pro ně $x^m \circ x^n = x^{m+n}$ a $(x^m)^n = x^{mn}$ pro všechna $m, n \in \mathbb{Z}$.

Uvedeme teď pro úplnost ještě jeden axiom.

! Definice.

Nechť (M, \circ) je grupoid. Řekneme, že je **komutativní** či **Abelův** (**commutative** or **Abelian**), jestliže pro všechna $x, y \in M$ platí $x \circ y = y \circ x$.

Můžeme tedy uvažovat komutativní pologrupy, komutativní monoidy a komutativní grupy. Čtenář se možná divil, že se tato vlastnost neobjevila dříve, ale pravda je taková, že jsme ji nepotřebovali, na vlastnosti zkoumaného prostoru má komutativita řádově menší dopad než naše axiomy. Potvrdí se to i ve zbytku této části, komutativitu potřebovat nebude.

Její hlavní význam na této úrovni je praktický, mnohé výpočty (a důkazy) se v případě

komutativní operace silně zjednoduší a my jsme proto samozřejmě rádi, když ji máme. (Při hloubějším ponoru do teorie grup začne být komutativita významná, ale v této knize tak daleko nedojdeme.)

Velká část našich příkladů jsou komutativní monoidy či grupy, ale pro některé to neplatí, třeba pro skládání zobrazení, funkcí a permutací, známý je také problém s prohazováním u maticového násobení. Takové příklady máme také rádi, protože se zajímavěji chovají.

Existuje jedna situace, kdy máme komutativitu zaručenu, a to je podmonoid generovaný prvkem. Pravidla pro počítání s mocninami g^k zaručují, že spolu komutují, viz cvičení 8a.7. Pro invertibilní prvky g tak $\langle g \rangle$ vždy tvoří komutativní grupu.

Je čas na další teorii. Pojem podgrupy vytvoříme stejným způsobem jako předtím, tedy nechceme, abychom přechodem k podmnožině o něco přišli.

Definice.

Nechť (G, \circ, e) je grupa a $N \neq \emptyset$ je podmnožina G . Řekneme, že N je **podgrupou** G , jestliže je (N, \circ, e) je grupa.

Oproti podmonoidům je zde jedno velké zjednodušení, ta specifikace e u (N, \circ, e) není podstatná. Podmnožina grupy, která chce zůstat grupou, si totiž nemůže vybrat svůj vlastní jednotkový prvek a musí také mít stejné inverzní prvky.

Fakt 8b.2.

Nechť $N \neq \emptyset$ je podmnožina grupy (G, \circ, e) . Předpokládejme, že pro nějaký prvek e_N je (N, \circ, e_N) grupa. Pak platí následující:

- (i) $e_N = e$.
- (ii) Nechť $x \in N$. Jestliže nějaký prvek $y \in N$ splňuje $x \circ y = y \circ x = e_N$, pak je y inverzní prvek k x v grupě G .

Důkaz (poučný): (i): Prvek e_N splňuje $e_N = e_N \circ e_N$ v N , tedy i v G . Každý prvek v G je invertibilní, tudíž podle Lemma 8a.11 aplikovaného na G a e_N platí $e_N = e$.

(ii): Teď už víme, že $e_N = e$, takže se předpoklad dá napsat jako $x \circ y = y \circ x = e$ a z jednoznačnosti inverze v grupě G máme $y = x^{-1}$. □

To má zajímavý dopad na to, jak poznávat, zda nějaká podmnožina N grupy (G, \circ, e) je její podmnožinou. Víme už, že není třeba kontrolovat přítomnost e , ale učitě zůstává nutnost ověřit, že je množina N uzavřená vůči operaci \circ . Přibude ještě něco navíc, ještě je třeba zajistit, aby každý prvek z N také našel v této množině svou inverzi (která určitě existuje, když je G grupa, jen ty inverze nesmíme při výběru N vynechat). Dostaneme tak kritérium pro vlastnost býti podgrupou, následující věta to potvrdí a navíc ukáže, že se dva testy dají spojit do jednoho.

Věta 8b.3.

Nechť (G, \circ, e) je grupa a $N \neq \emptyset$ podmnožina G . Následující tvrzení jsou ekvivalentní:

- (i) N je podgrupa (G, \circ, e) ;
- (ii) pro všechna $x, y \in N$ platí $x \circ y \in N$ a pro všechna $x \in N$ platí $x^{-1} \in N$;
- (iii) pro všechna $x, y \in N$ platí $x \circ y^{-1} \in N$.

Důkaz (poučný): (i) \Rightarrow (iii): Předpokládáme, že (N, \circ, e) je grupa. Nechť $x, y \in N$. Pak musí mít y inverzní prvek v N , víme, že je to totéž jako inverze y^{-1} vůči G . Prvky x, y^{-1} jsou z N , tudíž i $x \circ y^{-1} \in N$ podle uzavřenosti N na \circ .

(iii) \Rightarrow (ii): Protože je $N \neq \emptyset$, můžeme si vzít nějaké $x \in N$. Když aplikujeme (iii) na $y = x$, dostaneme, že $e = x \circ x^{-1} \in N$. Proto můžeme aplikovat (iii) na prvky e a x a dostaneme, že $x^{-1} = e \circ x^{-1} \in N$.

Jestliže $x, y \in N$, tak jsme právě ukázali, že i $y^{-1} \in N$, tudíž podle (iii) zase $x \circ y = x \circ (y^{-1})^{-1} \in N$.

(ii) \Rightarrow (i): Vezměme $x \in N$, pak podle (ii) $x^{-1} \in N$, tedy zase podle (ii) $e = x \circ x^{-1} \in N$.

Protože N splňuje $e \in N$ a $x \circ y \in N$ pro $x, y \in N$, je podle Věty 8a.4 (N, \circ, e) podmonoid (G, \circ, e) , tedy monoid. Podle podmínky o x^{-1} má každý prvek $x \in N$ v množině svou inverzi, proto je (N, \circ, e) grupa. □

V praxi se nejčastěji používá (iii), ale i (ii) se občas hodí, třeba teď.

Věta 8b.4.

Nechť je (M, \circ, e) monoid. Označme $N = \{x \in M; x \text{ invertibilní}\}$. Pak je (N, \circ, e) grupa.

Důkaz (rutinní): Tvrzení plyne z Věty 8b.3 (ii) a Věty 8a.9. □

Díky této větě snadno vyrobíme rozličné grupy. Můžeme například u čtvercových matic s maticovým násobením vybrat ty regulární neboli invertibilní a podle Věty 8b.4 tak dostaneme grupu. Víme také, že mezi zobrazeními z A na A jsou invertibilní jen bijekce, takže automaticky dostáváme, že množina bijekcí na určité množině A spolu se skládáním je grupa.

V předchozí kapitole jsme si ukázali zajímavé vlastnosti podmonoidů generovaných invertibilními prvky. V grupách jsou invertibilní všechny, čímž se situace pročistí. Shrňme si to nejpodstatnější.

Věta 8b.5.

Nechť (G, \circ, e) je grupa a $g \in G$.

- (i) Buď $\text{ord}(g) = \infty$ a $\langle g \rangle = \{g^n; n \in \mathbb{Z}\}$, kde jsou všechny prvky různé,
nebo $\text{ord}(g) < \infty$ a $\langle g \rangle = \{e, g, g^2, \dots, g^{\text{ord}(g)-1}\}$, kde jsou všechny prvky různé.
- (ii) Pro každé $x \in G$ platí $|x\langle g \rangle| = |\langle g \rangle|$.
- (iii) Pro všechna $x \in G$ platí: $x \in x\langle g \rangle$.
- (iv) Pro všechna $x, y \in G$ platí: $x\langle g \rangle = y\langle g \rangle$ právě tehdy, když $x^{-1} \circ y \in \langle g \rangle$, a to právě tehdy když $y^{-1} \circ x \in \langle g \rangle$.
- (v) Množina $\{x\langle g \rangle\}_{x \in G}$ je rozklad množiny G .

Důkaz: (i) plyne z Věty 8a.17.

(ii) plyne z Věty 8a.19.

(iii): $x = x \circ e = x \circ g^0 \in x\langle g \rangle$.

(iv): Nejprve ukážeme ekvivalence posledních dvou tvrzení. Nechť $x^{-1} \circ y \in \langle g \rangle$. Pak pro nějaké $k \in \mathbb{Z}$ je $x^{-1} \circ y = g^k$, proto také $[x^{-1} \circ y]^{-1} = [g^k]^{-1}$. Podle rozličných pravidel pro inverzi pak $y^{-1} \circ x = g^{-k} \in \langle g \rangle$. Opačný směr je zjevně obdobný.

Nyní předpokládejme, že $x\langle g \rangle = y\langle g \rangle$. Protože $y \in y\langle g \rangle = x\langle g \rangle$, existuje $k \in \mathbb{Z}$ takové, že $y = x \circ g^k$. Pak ovšem $x^{-1} \circ y = x^{-1} \circ x \circ g^k = e \circ g^k = g^k \in \langle g \rangle$.

Zbývá ukázat, že z předpokladů $x^{-1} \circ y \in \langle g \rangle$ a $y^{-1} \circ x \in \langle g \rangle$ plyne rovnost oněch dvou množin.

Z první inkluze máme $x^{-1} \circ y = g^k$ pro nějaké $k \in \mathbb{Z}$. Odtud $y = x \circ g^k$, proto pro všechna $z \in y\langle g \rangle$ platí:

$$z = y \circ g^m = x \circ g^k \circ g^m = x \circ g^{k+m} \in x\langle g \rangle.$$

Takže $y\langle g \rangle \subseteq x\langle g \rangle$.

Obdobně z $y^{-1} \circ x \in \langle g \rangle$ odvodíme $x\langle g \rangle \subseteq y\langle g \rangle$ a rovnost je dokázána.

(v) Každý $x \in G$ leží v $x\langle g \rangle$, proto $G = \bigcup_{x \in G} x\langle g \rangle$. Zbývá ukázat, že dotyčné množiny jsou buď shodné, nebo disjunktní. Vezměme tedy libovolné $x, y \in G$.

Jestliže $x\langle g \rangle \cap y\langle g \rangle = \emptyset$, jsme hotovi. Předpokládejme tedy, že existuje $z \in x\langle g \rangle \cap y\langle g \rangle$. Pak musí existovat $k, m \in \mathbb{Z}$ takové, že $z = x \circ g^k$ a $z = y \circ g^m$. Pak ovšem $x \circ g^k = y \circ g^m$, proto $x^{-1} \circ x \circ g^k = x^{-1} \circ y \circ g^m$ nebo $g^k = x^{-1} \circ y \circ g^m$, následně $g^k \circ g^{-m} = x^{-1} \circ y$ nebo $x^{-1} \circ y = g^{k-m}$. Podle (iv) z toho již plyne $x\langle g \rangle = y\langle g \rangle$. □

Čtenář si možná pomyslel, že některé kroky byly zbytečně složité. Proč jsme prostě rovnici $x \circ g^k = y \circ g^m$ nevynásobili prvkem $x^{-1} \circ g^{-m}$ a pak nepokrátili? Protože by to nefungovalo. V grupách obecně nelze jen tak násobit a krátit v rovnicích. Ještě se k tomu vrátíme.

Podívejme se teď blíže na ten rozklad. U monoidů jsme komentovali, že si množinu G vlastně umíme poskládat z rozličných posunutí jedné $\langle g \rangle$. V případě grupy navíc víme, že všechny tyto posuny jsou stejné velikosti. U konečných grup z toho dostaneme zajímavý a užitečný výsledek.

Důsledek 8b.6.

Nechť (G, \circ, e) je grupa, kde G je konečná množina. Pak pro každé $g \in G$ platí, že $\text{ord}(g)$ dělí $|G|$.

Důkaz (poučný): Pokud je G konečná, tak nemůže být $\langle g \rangle$ nekonečná množina, tudíž $\text{ord}(g) < \infty$ a jde o čísla, se kterými jde dále počítat.

Podle (iv) se G rozloží na několik disjunktních množin, díky konečnosti G jich bude konečně mnoho, řekněme k , všechny mají podle (ii) stejnou mohutnost $\text{ord}(g)$. Dostáváme tedy $|G| = k \cdot \text{ord}(g)$. \square

Tento výsledek hrál klíčovou roli v kapitole 7, jmenovitě v důkazu Eulerovy věty 7a.22. Ukážeme si ještě jeden praktický dopad, viz Věta 8a.16.

Důsledek 8b.7.

Nechť (G, \circ, e) je grupa, kde G je konečná množina, označme $n = |G|$. Pak pro každé $g \in G$ platí $g^n = e$. Je-li navíc n prvočíslo a $g \neq e$, pak $g^k = e$ právě tehdy, když k je násobek n .

Příklad 8b.h (pokračování 8a.b): Víme, že $(\mathbb{Z}_{14}, \odot, 1)$ je jen monoid. Podle Věty 8b.4 bychom měli dostat grupu tak, že si vybereme invertibilní prvky, tedy $G = \{1, 3, 5, 9, 11, 13\}$. Její multiplikativní tabulku získáme tak, že z původní tabulky pro \mathbb{Z}_{14} vyškrťáme nepoužité řádky a sloupce.

\odot	1	3	5	9	11	13
1	1	3	5	9	11	13
3	3	9	1	13	5	11
5	5	1	11	3	13	9
9	9	13	3	11	1	5
11	11	5	13	1	9	11
13	13	11	9	5	3	1

Jaké řády dokážeme získat? Pro $g = 1$ máme evidentně $1^k = 1$ a tedy $\langle 1 \rangle = \{1\}$, proto $\text{ord}(1) = 1$, což dělí $|G| = 6$.

Řetízek mocnin pro $g = 3$ nám již dříve ukázal, že $\langle 3 \rangle = G$ a $\text{ord}(3) = 6$, což opravdu dělí $|G| = 6$.

Případ $g = 5$ dává $5^2 = 11$, $5^3 = 13$, $5^4 = 9$, $5^5 = 3$, $5^6 = 1$, zase $\langle 5 \rangle = G$ a $\text{ord}(5) = 6$.

Pro $g = 9$ máme $9^1 = 9$, $9^2 = 9 \odot 9 = 11$, $9^3 = 9^2 \odot 9 = 11 \odot 9 = 1$. Proto $\langle 9 \rangle = \{9, 11, 1\}$ a $\text{ord}(9) = 3$, což zase dělí $|G| = 6$.

Ověřte si sami, že $\langle 11 \rangle = \{1, 9, 11\}$ a $\text{ord}(11) = 3$, také $\langle 13 \rangle = \{1, 13\}$ a $\text{ord}(13) = 2$.

Každý dělitel čísla 6 našel svůj prvek, ale obecně to nemusí být pravda.

Pro zajímavost se ještě podíváme na $(\mathbb{Z}_4, \oplus, 0)$, což je grupa.

Zvolme $g = 3$, pak $3^1 = 3$, $3^2 = 3 \oplus 3 = 2$, $3^3 = 3 \oplus 3 \oplus 3 = 2 \oplus 3 = 1$, $3^4 = 3^2 \oplus 3^2 = 2 \oplus 2 = 0 = e$. Máme tedy $\langle 3 \rangle = \{3, 2, 1, 0\} = \mathbb{Z}_4$ a $\text{ord}(3) = 4$.

Teď zvolme $g = 2$, pak $2^1 = 2$, $2^2 = 2 \oplus 2 = 0 = e$. Máme tedy $\langle 2 \rangle = \{2, 0\}$ a $\text{ord}(2) = 2$.

\triangle

8b.8 Poznámka: Ukázali jsme zajímavou souvislost mezi velikostí grupy a velikostí speciálních podgrup. Přesně stejný postup se dá udělat pro libovolnou podgrupu. Zkuste si znova přečíst úvahy, které k našemu výsledku vedly, ale s takovouto změnou:

- Nechť H je podgrupa G . Pak pro libovolné $x \in G$ je $T(h) = x \circ h$ bijekce z H na xH , mimo jiné tedy $|H| = |xH|$, viz Věta 8a.19 a Věta 8b.5 (ii).
- Pro libovolné $x, y \in G$ platí $xH = yH$ právě tehdy, když $x^{-1} \circ y \in H$, viz Věta 8b.5 (iii).
- Množina $\{xH\}_{x \in G}$ je rozklad množiny G , viz Věta 8b.5 (iv).

Důsledek: Jestliže je G konečná grupa a H její podgrupa, tak nutně $|H|$ dělí $|G|$.

Tomuto se říká Lagrangeova věta. Jak vidíte, dostáváme se zde již blízko k teorii grup, ale v našem dalším textu to nebudeme potřebovat. Pokud vás to zaujalo, přečtěte si nějakou pěknou knížku o algebře.

\triangle

Ve zbytku kapitoly se podíváme na alternativní způsoby, jak poznat, že je pologrupa vlastně grupou. První kritérium říká, že to lze poznávat podle toho, zda v rámci dotyčného prostoru umíme řešit lineární rovnice.

Věta 8b.9.

Nechť (M, \circ) je pologrupa. Je to grupa právě tehdy, když pro libovolné $a, b \in M$ existují řešení $x, y \in M$ rovnic $a \circ x = b$ a $y \circ a = b$.

Ta řešení jsou pak jednoznačná.

Důkaz (poučný): 1) \implies : Jestliže je M grupa, pak existuje jednotkový prvek e a inverzní prvek a^{-1} . Proto lze definovat $x = a^{-1} \circ b$ a snadno ověříme, že řeší první rovnici: $a \circ x = a \circ (a^{-1} \circ b) = (a \circ a^{-1}) \circ b = e \circ b = b$.

Podobně ověříme, že $y = b \circ a^{-1}$ řeší druhou rovnici. Jsou ta řešení jednoznačná? Nechť x je nějaké řešení první rovnice. Pak už nutně $x = e \circ x = (a^{-1} \circ a) \circ x = a^{-1} \circ (a \circ x) = a^{-1} \circ b$.

Podobně pro tu druhou rovnici.

2) \iff : Zvolme si nějaké $a \in M$. Pak v M existuje řešení rovnice $a \circ x = a$, označme ho e_r , a řešení $x \circ a = a$ označme e_l . Chceme ukázat, že jde o jediný prvek, a to jednotkový.

Na to ještě potřebujeme řešení rovnice $x \circ a = e_l$ označené $a_{l,-1}$ a řešení rovnice $a \circ x = e_r$ označené $a_{r,-1}$. Pak máme

$$e_l = a_{l,-1} \circ a = a_{l,-1} \circ (a \circ e_r) = (a_{l,-1} \circ a) \circ e_r = e_l \circ e_r = e_l \circ (a \circ a_{r,-1}) = (e_l \circ a) \circ a_{r,-1} = a \circ a_{r,-1} = e_r.$$

Označme tedy tento prvek e . Ukážeme, že je to jednotkový prvek.

Začneme tím, že $e^2 = e$. Máme totiž $e^2 \circ a = e \circ (e \circ a) = e \circ a = a$, proto také e^2 řeší rovnici $x \circ a = a$ a podle předchozího výpočtu s $e_l = e^2$ a $e_r = e$ máme $e^2 = e$.

Tedž vezměme libovolné $b \in M$. Pak existuje řešení rovnice $x \circ e = b$, označme jej c . Máme

$$b \circ e = (c \circ e) \circ e = c \circ e^2 = c \circ e = b.$$

Podobně ukážeme, že $e \circ b = b$. Máme potvrzeno, že (M, \circ, e) je monoid. Předpoklad o řešení rovnic pak umožňuje pro libovolné $a \in A$ najít řešení rovnic $a \circ x = e$ a $y \circ a = e$, tedy podle Věty 8a.6 získáme inverzní prvek k a . \square

Toto vede na zajímavý závěr. Mějme pologrupu (M, \circ) a uvažujme lineární rovnice. Věta říká, že když jsou takové rovnice řešitelné všechny, tak už jsou řešitelné jednoznačně. Z toho vyplývá, že když má nějaká rovnice více řešení, pak už musí také existovat rovnice, která řešení nemá.

Kapitolu ukončíme dalším způsobem, jak poznat grupu. Je to spíš taková kuriozitka.

V kapitole o monoidech jsme čtenáře varovali, že když zkoumáme, zda je nějaké x invertibilní, tak ještě nestačí, když najdeme „jednostrannou inverzi“, existence prvku y s vlastností $y \circ x = e$ nic neznamená. Musíme také vyzkoušet platnost $x \circ y = e$ a teprve pak jásat.

Když testujeme, zda je nějaký monoid (M, \circ, e) grupou, tak testujeme všechny prvky $x \in M$, zda jsou invertibilní. Ukáže se, že při takovémto hromadném testování už stačí zkoušet jen z jedné strany. Dokonce se to týká nejen axiomu (A3), ale i (A2).

Věta 8b.10.

Uvažujme množinu M s binární operací \circ . Předpokládejme, že následující podmínky jsou splněny:

$$(A1) \text{ pro všechny } x, y, z \in M \text{ platí } x \circ (y \circ z) = (x \circ y) \circ z;$$

$$(A2') \text{ existuje } e \in M \text{ takové, že } e \circ x = x \text{ pro všechna } x \in M;$$

$$(A3') \text{ pro každé } x \in M \text{ existuje } y \in M: y \circ x = e.$$

Pak už je (M, \circ, e) grupa.

Důkaz (poučný): 1) Nechť $x \in M$ a y je levá inverze k x dle (A3'). Ukážeme, že je to i pravá inverze. K tomu ještě najdeme prvek z jako levou inverzi prvku y dle (A3'). Pak pomocí (A2') počítáme

$$x \circ y = e \circ (x \circ y) = (z \circ y) \circ (x \circ y) = z \circ (y \circ x) \circ y = z \circ e \circ y = z \circ (e \circ y) = z \circ y = e.$$

Máme tedy podmínky $y \circ x = e$ a $x \circ y = e$, což bude znamenat, že libovolné $x \in M$ je invertibilní, ale až ve chvíli, kdy ukážeme, že e je opravdu jednotkový prvek.

2) Nechť $x \in M$ je libovolné. Podmínka (A2') dává $e \circ x = x$, potřebujeme to ještě zprava. Pro toto x najdeme levou inverzi y dle (A3'), která je dle 1) i pravou inverzí, a počítáme $x \circ e = x \circ (y \circ x) = (x \circ y) \circ x = e \circ x = x$. \square

Je samozřejmě možné udělat verze (A2'') a (A3'') s pravou identitou a pravou inverzí, zase by to stačilo na grupu. Tím končí kapitola o grupách.

Cvičení

Cvičení 8b.1 (rutinní, poučné): Uvažujte $G = \mathbb{Q} - \{0\}$ a operaci $x \circ y = 2xy$. Dokažte, že je to grupa, a najděte vzorec pro inverzní prvky.

Cvičení 8b.2 (rutinní, poučné): Nechť U je množina, uvažujme $G = P(U)$ a operaci **symetrický rozdíl** definovanou jako $A \div B = (A - B) \cup (B - A)$ (nakreslete si Vennův diagram).

Dokažte, že (G, \div) je grupa, identifikujte jednotkový prvek a inverzní prvky.

Nápověda: asociativita stačí Vennovým diagramem.

Cvičení 8b.3 (rutinní, poučné): Uvažujte monoid (\mathbb{Z}_9, \cdot) .

(i) Najděte grupu N jeho invertibilních prvků. Ověřte, že je opravdu uzavřená na násobení modulo 9.

(ii) Pro všechny prvky $x \in N$ určete $\langle x \rangle$ a $\text{ord}(x)$.

(iii) Nechť $g = 7$. Pro všechna $x \in N$ určete $x \langle g \rangle$, najděte odpovídající rozklad grupy N .

Cvičení 8b.4 (rutinní, poučné): Uvažujte matici $A = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$ v monoidu $(M_{2 \times 2}, \cdot)$.

Určete $\langle A \rangle$ a $\text{ord}(A)$.

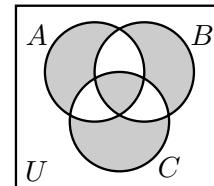
Řešení:

8b.1: $(x \circ y) \circ z = 4xyz = x \circ (y \circ z)$, $e = \frac{1}{2}$, $x^{-1} = \frac{1}{4x}$.

8b.2:

Vennův diagram $(A \div B) \div C = A \div (B \div C)$ je „květinka“, viz vpravo.

Ověřte: $A \div \emptyset = A$ a $A^{-1} = A$ pro libovolnou množinu.



8b.3: (i): $N = \{1, 2, 4, 5, 7, 8\}$. Ověření: třeba $7 \cdot 8 = 56 \bmod 9 = 2 \in N$, atd.

(ii) $\langle 1 \rangle = \{1\}$ a $\text{ord}(1) = 1$; $\langle 2 \rangle = \{2, 2^2 = 4, 2^3 = 8, 2^4 = 7, 2^5 = 5, 2^6 = 1\} = N$ a $\text{ord}(2) = 6$;

$\langle 4 \rangle = \{4, 4^2 = 7, 4^3 = 1\}$ a $\text{ord}(4) = 3$; $\langle 5 \rangle = \{5, 5^2 = 7, 5^3 = 8, 5^4 = 4, 5^5 = 2, 5^6 = 1\} = N$ a $\text{ord}(5) = 6$;

$\langle 7 \rangle = \{7, 7^2 = 4, 7^3 = 1\}$ a $\text{ord}(7) = 3$; $\langle 8 \rangle = \{8, 8^2 = 1\}$ a $\text{ord}(8) = 2$;

(iii) $1\langle 7 \rangle = \langle 7 \rangle = \{1, 4, 7\}$; $2\langle 7 \rangle = \{2, 8, 5\}$; $4\langle 7 \rangle = \{4, 7, 1\}$; $5\langle 7 \rangle = \{5, 2, 8\}$; $7\langle 7 \rangle = \{7, 1, 4\}$; $8\langle 7 \rangle = \{8, 5, 2\}$.
 $N = \{1, 4, 7\} \cup \{2, 5, 8\}$.

8b.4: $\langle A \rangle = \left\{ A, A^2 = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}, A^3 = E_2 \right\}$; $\text{ord}(A) = 3$.

8c. Další struktury

Tato kapitola je z větší části nepovinná a zbytek knihy lze bez ní číst, ale zajímavě doplňuje kapitolu 7c o polynomech a vůbec nabídne trochu nadhled. Víc ani nebylo cílem, tématy jen prolétneme, aby měl čtenář základní orientaci, oč vůbec jde.

Pokročilejší struktury vznikají například tak, že se přidá další operace.

Definice.

Uvažujme binární operace \oplus a \odot na množině R . Řekneme, že (R, \oplus, \odot) je **okruh (ring)**, jestliže splňuje následující axiomy:

- (A1) pro všechna $x, y, z \in R$ platí $x \oplus (y \oplus z) = (x \oplus y) \oplus z$, (asociativita \oplus)
- (A2) existuje $n \in R$ takové, že $x \oplus n = n \oplus x = x$ pro všechna $x \in R$, (jednotkový prvek pro \oplus)
- (A3) pro každé $x \in R$ existuje $y \in R$ takové že $x \oplus y = y \oplus x = n$, (inverzní prvky pro \oplus)
- (A4) pro každé $x, y \in R$ platí $x \oplus y = y \oplus x$, (komutativita \oplus)
- (A5) pro všechna $x, y, z \in R$ platí $x \odot (y \odot z) = (x \odot y) \odot z$, (asociativita \odot)
- (A6) existuje $e \in R$ takové, že $x \odot e = e \odot x = x$ pro všechna $x \in R$, (jednotkový prvek pro \odot)
- (A7) pro všechna $x, y, z \in R$ platí

$$x \odot (y \oplus z) = (x \odot y) \oplus (x \odot z) \quad \text{a} \quad (y \oplus z) \odot x = (y \odot x) \oplus (z \odot y). \quad (\text{distributivní zákon})$$

Řekneme, že okruh (R, \oplus, \odot) je **komutativní okruh (commutative ring)**, jestliže navíc splňuje axiom (A8) pro každé $x, y \in R$ platí $x \odot y = y \odot x$. (komutativita \odot)

I zde někteří autoři přidávají

(A0) pro všechna $x, y \in R$ platí $x \oplus y \in R$ a $x \odot y \in R$.

I zde je to zbytečné, ale nezaškodí to. Když se na ty podmínky pro okruh podíváme, vidíme, že je lze zkrátit následovně: (R, \oplus, n) je komutativní grupa, (R, \odot, e) je monoid a operace jsou spolehlivě svázány distributivním zákonem.

Z toho vyplývá, že jednotkové prvky i inverzní prvky jsou jednoznačně určeny. Inverzní prvky vzhledem k \oplus se tradičně značí $-x$, zatímco inverzní prvky vzhledem k \odot (alespoň ty, které náhodou existují) se značí x^{-1} . Víme z kapitoly 8a, že přinejmenším jeden existuje, $e^{-1} = e$. Mnoho autorů používá mnemotechnickou pomůcku a značí jednotkové prvky pomocí 0 a 1, někteří dokonce používají pro okruh značení $(R, +, \cdot)$, ale je třeba si uvědomit, že pak „0“ nemusí být číslo 0 a „1“ nemusí být číslo 1, jak ostatně hned uvidíme.

Nejjednodušší okruh je $R = \{0\}$ s operacemi $0 \oplus 0 = 0 \odot 0 = 0$. Pak také $n = e = 0$. Zrovna toto se některým autorům nelibí a přidávají do definice okruhu podmítku, že $n \neq e$, čímž vznikne okruh s více prvky. Dá se mimochodem ukázat, že jakmile má okruh více než jeden prvek, tak už nutně $n \neq e$ (viz poznámka po Faktu 8c.1). Tím se dostáváme k tomu, že autoři definuje okruhy různě, někteří třeba vynechávají axiom (A6), naše definice patří k těm nejobvyklejším.

Klasickým příkladem okruhů jsou samozřejmě naši starí známí $(\mathbb{R}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$, $(\mathbb{Z}, +, \cdot)$ a $(\mathbb{Z}_n, +, \cdot)$. Okruhy se v matematice používají docela hodně, protože jejich vlastnosti jsou vyváženým kompromisem mezi snahou mít co nejvíce axiomů, aby platilo hodně vlastností, a snahou mít jich co nejméně, aby se do výsledné formy vešlo co nejvíce aplikací.

O okruzích se toho dá říct velice mnoho. Tato kapitola je spíš uzavírací a přehledová, tak jen jako ukázku uvedeme pár vlastností, které máme u nám známých číselných množin (představujte si $n = 0, e = 1$) a fungují i obecně.

Fakt 8c.1.

Nechť (R, \oplus, \odot, n, e) je okruh. Pak platí následující:

- (i) pro každé $x \in R$ platí $n \odot x = x \odot n = n$;
- (ii) pro každé $x \in R$ platí $(-e) \odot x = (-x)$;
- (iii) pro každé $x, y \in R$ platí $(-x) \odot y = x \odot (-y) = -(x \odot y)$.

Důkaz (poučný): (i): Pro neutrální prvek platí $n = n \oplus n$. Pak máme díky distributivnímu zákonu rovnost $x \odot n = x \odot (n \oplus n) = (x \odot n) \oplus (x \odot n)$. Podle Lemma 8a.11 aplikovaného na invertibilní prvek $x \odot n$ v monoidu (R, \oplus, n) tedy $x \odot n = n$. Druhá rovnost se dokazuje obdobně.

(ii): Ověříme, že výraz nalevo splňuje definici inverzního prvku x vůči \oplus , použijeme $x = e \odot x$ a distributivní zákon:

$$x \oplus [(-e) \odot x] = [e \odot x] \oplus [(-e) \odot x] = [e \oplus (-e)] \odot x = n \odot x = n.$$

(iii): Ověříme, že výraz zcela nalevo splňuje definici inverzního prvku $x \odot y$ vůči \oplus :

$$[(-x) \odot y] \oplus [x \odot y] = [(-x) \oplus x] \odot y = n \odot y = n. \text{ Pro druhý výraz zleva se to dělá obdobně.}$$

□

Část (i) má zajímavý důsledek. Kdyby platilo $n = e$, tak už pro libovolné $x \in R$ máme $x = x \odot e = x \odot n = n$, tedy $R = \{n\}$. To potvrzuje naši poznámku výše, že jakmile má R víc než jeden prvek, tak už nutně $n \neq e$. Platí pak obecně i další věc, kterou známe z běžného násobení čísel.

Fakt 8c.2.

Nechť (R, \oplus, \odot, n, e) je okruh a $|R| \geq 2$. Pak n není invertibilní.

Důkaz (poučný): Sporem: Nechť existuje $y = n^{-1}$. Pak $y \odot n = e$, ale podle (i) výše také $y \odot n = n$. Dostáváme $e = n$, což ale platí jen pro jednoprvkové R .

□

Jsou ale také věci, které známe z \mathbb{R} či \mathbb{Z} , ale v obecných okruzích nefungují. V zásadě jsou okruhy dostatečně bohaté na to, abychom na nich zkusili vybudovat celou teorii dělitelnosti z kapitoly 6a (prvočísla atd.), a podle toho, jak daleko se dostaneme (funguje v nich Euklidův algoritmus? funguje v nich jednoznačnost rozkladu na prvočísla?), se okruhy klasifikují. Jsou o tom tlusté knihy z oboru algebra, takže do toho nebudeme štourat, nicméně na jednu zajímavou vlastnost jsme tu už narazili.

Definice.

Nechť (R, \oplus, \odot) je okruh. Řekneme, že prvek $x \in R$ je **dělitelem nuly (zero divisor)**, jestliže $x \neq n$ a existuje $y \neq n$ takový, že $x \odot y = n$.

Na takové prvky nejsme zvyklí, naopak rádi v algebraických úpravách z rovnice $x \cdot y = 0$ usuzujeme, že jeden z prvků musí být nula, ale obecně v okruzích to neplatí. Příklady už jsme potkali v kapitole 7, například při násobení modulo 4 platí $2 \cdot 2 = 0$, neboli v okruhu $(\mathbb{Z}_4, +, \cdot)$ je 2 dělitelem nuly. Podobně jsme v příkladě 8a.b viděli, že v okruhu \mathbb{Z}_{14} platí $2 \cdot 7 = 0, 7 \cdot 10 = 0$ a podobně.

To je v mnoha situacích velice nepříjemné, takže se vyplatí zavést speciální kategorii okruhů, kde toto nehrozí.

Definice.

Řekneme, že okruh (R, \oplus, \odot) je **obor integrity (integral domain)**, jestliže je to komutativní okruh bez dělitelů nuly.

Takže $(\mathbb{R}, +, \cdot)$ a $(\mathbb{Z}, +, \cdot)$ jsou dozajista obory integrity, zajímavé je, že jsme v kapitole 7a zjistili, že pro prvočísla p jsou i $(\mathbb{Z}_p, +, \cdot)$ obory integrity. V těch se hned lépe pracuje. Jeden způsob, jak se dělitelů nuly zbavit, je mít co nejvíce invertibilních prvků.

Fakt 8c.3.

Nechť je (R, \oplus, \odot) okruh. Jestliže je $x \in R$ invertibilní vůči \odot , pak není dělitelem nuly.

Důkaz (rutinní, poučný): Nechť $y \in R$ splňuje podmínu $x \odot y = n$. Ukážeme, že pak už nutně $y = n$.

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

□

Neplatí to ale naopak: Prvky, které nejsou děliteli nuly, nemusí být automaticky invertibilní. Pro příklad se stačí podívat na $(\mathbb{Z}, +, \cdot)$, třeba $x = 13$ není dělitelem nuly, ale ani nemá v \mathbb{Z} inverzní prvek vůči násobení.

Měli jsme Fakt 8a.10, který říkal, že invertibilní prvky je možno v rovnících krátit. Platí to i pro prvky, které nejsou dělitelé nuly. Jak jsme právě viděli, je jich potencionálně více, čímž se vysvětluje, proč můžeme krátit nenulová čísla v rovnících, i když pracujeme v \mathbb{Z} . Poučný je způsob, kterým se toto dokazuje, protože to výstižně ukazuje mechanismus práce v okruzích.

Normálně bychom v rovnici $cx = cy$ přesunuli oba prvky na levou stranu odečtením pravé strany, získáme $cx - cy = 0$, vytkneme na $c(x - y) = 0$ a pak vydělíme. V okruhu ovšem nemáme ani odčítání, ani dělení, takže se vše musí odehrávat přes inverzní prvky. Jdeme na to.

Předpokládejme tedy, že máme rovnici $c \odot x = c \odot y$ a c není dělitel nuly. Přičteme k obou stranám opačný prvek k pravé straně vůči \oplus a dostaneme $(c \odot x) \oplus [-(c \odot y)] = n$. Teď použijeme (iii) z Faktu 8c.1, máme $(c \odot x) \oplus (c \odot (-y)) = n$ a jsme zralí na distributivní zákon: $c \odot (x \oplus (-y)) = n$. Protože c není dělitel nuly, musí být nutně $x \oplus (-y) = n$, takže $-y$ splňuje podmínu inverzního prvku vůči x . Z jeho jednoznačnosti máme $-x = -y$. Pak také $-(-x) = -(-y)$, tedy $x = y$, zde jsme použili Větu 8a.9.

Nejlepší jsou ale stejně prvky invertibilní. Ideální je stav, kdy jsou invertibilní všechny kromě n , kde to nejde již z principu (s výjimkou triviálního jednoprvkového okruhu).

Definice.

Uvažujme binární operace \oplus a \odot na množině F . Řekneme, že (F, \oplus, \odot) je **těleso (field)**, jestliže je to komutativní okruh, tedy splňuje axiomy (A1) až (A8), a navíc splňuje axiom

(A9) pro každé $x \in F$, $x \neq n$ existuje $y \in F$ takové, že $x \odot y = y \odot x = e$. (inverzní prvky pro \odot)

Jinak řečeno, (F, \oplus, n) je komutativní grupa, $(F - \{n\}, \odot, e)$ je komutativní grupa a operace jsou svázány distributivním zákonem. Klasickými příklady těles jsou \mathbb{C} , \mathbb{R} či \mathbb{Q} , teď už \mathbb{Z} vypadává. V kapitole 7a jsme viděli, že pro prvočísla p je \mathbb{Z}_p těleso, pro computer science je samozřejmě velice zajímavé těleso \mathbb{Z}_2 . Díky existenci inverzních prvků vůči \odot už nejsou v tělesech žádní dělitelé nuly, takže těleso je i okruh integrity. Dobře se proto s nimi pracuje, s tělesy se dá dělat v zásadě všechno, co s \mathbb{R} , například lineární algebra.

Příklad 8c.a: Máme-li těleso F , můžeme nad ním definovat matice obvyklým způsobem, pro každé $n \in \mathbb{N}$ tak vznikne množina $M_n(F)$ matic o rozměru $n \times n$. Když ji vybavíme operacemi maticového sčítání a násobení, které se definuje běžným způsobem, jen namísto $+$ a \cdot používáme operace \oplus a \odot z tělesa, dostáváme tak okruh, který je nekomutativní pro $n \geq 2$. Asi není překvapením, že v computer science jsou zajímavé matice nad \mathbb{Z}_p pro prvočísla p , zejména okruh $M_n(\mathbb{Z}_2)$. Blíže viz kapitola 7c.

8c.4 Okruhy polynomů

Běžně pracujeme s polynomy nad \mathbb{R} , značenými $\mathbb{R}[x]$, setkali jsme se již také s polynomy nad \mathbb{Z}_n . Zde vybudujeme polynomy nad obecným okruhem (R, \oplus, \odot) . Pro účely této části se vyplatí značit jeho neutrální prvek 0 a jednotkový prvek 1, protože to zjednoduší zápis a čtenář už je snad dost zkušený na to, aby věděl, že „0“ nemusí být zrovna číslo 0, ale třeba matice ze samých nul či jiný neutrální prvek. Zavedeme také pro zjednodušení zápisu prioritu \odot před \oplus .

!

Definice.

Uvažujme okruh (R, \oplus, \odot) .

Termínem **polynom (polynomial)** nad R označujeme abstraktní výraz

$$p = \sum_{k=0}^n a_k x^k = a_0 x^0 + a_1 x^1 + \cdots + a_{n-1} x^{n-1} + a_n x^n,$$

kde $n \in \mathbb{N}_0$ a $a_i \in R$ pro všechna i .

Jako $R[x]$ označíme množinu všech polynomů nad R .

! Poznámka (velice důležitá): Je třeba si uvědomit, že to x má v polynomu ryze symbolický charakter. Sice mu říkáme **proměnná**, ale to je jen jméno pro symbol, klidně tam místo x může být hvězdička nebo třeba srdíčko: $p = 13 + 2\heartsuit^2$ by mohl být polynom nad \mathbb{Z}_{14} . Tento symbol, stejně jako symbol „+“ nemající nic společného se scítáním, slouží k oddělení jednotlivých koeficientů polynomu. Když vidíme x^2 , tak víme, že ten prvek z R před ním je v pořadí třetí koeficient polynomu. Šlo by to i jinak, klidně jsme namísto $a_0x^0 + a_1x^1 + a_2x^2$ mohli zavést zápis „ $a_0\heartsuit, a_1\heartsuit, a_2\heartsuit$ “ a celá teorie by dál fungovala. Jinak řečeno, to podstatné jsou koeficienty a ostatní symboly slouží jen jako připomínka, kde který koeficient je.

V některých aplikacích si proto množinu $R[x]$ ztotožníme s množinou $\bigcup_{n=0}^{\infty} R^k$ vektorů s koeficienty z R libovolné délky, například $p = 13 + 2x^2$ je jen pohodlný zápis pro vektor $(13, 0, 2)$. Zde je třeba přijmout úmluvu, že každý vektor lze prodloužit nulami na vektor libovolné délky a bude to pořád stejný vektor. Tím do nového jazyka zahrneme obvyklou úmluvu, že $p = 13 + 2x^2$ je totéž jako $p = 13 + 2x^2 + 0x^3 + 0x^4$, takže potřebujeme, aby i vektory $(13, 0, 2)$ a $(13, 0, 2, 0, 0)$ byly považovány za stejné. Celou teorii polynomů je pak možné vybudovat čistě v jazyce vektorů a nebude to ani o moc těžší.

Takovéto polynomy lze vlastně zavést nad libovolnou množinou M , která ani nemusí mít nějaké operace. Pak se toho ale moc nedá vymyslet. Zajímavé to je právě u množin s operacemi. Brzy ukážeme, že oném speciálním symbolům $+$ a x^k můžeme dát určitý význam, který odpovídá tomu, co známe z reálných polynomů. Jsou tedy vhodnější než srdíčka, ale dá trochu práce to všechno udělat pořádně.

△

Polynomy nad reálnými čísly zná každý, ukážeme zajímavější příklad.

! Příklad 8c.b: Uvažujme $R = M_2(\mathbb{Z})$, okruh všech matic 2×2 s celočíselnými prvky. Pak třeba

$$p = \begin{pmatrix} 3 & -23 \\ 1 & 14 \end{pmatrix} x^0 + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} x^1 + \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix} x^2 + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} x^3$$

je polynom nad R .

△

! Abychom zjednodušili zápis, uděláme následující úmluvy:

- členy $0x^k$, kde 0 je nulový prvek okruhu R , můžeme ze zápisu vynechat,
- mocninu $1x^k$, kde 1 je jednotkový prvek okruhu R , budeme psát jako x^k ,
- mocninu x^1 budeme psát jako x ,
- mocninu $x^0 = 1$ nebudeme při psaní uvádět.

Samozřejmě tato úmluva úzce souvisí s významem, který symbolu x^k později přiřadíme.

Při této příležitosti uděláme malou poznámku o zápisu. Každý autor řeší při zavádění polynomů dilema, zda radit mocniny od nejmenší k největší nebo naopak. Obojí má své pro i proti. Já jsem se v definici rozhodl brát členy od nejmenších, protože jsme si rozmysleli, že polynomy můžeme „prodlužovat“, a přišlo mi dobré začínat od části, která je pro polynom důležitá, případně prodloužení o nulové členy pak dělat ke konci. Této konvence se budu držet v teoretických úvahách. Na druhou stranu jsme u reálných polynomů zvyklí začínat nejvyšší mocninou, což budu také dělat často v příkladech. Konec konců, je to v zásadě jedno.

Příklad 8c.c: Polynom z předchozího příkladu lze zapsat i jako

$$p = \begin{pmatrix} 3 & -23 \\ 1 & 14 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix} x^2.$$

Mimo jiné jsme použili pravidlo o nulovém prvku $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ okruhu $R = M_2(\mathbb{Z})$.

△

! **Definice.**

Nechť R je okruh, nechť $p = \sum_{k=0}^n a_k x^k \in R[x]$.

Definujeme **stupeň (degree)** polynomu p jako $\text{st}(p) = -\infty$ jestliže pro všechna $k = 1, \dots, n$ je $a_k = 0$ (tedy $p = 0$), jinak jako $\text{st}(p) = \max\{k; a_k \neq 0\}$.

Takže třeba pro polynom p s maticemi z příkladu výše máme $\text{st}(p) = 2$. Definice stupně pro nulový polynom možná překvapí, ale když si ještě zavedeme pravidlo, že $-\infty + a = -\infty$ a $\max(-\infty, a) = a$ pro libovolné a (včetně $-\infty$), tak nám to bude skvěle fungovat.

Ted' se naučíme s polynomy pracovat. Nejprve si zadefinujeme klasické operace. Budou samozřejmě inspirovány tím, co dobře známe, připomeneme násobení reálných polynomů:

$$(a_0 + a_1x + \dots + a_nx^n) \cdot (b_0 + b_1x + \dots + b_mx^m) = [a_0b_0] + [a_0b_1 + a_1b_0]x + [a_0b_2 + a_1b_1 + a_2b_0]x^2 + \dots + [a_nb_m]x^{n+m}.$$

Jenže my ted' nemáme sčítání ani násobení, ale nějaké neznámé operace z okruhu R , takže musíme zapisovat opatrnejí.

! Definice.

Nechť (R, \oplus, \odot) je okruh. Pro polynomy $p = \sum_{k=0}^n a_k x^k$ a $q = \sum_{k=0}^m b_k x^k$ a $c \in R$ definujeme operace takto:

3) Definujeme **skalárni násobek polynom** jako

$$c \square p = \sum_{k=0}^n (c \odot a_i)x^k.$$

2) Doplněním mocnin typu $0x^i$ můžeme předpokládat, že $m = n$, pak definujeme **součet polynomů** jako

$$p \boxplus q = \sum_{k=0}^n (a_k \oplus b_k)x^k.$$

3) Definujeme **součin polynomů** jako

$$p \boxdot q = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i \odot b_j \right) x^k.$$

Zavedli jsme značení \boxplus a \boxdot pro operace mezi polynomy, abychom mohly ve výpočtech a důkazech správně indikovat, kterou operaci používáme. Jakmile naberejme trochu zkušenosti, přejdeme k obvyklé úmluvě, že se operace mezi polynomy značí stejnou značkou jako operace v R .

Význam definice vynikne v dlouhém zápisu. Jestliže $p = a_0 + a_1x + \dots + a_{n-1}x^{n-1} + a_nx^n$ a $q = b_0 + b_1x + \dots + b_{m-1}x^{m-1} + b_mx^m$, pak pro $m = n$ máme

$$p \boxplus q = [a_0 \oplus b_0] + [a_1 \oplus b_1]x + \dots + [a_{n-1} \oplus b_{n-1}]x^{n-1} + [a_n \oplus b_n]x^n,$$

pro libovolná m, n pak máme

$$\begin{aligned} p \boxdot q &= [a_0 \odot b_0] + [(a_0 \odot b_1) \oplus (a_1 \odot b_0)]x + [(a_0 \odot b_2) \oplus (a_1 \odot b_1) \oplus (a_2 \odot b_0)]x^2 + \dots \\ &\quad \dots + [(a_{n-1} \odot b_m) \oplus (a_n \odot b_{m-1})]x^{n-1} + [a_n \odot b_m]x^{n+m}. \end{aligned}$$

Všimněte si, že jde zase jen o manipulaci s koeficienty, takže tyto operace bychom mohli stejně tak definovat pro vektory koeficientů:

$$\begin{aligned} c \square (a_0, a_1, \dots, a_n) &= (c \odot a_0, c \odot a_1, \dots, c \odot a_n) \\ (a_0, a_1, \dots, a_n) \boxplus (b_0, b_1, \dots, b_n) &= (a_0 \oplus b_0, a_1 \oplus b_1, \dots, a_n \oplus b_n) \\ (a_0, a_1, \dots, a_n) \boxdot (b_0, b_1, \dots, b_m) &= (a_0 \odot b_0, (a_0 \odot b_1) \oplus (a_1 \odot b_0), \dots, a_n \odot b_m). \end{aligned}$$

Sice to funguje stejně, ale museli bychom si to pamatovali jako pravidla, zatímco ten zápis s „mocninami“ nám napovídá, jak operace dělat. Má tedy své výhody, jak ukáže příklad.

Příklad 8c.d: Uvažujme polynomy $p = \begin{pmatrix} 0 & -23 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix}x$ a $q = \begin{pmatrix} 3 & 13 \\ 0 & 14 \end{pmatrix} + \begin{pmatrix} 3 & -6 \\ 1 & 7 \end{pmatrix}x$ z okruhu polynomů $M_2(\mathbb{Z})[x]$ a $c = \begin{pmatrix} 2 & 0 \\ -1 & 0 \end{pmatrix} \in M_2(\mathbb{Z})$. Pak máme

$$\begin{aligned} c \square q &= \begin{pmatrix} 2 & 0 \\ -1 & 0 \end{pmatrix} \square \left[\begin{pmatrix} 0 & -23 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix}x \right] = \left[\begin{pmatrix} 2 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -23 \\ 1 & 0 \end{pmatrix} \right] + \left[\begin{pmatrix} 2 & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix} \right] x \\ &= \begin{pmatrix} 0 & -46 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & -13 \end{pmatrix} x, \end{aligned}$$

$$\begin{aligned} p \boxplus q &= \left[\begin{pmatrix} 0 & -23 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 3 & 13 \\ 0 & 14 \end{pmatrix} \right] + \left[\begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix} + \begin{pmatrix} 3 & -6 \\ 1 & 7 \end{pmatrix} \right] x \\ &= \begin{pmatrix} 3 & -10 \\ 1 & 14 \end{pmatrix} + \begin{pmatrix} 4 & -6 \\ 1 & 20 \end{pmatrix} x \end{aligned}$$

a

$$\begin{aligned}
 p \boxdot q &= \left[\begin{pmatrix} 0 & -23 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix} x \right] \boxdot \left[\begin{pmatrix} 3 & 13 \\ 0 & 14 \end{pmatrix} + \begin{pmatrix} 3 & -6 \\ 1 & 7 \end{pmatrix} x \right] \\
 &= \left[\begin{pmatrix} 0 & -23 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 13 \\ 0 & 14 \end{pmatrix} \right] + \left[\begin{pmatrix} 0 & -23 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & -6 \\ 1 & 7 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix} \begin{pmatrix} 3 & 13 \\ 0 & 14 \end{pmatrix} \right] x + \left[\begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix} \begin{pmatrix} 3 & -6 \\ 1 & 7 \end{pmatrix} \right] x^2 \\
 &= \begin{pmatrix} 0 & -322 \\ 3 & 13 \end{pmatrix} + \begin{pmatrix} -20 & -148 \\ 3 & -1086 \end{pmatrix} x + \begin{pmatrix} 3 & -6 \\ 13 & 91 \end{pmatrix} x^2.
 \end{aligned}$$

Protože víme, že násobení matic není komutativní, je třeba být opatrný na správné pořadí matic při roznásobování závorek. Z toho pak usoudíme, že násobení polynomů nebude obecně komutativní.

△

Vytvářením polynomů neztratíme základní vlastnosti.

Věta 8c.5.

Nechť (R, \oplus, \odot) je okruh. Pak je $(R[x], \boxplus, \boxdot)$ také okruh.

Jestliže je (R, \oplus, \odot) komutativní, pak je také $(R[x], \boxplus, \boxdot)$ komutativní.

Důkaz si v této přehledové kapitole odpustíme, je snadný, ale dlouhý, než se ověří všech 7 až 8 axiomů.

Teď se podíváme, které z vlastností známých pro běžné reálné polynomy platí i obecně. Začneme stupněm.

Fakt 8c.6.

Nechť R je okruh a $p, q \in R[x]$. Pak $\text{st}(p \boxplus q) \leq \max(\text{st}(p), \text{st}(q))$ a $\text{st}(p \boxdot q) \leq \text{st}(p) + \text{st}(q)$.

To první tvrzení nepřekvapí, sečtením polynomů se nemůže stupeň navýšit, zato se vyšší mocniny mohou ztratit, když se pokrátí, například $(x+1) + (-x+1) = 2$. Zato u násobení asi čtenář čekal rovnost, jenže tady je zádrhel s děliteli nuly. Například v okruhu \mathbb{Z}_4 máme $(2x+1) \cdot (2x+1) = 4x^2 + 4x + 1 = 0x^2 + 0x + 1 = 1$.

Takovýmto problémům se dá vyhnout, když se jako základ použije těleso.

Věta 8c.7.

Je-li R těleso, pak $R[x]$ je okruh integrity a pro $p, q \in R[x]$ platí $\text{st}(p \boxdot q) = \text{st}(p) + \text{st}(q)$.

Vůbec se dá říct, že pokud chceme, aby se polynomy chovaly tak, jak jsme zvyklí, tak potřebujeme začít s tělesem. Jako příklad si dokážeme, že funguje dělení polynomů se zbytkem.

Věta 8c.8.

Nechť F je těleso a p, d jsou polynomy z $F[x]$, nechť $d \neq 0$.

Pak existují polynomy $q, r \in F[x]$ takové, že $p = q \boxdot d \boxplus r$ a $\text{st}(r) < \text{st}(d)$.

Důkaz (drsný, poučný): Důkaz povedeme silnou indukcí na $\text{st}(p)$ a zároveň tím ukážeme algoritmus. Zkuste si rozmyslet, že je to pořád náš starý známý algoritmus na dělení polynomů se zbytkem, jen obecně nemůžeme mluvit o odčítání a dělení, místo toho používáme opačné a inverzní prvky.

(0) Nechť $\text{st}(p) = 0$. Pak je $p = a_0$ konstantní polynom. Jestliže $\text{st}(d) > 0$, pak $p = 0 \boxdot d \boxplus p$ a $\text{st}(p) < \text{st}(d)$, tedy lze vzít $r = p$. Jestliže $\text{st}(d) = 0$, tak je $d = b_0$ konstantní polynom a díky našemu předpokladu není nulový. b_0 má tedy inverzi vůči násobení v F a můžeme psát $a_0 = a_0 \odot b_0^{-1} \odot b_0 \oplus 0$ neboli $p = (a_0 \odot b_0^{-1}) \boxdot d \boxplus r$, kde $r = 0$, proto $\text{st}(r) = -\infty < 0 = \text{st}(d)$.

(1) Předpokládejme, že tvrzení platí pro všechny polynomy stupně nejvyšše n . Mějme polynom p stupně $n+1$ a nenulový polynom d . Jestliže $\text{st}(d) > \text{st}(p)$, pak položíme $p = 0 \boxdot d \boxplus p$ a je to.

Jinak nechť a_{n+1} je nejvyšší nenulový koeficient p a b_m je nejvyšší nenulový koeficient d , máme $m \leq n+1$. Pak polynom

$(-a_{n+1} \odot b_m^{-1})x^{n+1-m} \boxdot d$ má nejvyšší koeficient rovný $-a_{n+1}$ a je u mocniny $x^{m+n+1-m} = x^{n+1}$, takže po přičtení k p dostaneme, že polynom $h = p \boxplus (-a_{n+1} \odot b_m^{-1})x^{n+1-m} \boxdot d$ má určitě stupeň menší než $n+1$. To znamená, že na něj můžeme aplikovat indukční předpoklad a dostáváme polynomy q', r takové, že $\text{st}(r) < \text{st}(d)$ a $h = q' \boxdot d \boxplus r$. Pak

$$\begin{aligned}
 p &= h \boxplus (-a_{n+1} \odot b_m^{-1})x^{n+1-m} \boxdot d = (-a_{n+1} \odot b_m^{-1})x^{n+1-m} \boxdot d \boxplus q' \boxdot d \boxplus r \\
 &= [(-a_{n+1} \odot b_m^{-1})x^{n+1-m} \boxplus q'] \boxdot d \boxplus r
 \end{aligned}$$

a pořád $\text{st}(r) < \text{st}(d)$. Rozklad je hotov.

□

Když umíme dělit, můžeme také zavést pojem dělitelnosti, hledat společné násobky a dělitele a další věci, jako se to dělá v kapitole 6a. Obecně se dá říct, že když máme polynomy nad tělesem, tak vše funguje v zásadě stejně, a to dokonce včetně Bezoutovy věty a rozšířeného Euklidova algoritmu, i když si asi umíte představit, že dělit v Euklidovi polynomy se zbytkem, přičemž koeficienty jsou z nějakého exotického tělesa, je poněkud dobrodružnější, než když se to dělá s čísly.

Můžeme také zavést pojem inspirovaný prvočísly.

Definice.

Nechť R je okruh. Řekneme, že polynom $p \in R[x]$ je **ireducibilní (irreducible)**, jestliže neexistují polynomy $q, r \in R[x]$ takové, že $p = q \square r$, $\text{st}(q) < \text{st}(p)$ a $\text{st}(r) < \text{st}(p)$.

Pro reálné polynomy jsou ireducibilními polynomy všechny konstantní, lineární a pak ty kvadratické, které nemají reálné kořeny. Pokud máme polynomy nad tělesem, jsme schopni každý polynom rozložit na součin ireducibilních faktorů, podobně jako prvočíselný rozklad celých čísel.

Ted' se podíváme na poslední velké téma, z formálních polynomů vytvoříme zobrazení (či funkci, chcete-li) z R do R . Není v tom žádné překvapení, pro polynom $p = a_n x^n + \dots + a_1 x + a_0$ definujeme ono zobrazení předpisem

$$r \mapsto (a_n \odot r^n) \oplus \dots \oplus (a_1 \odot r) \oplus a_0.$$

Říkáme, že do polynomu dosazujeme $r \in R$, hodnotu značíme $p(r)$.

Příklad 8c.e: Uvažujme polynom $p = \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix} x + \begin{pmatrix} 0 & -23 \\ 1 & 0 \end{pmatrix}$ z prostoru $M_2(\mathbb{Z})[x]$. Můžeme do něj dosadit třeba $x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in M_2(\mathbb{Z})$ a dostaneme

$$p(x) = \begin{pmatrix} 1 & 0 \\ 0 & 13 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -23 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -22 \\ 1 & 0 \end{pmatrix}.$$

△

Pro toto dosazování platí běžná pravidla, dá se například dokázat, že pro polynomy p, q získáme dosazením do $p \boxplus q$ stejnou hodnotu, jako kdybychom dosadili do p a q zvlášť a výsledky spojili pomocí \oplus . Formálně řečeno: $(p \boxplus q)(r) = p(r) \oplus q(r)$, podobně $(p \square q)(r) = p(r) \odot q(r)$.

Je třeba si ale uvědomit, že zobrazení daná polynomu jsou obecně jinými objekty než polynomy, které jim daly vzniknout. Čtenář je samozřejmě zvyklý na polynomy reálné, kde se polynom a funkce z něj vzniklá berou jako totéž. To je odůvodněno základním faktrem, že když dva polynomy dělají stejné hodnoty coby reálné funkce, tak už musí jít o stejně polynomy coby výrazy, tedy musejí se rovnat všechny koeficienty. Bohužel, toto neplatí u polynomů nad obecnými tělesy. Může se stát, že dva polynomy, které mají různé koeficienty, dělají při dosazování $r \in R$ tvrdošíjně stále stejné hodnoty, takže z nich vzniká stejně zobrazení. Pro případ není třeba chodit daleko, viz kapitola 7c.

Proto je třeba obecně dbát na rozdíl mezi polynomem jako výrazem a zobrazením vzniklým pomocí onoho polynomu.

Definice.

Nechť R je okruh, uvažujme polynom $p \in R[x]$. Řekneme, že prvek $r \in R$ je **kořen (root)** polynomu p , jestliže po dosazení do p platí $p(r) = 0$.

Pro tělesa pak máme povědomé tvrzení.

Věta 8c.9.

Nechť F je těleso, uvažujme polynom $p \in F[x]$.

Prvek $r \in F$ je kořenem p právě tehdy, když polynom $x + (-r)$ dělí p .

Znamená to, že polynom, který má kořen, již nemůže být ireducibilní. Obměna tedy říká, že ireducibilní polynom nemá kořeny. Platí to i naopak? Obecně ne, dokonce ani u reálných polynomů, například polynom $(1+x^2)(1+x^2)$ reálné kořeny nemá, ale rozložit jej lze. Pro polynomy nižšího stupně už ale ireducibilita a neexistence kořenů souhlasí dokonce obecně.

Fakt 8c.10.

Nechť F je těleso, p je polynom z $F[x]$ stupně 2 nebo 3.
 p je irreducibilní v $F[x]$ právě tehdy, když nemá kořeny v F .

Důkaz (poučný): Neexistenci kořenů pro irreducibilní polynomy jsme už ukázali výše.

Předpokládejme naopak, že polynom p není irreducibilní, tedy $p = s \odot t$, kde s, t jsou polynomy se stupni menšími než $\text{st}(p)$. Protože $\text{st}(p) \leq 3$, musí být alespoň jeden z polynomů s, t stupně jedna, čili polynom ve tvaru $x + r$. Pak je $(-r) \in F$ jeho kořenem, tedy i kořenem p . \square

Tímto končíme stručný přehled vlastností polynomů nad okruhy.

8d. Bonus: Racionální čísla

Zde si položíme zdánlivě jednoduchou otázku: Co jsou to racionální čísla? Je svůdné odpovědět, že to jsou zlomky, jenž co je to vlastně za objekt? Zdá se přirozené definovat zlomek jako uspořádanou dvojici čísel, pro kterou jsme zvolili speciální zápis, jenž tento nápad má zásadní problém. Pokud přijmeme, že $\frac{1}{2}$ je zápis pro dvojici $(1, 2)$, co je pak $\frac{2}{4}$? Jako dvojice $(2, 4)$ je to zcela jiný objekt než $(1, 2)$, ale my bychom to rádi brali jako jednu věc. Je vidět, že to nebude tak snadné, ale s trohou práce se s tím vypořádáme.

V této kapitole korektně vybudujeme množinu racionálních čísel, přičemž si krásně ukážeme v akci pojmy z několika kapitol této knihy. Protože důkazy jsou často spíš technické (nahání se triviální rovnosti a nerovnosti), čtenář je případně může přeskočit a soustředit se na tok myšlenek.

8d.1 Budujeme zlomky

Začneme množinou, která se nabízí, provizorně si ji nazveme \mathbb{F} jako „fractions“:

$$\mathbb{F} = \{(p, q); p, q \in \mathbb{Z} \wedge q \neq 0\}.$$

Pro ušetření místa a znaků se domluvíme, že místo dvojic (p, q) budeme psát $\frac{p}{q}$.

Teď potřebujeme matematicky zachytit, že například zlomky $\frac{4}{6}$ a $\frac{6}{9}$ jsou vlastně stejná věc. Na to se přesně hodí pojem relace.

Definice 8d.2.

Definujeme relaci \sim na množině \mathbb{F} předpisem $\frac{p}{q} \sim \frac{u}{v}$ právě tehdy, když $pv = uq$.

Protože $4 \cdot 9 = 6 \cdot 6$, máme $\frac{4}{6} \sim \frac{6}{9}$, přesně jak bychom chtěli. Tato ekvivalence má geometrický význam. Každý zlomek $\frac{p}{q}$ neboli (p, q) reprezentuje určitou délku na reálné ose, tuto délku získáme tak, že si vezmeme úsek délky p a rozdělíme jej na q shodných částí (zde je nutno precizovat, co toto znamená, když je jedno či obě čísla záporné, vzniká tím orientace). Není obtížné si rozmyslet, že dva zlomky jsou v relaci \sim právě tehdy, když dávají stejnou geometrickou délku.

Čtenář již možná tuší, co od naší relace očekáváme.

Fakt 8d.3.

Relace \sim je na \mathbb{F} ekvivalence.

Důkaz (rutinní): 1) reflexivita: Nechť $\frac{p}{q} \in \mathbb{F}$. Protože určitě $pq = pq$, je i $\frac{p}{q} \sim \frac{p}{q}$.

2) symetrie: Nechť $\frac{p}{q}, \frac{u}{v} \in \mathbb{F}$ a předpokládejme, že $\frac{p}{q} \sim \frac{u}{v}$. Pak $pv = uq$, proto i $uq = pv$ a tedy $\frac{u}{v} \sim \frac{p}{q}$.

3) tranzitivita: Nechť $\frac{p}{q}, \frac{u}{v}, \frac{s}{t} \in \mathbb{F}$ a předpokládejme, že $\frac{p}{q} \sim \frac{u}{v}$ a $\frac{u}{v} \sim \frac{s}{t}$. Podle definice to znamená $pv = uq$ a $ut = vs$. Protože $t \neq 0$ a $q \neq 0$, můžeme obě rovnice vynásobit na tvar $pvt = uqt$ a $utq = vsq$, což dává $pvt = vsq$. Protože také $v \neq 0$, zkrátíme a dostaváme $pt = sq$ neboli $\frac{p}{q} \sim \frac{s}{t}$. \square

Množina \mathbb{F} se tedy rozpadne na třídy ekvivalence. Každá třída ekvivalence obsahuje všechny zlomky, které geometricky vyjadřují jednu konkrétní délku neboli určitou kvantitu.

Definice 8d.4.

Definujeme množinu \mathbb{Q} racionálních čísel jako množinu všech tříd ekvivalence relace \sim , značíme $\mathbb{Q} = \{[f]; f \in F\}$.

Teď můžeme přesně říct, jak vlastně v praxi fungují racionální čísla. Když si vezmeme třeba $\frac{4}{6}$, tak ve skutečnosti myslíme na celou třídu ekvivalence, tedy jakoby bereme i ostatní zlomky ekvivalentní k $\frac{4}{6}$, protože víme, že z hlediska praxe je jedno, kterého zástupce dotyčné třídy ekvivalence si vybereme. To jsme ovšem ještě nedokázali, máme co dohánět. Zatím díky kapitole 4a víme, že když si pro dotyčnou třídu vybereme jiného zástupce, dá nám stejnou třídu ekvivalence, tedy dosáhneme z něj na stejně zlomky. Ještě jsme ale nedokázali, že si můžeme zástupce vybírat i v případech, kdy se zlomky provádí rozličné běžné věci jako porovnávání, operace a podobně. Nejprve si samozřejmě musíme tyto věci zadefinovat.

Poznámka doplňující: V řeči faktorových množin můžeme napsat $\mathbb{Q} = \mathbb{F}/\sim$.

8d.5 Porovnáváme zlomky

Jak porovnáváme zlomky v praxi? Porovnáme kvantity, které reprezentují. Abychom to udělali správně matematicky, musíme toto porovnávání vyjádřit čistě pomocí zúčastněných čísel a známých matematických operací, což ale není problém, například $\frac{3}{5} < \frac{7}{8}$ se dá hravě přepsat na $3 \cdot 8 < 7 \cdot 5$. Tím je inspirována definice. Je v tom ale háček, my jsme se od $\frac{3}{5} < \frac{7}{8}$ k $3 \cdot 8 < 7 \cdot 5$ dostali vynásobením, ale to funguje jen v případech, kdy jsou jmenovatelé kladní. Vyřešíme to jednoduše, prostě si vezmeme vhodné zástupce ze tříd.

Definice 8d.6.

Definujeme uspořádání $<$ na \mathbb{Q} následujícím předpisem: třídy $x, y \in \mathbb{Q}$ splňují $x < y$ právě tehdy, když existují $\frac{p}{q} \in x$ a $\frac{u}{v} \in y$ takoví, že $p > 0, v > 0$ a $pv < uq$.

Dále budeme namísto „ $p > 0, v > 0$ “ zkráceně psát jen $p, v > 0$.

Protože u každé třídy dokážeme vybrat zástupce $\frac{p}{q}$ tak, aby $q > 0$, vztahuje se tato definice na všechny třídy a dokážeme tedy vždy rozhodnout, zda je nějaká dvojice tříd v relaci nebo ne. Aby byla definice korektní, musíme ještě ukázat, že toto rozhodnutí nezáleží na volbě zástupců.

Frac 8d.7.

Nechť $\frac{p}{q}, \frac{u}{v}, \frac{P}{Q}, \frac{U}{V} \in \mathbb{F}$, předpokládejme, že $q, Q, v, V > 0$, $[\frac{p}{q}] = [\frac{P}{Q}]$ a $[\frac{u}{v}] = [\frac{U}{V}]$. Pak $[\frac{p}{q}] < [\frac{u}{v}]$ právě tehdy, když $[\frac{P}{Q}] < [\frac{U}{V}]$.

Důkaz (rutinní): Díky symetrii situace stačí ukázat jeden směr, druhý se dokazuje obdobně (zkuste si to).

Předpokládejme tedy, že $[\frac{p}{q}] < [\frac{u}{v}]$ neboli $pv < uq$, po vynásobení číslem $Q > 0$ dostáváme $pvQ < uqQ$.

Z předpokladu $[\frac{p}{q}] = [\frac{P}{Q}]$ máme $pQ = Pq$, po vynásobení číslem $v \neq 0$ dostáváme $pQv = PvQ$. Spojením s odvozenou nerovností máme $PvQ < uqQ$, po zkrácení číslem $q > 0$ máme $Pv < uQ$ a jsme v polovině cesty.

Obdrženou nerovnost vynásobíme číslem $V > 0$, dostáváme $PvV < uQV$. Z předpokladu $[\frac{u}{v}] = [\frac{U}{V}]$ máme $uV = Uv$, po vynásobení číslem $Q \neq 0$ vyjde $uVQ = UvQ$. Spojením dostáváme $PvV < UvQ$, a když zkrátíme číslem $v > 0$, máme $PV < UQ$ neboli $[\frac{P}{Q}] < [\frac{U}{V}]$.

Všimněte si, že jsme v důkazu silně závislí na předpokladu $q, Q, v, V > 0$.

□

Dokázali jsme, že naše definice porovnávání nezáleží na volbě zástupců (s kladnými druhými složkami) ze tříd ekvivalence. Umíme tedy porovnávat zlomky. K naší úplné spokojenosti chybí ještě dvě věci.

Definice 8d.8.

Definujeme relaci $>$ na \mathbb{Q} jako $<^{-1}$.

Definujeme relaci \leq na \mathbb{Q} předpisem $x \leq y$ právě tehdy, když $[x < y \text{ nebo } x = y]$.

Definujeme relaci \geq na \mathbb{Q} jako \leq^{-1} .

Teď už máme k dispozici všechna porovnávání, na která jsme zvyklí, a dokážeme, že fungují dle očekávání.

Fakt 8d.9.

Relace \leq a \geq jsou lineární částečná uspořádání na \mathbb{Q} .

Relace $<$ a $>$ jsou ostrá uspořádání na \mathbb{Q} .

Důkaz (z povinnosti): 1) $<$ je ostré uspořádání:

Asymetrie: Nechť $x, y \in \mathbb{Q}$ splňují $x < y$. Zvolme $\frac{p}{q} \in x$, $\frac{u}{v} \in y$ tak, aby $q, v > 0$. Předpoklad dává $pv < uq$, pak ovšem nemůže zároveň platit $pv < uq$ a tedy ani $y < x$.

Tranzitivita: Nechť $x, y, z \in \mathbb{Q}$ splňují $x < y$ a $y < z$. Zvolme $\frac{p}{q} \in x$, $\frac{u}{v} \in y$ a $\frac{s}{t} \in z$ tak, aby $q, v, t > 0$. Dle předpokladu pak $pv < uq$ a $ut < sv$. Vynásobíme první nerovnost číslem $t > 0$ a druhou číslem $q > 0$, vyjde $pvt < uqt$ a $utq < svq$ neboli $pvt < svq$. Vydělíme číslem $v > 0$ a dostáváme $pt < sq$, tedy $x < z$ neboli $x \leq z$.

2) Z části 1) a Věty 4b.5 vyplývá, že \leq je částečné uspořádání.

Linearita: Mějme libovolné $x, y \in \mathbb{Q}$. Zvolme zástupce $\frac{p}{q} \in x$, $\frac{u}{v} \in y$ tak, aby $q, v > 0$. Jestliže $pv = uq$, tak $x = y$ a proto dle definice $x \leq y$. Jinak máme dvě různá celá čísla pv, uq , proto buď $pv < uq$, pak $x < y$ a tedy i $x \leq y$, nebo $pv > uq$, což znamená $uq < pv$ neboli $y < x$ neboli $y \leq x$. Ukázali jsme, že prvky x, y jsou porovnatelné.

3) Z 1), 2) a Věty 3c.3 vyplývá, že i $>$ je ostré a \geq je částečné uspořádání, důkaz linearity je obdobný. \square

Doporučujeme čtenáři, aby si jako cvičení zkusil dokázat přímo, že \leq je částečné uspořádání.

8d.10 Počítáme se zlomky

V praxi zlomky sčítáme tak, že sečteme libovolně zvolené zástupce. Podobně funguje odčítání, násobení i dělení.

Definice 8d.11.

Nechť $x, y \in \mathbb{Q}$. Definujeme

$$x + y = \left[\frac{pv + uq}{qv} \right],$$

$$x - y = \left[\frac{pv - uq}{qv} \right],$$

$$x \cdot y = \left[\frac{pu}{qv} \right],$$

$$x/y = \left[\frac{pv}{qu} \right],$$

kde $\frac{p}{q} \in x$ a $\frac{u}{v} \in y$ jsou zvoleny libovolně.

I zde musíme začít důkazem, že výsledek operací nezáleží na volbě zástupců.

Věta 8d.12.

Nechť $\frac{p}{q}, \frac{u}{v}, \frac{P}{Q}, \frac{U}{V} \in \mathbb{F}$, předpokládejme, že $\left[\frac{p}{q} \right] = \left[\frac{P}{Q} \right]$ a $\left[\frac{u}{v} \right] = \left[\frac{U}{V} \right]$. Pak

$$\left[\frac{p}{q} \right] + \left[\frac{u}{v} \right] = \left[\frac{P}{Q} \right] + \left[\frac{U}{V} \right],$$

$$\left[\frac{p}{q} \right] - \left[\frac{u}{v} \right] = \left[\frac{P}{Q} \right] - \left[\frac{U}{V} \right],$$

$$\left[\frac{p}{q} \right] \cdot \left[\frac{u}{v} \right] = \left[\frac{P}{Q} \right] \cdot \left[\frac{U}{V} \right],$$

$$\left[\frac{p}{q} \right] / \left[\frac{u}{v} \right] = \left[\frac{P}{Q} \right] / \left[\frac{U}{V} \right].$$

Důkaz (z povinnosti): 1) Sčítání: Z předpokladu máme $pQ = Pq$ a $uV = Uv$. Pak

$(pv + uq)(QV) = pvQV + uqQV = (pQ)vV + (uV)qQ = PqvV + UvqQ = PVqv + UQqv = (PV + UQ)(qv)$, tedy $\frac{pv+uq}{qv} \sim \frac{PV+UQ}{QV}$. Proto $\left[\frac{pv+uq}{qv} \right] = \left[\frac{PV+UQ}{QV} \right]$.

2) Ostatní operace si čtenář jistě rád dokáže, je to snadné, ale únavné. \square

Již tradičně konstatujeme, že operace odčítání a dělení ve skutečnosti nejsou plnoprávné operace, ale jen zkratky pro jiné činnosti, jmenovitě pro výpočty $\left[\frac{p}{q}\right] + \left(-\left[\frac{u}{v}\right]\right)$ a $\left[\frac{p}{q}\right] \cdot \left[\frac{u}{v}\right]^{-1}$. Klíčové operace sčítání a násobení samozřejmě splňují všechna pravidla, na která jsme zvyklí, viz kapitola 8c.

Věta 8d.13.

Prostor $(\mathbb{Q}, +, \cdot, \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right])$ je těleso.

Důkaz (rutinní): Důkazy jsou snadné, vycházejí z toho, že v definicích operací pro \mathbb{Q} jsme použili tradiční operace pro reálná čísla, která všechny potřebné vlastnosti mají. Ukážeme dvě vlastnosti jako inspiraci, čtenář si dodělá zbytek.

Komutativita ščítání: Nechť $x, y \in \mathbb{Q}$, zvolme zástupce $\frac{p}{q} \in x, \frac{u}{v} \in y$. Pak díky komutativitě sčítání a násobení v \mathbb{R} dostáváme

$$x + y = \left[\frac{pv + uq}{qv} \right] = \left[\frac{vp + qu}{vq} \right] = y + x.$$

Jednotkový prvek pro sčítání a násobení: Označme $0_Q = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$ a $1_Q = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$. Pak pro libovolné $x \in \mathbb{Q}$ zvolíme zástupce $\frac{p}{q} \in x$ a počítáme

$$\begin{aligned} x + 0_Q &= \left[\frac{p \cdot 1 + 0 \cdot q}{q \cdot 1} \right] = \left[\frac{p}{q} \right] = x, \\ x \cdot 1_Q &= \left[\frac{p \cdot 1}{q \cdot 1} \right] = \left[\frac{p}{q} \right] = x. \end{aligned}$$

Obdobně $0_Q + x = x$ a $1_Q \cdot x = x$.

Invertibilita pro násobení: Nechť $x \in \mathbb{Q}$, $x \neq 0_Q$. Zvolme libovolného zástupce $\frac{p}{q} \in x$. Pak jistě $p, q \neq 0$ (jinak by $x = 0_Q$), můžeme tedy uvažovat $y = \left[\begin{smallmatrix} q \\ p \end{smallmatrix}\right]$. Dostáváme

$$x \cdot y = \left[\frac{pq}{qp} \right] = \left[\frac{1}{1} \right] = 1_Q.$$

Obdobně $y \cdot x = 1_Q$, proto je x invertibilní vůči násobení a $y = x^{-1}$. □

8d.14 Zlomky a celá/reálná čísla

Pro zlomky coby třídy jsme definovali srovnání a operace. My jsme ovšem zvyklí vnímat zlomky jako rozšíření celých čísel, jinak řečeno, celá čísla by měla být obsažena v racionálních a operace by si měly odpovídat. Podobně by si měly odpovídat operace na zlomcích s operacemi a porovnáním u reálných čísel, protože bychom rádi vnímali zlomky coby podmnožinu reálných čísel. Jak se toto dá zařídit? Začneme čísly celými.

Zde je myšlenka jednoduchá, ztotožníme si celá čísla se zlomky, které označují celočíselné kvantity, tedy se zlomky typu $\frac{p}{1}$. Pro takovéto ztotožnění se v matematice používá pojem zobrazení.

Definujme $T: \mathbb{Z} \mapsto \mathbb{Q}$ předpisem $T(z) = \left[\begin{smallmatrix} z \\ 1 \end{smallmatrix}\right]$. Množina \mathbb{Z} se tímto zobrazením spojí s množinou $Z = T[\mathbb{Z}] \subseteq \mathbb{Q}$, jinak řečeno, když ze všech celých čísel vyrobíme zlomky a uvažujeme jejich třídy ekvivalence, pak dostáváme jistou podmnožinu Z množiny \mathbb{Q} . Na této podmnožině pak máme srovnání a operace tak, jak jsme je definovali pro třídy ekvivalence.

Tvrdíme ovšem, že tato množina Z je věrnou kopí celých čísel \mathbb{Z} , takže vyjde nestejno, jestli pracujeme s čísly $x, y \in \mathbb{Z}$ nebo jejich kopiemi $\left[\begin{smallmatrix} z \\ 1 \end{smallmatrix}\right], \left[\begin{smallmatrix} y \\ 1 \end{smallmatrix}\right] \in \mathbb{Q}$. Matematicky se tím dostáváme k isomorfismům, viz kapitola 13. Nebudeme se na tento jazyk odvolávat a raději řekneme přesně a otevřeně, co tím míníme.

Věta 8d.15.

Zobrazení T je bijekce $\mathbb{Z} \mapsto Z$. Dále pro všechna $x, y \in \mathbb{Z}$ platí:

- (i) $x = y$ právě tehdy, když $T(x) = T(y)$;
- (ii) $x \leq y$ právě tehdy, když $T(x) \leq T(y)$;
- (iii) $x < y$ právě tehdy, když $T(x) < T(y)$;
- (iv) $T(x+y) = T(x) + T(y)$;
- (v) $T(x-y) = T(x) - T(y)$;
- (vi) $T(x \cdot y) = T(x) \cdot T(y)$;
- (vii) $T(x/y) = T(x)/T(y)$.

Čtenář jistě vidí, že na levé straně jsou srovnání/operace ze \mathbb{Z} a na pravé z \mathbb{Q} coby množiny tříd ekvivalence. Vlastnosti (i) až (iii) říkají, že když porovnáváme dvě celá čísla, pak vyjde náležitost, jestli je porovnáváme jako celá čísla, nebo si porovnáme jejich kopie v \mathbb{Q} „po zlomkovsku“ dle naší definice.

U vlastností (iv) až (vii) pomůže, když si je vyjádříme ještě jinak. Protože T je bijekce, máme k dispozici i inverzní zobrazení $T^{-1}: Z \mapsto \mathbb{Z}$. Když jej aplikujeme třeba na (iv), dostáváme $T^{-1}(T(x+y)) = T^{-1}(T(x) + T(y))$ neboli $x+y = T^{-1}(T(x) + T(y))$. Jaká je interpretace? Jestliže chceme sečít dvě celá čísla, tak to můžeme udělat přímo, nebo si představíme, že jsou to zlomky, sečteme je „po zlomkovsku“ a výsledek pak zase vezmeme jako celé číslo, vyjde to náležitost.

Všimněte si mimochodem, že jsme při aplikování T^{-1} na třídu $T(x) + T(y) \in \mathbb{Q}$ mlčky předpokládali, že to jde, tedy že ten součet je zase z množiny Z . To vůbec není samozřejmé a musí se to u všech operací dokázat. Jazykem binárních operací má platit, že množina Z je uzavřená na sčítání, násobení a přechod k opačným a inverzním prvkům.

Důkaz (z povinnosti, možná poučný):

$$(i): x = y \iff x \cdot 1 = y \cdot 1 \iff \left[\begin{matrix} x \\ 1 \end{matrix} \right] = \left[\begin{matrix} y \\ 1 \end{matrix} \right] \iff T(x) = T(y).$$

(iii): Protože $1 > 0$, je $\frac{x}{1}$ vhodným zástupcem pro $\left[\begin{matrix} x \\ 1 \end{matrix} \right]$ při porovnávání, podobně u y . Proto

$$x < y \iff x \cdot 1 < y \cdot 1 \iff \left[\begin{matrix} x \\ 1 \end{matrix} \right] < \left[\begin{matrix} y \\ 1 \end{matrix} \right] \iff T(x) < T(y).$$

(ii): Je to kombinace (i) a (iii).

$$(iv): T(x+y) = \left[\begin{matrix} x+y \\ 1 \end{matrix} \right] = \left[\begin{matrix} x \cdot 1 + y \cdot 1 \\ 1 \cdot 1 \end{matrix} \right] = \left[\begin{matrix} x \\ 1 \end{matrix} \right] + \left[\begin{matrix} y \\ 1 \end{matrix} \right] = T(x) + T(y).$$

Při čtení zprava vidíme, že třída $T(x) + T(y)$ je dána zástupcem $\frac{x+y}{1}$, patří tedy zase do Z , množina Z je tedy uzavřená na sčítání.

Zbytek důkazu je obdobný.

□

Shrneme-li naše poznatky, tak množina \mathbb{Z} a její kopie Z mají i stejnou algebraickou strukturu, lze je tedy z praktického pohledu považovat za totéž. Více jsme o tom psali v kapitole 13.

Zapojení zlomků mezi reálná čísla probíhá podobně, v rámci reálných čísel si vytvoříme kopii našeho \mathbb{Q} pomocí přirozeného kandidáta, zobrazení $\left[\begin{matrix} p \\ q \end{matrix} \right] \mapsto (p/q)$. Pak se ukáže obdoba věty výše o porovnání a operacích. Nebudeme to zde dělat, je to rutinní práce a důkazy jsou obdobné. Abychom knihu nenatahovali, necháme to zvídavému (a pilnému) čtenáři, cíl kapitoly už byl dosažen: Nahlédli jsme do matematické kuchyně a viděli jsme některé probrané pojmy v akci.

9. Posloupnosti a součty, řady

Posloupnosti se tradičně studují v matematické analýze, nám se ale budou některé pojmy a výsledky hodit i zde. Uvedeme si proto poznatky relevantní pro diskrétní matematiku, ale často je jen zacitujeme, protože některé důkazy by bez analytických metod byly těžké až nemožné. Tam kde to rozumně jde, důkazy provedeme, ale čtenáři určitě pomůže, když už bude něco z analýzy znát. Na druhou stranu tady o posloupnostech probereme věci, které se v typických kursech diferenciálního počtu nedělají, ale velice se hodí v computer science.

9a. Posloupnosti

Posloupnosti jsou užitečným vyjadřovacím nástrojem v situacích, kdy nám přicházejí čísla jedno po druhém a je potřeba je zpracovat. Začneme formální definicí, ale rychle od ní utečeme k užitečnější představě.

! Definice.

Posloupnost je libovolné zobrazení z nějaké množiny $\{n_0, n_0 + 1, n_0 + 2, \dots\}$ do \mathbb{R} , kde $n_0 \in \mathbb{Z}$.

By a **sequence** we mean any mapping from a set $\{n_0, n_0 + 1, n_0 + 2, \dots\}$ into \mathbb{R} , where $n_0 \in \mathbb{Z}$.

! Takto se posloupnosti definují tadičně, protože se potřebujeme odkázat na nějakou známou matematickou strukturu. V praxi je ovšem chápeme jinak.

Vezměme si nějaké takové zobrazení T . U posloupností jsou podstatné především hodnoty, což jsou čísla $T(n_0)$, $T(n_0 + 1)$, $T(n_0 + 2)$, ... Takže ten správný pohled na posloupnost je, že jde o reálná čísla jdoucí jedno za druhým (pořadí je důležité), bude jednodušší si je prostě značit jako $a_{n_0}, a_{n_0+1}, a_{n_0+2}, \dots$

Budeme tedy posloupnosti zapisovat jako $\{a_k\}_{k=n_0}^{\infty}$, někdy jen $\{a_k\}$. Není to množina, ale uspořádaná množina, což znamená, že na pořadí záleží a mohou se opakovat prvky. Jedno a_k značí **člen** posloupnosti.

Někteří autoři pro zvýraznění uspořádanosti používají značení $(a_k)_{k=n_0}^{\infty} = (a_{n_0}, a_{n_0+1}, a_{n_0+2}, \dots)$, které pěkně naznačuje, že jde do značné míry o zobecnění vektorů na situaci s nekonečně mnoha souřadnicemi, ale zdá se, že je méně rozšířené, proto se budeme držet složených závorek.

! **Příklad 9a.a:** Uvažujme posloupnost danou formálně $T(n) = (-1)^n$ pro $n \geq 0$. Její hodnoty jsou $T(0) = (-1)^0 = 1$, $T(1) = (-1)^1 = -1$, $T(2) = (-1)^2 = 1$, ..., takže je to posloupnost $\{1, -1, 1, -1, 1, \dots\}$. Tuto posloupnost jsme zadali jako zobrazení, abychom viděli, jak by se to dělalo, ale normálně by se zadala jinak. Možností je více, nejtypičtější je předpis $\{(-1)^k\}_{k=0}^{\infty}$, někdy se používá i předpis „ $a_k = (-1)^k$ pro $k \geq 0$ “, pak tedy $a_0 = 1$, $a_1 = -1$, $a_2 = 1$, ...

Této posloupnosti se říká **alternující posloupnost**.

△

! Když posloupnost zadáme předpisem pro její k -tý člen, říkáme tomu **explicitní definice** posloupnosti. Níže ještě ukážeme definici rekurzí. Je důležité poznámenat, že u posloupnosti jsou důležité hodnoty, takže na zápisu až tak nezáleží. Asi nepřekvapí, že zápis $\{(-1)^k\}_{k=0}^{\infty}$, $\{(-1)^i\}_{i=0}^{\infty}$ nebo třeba $\{(-1)^n\}_{n=0}^{\infty}$ dávají totéž, ale flexibilita zápisu jde ještě dále.

Podívejme se na následující posloupnosti:

$$a_n = 2n + 13, n \geq -6 \text{ dává posloupnost } \{1, 3, 5, 7, \dots\}, \text{ zatímco}$$

$b_n = 2n - 1, n \geq 1$ dává posloupnost $\{1, 3, 5, 7, \dots\}$, čili je to tatáž posloupnost, i když vzorce vypadají jinak. Můžeme si tedy dovolit vzoreček pro posloupnost různě měnit, hlavní je, aby hodnoty zůstávaly stejné. Často bývá jeden určitý zápis lepší z hlediska praktických výpočtů. Tím se vracíme k tomu, že posloupnost je určena svými členy, vzoreček je jen pohodlný způsob, jak ty členy určit, ale není tím podstatným.

Často se posloupnosti „zadají“ tím, že se napíše několik prvních členů a pak se předpokládá, že posloupnost pokračuje „stejným způsobem“. My budeme částečný výpis členů používat jen pro ilustrační účely, protože nám například $\{2, 4, 6, 8, 10, \dots\}$ či $\{1, -1, 1, -1, 1, \dots\}$ umožňují získat o dotyčné posloupnosti dobrou představu, ale nebudeme takto posloupnosti definovat. Ono totiž není jasné, co to je ten „stejný způsob“.

! **Příklad 9a.b:** Uvažujme posloupnost začínající $\{-6, 6, -6, 6, \dots\}$. Jaká je to posloupnost? Uvažujme následující vzorce:

- $a_k = 6 \cdot (-1)^k$ pro $k = 1, 2, \dots$ neboli $\{6 \cdot (-1)^k\}_{k=1}^{\infty}$ dává $\{-6, 6, -6, 6, -6, 6, \dots\}$. Toto si asi lidé představí jako „stejně pokračování“.

Je ovšem také možné použít třeba $a_k = 6 \cdot (-1)^k$ pro $k = 13, 14, 15, \dots$, dostáváme stejnou posloupnost, jen jiným zápisem.

- $b_k = 8k^3 - 12k^2 - 8k + 6$ pro $k = -1, 0, 1, \dots$ neboli $\{8k^3 - 12k^2 - 8k + 6\}_{k=-1}^{\infty}$ dává $\{-6, 6, -6, 6, 90, 294, \dots\}$. I toto je přirozené pokračování, protože z matematického pohledu není polynom nijak horší než mocnina $6 \cdot (-1)^k$.
- $c_k = \begin{cases} -6(k/2)!, & k \text{ sudé;} \\ 6, & k \text{ liché} \end{cases}$ pro $k = 0, 1, 2, \dots$. Pak máme $\{-6, 6, -6, 6, -12, 6, \dots\}$.

Dá se vymyslet mnoho vzorečků, které dávají různé posloupnosti začínající stejně, členy $-6, 6, -6, 6$. Z matematického hlediska je tedy oblibená otázka z testů intelligence „jaký je další člen“ poněkud srovnání, z pohledu posloupností je správných řešení mnoho.

Pro naši teorii z toho vyplývá jednoznačný závěr, že definovat posloupnosti vypsáním „začátku“ není možné.

△

! Pro účely diskrétní matematiky se vyplatí definovat také **konečné posloupnosti** $\{a_k\}_{k=n_0}^{m_0}$. Ty pak slouží jako jeden z možných matematických modelů pro řetězce znaků čili slova. Vzhledem k tomu, že u počítačů pracujeme výhradně s konečnými řetězci znaků, má vůbec smysl zabývat se nekonečnými posloupnostmi? Ano, ze dvou důvodů. Za prvé, jsou situace, kdy nevíme dopředu, kolik znaků budeme mít ke zpracování. Je pak možné pracovat s množinou konečných posloupností všech možných délek, ale prodloužením na nekonečné posloupnosti si někdy ušetříme formální komplikace například u operací.

Druhý důvod je ten, že nás často zajímá otázka dlouhodobých trendů, pak nám přechod na nekonečné posloupnosti nabízí mocně nástroje.

Je evidentní, že představa posloupnosti jako nekonečného sledu objektů je velice obecná, například asi bude užitečné uvažovat posloupnosti znaků nějaké abecedy. Naše nástroje však fungují hlavně na posloupnosti s číselnými členy, na ty se tu zaměříme. Začneme tím, že si představíme několik základních posloupností, které se často vyskytují.

Příklad 9a.c: Jedna z nejslavnějších posloupností je ta, která se v Evropě objevila v roce 1202 v knize Liber abaci čili Kniha o počítání, díky které se do Evropy dostal indo-arabský systém zápisu čísel. Její autor, snad nejlepší středověký matematik zvaný Fibonacci, se v ní zmínil také o posloupnosti zadáne následujícím induktivním předpisem:

- (0) $F_1 = 1, F_2 = 1$.
- (1) Pro $n \geq 2$ je $F_{n+1} = F_n + F_{n-1}$.

Někdy se ještě dává $F_0 = 0$. Této posloupnosti se říká Fibonacciho posloupnost (a jejím členům Fibonacciho čísla, $1, 1, 2, 3, 5, 8, 13, 21, \dots$). V kapitole o rekurzivních vztazích (viz příklad) si povíme víc o tom, jak k ní přišel, zmínu si ale zaslouží, že již před 6. stoletím n.l. ji používali hindští učenci při práci s přízvuky v poezii.

Rekurzivní zadávání posloupností je velice populární a z kapitoly o indukci víme, že je to korektní způsob. Pro praktické použití ale bývá často nepříjemný a dáváme přednost přepisu do explicitního vyjádření. Problémem je, jak jej najít, ne vždy to umíme, někdy to dokonce ani není možné. Částečně se tomu budeme věnovat právě v kapitole o rekurentních vztazích. Pro Fibonacciho posloupnost tam odvodíme vzorec, jehož správnost teď dokážeme, jak jinak než indukcí, jmenovitě její modifikovanou verzí s návratem o dva kroky.

Tvrdíme, že vzorec $f_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n$ splňuje podmínky (0) a (1).

$$(0): f_1 = \frac{1}{\sqrt{5}} \frac{1+\sqrt{5}}{2} - \frac{1}{\sqrt{5}} \frac{1-\sqrt{5}}{2} = 1 = F_1;$$

$$f_2 = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^2 - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^2 = \frac{1}{\sqrt{5}} \frac{6+2\sqrt{5}}{4} - \frac{1}{\sqrt{5}} \frac{6-2\sqrt{5}}{4} = 1 = F_2.$$

(1) Zvolíme libovolné $n \geq 2$ a předpokládáme, že F_n a F_{n-1} jsou opravdu dány příslušným vzorcem, tedy $F_n = f_n$ a $F_{n-1} = f_{n-1}$. Pak podle definice F_{n+1} a indukčního předpokladu dostaneme

$$\begin{aligned} F_{n+1} &= F_n + F_{n-1} = f_n + f_{n-1} = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n + \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{n-1} - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^{n-1} \\ &= \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{n-1} \left[\frac{1+\sqrt{5}}{2} + 1 \right] - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^{n-1} \left[\frac{1-\sqrt{5}}{2} + 1 \right] = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{n-1} \frac{3+2\sqrt{5}}{2} - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^{n-1} \frac{3-2\sqrt{5}}{2} \\ &= \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{n-1} \frac{6+4\sqrt{5}}{2} - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^{n-1} \frac{6-4\sqrt{5}}{2} = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{n-1} \left(\frac{1+\sqrt{5}}{2} \right)^2 - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^{n-1} \left(\frac{1-\sqrt{5}}{2} \right)^2 \\ &= \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} = f_{n+1}. \end{aligned}$$

Uf. Takže to funguje.

Ještě se k této posloupnosti vrátíme, viz příklad či příklad , také příklad .

△

!

Definice.

Uvažujme posloupnost $\{a_k\}_{k=n_0}^{\infty}$.

Řekneme, že je to **aritmetická posloupnost (arithmetic sequence)**, jestliže existují $a, d \in \mathbb{R}$ tak, aby platilo $a_k = a + dk$ pro všechna $k = n_0, n_0 + 1, \dots$

Řekneme, že je to **geometrická posloupnost (geometric sequence)** jestliže existují $a, q \in \mathbb{R}$ tak, aby platilo $a_k = aq^k$ pro všechna $k = n_0, n_0 + 1, \dots$

Příklad 9a.d: Posloupnost $\{1, 3, 5, 7, 9, \dots\}$ všech lichých přirozených čísel je aritmetická, protože se dá zapsat jako $a_k = 2k + 1$ pro $k \in \mathbb{N}_0$.

Konstantní posloupnost $\{1, 1, 1, 1, \dots\}$ je aritmetická, protože se dá zapsat jako $a_k = 0k + 1$ pro $k \in \mathbb{N}_0$. Je ovšem také geometrická, protože se dá zapsat jako $a_k = 1^k$ pro $k \in \mathbb{N}_0$.

Alternující posloupnost z úplně prvního příkladu je také geometrická.

△

Bývá tradiční indexovat tyto posloupnosti od nuly. Nikterak se tím neomezujeme, každou aritmetickou a geometrickou posloupnost lze takto upravit. Například posloupnost $\{13^7, 13^8, 13^9, \dots\}$ je přirozené zapsat jako $a_k = 13^k$ pro $k \geq 7$ a v mnoha situacích to bude nejlepší. Jsou ale situace (v kapitole), kdy potřebujeme indexování od nuly, pak použijeme třeba toto: Číslo $a = 13^7$ se dá vytknout ze všech členů, pak vidíme, že naší posloupnost popisuje také vzorec $b_k = 13^7 \cdot 13^k$ pro $k \in \mathbb{N}_0$.

Obecně lze psát u aritmetické posloupnosti $a_k = (a + n_0d) + d(k - n_0)$, u geometrické zase $a_k = (aq^{n_0})q^{k-n_0}$ a zavedením nového indexu $n = k - n_0$ už indexujeme od nuly.

Jak už jsme naznačili, v mnoha situacích nezáleží na začátku indexace, pak ji nebudeme uvádět a píšeme $\{a_k\}_k$, popřípadě jen $\{a_k\}$ (když ve vzorci nebude jiné písmenko, takže bude index jasný). Budeme používat frázi „pro všechna k “ a myslit tím „pro všechna k z množiny indexů dotyčné posloupnosti“.

Aritmetické a geometrické posloupnosti poznáme snadno.

!

Fakt 9a.1.

Uvažujme posloupnost $\{a_k\}_k$.

Je to geometrická posloupnost právě tehdy, když existuje $q \in \mathbb{R}$ takové, že $\frac{a_{k+1}}{a_k} = q$ pro všechna k .

Je to aritmetická posloupnost právě tehdy, když existuje $d \in \mathbb{R}$ takové, že $a_{k+1} - a_k = d$ pro všechna k .

Například lichá čísla mají mezi sebou vždy rozdíl 2, proto tvoří aritmetickou posloupnost. Když je ale postupně dělíme, $\frac{3}{1}, \frac{5}{3}, \frac{7}{5}, \dots$, tak nedostáváme totéž, není to proto posloupnost geometrická. Naopak v tom příkladě s 13^k máme vždy $\frac{13^8}{13^7} = 13$, $\frac{13^9}{13^8} = 13$, $\frac{13^{10}}{13^9} = 13, \dots$, je to tedy posloupnost geometrická.

Tento fakt nás inspiruje k následujícím ekvivalentním definicím:

Rekurzivní definice aritmetické posloupnosti:

- (0) $a_0 = a$.
- (1) $a_{k+1} = a_k + d$ pro $k \in \mathbb{N}_0$.

Rekurzivní definice geometrické posloupnosti:

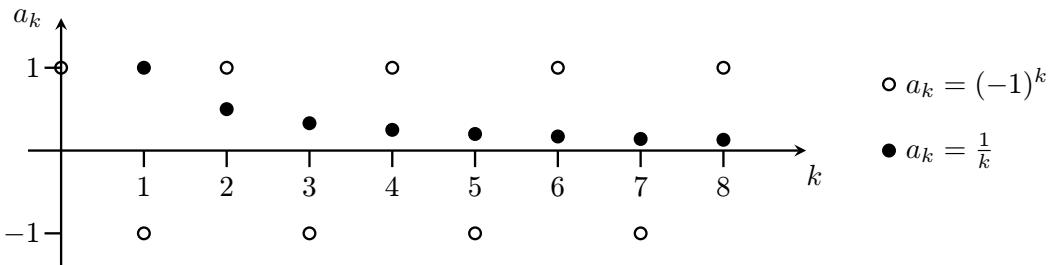
- (0) $a_0 = a$.
- (1) $a_{k+1} = a_k \cdot q$ pro $k \in \mathbb{N}_0$.

Snadno dokážeme indukcí, že tato definice souhlasí s naší původní definicí. Ukážeme to v rámci tréninku pro tu geometrickou, přesně řečeno dokážeme pro všechna $k \in \mathbb{N}_0$ vlastnost $V(k)$: k -tý člen takto rekurzivně dané posloupnosti splňuje $a_k = a_0q^k$.

(0) $k = 0$: $a_0 = a_0 \cdot q^0$.
 (1) Nechť $k \in \mathbb{N}_0$, předpokládáme $a_k = a_0q^k$. Pak podle rekurzivní definice platí $a_{k+1} = a_k \cdot q = a_0q^k \cdot q = a_0q^{k+1}$. Důkaz hotov.

! Další důležité typy posloupností jsou posloupnosti s mocninami $\{k^a\}$, například $\{k^3\}_{k=1}^{\infty} = \{1, 8, 27, 64, 125, 216, \dots\}$, a faktoriál jako posloupnost $\{k!\}_{k=1}^{\infty} = \{1, 2, 6, 24, 120, 720, \dots\}$.

Někdy pomůže si posloupnosti znázornit, jsou to vlastně zobrazení, čili máme k dispozici klasický graf. Na následujícím obrázku ukazujeme posloupnosti $\{(-1)^k\}_{k=0}^{\infty}$ a $\{\frac{1}{k}\}_{k=1}^{\infty}$.



Vlastnosti, které dále probereme, jsou na takovém grafu pěkně vidět.

U posloupností se dá studovat mnoho vlastností. Z hlediska praktického použití je docela zajímavá monotonie. Porovnáváme každý člen posloupnosti s tím bezprostředně následujícím, tedy členem a_k s členem a_{k+1} .

! Definice.

Uvažujme posloupnost $\{a_k\}$. Řekneme, že je tato posloupnost

- **rostoucí (increasing)**, jestliže $a_k < a_{k+1}$ pro všechna k ;
- **klesající (decreasing)**, jestliže $a_k > a_{k+1}$ pro všechna k ;
- **neklesající (non-decreasing)**, jestliže $a_k \leq a_{k+1}$ pro všechna k ;
- **nerostoucí (non-increasing)**, jestliže $a_k \geq a_{k+1}$ pro všechna k ;
- **monotonní (monotone)**, jestliže splňuje jednu z vlastností výše;
- **ryze monotonní (strictly monotone)**, jestliže je rostoucí nebo klesající.

Základní pojmy jsou rostoucí a klesající, které nutí posloupnost jít buď pořád nahoru, nebo pořád dolů (proto je v definici obecný kvantifikátor, kontrolujeme všechny dvojice, aby nám někde posloupnost neposkočila špatným směrem). Druhé dva pojmy rovněž nutí posloupnost jít stále jedním směrem, ale (někdy) také může zůstat stejná. Podíváme-li se na naše příklady, tak jsou skoro všechny monotonní, buď rostoucí (lichá čísla, 13^k) nebo klesající (viz $\frac{1}{k}$), měli jsme i konstantní posloupnost, která je zároveň nerostoucí i neklesající. Jedinou výjimkou je hned první příklad alternující posloupnosti, ta monotonní není.

Příklad 9a.e: Vyšetříme monotonii posloupnosti $\{\frac{1}{2k-1}\}_{k=1}^{\infty}$. První členy jsou $\{1, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \frac{1}{9}, \dots\}$, máme tedy podezření, že klesá. Dokážeme, že posloupnost $\{\frac{1}{2k-1}\}_{k=1}^{\infty}$ je klesající, podle definice.

Vezměme libovolné $k \in \mathbb{N}$. Platí $a_k > a_{k+1}$? Vyzkoušíme:

$$a_k > a_{k+1} \iff \frac{1}{2k-1} > \frac{1}{2(k+1)-1} \iff \frac{1}{2k-1} > \frac{1}{2k+1} \iff 2k+1 > 2k-1 \iff 2 > 0,$$

což je pravda. Všechny kroky v této úvaze jsou ekvivalentní, je tedy možné zpětně z pravdivého $2 > 0$ odvodit $a_k > a_{k+1}$, což dokazuje, že daná posloupnost je opravdu klesající. Mimořádne, všimněte si, že v důkazu jsme použili kladnost k , bez této znalosti by nešlo přejít od $\frac{1}{2k-1} > \frac{1}{2k+1}$ k $2k+1 > 2k-1$.

△

Poznamenejme, že v definici monotonie se testuje pro všechna k , čili to je zrovna případ, kdy záleží na tom, kde s indexy začneme. Například posloupnost $\{k^2 - 5k + 5\}_{k=1}^{\infty} = \{1, -1, -1, 1, 5, 11, 19, \dots\}$ není monotonní, protože na začátku klesne (tudíž nemůže být rostoucí či neklesající), a později zase vzroste (takže nemůže být ani klesající či nerostoucí). U stejněho vzorce ale můžeme začít později, pak $\{k^2 - 5k + 5\}_{k=2}^{\infty}$ už je neklesající a $\{k^2 - 5k + 5\}_{k=3}^{\infty}$ je rostoucí.

! Dalším důležitým pojmem je limita. Odpovídá na otázku, co se s členy posloupnosti děje, když po ní jdeme stále dál a dál. Například u posloupnosti $\{\frac{1}{k}\}_{k=1}^{\infty}$ se zdá, že se členy zmenšují k nule, zatímco u posloupnosti lichých čísel se členy stále bez omezení zvětšují, intuitivně bychom řekli, že jdou do nekonečna. Vymyslíme si pojmy, které takovéto chování vystihnu. Posloupnost jde do nekonečna, pokud dříve či později vyleze nad libovolně velkou zvolenou mez a zůstane nad ní. Jde k nule, jestliže pokaždé, když si zvolíme nějakou malou toleranci okolo nuly, tak do ní ta posloupnost dřív či později vleze a zůstane tam, blízko u nuly. Všimněte si, že teď nás vlastně ani nezajímá, co posloupnost dělá na začátku, ptáme se na její „konec“. Můžeme si tedy dovolit vynechat specifikaci, kde index začíná. Formální definice je kapku technická, ale jen upřesňuje to, co jsme si zde nastínili.

!

Definice.

Nechť $\{a_k\}$ je posloupnost.

Řekneme, že tato posloupnost jde do nekonečna, popřípadě že má limitu nekonečno, značeno $\lim(a_k) = \infty$ popřípadě $a_k \rightarrow \infty$, jestliže

pro každé $K > 0$ existuje k_0 tak, aby $a_k > K$ pro všechna $k \geq k_0$.

Řekneme, že tato posloupnost jde k nule, popřípadě že konverguje k nule, popřípadě že má limitu rovnou nule, značeno $\lim(a_k) = 0$ popřípadě $a_k \rightarrow 0$, jestliže

pro každé $\varepsilon > 0$ existuje k_0 tak, aby $|a_k| < \varepsilon$ pro všechna $k \geq k_0$.

Let $\{a_k\}$ be a sequence.

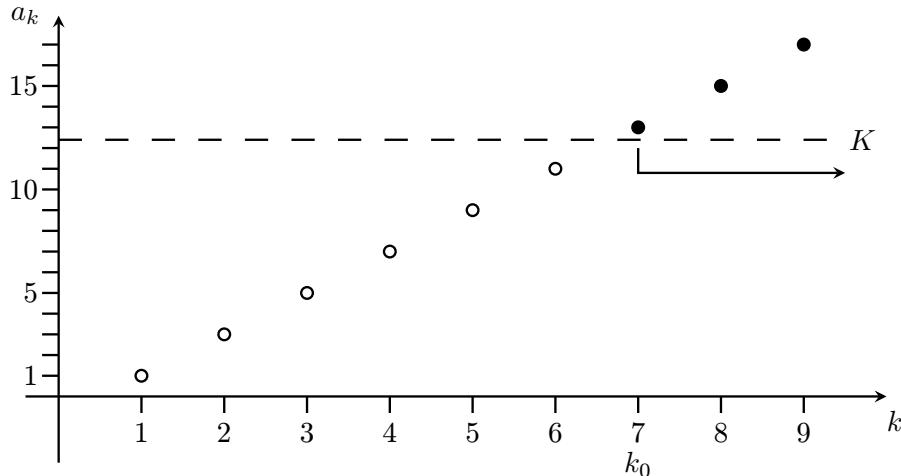
We say that it goes to infinity, or that its limit is infinity, denoted $\lim(a_k) = \infty$ or $a_k \rightarrow \infty$, if

for all $K > 0$ there is k_0 such that $a_k > K$ for all $k \geq k_0$.

We say that it goes to zero, or that it converges to zero, or that its limit is zero, denoted $\lim(a_k) = 0$ or $a_k \rightarrow 0$, if

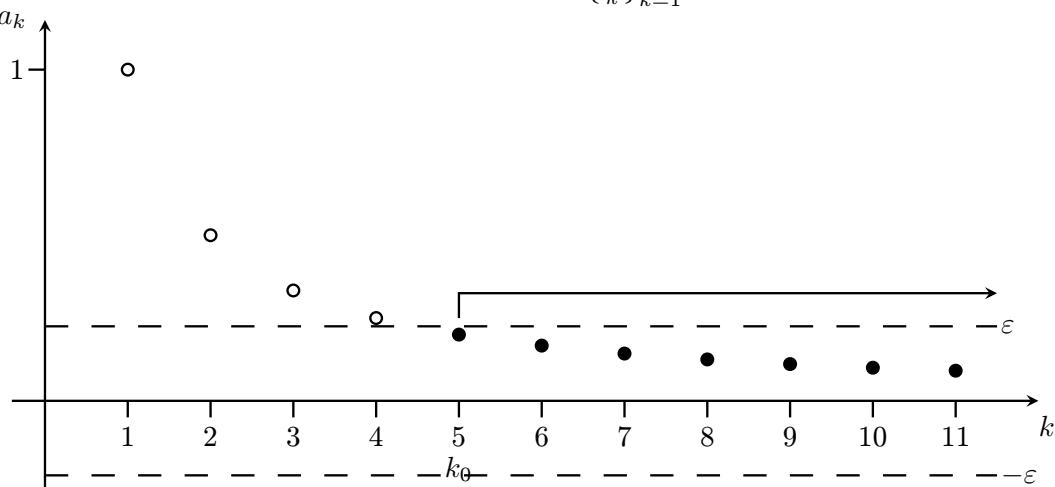
for all $\varepsilon > 0$ there is k_0 so that $|a_k| < \varepsilon$ for all $k \geq k_0$.

Ukážeme si význam definice na obrázku, nejprve nekonečnou limitu na příkladu posloupnosti $\{2k-1\}_{k=1}^{\infty}$ (kladná lichá čísla).



Kdykoliv nám někdo zadá hodnotu K (to je hladina, kterou máme překonat a již nad ní zůstat), dokážeme najít „odřezávací index“ k_0 takový, že pokud ignorujeme část posloupnosti před tímto indexem, tak již její zbývající členy zůstanou nad K .

Tedž si ilustrujeme nulovou limitu na příkladu posloupnosti $\{\frac{1}{k}\}_{k=1}^{\infty}$.



Kdykoliv nám někdo zadá hodnotu ε , tak se po nás chce splnění podmínky $|a_k| < \varepsilon$, tedy $-\varepsilon < a_k < \varepsilon$. Jinými slovy, hodnoty posloupnosti mají zůstat mezi hladinami ε a $-\varepsilon$, ovšem nikoliv všechny. Pro libovolně zadanou hodnotu ε musíme být schopni najít „odřezávací index“ k_0 takový, že pokud ignorujeme část posloupnosti před tímto indexem, tak již její zbývající členy zůstanou v cílové oblasti. Z obrázku se zdá, že to půjde. V definici je u epsilonu „prokáždítko“, takže bychom správně měli nechat protihráče, ať na nás zkouší všechny možné epsilony, ale zdá se, že stejně vždycky dokážeme odříznout začátek posloupnosti tak, aby její zbytek byl již v daném pruhu, jen pro hodně malá ε budeme muset odříznout větší začátek.

S definicí limity nebudeme příliš pracovat, bylo ale dobré se nad ní zamyslet, protože ilustruje důležitý princip. V mnoha situacích (například v příští kapitole) nás bude zajímat, jak daná posloupnost vypadá „na konci“ (v nekonečnu). Prakticky se to dělá tak, že ignorujeme začátek oné posloupnosti a na zbytku již dotyčná věc skoro platí (například posloupnost je skoro nula). Kolik posloupnosti odřízneme záleží na tom, jaké „skoro“ zrovna potřebujeme.

Čtenáře asi napadne, že konstantní posloupnost $\{1, 1, 1, \dots\}$ evidentně míří k 1, a definice limity se v analýze opravdu dělá obecněji, není třeba jít jen do nuly. Zde ale potřebujeme jen nulu a nekonečno, tak se na ně budeme soustředit. Je také zjevné, že ne každá posloupnost má limitu, například ta alternující posloupnost $\{1, -1, 1, -1, 1, -1\}$ na své cestě nikam nesměřuje a limitu nemá, to je život.

Ze zkušenosti s čísly si asi tipneme, že $k \rightarrow \infty$, $k^2 \rightarrow \infty$, $13^k \rightarrow \infty$, ale naopak $\frac{1}{k} \rightarrow 0$ či $\frac{1}{13^k} \rightarrow 0$. Pokud v tom cítíte jakousi dualitu mezi nekonečnem a nulou, tak máte opravdu.

Fakt 9a.2.

Pro každou posloupnost $\{a_k\}$ platí: $|a_k| \rightarrow \infty$ právě tehdy, když $\frac{1}{a_k} \rightarrow 0$.

Ekvivalentně, $a_k \rightarrow 0$ právě tehdy, když $\frac{1}{|a_k|} \rightarrow \infty$.

To známe; když dělíme malinkými čísly, dostáváme veliké výsledky, a také naopak. V další části pro nás bude důležité umět limity nejpopulárnějších posloupností, což je většinou snadné. Všimněte si, že $2^k \rightarrow \infty$, zatímco $(\frac{1}{2})^k = \frac{1}{2^k} \rightarrow 0$. U geometrické posloupnosti tedy záleží na velikosti základu. Uděláme si oficiální tvrzení.

Fakt 9a.3.

- (i) Nechť $a > 0$. Pak $k^a \rightarrow \infty$.
- (ii) Jestliže $q > 1$, pak $q^k \rightarrow \infty$.
Jestliže $|q| < 1$, pak $q^k \rightarrow 0$.
- (iii) $k! \rightarrow \infty$.
- (iv) $k^k \rightarrow \infty$.
- (v) Nechť $b > 0$. Pak $[\ln(k)]^b \rightarrow \infty$.

Vidíme, že většina výrazů utíká do nekonečna, v příští kapitole je budeme mezi sebou navzájem porovnávat a ptát se, kdo utíká do nekonečna rychleji. Nejprve ale jedna poznámka.

Ve Faktu jsme použili přirozený logaritmus, ale v computer science se často dává přednost logaritmům jiným, třeba dvojkovému. I pro ně platí tvrzení z Faktu, protože máme přepis $\log_a(k) = \frac{1}{\ln(a)} \ln(k)$. Pak také máme rovnost $[\log_a(k)]^b = \frac{1}{\ln^b(a)} [\ln(k)]^b$, takže se při změně základu jen modifikuje rychlosť růstu konstantou, při libovolném základě $a > 1$ logaritmus pořád utíká do nekonečna.

Cvičení

Cvičení 9a.1 (rutinní): Jsou dány následující posloupnosti rozličnými předpisy pro $k \in \mathbb{N}$. Najděte vždy prvních řekněme deset členů.

- (i) (0) $a_1 = 1, a_2 = -1, a_3 = 1, \dots$ (1) $a_{k+1} = a_k - a_{k-2}$;
- (ii) $a_k = \lfloor \sqrt{k} \rfloor$;
- (iii) a_k je počet písmen v číslovce označující k ;
- (iv) a_k je největší celé číslo, jehož binární zápis má k bitů;
- (v) a_k je největší n takové, že $n! \leq k$.

Cvičení 9a.2 (dobré, poučné): Najděte alespoň tři různá pravidla pro definici posloupnosti tak, aby její první tři členy byly

- (i) 1, 2, 4; (ii) 1, 2, 3.

Pro každé takové pravidlo dopočítejte další dva členy.

Cvičení 9a.3 (dobré): Pro následující seznamy celých čísel najděte jednoduchý vzorec či pravidlo, který je generuje, a pomocí něj odhadněte další člen.

- | | |
|--|--|
| (i) 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, ...; | (ix) 3, 6, 11, 18, 27, 38, 51, 66, 83, 102, ...; |
| (ii) 1, 2, 2, 3, 4, 4, 5, 6, 6, 7, 8, 8, ...; | (x) 1, 8, 27, 125, 216, ...; |
| (iii) 1, 0, 2, 0, 4, 0, 8, 0, 16, 0, ...; | (xi) 2, 3, 7, 25, 121, 721, 5041, 40321, ...; |
| (iv) 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, ...; | (xii) 2, 16, 54, 128, 250, 432, 686, ...; |
| (v) 1, -3, 9, -27, 81, -243, 729, ...; | (xiii) 1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010, 1011, ...; |

- (vi) $15, 8, 1, -6, -13, -20, -27, \dots$;
 (vii) $3, 6, 12, 24, 48, 96, 192, \dots$;
 (viii) $1, 4, 9, 16, 25, 36, 49, 64, \dots$;

- (xiv) $0, 2, 8, 26, 80, 242, 728, 2186, 6560, 19682, \dots$;
 (xv) $1, 3, 15, 105, 945, 10395, 125125, 2027025, 34459425, \dots$;
 (xvi) $2, 4, 16, 256, 65536, 4294967296, \dots$.

Cvičení 9a.4 (rutinní):

Odhadněte, které z následujících posloupností jsou monotonní, a svůj odhad dokažte.

- (i) $\left\{ \frac{k-1}{k+1} \right\}_{k=1}^{\infty}$; (iii) $\left\{ \frac{3}{k!} \right\}_{k=2}^{\infty}$; (v) $\left\{ \frac{2^{k-5}}{3^k} \right\}_{k=1}^{\infty}$;
 (ii) $\left\{ \frac{k-4}{2^k} \right\}_{k=1}^{\infty}$; (iv) $\left\{ \frac{2^k}{k!} \right\}_{k=1}^{\infty}$; (vi) $\{2k + (-1)^k\}_{k=1}^{\infty}$.

Cvičení 9a.5 (drsně²): Dokažte, že posloupnost $1, 2, 2, 3, 3, 3, 4, 4, 4, 4$ (každé k je přesně k -krát) je dána vzorcem $a_n = \lfloor \sqrt{2n} + \frac{1}{2} \rfloor$.

Řešení:

9a.1: (i): $1, -1, 1, 0, 1, 0, -1, -1, 0, 1, 2, 2, 1, -1, -3, -4, -3, 0, 4, 7, 7, 3, -4, -11, -14, -10, 1, 15$.

Nevidím pro tohle nějaký rozumný vzoreček, je to zajímavá posloupnost. Vzorec ale určitě existuje, viz kapitola .

(ii): $1, 1, 1, 2, 2, 2, 2, 2, 3, 3$. (iii): $5, 3, 3, 5, 3, 4, 4, 3, 5, 5$.

(iv): $1, 3, 7, 15, 31, 63, 127, 255, 511, 1023$. (v): $1, 2, 2, 2, 3, 3, 3, 3, 3, 3$.

9a.2: (i): Tak třeba A) $a_{k+1} = 2a_k$, pak $1, 2, 4, 8, 16$; B) $a_{k+1} = a_k + k$ pro $k \in \mathbb{N}$, pak $1, 2, 4, 7, 11$;

C) posloupnost všech přirozených čísel, která nejsou dělitelná třemi, pak $1, 2, 4, 5, 7$.

(ii): Tak třeba A) $a_k = k$ pro $k \in \mathbb{N}$, pak $1, 2, 3, 4, 5$, vzorec $a_{k+1} = a_k + 1$ dává totéž;

B) $a_{k+1} = a_k + a_{k-1}$, pak $1, 2, 3, 5, 8$; C) $a_k = k^3 + 2$ pro $k \geq -1$, pak $1, 2, 3, 10, 29$.

9a.3: (i): 1. (ii): 9. (iii): 32. (iv): 47, $a_{k+1} = a_k + 4$. (v): -2187 , $a_{k+1} = a_k \cdot (-3)$. (vi): -34 , $a_{k+1} = a_k - 7$. (vii):

384 , $a_{k+1} = 2a_k$, $a_k = 3 \cdot 2^{k-1}$. (viii): 81 , $a_{k+1} = k^2$. (ix): 123 , $a_{k+1} = a_k + 2k + 1$, tedy symbolicky $+3, +5, +7, \dots$

(x): 343 , $a_{k+1} = k^3$. (xi): 62881 , $a_{k+1} = k! + 1$. (xii): 1024 , $a_{k+1} = 2k^3$. (xiii): 1100 , $a_{k+1} = a_k + 1$, ale binárně!. (xiv): 59048 , $a_{k+1} = 3a_k + 2$. (xv): 654729075 , $a_{k+1} = a_k \cdot (2k + 1)$, symbolicky $\cdot 3, \cdot 5, \cdot 7, \dots$ (xvi): $1.844 \cdot 10^{19}$, $a_{k+1} = 2^{2^k}$.

9a.4: (i): rostoucí. Ekvivalentní úpravy pro libovolné $k \geq 1$:

$$a_k < a_{k+1} \iff \frac{k-1}{k+1} < \frac{k}{k+2} \iff (k-1)(k+2) < k(k+1) \iff k^2 + k - 2 < k^2 + k \iff -2 < 0 \text{ pravda.}$$

(ii): není monotonní. $a_1 = -\frac{3}{2}$, $a_2 = -\frac{1}{2}$, $a_1 < a_2$ vzroste tedy už nemůže být klesající nebo nerostoucí.

$a_6 = \frac{1}{2^5} = \frac{4}{2^7}$, $a_7 = \frac{3}{2^7}$, $a_6 > a_7$ klesne, vyloučí rostoucí a neklesající.

(iii): klesající. Ekvivalentní úpravy pro libovolné $k \geq 2$:

$$a_k > a_{k+1} \iff \frac{3}{k!} > \frac{3}{(k+1)!} \iff k+1 > 1 \iff k > 0 \text{ pravda.}$$

(iv): $\{2, 2, \frac{4}{3}, \frac{2}{3}, \frac{4}{15}, \dots\}$, nerostoucí.

$$a_k \geq a_{k+1} \iff \frac{2^k}{k!} \geq \frac{2^{k+1}}{(k+1)!} \iff k+1 \geq 2 \iff k \geq 1 \text{ pravda.}$$

(v): $\{-1, -\frac{1}{9}, \frac{1}{27}, \frac{1}{27}, \frac{5}{243}, \dots\}$ není monotonní, $a_1 < a_2$ vyloučuje, aby byla klesající či nerostoucí, $a_4 > a_5$ vyloučuje rostoucí a neklesající.

(vi): $\{1, 5, 5, 9, 9, 13, 13, 17, \dots\}$ neklesající. Ekvivalentní úpravy pro libovolné $k \geq 1$:

$$a_k \leq a_{k+1} \iff 2k + (-1)^k \leq 2(k+1) + (-1)^{k+1} \iff (-1)^k \leq 2 + (-1)^{k+1} \iff (-1)^k + (-1)^{k+1} \leq 2 \text{ pravda.}$$

9a.5: Poslední výskyt čísla k je na místě posloupnosti daném $1 + 2 + \dots + k$. Proto je číslo k je rovno a_n pro n splňující $\frac{1}{2}(k-1)k + 1 \leq n \leq \frac{1}{2}k(k+1)$, tedy pro $k^2 - k + 1 \leq 2n \leq k^2 + k$. Vyřešíme pro k a dostaneme, že člen a_n je roven číslu $k \in \mathbb{N}$ splňujícímu $\sqrt{2n - \frac{3}{4} + \frac{1}{2}} \leq k \leq \sqrt{2n + \frac{1}{4}} - \frac{1}{2}$. Pak si s tím trochu pohrejte, já například vidím $a_N = \lceil \sqrt{2n + \frac{1}{4}} - \frac{1}{2} \rceil$ a $a_N = \lfloor \sqrt{2n - \frac{3}{4} + \frac{1}{2}} \rfloor$, vzoreček ze cvičení je ale výrazně jednodušší, tak to chce k němu dojít.

9b. Porovnávání rychlosti růstu

V této kapitole budeme porovnávat rychosti růstu výrazů typu $[\ln(k)]^a, k^b, q^k, k!$ a k^k , které (pro $a, b > 0, q > 1$) jdou všechny do nekonečna, ale každý typ jinak rychle. Abychom viděli, odkud přišly naše definice, podíváme se na motivační příklad.

! Příklad 9b.a: Porovnáme rychlost růstu výrazů (posloupností) $k!$ a k^3 . Začneme tabulkou s několika prvními hodnotami.

$k:$	1	2	3	4	5	6	7	8
$k^3:$	1	8	27	64	125	218	343	512
$k!:$	1	2	6	24	120	720	5040	40320

Vidíme, že zpočátku rostla rychleji třetí mocina, ale pak ji faktoriál předběhl a roste do nekonečna rychleji. Trocha experimentování s dalšími mocninami naznačí následující:

- Pro každý exponent $a > 0$ existuje index k_0 takový, že $k! > k^a$ pro $k \geq k_0$.

Budeme říkat, že pro „dostatečně velká“ k je faktoriál větší než mocniny.

V čem jsou tyto úvahy pro nás užitečné? Mají přímý dopad na rozhodování při výběru algoritmů. Většina algoritmů funguje tak, že je schopna přijmout vstupní data rozličných velikostí a délka trvání algoritmu pak nějakým způsobem závisí na této velikosti, většinou se to odhaduje podle toho, kolik operací musí algoritmus vykonat na splnění úkolu (podobně se počítá i nárok na paměť a podobně). U mnohých problémů dopředu přesně nevíme, jak velký datový soubor bude zpracováván, jen víme, že bude hodně veliký a že navíc časem i poroste. Pak přichází na scénu ono porovnání „pro velká k “.

Třeba v příkladě uvidíme, že počítání determinantu podle definice vyžaduje cca $k!$ operací, zatímco převod na trojúhelníkovou matici jen asi k^3 . Pokud počítáme matice 2×2 a 3×3 , pak vychází lépe ten faktoriál, což souhlasí, pro takto malé matice máme příjemná pravidla. Jestliže ale máme čekat velké matice, pak je verze s náročností k^3 výrazně lepším kandidátem.

Zde je ovšem zajímavá otázka: Vyplatí se investovat do nalezení programu s menší náročností, nebo do silnějšího počítače? Pokud počítač stokrát urychlíme, pak nebude porovnávat $k!$ s k^3 , ale $\frac{1}{100}k!$ s k^3 neboli $k!$ s $100k^3$.

Položme si tedy otázku, o kolik větší je faktoriál než třetí mocnina, když k roste. To se nejlépe vidí z podílu. Pro $k \geq 4$ můžeme odhadovat:

$$\frac{k!}{k^3} = \frac{1 \cdot 2 \cdots (k-3)(k-2)(k-1)k}{k \cdot k \cdot k} \geq \frac{(k-3)(k-2)(k-1)k}{k \cdot k \cdot k} = \frac{k-3}{k} \frac{k-2}{k} \frac{k-1}{k} \cdot k.$$

Čtenář znalý analýzy již ví, že pro velké hodnoty k jsou podíly typu $\frac{k-a}{k}$ téměř rovny jedné. My nechceme na analýzu příliš spoléhat, proto dokážeme, že když je $k \geq 6$, tak jistě $\frac{k-3}{k} \geq \frac{1}{2}$:

$$k \geq 6 \implies 2k - 6 \geq k \implies 2(k-3) \geq k \implies \frac{k-3}{k} \geq \frac{1}{2}.$$

Pro tato k tedy máme $\frac{1}{2} \leq \frac{k-3}{k} \leq \frac{k-2}{k} \leq \frac{k-2}{k}$ a dostáváme odhad

$$\frac{k!}{k^3} \geq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot k = \frac{1}{8}k.$$

Co to znamená? Že pokud se rozhodneme urychlit tu třetí mocninu nějakým faktorem A , tak dozajista

$$\frac{k!}{Ak^3} \geq \frac{1}{8A}k,$$

takže pokud si počkáme na dostatečně velké k , tak jistě bude podíl větší než 1 neboli $k! > Ak^3$. Jinými slovy, nejenže je faktoriál (pro velká k) větší než třetí mocnina, on je dokonce větší než libovolný násobek třetí mocniny (čím větší násobek, tím déle si musíme počkat, než faktoriál vyhraje, ale dočkáme se).

Z praktického hlediska to říká, že pokud si máme u úlohy, kde očekáváme narůstající objem vstupních dat, vybrat mezi urychlením algoritmu na lepší typ (třeba k^3 namísto faktoriálu) nebo urychlením počítače, tak je rozhodně výhodnější zlepšit algoritmus.

△

Podívejme se ještě jednou na odhad, který jsme v příkladě odvodili. Zjistili jsme, že $k!$ je pro rostoucí k větší než libovolný násobek třetí mocniny. Nepřesně řečeno to znamená, že v nekonečnu je faktoriál nekonečně krát větší než k^3 . Matematicky řečeno jsme ukázali, že pro libovolnou konstantu K dokážeme přejít k velkým hodnotám k tak, aby už $\frac{k!}{k^3} \geq K$. V řeči limit to znamená, že $\frac{k!}{k^3} \rightarrow \infty$, popřípadě že $\frac{k^3}{k!} \rightarrow 0$ (viz předchozí sekce).

Třetí mocnina samozřejmě není ničím výjimečná. Obdobným způsobem se ukáže, že pro libovolné $a > 0$ je faktoriál nekonečně krát větší než k^a , přesně řečeno $\frac{k!}{k^a} \rightarrow \infty$, viz cvičení .

Přesně takový vztah totální dominance jednoho výrazu nad druhým nás zajímá a dáme mu jméno. A protože vztah, který uvažujeme, je jakási totální nerovnost, zavedeme k ní i nerovnost opačnou. Samozřejmě to není třeba (stejně jako nám k porovnávání stačí jen směr \geq), ale občas to někomu přijde užitečné (tak jako \leq).

Z hlediska formálního budeme mluvit o posloupnostech. Není to nic divného, když se podíváme na chování algoritmu globálně, přes všechna možná k , tak dostáváme posloupnost hodnot. Pro zjednodušení se soustředíme na vzájemné srovnání dvou výrazů, které jdou do nekonečna, což pak znamená, že můžeme i předpokládat, že jsou kladné. Pokud by totiž nějaký výraz pro malá k dával záporné hodnoty, tak prostě počátek takového posloupnosti ignorujeme, stejně nás zajímá jen chování pro velká k .

!

Definice.

Nechť $\{a_k\}$, $\{b_k\}$ jsou posloupnosti kladných čísel splňující $a_k \rightarrow \infty$, $b_k \rightarrow \infty$.

Řekneme, že a_k je $o(b_k)$, psáno také $a_k = o(b_k)$, jestliže $\frac{a_k}{b_k} \rightarrow 0$ neboli $\frac{b_k}{a_k} \rightarrow \infty$.

Řekneme, že a_k je $\omega(b_k)$, psáno také $a_k = \omega(b_k)$, jestliže $\frac{a_k}{b_k} \rightarrow \infty$ neboli $\frac{b_k}{a_k} \rightarrow 0$.

Naše úvodní úvahy teď můžeme zachytit zápisem $k! = o(k^3)$ nebo také $k^3 = \omega(k!)$.

Toto značení pochází z oblasti fyziky, matematické analýzy a teorie čísel, hodně se také používá při analýze algoritmů. Čte se „ a_k je malé ó b_k “, podobně s omegou. Nejen v diskrétní matematice je také velmi rozšířená fráze **posloupnost** $\{a_k\}$ **roste asymptoticky rychleji než posloupnost** $\{b_k\}$, také by šlo říct **asymptotický řád růstu** a_k je větší než asymptotický řád růstu b_k a další obdobné formulace. Když se objeví slova „asymptoticky“, „rychlosť růstu“ či „řád růstu“ v nějaké podobě, tak jde s vysokou pravděpodobností právě o vztah $a_k = o(b_k)$.

Již z definice je jasné, že pojmy o a ω jsou duální,

- $a_k = o(b_k)$ právě tehdy, když $b_k = \omega(a_k)$.

Symbol ω je tedy používán výrazně méně. Všimněte si, že značka $=$ zde neznamená klasickou rovnost, ale je to součástí specifického značení, je to jakási zkratka pro slovo „je“. Někteří autoři to vnímají jinak, berou $o(b_n)$ jako množinu všech posloupností $\{a_k\}$ splňujících $\lim(\frac{a_k}{b_k}) = 0$. Pak to není jen součást značení, ale má to skutečný matematický význam a namísto našeho značení s „=“ píšou $a_k \in o(b_k)$.

Teď přichází na řadu prozkoumání vzájemného vztahu populárních výrazů. Než zformulujeme příslušné tvrzení, uděláme ještě jeden motivační příklad, abychom docenili praktický dopad takového srovnání.

! **Příklad 9b.b:** Představme si, že máme počítač, který vykoná milion určitých kroků za sekundu (tedy jeden trvá tisícinu milisekundy), a zkoušme na něm algoritmy, které pracují se vstupy o velikosti k . Každý algoritmus potřebuje na zpracování jiný počet kroků, podle toho, jakým vzorcem náročnost závisí na k . V tabulce si porovnáme několik typických náročností přepočítaných na čas.

Na začátku tabulky zvyšujeme k povlovně, od stovky dál zvyšujeme velikost dat vždy desetkrát. Mimochodem, velikost vstupních dat 10^8 není žádné přehánění, například při řešení klimatických modelů, proudění kolem letadel či podobných hrátkách s přírodou nejsou matice o tak velkých rozdílech ničím výjimečným.

Čas je udáván v milisekundách, pokud není řečeno jinak, pak je „s“ sekunda, „m“ minuta, „h“ hodina, „d“ den a „r“ rok.

Při prohlížení tabulky začneme nejprve většinou řádky s tečkami, kde porovnáváme algoritmy s lineární a kvadratickou náročností. Porovnání časů potvrzuje naši zkušenosť, že pro malá k zase tak velký rozdíl mezi přímou a parabolou není, ale jak zvětšujeme proměnnou, tak se parabola odpichne a uhání k nekonečnu výrazně rychleji; ke konci tabulky dobíhá lineární algoritmus pořád ještě v řádu minut, zatímco kvadratický už vyžaduje stovky let.

Mnoho algoritmů má lineární časovou náročnost, což v zásadě znamená, že zvětšíme-li velikost dat desetkrát, délka trvání se zvětší také desetkrát (přímá úměrnost). Příkladem budiž trvání komplikace videa v závislosti na jeho délce.

Kvadratickou náročnost zná každý z nás. V mozku-procesoru umíme rychle násobit čísla až po 9, pro násobení větších (k -ciferných) čísel máme algoritmus, který vyžaduje k^2 oněch jednoduchých násobení. Podobně to funguje i v počítači.

$k =$	5	10	15	20	30	50	100	1000	10000	10^5	10^6	10^7	10^8
$\ln(k)$:	0.0016	0.0023	0.0027	0.003	0.0034	0.004	0.0046	0.007	0.009	0.01	0.014	0.016	0.018
• k :	0.005	0.01	0.015	0.02	0.03	0.05	0.1	1	10	0.1s	1s	10s	1.7m
$k \ln(k)$:	0.008	0.023	0.04	0.06	0.1	0.2	0.46	6.9	92	1.1s	13s	2.7m	31m
$k^{1.585}$:	0.013	0.038	0.073	0.12	0.22	0.49	1.5	57	2.2s	1.4m	54m	1.4d	55d
$\frac{1}{100} k^2$:	0.0002	0.001	0.002	0.004	0.009	0.025	0.01	10	1s	1.7m	2.8h	11.6d	3.2r
• k^2 :	0.025	0.1	0.2	0.4	0.9	2.5	10	1s	100s	28m	11.6d	3.2r	317r
2^k :	0.03	1	32	1s	18m	35.5r	$10^{16}r$	$10^{287}r$					
$\frac{1}{1000} k!$:	0.1	3	21m	77r	$10^{16}r$	$10^{48}r$	$10^{141}r$						

Teď se podíváme na zajímavé modifikace kvadratického růstu. Hned nad příslušným řádkem vidíme data pro jiný počítač, který se nám nemalými náklady podařilo stokrát urychlit. Jak se dá čekat, pro menší hodnoty k máme dokonce lepší výsledky než lineární případ, ale pro velká k se nakonec zase dostáváme k dlouhým běhům.

Ale násobení k -místných čísel se dá udělat i fintou, která sníží náročnost z k^2 na přibližně $k^{1.585}$, viz příklad . Jak ukazuje příslušný řádek tabulky, i zde na začátku prohráváme se stokrát urychleným kvadratickým růstem, ale podle očekávání nakonec menší mocnina vychází výrazně rychleji i při použití původního pomalého počítače.

Kvadratickou náročnost vyžaduje také uspořádání seznamu o k položkách pomocí vzájemného porovnávání. Zde existují populární alternativy, například merge-sort, který má náročnost jen $k \ln(k)$ (viz cvičení). Příslušný řádek je hned pod řádkem lineárním a vidíme, že se až tak moc neliší (u obou mluvíme na konci tabulky o minutách), takže to je opravdu dost dobré, na hlavu porázíme urychlený kvadrát i mocninu $k^{1.585}$.

Pro úplnost jsme přidali několik situací jiného typu. Na prvním řádku je algoritmus logaritmické náročnosti, což je velice nenáročný růst. Logaritmus pro velké k „uvadá“ a roste čím dál pomaleji, takové algoritmy máme nejraději, třeba binární vyhledávání, viz příklad .

Naopak dole máme řádek s geometrickou náročností a tisícásobně urychlený faktoriál. V obou případech jsme tabulkou ani nedokončili, ono to ostatně nemá smysl v okamžiku, kdy vyskakují délky běhu programu mnohonásobně překračující dosavadní trvání vesmíru. To už jsou hodnoty, kde ani miliarda-násobné urychlení počítace nic nesvede. Jsou problémy, ve kterých faktoriál jako náročnost vyskočí, například počítání determinantu matice podle definice, tabulka ukazuje, že od těch raději ruce pryč (determinant se dá urychlit na mocninu, viz příklad). Podobně nepříjemné jsou i geometrické náročnosti, například lámání šifry RSA závisí na délce zvoleného klíče zhruba geometrickým způsobem, díky čemuž je tato šifra (zatím) bezpečná, jen málo informací si svou hodnotu uchová desítky let, které jsou zatím třeba na vyuštění i na těch nejsilnějších počítačích. V okamžiku, kdy výkon počítačů vzroste natolik, že šifra přestane být bezpečnou, stačí díky rychlému růstu náročnosti nepatrně zvětšit délku klíče a počítace už zase čelí výpočtem na dekády.

△

Potvrďme si oficiálně, co jsme z příkladu vytušili.

!

Věta 9b.1.

- (i) Nechť $a, b > 0$ a $q > 1$. Pak platí $[\ln(k)]^a$ je $o(k^b)$, k^b je $o(q^k)$, q^k je $o(k!)$ a $k!$ je $o(k^k)$.
- (ii) Jestliže $0 < a < b$, pak $[\ln(k)]^a$ je $o([\ln(k)]^b)$ a k^a je $o(k^b)$.
- (iii) Jestliže $1 < q < r$, pak q^k je $o(r^k)$.

Jaký tedy máme obrázek? Máme pět skupin výrazů, logaritmy $[\ln(k)]^a$, mocniny k^b , geometrické výrazy q^k , faktoriál $k!$ a obecnou mocninu k^k . Každý výraz z jedné z těchto skupin je pro velká k větší než jakýkoliv výraz ze skupin jmenovaných dříve, bez ohledu na konstanty. Takže například pokud si dostatečně dlouho počkáme, tak $(1.1)^k > Ak^{1000000}$ pro libovolnou volbu konstanty $A > 0$.

V rámci každé skupiny pak o hierarchii rozhodují hodnoty parametru, třeba 3^k roste asymptoticky rychleji než e^k , což roste asymptoticky rychleji než 2^k , podobně k^7 roste asymptoticky rychleji než k^4 a to roste asymptoticky rychleji než \sqrt{k} .

Těmto vztahům se často říká „škála mocnin“ a její znalost umožní rychle porovnávat růsty různých výrazů. Důkaz jen naznačíme, aby měl čtenář představu.

Důkaz (náznak): (i): Začneme zprava, budeme zkoumat podíly.

1) Pro $k > 1$ máme shodný počet členů v čitateli i jmenovateli a můžeme je spárovat.

$$\frac{k^k}{k!} = \frac{k \cdot k \cdot k \cdots k}{1 \cdot 2 \cdot 3 \cdots k} = \frac{k}{1} \frac{k}{2} \frac{k}{3} \cdots \frac{k}{k} \geq \frac{k}{1} \cdot 1 \cdot 1 \cdots 1 = k.$$

Podíl tedy dosahuje libovolně velkých hodnot, proto $\frac{k^k}{k!} \rightarrow \infty$.

2) I zde rozdělíme podíl do páru, protože počty součinitelů souhlasí.

$$\frac{k!}{q^k} = \frac{1}{q} \cdot \frac{2}{q} \cdot \frac{3}{q} \cdots \frac{k-1}{q} \cdot \frac{k}{q}.$$

Ted' je třeba si uvědomit, že q je konstantní, zatímco k se mění. Vidíme, že zlomky jsou vlastně pořád stejné, pro všechna k začínáme stejně, pokud k zvětšíme, tak se k součinu přidají nové členy na konec. Tyto nově přidávané členy jsou typu $\frac{k}{q}$, tedy jsou stále větší, zatímco začátek zůstává stejný. Jak je tento začátek velký?

Nejprve jsou zlomky, které jsou malé, jako $\frac{1}{q}$, ale dříve či později se čitatel zvětší natolik, že už budou výsledné zlomky typu $\frac{n}{q}$ větší než 1, časem dokonce dospějeme tak daleko, že budou větší než 2. My přesně víme, kdy se tak stane, až číslo v čitateli překročí $[2q]$. Pokud tedy bude k větší než číslo $K = [2q]$, tak výraz $\frac{k!}{q^k}$ vždy začíná výrazem $\frac{1}{q} \cdot \frac{2}{q} \cdots \frac{K-1}{q} \cdot \frac{K}{q} = A$, jehož hodnota závisí čistě na q . Tento výraz bude násoben dalšími zlomky typu $\frac{n}{q}$, které jsou ovšem všechny větší než 2 a je jich přesně $k - K$. Můžeme proto pro $k > K$ odhadovat

$$\begin{aligned} \frac{k!}{q^k} &= \frac{1}{q} \cdot \frac{2}{q} \cdots \frac{K}{q} \cdot \frac{K+1}{q} \cdots \frac{k-1}{q} \cdot \frac{k}{q} = A \cdot \frac{K+1}{q} \cdots \frac{k-1}{q} \cdot \frac{k}{q} \\ &> A \cdot 2 \cdots 2 = A \cdot 2^{k-K} = \frac{A}{2^K} \cdot 2^k \rightarrow \infty. \end{aligned}$$

3) Vztahy $k^a = o(q^k)$ a $[\ln(k)]^b = o(k^a)$ se dokazují metodami matematické analýzy na pár řádcích. Je nicméně zajímavé, že existuje relativně elementární důkaz že k^a je $o(q^k)$, ukážeme si jej.

Nejprve odvodíme, že k je $o(2^k)$. Nechť je tedy $A > 0$ libovolné, ukážeme, že $2^k > Ak$ pro dostatečně velká k .

Označme $d_k = 2^k - Ak$. Protože $2^k \rightarrow \infty$, určitě existuje $K \in \mathbb{N}$ takové, že $2^k > A + 1$ pro $k \geq K$. Pro tyto k pak můžeme odhadovat

$$d_{k+1} = 2^{k+1} - A(k+1) = 2^k + 2^k - Ak - A = 2^k - Ak + 2^k - A = d_k + (2^k - A) > d_k + 1.$$

Co to znamená? Ať už je rozdíl $2^K - AK$ jakýkoliv, počínaje tímto indexem se s každým zvýšením k rozdíl zvětší alespoň o jedničku. To znamená, že dříve či později začnou být d_k kladné, tedy existuje k_0 takové, že pro $k \geq k_0$ je $d_k > 0$, tedy $2^k > Ak$.

Tento důkaz lze upravit pro libovolné q^k s $q > 1$, viz cvičení .

Pak už snadno dokážeme, že pro libovolné $a > 0$ máme

$$\frac{q^k}{k^a} = \frac{(q^{k/a})^a}{k^a} = \left(\frac{(q^{1/a})^k}{k}\right)^a \rightarrow \infty^a = \infty.$$

(ii): $\frac{k^b}{k^a} = k^{b-a} \rightarrow \infty$, neboť $b - a > 0$. Podobně pro logaritmy.

(iii): $\frac{r^k}{q^k} = \left(\frac{r}{q}\right)^k \rightarrow \infty$, neboť $\frac{r}{q} > 1$, je to zase geometrická posloupnost.

□

Teď si zavedeme další pojmy, které nám umožní přibližně porovnávat výrazy podle velikosti pro velká k . Definice jsou inspirovány praktickými požadavky. Jeden je, že srovnání by mělo být přibližné. Například je v zásadě jedno, jestli něco trvá 10 let nebo 10 let a 2 hodiny. Členy méně důležité tedy budeme chtít zanedbávat.

Druhý požadavek je, aby naše porovnávání ignorovalo situaci, kdy počítač několikrát urychlíme. To znamená, že pokud jeden výraz roste určitou rychlostí a druhý je stále řekněme přibližně dvakrát větší, tak chceme, aby definice řekla, že budou mít stejnou asymptotickou rychlosť růstu.

Opět se omezíme na výrazy, které jdou do nekonečna.

!

Definice.

Nechť $\{a_k\}, \{b_k\}$ jsou posloupnosti kladných čísel splňující $a_k \rightarrow \infty, b_k \rightarrow \infty$.

Řekneme, že a_k je $O(b_k)$, jestliže $\exists K > 0$ a $k_0 \in \mathbb{N}$ aby $\forall k \geq k_0: a_k \leq Kb_k$.

Řekneme, že a_k je $\Omega(b_k)$, jestliže $\exists L > 0$ a $k_0 \in \mathbb{N}$ aby $\forall k \geq k_0: a_k \geq Lb_k$.

Řekneme, že a_k je $\Theta(b_k)$ nebo že $a_k \asymp b_k$, jestliže $\exists K, L > 0$ a $k_0 \in \mathbb{N}$ aby $\forall k \geq k_0: Lb_k \leq a_k \leq Kb_k$.

První pojem je odhad shora, kdy jakoby říkáme, že $a_k \leq b_k$, ale té b_k můžeme trochu pomocí vynásobením konstantou, aby se nad a_k dostala. Podobně funguje odhad zdola v druhém pojmu. Čteme a_k je velké ó b_k .

Třetí pojem je právě tím, kterým chceme říct, že se věci v zásadě rovnají. Čteme a_k je téta b_k , ale v diskrétní matematice (a dalších aplikacích) často slyšíme něco jako **asymptotická rychlosť růstu** a_k je b_k .

Někteří autoři mají ty definice jinak, bez odřezávání s k_0 , například takto:

- $a_k = \Theta(b_k)$ jestliže $\exists K, L > 0$ aby $\forall k: Lb_k \leq a_k \leq Kb_k$.

Dá se snadno ukázat, že takováto definice je vlastně stejná jako ta naše. Je snažší na čtení (a občas i na použití), ale neumožňuje odřezávání začátků, čímž se zase komplikuje hledání správného K či L . Níže to uvidíme na konkrétních příkladech.

Hned z definice vidíme, že platí následující:

- $a_k = O(b_k)$ právě tehdy, když $b_k = \Omega(a_k)$ [protože $a_k \leq Kb_k \iff b_k \geq (1/K)a_k$],
- $a_k = \Theta(b_k)$ právě tehdy, když $a_k = O(b_k)$ a $a_k = \Omega(b_k)$, což je právě tehdy, když $a_k = O(b_k)$ a $b_k = O(a_k)$.

Má to smysl, dva výrazy se „skoro rovnají“ (rychlosť růstu), jestliže je jeden „skoro menší“ než druhý a také naopak.

Uvedeme si některé souvislosti mezi novými pojmy i předchozím srovnáním o a ω . Pokud má čtenář už trochu představu o významu rozličných srovnání, tak by mu následující tvrzení měla přijít přirozená.

Fakt 9b.2.

Uvažujme posloupnosti $\{a_k\}, \{b_k\}$ kladných čísel jdoucí do nekonečna.

- (i) Jestliže $a_k = o(b_k)$, pak $a_k = O(b_k)$ a nemůže platit $a_k = \Omega(b_k)$ ani $a_k = \Theta(b_k)$.
- (ii) Jestliže $a_k = \omega(b_k)$, pak $a_k = \Omega(b_k)$ a nemůže platit $a_k = O(b_k)$ ani $a_k = \Theta(b_k)$.
- (iii) Nemůže platit zároveň $a_k = o(b_k)$ a $b_k = o(a_k)$.
- (iv) Jestliže $a_k = \Theta(b_k)$, pak nemůže platit $a_k = o(b_k)$ ani $b_k = o(a_k)$.

Důkaz (poučný): (i) Zvolme libovolné $K > 0$. Z předpokladu máme platnost $\frac{a_k}{b_k} \rightarrow 0$ a podle definice limity musí existovat k_0 takové, aby pro $k \geq k_0$ platilo $\frac{a_k}{b_k} < K$. Pro tato k pak máme i $a_k \leq Kb_k$, což dokazuje $a_k = O(b_k)$.

Pokud by zároveň platilo $a_k = \Omega(b_k)$, tak pro nějaké $L > 0$ a k_0 platí $a_k \geq Lb_k$ neboli $\frac{a_k}{b_k} \geq L$ pro všechna $k \geq k_0$, což ale znemožňuje splnění definicne $\frac{a_k}{b_k} \rightarrow 0$ pro $\varepsilon = L$.

(ii) Důkaz je obdobný.

(iii) Toto by znamenalo, že $\frac{a_k}{b_k} \rightarrow 0$ a zároveň $\frac{a_k}{b_k} \rightarrow \infty$, což je nemožné, jedna posloupnost nemůže mít dvě různé limity.

(iv) Z definice $a_k = \Theta(b_k)$ najdeme $K, L > 0$ a k_0 takové, že pro všechna (dostatečně velká) $k \geq k_0$ platí $L \leq \frac{a_k}{b_k} \leq K$. Jestliže ovšem pro všechna $k \geq k_0$ platí $\frac{a_k}{b_k} > L$, pak pro $\varepsilon = L$ není možné splnit podmínu z definice limity 0, tudíž neplatí, že $\frac{a_k}{b_k} \rightarrow 0$ neboli neplatí, že $a_k = o(b_k)$. Z $\frac{a_k}{b_k} < K$ pro všechna k zase vyloučíme $\frac{a_k}{b_k} \rightarrow \infty$ a neplatí $a_k = \omega(b_k)$ neboli neplatí $b_k = o(a_k)$.

□

Příklad 9b.c:

1) a) $k = o(k^2)$ neboli $k^2 = \omega(k)$.

b) Platí i $k = O(k^2)$, ale neplatí $k = \Omega(k^2)$ ani $k = \Theta(k^2)$.

Pojem Θ tedy „pozná“, že rychlosti k a k^2 nejsou souměřitelné.

Důkazy: a) $\frac{k^2}{k} = k \rightarrow \infty$, proto $k = o(k^2)$.

b) Platí $k \leq 1 \cdot k^2$ pro všechna $k \in \mathbb{N}$, proto lze zvolit $K = 1$, $k_0 = 1$ a máme $k = O(k^2)$.

Platí také $k = \Omega(k^2)$? Aby pro nějaké L platilo $k \geq Lk^2$ pro všechna (dostatečně velká) k , muselo by platit i $\frac{k}{k^2} \geq L$ pro všechna (dostatečně velká) k , ale to nejde, protože $\frac{k}{k^2} \rightarrow 0$, jinými slovy se tento podíl nakonec dostane pod jakoukoliv hladinu L , kterou zkusíme.

Nemůže proto platit ani $k = \Omega(k^2)$. Plyne to ostatně z části a) a Faktu výše.

2) a) $13k = \Theta(k)$.

b) $13k = O(k)$, $13k = \Omega(k)$, $k = O(13k)$, $k = \Omega(13k)$

c) Neplatí $13k = o(k)$ ani $13k = \omega(k)$.

Pojmy tedy dle našeho přání nepovažují násobení konstantu za podstatnou změnu rychlosti růstu.

Důkazy: a) Zvolíme $L = 1$ a $K = 13$, pak pro každé k máme $L \cdot k \leq 13k \leq K \cdot k$, stačí proto zvolit $k_0 = 1$. Zde tedy v pohodě projde i důkaz pomocí alternativní, neodřezávací definice.

b) plyne automaticky z a).

c) Protože $\frac{13k}{k} = 13$, tak neplatí ani $\frac{13}{k} \rightarrow 0$, ani $\frac{13k}{k} \rightarrow \infty$.

3) a) $100k + 50 = o(k^2)$.

b) $100k + 50 = O(k^2)$.

c) Neplatí $100k + 50 = \Theta(k^2)$.

d) $100k + 50 = \Theta(k)$. Neboli asymptotická rychlosť růstu $100k + 50$ je k .

Nejprve pojmy správně poznaly, že všechny členy výrazu $100k + 50$ rostou výrazně pomaleji než k^2 , nepomohlo jim ani násobení konstantou.

Pojem Θ pak ukázal, že umí zanedbávat nejen násobení číslem, ale také přítomnost méně důležitých členů, správně rozpoznal, že v daném výrazu je tím podstatným člen k .

Důkaz: a) $\frac{100k+50}{k^2} = \frac{100}{k} + \frac{50}{k^2} \rightarrow 0$.

b) Vyplývá automaticky z a), ale klidně ukážeme i přímý důkaz. Potřebujeme najít konstantu K takovou, aby $100k + 50 \leq Kk^2$ (pro velká k). Protože $k \leq k^2$, určitě máme $100k \leq 100k^2$, takže volba $K = 100$ by se postarala o první část, ale ještě bude zlobit ta druhá. Spravíme to navýšením K .

Určitě platí $50 = 50 \cdot 1 \leq 50k^2$, takže pokud přidáme k prvotní volbě $K = 100$ ještě padesátku, mělo by to stačit k pokrytí obou částí. Volíme tedy $K = 150$ a teď už opravdu pro všechna $k \in \mathbb{N}$ máme

$$100k + 50 \leq 100k^2 + 50k^2 = 150k^2 = Kk^2.$$

Vidíme, že jsme ani nemuseli odřezávat začátek posloupnosti, takže náš důkaz platí i pro alternativní definici pojmu $O(k^2)$.

Jako alternativu si ukážeme, že namísto zvětšování K lze použít možnost odřezávání, pokud pracujeme s definicí, která to povoluje. Víme, že k^2 je nekonečně krát větší než k , takže pokud si počkáme, tak určitě převáží přímo celý člen $100k$. Vidíme například, že pokud je $k \geq 100$, tak už máme $100k \leq k \cdot k = k^2$, takže by fungovala i volba $K = 1$ a odříznutí $k_0 = 100$.

Ještě se ale musíme postarat o tu padesátku. Jedna možnost je, že si prostě počkáme o něco déle. Trocha experimentování ukáže, že po dosazení $k = 101$ už opravdu platí $100k + 50 \leq k^2$, a dá se ukázat, že to platí

i pro všechna následující k . Lze tedy zvolit $K = 1$ a odřezávací bod $k_0 = 101$, pro všechna $k \geq k_0$ pak platí $100k + 50 \leq 1 \cdot k^2$, to už ale není vidět tak jasné, muselo by se to dokázat analytickými metodami.

Nejjednodušší je často kombinovat oba přístupy, přes odřezávání a přes zvětšování pomocí K . Už jsme viděli, že pro $k \geq k_0 = 100$ máme $100k \leq k^2$, pro takováto k ale také evidentně máme $50 \leq 100 = k \leq k^2$. Můžeme tedy zvolit $K = 2$ a odhadovat, že pro $k \geq 100 = k_0$ je

$$100k + 50 \leq k^2 + k^2 = K \cdot k^2.$$

c) Protože $100k + 50 = o(k^2)$, tak už je vyloučeno $100k + 50 = \Omega(k^2)$ a tedy i $100k + 50 = \Theta(k^2)$ (viz Fakt výše).

d) Hledáme konstanty L, K tak, aby $L \cdot k \leq 100k + 50 \leq K \cdot k$ pro všechna či pro dostatečně velká k (podle verze definice Θ). Je jasné, že volba $L = 1$ funguje pro všechna $k \in \mathbb{N}$. U horního odhadu zase zvolíme kombinaci přístupů přes odřezávání a přes zvětšování. Evidentně $50 \leq k$, pokud se chytře omezíme na velká k , v tomto případě stačí zvolit $k_0 = 50$. Také máme (pro všechna k) odhad $100k \leq 100 \cdot k$. Dáme to dohromady, pro $k \geq 50$ platí

$$13 \leq k^2 \implies 100k + 50 \leq 100k + k = 101k,$$

tedy stačí zvolit $K = 101$.

4) a) $2k^2 + 13 = \Theta(k^2)$.

b) Neplatí $2k^2 + 13 = o(k^2)$ ani $2k^2 + 13 = \omega(k^2)$.

Vidíme, že Θ „poznalo“, že o rychlosti růstu $2k^2 + 13$ rozhoduje výraz k^2 .

Důkaz: a) Hledáme K, L tak, aby platil odhad $Lk^2 \leq 2k^2 + 13 \leq Kk$ pro všechna (velká) k . Je hned vidět, že $L = 1$ bude fungovat pro všechna $k \in \mathbb{N}$, což mimochodem dokazuje, že $2k^2 + 13 = \Omega(k^2)$ neboli $k^2 = O(2k^2 + 13)$.

Teď hledáme K neboli vlastně chceme ukázat, že také $2k^2 + 13 = O(k^2)$. Volba $K = 2$ by nám zajistila dobrý odhad pro první část výrazu, u druhé bude nejjednodušší si trošku počkat. Pro $k \geq 4$ totiž určitě máme

$$2k^2 + 13 \leq 2k^2 + k^2 = 3k^2,$$

takže stačí zvolit $k_0 = 4$ a $K = 3$.

Pokud bychom nechtěli odřezávat, tak bychom odhadovali třeba takto: Protože $1 \leq k^2$, tak určitě

$$2k^2 + 13 = 2k^2 + 13 \cdot 1 \leq 2k^2 + 13k^2 = 15k^2.$$

Volba $K = 15$ tedy zaručí platnost potřebného odhadu pro všechna k .

b) Plyne to z a). Dá se také ukázat, že $\frac{2k^2+13}{k^2} \rightarrow 2$, takže tento podíl nemá ani limitu 0, ani limitu ∞ . Proto neplatí $2k^2 + 13 = o(k^2)$ ani $2k^2 + 13 = \omega(k^2)$.

△

V praxi se ovšem $a_k = \Theta(b_k)$ obvykle dokazuje jinak než hledáním K, L , které by fungovaly. Východiskem je zkoumání podílu $\frac{a_k}{b_k}$. Pokud má limitu 0 či nekonečno, tak už víme, co to znamená. Pokud má limitu jinou, pak také dostáváme podstatnou informaci díky následujícímu důležitému tvrzení:

• Jestliže má $\frac{a_k}{b_k}$ nenulovou (a konečnou) limitu, tak $a_k = \Theta(b_k)$.

Na hledání limit nám analýza nabízí mocné nástroje, takže důkaz je pak často na jednom řádku. Například to, že $100k + 50 = \Theta(k)$ (viz příklad výše) se dokáže hravě výpočtem

$$\lim\left(\frac{100k + 50}{k}\right) = \lim\left(100 + \frac{50}{k}\right) = 100 + 0 = 100.$$

Tento výsledek říká, že pro velké hodnoty k je $100k + 50$ v zásadě stokrát větší než k , takže rostou řádově stejně rychle.

Klasický problém je, když nám někdo dá kombinaci různých typů a my máme rozhodnout, jak se chová celý výraz (jakou má asymptotickou rychlosť růstu). Pak se použije následující trik: Jestliže se sčítají dvě části a ve vzájemném porovnávání jedna z nich prohraje, tak ji lze vynechat, aniž bychom tím pro velká k ovlivnili daný výraz. Formálně:

Fakt 9b.3.

Uvažujme posloupnosti $\{a_k\}, \{b_k\}$ kladných čísel jdoucí do nekonečna.

Jestliže $b_k = o(a_k)$, pak $a_k + b_k = \Theta(a_k)$.

Indukcí to pak rozšíříme na více sčítanců, čímž dostaneme algoritmus pro hledání dominantního typu v kombinaci více členů.

Příklad 9b.d: Jakou asymptotickou rychlosť růstu má výraz $3^k - 150k^{17} + \sqrt{k} - 200 \cdot e^k + \log_5(k)$?

Vidíme, že se ve výrazu sčítají členy ze tří kategorií: geometrické výrazy 3^k a e^k , mocniny \sqrt{k} a k^{17} a logaritmus. Víme, že z těchto tří skupin rostou nejrychleji geometrické výrazy, proto lze podle Faktu výše ty ostatní ignorovat.

O dominanci se tedy poperou výrazy 3^k a e^k . V rámci jedné skupiny rozhoduje velikost parametru, zde $3 > e$ a je jasno.

Závěr: $3^k - 150k^{17} + \sqrt{k} - 200 \cdot e^k + \log_5(k) = \Theta(3^k)$.

Slovně, asymptotická rychlosť růstu daného výrazu je 3^k . Někdy se také říká, že daný výraz je typu 3^k .

Zkuste si vzít nějaký počítací přístroj a dosadit do obou výrazů třeba $k = 10^6$, uvidíte, že se výsledné hodnoty moc neliší.

△

Jak tedy vypadá postup? Při hledání asymptotické rychlosťi růstu daného výrazu se nejprve pohybujeme na úrovni kategorií, porovnáme ty, které jsou přítomny a pak se dále v úvahách omezíme jen na ty části výrazu, které patří do nejvyšší přítomné kategorie. Mezi nimi pak rozhodne velikost parametru, čímž se vybere tzv. „dominantní člen“, který určuje chování celého výrazu.

Příklad 9b.e:

- a) $k^2 - 257k = \Theta(k^2)$.
- b) $3^k + 7k^{527} - \pi(-2)^k = \Theta(3^k)$.
- c) $20k! + 160k^{13} - 3^k = \Theta(k!)$.

△

Často srovnáváme jen mocniny, z výsledků výše pak okamžitě plyne následující závěr:

! Důsledek 9b.4.

Jestliže je $p(k)$ polynom stupně n a $a_n > 0$, pak $p(k) = \Theta(k^n)$.

Stručně řečeno, pro velká k se polynom chová stejně jako jeho nejvyšší mocnina.

! Zdálo by se, že už dokážeme v pohodě vyhodnocovat rychlosť růstu výrazů, ale není tomu tak, naše škála totiž nepostihuje všechny výrazy. Kam na ní například zapadne výraz $k \ln^3(k)$? Určitě $k \ln^3(k) = \omega(k)$, ale jak se ten výraz porovná třeba s $k^{1.01}$? Kam přijde na naší škále výraz $e^{\sqrt{k}}$ nebo $\ln(e^k + 13k)$?

Ve skutečnosti ona škála představuje jen malou část toho, co se může vyskytnout, na druhou stranu si s ní překvapivě často vystačíme. Když náhodou narazíme na něco, co do ní nepatří, pak hodně záleží na zkušenosti a intuici, protože univerzální metody pro zkoumání nejsou. Jako ukázku prozkoumáme dva výrazy.

Příklad 9b.f: Jaká je asymptotická rychlosť růstu $k \ln(k)$?

Protože $\frac{k \ln(k)}{k} \rightarrow \infty$, tak určitě $k \ln(k)$ je $\omega(k)$.

My ale víme, že pro libovolné kladné a platí $\ln(k) = o(k^a)$, proto $k \ln(k)$ roste asymptoticky pomaleji než $k \cdot k^a = k^{1+a}$. Důkaz:

$$\frac{k \ln(k)}{k^{1+a}} = \frac{\ln(k)}{k^a} \rightarrow 0.$$

Dostáváme tedy zajímavý obrázek. Máme kategorie mocnin k^a , kterou si člověk intuitivně představuje jako souvislou, začneme malými mocninami, třeba $k^{0.13}$, a jak postupně mocninu zvyšujeme, dostáváme stále rychleji rostoucí mocniny. Teď jsme ale zjistili, že do té škály dokážeme zasunout výraz $k \ln(k)$, a to bezprostředně za k^1 . Tak bezprostředně, že sebemenší zvýšení mocniny nám již dá něco, co dominuje výrazu $k \ln(k)$.

Ve skutečnosti je to ještě zajímavější. Za k je schována celá kategorie. Dá se totiž snadno dokázat, že pro všechna $b > 0$ máme $k \ln^b(k) = \omega(k)$, ale $k \ln^b(k) = o(k^a)$ pro libovolné $a > 1$. V rámci této vsunuté kategorie určujeme dominanci tradičně podle mocniny b .

△

Poznamenejme, že mnohé významné algoritmy mají právě náročnost $k \ln(k)$, takže znalost rychlosťi růstu tohoto výrazu je vysoce užitečná.

Příklad 9b.g: Jaká je asymptotická rychlosť růstu $e^{\sqrt{k}}$?

Výraz e^k patří do kategorie geometrických posloupností q^k . Všimněme si, že všechny tyto výrazy lze převést na exponenciálu, protože $q^k = e^{\ln(q)k}$. Tuto kategorii (kde bereme $q > 1$) si tedy můžeme představit jako množinu exponenciál e^{Ak} pro $A = \ln(q) > 0$.

Z toho bychom odhadli, že $e^{\sqrt{k}}$ do této kategorie nepatří, protože $\sqrt{k} = k^{1/2} = o(k)$, tudíž se dá čekat, že i $e^{\sqrt{k}} = o(e^{Ak})$. Dokážeme to: Pro $A > 0$ máme

$$\frac{e^{\sqrt{k}}}{e^{Ak}} = e^{\sqrt{k}-Ak} \rightarrow e^{-\infty} = 0 \quad \text{neboť} \quad \sqrt{k} - Ak \rightarrow -\infty.$$

Máme tedy $e^{\sqrt{k}} = o(q^k)$ pro $q > 1$. Nižší kategorií jsou mocniny k^a pro $a > 0$. Zkusíme tedy s nimi náš výraz provnat. Pokud čtenář zná analýzu, pak pomocí l'Hospitalova pravidla hravě dokáže, že $\frac{e^{\sqrt{k}}}{k^a} \rightarrow \infty$, což potvrzuje, že $e^{\sqrt{k}} = \omega(k^a)$. Tento výraz je tedy někde mezi kategorií mocnin a kategorií geometrických posloupností (exponenciál).

Protože se snažíme co nejméně záviset na jiném kursu, alespoň naznačíme, proč by ona limita měla jít do nekonečna. Nejprve si upravíme $\frac{e^{\sqrt{k}}}{k^a} = \frac{e^{\sqrt{k}}}{(\sqrt{k})^{2a}}$ a pak si označíme $m = \sqrt{k}$ (tedy používáme substituci). Dostáváme $\frac{e^{\sqrt{k}}}{k^a} = \frac{e^m}{m^{2a}}$. Když teď pošleme $k \rightarrow \infty$, tak také $m \rightarrow \infty$ a naše standardní škála nekonečen už odpoví, že $\frac{e^m}{m^{2a}} \rightarrow \infty$. \triangle

Dobrá zpráva je, že při běžné analýze algoritmů si v zásadě vystačíme s oněmi základními čtyřmi kategoriemi a $k \ln(k)$, což už všechno známe. Opravdu? To je otázka provokativní, a zaprovokujeme ještě více. Umíme vůbec do těch výrazů dosadit číslo?

Kupodivu to není tak snadné, jak to vypadá. Zatímco mocniny a exponenciály zvládne v pohodě i obyčejná kalkulačka, dosazovat do faktoriálu velká čísla je adrenalinovým sportem. Například při výpočtu 1000000! bychom potřebovali udělat milion násobení, přičemž by se ke konci pracovalo s dost velkými čísly, i velké počítače by se rádně zapotily.

Jenže my vlastně nepotřebujeme přesnou hodnotu, stejně při analýze výrazů bereme všechno přibližně. Pak se nabízí několik velice užitečných odhadů pro faktoriál.

Fakt 9b.5.

Pro $k \geq 6$ platí $\frac{k^k}{3^k} < k! < \frac{k^k}{2^k}$.

Důkaz (rutinní s výjimkou): Tou výjimkou je netriviální fakt, že zlomek $(1 + \frac{1}{k})^k = (\frac{k+1}{k})^k$ je pro $k \in \mathbb{N}$ vždy mezi 2 a 3, na což jsou rozličné triky, které sem spíš nepatří, a dá se to najít v každé tlustší učebnici analýzy. Pro nás to znamená, že $\frac{1}{3} \leq (\frac{k}{k+1})^k \leq \frac{1}{2}$ neboli $3(\frac{k}{k+1})^k \geq 1$ a $2(\frac{k}{k+1})^k \leq 1$.

Dokážeme teď indukcí $V(n)$: $\frac{k^k}{3^k} < k!$.

(0) $V(6)$ říká $2^6 < 6!$, což určitě platí.

(1) Pro jisté (libovolné) $k \geq 6$ předpokládejme, že $\frac{k^k}{3^k} < k!$. Potřebujeme ukázat, že $\frac{(k+1)^{k+1}}{3^{k+1}} < (k+1)!$. Máme $(k+1)! = (k+1) \cdot k! > (k+1) \frac{k^k}{3^k} = \frac{3(k+1)k^k}{(k+1)^k} \frac{(k+1)^k}{3^{k+1}} = 3\left(\frac{k}{k+1}\right)^k \frac{(k+1)^{k+1}}{3^{k+1}} \geq \frac{(k+1)^{k+1}}{3^{k+1}}$.

Důkaz hotov.

Tedě dokážeme indukcí $W(n)$: $k! < \frac{k^k}{2^k}$.

(0) $W(6)$ říká $6! < 3^6$, což určitě platí.

(1) Pro jisté $k \geq 6$ předpokládejme, že $k! < \frac{k^k}{2^k}$. Potřebujeme ukázat, že $(k+1)! < \frac{(k+1)^{k+1}}{2^{k+1}}$. Máme

$$(k+1)! = (k+1) \cdot k! < (k+1) \frac{k^k}{2^k} = \frac{2(k+1)k^k}{(k+1)^k} \frac{(k+1)^k}{2^{k+1}} = 2\left(\frac{k}{k+1}\right)^k \frac{(k+1)^{k+1}}{2^{k+1}} \leq \frac{(k+1)^{k+1}}{2^{k+1}}.$$

Důkaz hotov.

□

Faktoriál je tedy někde mezi $(\frac{k}{3})^k$ a $(\frac{k}{2})^k$. Následující tvrzení říká, že když si místo 2 nebo 3 v tomto odhadu dáme e , tak už víceméně dostaneme faktoriál.

Věta 9b.6. (Stirlingův vzorec)

Pro velká k platí $k! \sim \sqrt{2\pi k} \left(\frac{k}{e}\right)^k$.

Je to výsledek těžký, ale stojí za to, ta approximace je opravdu vynikající. Dokonce až neuvěřitelně. Normálně když člověk slyší, že něco je approximační vzorec, tak čeká dobré approximace pro větší čísla, řekněme v rádu stovek či tisíců, někdy si musí počkat ještě déle, ale tento vzorec se už trefuje hodně blízko dokonce od začátku. Ukážeme to v tabulce, kde jsme dali i procentuální chybu vzhledem k základu, která o přesnosti vypovídá nejvíce.

k :	1	2	3	4	5	6	7	8	9	10	20	30
k! :	1	2	6	24	120	720	5040	40320	362880	3628800	$2.433 \cdot 10^{18}$	$2.653 \cdot 10^{32}$
Stirling :	0.92	1.9	5.8	23.5	118	710	4980	39902	359536	3598696	$2.423 \cdot 10^{18}$	$2.645 \cdot 10^{32}$
chyba % :	8	5	3.3	2	1.7	1.4	1.2	1	0.9	0.8	0.4	0.3

Cvičení

Cvičení 9b.1 (dobré): (i) Ukažte, že $k^2 < 2^k$ pro $k \geq 5$. Bude se vám hodit, že pro $k \geq 5$ je $k^2 - 2k - 1 = (k-1)^2 - 2 > 0$.

(ii) Ukažte, že $k^3 < 2^k$ pro $k \geq 5$. Bude sevám hodit, že pro $k \geq 5$ je $k^3 - 3k^2 - 3k - 1 > 0$, což odůvodníme například takto:

$$k^3 - 3k^2 - 3k - 1 \geq k^3 - 3k^2 - 4k = k(k^2 - 3k - 4) = k(k-4)(k+1) > 0.$$

Cvičení 9b.2 (dobré): Dokažte indukcí, že pro $k \geq 4$ platí $2^k < k!$.

Cvičení 9b.3 (poučné): Dokažte, že pro každé $a > 0$ platí $\frac{k!}{k^a} \rightarrow \infty$.

Návod: Nejprve pro $a \in \mathbb{N}$ imitujte důkaz z příkladu .

Pro necelá a použijte srovnání nerovnosti.

Cvičení 9b.4 (poučné): Dokažte, že pro libovolné $q > 1$ platí, že pro každé $A > 0$ existuje k_0 tak, aby $q^k > Ak$ pro $k \geq k_0$.

Cvičení 9b.5 (rutinní): Seřaďte podle asymptotické rychlosti růstu výrazy $5k + \ln(k)$, $k^2 - 100k$, $k \ln(k)$, $2k^3$, \sqrt{k} .

Cvičení 9b.6 (rutinní): Najděte asymptotické rychlosti růstu následujících výrazů.

- | | |
|------------------------------|--------------------------|
| (i) $\sqrt{k} + \log_2(k)$; | (iii) $2^k + k^2 + 2k$; |
| (ii) $k^3 + 13k^2 + 14$; | (iv) $2^k + k!$. |

Cvičení 9b.7 (poučné): Dokažte podle definice následující tvrzení:

- | | |
|---|------------------------------------|
| (i) $100k^2 = O(k^4)$, $100k^2 = o(k^4)$; | (iii) $k^4 = o(k!)$; |
| (ii) $3k + 7 = \Theta(k)$; | (iv) $k + 3 \sin(k) = \Theta(k)$. |

Řešení:

9b.1: (i): (0) $k = 5$: $5^2 = 25 < 128 = 2^5$.

(1) $k \geq 5$, předpoklad $k^2 < 2^k$. Pak $2^{k+1} = 2 \cdot 2^k > 2 \cdot k^2 = k^2 + k^2 = (k^2 + 2k + 1) + (k^2 - 2k - 1) = (k+1)^2 + (k^2 - 2k - 1) > (k+1)^2$ dle vztahu z návodu.

Alternativa: převést na rozdíl. Předpoklad: $2^k - k^2 > 0$. Pak

$$2^{k+1} - (k+1)^2 = 2 \cdot 2^k - k^2 - 2k - 1 = 2(2^k - k^2) + k^2 - 2k - 1 > 2 \cdot 0 + 0 = 0.$$

(ii): (0) $k = 5$: $5^3 = 125 < 128 = 2^5$.

(1) $k \geq 5$, předpoklad $k^3 < 2^k$. Pak $2^{k+1} = 2 \cdot 2^k > 2 \cdot k^3 = k^3 + k^3 = (k^3 + 3k^2 + 3k + 1) + (k^3 - 3k^2 - 3k - 1) = (k+1)^3 + (k^3 - 3k^2 - 3k - 1) > (k+1)^3$ dle vztahu z návodu.

Alternativa: převést na rozdíl. Předpoklad: $2^k - k^3 > 0$. Pak

$$2^{k+1} - (k+1)^3 = 2 \cdot 2^k - k^3 - 3k^2 - 3k - 1 = 2(2^k - k^3) + k^3 - 3k^2 - 3k - 1 > 2 \cdot 0 + 0 = 0.$$

9b.2: (0) $k = 4$: $16 < 24$ platí.

(1) $n \geq 4$, předpoklad $2^k < k!$. Pak $(k+1)! = (k+1) \cdot k! > k \cdot 2^k = \frac{k}{2} \cdot 2^{k+1} > 2^{k+1}$, protože pro $k \geq 4$ určitě platí $\frac{k}{2} > 1$.

9b.3: Nechť $a \in \mathbb{N}$. Nejprve dokažte, že pro $k \geq a$ platí $\frac{k-a}{k} \geq \frac{1}{2}$. Pak

$$\frac{k!}{k^a} = \frac{1 \cdot 2 \cdots (k-a) \cdots (k-1)k}{k \cdots k} \geq \frac{k-a}{k} \cdots \frac{k-1}{k} \cdot k \geq \left(\frac{1}{2}\right)^a \cdot k \rightarrow \infty.$$

Pro obecné $a > 0$ je $k^a \leq k^{\lceil a \rceil}$.

9b.4: Pro $A > 0$ libovolné označíme $d_k = q^k - Ak$ a najdeme $K \in \mathbb{N}$ takové, že $q^k > \frac{A+1}{q-1}$ pro $k \geq K$ (neboť $q^k \rightarrow \infty$). Pro tyto k pak $(q-1)q^k - A > 1$ a

$$d_{k+1} = q^{k+1} - A(k+1) = q^k + (q-1)q^k - Ak - A = q^k - Ak + (q-1)q^k - A = d_k + (q-1)q^k - A > d_k + 1.$$

9b.5: Pořadí je \sqrt{k} , $5k + \ln(k)$, $k \ln(k)$, $k^2 - 100k$, $2k^3$.

9b.6: (i): $\Theta(\sqrt{k})$; (ii): $\Theta(k^3)$; (iii): $\Theta(2^k)$; (iv): $\Theta(k!)$.

9b.7: (i): $100k^2 = o(k^4)$: $\frac{100k^2}{k^4} = \frac{100}{k^2} \rightarrow 0$. $100k^2 = O(k^4)$: plyne z předchozího nebo přímo: $K = 1$, $k_0 = 10$, pak $k \geq k_0 \implies k^2 \geq 100 \implies 100k^2 \leq k^2 \cdot k^2 = 1 \cdot k^4$.

(ii): $L = 1$, $K = 4$, $k_0 = 7$, pak $k \geq k_0 \implies 1 \cdot k \leq 3k + 7$ a $k \geq k_0 \implies 3k + 7 \leq 3k + k = 4k$.

(iii): Předpoklad $k \geq 5$, pak $\frac{k!}{k^4} = 1 \cdot 2 \cdots (k-5) \frac{k-4}{k} \frac{k-3}{k} \frac{k-2}{k} \frac{k-1}{k} k \geq \frac{k-4}{k} \frac{k-3}{k} \frac{k-2}{k} \frac{k-1}{k} k \rightarrow 1 \cdot 1 \cdot 1 \cdot 1 \cdot \infty = \infty$.

(iv): $L = \frac{1}{2}$, $K = 2$, $k_0 = 6$, pak $k \geq k_0 \implies k + 3 \sin(k) \leq k + 3 \leq k + k = 2k$ a díky $3 \leq \frac{1}{2}k$ také $k \geq k_0 \implies k + 3 \sin(k) \geq k - 3 \geq k - \frac{1}{2}k = \frac{1}{2}k$.

9c. Sumy

Uvažujme nějakou konečnou posloupnost $\{a_k\}_{k=n}^m$. Znamená to, že máme nějaká čísla. Co s nimi můžeme dělat? Třeba sečít. Abychom nemuseli pořád psát $a_n + a_{n+1} + \dots + a_m$, zavedeme pro to značení $\sum_{k=n}^m a_k$. Má to jeden drobný zádrhel, tři tečky nejsou zrovna přesné matematické vyjádření. Pokud chceme sumu definovat řádně, musíme to udělat indukcí. Budeme definovat sumy libovolné velikosti, takže si na začátku rovnou vezmeme nekonečnou posloupnost.

! Definice.

Nechť $\{a_k\}_{k=n}^\infty$ je posloupnost. Definujeme

$$\begin{aligned}\sum_{k=n}^n a_k &= a_n, \\ \sum_{k=n}^{m+1} a_k &= \left(\sum_{k=n}^m a_k \right) + a_{m+1} \quad \text{pro } m \geq n.\end{aligned}$$

Pokud $m < n$, pak definujeme $\sum_{k=n}^m a_k = 0$.

Názvosloví: **dolní mez** n , **horní mez** m , **sumační značení sigma** Σ , **sumační index** k .

! Co taková suma vlastně znamená? Reprezentuje určité číslo, které závisí na sčítané posloupnosti, na mezích, ale vůbec ne na k , to je jen pracovní proměnná, která plní svou roli uvnitř sumy, ven ale nepronikne. Je to pěkně vidět z těchto příkladů:

$$\begin{aligned}\sum_{k=1}^4 (2k+1) &= (2 \cdot 1 + 1) + (2 \cdot 2 + 1) + (2 \cdot 3 + 1) + (2 \cdot 4 + 1) = 3 + 5 + 7 + 9 = 24, \\ \sum_{k=13}^{15} a_k &= a_{13} + a_{14} + a_{15}, \\ \sum_{k=n}^m k^2 &= n^2 + (n+1)^2 + (n+2)^2 + \dots + (m-1)^2 + m^2.\end{aligned}$$

Jak vidíme, v žádném výsledku napravo k vůbec není, samozřejmě ale vidíme, že ostatní prvky sumy (členy posloupnosti, meze) tam už vidět jsou, na nich výsledek přirozeně závisí. Protože je k čistě interní a pracovní věc, můžeme si ji nazvat jak chceme, takže $\sum_{k=n}^m a_k = \sum_{i=n}^m a_i = \sum_{h=n}^m a_h = \dots$

! Dokonce můžeme sumační index posouvat, to se někdy hodí. Funguje to jako standardní substituce: Když máme sumu s indexem k , tak si můžeme zvolit nový index, třeba i , který je s k svázán vzorcem typu $i = k + A$. Všechny výskyty k v sumě se pak musí nahradit i , ale přesně podle toho substitučního vzorce, takže namísto k píšeme $k = i - A$, meze se musí také změnit. Například dolní mez je dána podmínkou $k = n$, takže nejmenší hodnota i je $i = k + A = n + A$. Ukážeme příklad:

$$\sum_{k=3}^7 a_k = \left| \begin{array}{l} i = k - 2 \\ k = i + 2 \\ k = 3 \mapsto i = 3 - 2 = 1 \\ k = 7 \mapsto i = 7 - 2 = 5 \end{array} \right| = \sum_{i=1}^5 a_{i+2}.$$

Je to opravdu totéž? Porovnáme ty sumy:

$$\begin{aligned}\sum_{k=3}^7 a_k &= a_3 + a_4 + a_5 + a_6 + a_7, \\ \sum_{i=1}^5 a_{i+2} &= a_{1+2} + a_{2+2} + a_{3+2} + a_{4+2} + a_{5+2} = a_3 + a_4 + a_5 + a_6 + a_7.\end{aligned}$$

Jaké operace budeme se sumami provádět? Jako inspiraci se podíváme na dva příklady s jednoduchými sumami, které si vždy přepíšeme do „dlouhého“ zápisu se všemi členy. Sumu můžeme chtít vynásobit číslem a pak zjednodušit pomocí distributivního zákona:

$$c \sum_{k=1}^3 a_k = c(a_1 + a_2 + a_3) = ca_1 + ca_2 + ca_3 = \sum_{k=1}^3 (ca_k).$$

Můžeme také chtít sumy sčítat, teď pro změnu použijeme komutativitu a asociativitu sčítání.

$$\sum_{k=1}^3 a_k + \sum_{k=1}^3 b_k = (a_1 + a_2 + a_3) + (b_1 + b_2 + b_3) = (a_1 + b_1) + (a_2 + b_2) + (a_3 + b_3) = \sum_{k=1}^3 (a_k + b_k).$$

Obecně platí, že když máme nějaký vztah se sumami, který nám není úplně jasný, tak se vyplatí zkoušit si nějakou kratší sumu rozvinout do dlouhého zápisu a pak je většinou hned vidět, co se děje.

Výše zmíněné dva příklady nás inspirují k obecné definici. Má to ale háček. Pokud například zkoušíme sečíst $\sum_{k=1}^2 a_k$ a $\sum_{k=1}^1 b_k$, nedokážeme už výsledný výraz $(a_1 + b_1) + a_2$ zapsat jednou sumou. Jinými slovy, abychom mohli sčítat, potřebujeme stejné meze u zúčastněných sum.

!

Definice.

Uvažujme sumy $\sum_{k=n}^m a_k$, $\sum_{k=n}^m b_k$ a $c \in \mathbb{R}$. Definujeme

$$c \cdot \left(\sum_{k=n}^m a_k \right) = \sum_{k=n}^m (c \cdot a_k),$$

$$\left(\sum_{k=n}^m a_k \right) + \left(\sum_{k=n}^m b_k \right) = \sum_{k=n}^m (a_k + b_k).$$

Všimněte si, že obě rovnosti lze číst oběma směry. První z nich při čtení zprava doleva říká, že ze sumy lze vytknout společný násobící faktor. Druhá při čtení zprava doleva říká, že sumu se členy, které umíme napsat jako součty (stejným způsobem), lze roztrhnout na dvě.

Pokud máme dvě sumy, potřebujeme je sečíst a jejich meze indexů nejsou stejné, pak jsou dvě možnosti. Jestliže mají obě sumy stejnou „délku“ (stejný počet členů), tak se dá u jedné z nich indexace posunout, aby už meze souhlasily. Pokud jsou počty členů různé, tak už posun nepomůže. Pokud je ale opravdu nutně potřebujeme sečíst do jedné sumy, budeme muset z té delší nějaké členy odtrhnout. To není problém, sumy se dají rozpojovat a spojovat, pokud indexace navazuje, třeba

$$\sum_{k=1}^3 a_k + \sum_{k=4}^5 a_k = (a_1 + a_2 + a_3) + (a_4 + a_5) = a_1 + a_2 + a_3 + a_4 + a_5 = \sum_{k=1}^5 a_k,$$

$$\sum_{k=12}^{27} c_k = \sum_{k=12}^{20} c_k + \sum_{k=21}^{27} c_k.$$

Posvětíme si to oficiálně:

!

Fakt 9c.1.

Nechť $\{a_k\}_{k=n_0}^{\infty}$ je posloupnost. Pro libovolné $m, n, p \in \mathbb{Z}$ splňující $n_0 \leq n \leq m \leq p$ platí

$$\sum_{k=n}^p a_k = \sum_{k=n}^m a_k + \sum_{k=m+1}^p a_k.$$

Tuto rovnost je možné používat oběma směry, tedy rozdělovat sumu na části či sumy spojovat do jedné. Lze pak i odečítat, například $\sum_{k=1}^8 a_k - \sum_{k=4}^8 a_k = \sum_{k=1}^3 a_k$ (rozepište si to, pokud to ještě nevidíte).

Příklad 9c.a: Chceme sečíst $\sum_{k=1}^3 a_k$ a $\sum_{k=2}^6 b_k$. Nejprve v druhé sumě posuneme index tak, aby indexace začínala jedničkou, to zajistí substituce $i = k - 1$, pak nejmenší hodnota $k = 2$ přejde na novou dolní mez $i = 2 - 1 = 1$. Protože má druhá suma více členů, nebude nám ale souhlasit horní mez, tak členy navíc odebereme. Protože chceme sčítat, vrátíme se v druhé sumě od indexu i zpět k indexu k , abychom měli stejné písmenka. Není to nutné, ale méně to mate.

$$\sum_{k=1}^3 a_k + \sum_{k=2}^6 b_k = \sum_{k=1}^3 a_k + \sum_{i=1}^5 b_{i+1} = \sum_{k=1}^3 a_k + \sum_{k=1}^5 b_{k+1} = \sum_{k=1}^3 a_k + \sum_{k=1}^3 b_{k+1} + b_5 + b_6 = \sum_{k=1}^3 (a_k + b_{k+1}) + b_4 + b_5.$$

△

Je důležité umět se sumami hbitě pracovat, ale jak jsme právě viděli, jde vlastně o staré dobré algebraické triky, jen v novém kabátě, takže to snad nebude problém.

Dobrá otázka je, kolik je součet dané sumy. Pokud je krátká, tak vždycky můžeme sundat boty a začít sčítat na prstech, ale pro delší sumy či sumy s proměnnýmimezemi je dobré znát nějaké vzorce. Začneme součtem asi nejužitečnější posloupnosti.

! Věta 9c.2. (součet geometrické posloupnosti)

Uvažujme $q \in \mathbb{R}$. Pak

$$\sum_{k=0}^n q^k = \begin{cases} \frac{1-q^{n+1}}{1-q}, & q \neq 1; \\ n+1, & q = 1. \end{cases}$$

Důkaz (poučný): Označme $s_n = \sum_{k=0}^n q^k = 1 + q + q^2 + \dots + q^n$. Pak $q \cdot s_n = q + q^2 + q^3 + \dots + q^{n+1}$, proto

$qs_n - s_n = q^{n+1} - 1$, tedy $(q-1)s_n = q^{n+1} - 1$. Jestliže je $q \neq 1$, můžeme vydělit a máme $s_n = \frac{q^{n+1}-1}{q-1}$.

Kdyby $q = 1$, tak rovnou vidíme, že $\sum_{k=0}^n 1 = 1 + 1 + \dots + 1$, sčítáme 1 celkem $(n+1)$ -krát (opravdu? začínáme s indexem 0, tím to vyskočí o jedno, zkuste si pár sum), takže $s_n = n+1$.

□

Tento základní vzorec pak umožní zpracovat i sumy obecnějších geometrických výrazů. Sčítat $\sum_{k=0}^n aq^k$ je snadné, stačí to a vytknout a můžeme použít již známý vzorec. Zajímavější případ je, pokud indexace nezačíná nulou. Pak se dá s úspěchem použít vytýkací trik. Je velice snadný, jak uvidíme z následujícího příkladu, jako obvykle nám výrazně pomůže dlouhý zápis:

$$\sum_{k=13}^{16} aq^k = aq^{13} + aq^{14} + aq^{15} + aq^{16} = q^{13}(a + aq + aq^2 + aq^3) = q^{13} \sum_{k=0}^3 aq^k.$$

Obecně dojdeme ke vzorci (pro $q \neq 1$)

$$\sum_{k=N}^n aq^k = q^N a \frac{1 - q^{n-N+1}}{1 - q}.$$

Ukážeme si formální důkaz, použijeme v něm vytáknutí společného násobku ze sumy a substituci:

$$\sum_{k=N}^n aq^k = \sum_{k=N}^n aq^N q^{k-N} = aq^N \sum_{k=N}^n q^{k-N} = \left| \begin{array}{l} i = k - N \\ k = 3 \mapsto i = N - N = 0 \\ k = n \mapsto i = n - N \end{array} \right| = aq^N \sum_{i=0}^{n-N} q^i = q^N a \frac{1 - q^{n-N+1}}{1 - q}.$$

Dalším populárním typem výrazu, který často sčítáme, jsou mocniny.

Věta 9c.3. (součty mocnin)

Následující vzorce platí pro všechna $n \in \mathbb{N}$:

- (i) $\sum_{k=1}^n 1 = n$;
- (ii) $\sum_{k=1}^n k = \frac{1}{2}n(n+1)$;
- (iii) $\sum_{k=1}^n k^2 = \frac{1}{6}n(n+1)(2n+1)$;
- (iv) $\sum_{k=1}^n k^3 = \frac{1}{4}n^2(n+1)^2$.

Důkaz (poučný): Takovéto vzorce se evidentně dokazují matematickou indukcí, (ii) a (iii) máme jako cvičení v kapitole o indukci, (iv) si laskavý čtenář během 17 vterin dodělá sám. Správná otázka ale zní, jak se na ty vzorce přijde. Tak (i) je snadné, prostě sčítáme n jedniček. Což takhle (ii)?

Vzorec pro tento součet vymyslel Gauss, když byl malé nechutně chytré dítě. Učitel jej nechal sečíst prvních 100 čísel, ať má od něj chvíli pokoj, takže byl notně překvapen, když mu malý Johann za chvíli hlásil výsledek. Jak na to přišel? Představme si takový součet, pro začátek pro sudé n :

$$1 + 2 + 3 + \dots + (n-2) + (n-1) + n.$$

Všimněte si, že první a poslední číslo dají dohromady $n + 1$. Také druhé číslo zleva a zprava dají dohromady $n + 1$. Také třetí ... Takových dvojic je přesně $\frac{n}{2}$, takže součet je $\frac{n}{2}(n + 1)$.

Co když je n liché? Pak máme $\frac{1}{2}(n - 1)$ dvojic a prostřední číslo zůstane osamocené, je to číslo $\frac{1}{2}(n + 1)$ (zkuste si to na nějakém příkladě). Celkový součet je tedy $\frac{1}{2}(n - 1)(n + 1) + \frac{1}{2}(n + 1) = \frac{1}{2}n(n + 1)$.

Jiný trik jak to vidět: Napíšeme si tu sumu dvakrát pod sebe, jednou jako $s = 1 + 2 + \dots + (n - 1) + n$, podruhé jako $s = n + (n - 1) + \dots + 2 + 1$, když sečteme, dostaneme nalevo $2s$, napravo n dvojic se součtem $n + 1$, teď už nemusíme rozlišovat mezi sudými a lichými n .

Některí autoři tu příhodu s Gaussem považují za bajku, takže si ukážeme jiný postup.

Začíná takto: Víme, že $(k + 1)^2 - k^2 = 2k + 1$. Co dostaneme, když takovéto členy začneme sčítat? Nejprve příklad:

$$\sum_{k=1}^3 [(k+1)^2 - k^2] = [2^2 - 1^2] + [3^2 - 2^2] + [4^2 - 3^2] = 4^2 - 1^2.$$

Jinými slovy, všechny „prostřední“ členy se pokrátí. Obecně máme toto:

$$\sum_{k=1}^n [(k+1)^2 - k^2] = (n+1)^2 - 1^2,$$

ale také

$$\sum_{k=1}^n [(k+1)^2 - k^2] = \sum_{k=1}^n [2k + 1] = 2 \sum_{k=1}^n k + \sum_{k=1}^n 1 = 2 \sum_{k=1}^n k + n.$$

Proto

$$(n+1)^2 - 1 = 2 \sum_{k=1}^n k + n \implies \sum_{k=1}^n k = \frac{1}{2}[(n+1)^2 - 1 - n] = \frac{1}{2}n(n+1).$$

(iii): Podobný trik.

$$\sum_{k=1}^n [(k+1)^3 - k^3] = (n+1)^3 - 1^3,$$

ale také

$$\sum_{k=1}^n [(k+1)^3 - k^3] = \sum_{k=1}^n [3k^2 + 3k + 1] = 3 \sum_{k=1}^n k^2 + 3 \cdot \sum_{k=1}^n k + \sum_{k=1}^n 1 = 3 \sum_{k=1}^n k^2 + \frac{3}{2}n(n+1) + n.$$

Máme tedy

$$(n+1)^3 - 1 = 3 \sum_{k=1}^n k^2 + \frac{3}{2}n(n+1) + n \implies \sum_{k=1}^n k^2 = \frac{1}{3}((n+1)^3 - 1 - \frac{3}{2}n(n+1) - n) = \frac{1}{6}n(n+1)(2n+1).$$

(iv): Podobný trik. Všimněte si, že při nalezení součtu $\sum k^2$ jsme potřebovali znát součet $\sum k$ a součet $\sum 1$. Teď si budeme hrát s $(k+1)^4 - k^4$ a budeme potřebovat znát součty $\sum k^2$, $\sum k$ a $\sum 1$. Přenecháme to čtenáři, pokud se ještě nudí, může si zkousit nalézt $\sum k^4$.

□

Gaussův trik je užitečný, protože díky němu získáme okamžitě také součet sumy, která má ustříhnutý začátek: $\sum_{k=m}^n k = \frac{1}{2}(n-m+1)(n+m)$. Samozřejmě to lze vždy získat rozdílem dvou sum, třeba

$$\sum_{k=m}^n k^2 = \sum_{k=1}^n k^2 - \sum_{k=1}^{m-1} k^2 = \frac{1}{6}n(n+1)(2n+1) - \frac{1}{6}(m-1)m(2m-1).$$

Alternativní způsob, jak najít uzavřené vzorce pro rozličné součty, čtenář najde v kapitole , viz příklad a cvičení .

Příklad 9c.b:

$$\sum_{k=50}^{150} k = \sum_{k=1}^{150} k - \sum_{k=1}^{49} k = \frac{1}{2}150 \cdot 151 - \frac{1}{2}49 \cdot 50 = 25 \cdot (3 \cdot 151 - 49) = 10100.$$

△

Ještě jednu sumu se naučíme sčítat, a to aritmetickou.

Věta 9c.4. (součet aritmetické posloupnosti)

Uvažujme čísla $a, d \in \mathbb{R}$. Pak

$$\sum_{k=0}^n (a + dk) = (n+1)a + \frac{1}{2}n(n+1)d.$$

Důkaz (rutinní): Toto je snadné,

$$\sum_{k=0}^n (a + dk) = \sum_{k=0}^n a + d \sum_{k=0}^n k = (n+1)a + \frac{1}{2}n(n+1)d.$$

□

Trochu jiný (a obecnější) výsledek lze dostat zase pomocí Gaussova triku. Zkuste si na několika příkladech ověřit, že jestliže $\{a_k\}$ je libovolná aritmetická posloupnost, pak zase napsáním dvou kopií pod sebe, jen v opačném pořadí, dostáváme dvojice se stále stejným součtem. Dostáváme tak vzorec $\sum_{k=N}^n a_k = \frac{1}{2}(n-N+1)(a_N + a_n)$.

Se sumami se dají dělat zajímavé věci. Jedna je možnost sčítat přes indexy, které netvoří souvislou posloupnost od dolní meze k horní. Pokud si zvolíme konečnou množinu $M \subseteq \mathbb{Z}$, můžeme definovat $\sum_{k \in M} a_k$. Význam je zjevný, například volba $M = \{13, 23\}$ dává $\sum_{k \in M} a_k = a_{13} + a_{23}$. Pro úplnost ještě zadefinujeme, že $\sum_{k \in \emptyset} a_k = 0$.

Ještě zajímavější to je, pokud je součástí sumy další suma. Ty se dají vždy rozbalit, někdy to jde lépe zevnitř, jindy zvenčí.

Příklad 9c.c: a) U výrazu $\sum_{i=1}^2 \sum_{j=1}^3 i^2 j$ to vyjde v zásadě natejno. Nejprve ukážeme postup, kdy začneme rozepisovat vnější sumu, pak postup, kdy začneme zevnitř.

$$\begin{aligned} \sum_{i=1}^2 \sum_{j=1}^3 i^2 j &= \sum_{j=1}^3 1^2 j + \sum_{j=1}^3 2^2 j = (1^2 \cdot 1 + 1^2 \cdot 2 + 1^2 \cdot 3) + (2^2 \cdot 1 + 2^2 \cdot 2 + 2^2 \cdot 3) = 30, \\ \sum_{i=1}^2 \sum_{j=1}^3 i^2 j &= \sum_{i=1}^2 (i^2 \cdot 1 + i^2 \cdot 2 + i^2 \cdot 3) = \sum_{i=1}^2 6i^2 = 6 \cdot 1^2 + 6 \cdot 2^2 = 30. \end{aligned}$$

b) U následující sumy je rozvíjení zvenčí výrazně snažší, protože pak se dozvímme, jaké jsou vlastně meze vnitřní sumy. Pokud začneme zevnitř, budeme muset sčítat $\sum_j j$ s proměnnýmimezemi.

$$\begin{aligned} \sum_{i=1}^3 \sum_{j=i}^3 (i+j) &= \sum_{j=1}^3 (1+j) + \sum_{j=2}^3 (2+j) + \sum_{j=3}^3 (3+j) \\ &= [(1+1) + (1+2) + (1+3)] + [(2+2) + (2+3)] + [(3+3)] = 24 \\ \sum_{i=1}^3 \sum_{j=i}^3 (i+j) &= \sum_{i=1}^3 \left[i \cdot \sum_{j=i}^3 1 + \sum_{j=i}^3 j \right] = \sum_{i=1}^3 \left[i \cdot (3-i+1) + \frac{1}{2}(3-i+1)(i+3) \right] = \frac{1}{2} \sum_{i=1}^3 (9i - 3i^2 + 12) \\ &= \frac{1}{2} \left[9 \sum_{i=1}^3 i - 3 \sum_{i=1}^3 i^2 + 12 \sum_{i=1}^3 1 \right] = \frac{1}{2} \left[9 \cdot \frac{1}{2} \cdot 3 \cdot 4 - 3 \cdot \frac{1}{6} \cdot 3 \cdot 4 \cdot 7 + 12 \cdot 3 \right] = \frac{1}{2} \cdot 48 = 24. \end{aligned}$$

c) Pokud ale vnější suma nemá konkrétní meze, pak rozvíjením zvenčí stejně nic nezískáme, ani to dokonce nejde (když nevíme konkrétně, jaké indexy vnější suma používá). Pak nezbývá než se s tím poprat zevnitř.

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^i \sum_{k=1}^j 1 &= \sum_{i=1}^n \sum_{j=1}^i j = \sum_{i=1}^n \frac{1}{2}i(i+1) = \frac{1}{2} \left[\sum_{i=1}^n i^2 + \sum_{i=1}^n i \right] \\ &= \frac{1}{2} \left[\frac{1}{6}n(n+1)(2n+1) + \frac{1}{2}n(n+1) \right] = \frac{1}{6}n(n+1)(n+2). \end{aligned}$$

△

Posloupnosti je možné také násobit. Neformálně jsme to už dělali v kapitole .

!

Definice.

Nechť $\{a_k\}_{k=n}^{\infty}$ je posloupnost. Definujeme

$$\prod_{k=n}^n a_k = a_n,$$

$$\prod_{k=n}^{m+1} a_k = \left(\prod_{k=n}^m a_k \right) \cdot a_{m+1} \quad \text{pro } m \geq n.$$

Pokud $m < n$, pak definujeme $\prod_{k=n}^m a_k = 1$.

Takže například $\prod_{k=3}^7 a_k = a_3 \cdot a_4 \cdot a_5 \cdot a_6 \cdot a_7$. Toto značení je velice pohodlné hlavně v případě, kdy je počet činitelů dán obecně, například můžeme psát $n! = \prod_{k=1}^n k$. Všimněte si, že díky definici prázdného součinu jako jedničky tento vzorec funguje i pro $n = 0$.

Cvičení

Cvičení 9c.1 (rutinní): Spočítejte následující sumy přímým výpočtem, bez použití vzorců.

(i) $\sum_{k=1}^5 (k+1)$;	(iv) $\sum_{k=0}^8 (2^{k+1} - 2^k)$;	(vii) $\sum_{k=0}^4 (3k+2)$;
(ii) $\sum_{i=0}^4 (-3)^i$;	(v) $\sum_{k=0}^4 3 \cdot 2^k$;	(viii) $\sum_{i=1}^4 \sum_{j=2}^i (i-j)$;
(iii) $\sum_{k=1}^{10} 13$;	(vi) $\sum_{k=0}^8 (1 + (-1)^k)$;	(ix) $\sum_{i=1}^4 \sum_{j=i}^4 j$.

Cvičení 9c.2 (rutinní): Spočítejte následující sumy pomocí vhodných vzorců.

(i) $\sum_{k=0}^{25} \left(\frac{1}{2}\right)^k$;	(iv) $\sum_{k=23}^{32} \left(\frac{3}{2}\right)^k$;	(vii) $\sum_{k=100}^{200} (2k+1)$;
(ii) $\sum_{k=3}^{33} 2^k$;	(v) $\sum_{k=0}^{12} (3^k - 2^k)$;	(viii) $\sum_{i=1}^{10} \sum_{j=i}^{20} 12$;
(iii) $\sum_{k=10}^{100} 10^k$;	(vi) $\sum_{k=0}^{\infty} \left(\frac{4}{5}\right)^k$;	(ix) $\sum_{i=1}^{10} \sum_{j=1}^i 12j$.

Cvičení 9c.3 (rutinní): Slučte následující sumy do jedné.

(i) $3 \sum_{k=-2}^7 k + \sum_{i=-2}^7 (1-i)$;	(iii) $\sum_{i=-2}^{10} i^2 - \sum_{j=1}^{13} j^2$;
(ii) $\sum_{n=1}^{123} \frac{1}{n} + \sum_{j=1}^{123} \frac{j-1}{j}$;	(iv) $\sum_{i=3}^{13} i - \sum_{m=0}^{13} m^2$.

Cvičení 9c.4 (dobrý, poučný): Dokažte, že pro libovolné $k, m \in \mathbb{N}_0$ platí

$$(x^{(m-1)k} + x^{(m-2)k} + \cdots + x^{3k} + x^{2k} + x^k + 1)(x^k - 1) = x^{km} - 1.$$

Nápowěda: Zapište si ten levý polynom jako sumu, pak roznásobte ten pravý a sjednoťte sumy.

Všimněte si, že pro $k = 1$ dostáváte $(x^m - 1) = (x-1)(x^{m-1} + x^{m-2} + \cdots + x + 1)$, zobecnění známého vzorce $x^2 - 1 = (x-1)(x+1)$.

Řešení:

9c.1: (i): $2 + 3 + 4 + 5 + 6 = 20$; (ii): $1 - 3 + 9 - 27 + 81 = 61$; (iii): $13 + 13 + \cdots + 13 = 10 \cdot 13 = 130$;
(iv): $(2^1 - 2^0) + (2^2 - 2^1) + (2^3 - 2^2) + (2^4 - 2^3) + (2^5 - 2^4) + (2^6 - 2^5) + (2^7 - 2^6) + (2^8 - 2^7) + (2^9 - 2^8) = 2^9 - 1 = 511$;
(v): $= 3(1 + 2 + 4 + 8 + 16) = 3 \cdot 31 = 93$; (vi): $2 + 0 + 2 + 0 + 2 + 0 + 2 + 0 + 2 = 10$; (vii): $2 + 5 + 8 + 11 + 14 = 40$;
(viii): $\sum_{j=2}^1 (1-j) + \sum_{j=2}^2 (2-j) + \sum_{j=2}^3 (3-j) + \sum_{j=2}^4 (4-j) = 0 + 0 + (1+0) + (2+1+0) = 4$;
(ix): $\sum_{j=1}^4 j + \sum_{j=2}^4 j + \sum_{j=3}^4 j + \sum_{j=4}^4 j = (1+2+3+4) + (2+3+4) + (3+4) + 4 = 30$.

9c.2: (i): $\frac{1 - (\frac{1}{2})^{26}}{1 - \frac{1}{2}} = 2(1 - (\frac{1}{2})^{26}) = 2 - \frac{1}{2^{25}}$; (ii): $2^3 \sum_{k=0}^{30} 2^k = 2^3 \frac{1 - 2^{31}}{1 - 2} = 2^3(2^{31} - 1) = 17179869176$

nebo $\sum_{k=0}^{33} 2^k - 2^0 - 2^1 - 2^2 = \frac{1-2^{34}}{1-2} - 1 - 2 - 4 = 2^{34} - 8 = 17179869176$;

(iii): $10^{10} \sum_{k=0}^{90} 10^k = 10^{10} \frac{1-10^{91}}{1-10} = 10^{10} \frac{1}{9} (10^{91} - 1)$; (iv): $(\frac{3}{2})^{23} \sum_{k=0}^9 (\frac{3}{2})^k = (\frac{3}{2})^{23} \frac{1 - (\frac{3}{2})^{10}}{1 - \frac{3}{2}} = 2(\frac{3}{2})^{23} ((\frac{3}{2})^{10} - 1)$;

(v): $\sum_{k=0}^{12} 3^k - \sum_{k=0}^{12} 2^k = \frac{1-3^{13}}{1-3} - \frac{1-2^{13}}{1-2} = \frac{1}{2}(3^{13} - 1) + 2^{13} - 1 = 805352$; (vi): $\frac{1}{1 - \frac{4}{5}} = 5$;

(vii): $2 \left(\sum_{k=1}^{200} k - \sum_{k=1}^{99} k \right) + \sum_{k=100}^{200} 1 = 2 \left(\frac{1}{2} 200 \cdot 201 - \frac{1}{2} 99 \cdot 100 \right) - (200 - 100 + 1) \cdot 1 = 30199$;

(viii): $\sum_{i=1}^{10} (20 - i + 1) \cdot 12 = 12 \left(\sum_{i=1}^{10} 21 - \sum_{i=1}^{10} i \right) = 12(10 \cdot 21 - \frac{1}{2} 10 \cdot 11) = 1860$;

(ix): $\sum_{i=1}^{10} 12 \sum_{j=1}^i j = \sum_{i=1}^{10} 12 \frac{1}{2} i(i+1) = \sum_{i=1}^{10} 6i^2 + \sum_{i=1}^{10} 6i = 6 \cdot \frac{1}{6} 10 \cdot 11 \cdot 21 + 6 \cdot \frac{1}{2} 10 \cdot 11 = 5640$.

9c.3: (i): $= 3 \sum_{k=-2}^7 k + \sum_{k=-2}^7 (1-k) = \sum_{k=-2}^7 (3k + 1 - k) = \sum_{k=-2}^7 (2k + 1)$.

(ii): $= \sum_{n=1}^{123} \frac{1}{n} + \sum_{n=1}^{123} \frac{n-1}{n} = \sum_{n=1}^{123} \left(\frac{1}{n} + \frac{n-1}{n} \right) = \sum_{n=1}^{123} 1 = 123$.

(iii): $= \begin{vmatrix} j = i+3 \\ i = j-3 \\ i = -2 \mapsto j = 1 \\ i = 10 \mapsto j = 13 \end{vmatrix} = \sum_{j=1}^{13} (j-3)^2 - \sum_{j=1}^{13} j^2 = \sum_{j=1}^{13} [(j-3)^2 - j^2] = \sum_{j=1}^{13} (9-6j)$.

(iv): $= \begin{vmatrix} m = i-2 \\ i = m+2 \\ i = 3 \mapsto m = 1 \\ i = 13 \mapsto m = 11 \end{vmatrix} = \sum_{m=1}^{11} (m+2) - \sum_{m=1}^{13} m^2 = \sum_{m=1}^{11} (m+2) - \left(\sum_{m=1}^{11} m^2 + 12^2 + 13^2 \right)$
 $= \sum_{m=1}^{11} (m+2 - m^2) - 12^2 - 13^2$.

9c.4:

$$\begin{aligned} \left(\sum_{i=0}^{m-1} x^{ik} \right) (x^k - 1) &= \left(\sum_{i=0}^{m-1} x^{ik} \right) x^k - \sum_{i=0}^{m-1} x^{ik} = \sum_{i=0}^{m-1} x^{(i+1)k} - \sum_{i=0}^{m-1} x^{ik} = \sum_{j=1}^m x^{jk} - \sum_{i=0}^{m-1} x^{ik} \\ &= x^{mk} + \sum_{j=1}^{m-1} x^{jk} - \sum_{i=1}^{m-1} x^{ik} - x^{0 \cdot k} = x^{mk} - 1. \end{aligned}$$

9d. Řady

Tato kapitola je silně doplňková. Čtenář ji bude potřebovat v zásadě jen tehdy, pokud bude chtít řešit rekurentní rovnice pomocí generujících funkcí (kapitola).

Jedna z věcí, kterou s čísly běžně děláme, je jejich sečtení. Když máme nekonečnou posloupnost, tak máme čísel nekonečně mnoho, nicméně to neodradí odhodlaného průkopníka od pokusu sečít i je. Zjevně to ale nebude nic lehkého, počínaje filosofickou otázkou, zda vůbec lze v konečném čase, který nám v životě zbývá, provést nekonečně mnoho sčítání.

Matematická analýza k tomuto problému přistupuje tradičním způsobem, snažíme se vyjít od toho, co umíme. Myšlenka je jednoduchá, začneme prostě čísla postupně sčítat a díváme se, jak se chovají obdržené mezivýsledky. Pokud se po čase v zásadě přestávají měnit, tak usoudíme, že jsme se dostali k číslu, které představuje součet úplně všech členů dotyčné posloupnosti.

Definice.

Nechť $\{a_k\}_{k=n}^{\infty}$ je posloupnost. Pro $N \in \mathbb{N}$, $N \geq n$ definujeme **částečné součty** jako $s_N = \sum_{k=n}^N a_k$.

Řekneme, že **řada** $\sum_{k=n}^{\infty} a_k$ **konverguje** k číslu A , značeno $\sum_{k=n}^{\infty} a_k = A$, jestliže $\lim_{N \rightarrow \infty} (s_N) = A$. Jinak řekneme, že dotyčná řada **diverguje**.

Pokud $\lim_{N \rightarrow \infty} (s_N) = \infty$, pak značíme $\sum_{k=n}^{\infty} a_k = \infty$.

Let $\{a_k\}_{k=n}^{\infty}$ be a sequence. For $N \in \mathbb{N}$, $N \geq n$ we define **partial sums** by $s_N = \sum_{k=n}^N a_k$.

We say that the series $\sum_{k=n}^{\infty} a_k$ **converges** to a number A , denoted $\sum_{k=n}^{\infty} a_k = A$, if $\lim_{N \rightarrow \infty} (a_N) = A$. Otherwise we say that the series **diverges**.

Příklad 9d.a: 1) Uvažujme řadu $\sum_{k=1}^{\infty} 0 = 0 + 0 + 0 + 0 + \dots$.

Tipneme si, že by součet měl vyjít 0. Pozná to naše definice?

Vezměme $N \in \mathbb{N}$, příslušný částečný součet pak je $s_N = \sum_{k=1}^N 0 = 0 + 0 + \dots + 0 = 0$ (sčítáme N nul). Když pošleme

N do nekonečna, tak evidentně $s_N \rightarrow 0$, a proto podle definice zkoumaná řada konverguje, navíc $\sum_{k=1}^{\infty} 0 = 0$.

2) Uvažujme řadu $\sum_{k=1}^{\infty} 1 = 1 + 1 + 1 + 1 + \dots$.

Tipneme si, že by součet měl vyjít ∞ . Pozná to naše definice?

Vezměme $N \in \mathbb{N}$, příslušný částečný součet pak je $s_N = \sum_{k=1}^{\infty} 1 = 1 + 1 + \dots + 1 = N$ (sčítáme N jedniček).

Když pošleme N do nekonečna, tak evidentně $s_N \rightarrow \infty$, a proto podle definice zkoumaná řada diverguje, navíc $\sum_{k=1}^{\infty} 1 = \infty$.

3) Uvažujme řadu $\sum_{k=0}^{\infty} (-1)^k = (-1)^0 + (-1)^1 + (-1)^2 + (-1)^3 + \dots = 1 - 1 + 1 - 1 + \dots$

Tady není jasné, co čekat, zeptáme se definice. Nejdříve se podíváme na pár částečných součtů pro inspiraci.

$$s_2 = \sum_{k=0}^2 (-1)^k = (-1)^0 + (-1)^1 + (-1)^2 = 1 - 1 + 1 = 1, \quad s_3 = \sum_{k=0}^3 (-1)^k = 1 - 1 + 1 - 1 = 0,$$

$$s_4 = \sum_{k=0}^4 (-1)^k = 1 - 1 + 1 - 1 + 1 = 1, \quad s_5 = \sum_{k=0}^5 (-1)^k = 1 - 1 + 1 - 1 + 1 - 1 = 0.$$

A teď obecně. Vezměme $N \in \mathbb{N}$, příklady a zamýšlení naznačují, že příslušný částečný součet pak je $s_N = \sum_{k=0}^N (-1)^k = 0$ pro N liché a $s_N = 1$ pro N sudé. Posloupnost $\{s_N\} = \{1, 0, 1, 0, 1, 0, \dots\}$ nemá limitu, tudíž ani

příslušná řada nemůže konvergovat. Závěr je, že $\sum_{k=0}^{\infty} (-1)^k$ diverguje.

Na rozdíl od předchozího příkladu teď nemáme součet nekonečno, takže už k tomu není co dodat.

4) Uvažujme řadu $\sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$.

Jak vypadají částečné součty?

$$s_2 = \sum_{k=1}^2 \left(\frac{1}{2}\right)^k = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}, \quad s_3 = \sum_{k=1}^3 \left(\frac{1}{2}\right)^k = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8},$$

$$s_4 = \sum_{k=1}^4 \left(\frac{1}{2}\right)^k = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16}.$$

Odhadneme, že pro $N \in \mathbb{N}$ platí $s_N = \frac{2^N - 1}{2^N} = 1 - \frac{1}{2^N}$. Dokážeme to matematickou indukcí:

$$(0) \text{ Pro } N = 1: s_1 = \sum_{k=1}^1 \left(\frac{1}{2}\right)^k = \frac{1}{2} = 1 - \frac{1}{2^1}.$$

(1) Vezměme libovolné $n \in \mathbb{N}$ a předpokládejme, že $s_N = 1 - \frac{1}{2^N}$. Pak

$$s_{N+1} = \sum_{k=1}^{N+1} \left(\frac{1}{2}\right)^k = \sum_{k=1}^N \left(\frac{1}{2}\right)^k + \frac{1}{2^{N+1}} = 1 - \frac{1}{2^N} + \frac{1}{2^{N+1}} = 1 - \frac{2-1}{2^{N+1}} = 1 - \frac{1}{2^{N+1}}.$$

Máme potvrzeno, že $s_N = 1 - \frac{1}{2^N}$, a protože $2^N \rightarrow \infty$, dostáváme $\lim(s_N) = 1 - 0 = 1$.

Závěr: $\sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k$ konverguje a $\sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = 1$.

5) Uvažujme řadu $\sum_{k=1}^{\infty} \frac{1}{k} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \dots$.

Jak vypadají částečné součty?

$$s_2 = \sum_{k=1}^2 \frac{1}{k} = 1 + \frac{1}{2} = \frac{3}{2}, \quad s_3 = \sum_{k=1}^3 \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} = \frac{11}{6},$$

$$s_4 = \sum_{k=1}^4 \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} = \frac{25}{12}, \quad s_5 = \sum_{k=1}^5 \frac{1}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} = \frac{137}{60}.$$

Nevím jak vy, ale já v tom nic pěkného nevidím. Je to tím, že rozumné vyjádření pro s_n opravdu neexistuje, tahle řada je docela drsná. Zajímavým vtipným trikem, popřípadě snadným analytickým výpočtem se dá ukázat, že tato řada diverguje a $\sum_{k=1}^{\infty} \frac{1}{k} = \infty$.

△

První tři příklady ukazují možné chování řad. Řady konvergují či divergují, a ty divergující mohou divergovat zajímavě (nasčítat se do nekonečna či mínus nekonečna) nebo nějakou oscilací, kdy už o výsledku sčítání nejde říct vůbec nic.

Čtvrtý příklad ukazuje, že konvergovat mohou i jiné řady než ta triviální nulová. V této souvislosti je dobré připomenout jeden výsledek zmatematické analýzy, že nutnou podmínkou konvergence řady je, aby její jednotlivé členy jako posloupnost šly k nule. Pak už je jasné, proč v případech 2) a 3) máme divergenci. Není to ale podmínka postačující, jak ukazuje příklad 5).

Ve skutečnosti je to tak, že konvergentní řady jsou ty, jejichž členy jdou k nule a to dostatečně rychle, přičemž význam „dostatečně“ závisí mimo jiné i na tom, jaká se v řadě vyskytuje znaménka. Porovnejme následující známé výsledky z analýzy:

$$\sum_{k=1}^{\infty} \frac{1}{k} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots = \infty,$$

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = \frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \text{ konverguje,}$$

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \dots \text{ konverguje.}$$

Rychlosť klesání členů posloupnosti $\left\{ \frac{1}{k} \right\}$ ještě není dostatečná k tomu, aby je šlo sečít se zdárným výsledkem (konvergentní řada). Pokud ale u každého druhého členu změníme znaménka, tak už se tato čísla rozumně nasčítají, mimochodem, výsledek je $\ln(2)$. Členy posloupnosti $\left\{ \frac{1}{k^2} \right\}$ klesají k nule tak rychle, že i když jím necháme plusy a posčítáme je, dostaneme rozumný výsledek.

Rozpozнат konvergenci řady je značně náročný problém a nejsou pro to jednoduché mechanismy. Analýza nabízí celou řadu nástrojů, ale nebudeme to zde potřebovat. Ukážeme si jeden typ řady, který umíme zvládnout, vrátme se na chvíli k příkladu 4). Můžeme si všimnout, že vlastně sčítáme geometrickou posloupnost, a pro částečný součet jsme měli výsledek, podle Věty máme $\sum_{k=0}^N q^k = \frac{1-q^{N+1}}{1-q}$ pro $q \neq 1$. Ověřte si, že to souhlasí s výsledkem, který jsme tam odvodili—pozor na odlišný začátek indexace, řadě v příkladě 4) chyběl první člen $k=0$ neboli jednička.

Když v tom vzorci pošleme $N \rightarrow \infty$, dostaneme obecný výsledek. Z něj se dá odvodit ještě jeden, který se bude hodit, tak jej přidáme.

Věta 9d.1.

- (i) $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$ pro $|q| < 1$;
- (ii) $\sum_{k=0}^{\infty} (k+1)q^k = \frac{1}{(1-q)^2}$ pro $|q| < 1$.

Řadám typu $\sum_{k=0}^{\infty} q^k$ říkáme **geometrická řada** a jsou velice užitečné. Všimněte si mimochodem, že první čtyři příklady výše spadají do této kategorie.

Operace s řadami.

S řadami se dá manipulovat podobně jako se sumami. Můžeme posouvat indexy substitucí, můžeme řady sčítat a násobit číslem.

Definice.

Uvažujme řady $\sum_{k=0}^{\infty} a_k$ a $\sum_{k=0}^{\infty} b_k$, nechť $c \in \mathbb{R}$. Definujeme operace

$$c\left(\sum_{k=0}^{\infty} a_k\right) = \sum_{k=0}^{\infty} (ca_k),$$

$$\sum_{k=0}^{\infty} a_k + \sum_{k=0}^{\infty} b_k = \sum_{k=0}^{\infty} (a_k + b_k).$$

Tato definice je čistě formální, je to návod, jak sestavovat nové řady pomocí určitých pravidel. Někdo nám dá dvě řady, rozhodneme se je sečíst, tak si z obou vytáhneme koeficienty, po dvou sečteme a z výsledků sestavíme novou řadu. Dobrá otázka zní, co se děje se součty řad, když s nimi provádíme tyto operace. Mají součty nově vyrobených řad něco společného se součty řad původních?

Obecně to je občas zajímavé. Pokud například sečteme dvě řady, $\sum_{k=1}^{\infty} 1$ a $\sum_{k=1}^{\infty} (-1)$, které divergují, dostaneme řadu novou $\sum_{k=1}^{\infty} [+1(-1)] = \sum_{k=1}^{\infty} 0$, která již konverguje. Pokud ale pracujeme čistě s konvergentními řadami, pak už vše funguje tak, jak bychom rádi, součet vzniklé řady se dá odvodit přirozeným způsobem z informace o řadách, se kterými jsme začali.

Věta 9d.2.

Uvažujme konvergentní řady $\sum_{k=0}^{\infty} a_k = A$ a $\sum_{k=0}^{\infty} b_k = B$, nechť $c \in \mathbb{R}$. Pak konvergují i řady $\sum_{k=0}^{\infty} (ca_k) = cA$ a $\sum_{k=0}^{\infty} (a_k + b_k) = A + B$.

9d.3 Mocninné řady

Připomeňme geometrickou řadu $\sum_{k=0}^{\infty} q^k$, pro jejíž součet máme vzorec. Dá se na to nahlížet tak, že vlastně máme řadu s parametrem q a podle toho, co dosadíme, dostáváme buď číslo (řada konverguje), nebo se dozvíme, že takové dosazování k číslu nevede. Když se s q omezíme na interval $(-1, 1)$, dostáváme předpis, který každému q přiřadí jisté číslo—jinými slovy, vznikla nám funkce. Protože bývá zvykem značit proměnnou jako x , můžeme napsat, že máme funkci $x \mapsto \sum_{k=0}^{\infty} x^k$, jejíž definičním oborem je interval $(-1, 1)$, kde dokonce pro ni máme i pohodlnější vyjádření: $x \mapsto \frac{1}{1-x}$.

Není důvod se omezovat jen na sčítání výrazů x^k , může chtít sčítat třeba $\sum_{k=1}^{\infty} \frac{\operatorname{tg}(k)}{x^{2+k}}$ pro různé hodnoty (volby) čísla x a položit si dvě zásadní otázky: Pro která x tak dostaneme konvergentní řadu? A když se na taková x omezíme, dokážeme pro součet řady najít rozumný vzorec?

Řady funkcí jsou vysoce náročná oblast matematické analýzy nabízející velice málo jednoduchých odpovědí. Je to tím, že funkci je strašně hodně a s velice rozličným chováním, takže když se jich navíc pokoušíme sčítat nekonečně mnoho, dá se čekat spousta různých problémů. V takovýchto nepřehledných situacích se matematici tradičně zaměří na určitou menší skupinku objektů s pěkným chováním, které by zároveň byly užitečné. V teorii řad funkcí je takovou klíčovou skupinou skupina řad, které by šlo lidově označit za polynomy nekonečného stupně.

Definice.

Pojem **mocninná řada (power series)** označuje libovolnou řadu ve tvaru $\sum_{k=0}^{\infty} a_k x^k$, kde $a_k, k \in \mathbb{N}_0$ jsou pevně zvolená čísla (**koeficienty řady**) a x je proměnná.

Dvě mocninné řady jsme už viděli, $\sum_{k=0}^{\infty} x^k$ (zde jsou všechny koeficienty $a_k = 1$) a $\sum_{k=0}^{\infty} (k+1)x^k$.

Poznámka: Správně bychom měli říct „mocninná řada se středem v 0“, protože v analýze se hovoří obecně o mocninných řadách se středem c ve tvaru $\sum_{k=0}^{\infty} a_k(x - c)^k$. Pro naše účely by tato větší obecnost nic nepřinesla, je proto rozumné se zaměřit na nejjednodušší verzi.

△

Je-li dána mocninná řada, zásadní otázka zní: Pro které hodnoty x bude po dosazení výsledná řada (tedy už s reálnými čísly) konvergovat? Vždycky je možné najít alespoň jedno x , pro které to vyjde: Dosadíme-li do libovolné mocninné řady $x = 0$, dostáváme $\sum_{k=0}^{\infty} 0 = 0$.

Ta otázka tedy ve skutečnosti zní, zda existují i jiná x , která dávají konvergentní řadu. Analytici dokazují, že množina takovýchto x je velice pěkná, vždy se jedná o interval kolem počátku. Je to tedy interval s krajními body $-\varrho$ a ϱ , tomuto číslu se říká poloměr konvergence řady.

Například již víme, že řada $\sum_{k=0}^{\infty} x^k$ konverguje pro $x \in (-1, 1)$, má tedy poloměr konvergence $\varrho = 1$. Všechna $\varrho \geq 0$ jsou možná, například existuje řada, pro kterou $\varrho = \infty$, tedy řada, která konverguje pro všechna reálná x , naopak jsou řady s $\varrho = 0$, takové konvergují jen na intervalu $\langle 0, 0 \rangle = \{0\}$, tedy kromě povinné nuly už jiné x dosadit rozumně nelze.

Operace s mocninnými řadami.

Mocninné řady jsou také řady, takže je umíme sčítat a násobit číslem. Jak to dopadne?

$$\begin{aligned} c \left(\sum_{k=0}^{\infty} a_k x^k \right) &= \sum_{k=0}^{\infty} c a_k x^k = \sum_{k=0}^{\infty} (c a_k) x^k, \\ \sum_{k=0}^{\infty} a_k x^k + \sum_{k=0}^{\infty} b_k x^k &= \sum_{k=0}^{\infty} (a_k x^k + b_k x^k) = \sum_{k=0}^{\infty} (a_k + b_k) x^k. \end{aligned}$$

Vidíme, že vznikají zase mocninné řady. U nich se nové koeficienty získají násobením či sečtením koeficientů původních řad, přesně jako u operací s polynomem. Jak tomu bude s výsledky u konvergujících řad?

Pokud výchozí řada/řady konvergují s poloměrem konvergence $\varrho > 0$, tak už jejich součtem není jen číslo, ale funkce na určitém intervalu. Odpovídají si funkce na obou stranách rovností výše? Naštěstí ano, jinak by ta teorie moc užitečná nebyla.

Věta 9d.4.

Uvažujme mocninné řady $\sum_{k=0}^{\infty} a_k x^k$ a $\sum_{k=0}^{\infty} b_k x^k$, nechť $c \in \mathbb{R}$. Předpokládejme, že obě konvergují na nějakém intervalu I a $\sum_{k=0}^{\infty} a_k x^k = f(x)$, $\sum_{k=0}^{\infty} b_k x^k = g(x)$ na I . Pak mocninné řady $\sum_{k=0}^{\infty} (c a_k) x^k$ a $\sum_{k=0}^{\infty} (a_k + b_k) x^k$ také konvergují na I a platí tam $\sum_{k=0}^{\infty} (c a_k) x^k = c f(x)$ a $\sum_{k=0}^{\infty} (a_k + b_k) x^k = f(x) + g(x)$.

Příklad 9d.b: Již víme, že $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ pro $x \in (-1, 1)$. Dá se ukázat, že pro x z tohoto intervalu platí také vztah $\sum_{k=1}^{\infty} \frac{1}{k} x^k = -\ln(1-x)$. Nemůžeme ale tyto dvě řady rovnou sečíst, protože nemají stejnou indexaci.

Pokud bychom se rozhodli vyřešit problém posunem indexu u druhé řady, nastane problém jinde:

$$\begin{aligned} \sum_{k=0}^{\infty} x^k + \sum_{k=1}^{\infty} \frac{1}{k} x^k &= \left| \begin{array}{l} i = k-1 \\ k = i+1 \\ k = 1 \mapsto i = 1-1 = 0 \end{array} \right| = \sum_{k=0}^{\infty} x^k + \sum_{i=0}^{\infty} \frac{1}{i+1} x^{i+1} \\ &= \sum_{k=0}^{\infty} x^k + \sum_{k=0}^{\infty} \frac{1}{k+1} x^{k+1} = \sum_{k=0}^{\infty} (x^k + \frac{1}{k+1} x^{k+1}). \end{aligned}$$

V sumě se nám sešly rozdílné mocniny, takže výslednou řadu nelze upravit do tvaru mocninné řady. To není dobré, proto indexy sjednotíme jinou z oblíbených metod, prostě z té první řady vynecháme její první člen (daný indexem $k = 0$), protože v druhé také není.

$$\sum_{k=0}^{\infty} x^k + \sum_{k=1}^{\infty} \frac{1}{k} x^k = \left(x^0 + \sum_{k=1}^{\infty} x^k \right) + \sum_{k=1}^{\infty} \frac{1}{k} x^k = x^0 + \sum_{k=1}^{\infty} x^k + \sum_{k=1}^{\infty} \frac{1}{k} x^k = 1 + \sum_{k=0}^{\infty} (\frac{1}{k} + 1) x^k.$$

Vznikla opravdu mocninná řada a podle věty výše konverguje přinejmenším na intervalu $(-1, 1)$, kde je jejím součtem funkce $\frac{1}{1-x} - \ln(1-x)$.

△

Daná mocninná řada se dá modifikovat i více způsoby než jen vynásobením číslem. Shrňme si populární úpravy.

Věta 9d.5.

Uvažujme mocninnou řadu $\sum_{k=0}^{\infty} a_k x^k$, která konverguje na nějakém intervalu I a platí tam $\sum_{k=0}^{\infty} a_k x^k = f(x)$.

Nechť $c \in \mathbb{R}$ a $N \in \mathbb{N}_0$. Pak na vnitřku intervalu I konvergují i následující řady:

$$\sum_{k=0}^{\infty} c \cdot a_k x^k = c \cdot f(x);$$

$$\sum_{k=0}^{\infty} c^k a_k x^k = f(cx);$$

$$\sum_{k=0}^{\infty} a_k x^{k+N} = x^N f(x);$$

$$\sum_{k=0}^{\infty} k a_k x^{k-1} = f'(x).$$

$$\sum_{k=0}^{\infty} k a_k x^k = x f'(x).$$

Druhý vztah je vlastně obyčejná substituce za proměnou: $\sum_{k=0}^{\infty} c^k a_k x^k = \sum_{k=0}^{\infty} a_k (cx)^k$, u třetího vztahu zase stačí vytknout:

$$\sum_{k=0}^{\infty} a_k x^{k+N} = \sum_{k=0}^{\infty} x^N a_k x^k = x^N \sum_{k=0}^{\infty} a_k x^k = x^N f(x)$$

Zajímavý je čtvrtý vztah. Všimněte si, že vznikne zderivováním původní rovnosti $\sum_{k=0}^{\infty} a_k x^k = f(x)$, přičemž nalevo namísto derivování celé řady derivujeme jen její členy. Zapsáno formálně, $\left[\sum_{k=0}^{\infty} a_k x^k \right]' = \sum_{k=0}^{\infty} [a_k x^k]'$. Podobně lze ukázat, že řady lze takto i integrovat, viz nějaká kniha o analýze.

Celkové poučení z této věty se dá shrnout tak, že pokud máme řady konvergující na otevřeném intervalu, tak s nimi můžeme v zásadě zacházet jako s polynomy (sčítat, vynásobit mocninou, derivovat člen po členu atd). Tyto triky se nám budou hodit v kapitole

Příklad 9d.c: Víme, že $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ na intervalu $(-1, 1)$.

Když tuto rovnost zderivujeme napravo i nalevo, dostáváme (čteno zprava)

$$\begin{aligned} \frac{1}{(1-x)^2} &= \left[\frac{1}{1-x} \right]' = \left[\sum_{k=0}^{\infty} x^k \right]' = \sum_{k=0}^{\infty} [x^k]' = \sum_{k=0}^{\infty} k x^{k-1} \\ &= \sum_{k=1}^{\infty} k x^{k-1} = \left| \begin{array}{c} i = k-1 \\ k = i+1 \\ k = 1 \mapsto i = 0 \end{array} \right| = \sum_{i=0}^{\infty} (i+1) x^i. \end{aligned}$$

Při přechodu na nový řádek jsme použili pozorování, že pro $k = 0$ vychází $0 \cdot x^{-1} = 0$, tudíž lze tento člen a tento index z řady vynechat.

Jaký je závěr? Pomocí známého vzorce pro součet geometrické řady jsme dokázali vzorec (ii) z Věty .

△

Pokud pro nějakou funkci najdeme mocninnou řadu, která ji má jako svůj součet, tak říkáme, že jsme danou

funkci rozvinuli v mocninnou řadu. Nejpopulárnější rozvoje jsou tyto:

$$\begin{aligned}\frac{1}{1-x} &= \sum_{k=0}^{\infty} x^k = 1 + x + x^2 + x^3 + x^4 + \dots, \quad x \in (-1, 1); \\ e^x &= \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots, \quad x \in \mathbb{R}; \\ \sin(x) &= \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots, \quad x \in \mathbb{R}; \\ \cos(x) &= \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots, \quad x \in \mathbb{R}; \\ \ln(1+x) &= \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots, \quad x \in (-1, 1).\end{aligned}$$

Další rozvoje se dají získat z těchto pomocí úprav z věty výše.

Příklad 9d.d: Najdeme rozvoje pro následující funkce:

$$\begin{aligned}\frac{x}{1-x} &= x \cdot \frac{1}{1-x} = x \cdot \sum_{k=0}^{\infty} x^k = \sum_{k=0}^{\infty} x^{k+1} = \sum_{i=1}^{\infty} x^i; \\ \frac{1}{1+x^2} &= \frac{1}{1-(-x^2)} = |y = -x| = \frac{1}{1-y} = \sum_{k=0}^{\infty} y^k \\ &= \sum_{k=0}^{\infty} (-x^2)^k = \sum_{k=0}^{\infty} (-1)^k x^{2k} = 1 - x^2 + x^4 - x^6 + \dots; \\ \arctg(x) &= \int \frac{1}{1+x^2} dx = \int \sum_{k=0}^{\infty} (-x^2)^k dx = \sum_{k=0}^{\infty} \int (-1)^k x^{2k} dx = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} x^{2k+1} + C.\end{aligned}$$

Nevíme, která hodnota C je správná, tak zkusíme dosadit nějaké konkrétní x do levé i pravé strany a uvidíme. Třeba volba $x = 0$ dává $\arctg(0) = \sum 0 + C$ neboli $0 = C$. Máme tedy

$$\arctg(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} x^{2k+1} = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots.$$

Pokud dosadíme $x = 1$, dostáváme $\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots$ neboli $\pi = 4 - \frac{4}{3} + \frac{4}{5} - \frac{4}{7} + \frac{4}{9} - \dots$, což je východisko pro některé populární metody výpočtu čísla π s libovolnou přesností.

△

10. Rekurentní vztahy

Kapitolu uvedeme populárním příkladem.

Příklad 10.a: Tento problém je znám po názvem Hanojské věže. Představte si tři tyčky, na jedné je navlečeno n disků (s dírkou uprostřed) pěkně podle velikosti od největšího dole po nejmenší nahoru. (Chtěl jsem udělat obrázek, ale místo toho vás pošlu do nejbližšího hračkářství, kde v oddělení pro mrňata určitě tyčku s kolečky mají.) Cílem je dostat tyto disky do stejné pozice, ale na jiné tyčce, přičemž jediný povolený tah je přesunout právě jeden disk buď na prázdnou tyč nebo na disk na jiné tyči, který je ale větší; jinými slovy, nelze nosit více disků najednou a nelze položit větší na menší (a nelze je odložit někde úplně mimo). Je možné tuto úlohu vyřešit?

Ať už vymyslíme jakýkoliv způsob, nakonec musí přijít okamžik, kdy přesouváme dolní, největší disk na cílovou tyč. Abychom to mohli udělat, je nutné všechny disky nad ním dát někam jinam, ale žádný z nich nesmí přijít na cílovou tyč, to bychom totiž ten největší nemohli dát na něj, konec konců, my ten největší stejně chceme dát až dolů. Vidíme tedy, že nutnou přípravou pro přesunutí největšího disku je, aby všechny ostatní byly na té třetí tyči, a to samozřejmě podle velikosti (jinak to nejde). Shrnuji, chceme-li přenést celou hromadu řekněme na tyč 2, musíme tam dát dolů největší disk, což vyžaduje přenesení pyramidy disků nad ním na tyč číslo 3.

Dostáváme tím jasnou rekurzi. Začneme s pyramidou n disků na tyči 1, a abychom ji přenesli na tyč 2, musíme nejprve přenést horních $n - 1$ disků na tyč 3. Tuto menší pyramidku o $n - 1$ discích přeneseme tak, že její největší disk chceme přenést na cílovou tyč 3, ale na to potřebujeme to, co je nad ním, tedy pyramidku velikosti $n - 2$, přenést na tyč 1 či 2 atd. Dříve či později dojdeme k tomu, že máme někam přenést jeden disk, a to ten nejmenší, což lze bez problémů.

Proveditelnost by tedy mělo jít dokázat indukcí. Jediná trochu nejasná věc je, že uprostřed řešení budeme v situaci, kdy máme přenést řekněme pyramidku s 13 disky, ale dalších 5 disků z předchozího rekurzivního rozkladu už se někde potuluje. Nedojde při pokusu o skutečnou realizaci našeho algoritmu ke konfliktu s pravidly? Naštěstí ne. Všechny ty disky z předchozího rozkladu jsou totiž větší než ty v naší pyramidce, tudíž je můžeme v dané chvíli považovat za podlahu, v přesouvání té pyramidky nás neomezí. Raději to zapracujeme do našeho důkazu.

Zkusíme tedy dokázat indukcí, že dokážeme přenést pyramidku n disků z libovolné tyče a na libovolnou jinou tyč b , s tím, že na tyčích b a c již mohou být nějaké větší disky.

(0) $n = 1$: Jeden disk určitě přeneseme na cílovou tyč, přičemž nám nebude vadit, když už tam bude nějaký větší disk.

(1) Předpokládáme, že pyramidku o velikosti n umíme. Mějme pyramidku o $n + 1$ discích na tyči a , potřebujeme ji dostat na tyč b , přičemž na tyčích b a c už jsou třeba nějaké disky větší než ty v naší pyramidce. Nejprve použijeme indukční předpoklad a přesuneme horních n disků na tyč c (v tom nám případný větší disk dole nebude vadit), pak přesuneme spodní disk naší pyramidky na tyč b (ani v tom nám případný větší disk nebude vadit), načež opět využijeme indukční předpoklad a přesuneme horních n disků naší pyramidy z tyče c na tyč b , kde už leží disk s číslem $n + 1$, který je větší než disky pyramidky nad ním, i to je v pořádku.

Tím je důkaz hotov.

To bylo snadné. Mnohem zajímavější je otázka, kolik přesunů disků („tahů“) na to budeme potřebovat. Když označíme H_n počet tahů, které náš algoritmus spotřebuje na přesun pyramidky o n discích, tak nám zkušenosti z kapitoly o indukci naznačují, že pro H_n dostaneme rekurzivní vztah. Je jasné, že $H_1 = 1$. Postup v kroku (1) pak říká, že $H_{n+1} = H_n + 1 + H_n = 2H_n + 1$.

Podobné funkce jsme zkoumali v kapitole . Nejprve jsme si vždy spočítali několik prvních hodnot a pak z toho uhádli vzorec. Zde máme $H_1 = 1$, $H_2 = 3$, $H_3 = 7$, $H_4 = 15$, $H_5 = 31$. Vidíte nějaký vzorec? Pokud ne, zkusíme si ještě jiný přístup, který na to jde z opačné strany, použijeme k nalezení H_n naši rekurzi, trochu optimismu a Větu .

$$\begin{aligned} H_n &= 2H_{n-1} + 1 = 2(2H_{n-2} + 1) + 1 = 2^2H_{n-2} + 2 + 1 \\ &= 2^2(2H_{n-3} + 1) + 2 + 1 = 2^3H_{n-3} + 2^2 + 2 + 1 \\ &= 2^3(2H_{n-4} + 1) + 2^2 + 2 + 1 = 2^4H_{n-4} + 2^3 + 2^2 + 2 + 1 = \dots \\ &= 2^{n-1}H_1 + 2^{n-2} + \dots + 2^2 + 2 + 1 = \sum_{i=0}^{n-1} 2^i = \frac{1 - 2^n}{1 - 2} = 2^n - 1. \end{aligned}$$

Ta část se třemi tečkami je samozřejmě podezřelá, to byl ten optimismus. Proto jsme zatím nedokázali, že máme správnou odpověď, ale máme už rozumného kandidáta, pro kterého správnost dokážeme snadno indukcí:

(0) $H_1 = 2^1 - 1 = 1$, to souhlasí.

(1) Předpokládejme, že pro nějaké $n \geq 1$ máme $H_n = 2^n - 1$. Pak

$$H_{n+1} = 2H_n + 1 = 2 \cdot (2^n - 1) + 1 = 2^{n+1} - 1.$$

Souhlasí, náš vzorec je správný.

Řešení hádanky zvané Hanojské věže pro n disků tedy zabere $2^n - 1$ přesunů, pokud použijeme náš algoritmus. Zajímavé je, že lze dokázat (to už je těžší), že hádanku nelze vyřešit za méně tahů, náš algoritmus je tedy optimální řešení.

K této úloze se váže legenda o jakémse klášteře ve Vietnamu, kde mniší zkouší už několik staletí vyřešit tuto úlohu s 64 disky (samozřejmě zlatými a velkými), a až to udělají, tak svět skončí nebo něco. Pokud následují optimální strategii a jeden disk přenesou za sekundu (což je hodně optimistické, jak asi sami znáte, když někdy doma přenášíte velké kusy zlata), tak jim to zabere $2^{64} - 1 \sim 18 \cdot 10^{18}$ vteřin, což vychází na nějakých 600 miliard let. Zatím tedy asi nemá smysl rozfofrovat penzijní fond.

S tím souvisí ještě jedna zajímavost. Jak by vlastně mniši ten ideální rekurzivní algoritmus dělali? Museli by si zapisovat do svitků, ve které fázi kterého podprogramu zrovna jsou a kam se mají vracet, to nevypadá moc prakticky. Naštěstí existuje jednoduchý návod.

Algoritmus: V prvním kroku vezměte nejmenší disk a přemístěte jej směrem doprava na nejbližší tyč, kam vám to pravidla dovolí, přičemž se to bere cyklicky (z poslední tyče se vracíte na první).

V druhém kroku vezměte takový jiný než nejmenší disk, kterým je možné v rámci pravidel táhnout, a také jej přesuňte doprava (cyklicky) na nejbližší možnou tyč.

Tyto dva kroky teď opakujte, v druhém kroku bude vždy jen jediný disk jiný než nejmenší, který je možné přesunout jinam.

Pokud je n liché, celá pyramidka se přesune o jedno doprava, pokud je n sudé, přesune se o jedno doleva (cyklicky). Chcete-li to naopak, šoupejte disky na opačnou stranu.

Zkusíme si to se čtyřmi disky:

Je vidět, že v každém kroku, kdy nenosíme disk 1, máme opravdu jen jednu jinou možnost tahu.

Mimochodem, zkoumají se i varianty tohoto problému. Jedna z nich (Reve's puzzle) je, že ty tyčky jsou čtyři, a zajímavé na ní je, že se dodnes neví, kolik je minimum tahů k přemístění n disků. Kandidátem je algoritmus z roku 1939, ale zatím (2011) se nepovedlo dokázat, že je to optimální strategie.

Jako doplňkové čtení silně doporučují povídka A.C. Clarka *Devět miliard božích jmen*.

△

Induktivně definované funkce či posloupnosti dostáváme, když se při popisu situace setkáváme se vztahem, který nějak kombinuje přítomnost s minulostí. Takový vztah má své jméno.

! Definice

Rekurentní vztah či **rekurzivní vztah** (recurrence relation) pro posloupnost $\{a_k\}$ je libovolná rovnice typu $F(a_n, a_{n-1}, a_{n-2}, \dots, a_0) = 0$, kde F je nějaká funkce.

! Třeba podstata problému Hanojských věží se dá vyjádřit vztahem $H_n - 2H_{n-1} - 1 = 0$. Je to vztah výjimečně pěkný, rozhodně lepší než třeba vztah $a_n = \sin(a_{n-1}a_{n+1})\sqrt{\frac{a_{n-2}^4 + 1}{a_{n-1}^4 + 1}}$, se kterým bychom upřímně řečeno nehnuli.

Hlavním problémem rekurentních vztahů je, že nenabízí rozumný způsob, jak počítat neznámé hodnoty. Například v rekurentním vztahu $a_n^8 - a_n a_{n-1} + a_{n-2} = 0$ nám příliš nepomůže, když známe třeba $a_1 = a_2 = 1$, protože rovnici $a_3^8 - a_3 + 1 = 0$ vyřešit neumíme.

Není proto divu, že se většina pojednání o rekurentních vztazích hned na začátku omezí jen na ty vztahy, které lze upravit do tvaru $a_n = G(a_{n-1}, a_{n-2}, \dots, a_0)$. Často se takovýto vztah píše s posunutým indexem, $a_{n+1} = G(a_n, a_{n-1}, \dots, a_0)$, to „ $n+1$ “ psychologicky naznačuje, že něco máme (a_0 až a_n) a chceme další člen. Takto jsme to dělávali v kapitole . Vztah tohoto typu je mnohem perspektivnější, protože jakmile známe několik počátečních hodnot, dokážeme iterací počítat nové a nové hodnoty. Tím ale narázíme na další problém. Výpočet konkrétních hodnot tímto způsobem je velice drahý, na a_{10000} potřebujeme spočítat všechna předchozí a_n .

Hlavním tématem proto bývá hledání způsobu, jak pro rekurentně zadánou posloupnost (či funkci) najít explicitní vyjádření pro a_n vzorcem v uzavřeném tvaru (elementární funkce spojené algebraickými operacemi či skládáním), viz ten příklad s věžemi. Někdy je to z principu nemožné, někdy by to snad možné i bylo, ale neumíme to najít. Abychom dostali rozumné odpovědi, budeme se muset omezit na rekurentní vztahy velice jednoduchého (ale pořád silně užitečného) typu. Pro ilustraci ukážeme ještě několik příkladů.

Příklad 10.b: Připomeňme si příklad , kde jsme rekurzí dokázali tapetovatelnost šachovnice triminy. Otevřená otázka zůstala, kolik trimin je na šachovnici o straně 2^n potřeba, označili jsme to t_n . Algoritmus vedl na rovnice $t_1 = 1$ a $t_{n+1} = 4t_n + 1$. Zkusíme postup z hanojského příkladu .

$$\begin{aligned} t_n &= 4t_{n-1} + 1 = 4(4t_{n-2} + 1) + 1 = 4^2t_{n-2} + 4 + 1 \\ &= 4^2(4t_{n-3} + 1) + 4 + 1 = 4^3t_{n-3} + 4^2 + 4 + 1 \\ &= 4^3(4t_{n-4} + 1) + 4^2 + 4 + 1 = 4^4t_{n-4} + 4^3 + 4^2 + 4 + 1 = \dots \\ &= 4^{n-1}t_1 + 4^{n-2} + \dots + 4^2 + 4 + 1 = \sum_{i=0}^{n-1} 4^i = \frac{1 - 4^n}{1 - 4} = \frac{1}{3}(4^n - 1). \end{aligned}$$

Ověření správnosti tohoto vzorce indukcí necháme na čtenáři.

△

Vidíme, že jsme schopni tímto postupem relativně rychle nalézat vzorce pro funkce či posloupnosti zadané vztahem $a_{n+1} = a \cdot a_n + b$ a počáteční hodnotou $a_{n_0} = A$. Bohužel pro komplikovanější vztahy už to není příliš perspektivní.

Příklad 10.c: Mějme induktivní definici funkce $f(1) = 3$, $f(2) = -1$ a $f(n) = f(n-1) + 6f(n-2)$ pro $n \geq 3$. Dostáváme $f(3) = 17$, $f(4) = 11$, $f(5) = 113$, $f(6) = 209$. Vidíte v tom nějaký vzorec? Já ne.

Zkusme přístup z předchozích příkladů:

$$\begin{aligned} f(n) &= f(n-1) + 6f(n-2) = [f(n-2) + 6f(n-3)] + 6f(n-2) = 7f(n-2) + 6f(n-3) \\ &= 7[f(n-3) + 6f(n-4)] + 6f(n-3) = 13f(n-3) + 42f(n-4) \\ &= 13[f(n-4) + 6f(n-5)] + 42f(n-4) = 55f(n-4) + 78f(n-5) = ? \end{aligned}$$

Vidíte z toho něco? Asi bude lepší počkat na kapitolu , kde se podobné příklady naučíme řešit na dvou řádcích.
△

10a. Lineární rekurentní rovnice

Hodně rekurentních vztahů se dá přepsat do tvaru, kdy G závisí jen na stále stejném počtu k předchozích členů a navíc lineárním způsobem: $a_n = d_1(n)a_{n-1} + d_2(n)a_{n-2} + \dots + d_k(n)a_{n-k}$.

Z praktického důvodu bude lepší takovéto rovnice psát trochu jinak.

!

Definice.

Lineární rekurentní rovnice, popřípadě **lineární rekurzivní rovnice řádu** $k \in \mathbb{N}_0$ je libovolná rovnice ve tvaru

$$a_{n+k} + c_{k-1}(n)a_{n+k-1} + \dots + c_2(n)a_{n+2} + c_1(n)a_{n+1} + c_0(n)a_n = b_n, \quad n \geq n_0,$$

kde $n_0 \in \mathbb{Z}$, $c_i(n)$ pro $i = \{0, \dots, k-1\}$ (tzv. **koeficienty** rovnice) jsou nějaké funkce $\mathbb{Z} \mapsto \mathbb{R}$, přičemž $c_0(n)$ není identicky nulová funkce, a $\{b_n\}_{n=n_0}^\infty$ (tzv. **pravá strana rovnice**) je pevně zvolená posloupnost reálných čísel.

Jestliže $b_n = 0$ pro všechna $n \geq n_0$, pak se příslušná rovnice nazývá **homogenní**.

By a **linear recurrence equation** of order $k \in \mathbb{N}_0$ we mean any equation of the form

$$a_{n+k} + c_{k-1}(n)a_{n+k-1} + \dots + c_2(n)a_{n+2} + c_1(n)a_{n+1} + c_0(n)a_n = b_n, \quad n \geq n_0,$$

where $n_0 \in \mathbb{Z}$, $c_i(n)$ for $i = \{0, \dots, k-1\}$ (**coefficients** of the equation) are some functions $\mathbb{Z} \mapsto \mathbb{R}$ with $c_0(n)$ not identically zero, and $\{b_n\}_{n=n_0}^\infty$ (the **right hand-side** of the equation) is a fixed sequence of real numbers.

If $b_n = 0$ for all $n \geq n_0$, then the equation is called **homogeneous**.

Příklady ze začátku této kapitoly sem samozřejmě patří, stačí ve vztazích vhodně posunout indexy. Hanojský vztah se dá přepsat jako $H_{n+1} - 2H_n = 1$, $n \geq 1$, jde o lineární rekurentní rovnici 1. řádu, podobně se dá triminová rovnice přepsat jako $t_{n+1} - 4t_n = 1$, $n \geq 1$. Ve příkladě jsme měli rovnici $f(n+2) - f(n+1) - 6f(n) = 0$, $n \geq 1$, je to tedy lineární rekurentní rovnice 2. řádu. Naopak $a_n - a_{n-1}a_{n-2} = 0$ lineární není.

U rekurentních rovnic je drobný problém se zápisem. Viděli jsme již tři formální způsoby, jak rovnice napsat. V příkladě jsme použili zápis $f_n = f_{n-1} + 6f_{n-2}$ typu $f_n = G(f_{n-1}, f_{n-2})$, zmínili jsme také intuitivnější formu $f_{n+1} = f_n + 6f_{n-1}$ neboli $f_{n+1} = G(f_n, f_{n-1})$. Definice lineárních rekurentních rovnic teď po nás chce zapsat tento vztah jako $f_{n+2} - f_{n+1} - 6f_n = 0$, kdy jsme to „referenční n “ dali naopak ke členu s nejmenším indexem ve vztahu.

Každý z těchto způsobů má své výhody i nevýhody a autoři učebnic si tedy vybírají podle svého gusta, rozhodně to není jednotné. Co z toho pro nás plyne? Když otevřeme nějakou knihu o takovýchto rovnicích, tak se musíme dobře podívat, v jakém tvaru je autor chce mít, abychom správně chápali jeho tvrzení. Nepříjemné je to u teorie, kdy je třeba při přechodu z jedné knihy do druhé překládat vzorce do jiného značení.

Tento přechod se dělá snadno „posunem indexu“, což je téma pro poznámku níže. Posouvání je docela běžné, mimo jiné proto, že při popisu reálného problému se nejlépe vytváří rovnice v přirozeném tvaru $a_{n+1} = \dots$, který je ale asi nejméně vhodný pro další zpracování pomocí teorie rekurentních rovnic.

Stojí za zmínku, že vlastně nejde o rovnici jednu, ale o nekonečně mnoha rovnic, například ten hanojský vztah vlastně známená rovnice $H_2 - 2H_1 = 1$, $H_3 - 2H_2 = 1$, $H_4 - 2H_3 = 1$, $H_5 - 2H_4 = 1$ a tak dále. Je to triviální, ale až budeme mluvit o „řešení rovnice“, je dobré si toho být vědom, pod slovem „rovnice“ se jich skrývá mnoho. Proto je také při zadávání rovnice podstatná ta poznámka o indexu za čárkou, třeba rovnice $a_{n+1} - a_n = 1$, $n \geq 1$ je formálně jiná než rovnice $a_{n+1} - a_n = 1$, $n \geq 3$, protože výsledné množiny rovnic se nerovnají. Dá se očekávat, že pak se budou lišit i řešení.

Je ovšem nutno přiznat, že zatímco znalost indexu je podstatná při zápisu řešení, z pohledu praktického v tom až tak velký rozdíl není. Pokud totiž najdeme nějaké řešení rovnice $a_{n+1} - a_n = 1$, $n \geq 1$, tak vynecháním prvních dvou členů získáme automaticky řešení rovnice $a_{n+1} - a_n = 1$, $n \geq 3$; naopak řešení rovnice $a_{n+1} - a_n = 1$, $n \geq 3$ jistě půjde „prodloužením začátku“ upravit na řešení té první rovnice. Dále uvidíme, že v postupu řešení se znalost vymezení $n \geq n_0$ příliš neobjevuje.

Poznámka: Máme-li posloupnost $\{a_n\}$, lze vytvořit z rozdílu následujících členů posloupnost novou, definovanou jako $\Delta a_n = a_n - a_{n-1}$. Tomuto se říká *diference posloupnosti* $\{a_n\}$ a existují úlohy, které s tímto pojmem pracují a vedou na rovnice typu $a_{n+1} = \Delta a_n + 1$ a podobně. Těmto rovnicím se přirozeně říká diferenční rovnice. Některí autoři pak tento název přenášejí na všechny rekurentní rovnic a mluví o lineárních diferenčních rovnicích.

△

! 10a.1 Poznámka o posunu indexu: Jak indexy posouváme? Na první pohled snadno, prostě všechny výskyty indexační proměnné zvýšíme či snížíme o stejně číslo. U dané rovnice $f_n = f_{n-1} + 6f_{n-2}$ přičtením dvojky ke všem n dostáváme $f_{n+2} = f_{n+1} + 6f_n$ neboli $f_{n+2} - f_{n+1} - 6f_n = 0$, přesně jak potřebujeme pro tuto knihu. Je přitom ale třeba mít na paměti dvě věci. Ukážeme si je na rovnici $a_{n+1} = a_n - n^2 a_{n-3} + 2n$, $n \geq 7$, kde potřebujeme zvýšit index o tři.

Za prvé, když se mění indexační proměnná, mění se opravdu všude, nejen v místech indexu. Pokud bychom naší vzorovou rovnici přepsali jako $a_{n+4} = a_{n+3} - n^2 a_n + 2n$, bylo by to špatně, viz n^2 a $2n$. Správný převod je $a_{n+4} = a_{n+3} - (n+3)^2 a_n + 2(n+3)$. Opomenutí takového posunu je častá chyba u zkoušek.

Za druhé, indexační proměnná se mění i ve specifikaci rozsahu. To souvisí s tím, že vlastně mluvíme o soustavě rovnic. Původní zadání $a_{n+1} = a_n - n^2 a_{n-3} + 2n$ pro $n \geq 7$ obsahovalo tyto rovnice:

$$\begin{aligned} a_8 &= a_7 - 7^2 a_4 + 14 \\ a_9 &= a_8 - 8^2 a_5 + 16 \\ a_{10} &= a_9 - 9^2 a_6 + 18 \\ a_{11} &= a_{10} - 10^2 a_7 + 20 \quad \text{atd.} \end{aligned}$$

My musíme dosáhnout toho, aby rovnice po posunu indexu dávala stejnou množinu. Na to doporučíme dva způsoby.

Jedna možnost je spolehnout se na selský rozum. Ze zadání „ $a_{n+1} = a_n - n^2 a_{n-3} + 2n$, $n \geq 7$ “ vidíme, že nejmenší možný index, který se v rovnicích může vyskytnout, je $7 - 3 = 4$. Musíme zajistit, aby tomu tak bylo i v naší přeindexované rovni, což se evidentně udělá volbou $n \geq 4$. Ověříme, po dosazení $n = 4$ do přepsané rovnice pak první rovnost vychází $a_8 = a_7 - 7^2 a_4 + 14$, což je správně, viz ten sloupeček výše. Řešíme tedy rovnici $a_{n+4} - a_{n+3} + (n+3)^2 a_n = 2n + 6$, $n \geq 4$, je to lineární rekurentní rovnice 4. rádu.

Další možnost je při přečíslování rovnic použít formální substituci, viz příklad .

△

Když už jsme u zápisu, co jsou ty tři tečky v definici rekurentní rovnice? V matematice se rozumí, že takto definované výrazy v sobě ve skutečnosti schovávají indukci, v tomto případě se dá vyjádřit pomocí sumačního

znaménka:

$$a_{n+k} + \sum_{i=0}^{k-1} c_i(n) a_{n+i} = b_n.$$

Trochu nevýhoda takového značení je, že si jej člověk musí zase v hlavě překládat do dlouhých součtů, zejména chce-li to použít v konkrétní situaci, takže z důvodu názornosti se budeme v kritických chvílích snažit používat spíš to delší značení. Na druhou stranu je v tomto sumačním značení pěkně vidět základní referenční index n a řád k , navíc je to podstatně kratší na psaní, takže v teoretických výpočtech (zejména těch brutálnějších) se sumační zápis vyloženě vyplatí.

Dodejme ještě, že ta podmínka $c_0 \neq 0$ je podstatná pro určení řádu rovnice, jinak by tam nula nevadila. Například rekurentní rovnici 1. řádu $a_{n+1} - 13a_n = 0$ lze také psát jako $a_{n+3} - 13a_{n+2} = 0$, což vlastně odpovídá obecnému tvaru $a_{n+3} + c_2a_{n+2} + c_1a_{n+1} + c_0a_n = 0$, ale zde $c_0 = 0$, a tak to rovnice 3. řádu není. V principu by šlo používat i tu druhou variantu (pokud tomu člověk dobře rozumí), ale my se zde budeme držet „správného“ zápisu.

Abychom tu podmínku, že $c_0(n)$ není nulová funkce, nemuseli pořád psát, tak namísto toho řekneme, že daná rovnice je řádu k , a bude to totéž. Jsou ale situace, kde řád tak podstatný není, pak předpoklad o řádu rovnice psát nebudem.

V této kapitole teoreticky prozkoumáme, jak se řešení lineárních rekurentních rovnic chovají, konkrétní metody pro jejich hledání pak odvodíme v další kapitole. Nejprve abychom si měli říct, co vlastně při řešení rovnic hledáme.

!

Definice.

Nechť je dána lineární rekurentní rovnice

$$a_{n+k} + c_{k-1}(n)a_{n+k-1} + \cdots + c_1(n)a_{n+1} + c_0(n)a_n = b_n, \quad n \geq n_0.$$

Jako její **řešení** označíme libovolnou posloupnost $\{a_n\}_{n=n_0}^{\infty}$ takovou, že po dosazení odpovídajících členů do dané rovnice dostáváme pro všechna $n \geq n_0$ pravdivý výrok.

Ukážeme jednoduchý příklad.

Příklad 10a.a: Posloupnost $a_n = 13(n-1)!$, $n \geq 1$ je řešením rovnice

$$a_{n+1} - na_n = 0, \quad n \geq 1,$$

protože když pro libovolné $n \in \mathbb{N}$ dosadíme dotyčný vzorec do rovnice za a_n a a_{n+1} , tak dostaneme pravdivý výrok $13n! - n \cdot 13(n-1)! = 0$.

△

Mimochodem, tato rovnice už je dost těžší a zde se takové řešit nenaučíme.

!**Příklad 10a.b:** Dlouhodobé pozorování volně žijících králíků v Austrálii ukázalo, že se jejich počet každých šest měsíců zdvojnásobí. Vše začalo v roce 1859, kdy jich bylo vypuštěno 24. Kolik jich bylo v roce 1869?

Označme si jako a_n počet králíků v roce $1859 + n$, takže nás zajímá a_n pro $n \geq 0$ a víme, že $a_0 = 24$. Ze vstupních dat vyplývá lineární rekurentní rovnice, kterou se růst králíků řídí: Je-li jich jeden rok a_n , pak za šest měsíců je jich $2a_n$ a za dalších šest měsíců (tedy za rok od a_n) je jich $2 \cdot 2a_n = 4a_n$. Dostáváme $a_{n+1} = 4a_n$, tedy $a_{n+1} - 4a_n = 0$ pro $n \geq 0$.

Hledáme řešení této rovnice 1. řádu, zároveň chceme, aby splňovalo $a_0 = 24$. Spočítáme začátek této posloupnosti:

$$\begin{aligned} a_0 &= 24, \\ a_1 &= 4a_0 = 4 \cdot 24, \\ a_2 &= 4a_1 = 4 \cdot (4 \cdot 24) = 4^2 \cdot 24, \\ a_3 &= 4a_2 = 4 \cdot (4^2 \cdot 24) = 4^3 \cdot 24. \end{aligned}$$

Odhadneme, že posloupnost $a_n = 24 \cdot 4^n$ je hledanou posloupností.

Zkouška: $a_0 = 24 \cdot 4^0 = 24$, počáteční hodnota je v pořádku.

Pro libovolné $n \geq 0$ pak $a_{n+1} - 4a_n = 24 \cdot 4^{n+1} - 4 \cdot 24 \cdot 4^n = 0$, přesně jak bylo požadováno.

Pak už hravě zjistíme, že v roce 1869 tam teoreticky bylo $a_{10} = 24 \cdot 4^{10}$ neboli cca 25 milionů králíků. To už je skoro dost, není někde chyba? Historie říká, že od roku 1869 zabíjeli v Austrálii asi 2 milony králíků ročně, aniž by to nějak znatelněji ovlivnilo jejich populaci, takže těch 24 milionů najednou nevypadá nereálně. Pokusy o klasickou decimaci (odstrel, jedy) růst mírně zpomalily, ale i tak byl v roce 1950 jejich stav odhadován na 600 milionů. Moderní metody s tím drobet pohly, dnes je stav odhadován na 300 milionů, což je ale pořád slušné číslo na 24 prapředků.

Pokročilá poznámka: Čtenář asi cítí, že tento model není zrovna nejspolehlivější. Zádrhel je ve výchozím předpokladu, populace se za šest měsíců bude těžko přesně zdvojnásobovat, spíš půjde o průměrnou hodnotu. Jde tedy o statistický údaj a statistické údaje se nedají používat na malé populace, fungují jen na velkých množstvích. Kdybychom zkoumali populace o milionech králíků (což později děláme), tak by ještě výsledky mohly být docela spolehlivé, ale my jsme začali s 24 králíky, což je hodně málo. U tak malé populace se klidně může stát, že za šest měsíců jich nebude $2 \cdot 24 = 48$, ale třeba 34 či 50, čímž se celý další vývoj vykolejí dost jinam. Při modelování reálných situací je vždy třeba být opatrný, zda výsledek, který nám matematika dala, lze aplikovat na popisovanou situaci. Klíčová je zde věrnost modelu, matematická odpověď na matematickou otázku je samozřejmě spolehlivá.

Ke králíkům se vrátíme v příkladě, kde se podíváme za rok 1869, a také v příkladě, kde ukážeme jiný přístup k jejich problematice.

△

! Vraťme se k příkladu o králících. Když si místo počátečního počtu 24 dáme počet jiný, třeba obecně číslo c , pak stejným postupem dostaneme řešení $a_n = c \cdot 4^n$. Jinak řečeno, pokud uvažujeme jen danou rovnici $a_{n+1} = 4a_n$, pak ji řeší všechny posloupnosti typu $a_n = c \cdot 4^n$, kde c je nějaká reálná konstanta (tohle si zase ověřte dosazením). Máme tedy nekonečně mnoho řešení daných jedním vzorečkem, ve kterém si můžeme úpravou parametru vybírat jedno konkrétní řešení, které vyhovuje doplnkovým požadavkům.

Z hlediska filosofického to dává smysl. Rovnice samotná je něco jako přírodní zákon popisující množení králíků, tudíž se dá čekat, že existuje mnoho situací, které tomuto zákonu vyhovují, stejně jako gravitační zákon umožňuje mnoho různých způsobů padání objektů (kladivo padající ze střechy mrakodrapu urazí trochu jinou dráhu než chleba padající ze stolu). Takovéto situace se liší svobodou, v našem případě králíků lze mluvit o situaci s jedním stupněm volnosti, což vidíme podle jednoho parametru. Logicky pak do toho zapadá, že když si přidáme jeden další požadavek, tak už tuto volnost ztratíme a dostaneme jedno konkrétní řešení.

Brzy ukážeme, že když je rovnice řádu k , tak má k stupňů volnosti, neboli lze najít řešení, ve kterém je k parametrů, které si můžeme zvolit dle libosti. Takovému řešení se říká **obecné řešení**.

Pokud začneme klást požadavky, tak s každým požadavkem jeden stupeň volnosti ubyde (pokud je klademe rozumně, aby si třeba neodporovaly nebo některé z nich neříkaly totéž). Pokud si rozumně zadáme k podmínek, tak už zbyde jen jedno řešení. Každému takovému konkrétnímu řešení říkáme **partikulární řešení**. Pokud bychom si například u těch králíků zadali, že $a_5 = 40000$, tak se najde přesně jedno partikulární řešení, které tomu bude vyhovovat, a my z něj vyčteme, s kolika králíky bychom měli začít, aby z nich za 5 let bylo 40000.

Máme-li rovnici druhého řádu, pak bude mít obecné řešení dva parametry, a pokud si zadáme třeba $a_0 = 1$ a $a_{10} = 13$, tak už tím určíme řešení jednoznačně. Podmínky si tedy můžeme zadávat všelijak, ale nejčastěji to děláme tak, jak jsme zvyklí z kapitoly: Řekneme, jak má hledaná posloupnost začít.

Například rovnici $a_{n+2} - 2a_{n+1} + a_n = n + 1$ řádu 2 si můžeme přepsat (po posunu indexu $n + 1 \mapsto n$) do intuitivního tvaru $a_{n+1} = 2a_n - a_{n-1} + n$. Když si zadáme, kolik je a_0 a a_1 , tak už se ostatní hodnoty dají dopočítat neboli rovnice má jediné řešení.

Této myšlence dáme oficiální název.

! Definice.

Nechť je dána lineární rekurentní rovnice řádu k

$$a_{n+k} + c_{k-1}(n)a_{n+k-1} + \cdots + c_1(n)a_{n+1} + c_0(n)a_n = b_n, \quad n \geq n_0.$$

Za **počáteční podmínky (initial conditions)** pro tuto rovnici považujeme libovolnou soustavu rovnic $a_{n_0} = A_0, a_{n_0+1} = A_1, \dots, a_{n_0+k-1} = A_{k-1}$, kde $A_i \in \mathbb{R}$ jsou pevně zvolená čísla.

Takže pro rovnici řádu k definujeme k podmínek, zkušenost pak říká, že řešení nezbyde žádná svoboda. Potvrdí nám to oficiálně následující věta.

! Věta 10a.2. (o existenci a jednoznačnosti)

Každá lineární rekurentní rovnice má nějaké řešení.

Je-li dána lineární rekurentní rovnice řádu k a příslušné počáteční podmínky, pak existuje jediné řešení této rovnice, které splňuje dané počáteční podmínky.

Věty o existenci a jednoznačnosti bývají v matematických teoriích o rovnicích klíčové. Všimněte si, že druhá část věty vlastně říká dvě věci, jednak že řešení existuje a pak že je jediné. Není to přitom opakování z první části. První část věty totiž jenom říká, že nějaké řešení existuje, ale nezaručí, že nutně musí splňovat počáteční

podmínky, které zrovna potřebujeme. Až druhá část říká, že když si libovolně vybereme počáteční podmínky, tak už k nim řešení existuje, a pak také dodá, že je jediné možné.

Důkaz (poučný): 1) Dokážeme druhou část věty. Uvažujme rovnici $a_{n+k} + \sum_{i=0}^{k-1} c_i(n)a_{n+i} = b_n$ a počáteční podmínky $a_{n_0} = A_0, a_{n_0+1} = A_1, \dots, a_{n_0+k-1} = A_{k-1}$. Nejprve ukážeme, že existuje řešení této úlohy. Sestrojíme jej pomocí strukturální indukce pro $n \in \{n_0, n_0 + 1, n_0 + 2, \dots\}$.

(0) Základní krok: Pro $n = n_0, n_0 + 1, \dots, n_0 + k - 1$ definujeme $a_n = A_{n-n_0}$.

(1) Nechť $n \in \mathbb{N}$, $n \geq n_0 + k - 1$. Předpokládejme, že jsou definovány $a_{n_0}, a_{n_0+1}, \dots, a_n$. Pak definujeme $a_{n+1} = b_{n+1-k} - \sum_{i=0}^{k-1} c_i(n+1-k)a_{n+1-k+i}$. Na pravé straně používáme jen $a_{n+1-k}, a_{n+2-k}, \dots, a_n$, které dle předpokladu máme, neboť díky $n \geq n_0 + k - 1$ všechny používané indexy splňují $n + 1 - k + i \geq n_0$. Definice je proto korektní.

Tím máme vytvořenu posloupnost $\{a_n\}_{n=n_0}^\infty$. Potřebujeme ukázat, že je to řešení.

Platnost počátečních podmínek je zjevná dle definice v kroku (0). Nyní ověříme platnost rovnice. Zvolme tedy libovolné $N \geq n_0$. Pak $N + k > n_0 + k - 1$, proto byl při definici a_{N+k} použit krok (1) pro $n = N + k - 1$. Měli jsme tedy definici $a_{N+k} = b_N - \sum_{i=0}^{k-1} c_i(N)a_{N+i}$, což po přepsání dává $a_{N+k} + \sum_{i=0}^{k-1} c_i(N)a_{N+i} = b_N$ a řešená rovnice je splněna.

Nyní je třeba ukázat jednoznačnost. Vezměme tedy jiné řešení $\{\tilde{a}_n\}$ naší úlohy (rovnice a počátečních podmínek). Ukážeme silnou indukcí, že se již musí shodovat s naším $\{a_n\}$. Přesně, pro $n \geq n_0$ dokážeme $V(n)$: $\tilde{a}_n = a_n$.

(0) Pro $n = n_0, n_0 + 1, \dots, n_0 + k - 1$ musí podle počátečních podmínek platit $\tilde{a}_n = A_{n-n_0} = a_n$.

(1) Nechť $n \in \mathbb{N}$, $n \geq n + k - 1$. Předpokládejme, že pro $m = n_0, n_0 + 1, \dots, n$ platí $\tilde{a}_m = a_m$. Protože je $\{\tilde{a}_n\}$ řešení, musí platit $\tilde{a}_{N+k} + \sum_{i=0}^{k-1} c_i(N)\tilde{a}_{N+i} = b_N$. Když to aplikujeme s $N = n + 1 - k$, dostáváme $\tilde{a}_{n+1} + \sum_{i=0}^{k-1} c_i(n+1-k)\tilde{a}_{n+1-k+i} = b_{n+1-k}$ neboli $\tilde{a}_{n+1} = b_{n+1-k} - \sum_{i=0}^{k-1} c_i(n+1-k)\tilde{a}_{n+1-k+i}$. Použijeme indukční předpoklad a podmínu (1) z definice $\{a_n\}$ a dostáváme

$$\tilde{a}_{n+1} = b_{n+1-k} - \sum_{i=0}^{k-1} c_i(n+1-k)\tilde{a}_{n+1-k+i} = b_{n+1-k} - \sum_{i=0}^{k-1} c_i(n+1-k)a_{n+1-k+i} = a_{n+1}.$$

Tím je indukční krok dokázán a tedy i jednoznačnost nalezeného řešení.

2) Teď dokážeme první část věty, tedy existenci nějakého řešení libovolné lineární rekurentní rovnice. To je ale snadné. Když je dána lineární rekurentní rovnice rádu k , tak si prostě zvolíme nějaké počáteční podmínky, třeba pro jednoduchost $a_{n_0} = a_{n_0+1} = \dots = a_{n_0+k-1} = 0$, a podle 1) k nim najdeme řešení, máme proto nějaké řešení dané rovnice. □

Důsledek 10a.3.

Nechť $\{a_m\}_{m=n_0}^\infty$ a $\{\tilde{a}_m\}_{m=n_0}^\infty$ jsou dvě řešení téže lineární rekurentní rovnice rádu k . Jestliže se shoduje prvních k členů těchto řešení, pak se shodují celé posloupnosti.

Důkaz (poučný): Označme $A_0 = a_{n_0}, A_1 = a_{n_0+1}, \dots, A_{k-1} = a_{n_0+k-1}$. Pak $\{a_m\}$ řeší onu lineární rekurentní rovnici a také splňuje počáteční podmínky A_i . Díky shodnosti prvních k členů ovšem i řešení $\{\tilde{a}_m\}$ splňuje tyto počáteční podmínky, proto podle Věty o jednoznačnosti musí jít o shodné posloupnosti. □

Věta nám zaručila existenci řešení, ale neporadila, jak jej najít vyjádřené explicitním vzorečkem. Abychom tuto otázku dokázali uspokojivě zodpovědět, musíme vědět více o tom, jakou strukturu řešení mají. Následující věty nepřekvapí nikoho, kdo se již s nějakými lineárními rovnicemi setkal. Jako obvykle ukážeme, že řešení homogenních rovnic utváří vektorový (lineární) prostor. Vhodné posuny tohoto prostoru pak dají řešení dané rovnice pro nenulové pravé strany. To je v kostce obsah zbytku této kapitoly, začneme posledně zmíněným faktem.

!

Definice.

Uvažujme lineární rekurentní rovnici

$$a_{n+k} + c_{k-1}(n)a_{n+k-1} + \cdots + c_2(n)a_{n+2} + c_1(n)a_{n+1} + c_0(n)a_n = b_n, \quad n \geq n_0.$$

Pak se lineární rekurentní rovnice

$$a_{n+k} + c_{k-1}(n)a_{n+k-1} + \cdots + c_2(n)a_{n+2} + c_1(n)a_{n+1} + c_0(n)a_n = 0, \quad n \geq n_0$$

nazývá k ní **přidružená homogenní rovnice**.

Přichází první inzerovaná věta.

!

Věta 10a.4. (o struktuře řešení lineární rekurentní rovnice)

Nechť je dána lineární rekurentní rovnice

$$a_{n+k} + c_{k-1}(n)a_{n+k-1} + \cdots + c_1(n)a_{n+1} + c_0(n)a_n = b_n, \quad n \geq n_0$$

a nějaké její řešení $\{a_{p,n}\}_{n=n_0}^{\infty}$.

Posloupnost $\{a_n\}_{n=n_0}^{\infty}$ je řešením této rovnice právě tehdy, pokud se dá napsat jako $\{a_n\} = \{a_{p,n}\} + \{a_{h,n}\}$, kde $\{a_{h,n}\}_{n=n_0}^{\infty}$ je nějaké řešení přidružené homogenní rovnice.

! Tento důkaz je v zásadě stejný jako u lineárních rovnic v předchozích kapitolách. Zda je posloupnost řešením dokážeme prostě tak, že ji dosadíme do příslušné rovnice; také víme, že praktičtější je dosadit zkoumanou posloupnost do levé (komplikovanější) strany rovnosti a postupně se úpravami propracovat k pravé.

Důkaz (rutinní, poučný): Dokazujeme oba směry ekvivalence.

1) \iff : Nechť $\{a_{h,n}\}_{n=n_0}^{\infty}$ je nějaké řešení přidružené homogenní rovnice a $\{a_n\} = \{a_{p,n}\} + \{a_{h,n}\}$. Chceme ukázat, že jsme dostali řešení dané rovnice. Dosadíme tedy do rovnice, začneme levou stranou. Pro libovolné $n \geq n_0$ máme

$$\begin{aligned} a_{n+k} + c_{k-1}(n)a_{n+k-1} + \cdots + c_1(n)a_{n+1} + c_0(n)a_n \\ = (a_{p,n+k} + a_{h,n+k}) + c_{k-1}(n)(a_{p,n+k-1} + a_{h,n+k-1}) + \cdots + c_1(n)(a_{p,n+1} + a_{h,n+1}) + c_0(n)(a_{p,n} + a_{h,n}) \\ = [a_{p,n+k} + c_{k-1}(n)a_{p,n+k-1} + \cdots + c_1(n)a_{p,n+1} + c_0(n)a_{p,n}] \\ \quad + [a_{h,n+k} + c_{k-1}(n)a_{h,n+k-1} + \cdots + c_1(n)a_{h,n+1} + c_0(n)a_{h,n}] \\ = b_n + 0 = b_n. \end{aligned}$$

Takže $\{a_{p,n} + a_{h,n}\}_{n=n_0}^{\infty}$ řeší danou rovnici.

2) \implies : Nechť $\{a_n\}_{n=n_0}^{\infty}$ je nějaké řešení dané rovnice. Definujme $a_{h,n} = a_n - a_{p,n}$. Pak evidentně platí rovnost $\{a_n\} = \{a_{p,n}\} + \{a_{h,n}\}$ a zbývá ukázat, že $\{a_{h,n}\}_{n=n_0}^{\infty}$ řeší přidruženou homogenní rovnici. Dosadíme do levé strany, použijeme sumiční zápis, abychom ukázali, jak se tím věci pěkně zjednoduší.

$$\begin{aligned} a_{h,n+k} + \sum_{i=0}^{k-1} c_i(n)a_{h,n+i} &= (a_{n+k} - a_{p,n+k}) + \sum_{i=0}^{k-1} c_i(n)(a_{n+i} - a_{p,n+i}) \\ &= a_{n+k} - a_{p,n+k} + \sum_{i=0}^{k-1} c_i(n)a_{n+i} - \sum_{i=0}^{k-1} c_i(n)a_{p,n+i} \\ &= \left(a_{n+k} + \sum_{i=0}^{k-1} c_i(n)a_{n+i} \right) - \left(a_{p,n+k} + \sum_{i=0}^{k-1} c_i(n)a_{p,n+i} \right) = b_n - b_n = 0. \end{aligned}$$

□

! Když to zapíšeme množinově, tak množina všech řešení dané lineární rekurentní rovnice je

$$\{\{a_{p,n}\} + \{a_{h,n}\}; \{a_{h,n}\} \text{ řeší přidruženou homogenní rovnici}\}.$$

Z toho plyne, že abychom uměli řešit lineární rekurentní rovnice, tak potřebujeme umět opravdu dobře řešit homogenní rovnice, zatímco pro ty nehomogenní stačí nějak uhodnout alespoň jedno řešení. Přesně tuto strategii teď budeme následovat, nejprve se blíže podíváme na to, jak vypadají všechna řešení homogenní rovnice. Na to máme další větu, tentokrát už budeme pracovat jen se sumami, život je krátký.

!

Věta 10a.5. (o struktuře prostoru řešení homogenní lineární rekurentní rovnice)

Nechť je dána homogenní lineární rekurentní rovnice řádu k

$$a_{n+k} + \sum_{i=0}^{k-1} c_i(n) a_{n+i} = 0, \quad n \geq n_0.$$

Pak množina M všech jejích řešení je vektorový prostor dimenze k .

Důkaz (drsný, poučný): 1) Nejprve dokážeme, že M je vektorový prostor. Vezměme dva prvky $\{a_n\}$ a $\{\tilde{a}_n\}$ tohoto prostoru (jde tedy o řešení dané rovnice), nechť $u, v \in \mathbb{R}$. Pak příslušná lineární kombinace $u\{a_n\} + v\{\tilde{a}_n\}$ splňuje pro každé $n \geq n_0$

$$\begin{aligned} (ua_{n+k} + v\tilde{a}_{n+k}) + \sum_{i=0}^{k-1} c_i(n)(ua_{n+i} + v\tilde{a}_{n+i}) &= ua_{n+k} + v\tilde{a}_{n+k} + u \sum_{i=0}^{k-1} c_i(n)a_{n+i} + v \sum_{i=0}^{k-1} c_i(n)\tilde{a}_{n+i} \\ &= u \left(a_{n+k} + \sum_{i=0}^{k-1} c_i(n)a_{n+i} \right) + v \left(\tilde{a}_{n+k} + \sum_{i=0}^{k-1} c_i(n)\tilde{a}_{n+i} \right) = u \cdot 0 + v \cdot 0 = 0, \end{aligned}$$

takže také řeší danou rovnici a tedy $u\{a_n\} + v\{\tilde{a}_n\} \in M$. M je proto lineární podprostor vektorového prostoru všech posloupností, tudíž vektorový prostor.

2) Teď ukážeme, že dimenze M je k .

Nechť $\{a_{1,n}\}$ je řešení dané rovnice a počátečních podmínek $a_{n_0} = 1, a_{n_0+1} = 0, \dots, a_{n_0+k-1} = 0$.

Nechť $\{a_{2,n}\}$ je řešení dané rovnice a počátečních podmínek $a_{n_0} = 0, a_{n_0+1} = 1, a_{n_0+2} = 0, \dots, a_{n_0+k-1} = 0$.

Postupujeme obdobně dále a dostaneme k řešení, přičemž $\{a_{i,n}\}$ řeší počáteční podmínky $a_{n_0+i-1} = 1$ a $a_{n_0+j-1} = 0$ pro $j \neq i, 1 \leq j \leq k$. Tato řešení existují dle Věty . Všimněte si, že řešení se doplňují. První z těchto řešení začíná $\{1, 0, 0, 0, \dots\}$, druhé začíná $\{0, 1, 0, 0, \dots\}$, třetí začíná $\{0, 0, 1, 0, 0, \dots\}$, takže kdykoliv se u těchto posloupností podílíme na stejný člen (první, druhý až k -tý), tak vždy pouze jedna z nich má hodnotu 1 a ostatní mají hodnotu 0. To bude za chvíli velice důležité.

a) Tvrdíme, že tato řešení jsou lineárně nezávislá. Předpokládejme, že $\sum_{i=1}^k u_i \{a_{i,n}\} = \{0\}$, neboli nechť se jistá lineární kombinace těchto řešení rovná nulové posloupnosti, tedy nulovému prvku prostoru M . Potřebujeme ukázat, že pak nutně $u_i = 0$ pro každé i . Vezměme tedy j z množiny $\{1, 2, \dots, k\}$ a podívejme se v té rovnosti nalevo i napravo na j -tý prvek posloupnosti, neboli na prvek posloupnosti s indexem $n_0 + j - 1$. Každý prvek nulové posloupnosti je prostě 0, proto

$$\sum_{i=1}^k u_i a_{i,n_0+j-1} = 0$$

Na výraz nalevo aplikujeme počáteční podmínky pro jednotlivá řešení. Jak už jsme diskutovali, jen jediné z těchto řešení má nenulový svůj j -tý člen (člen s indexem $n_0 + j - 1$), jde o řešení $\{a_{j,n}\}$ a ten člen je roven $a_{j,n_0+j-1} = 1$. Proto ta rovnice ve skutečnosti zní $u_j = 0$, přesně jak jsme potřebovali.

Dokázali jsme, že z rovnosti $\sum_{i=1}^k u_i \{a_{i,n}\} = \{0\}$ už nutně vyplývá $u_i = 0$ pro všechna i , proto je množina $\{\{a_{i,n}\}_{n=n_0}^\infty; i = 1, 2, \dots, k\}$ lineárně nezávislá. Z toho mimo jiné plyne, že $\dim(M) \geq k$.

b) Nyní dokážeme, že oněch k posloupností také generuje M , tedy je to vlastně báze. Vezměme nějaké libovolné řešení $\{a_n\} \in M$, potřebujeme ukázat, že se dá vyjádřit jako lineární kombinace našich k posloupností. Podívejme se na lineární kombinaci $\{\tilde{a}_n\} = \sum_{i=1}^k a_{n_0+i-1} \{a_{i,n}\}$ (zde tedy používáme prvních k členů onoho daného řešení jako koeficienty lineární kombinace). Nechť j je nějaké číslo mezi 1 a k . Jaký je j -tý člen posloupnosti $\{\tilde{a}_n\}$, tj. člen \tilde{a}_{n_0+j-1} ? Jak jsme již diskutovali, ona řešení zahrnutá v dané lineární kombinaci jsou nenulová vždy pouze jednou, takže dostáváme

$$\tilde{a}_{n_0+j-1} = \sum_{i=1}^k a_{n_0+i-1} a_{i,n_0+j-1} = a_{n_0+j-1} \cdot 1 = a_{n_0+j-1}.$$

To znamená, že posloupnosti $\{\tilde{a}_n\}$ a $\{a_n\}$ jsou obě řešením dané rovnice a mají také shodné všechny ze svých prvních k -členů, proto se podle Důsledku rovnají, tedy $\{a_n\} = \{\tilde{a}_n\} = \sum_{i=1}^k a_{n_0+i-1} \{a_{i,n}\}$.

Ukázali jsme, že libovolné řešení dané rovnice (tj. libovolný prvek množiny M) lze vyjádřit jako lineární kombinaci posloupností $\{\{a_{i,n}\}_{n=n_0}^\infty; i = 1, 2, \dots, k\}$.

je báze prostoru M a proto $\dim(M) = k$.

□

! Jaký to má důsledek? Lineární algebra nabízí zajímavou zkratku. Stačí najít nějakou bázi prostoru těchto řešení, jinými slovy, stačí najít nějakých k řešení $\{a_{i,n}\}_{n=n_0}^{\infty}$ této homogenní rovnice tak, aby byly lineárně nezávislé, a už známe všechna řešení, jmenovitě se dají vyjádřit jako lineární kombinace $\sum_{i=1}^k u_i \{a_{i,n}\} = \left\{ \sum_{i=1}^k u_i a_{i,n} \right\}_{n=n_0}^{\infty}$. Všimněte si, že se tam objevuje k parametrů u_i , což přesně souhlasí, dostáváme tak obecné řešení dané homogenní rovnice.

Na něco podobného jsme narazili u králíků (příklad). Rovnice $a_{n+1} - 4a_n = 0$ je homogenní rovnice 1. rádu, podle této věty je prostor jejich řešení jednorozměrný. My jsme zjistili, že všechny posloupnosti $\{c4^n\} = c\{4^n\}$ jsou řešeními, toto jedno řešení $\{4^n\}$ tedy představuje jednoprvkovou bázi prostoru řešení a vše souhlasí.

To je krásné, ale zatím je to jen samá teorie, pořád ještě nevíme, jak vlastně nějaké to řešení najít. Abychom to napravili, musíme se omezit na ještě pěknější typ rovnic.

Cvičení

Cvičení 10a.1 (rutinní): Dokažte, že daná posloupnost řeší danou rovnici, případně i s počátečními podmínkami:

- (i) $\{2^n\}_{n=3}^{\infty}$; $a_{n+2} - 4a_n = 0$, $n \geq 3$;
- (ii) $\{2^n\}_{n=1}^{\infty}$; $a_{n+2} + a_{n+1} - a_n = 5 \cdot 2^n$, $n \geq 1$; $a_1 = 2$, $a_2 = 4$;
- (iii) $\{n2^n + 1\}_{n=0}^{\infty}$; $a_{n+2} - 3a_{n+1} + 2a_n = 2 \cdot 2^n$, $n \geq 0$; $a_0 = 1$, $a_1 = 3$;
- (iv) $\{2^n + 3^n\}_{n=0}^{\infty}$; $a_{n+2} - 5a_{n+1} + 6a_n = 0$, $n \geq 0$;
- (v) $\{2^n + n\}_{n=1}^{\infty}$; $a_{n+2} - a_{n+1} - 2a_n = 1 - 2n$, $n \geq 1$; $a_1 = 3$, $a_2 = 6$.

Řešení:

10a.1: Stačí dosadit, nejlépe začít levou stranou:

- (i): $a_{n+2} - 4a_n = 2^{n+2} - 4 \cdot 2^n = 4 \cdot 2^n - 4 \cdot 2^n = 0$ pro $n \geq 3$;
- (ii): $2^{n+2} + 2^{n+1} - 2^n = 4 \cdot 2^n + 2 \cdot 2^n - 2^n = 5 \cdot 2^n$ pro $n \geq 1$, počáteční podmínky: $a_1 = 2^1 = 2$, $a_2 = 2^2 = 4$ souhlasí;
- (iii): $a_{n+2} - 3a_{n+1} + 2a_n = (2^{n+2}(n+2) + 1) - 3(2^{n+1}(n+1) + 1) + 2(2^n n + 1)$
 $= 4n2^n + 8 \cdot 2^n + 1 - 6n2^n - 6 \cdot 2^n - 3 + 2n2^n + 2 = 2 \cdot 2^n$ pro $n \geq 0$, počáteční podmínky: $a_0 = 2^0 \cdot 0 + 1 = 1$, $a_1 = 2^1 \cdot 1 + 1 = 3$ souhlasí;
- (iv): $a_{n+2} - 5a_{n+1} + 6a_n = (2^{n+2} + 3^{n+2}) - 5(2^{n+1} + 3^{n+1}) + 6(2^n + 3^n) = 4 \cdot 2^n + 9 \cdot 3^n - 10 \cdot 2^n - 15 \cdot 3^n + 6 \cdot 2^n + 6 \cdot 3^n = 0$ pro $n \geq 0$;
- (v): $a_{n+2} - a_{n+1} - 2a_n = (2^{n+2} + (n+2)) - (2^{n+1} + (n+1)) - 2(2^n + n) = 4 \cdot 2^n + n + 2 - 2 \cdot 2^n - n - 1 - 2 \cdot 2^n - 2n = 1 - 2n$ pro $n \geq 1$, počáteční podmínky: $a_1 = 2^1 + 1 = 3$, $a_2 = 2^2 + 2 = 6$ souhlasí.

10b. Rovnice s konstantními koeficienty

Zde se konečně naučíme rovnice řešit. Nejprve se ale musíme vzdát možnosti, že by koeficienty rovnice závisely na n .

!

Definice.

Lineární rekurentní rovnice s konstantními koeficienty je libovolná rovnice ve tvaru

$$a_{n+k} + c_{k-1}a_{n+k-1} + \cdots + c_1a_{n+1} + c_0a_n = b_n, \quad n \geq n_0,$$

kde $n_0 \in \mathbb{Z}$, $c_i \in \mathbb{R}$ pro $i = 0, \dots, k-1$ jsou pevně zvolená čísla a $\{b_n\}_{n=n_0}^{\infty}$ je pevně zvolená posloupnost reálných čísel.

S takovými rovnicemi se už dá něco dělat. Všimněte si, že příklady , a jsou tohoto typu.

!

Definice.

Nechť je dána lineární rekurentní rovnice s konstantními koeficienty

$$a_{n+k} + c_{k-1}a_{n+k-1} + \cdots + c_1a_{n+1} + c_0a_n = b_n, \quad n \geq n_0.$$

Její **charakteristický polynom** (**characteristic polynomial**) je definován jako polynom

$$p(\lambda) = \lambda^k + c_{k-1}\lambda^{k-1} + \cdots + c_1\lambda + c_0.$$

Kořeny charakteristického polynomu se nazývají **charakteristická čísla**, popřípadě **vlastní čísla** dané rovnice (**characteristic numbers/roots or eigenvalues**).

K získání charakteristických čísel potřebujeme vyřešit rovnici $\lambda^k + c_{k-1}\lambda^{k-1} + \dots + c_1\lambda + c_0 = 0$, které se také říká **charakteristická rovnice (characteristic equation)**.

V příkladě s králíky jsme měli rovnici $a_{n+1} - 4a_n = 0$ a řešení $\{4^n\}$, podle nové definice také máme charakteristický polynom $\lambda - 4$ a charakteristické číslo $\lambda = 4$. Tato shoda není náhoda.

! Fakt 10b.1.

Jestliže je λ charakteristickým číslem dané homogenní lineární rekurentní rovnice s konstantními koeficienty

$$a_{n+k} + c_{k-1}a_{n+k-1} + \dots + c_1a_{n+1} + c_0a_n = 0, \quad n \geq n_0,$$

pak je geometrická posloupnost $\{\lambda^n\}_{n=n_0}^\infty$ jejím řešením.

Důkaz (poučný): Dosadíme posloupnost $a_n = \lambda^n$ do dané rovnice (začneme levou stranou). Protože je λ nulovým bodem charakteristického polynomu, dostaneme pro libovolné $n \geq n_0$

$$\begin{aligned} a_{n+k} + c_{k-1}a_{n+k-1} + \dots + c_1a_{n+1} + c_0a_n &= \lambda^{n+k} + c_{k-1}\lambda^{n+k-1} + \dots + c_1\lambda^{n+1} + c_0\lambda^n \\ &= \lambda^n(\lambda^k + c_{k-1}\lambda^{k-1} + \dots + c_1\lambda + c_0) = \lambda^n p(\lambda) = \lambda^n \cdot 0 = 0. \end{aligned}$$

Tato posloupnost tedy řeší danou rovnici. □

Jak víme, polynom stupně k má k kořenů (pokud bereme i komplexní a včetně násobnosti). Pro nalezení báze řešení rovnice řádu k zase potřebujeme k posloupností. To vypadá slibně, nejprve snadný případ:

! Věta 10b.2.

Uvažujme homogenní lineární rekurentní rovnici s konstantními koeficienty řádu k

$$a_{n+k} + c_{k-1}a_{n+k-1} + \dots + c_1a_{n+1} + c_0a_n = 0, \quad n \geq n_0.$$

Jestliže má k různých charakteristických čísel λ_i , pak posloupnosti $\{\lambda_i^n\}_{n=n_0}^\infty$ tvoří bázi prostoru řešení dané rovnice.

To je velice příjemné tvrzení, protože obecné řešení této rovnice pak je $\sum_{i=1}^k u_i \{\lambda_i^n\} = \left\{ \sum_{i=1}^k u_i \lambda_i^n \right\}$.

Důkaz (drsný): Dané posloupnosti tvoří řešení dle předchozího tvrzení, stačí tedy dokázat, že jsou nezávislé, a bude to i báze, neboť je jich tolik, kolik je dimenze prostoru řešení (Věta).

Potřebujeme ukázat, že rovnice $\sum_{i=1}^k u_i \{\lambda_i^n\} = \{0\}$ nutně vede na $u_i = 0$ pro všechna i . Jako obvykle se stačí podívat na prvních k členů zúčastněných posloupností, dostaneme následující soustavu rovnic:

$$\begin{array}{llllllll} u_1\lambda_1^{n_0} & +u_2\lambda_2^{n_0} & +\dots+u_k\lambda_k^{n_0} & = 0 & u_1 & +u_2 & +\dots+u_k & = 0 \\ u_1\lambda_1^{n_0+1} & +u_2\lambda_2^{n_0+1} & +\dots+u_k\lambda_k^{n_0+1} & = 0 & u_1\lambda_1 & +u_2\lambda_2 & +\dots+u_k\lambda_k & = 0 \\ u_1\lambda_1^{n_0+2} & +u_2\lambda_2^{n_0+2} & +\dots+u_k\lambda_k^{n_0+2} & = 0 & \implies u_1\lambda_1^2 & +u_2\lambda_2^2 & +\dots+u_k\lambda_k^2 & = 0 \\ \vdots & \vdots \\ u_1\lambda_1^{n_0+k-1} & +u_2\lambda_2^{n_0+k-1} & +\dots+u_k\lambda_k^{n_0+k-1} & = 0 & u_1\lambda_1^{k-1} & +u_2\lambda_2^{k-1} & +\dots+u_k\lambda_k^{k-1} & = 0 \end{array}$$

Potřebujeme ukázat, že jejím jediným řešením je to triviální, což znamená, že potřebujeme ukázat, že matice soustavy

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_k \\ \lambda_1^2 & \lambda_2^2 & \dots & \lambda_k^2 \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{k-1} & \lambda_2^{k-1} & \dots & \lambda_k^{k-1} \end{pmatrix}$$

je regulární. To je ale pro navzájem různá λ_i standardní fakt z lineární algebry, jde o tzv. Vandermondův determinant. □

Příklad 10b.a: V příkladě jsme se dívali na funkci zadanou induktivně podmínkami $f(1) = 3$, $f(2) = -1$ a $f(n) = f(n-1) + 6f(n-2)$ pro $n \geq 3$. V řeči posloupností to je $f_n = f_{n-1} + 6f_{n-2}$, do správného tvaru lineární rovnice to převedeme například substitucí $m = n - 2$ neboli $n = m + 2$. Dostáváme homogenní lineární rovnici $f_{m+2} - f_{m+1} - 6f_m = 0$ pro $m \geq 1$, hledáme řešení splňující počáteční podmínky $f_1 = 3$, $f_2 = -1$. Protože je rovnice 2. řádu, počet podmínek souhlasí.

Řešení: Nejprve najdeme obecné řešení. Určíme charakteristický polynom:

$$p(\lambda) = \lambda^2 - \lambda - 6 = (\lambda - 3)(\lambda + 2).$$

Z rovnice $(\lambda - 3)(\lambda + 2) = 0$ dostáváme charakteristická čísla $\lambda = 3, -2$ a obecné řešení $\{u \cdot 3^n + v \cdot (-2)^n\}_{n=1}^{\infty}$, tedy $f_n = 3^n u + (-2)^n v$ pro libovolná $u, v \in \mathbb{R}$.

Ted' mezi těmito nekonečně mnoha řešeními najdeme to, které splňuje počáteční podmínky. Ty vyžadují následující:

$$\begin{aligned} f_1 &= 3^1 u + (-2)^1 v = 3 \\ f_2 &= 3^2 u + (-2)^2 v = -1 \end{aligned} \implies \begin{aligned} 3u - 2v &= 3 \\ 9u + 4v &= -1. \end{aligned}$$

Odtud hravě vykutáme, že $u = \frac{1}{3}$, $v = -1$. Řešení dané úlohy tedy je $\left\{ \frac{1}{3} \cdot 3^n + (-1) \cdot (-2)^n \right\}_{n=1}^{\infty}$ neboli (přejdeme k funkcím dle zadání) $f(n) = 3^{n-1} - (-2)^n$ pro $n \geq 1$.

Po vyřešení úlohy bývá dobré udělat zkoušku, tj. ověřit, že to, co jsme našli, je opravdu řešení.

Takže: $f(1) = 3^0 - (-2) = 3$ a $f(2) = 3^1 - (-2)^2 = -1$, to souhlasí, ted' se podíváme na induktivní rovnici:

$$\begin{aligned} f(n) + 6f(n-1) &= [3^{n-1} - (-2)^n] + 6[3^{n-2} - (-2)^{n-1}] = 3^{n-1} - (-2)^n + 2 \cdot 3^{n-1} + 3 \cdot (-2)^n \\ &= 3 \cdot 3^{n-1} + 2 \cdot (-2)^n = 3^{(n+1)-1} - (-2)^{n+1} = f(n+1). \end{aligned}$$

Poznámka: Pokud bychom při zápisu obecného řešení chtěli použít metody z lineární algebry, mohli bychom množinu všech řešení zapsat například těmito způsoby:

$$\left\{ \{3^{n-1} u + (-2)^n v\}_{n=1}^{\infty}; u, v \in \mathbb{R} \right\} = \left\{ u\{3^{n-1}\}_{n=1}^{\infty} + v\{(-2)^n\}_{n=1}^{\infty}; u, v \in \mathbb{R} \right\} = \langle \{3^{n-1}\}_{n=1}^{\infty}, \{(-2)^n\}_{n=1}^{\infty} \rangle.$$

△

! Nejčastěji řešíme dva druhy úloh.

1) Chceme obecné řešení dané rovnice (tedy vlastně chceme znát množinu všech řešení). Pak najdeme bázi prostoru řešení pomocí charakteristických čísel a obecné řešení dostaneme jako lineární kombinaci této báze.

2) Chceme řešení rovnice, které navíc splňuje dané počáteční podmínky. Jak jsme právě viděli, tento druhý typ úlohy se obvykle řeší ve dvou krocích. Nejprve se najde obecné řešení podle postupu 1) a pak se určí hodnoty konstant tak, aby výsledné řešení splňovalo počáteční podmínky. Tento druhý krok je v zásadě triviální, prostě si napíšeme, co od obecného řešení chceme (počáteční podmínky), pak vyřešíme vzniklých k rovnic o k neznámých a je to.

Příklad 10b.b: Zde zkusíme jiný pohled na králíky. Na začátku prvního měsíce dostaneme párek čerstvě narozených králíků. Nechť F_n je počet párů na začátku n -tého měsíce od obdržení. Podle jakých pravidel se králíci množí? Uvažujme tyto zásady:

- králíci se začnou množit ve chvíli, kdy jsou jim 2 měsíce;
- když se začnou množit, tak pak mají každý měsíc jeden pár mladých;
- králíci nikdy neumřou (jsou to matematictí králíci).

Jak to pak s jejich počty vypadá? Na začátku 1. měsíce je $F_1 = 1$ pár. Pořád je ještě mladý, takže i na začátku druhého měsíce je jen $F_2 = 1$ pár. Pak se ale začne množit a na začátku dalšího měsíce už má mladé, proto $F_3 = 1 + 1 = 2$. Na začátku 4. měsíce se ten první pár znova zmnožil, ale druhý je ještě mladý, takže $F_4 = 2 + 1 = 3$. A tak dále, jak to vlastně funguje?

Chceme-li znát, kolik bude králíků na začátku měsíce $n+1$, tak výchozím stavem je samozřejmě stav předchozí, tedy F_n . K tomu se musí přičíst nové přírůstky, což se přesně rovná počtu párů, kterým je na začátku měsíce $n+1$ alespoň dva měsíce, neboli počtu párů, které už tu byly na začátku měsíce $n-1$. Dostáváme $F_{n+1} = F_n + F_{n-1}$, což je rovnice udávající Fibonacciho posloupnost, viz příklad . Přesně takovou úvahou (včetně nesmrtelnosti králíků) k ní ctihodný Leonardo z Pisy známý též jako Fibonacci příšel.

Přepíšeme si to jako $F_{n+2} - F_{n+1} - F_n = 0$ a tuto homogenní lineární rekurentní rovnici 2. řádu s počátečními podmínkami $F_1 = 1$, $F_2 = 1$ hravě vyřešíme.

Nejprve obecné řešení: Z $p(\lambda) = \lambda^2 - \lambda - 1 = 0$ dostáváme charakteristická čísla $\lambda = \frac{1 \pm \sqrt{5}}{2}$. Obecné řešení dané rovnice je tedy $F_n = u\left(\frac{1+\sqrt{5}}{2}\right)^n + v\left(\frac{1-\sqrt{5}}{2}\right)^n$.

Aplikujeme počáteční podmínky:

$$\begin{aligned} F_1 &= u\frac{1+\sqrt{5}}{2} + v\frac{1-\sqrt{5}}{2} = 1 & \implies u(1+\sqrt{5}) + v(1-\sqrt{5}) &= 2 \\ F_2 &= u\left(\frac{1+\sqrt{5}}{2}\right)^2 + v\left(\frac{1-\sqrt{5}}{2}\right)^2 = 1 & \implies u(3+\sqrt{5}) + v(3-\sqrt{5}) &= 2 \end{aligned}$$

Když rovnice odečteme, dostaneme $u + v = 0$, odtud $v = -u$, dosadíme do první rovnice a dostaneme $2u\sqrt{5} = 2$, tedy $u = \frac{1}{\sqrt{5}} = -v$.

Závěr: Fibonacciho posloupnost je dána explicitním vzorcem $F_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2} \right)^n$.

Tato posloupnost možná není nejlepším modelem pro množení králíků, ale objevuje se při zkoumání jevů překvapivě často. Již jsme ji viděli při odhadu rychlosti Euklidova algoritmu (Věta) a ještě ji potkáme jak v této kapitole, tak v příkladě.

Zajímavé je číslo $\varphi = \frac{1+\sqrt{5}}{2} = 1.618\dots$. Je to přesně hodnota zlatého řezu, další úžasná náhoda. Protože je to to větší z charakteristických čísel, je dominantní a tedy v jazyce kapitoly můžeme říct, že $F_n = \Theta(\varphi^n)$.

Možná vám vrtá hlavou, co by se stalo, kdybychom do toho zapracovali smrtelnost králíků. Zkusme třeba toto:

- na konci čtvrtého měsíce králíci umírají.

Jaký model z toho vyjde teď? To je trochu komplikovanější, je to možná lépe vidět, když si zavedeme posloupnost b_n , která říká, kolik párů králíků se narodilo na začátku měsíce n . Toto číslo se rovná počtu párů, kterým je v té chvíli dva či tři měsíce, protože ti mladší ještě nemohou a ti starší už mají přesně opačné starosti. Dostáváme vztah $b_n = b_{n-2} + b_{n-3}$.

Aktuální stav na začátku měsíce n , označme jej c_n , je dán součtem počtu králíků, kteří se právě narodili či jsou měsíc, dva a tři měsíce staří, tedy $c_n = b_n + b_{n-1} + b_{n-2} + b_{n-3}$. Když na všechny b_i aplikujeme již odvozený vztah, dostaneme

$$\begin{aligned} c_n &= (b_{n-2} + b_{n-3}) + (b_{n-3} + b_{n-4}) + (b_{n-4} + b_{n-5}) + (b_{n-5} + b_{n-6}) \\ &= (b_{n-2} + b_{n-3} + b_{n-4} + b_{n-5}) + (b_{n-3} + b_{n-4} + b_{n-5} + b_{n-6}) = c_{n-2} + c_{n-3}. \end{aligned}$$

Posun indexu dává $c_{n+1} = c_{n-1} + c_{n-2}$, je to tedy podobné Fibonacciho vztahu, počátek posloupnosti dokonce s Fibonacciho posloupností souhlasí, $c_1 = c_2 = 1$, $c_3 = 2$, protože v té době se ještě smrtelnost neprojevila. Dostáváme posloupnost $\{1, 1, 2, 2, 3, 4, 5, 7, 9, 12, 16, 21, \dots\}$, která evidentně (a logicky) roste pomaleji než ta pro Fibonacciho nesmrtelné králíky. Jak rychle vlastně roste?

Přepis dává homogenní lineární rekurentní rovnici $c_{n+3} - c_{n+1} - c_n = 0$ třetího rádu. Bohužel, její charakteristická rovnice $\lambda^3 - \lambda - 1 = 0$ nemá žádné „pěkné“ kořeny. To ale nevadí. K určení rychlosti růstu nám vlastně stačí znát kořen, který je v absolutní hodnotě největší, ten se dá určit i přibližně některou z oblíbených metod (Newtonova, bisekce), vychází $1.325\dots$. To znamená, že c_n je přibližně $\Theta((1.325)^n)$.

△

Zatím jsme měli v příkladech štěstí a vždy jsme dostali k různých charakteristických čísel. Na to se ale nedá spoléhat a v případě opakování kořene zatím nevíme, co dělat. Navíc ani různé kořeny ještě neznamenají úspěch, protože ony mohou být komplexní, zatímco my zde řešíme rovnice v reálném oboru (konec konců uvažujeme jen reálné koeficienty rovnic). S těmito dvěma problémy se musíme naučit vyrovnat, začneme tím prvním.

Fakt 10b.3.

Nechť je dána homogenní lineární rekurentní rovnice s konstantními koeficienty. Jestliže je λ její charakteristické číslo a má násobnost m jako kořen charakteristického polynomu, pak posloupnosti $\{\lambda^n\}, \{n\lambda^n\}, \dots, \{n^{m-1}\lambda^n\}$ jsou řešením dané rovnice a tvoří lineárně nezávislou množinu.

Pro zkrácení budeme v takové situaci prostě říkat, že λ je charakteristické číslo násobnosti m .

Důkaz (náznak, poučný): Důkaz tohoto faktu je obecně drobet dobrodružnější, proto ukážeme, proč to platí pro dvojnásobný kořen. Použijeme obecný postup, který funguje i pro charakteristická čísla vyšší násobnosti.

Mějme tedy rovnici $a_{n+k} + \sum_{i=0}^{k-1} c_i a_{n+i} = 0$, kde $c_0 \neq 0$. Její charakteristický polynom je $p(\lambda) = \lambda^k + \sum_{i=0}^{k-1} c_i \lambda^i$.

Jestliže je λ_0 alespoň dvojnásobný kořen, pak teorie polynomů (nebo obecně funkcí) říká, že platí nejen $p(\lambda_0) = 0$, ale také $p'(\lambda_0) = 0$ (zde p' je derivace). To budeme záhy potřebovat.

1) Potřebujeme dokázat, že posloupnosti $\{\lambda_0^n\}$ a $\{n\lambda_0^n\}$ řeší danou rovnici. Pro první posloupnost už to víme,

zbývá tedy ověřit, že i ta druhá funguje. Dosadíme do levé strany rovnice.

$$\begin{aligned}
 a_{n+k} + \sum_{i=0}^{k-1} c_i a_{n+i} &= (n+k)\lambda_0^{n+k} + \sum_{i=0}^{k-1} c_i(n+i)\lambda_0^{n+i} = n\lambda_0^{n+k} + k\lambda_0^{n+k} + \sum_{i=0}^{k-1} c_i n \lambda_0^{n+i} + \sum_{i=0}^{k-1} c_i i \lambda_0^{n+i} \\
 &= n\lambda_0^{n+k} + \sum_{i=0}^{k-1} c_i n \lambda_0^{n+i} + k\lambda_0^{n+k} + \sum_{i=0}^{k-1} c_i i \lambda_0^{n+i} \\
 &= n\lambda_0^n \left(\lambda_0^k + \sum_{i=0}^{k-1} c_i \lambda_0^i \right) + \lambda_0^{n+1} \left(k\lambda_0^{k-1} + \sum_{i=0}^{k-1} c_i i \lambda_0^{i-1} \right) \\
 &= n\lambda_0^n p(\lambda_0) + \lambda_0^{n+1} p'(\lambda_0) = n\lambda_0^n \cdot 0 + \lambda_0^{n+1} \cdot 0 = 0.
 \end{aligned}$$

Posloupnost $\{n\lambda_0^n\}$ tedy splňuje danou rovnici.

2) Teď ještě potřebujeme dokázat lineární nezávislost těch dvou posloupností. Podobně jako u předchozího důkazu nezávislosti se to nakonec převede na otázku, zda je matice $\begin{pmatrix} 1 & 0 \\ \lambda & \lambda \end{pmatrix}$ regulární. Protože $c_0 \neq 0$, musí být i $\lambda_0 \neq 0$ a matice regulární je.

3) Obecně by se muselo dokázat, že $\{n^{m-1}\lambda^n\}$ je řešením, pokud má λ násobnost alespoň m , což by se dělalo podobně, jen to dá víc práce, použilo by se přitom faktu, že pro kořen λ_0 násobnosti m se v bodě λ_0 vynulují i derivace p' , p'' , ..., $p^{(m-1)}$.

Nezávislost by se pak redukovala na otázku regulárnosti matice

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \lambda & \lambda & \lambda & \dots & \lambda \\ \lambda^2 & 2\lambda^2 & 2^3\lambda^2 & \dots & 2^{m-1}\lambda^2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \lambda^{m-1} & (m-1)\lambda^{m-1} & (m-1)^3\lambda^{m-1} & \dots & (m-1)^{m-1}\lambda^{m-1} \end{pmatrix}$$

I to je něco, na co nám kladně odpoví lineární algebra.

□

Vidíme tedy, že pokud má některý kořen charakteristického polynomu vyšší násobnost m , tak sice bude o $m-1$ méně různých kořenů ve Větě, ale stejný počet nezávislých řešení zase přibyde díky tomu násobení členem n^i . Celkový počet řešení je tedy pořád správný, ale ještě zbývá dokázat, že tato nová řešení typu $\{n^i\lambda^n\}$ nezkazí nezávislost po přidání k posloupnostem pocházejícím od ostatních λ .

! Věta 10b.4.

Nechť je dána homogenní lineární rekurentní rovnice s konstantními koeficienty rádu k . Nechť jsou $\lambda_1, \dots, \lambda_M$ její různá charakteristická čísla, přičemž každé λ_i má násobnost $m_i \in \mathbb{N}$. Pak je množina

$$\{\{\lambda_1^n\}, \{n\lambda_1^n\}, \dots, \{n^{m_1-1}\lambda_1^n\}, \{\lambda_2^n\}, \{n\lambda_2^n\}, \dots, \{n^{m_2-1}\lambda_2^n\}, \dots, \{\lambda_M^n\}, \{n\lambda_M^n\}, \dots, \{n^{m_M-1}\lambda_M^n\}\}$$

bází prostoru řešení dané rovnice.

Vlastně už toho hodně máme dokázaného, zbývá lineární nezávislost výsledné množiny řešení, která se zase redukuje na regularitu jisté matice a to je práce pro lineární algebru (tentokrát opravdu drsná). Stejně jako autoři většiny učebnic na toto téma i my toho čtenáře (a sebe) ušetříme. Opravdu zvědavého čtenáře odkážeme na nějakou podrobnější učebnici diferenciálních rovnic, kde se u homogenních lineárních rovnic dělá prakticky totéž.

! Příklad 10b.c: Najdeme obecné řešení rovnice $a_{n+3} - a_{n+2} - a_{n+1} + a_n = 0$, $n \geq -2$.

Řešení: Je to homogenní lineární rekurentní rovnice s konstantními koeficienty, Věta tedy dává návod k řešení. Charakteristický polynom je $p(\lambda) = \lambda^3 - \lambda^2 - \lambda + 1$. Kořeny polynomu třetího stupně se sice dají získat pomocí vzorců, ale ty jsou díky své komplikovanosti vysoko nepopulární. Toto je školní příklad, nenašel by se nějaký pěkný kořen? Zkusíme dosadit malá celá čísla, 0 nefunguje, ale 1 ano. Máme první kořen a teď odloupneme z p příslušný kořenový faktor:

$$p(\lambda) = (\lambda - 1)(\lambda^2 - 1).$$

A je to jasné,

$$p(\lambda) = (\lambda - 1)(\lambda - 1)(\lambda + 1) = (\lambda - 1)^2(\lambda + 1).$$

Máme tedy charakteristická čísla $\lambda = 1$ (dvojnásobné) a $\lambda = -1$ (jednoduché). Podle Věty proto bude množina

$$\{\{1^n\}_{n=-2}^{\infty}, \{n1^n\}_{n=-2}^{\infty}, \{(-1)^n\}_{n=-2}^{\infty}\}$$

bází prostoru všech řešení, dostáváme obecné řešení $\{u \cdot 1^n + v \cdot n1^n + w \cdot (-1)^n\}_{n=-2}^{\infty} = \{u + vn + w(-1)^n\}_{n=-2}^{\infty}$ pro $u, v, w \in \mathbb{R}$.

Pro úplnost uděláme zkoušku, dosadíme toto řešení do levé strany dané rovnice. Pro $n \geq -2$ máme

$$\begin{aligned} a_{n+3} - a_{n+2} - a_{n+1} + a_n &= (u + v(n+3) + w(-1)^{n+3}) - (u + v(n+2) + w(-1)^{n+2}) \\ &\quad - (u + v(n+1) + w(-1)^{n+1}) + (u + vn + w(-1)^n) \\ &= u + vn + 3v + w(-1)^3(-1)^n - u - vn - 2v - w(-1)^2(-1)^n \\ &\quad - u - vn - v - w(-1)^1(-1)^n + u + vn + w(-1)^n = 0. \end{aligned}$$

△

Teď už tedy teoreticky umíme vyřešit všechny homogenní lineární rovnice s konstantními koeficienty, pokud jsme spokojeni s případnými komplexními posloupnostmi. Jenže my spokojeni nejsme, když zde mluvíme o rovnicích s reálnými koeficienty, tak také očekáváme reálná řešení. Potřebujeme tedy ještě jeden trik.

Je známo, že jestliže je $\lambda = \alpha + \beta i$ kořenem polynomu p , pak je jeho kořenem i $\lambda^* = \alpha - \beta i$. V bázi se tedy objeví i dvojice $\{(\alpha + \beta i)^n\}, \{(\alpha - \beta i)^n\}$. Klíčem je podívat se na jejich lineární kombinace $u(\alpha + \beta i)^n + v(\alpha - \beta i)^n$, ukáže se totiž, že mezi nimi je i dostatečný počet reálných posloupností na to, aby nám daly dvojrozměrný vektorový prostor (nad reálnými čísly), což je přesně to, co potřebujeme.

Dělá se to takto: Napíšeme $\alpha \pm \beta i = r[\cos(\varphi) \pm i \sin(\varphi)]$, pak $(\alpha \pm \beta i)^n = r^n[\cos(n\varphi) \pm i \sin(n\varphi)]$, a proto

$$\begin{aligned} (\alpha + \beta i)^n + (\alpha - \beta i)^n &= 2r^n \cos(n\varphi), \\ (\alpha + \beta i)^n - (\alpha - \beta i)^n &= 2ir^n \sin(n\varphi). \end{aligned}$$

Vidíme tedy, že pokud vezmeme lineární kombinaci $\frac{1}{2}\{\lambda^n + (\lambda^*)^n\}$, tak dostaneme řešení dané rovnice ve tvaru $\{r^n \cos(n\varphi)\}$, což už je reálná posloupnost, zatímco lineární kombinace $\frac{1}{2i}\{\lambda^n - (\lambda^*)^n\}$ dává další reálné řešení $\{r^n \sin(n\varphi)\}$. Dá se také ukázat, že posloupnosti $\{r^n \cos(n\varphi)\}$ a $\{r^n \sin(n\varphi)\}$ jsou nezávislé. Jaké je tedy ponaučení?

! Komplexní kořeny vždy chodí po dvou, takže pokud dostaneme charakteristické číslo $\lambda = r[\cos(\varphi) + i \sin(\varphi)]$, které není reálné, tak použijeme dvojici řešení $\{r^n \cos(n\varphi)\}$ a $\{r^n \sin(n\varphi)\}$. Je-li to kořen vícenásobný, tak obvyklým způsobem dále přihodíme $\{nr^n \cos(n\varphi)\}$ a $\{nr^n \sin(n\varphi)\}$, případně $\{n^2r^n \cos(n\varphi)\}$ a $\{n^2r^n \sin(n\varphi)\}$ až po $\{n^{m-1}r^n \cos(n\varphi)\}$ a $\{n^{m-1}r^n \sin(n\varphi)\}$, kde m je násobnost kořene λ . Tato řešení (kterých je $2m$) zároveň zastupují řešení pro sdružený kořen λ^* .

Teď už tedy opravdu umíme vyřešit všechny homogenní lineární rekurentní rovnice s konstantními koeficienty. Mohli bychom to zase potvrdit větou, ale stejně bychom ji nedokazovali, tak raději rovnou zformulujeme algoritmus.

S Algoritmus 10b.5. pro řešení homogenní lineární rekurentní rovnice $a_{n+k} + \sum_{i=0}^{k-1} c_i a_{n+i} = 0, n \geq n_0$ řádu k .

1. Sestavte charakteristický polynom $p(\lambda) = \lambda^k + \sum_{i=0}^{k-1} c_i \lambda^i$.

Řešením rovnice $p(\lambda) = 0$ najděte všechna charakteristická čísla dané rovnice.

2. Sestavte množinu posloupností B takto:

- pro každé reálné charakteristické číslo λ přidejte do B posloupnost $\{\lambda^n\}_{n=n_0}^{\infty}$;
- pro každé reálné charakteristické číslo λ , jehož násobnost je $m > 1$, přidejte do B rovněž posloupnosti $\{n\lambda^n\}_{n=n_0}^{\infty}, \dots, \{n^{m-1}\lambda^n\}_{n=n_0}^{\infty}$;
- pro každé komplexní charakteristické číslo $\lambda = r[\cos(\varphi) + i \sin(\varphi)]$, které není reálné, přidejte do B posloupnosti $\{r^n \cos(n\varphi)\}_{n=n_0}^{\infty}$ a $\{r^n \sin(n\varphi)\}_{n=n_0}^{\infty}$; pro jeho komplexně sdružené číslo λ^* již do B nic nepřidáváme;
- pro každé komplexní charakteristické číslo $\lambda = r[\cos(\varphi) + i \sin(\varphi)]$, které není reálné a jehož násobnost je $m > 1$, přidejte do B posloupnosti $\{nr^n \cos(n\varphi)\}_{n=n_0}^{\infty}, \dots, \{n^{m-1}r^n \cos(n\varphi)\}_{n=n_0}^{\infty}$ a $\{nr^n \sin(n\varphi)\}_{n=n_0}^{\infty}, \dots, \{n^{m-1}r^n \sin(n\varphi)\}_{n=n_0}^{\infty}$; pro jeho komplexně sdružené číslo λ^* již do B nic nepřidáváme. Množina B je bází prostoru řešení.

3. Označíme-li $B = \{\{a_{1,n}\}, \dots, \{a_{k,n}\}\}$, pak je obecné řešení dané rovnice určeno vzorcem $\left\{ \sum_{i=1}^k u_i a_{i,n} \right\}_{n=n_0}^{\infty}$ pro $u_1, \dots, u_k \in \mathbb{R}$.

4. Jsou-li dány počáteční podmínky, pak do nich za příslušná a_j pro $j = n_0, \dots, n_0 + k - 1$ dosadíme vzorce $a_j = \sum_{i=1}^k u_i a_{i,j}$ a vyřešíme vzniklých k rovnic pro k neznámých u_i . Ty po dosazení do obecného řešení určí příslušné partikulární řešení.

△

Hned si to ukážeme na příkladě.

Příklad 10b.d: Najdeme řešení rovnice $a_{n+3} - a_{n+2} + a_{n+1} - a_n = 0$, $n \geq 13$ s počátečními podmínkami $a_{13} = 1$, $a_{14} = 0$, $a_{15} = 1$.

Řešení: Je to homogenní lineární rekurentní rovnice 3. řádu s konstantními koeficienty a máme tři počáteční podmínky, je to tedy korektně zadaná úloha a použijeme algoritmus.

1) Nejprve najdeme obecné řešení. Charakteristický polynom je $p(\lambda) = \lambda^3 - \lambda^2 + \lambda - 1$. Zkusíme nějaký kořen uhádnout, trefíme se s 1. Máme první kořen a rozložíme p příslušným způsobem:

$$p(\lambda) = (\lambda - 1)(\lambda^2 + 1).$$

A je to jasné,

$$p(\lambda) = (\lambda - 1)(\lambda - i)(\lambda + i).$$

Máme tedy jednoduchá charakteristická čísla $\lambda = 1$ a $\lambda = \pm i = 1 \cdot [\cos(\frac{\pi}{2}) \pm i \sin(\frac{\pi}{2})]$. Podle Algoritmu proto bude množina

$$\left\{ \{1^n\}_{n=13}^{\infty}, \{1^n \cos(n\frac{\pi}{2})\}_{n=13}^{\infty}, \{1^n \sin(n\frac{\pi}{2})\}_{n=13}^{\infty} \right\}$$

bází prostoru všech řešení, dostáváme obecné řešení

$$\left\{ u \cdot 1^n + v \cdot 1^n \cos(n\frac{\pi}{2}) + w \cdot 1^n \sin(n\frac{\pi}{2}) \right\}_{n=13}^{\infty} = \left\{ u + v \cos(n\frac{\pi}{2}) + w \sin(n\frac{\pi}{2}) \right\}_{n=13}^{\infty} \text{ pro } u, v, w \in \mathbb{R}.$$

2) Teď je třeba najít řešení vyhovující počátečním podmínkám. Dosadíme do nich obecné řešení:

$$\begin{aligned} a_{13} &= u + v \cos(13\frac{\pi}{2}) + w \sin(13\frac{\pi}{2}) = 1 & u + w &= 1 \\ a_{14} &= u + v \cos(14\frac{\pi}{2}) + w \sin(14\frac{\pi}{2}) = 0 & \implies u - v &= 0 \\ a_{15} &= u + v \cos(15\frac{\pi}{2}) + w \sin(15\frac{\pi}{2}) = 1 & u - w &= 1. \end{aligned}$$

Odtud $u = v = 1$ a $w = 0$. Dosazením do obecného řešení dostáváme hledané partikulární řešení

$$\{1 + \cos(n\frac{\pi}{2})\}_{n=13}^{\infty}.$$

Vypadá trochu komplikovaně, ve skutečnosti je to periodická posloupnost $\{1, 0, 1, 2, 1, 0, 1, 2, 1, 0, 1, 2, 1, \dots\}$.

△

Krásný algoritmus. Bohužel funguje jen v případě, že máme snadné rovnice. Jakmile má rovnice řád vyšší než dva, je malá šance, že získáme kořeny charakteristické rovnice. Pak nezbývá než hledat kořeny pomocí numerických metod, které ale dávají hodnoty jen přibližné, nastávají problémy s násobností a vůbec je to nepříjemné. Proto bývá jednodušší hledat numericky přímo řešení.

10b.6 Nehomogenní rovnice

Tak jsme se naučili, jak najít všechna řešení homogenní lineární rovnice. Víme už také, že v případě, že je rovnice nehomogenní, nám stačí nějak najít jedno její řešení, a už zase budeme umět najít všechny. Bohužel, k nalezení toho jednoho řešení obecný algoritmus neexistuje. Zde je třeba spoléhat na náhodu, například existuje jistý typ pravé strany b_n , pro který již řešení umíme uhodnout.

Definice.

Řekneme, že posloupnost $\{b_n\}_{n=n_0}^{\infty}$ je **kvazipolynom** (quasipolynomial), jestliže existuje $\lambda \in \mathbb{R}$ a polynom $P(n)$ takový, že $b_n = P(n)\lambda^n$ pro všechna $n \geq n_0$.

Naštěstí pro nás vede mnoho příkladů právě na kvazipolynomiální pravou stranu a můžeme použít následující tvrzení.

Věta 10b.7. (řešení pro kvazipolynomiální pravou stranu)

Uvažujme rovnici

$$a_{n+k} + c_{k-1}a_{n+k-1} + \cdots + c_1a_{n+1} + c_0a_n = b_n, \quad n \geq n_0.$$

Předpokládejme, že existují $\lambda \in \mathbb{R}$ a polynom P takový, že $b_n = P(n)\lambda^n$ pro všechna $n \geq n_0$. Nechť m je násobnost tohoto čísla λ jako charakteristického čísla přidružené homogenní rovnice, přičemž $m = 0$ v případě, že toto λ vůbec charakteristickým číslem není.

Pak existuje polynom $Q(n)$ stupně stejném jako P takový, že $\{n^m Q(n)\lambda^n\}$ je řešením dané rovnice.

Důkaz je poněkud komplikovanější a vynecháme jej.

Všimněte si, že kvazipolynomy zahrnují i polynomy, volba $\lambda = 1$ totiž dává $b_n = P(n)$. Toto pozorování se ještě bude hodit.

Na první pohled to vypadá jako další teoretické tvrzení o existenci řešení, které při praktickém hledání moc nepomůže, ale není tomu tak. Je pravda, že nám věta neřekla, který polynom Q bude fungovat, ale množina nabízených možností je natolik malá, že si z ní již dokážeme vybrat to správné $Q(n)$.

Slouží k tomu tzv. **metoda odhadu či metoda neurčitých koeficientů (method of undetermined coefficients)**. Uhodneme řešení jako $\{n^m Q(n) \lambda^n\}$, kde jsme za Q napsali obecný polynom vhodného stupně, jeho koeficienty tedy budou sloužit jako parametry. Zbývá najít hodnoty těchto parametrů tak, aby vzniklá posloupnost byla řešením, a to se dělá dosazením posloupnosti do dané rovnice. Tím zjistíme, který polynom funguje, a jsme hotovi. Nejlépe to vysvětlí příklad.

! Příklad 10b.e: Najdeme obecné řešení rovnice $a_{n+2} - 2a_{n+1} - 3a_n = -9n2^n$, $n \geq 0$.

Využijeme Větu a začneme tím snažším.

1) Najdeme obecné řešení rovnice $a_{n+2} - 2a_{n+1} - 3a_n = 0$, $n \geq 0$. Charakteristická rovnice $\lambda^2 - 2\lambda - 3 = 0$ dává charakteristická čísla $-1, 3$ a obecné řešení $\{u \cdot (-1)^n + v \cdot 3^n\}_{n=0}^{\infty}$.

2) Teď musíme najít nějaké řešení dané rovnice. Máme štěstí, pravá strana je kvazipolynom, kde $\lambda = 2$ a $P(n) = -9n$ je polynom stupně 1. Protože číslo $\lambda = 2$ není charakteristickým číslem přidružené homogenní rovnice (viz první část), bude $m = 0$ a člen $n^0 = 1$ vlastně z řešení vypadne, to je také příjemné. Vidíme tedy, že určitě existuje nějaké řešení ve tvaru $\{n^0 \cdot Q(n) \cdot 2^n\} = \{Q(n)2^n\}$, kde Q je jistý polynom stupně 1. Takový obecný polynom je dán jako $Q(n) = An + B$, takže Věta vlastně garantuje, že existují nějaké konstanty A, B takové, že $\{(An + B)2^n\}$ je řešením dané rovnice. Tyto neznámé konstanty najdeme tak, že dotyčnou posloupnost prostě dosadíme do dané rovnice a uvidíme, které konstanty povedou k jejímu splnění.

Chceme tedy, aby pro všechna $n \geq 0$ platilo

$$\begin{aligned} a_{n+2} - 2a_{n+1} - 3a_n &= -9n2^n \\ (A(n+2) + B)2^{n+2} - 2(A(n+1) + B)2^{n+1} - 3(An + B)2^n &= -9n2^n \\ (An + 2A + B) \cdot 4 - 2(An + A + B) \cdot 2 - 3(An + B) &= -9n \\ 4An + 8A + 4B - 4An - 4A - 4B - 3An - 3B &= -9n \\ -3An + 4A - 3B &= -9n \\ [-3A]n + [4A - 3B] &= -9n + 0. \end{aligned}$$

Dostali jsme rovnost dvou polynomů a víme, že dva polynomy se rovnají, jen pokud se rovnají jejich koeficienty (to vyplývá z toho, že jednotlivé mocniny jsou jako funkce lineárně nezávislé). Proto máme rovnice $-3A = -9$ a $4A - 3B = 0$, odkud obratem ruky dostaneme $A = 3$, $B = 4$. Právě jsme zjistili, že daná rovnice má řešení $\{(3n + 4)2^n\}$.

Protože díky větě o struktuře řešení víme, že obecné řešení se dá získat jako jedno partikulární plus obecné homogenní, můžeme napsat odpověď: Obecné řešení dané rovnice je $\{(3n + 4)2^n + (-1)^n u + 3^n v\}_{n=0}^{\infty}$.

△

Toto je typický průběh řešení lineární rekurentní rovnice s konstantními koeficienty a kvazipolynomiální pravou stranou. Všimněte si, že charakteristická čísla závisí čistě na levé straně dané rovnice, takže nám ji popisují, říkají nám, jak se chová. Naopak to číslo λ nám popisuje zásadní chování pravé strany (polynom už ji jen modifikuje). Jestliže se λ a charakteristická čísla nepřekrývají, pak to znamená, že se levá a pravá strana neovlivňují a řešení dostaneme v jednodušším tvaru. Pokud by se ale λ rovnala některému charakteristickému číslu, pak to ukazuje, že levá a pravá strana mají něco společného, můžeme si představit, že spolu rezonují, že mezi nimi existuje nějaká vazba. To se pak musí projevit v našem odhadnutém řešení, přidáváme tam násobící faktor n , a to tolikrát, kolik je násobnost λ , čili jak moc se ty dvě strany rovnice ovlivňují. Dá se říci, že to n^m je korekční faktor pro případy vzájemného ovlivnění levé a pravé strany.

! Příklad 10b.f: Najdeme řešení rovnice $a_n = 2a_{n-1} + 3 \cdot 2^n$, $n \geq 1$ s podmínkou $a_0 = 13$.

Nejprve si ji přepíšeme tak, aby neznámé byly na levé straně: $a_n - 2a_{n-1} = 3 \cdot 2^n$. Teď ji přepíšeme do standardního tvaru, tedy zvětšíme všechna n o jedničku: $a_{n+1} - 2a_n = 3 \cdot 2^{n+1}$. Nejmenší a_i zmíněné v původní rovnici pro $n = 1$ bylo a_0 , aby to platilo i pro naší přepsanou rovnici, musíme brát $n \geq 0$.

Pokud dáváte přednost formálnímu přístupu, pak v rovnici $a_n - 2a_{n-1} = 3 \cdot 2^n$ použijte substituci $m = n - 1$, dostanete $n = m + 1$ a tedy $a_{m+1} - 2a_m = 3 \cdot 2^{m+1}$. Tato rovnice má být platná pro $n \geq 1$, tedy pro $m + 1 \geq 1$, což znamená $m \geq 0$.

Každopádně máme lineární rekurentní rovnici s konstantními koeficienty $a_{n+1} - 2a_n = 3 \cdot 2^{n+1}$, $n \geq 0$ a nejprve budeme hledat obecné řešení. Jako obvykle začneme přidruženou homogenní rovnicí.

1) Rovnice $a_{n+1} - 2a_n = 0$ má charakteristický polynom $p(\lambda) = \lambda - 2$ a tudíž charakteristické číslo $\lambda = 2$. Dostáváme obecné „homogenní řešení“ $a_{h,n} = 2^n u$ pro $u \in \mathbb{R}$.

2) Teď potřebujeme najít nějaké řešení rovnice $a_{n+1} - 2a_n = 6 \cdot 2^n$, pravou stranu jsme si přepsali, aby bylo vidět, že je to kvazipolynom. Jeho parametry jsou $\lambda = 2$ a $P(n) = 6$. Hned vidíme, že toto λ se překrývá s charakteristickými čísly levé strany, jmenovitě jednou, proto $m = 1$ a v námi uhodnutém řešení bude muset být opravný člen. Stupeň polynomu P je nula, proto podle Věty musí existovat polynom Q stupně 0 takový, že $\{n^1 Q(n) 2^n\}$ je řešení naší rovnice. Obecný polynom nultého stupně má tvar $Q(n) = A$ pro nějaké $A \in \mathbb{R}$, dostáváme tedy následující uhodnuté řešení: $a_n = An 2^n$.

Zbývá najít A , takže dosadíme toto odhadnuté řešení do rovnice a uvidíme, jaké A bude fungovat:

$$\begin{aligned} a_{n+1} - 2a_n &= 6 \cdot 2^n \\ A(n+1)2^{n+1} - 2 \cdot An2^n &= 6 \cdot 2^n \\ 2A(n+1) - 2An &= 6 \\ 2A = 6 &\implies A = 3. \end{aligned}$$

Máme tedy žádané řešení, je to $a_{p,n} = 3n 2^n$. Když jej přičteme k $a_{h,n}$, dostaneme obecné řešení.

Závěr: Obecné řešení rovnice je $\{3n 2^n + 2^n u\}_{n=0}^{\infty}$ pro $u \in \mathbb{R}$.

Protože jsme přeci jen řešili trochu jinou rovnici než tu původní, raději uděláme zkoušku pro rovnici ze zadání: Nejmenší hodnota n , která se v zadané rovnici vyskytuje, je $n-1$ pro $n=1$, což je $n=0$. Naše posloupnost tedy začíná správným indexem. Teď dosadíme, začneme komplikovanější pravou stranou:

$$\begin{aligned} 2a_{n-1} + 3 \cdot 2^n &= 2(3(n-1)2^{n-1} + 2^{n-1}u) + 3 \cdot 2^n = 3(n-1)2^n + 2^n u + 3 \cdot 2^n \\ &= 3 \cdot (n-1+1)2^n + 2^n u = 3n 2^n + 2^n u = a_n. \end{aligned}$$

Rovnice je splněna.

3) Teď najdeme partikulární řešení, které splňuje danou počáteční podmínu:

$$13 = a_0 = 3 \cdot 0 \cdot 2^0 + 2^0 u = 0 + u \implies u = 13.$$

Zadaná úloha má řešení $\{(3n+13)2^n\}_{n=0}^{\infty}$.

Poznámka: Co by se stalo, kdybychom zapomněli na opravný faktor? Mysleli bychom si naivně, že existuje řešení ve tvaru $a_n = A \cdot 2^n$. Po dosazení do rovnice bychom dostali

$$\begin{aligned} a_{n+1} - 2a_n &= 6 \cdot 2^n \\ A2^{n+1} - 2 \cdot A2^n &= 6 \cdot 2^n \\ 2A - 2A &= 6 \\ 0 &= 6. \end{aligned}$$

A máme smůlu. Takže ty opravné faktory opravdu k něčemu jsou.

△

Shrneme si postup.

S Algoritmus 10b.8. pro nalezení řešení rovnice $a_{n+k} + c_{k-1}a_{n+k-1} + \cdots + c_1a_{n+1} + c_0a_n = b_n$, $n \geq n_0$, kde $b_n = P(n)\lambda^n$, $c_i \in \mathbb{R}$ a $c_0 \neq 0$ (tedy řád k).

1. Nejprve řešete přidruženou homogenní rovnici $a_{n+k} + c_{k-1}a_{n+k-1} + \cdots + c_1a_{n+1} + c_0a_n = 0$.
- a) Najděte všechna charakteristická čísla λ_j s násobnostmi m_j řešením rovnice $\lambda_k + c_{k-1}\lambda^{k-1} + \cdots + c_1\lambda + c_0 = 0$.
- b) Podle Algoritmu sestavte bázi prostoru řešení $B = \{\{a_{i,n}\}_{n=n_0}^{\infty}; i = 1, \dots, k\}$.
- c) Obecné řešení přidružené homogenní rovnice je $\{a_{h,n}\} = \left\{ \sum_{i=1}^k u_i a_{i,n} \right\}$ pro $u_i \in \mathbb{R}$.

Pokud byla zadaná rovnice již homogenní, jděte na 3.

2. Pokud nebyla zadaná rovnice homogenní, zkontrolujte, že je pravá strana kvazipolynom, tedy $b_n = P(n)\lambda^n$ pro nějaké $\lambda \in \mathbb{R}$ a polynom P .

- a) Porovnejte λ s charakteristickými čísly λ_j z kroku 1. Pokud se žádnému nerovná, položte $m = 0$. Pokud pro nějaké j platí $\lambda = \lambda_j$, položte $m = m_j$ (násobnost dotyčného charakteristického čísla).
- b) Sestavte obecný polynom Q stupně stejněho jako P , tradičně se používá $Q(n) = A + Bn + \cdots$.
- c) Uhádněte řešení $a_n = n^m Q(n)\lambda^n$. Dosadte jej do dané rovnice a po zkrácení λ zjednodušte levou stranu do tvaru polynomu. Porovnáním koeficientů polynomů na levé a pravé straně získáte kolik je neznámých koeficientů v Q .
- d) Vyřešte tyto rovnice a obdržené konstanty dosadte zpět do Q . Získáte jedno konkrétní řešení $a_{p,n}$.
- e) Obecné řešení dané úlohy je $\left\{ a_{p,n} + \sum_{i=1}^k u_i a_{i,n} \right\}_{n=n_0}^{\infty}$ či $a_n = a_{p,n} + \sum_{i=1}^k u_i a_{i,n}$ pro $n \geq n_0$.

3. Pokud byly s rovnicí zadány také počáteční podmínky, dosaďte za a_j v těchto podmírkách vzorce pro a_j z obecného řešení, které jste našli. Získáte k rovnic pro k neznámých u_1, \dots, u_k . Vyřešte tuto soustavu, získaná u_i dosaďte do vzorce pro obecné řešení a dostanete tak partikulární řešení pro zadанou úlohu.

△

S Mechanismus vzniku odhadu je vlastně velice snadný. Geometrický faktor λ^n opíšeme, polynom také, jen jej změníme na obecný. V případě shody λ na pravé a levé straně ještě přidáme korekční faktor. Vyzkoušíme to v následující tabulce. Ve sloupcích jsou levé strany rovnic a v řádcích napravo jsou pravé strany. Například druhé pole zleva v prvním řádku odpovídá rovnici $a_{n+2} - 3a_{n+1} + 2a_n = n2^n$, vyplnili jsme příslušný odhad.

$a_{n+2} - 9a_n =$ [$\lambda = -3, 3$]	$a_{n+2} - 3a_{n+1} + 2a_n =$ [$\lambda = 1, 2$]	$a_{n+2} - 4a_{n+1} + 4a_n =$ [$\lambda = 2$ (2×)]	$L = / = b_n$
$(An + B)2^n$	$n(An + B)2^n$	$n^2(An + B)2^n$	$= n2^n$ [$\lambda = 2$]
$(An^2 + Bn + C)(-1)^n$	$(An^2 + Bn + C)(-1)^n$	$(An^2 + Bn + C)(-1)^n$	$= n^2(-1)^n$ [$\lambda = -1$]
$An + B$	$n(An + B)$	$An + B$	$= 2n - 5$ [$\lambda = 1$]
$nA(-3)^n$	$A(-3)^n$	$A(-3)^n$	$= (-3)^n$ [$\lambda = -3$]

Je dobré se na to podívat po řádcích. V prvním řádku krásně vidíme, jak se vždy v odhadech řešení zachovává geometrická posloupnost a zobecněný polynom z pravé strany, někdy pak přidáváme opravný faktor podle toho, kolikrát se λ napravo najde také jako charakteristické číslo levé strany. Zajímavý je třetí řádek, tam žádný geometrický člen nebyl a toto se také v odhadech zachovalo, při porovnávání je pak třeba použít $\lambda = 1$, protože pravou stranu lze přepsat jako $(2n - 5) \cdot 1^n$. Podobně v posledním řádku máme vlastně $1 \cdot (-3)^n$, tudiž je tam jakoby polynom stupně nula, což se zobecní jako konstanta.

Ukážeme teď jeden zajímavý příklad a pak si zkusíme působnost tohoto algoritmu trochu rozšířit.

! Příklad 10b.g: Zde se podíváme, kolik operací „stojí“ výpočet determinantu matice o rozměru $n \times n$.

a) Podle definice se mají sčítat součiny typu $a_{1\pi(1)} \cdot a_{2\pi(2)} \cdots a_{n\pi(n)}$, kde suma jde přes všechny permutace π množiny $\{1, 2, \dots, n\}$. Těchto permutací je $n!$ a jeden takový součin stojí $n - 1$ násobení plus jedno přičtení k ostatním, tedy výpočet determinantu podle definice vyžaduje celkem $n \cdot n!$ operací, a to jsme ještě nezapočítali vytváření těch premutací. Jde tedy o zcela neperspektivní způsob, jakmile se mohou objevit trochu větší maticy.

My takto počítáme determinanty 2×2 a 3×3 tužkou na papír, kdy to ještě jde. Pro trochu větší matice se s oblibou používá rozvoj podle řádku či sloupce, který je případně opakován, dokud se nedojde k malým determinantům. Byla by tato metoda dobrým východiskem pro algoritmus?

b) Nechť a_n je náročnost počítání determinantu rozvojem podle prvního sloupce, je-li velikost matice $n \times n$. K jeho spočítání si nejprve musíme spočítat n doplňkových determinantů, které mají o jedno menší velikost, takže na to potřebujeme $n \cdot a_{n-1}$ operací. Ty pak potřebujeme střídavě přičítat a odčítat, dostaneme $a_n = n \cdot a_{n-1} + n$ operací. Pro matici 1×1 samozřejmě potřebujeme $a_1 = 1$, což je počáteční podmínka.

Když to přepíšeme, máme $a_{n+1} - (n+1)a_n = n+1$, což je sice krásná lineární rekurentní rovnice, ale bohužel nemá konstantní koeficienty, takže jsme se ji vyřešit nenaučili. To ale nakonec nebude tak velký problém, my teď totiž ukážeme, že i tento algoritmus je natolik náročný, že nás nebude zajímat.

Dokážeme indukcí, že každé řešení a_n této rovnice splňuje $a_n \geq n!$.

(0) $a_1 = 1 = 1!$, v pořádku.

(1) Předpokládejme, že $a_n \geq n!$. Pak $a_{n+1} = (n+1)a_n + n+1 \geq (n+1)n! + n+1 \geq (n+1)!$.

Oproti výpočtu podle definice jsme si tedy moc nepolepšili a jako obecná metoda to není vhodný nápad, nicméně stojí za zmínu, že pokud dopředu víme, že hodně členů matice je nulových, pak se náročnost prudce snižuje a rozvoj podle sloupce/řádku se stává zajímavou volbou, viz cvičení.

c) Zkusíme teď jiný postup, uděláme zase rozvoj podle prvního sloupce, ale tentokrát si nejdříve vyrobíme nuly pod levým horním rohem, abychom pak nemuseli počítat tolik subdeterminantů. Kolik operací bude stát takovýto výpočet determinantu matice $n \times n$, označme to d_n ?

Zkusme totálně pesimistickou variantu. Vlevo nahoře může být nula. Pak nejprve projdeme ostatní členy v prvním sloupci, abychom našli něco nenulového (pesimisticky n operací, bude to až dole), pak prohodíme první a poslední řádek (n operací, musíme prohodit všechny členy v řádcích) a někde si zapamatujeme, že tím pádem počítáme míinus ten determinant. Příprava prvního řádku nás tedy pesimisticky stála $2n$ operací. No a pak odečítáme

příslušný násobek prvního řádku od ostatních, každý řádek nás stojí 2 operace na člen (násobení, odečtení), tedy zase $2n$ operací na řádek, musíme jich takto upravit $n - 1$. Sečteno: vyžaduje to $2n + 2n(n - 1) = 2n^2$ operací.

Teď je tedy matice připravena a rozvíjíme podle členu vlevo nahoře, takže musíme spočítat jeden determinant matice o řád menší a vynásobit jej tím levým horním členem. Dostáváme tak rovnici $d_n = d_{n-1} + 2n^2 + 1$. Opět si ji přepíšeme: Chceme vyřešit $d_{n+1} - d_n = 2(n+1)^2 + 1$ s počáteční podmínkou $d_1 = 1$.

Toto už je lineární rekurentní rovnice s konstantními koeficienty, takže nasadíme standardní metody. Nejprve homogenní rovnice: $d_{n+1} - d_n = 0$ dává charakteristický polynom $\lambda - 1$ a tudíž vychází charakteristické číslo $\lambda = 1$. Obecné řešení této homogenní rovnice je konstantní posloupnost $\{u \cdot 1^n\}_{n=1}^{\infty} = \{u\}_{n=1}^{\infty}$.

Teď potřebujeme partikulární řešení, přepis $d_{n+1} - d_n = 2n^2 + 4n + 3 = (2n^2 + 4n + 3)1^n$ ukazuje, že máme kvazipolynomiální pravou stranu, která má $\lambda = 1$, což se shoduje s charakteristickým číslem (jednoduchým) levé strany. V námi odhadnutém řešení tedy bude muset být opravný člen n^1 a také obecný polynom Q stupně 2. Odhadneme proto řešení ve tvaru $d_n = n^1(An^2 + Bn + C)1^n = An^3 + Bn^2 + Cn$. Dosadíme jej do dané rovnice a levou stranu pak upravíme na polynom:

$$\begin{aligned} d_{n+1} - d_n &= 2n^2 + 4n + 3 \\ (A(n+1)^3 + B(n+1)^2 + C(n+1)) - (An^3 + Bn^2 + Cn) &= 2n^2 + 4n + 3 \\ An^3 + A3n^2 + A3n + A + Bn^2 + B2n + B + Cn + C - An^3 - Bn^2 - Cn &= 2n^2 + 4n + 3 \\ (3A)n^2 + (3A + 2B)n + (A + B + C) &= 2n^2 + 4n + 3. \end{aligned}$$

Porovnáním koeficientů u stejných mocnin na levé a pravé straně dostáváme soustavu rovnic $3A = 2$, $3A + 2B = 4$, $A + B + C = 3$, z ní hravě odvodíme $A = \frac{2}{3}$, $B = 1$, $C = \frac{4}{3}$.

Dostáváme tedy partikulární řešení $\{\frac{2}{3}n^3 + n^2 + \frac{4}{3}n\}$ a obecné řešení $\{\frac{2}{3}n^3 + n^2 + \frac{4}{3}n + u\}$. My ovšem hledáme řešení splňující počáteční podmítku $d_1 = 1$, takže chceme $\frac{2}{3} + 1 + \frac{4}{3} + u = 1$, což dává $u = -2$.

Závěr: Daná rovnice má řešení $\{\frac{2}{3}n^3 + n^2 + \frac{1}{3}n - 2\}_{n=1}^{\infty}$.

Popravdě řečeno, většinu práce jsme si mohli odpustit. Nás u algoritmů zajímá jejich asymptotická rychlosť, přičemž vidíme, že homogenní řešení jsou konstantní posloupnosti, takže růst řešení určí spíš ten odhad. V něm byl dominantním členem n^3 , to už jsme viděli i předtím, než jsme konstanty našli (protože A v takových případech vyjde nenulové). Nemuseli jsme tedy konstanty nacházet, už v okamžiku odhadu bylo vidět, že studovaný algoritmus potřebuje řádově n^3 operací neboli $d_n = \Theta(n^3)$.

Když si představíte, jak tento algoritmus funguje rekurzivně přes celou matici, tak byste měli dospět k názoru, že to vlastně není nic jiného, než že matici upravíme na trojúhelníkový tvar a pak vynásobíme členy na diagonále. Je to proto úzce svázáno s Gaussovou eliminací a o té je známo, že potřebuje řádově n^3 operací. Náš výpočet tedy dal správnou odpověď. Náročnost n^3 je upřímně řečeno také dost drsná, ale oproti faktoriálu je to totální revoluce.

△

Existují pravé strany, které jsou velice blízké kvazipolynomům. Například $b_n = n \cdot 3^n + 7 \cdot 2^n$ vypadá hodně jako kvazipolynom, ale tento výraz nelze sjednotit pomocí jediného λ^n , je jich tam víc. Zde pomůže princip, který je velice obecný a zase jej najdeme u všech typů lineárních rovnic.

!

Věta 10b.9. (o superpozici)

Nechť $k \in \mathbb{N}$, uvažujme funkce $c_0(n), c_1(n), \dots, c_{k-1}(n): \mathbb{Z} \mapsto \mathbb{R}$.

Jestliže posloupnost $\{a_n\}_{n=n_0}^{\infty}$ řeší rovnici $a_{n+k} + \sum_{i=0}^{k-1} c_i(n)a_{n+i} = b_n$, $n \geq n_0$

a posloupnost $\{\tilde{a}_n\}_{n=n_0}^{\infty}$ řeší rovnici $a_{n+k} + \sum_{i=0}^{k-1} c_i(n)a_{n+i} = \tilde{b}_n$, $n \geq n_0$,

pak posloupnost $\{a_n + \tilde{a}_n\}_{n=n_0}^{\infty}$ řeší rovnici $a_{n+k} + \sum_{i=0}^{k-1} c_i(n)a_{n+i} = b_n + \tilde{b}_n$, $n \geq n_0$.

Důkaz je snadný a necháme jej jako cvičení.

Indukcí se to dá snadno rozšířit na libovolný (konečný) počet sčítanců na pravé straně. Věty o superpozici jsou jakýmsi evergreenem všech oblastí matematiky zabývajícími se lineárními rovnicemi a umožňují nám rozdělit nepříjemnou pravou stranu na příjemnější sčítance, rovnice pak řešíme pro každý z nich zvlášť.

V našem případě jde o součet kvazipolynomů. Pak nemusíme se sčítáním jednotlivých částečných řešení čekat až na konec postupu, ale lze dávat dohromady již jednotlivé odhady. Ukážeme si to na příkladě.

! **Příklad 10b.h:** Vyřešíme rovnici $a_{n+2} - a_n = 3 \cdot 2^n - 12$, $n \geq 0$ s počátečními podmínkami $a_0 = 1$, $a_1 = 0$.

Řešení: Jde o lineární rekurentní rovnici s konstantními koeficienty, takže aplikujeme standardní algoritmus. Nejprve řešíme homogenní rovnici: $a_{n+2} - a_n = 0$ dává charakteristickou rovnici $\lambda^2 - 1 = 0$, jsou tedy jednoduchá charakteristická čísla $\lambda = \pm 1$ a obecné řešení $\{u \cdot (-1)^n + v \cdot 1^n\}_{n=0}^{\infty} = \{(-1)^n u + v\}_{n=0}^{\infty}$.

Ted potřebujeme nějaké partikulární řešení. Bohužel pravá strana není kvazipolynomiální, protože se nedá napsat jako polynom krát nějaké λ^n . Je to ale součet dvou kvazipolynomů, $3 \cdot 2^n$ a $-12 \cdot 1^n$. Pro každý z nich teď odhadneme řešení.

Kvazipolynom $3 \cdot 2^n$ má polynom $P(n) = 3$ stupně 0, budeme tedy potřebovat $Q(n) = A$, a speciální číslo $\lambda = 2$, které nepatří mezi charakteristická čísla levé strany, takže není třeba dělat korekci, $m = 0$. Proto uhodneme řešení $A2^n$.

Kvazipolynom $-12 \cdot 1^n$ má polynom $P(n) = -12$ stupně 0, budeme potřebovat $Q(n) = B$ (zde jsme použili jiné písmeno, abychom mohli vznikající odhad přičítat k předchozímu), a speciální číslo $\lambda = 1$, které je charakteristickým číslem levé strany, a to jednou, bude tedy korekce s $m = 1$. Proto uhodneme řešení $n^1 \cdot B \cdot 1^n = Bn$.

Pro celou danou pravou stranu tak dostaváme řešení ve tvaru $a_n = A2^n + Bn$. Konstanty A, B určíme dosazením našeho odhadu do dané rovnice:

$$\begin{aligned} a_{n+2} - a_n &= 3 \cdot 2^n - 12 \\ (A2^{n+2} + B(n+2)) - (A2^n + Bn) &= 3 \cdot 2^n - 12 \\ 4A2^n + Bn + 2B - A2^n - Bn &= 3 \cdot 2^n - 12 \\ 3A2^n + 2B &= 3 \cdot 2^n - 12. \end{aligned}$$

Na rozdíl od předchozích příkladů se nám teď nepodařilo zkrátit λ^n , což není překvapení, protože vlastně máme v úloze dvě různé lambdy. Není to ale žádný problém, mocniny 2^n a 1^n jsou lineárně nezávislé, takže jen zobecníme předchozí metodu. Budeme navzájem porovnávat koeficienty na levé a pravé straně u stejných mocnin n^i a λ^n . Vidíme tam mocninu $n^0 2^n$, na levé straně je u ní $3A$ a na pravé straně je u ní 3 . Proto $A = 1$. Podobně vidíme na levé straně u mocniny $n^0 1^n$ koeficient $2B$ a na pravé straně -12 , proto $B = -6$. Tyto konstanty dosadíme do našeho odhadnutého řešení a nalezli jsme partikulární řešení $\{2^n - 6n\}_{n=0}^{\infty}$.

Ted použijeme větu o struktuře řešení, obecné řešení dané rovnice je $\{2^n - 6n + (-1)^n u + v\}_{n=0}^{\infty}$.

Zbývá zpracovat počáteční podmínky:

$$\begin{aligned} a_0 &= 1 + u + v = 1 & u + v &= 0 \\ a_1 &= 2 - 6 - u + v = 0 & -u + v &= 4 \end{aligned} \implies \begin{aligned} u &= -2, v = 2. \end{aligned}$$

Řešení úlohy je $\{2^n - 6n - 2 \cdot (-1)^n + 2\}_{n=0}^{\infty}$.

△

Všimněte si, že větu jsme formulovali pro všechny lineární rovnice, nejen ty s konstantními koeficienty, takže jsme ji mohli zařadit do kapitoly. Dokonce by trochu pomohla, protože jsme při důkazu Věty používali speciální případ superpozice, kdy je jedna pravá strana nulová. Čtenář by ale asi nevěděl, odkud se tam bere a co s ní, tady je na správném místě.

Příklad 10b.i (pokračování): Vraťme se ke králíkům. Odhadli jsme, že poté, co se králíci deset let nerušeně množili, bylo jich v roce 1869 cca $24 \cdot 10^6$. Podíváme se na dva scénáře. Označme jako a_n počet králíků n let po roce 1869, a to v milionech (ušetříme šest nul).

a) Pokud by se králíci nerušeně množili dál, pak by jejich populace zase splňovala rovnici $a_{n+1} = 4a_n$ neboli $a_{n+1} - 4a_n = 0$ a počáteční podmínsku $a_0 = 24$.

Tato homogenní rovnice má charakteristické číslo $\lambda = 4$ a obecné řešení $a_n = u \cdot 4^n$, počáteční podmínka pak dává $a_n = 24 \cdot 4^n$.

b) Druhý scénář využívá informace, že se střílelo $2 \cdot 10^6$ králíků ročně. Můžeme si to interpretovat tak, že je vystřílíme na konci roku, což dává upravený vztah $a_{n+1} = 4a_n - 2$, tedy $a_{n+1} - 4a_n = -2$, zase s počáteční podmínkou $a_0 = 24$.

Toto je standardní lineární rekurentní rovnice, u níž nejprve najdeme řešení přidružené homogenní úlohy. To už jsme udělali ve scénáři a), dostali jsme $a_{h,n} = 4^n u$.

Ted potřebujeme nějaké partikulární řešení. Pravá strana $-2 = -2 \cdot 1^n$ je kvazipolynom, který má polynom $P(n) = -2$ stupně 0, proto použijeme $Q(n) = A$, a číslo $\lambda = 1$, které není stejné jako charakteristické číslo levé strany. Proto $m = 0$ a korekce nebude třeba. Dostaváme odhadnuté řešení $a_n = A1^n = A$. Správné A najdeme dosazením a_n do dané rovnice:

$$\begin{aligned} a_{n+1} - 4a_n &= -2 \\ A - 4A &= -2 \\ -3A &= -2 \implies A = \frac{2}{3}. \end{aligned}$$

Máme tedy obecné řešení naší rovnice ve tvaru $a_n = \frac{2}{3} + 4^n u$. Když použijeme počáteční podmínu $a_0 = 24$, dostaneme $\frac{2}{3} + u = 24$, tedy $u = 23 + \frac{1}{3}$.

Druhý scénář dává růst populace $a_n = \frac{2}{3} + (23 + \frac{1}{3})4^n$. Vidíme, že populace roste pomaleji, ale pořád rychlostí geometrické posloupnosti 4^n , takže k nějaké zásadnější změně nedošlo.

Zajímavá úloha: Zkuste vyřešit stejnou úlohu, ale teď s úbytkem L , tedy vyřešte rovnici $a_{n+1} - 4a_n = -L$ s podmínkou $a_0 = 24$. Stejným postupem dostanete vzorec, ve kterém se objeví parametr L , a položte si otázku, jak velké by L muselo být, aby již a_n nerostlo rychlostí 4^n (nápočeda: vzorec vyjde $a_n = \frac{L}{3} + (24 - \frac{L}{3})4^n$). Je vidět, že byste museli na konci každého roku vystřílet 72 milionů králíků. Možná vás napadne, že chyba je v tom, že je střílíme až na konci roku, když se rozmnoží. Bohužel si člověk nepomůže ani střílením třeba hned na začátku roku. Pak je správná rovnice dána $a_{n+1} = 4(a_n - L) = 4a_n - 4L$, je to tedy vlastně původní příklad, jen se efekt L čtyřikrát zvětší. Na zabránění geometrického růstu by bylo proto třeba střílet minimálně $\frac{72}{4} = 18$ miliónů ročně, což je pořád nerealistické.

Mnohem perspektivnější je změnit tu čtyřku neboli rozmnožovací konstantu, což je v zásadě základ moderních metod boje proti škůdcům.

Poznámka: Problém s tím, kdy králíky střílíme, se dá řešit tak, že pracujeme s menšími časovými intervaly. Můžeme například uvažovat stav po týdnech b_n , dostáváme pak rovnici $b_{n+1} = Kb_n - L$, kde L je hodnota týdenního odstřelu a K je konstanta přirozené množivosti, která samozřejmě souvisí s tou čtyřkou za rok. Pokud bychom nestříleli, bylo by za 52 týdnů K^{52} krát více králíků, proto musí platit $K^{52} = 4$. Odtud již K snadno určíme, rovnici lze vyřešit a roční stavy dostaneme jednoduchým převodem $a_n = b_{52n}$. Tento model je již mnohem realističtější. Lze dokonce počítat i stav po dnech, pak už je ale lepší přejít rovnou na diferenciální rovnice, kdy měříme čas na reálné ose, čímž se dostáváme mimo oblast diskrétní matematiky, toto je hájemství matematiky spojité neboli matematické analýzy.

△

! 10b.10 Poznámka: Onen první model neomezeného růstu králíků daný vztahem $a_{n+1} = Ka_n$ patří přes svou jednoduchost k jedněm z nejzajímavějších a nejpoužívanějších modelů, říká se mu exponenciální model, protože vede na řešení $c \cdot K^n = c \cdot e^{\ln(K)n}$. I když často vykazuje jen přibližnou shodu s reálným jevem, lze jej použít k rychlému orientačnímu náhledu, což bývá velice užitečné. Některé populární situace:

- Máte na účtu $a_0 = C$ peněz, každý rok se vám to zúročí hodnotou r procent. Pak vývoj financí popisuje (dokonce přesně) vztah $a_{n+1} = (1 + \frac{r}{100})a_n$.
- Máte-li na začátku kolonii $a_0 = C$ bakterií v Petriho misce, pak jejich počet po jednotlivých hodinách je dán vztahem $a_{n+1} = Ka_n$. Výsledná exponenciála vykazuje dobrou shodu s realitou, dokud není počet natolik velký, že baktérie začnou soupeřit o zdroje živin.
- Totéž lze říci o libovolné populaci, která se volně množí, neužírají z ní predátoři a má dostatek zdrojů. I lidstvo se tak jeden čas porůznu množilo.
- Když se rozpadne atom uranu 235, vypustí dva neutrony. Když neutron vhodně vletí do dalšího atomu uranu, ten se rozpadne a vypustí dva neutrony. Počet rozpadlých jader po n krocích je dán $a_{n+1} = 2a_n$, což vede na 2^n a říká se tomu jaderný výbuch. Ve skutečnosti je počet vypuštěných neutronů proměnný, udává se v průměru okolo 2.5, ale ten násobící faktor v rovnici bývá menší, protože určité procento neutronů se netrefí do dalšího atomu a uletí bez užitku pryč. V jaderných elektrárnách se toto snaží podporovat, dokonce neutrony i lapají. Tento příklad je samozřejmě zjednodušený, ale vystihuje podstatu.
- Radioaktivní prvky fungují tak, že se jednotlivé atomy neustále náhodně rozhodují, zda se rozpadnou nebo ne. Jestliže máme a_n atomů, tak se z nich během sekundy určité procento p rozpadne, takže pak už jich máme $a_{n+1} = (1 - \frac{p}{100})a_n$. Vede to na řešení $a_n = c(1 - \frac{p}{100})^n$, které platí i pro hmotnost či hustotu zastoupení v jiném materiálu a velice dobře odpovídá reálnému stavu (většinou se ale zase pracuje se spojitým časem a vyjde exponenciála $c e^{\lambda t}$). Na tomto vzorci je mimo jiné založena metoda měření stáří organických látek ve vykopávkách.

△

! Příklad 10b.j: V kapitole o indukci jsme si dokazovali rozličné součtové vzorce (viz např. cvičení), ale poznamenali jsme, že na to, abychom dokázali indukci jejich správnost, je musíme odněkud získat. Jedna možnost byla naznačena v důkazu Věty , zde si ukážeme další.

Najdeme vzorec pro $\sum_{k=1}^n k = 1 + 2 + \dots + n$.

Hledanou hodnotu si označíme $s_n = \sum_{k=1}^n k$. Pak platí rekurentní vztah $s_{n+1} = s_n + (n+1)$, máme tedy rovnici $s_{n+1} - s_n = n + 1$, evidentně platí počáteční podmínka $s_1 = 1$.

Tuto úlohu vyřešíme standardním způsobem. Nejprve najdeme obecné řešení přidružené homogenní rovnice $s_{n+1} - s_n = 0$. Charakteristická rovnice $\lambda - 1 = 0$ dává charakteristické číslo $\lambda = 1$ a konstantní posloupnost $\{u \cdot 1^n\}_{n=1}^{\infty} = \{u\}_{n=1}^{\infty}$ jako obecné řešení homogenní rovnice.

Teď najdeme jedno partikulární řešení. Pravá strana je kvazipolynomiální, $n + 1 = (n + 1)1^n$, proto lze použít metodu odhadu. Zde je speciální číslo $\lambda = 1$ také charakteristickým číslem (jednoduchým), máme tedy korekční faktor $m = 1$ a odhadneme řešení $s_n = n^1(A_n + B)1^n = An^2 + Bn$. Dosadíme do řešené rovnice, upravíme na polynom a porovnáme koeficienty:

$$\begin{aligned} s_{n+1} - s_n &= n + 1 \\ (A(n+1)^2 + B(n+1)) - (An^2 + Bn) &= n + 1 \\ 2An + (A + B) &= n + 1 \implies 2A = 1, A + B = 1 \implies A = B = \frac{1}{2}. \end{aligned}$$

Máme tedy partikulární řešení $\{\frac{1}{2}n^2 + \frac{1}{2}n\}$ a obecné řešení $\{\frac{1}{2}n^2 + \frac{1}{2}n + u\}_{n=1}^{\infty}$. Teď ještě najdeme hodnotu u tak, aby tato posloupnost splňovala počáteční podmínu:

$$s_1 = \frac{1}{2} + \frac{1}{2} + u = 1 \implies u = 0.$$

Vychází $s_n = \frac{1}{2}n^2 + \frac{1}{2}n = \frac{n(n+1)}{2}$, což je onen známý vzorec.

△

Příklad 10b.k: Máme k dispozici dva různé signály lišící se délou (pípnutí, třeba jako u Morseovky), jeden trvá 1 ms, druhý 2 ms. Kolik různých zpráv se dá poslat za n ms?

Toto je kombinatorická úloha, kterými se zabýváme v kapitole , je to tedy jakási upoutávka na to, co přijde, a zároveň připomínka, že v dotyčné kapitole čtenář najde další příklady s rekurentními vztahy.

Označme jako a_n počet různých zpráv, které lze vyslat za n milisekund. Uvažujme množinu všech takových zpráv. Tato množina se dá rozdělit na dvě části podle toho, jestli ve zprávě je jako poslední krátký či dlouhý signál. Počet zpráv, které končí krátkým signálem, je stejný, jako počet zpráv o délce $n - 1$. Počet zpráv, které končí dlouhým signálem, je stejný jako počet zpráv o délce $n - 2$. Dostáváme tedy rovnici $a_n = a_{n-1} + a_{n-2}$ pro $n \geq 2$. Je to zase Fibonacciho vztah!

Jaké jsou počáteční podmínky? Je jen jedna zpráva dlouhá 1 ms, $a_1 = 1$, zato do dvou ms se vejde buď jeden dlouhý signál nebo dva krátké, tedy dvě různé zprávy, $a_2 = 2$.

Když si rovnici upravíme, $a_{n+2} - a_{n+1} - a_n = 0$, dostáváme homogenní lineární rovnici 2. řádu s konstantními koeficienty, kterou řešíme jako v příkladě . Dostáváme obecné řešení $a_n = u\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} + v\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}$.

Počáteční podmínky nám pak dají hledané partikulární řešení $a_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}$. Je to vlastně posunutá Fibonacciho posloupnost, $a_n = F_{n+1}$. Pro ilustraci, je možné vyslat třeba 89 různých zpráv během 10 ms a 987 různých zpráv během 15 ms.

△

Příklad 10b.l: Zde si přiblížíme tzv. Josefův problém: Celkem n lidí stojí v kruhu a postupně jsou vyřazováni takto: Jeden se určí jako první, od něj se pak odpočítává a každý $k-tý člověk vypadává z kruhu. Počítá se pořad dokola stále se zmenšujícího kruhu, dokud nezůstane jeden člověk. Otázka zní: Na jakém místě (v původním rozestavení) stál ten, kdo nakonec zůstal?$

O jméno tohoto problému se postaral historik Flavius Josephus. Jednou se prý údajně sešlo v jeskyni 41 poražených vzbouřenců, kteří chtěli tímto způsobem spáchat postupně sebevraždu metodou $k = 3$. Mezi nimi byl i Flavius se svým otrokem, ale z nějakého důvodu se neholali přidat k ostatním s tou sebevraždou. Vybrali si proto chytře místa tak, aby zůstali jako poslední dva, a pak se nenápadně vytratili. O věrohodnosti této příhody lze oprávněně pochybovat, však to byl historik.

My se podíváme na verzi s $k = 2$. Máme n lidí v kruhu očíslovaných od 1 do n , rozpočítávají se první druhý a každý se sudým číslem vypadává ze hry, dokud se nedojede na konec kruhu, pak to začne být zajímavé. Protože se to bude hodit, zkuste si to rozmyslet. Začátek kruhu po prvním průchodu vypadá jako 1, 3, 5, 7, ... Pokud je n sudé, tak jako poslední vypadl pán číslo n a po něm vypadne pán číslo 3. Pokud je liché, tak poslední byl na řadě pán $n - 1$, pán n (zatím) zůstává a jako další vypadává 1.

Označíme jako $J(n)$ pořadové číslo toho, kdo nakonec zůstane jako jediný, pokud se začne s n lidmi. Předchozí rozbor nabízí možnost vytvořit rekurentní vztah.

Jestliže byl na začátku sudý počet $n = 2m$, pak se po prvním průchodu začne vybírat z lidí 1, 3, 5, ..., $(2m - 1)$, kterých je m , zase metodou každý druhý. Při tom se ale počítá vzhledem k novému pořadí K . V jakém vztahu je toto k původnímu pořadí k ? Ten, kdo je nyní druhý, měl původně číslo 3. Nově třetí měl 5, nově čtvrtý měl 7. Vidíme, že vzorec je $k = 2K - 1$. Víme vše, co potřebujeme k následující úvaze. Pokud víme, že v probraném kruhu

o m lidech nakonec zůstane člověk na pozici K , tedy $J(m) = K$, pak tento člověk zůstane i z původních $n = 2m$ lidí a měl původně pozici $J(2m) = 2K - 1$. Dostáváme tedy pro sudý počet lidí rekurentní vztah $J(2m) = 2J(m) - 1$.

Pokud jsme začali s lichým počtem lidí $n = 2m + 1$, pak po prvním kole zbydou lidé 1, 3, 5, 7, ..., $(2m + 1)$, kterých je $m + 1$. Další vypadne jednička, máme teď m lidí 3, 5, 7, ..., $(2m + 1)$ a z nich se vybírá metodou každý druhý. Opět potřebujeme najít převodní vztah. Nový první měl číslo 3, nový druhý měl číslo 5 atd., dostáváme $k = 2K + 1$. Jestliže tedy z nového polovičního kruhu nakonec zůstane člověk na nové pozici $K = J(m)$, pak to je podle původní pozice člověk $J(2m + 1) = 2K + 1$. Pro lichý počet lidí proto dostáváme rekurentní vztah $J(2m + 1) = 2J(m) + 1$. Je to nepříjemné, nemáme jednotnou rovnici pro všechna čísla.

Takovéto situace jsme se řešit nenaučili, nezbývá, než to nějak odhadnout. Jakou hodnotu má J pro malá n ? $J(1) = 1$, $J(2) = 1$, $J(3) = 3$, $J(4) = 1$, $J(5) = 3$, $J(6) = 5$, $J(7) = 7$, $J(8) = 1$, $J(9) = 3$, $J(10) = 5$, $J(11) = 7$, $J(12) = 9$, $J(13) = 11$, $J(14) = 13$, $J(15) = 15$, $J(16) = 1$.

Vidíme, že když je n ve tvaru 2^a , pak je $J(n) = 1$, pak se přidává po dvou. Hypotéza: Jestliže $n = 2^a + b$, kde $b < 2^a$ a $b \in \mathbb{N}_0$, pak $J(n) = 1 + 2b$. Dokážeme to silnou indukcí.

První počáteční hodnoty se shodují. Teď uděláme indukční krok.

Uvažujme tedy nějaké $n = 2^a + b$, kde $b < 2^a$ a $b \in \mathbb{N}_0$. Předpokládáme, že vzorec pro J z naší hypotézy je platný pro všechna menší čísla.

Jestliže je n sudé, tedy $n = 2m = 2^a + b$, pak musí být b sudé a $m = 2^{a-1} + \frac{1}{2}b$, kde $\frac{1}{2}b \in \mathbb{N}_0$ a $\frac{1}{2}b < 2^{a-1}$. Můžeme použít indukční předpoklad, $J(2^{a-1} + \frac{1}{2}b) = 2 \cdot \frac{1}{2}b + 1 = b + 1$ a tedy

$$J(n) = J(2m) = 2J(m) - 1 = 2(b + 1) - 1 = 2b + 1,$$

souhlasí.

Jestliže je n liché, tedy $n = 2m + 1 = 2^a + b$, pak musí být b liché a $m = 2^{a-1} + \frac{1}{2}(b - 1)$, kde $\frac{1}{2}(b - 1) \in \mathbb{N}_0$ a $\frac{1}{2}(b - 1) < 2^{a-1}$. Můžeme použít indukční předpoklad, $J(2^{a-1} + \frac{1}{2}(b - 1)) = 2 \cdot \frac{1}{2}(b - 1) + 1 = b$ a tedy

$$J(n) = J(2m + 1) = 2J(m) + 1 = 2b + 1,$$

souhlasí. Náš vzorec je tedy správně.

△

Tím končí kapitola o lineárních rekurentních rovnicích. Nakonec ještě jedna nepoviná poznámka.

Týká se zásadního problému, že naše krásné metody jsou k ničemu, pokud pravá strana není kvazipolynom, a dokonce ani homogenní rovnici neumíme vyřešit, pokud nejsou koeficienty rovnice konstantní. Touto problematikou se zabývají pokročilé teorie, my si zde ukážeme dva případy, kdy se dají pomocí chytrého převodu použít naše metody.

a) Pokud se v rekurentním vztahu jen násobí/dělí, pak se dá rovnice převést na lineární pomocí logaritmu. Ukážeme to na příkladě.

Máme rovnici $a_{n+1} = \frac{a_n^3}{a_{n-1}^2}$. Po jejím zlogaritmování dostáváme $\ln(a_{n+1}) = 3\ln(a_n) - 2\ln(a_{n-1})$. Když si označíme $b_n = \ln(a_n)$, dostáváme tak rovnici $b_{n+1} = 3b_n - 2b_{n-1}$ neboli $b_{n+2} - 3b_{n+1} + 2b_n = 0$. Charakteristická rovnice $\lambda^2 - 3\lambda + 2 = 0$ dává $\lambda = 1, 2$ a řešení $b_n = u1^n + v2^n$. Pak $a_n = e^{b_n} = e^{u+2^n} = e^u \cdot (e^v)^{2^n}$.

b) Zde se podíváme, co se dá dělat s lineární rekurentní rovnicí 1. rádu, pokud nemá konstantní koeficienty.

Uvažujme následující rekurentní vztah: $f(n)a_{n+1} + g(n)a_n = h(n)$.

Vytvoříme funkce $Q(n) = \frac{f(1)f(2)\cdots f(n-1)}{g(1)g(2)\cdots g(n-1)g(n)}$, všimněte si, že $Q(n+1) = Q(n) \frac{f(n)}{g(n+1)}$. Když použijeme substituci $b_n = g(n)Q(n)a_n$ neboli dosazujeme do rovnice $a_n = \frac{b_n}{g(n)Q(n)}$, tak máme

$$\begin{aligned} f(n) \frac{b_{n+1}}{g(n+1)Q(n+1)} + g(n) \frac{b_n}{g(n)Q(n)} &= h(n) \implies b_{n+1} \frac{f(n)g(n+1)}{g(n+1)Q(n)f(n)} + \frac{b_n}{Q(n)} = h(n) \\ \implies b_{n+1} + b_n &= Q(n)h(n). \end{aligned}$$

Dostáváme tedy lineární rekurentní rovnici s konstantními koeficienty, kterou při troše štěstí (když vyjde pravá strana kvazipolynomální) dokážeme vyřešit. Dokonce lze napsat obecný vzorec. Jestliže začneme počáteční hodnotou $b_1 = -C$, pak $b_2 = Q(1)h(1) + C$, $b_3 = Q(2)h(2) - Q(1)h(1) - C$ atd., indukcí ukážeme vztah

$$b_n = (-1)^n C - \sum_{i=1}^{n-1} (-1)^{n+i} Q(i)h(i). \text{ Substituční rovnice pak dává } a_n = \frac{(-1)^n C - \sum_{i=1}^{n-1} (-1)^{n+i} Q(i)h(i)}{g(n)Q(n)}.$$

Bohužel, míra nutného štěstí je docela vysoká, například rovnice $a_{n+1} - (n+1)a_n = n+1$, kterou jsme dostali v příkladě, dává $f(n) = 1$, $g(n) = -(n+1)$, $h(n) = n+1$, odtud $Q(n) = \frac{(-1)^n}{(n+1)!}$ a dostáváme rovnici

$b_{n+1} + b_n = \frac{(-1)^n}{n!}$. Tím jsme skončili, kvazipolynom na pravé straně není a obecný součtový vzorec výše nám také k přesnému řešení v uzavřeném tvaru nepomůže.

Cvičení

Cvičení 10b.1 (rutinní): Každé prázdné pole tabulky reprezentuje rekurentní rovnici, jejíž levou stranu najdete v záhlaví sloupce a pravou stranu napravo v označení řádku. Například první pole druhého (prázdného) řádku dává rovnici $a_{n+2} - 4a_{n+1} + 3a_n = (-1)^n$.

Pro každé pole najděte odhad tvaru partikulárního řešení (obecný, s konstantami A, B, \dots , nemusíte to dál řešit).

$a_{n+2} - 4a_{n+1} + 3a_n =$ [$\lambda = 1, 3$]	$a_{n+2} - 4a_n =$ [$\lambda = -2, 2$]	$a_{n+2} + 2a_{n+1} + a_n =$ [$\lambda = -1 (2\times)$]	$L = / = b_n$
			$= n 2^n$ [$\lambda = 2$]
			$= (-1)^n$ [$\lambda = -1$]
			$= (n-2) 3^n$ [$\lambda = 3$]
			$= n^2 - 1$ [$\lambda = 1$]
			$= 2^n - 2n \cdot (-1)^n$ [$\lambda = 2, -1$]
			$= 1 - (-2)^n$ [$\lambda = 1, -2$]

Cvičení 10b.2 (rutinní, zkouškové): Najděte obecná řešení následujících rovnic:

- (i) $a_{n+2} - 6a_{n+1} + 8a_n = 0$, $n \geq 0$;
- (ii) $a_{n+2} - 4a_n = 0$, $n \geq 1$;
- (iii) $a_{n+2} + a_{n+1} - 2a_n = 0$, $n \geq 2$;
- (iv) $a_{n+2} - 6a_{n+1} + 9a_n = 0$, $n \geq -2$;
- (v) $a_{n+2} + 9a_n = 0$, $n \geq 0$;
- (vi) $a_{n+3} - 2a_{n+2} - a_{n+1} + 2a_n = 0$, $n \geq 1$;
- (vii) $a_{n+2} - a_n = 18n 2^n$, $n \geq 0$;
- (viii) $a_{n+2} + 2a_{n+1} - 3a_n = (5n+12)2^n$, $n \geq 0$;
- (ix) $a_{n+2} - 4a_{n+1} + 4a_n = 13 \cdot 3^n - 3$, $n \geq 1$;
- (x) $a_{n+1} = a_n + 2a_{n-1} + 3 \cdot 2^n - 2 \cdot (-2)^n$, $n \geq 2$;
- (xi) $a_{n+3} + 3a_{n+2} - 4a_n = 16 \cdot 2^n + 9$, $n \geq 0$;
- (xii) $a_{n+1} = a_{n-1} + n - 1$, $n \geq 2$.

Cvičení 10b.3 (rutinní, zkouškové): Najděte řešení následujících úloh s počátečními podmínkami:

- (i) $a_{n+1} = 6a_{n-1} - a_n$, $n \geq 3$; $a_2 = -6$, $a_3 = 78$;
- (ii) $a_{n+2} - a_n = (8n+18)3^n$, $n \geq 1$; $a_1 = 16$, $a_2 = 31$;
- (iii) $a_{n+1} = 4a_n - 5a_{n-1} + 2a_{n-2}$, $n \geq 2$; $a_0 = 1$, $a_1 = 1$, $a_2 = 2$;
- (iv) $a_{n+2} + 4a_n = 0$, $n \geq 0$; $a_0 = 0$, $a_1 = 1$;
- (v) $a_{n+1} = 2a_n - a_{n-1} - 4 \cdot (-1)^n$, $n \geq 0$; $a_{-1} = 0$, $a_0 = 2$;
- (vi) $a_{n+2} + 2a_{n+1} - 3a_n = 5 \cdot 2^n + 8$, $n \geq 1$; $a_1 = 5$, $a_2 = 9$;
- (vii) $a_{n+1} = a_n + 4a_{n-1} - 4a_{n-2} + (6n-7)(-1)^n$, $n \geq 2$; $a_0 = 2$, $a_1 = 4$, $a_2 = 7$.

Cvičení 10b.4 (rutinní, dobré, zkouškové): Uvažujte funkce dané následujícími rekurentními rovnicemi. Pro každou z nich určete asymptotickou rychlosť růstu (viz značení Θ z kapitoly) pomocí metod z této kapitoly, ale bez toho, abyste funkce opravdu počítali.

- (i) $f(n+1) = 2f(n) + 2n$;
- (ii) $f(n+1) = f(n) + 2n$;
- (iii) $f(n+1) = 4f(n) - 3f(n-1) + 2n$;
- (iv) $f(n+1) = 4f(n) - 3f(n-1) + n2^n$;
- (v) $f(n+1) = 4f(n) - 3f(n-1) + 3^n$;
- (vi) $f(n+1) = 4f(n) - 3f(n-1) + 4^n$.

Cvičení 10b.5 (rutinní, dobré, zkouškové): Najděte pomocí rekurentních vztahů explicitní vzorce v uzavřeném tvaru pro následující sumy:

- (i) $1 + 4 + 7 + 10 + \dots + (3n+1) = \sum_{k=0}^n (3k+1)$;
- (ii) $1 + \lambda + \lambda^2 + \dots + \lambda^n = \sum_{k=0}^n \lambda^k$ pro $\lambda \neq 1$;
- (iii) $1 + 3 + 6 + 10 + \dots + \frac{n(n+1)}{2} = \sum_{k=1}^n \frac{k(k+1)}{2}$;
- (iv) $1^2 + 3^2 + \dots + (2n+1)^2 = \sum_{k=0}^n (2k+1)^2$.

Cvičení 10b.6 (poučné): Nástupní plat je 20,000. Ve smlouvě je každoroční zvýšení platu o 5 procent kvůli inflaci plus zvýšení o věrnostní částku 1000. Kolik je plat po n letech?

Cvičení 10b.7 (poučné): Vezmeme si hypotéku H korun. Je úročena měsíčně úrokem r procent, na konci měsíce se platí splátka S . Nechť $H(k)$ je dlužná částka po k měsících splácení.

(i) Najděte rekurentní vztah pro $H(k)$, určete počáteční podmínu a vzniklou úlohu vyřešte.

(ii) Určete dlužnou částku po 2 letech, jsou-li údaje $H = 3 \cdot 10^6$, $r = 0.5$ (tedy půl procenta za měsíc), $S = 13000$. Kolik je roční úroková míra?

(iii) Obecně najděte vzorec, jak vysoká musí být splátka S , aby dlužná částka ubývala.

Cvičení 10b.8 (dobré, poučné): Nechť A_n je matice daná $a_{ii} = 2$ pro všechna i , $a_{ij} = 1$ pro $|i - j| = 1$ a $a_{ij} = 0$ jinak (tedy 2 na diagonále, 1 v místech hned vedle diagonály a 0 jinde). Najděte rekurentní vzorec pro $d_n = \det(A_n)$.

Cvičení 10b.9 (rutinní, poučný): Dokažte, že když posloupnost $\{a_n\}_{n=n_0}^\infty$ řeší rovnici $a_{n+k} + \sum_{i=0}^{k-1} c_i a_{n+i} = b_n$ a posloupnost $\{\tilde{a}_n\}_{n=n_0}^\infty$ řeší rovnici $a_{n+k} + \sum_{i=0}^{k-1} c_i a_{n+i} = \tilde{b}_n$ (stejná levá strana), pak posloupnost $\{a_n + \tilde{a}_n\}_{n=n_0}^\infty$ řeší rovnici $a_{n+k} + \sum_{i=0}^{k-1} c_i a_{n+i} = b_n + \tilde{b}_n$.

Řešení:

10b.1:

$a_{n+2} - 4a_{n+1} + 3a_n =$ [$\lambda = 1, 3$]	$a_{n+2} - 4a_n =$ [$\lambda = -2, 2$]	$a_{n+2} + 2a_{n+1} + a_n =$ [$\lambda = -1$ (2×)]	$L = / = b_n$
$(An + B)2^n$	$n(An + B)2^n$	$(An + B)2^n$	$= n2^n$ [$\lambda = 2$]
$A(-1)^n$	$A(-1)^n$	$n^2 A(-1)^n$	$= (-1)^n$ [$\lambda = -1$]
$n(An + B)3^n$	$(An + B)3^n$	$(An + B)3^n$	$= (n-2)3^n$ [$\lambda = 3$]
$n(An^2 + Bn + C)$	$(An^2 + Bn + C)$	$(An^2 + Bn + C)$	$= n^2 - 1$ [$\lambda = 1$]
$A2^n + (Bn + C)(-1)^n$	$nA2^n + (Bn + C)(-1)^n$	$A2^n + n^2(Bn + C)(-1)^n$	$= 2^n - 2n \cdot (-1)^n$ [$\lambda = 2, -1$]
$nA + B(-2)^n$	$A + nB(-2)^n$	$A + B(-2)^n$	$= 1 - (-2)^n$ [$\lambda = 1, -2$]

10b.2: (i): $(\lambda - 2)(\lambda - 4) = 0$, $\{2^n u + 4^n v\}_{n=0}^\infty$; (ii): $(\lambda - 2)(\lambda + 2) = 0$, $\{2^n u + (-2)^n v\}_{n=1}^\infty$; (iii): $(\lambda + 2)(\lambda - 1) = 0$, $\{u + (-2)^n v\}_{n=2}^\infty$; (iv): $(\lambda - 3)^2 = 0$, $\{(un + v)3^n\}_{n=-2}^\infty$; (v): $(\lambda - 3i)(\lambda + 3i) = 0$, $\{(u \cos(n\frac{\pi}{2}) + v \sin(n\frac{\pi}{2}))3^n\}_{n=0}^\infty$; (vi): $(\lambda - 1)(\lambda + 1)(\lambda - 2) = 0$, $\{u + (-1)^n v + 2^n w\}_{n=1}^\infty$; (vii): $(\lambda - 1)(\lambda + 1) = 0$, $a_{h,n} = u + (-1)^n v$; odhad $a_n = (An + B)2^n$, $\{(6n - 16)2^n + u + (-1)^n v\}_{n=0}^\infty$; (viii): $(\lambda - 1)(\lambda + 3) = 0$, $a_{h,n} = u + (-3)^n v$; odhad $a_n = (An + B)2^n$, $\{n2^n + u + (-3)^n v\}_{n=0}^\infty$; (ix): $(\lambda - 2)^2 = 0$, $a_{h,n} = n2^n u + 2^n v$; odhad $a_n = A3^n + B$, $\{13 \cdot 3^n - 3 + n2^n u + 2^n v\}_{n=1}^\infty$; (x): přepis: $a_{n+2} - a_{n+1} - 2a_n = 6 \cdot 2^n + 4 \cdot (-2)^n$, $n \geq 1$; $(\lambda - 2)(\lambda + 1) = 0$, $a_{h,n} = 2^n u + (-1)^n v$; odhad $a_n = n^1 A2^n + B(-2)^n$, $\{n2^n + (-2)^n + 2^n u + (-1)^n v\}_{n=1}^\infty$; (xi): $(\lambda - 1)(\lambda + 2)^2 = 0$, $a_{h,n} = u + (-2)^n v + n(-2)^n w$; odhad $a_n = A2^n + n^1 B$, $\{n + 2^n + u + (-2)^n v + n(-2)^n w\}_{n=0}^\infty$; (xii): přepis: $a_{n+2} - a_n = n = n \cdot 1^n$, $n \geq 1$; $(\lambda - 1)(\lambda + 1) = 0$, $a_{h,n} = u + (-1)^n v$; odhad $a_n = n(An + B)$, $\{\frac{1}{4}n^2 - \frac{1}{2}n + u + (-1)^n v\}_{n=0}^\infty$.

10b.3: (i): přepis: $a_{n+2} + a_{n+1} - 6a_n = 0$, $n \geq 2$; $(\lambda - 2)(\lambda + 3) = 0$, $\{2^n u + (-3)^n v\}_{n=2}^\infty$, poč. podm. dávají $\{3 \cdot 2^n - 2(-3)^n\}_{n=2}^\infty$; (ii): $(\lambda - 1)(\lambda + 1) = 0$, $a_{h,n} = u + (-1)^n v$; odhad $a_n = (An + B)3^n$, $\{n3^n + u + (-1)^n v\}_{n=1}^\infty$, poč. podm. dávají $\{n3^n + 13\}_{n=1}^\infty$; (iii): přepis: $a_{n+3} - 4a_{n+2} + 5a_{n+1} - 2a_n = 0$, $n \geq 0$; $(\lambda - 1)^2(\lambda - 2) = 0$; $\{u + nv + 2^n w\}_{n=0}^\infty$; poč. podm. dávají $\{2^n - n\}_{n=0}^\infty$; (iv): $(\lambda - 2i)(\lambda + 2i) = 0$, $\{(u \cos(n\frac{\pi}{2}) + v \sin(n\frac{\pi}{2}))2^n\}_{n=0}^\infty$; poč. podm. dávají $\{2^{n-1} \sin(n\frac{\pi}{2})\}_{n=0}^\infty$, je to $\{0, 1, 0, -4, 0, 16, 0, -64, 0, \dots\}$; (v): přepis: $a_{n+2} - 2a_{n+1} + a_n = 4 \cdot (-1)^n$, $n \geq -1$; $(\lambda - 1)^2 = 0$, $a_{h,n} = u + nv$; odhad $a_n = A(-1)^n$, $\{(-1)^n + u + nv\}_{n=-1}^\infty$, poč. podm. dávají $\{(-1)^n + 1\}_{n=-1}^\infty$, je to $\{0, 2, 0, 2, 0, 2, 0, 2, \dots\}$;

(vi): $(\lambda - 1)(\lambda + 3) = 0$, $a_{h,n} = u + (-3)^n v$; odhad $a_n = A2^n + n^1 B$, $\{2^n + 2n + u + (-3)^n v\}_{n=1}^\infty$, poč. podm. dávají $\{2^n + 2n + 1\}_{n=1}^\infty$;

(vii): přepis: $a_{n+3} - a_{n+2} - 4a_{n+1} + 4a_n = (6n + 5)(-1)^n$, $n \geq 0$; $(\lambda - 1)(\lambda - 2)(\lambda + 2) = 0$, $a_{h,n} = u + 2^n v + (-2)^n w$; odhad $a_n = (An + B)(-1)^n$, $\{(n+1)(-1)^n + u + 2^n v + (-2)^n w\}_{n=0}^\infty$, poč. podm. dávají $\{(n+1)(-1)^n + 2^{n+1} - (-2)^n\}_{n=0}^\infty$.

10b.4: (i): char. čísla: $\lambda = 2$, proto hom. řeš. $f(n) = 2^n u$. Odhad pravé strany: $f(n) = An + B$, proto obecné řešení bude tvaru $f(n) = 2^n + An + B$. Protože $2^n \gg (2n + 1)$, bude $f(n) = \Theta(2^n)$.

(ii): char. čísla: $\lambda = 1$, proto hom. řeš. $f(n) = u$. Odhad pravé strany: $f(n) = n(An + B)$, proto obecné řešení bude tvaru $f(n) = u + An^2 + Bn$. Máme $f(n) = \Theta(n^2)$.

(iii): char. čísla: $\lambda = 1, 3$, proto hom. řeš. $f(n) = u + 3^n v$. Odhad pravé strany: $f(n) = n(An + B)$, proto obecné řešení bude tvaru $f(n) = u + 3^n v + An^2 + Bn$. Protože $3^n \gg (An^2 + Bn + u)$, bude $f(n) = \Theta(3^n)$.

(iv): char. čísla: $\lambda = 1, 3$, proto hom. řeš. $f(n) = u + 3^n v$. Odhad pravé strany: $f(n) = (An + B)2^n$, proto obecné řešení bude tvaru $f(n) = u + 3^n v + (An + B)2^n$. Protože $3^n \gg (An2^n + B2^n + u)$, bude $f(n) = \Theta(3^n)$.

(v): char. čísla: $\lambda = 1, 3$, proto hom. řeš. $f(n) = u + 3^n v$. Odhad pravé strany: $f(n) = An3^n$, proto obecné řešení bude tvaru $f(n) = u + 3^n v + An3^n$. Máme $f(n) = \Theta(n3^n)$.

(vi): char. čísla: $\lambda = 1, 3$, proto hom. řeš. $f(n) = u + 3^n v$. Odhad pravé strany: $f(n) = A \cdot 4^n$, proto obecné řešení bude tvaru $f(n) = u + 3^n v + A \cdot 4^n$. Protože $4^n \gg (3^n + u)$, bude $f(n) = \Theta(4^n)$.

10b.5: (i): $s_{n+1} = s_n + (3n + 4)$ a $s_0 = 1$; $s_{h,n} = u$, odhad $s_n = n(An + B) = An^2 + Bn$, po dosazení $A = \frac{3}{2}$, $B = \frac{5}{2}$, $s_n = \frac{3}{2}n^2 + \frac{5}{2}n + u$. Poč. podm. dá $u = 1$, $s_n = \frac{(n+1)(3n+2)}{2}$.

(ii): $s_{n+1} = s_n + \lambda^{n+1}$ a $s_0 = 1$; $s_{h,n} = u$, odhad pro $s_{n+1} - s_n = \lambda \cdot \lambda^n$ a $\lambda \neq 1$ je $s_n = A\lambda^n$, po dosazení $A = \frac{\lambda}{\lambda-1}$, $s_n = \frac{\lambda}{\lambda-1}\lambda^n + u$. Poč. podm. dá $u = -\frac{1}{\lambda-1}$, $s_n = \frac{1-\lambda^{n+1}}{1-\lambda}$.

(iii): $s_{n+1} = s_n + \frac{(n+1)(n+2)}{2}$ a $s_1 = 1$; $s_{h,n} = u$, odhad $s_n = n(An^2 + Bn + C) = An^3 + Bn^2 + Cn$, po dosazení $A = \frac{1}{6}$, $B = \frac{1}{2}$, $C = \frac{1}{3}$, $s_n = \frac{1}{6}n^3 + \frac{1}{2}n^2 + \frac{1}{3}n + u$. Poč. podm. dá $u = 0$, $s_n = \frac{n(n+1)(n+2)}{6}$.

(iv): $s_{n+1} = s_n + (2n + 3)^2$ a $s_0 = 1$; $s_{h,n} = u$, odhad $s_n = n(An^2 + Bn + C) = An^3 + Bn^2 + Cn$, po dosazení $A = \frac{4}{3}$, $B = 4$, $C = \frac{11}{3}$, $s_n = \frac{4}{3}n^3 + 4n^2 + \frac{11}{3}n + u$. Poč. podm. dá $u = 1$, $s_n = \frac{(n+1)(2n+1)(2n+3)}{3}$.

10b.6: $P(n+1) = 1.05 \cdot P(n) + 1000$, $P(0) = 20000$.

Homogenní: $P(n+1) - 1.05 \cdot P(n) = 0$, $\lambda = 1.05$, $P_h(n) = (1.05)^n$.

Odhad kvazipolynomiální pravé strany: $P(n) = A$, dosadit, $-0.05A = 1000$, $A = -20000$, proto $P(n) = u(1.05)^n - 20000$. Poč. podmínka dá $40000 = u$. Proto $P(n) = 40000 \cdot (1.05)^n - 20000$.

10b.7: (i): $H(k+1) = (1 + \frac{r}{100})H(k) - S$, $H(0) = H$; označme $R = 1 + \frac{r}{100}$.

Homogenní: $H(k+1) - RH(k) = 0$, $\lambda = R$, $H_h(k) = uR^k$. Toto popisuje růst dluhu bez splátek.

Odhad kvazipolynomiální pravé strany: $H(k) = A$, dosadit, $(1 - R)A = -S$, $A = \frac{S}{R-1} \frac{100S}{r}$, $H(k) = uR^k + \frac{100S}{r}$.

Poč. podmínka: $H - \frac{100S}{r} = u$. Proto $H(k) = (H - \frac{100S}{r})(1 + \frac{r}{100})^k + \frac{100S}{r}$.

(ii): $H(k+1) - 1.005H(k) = 15000$, $H(0) = 3000000$;

Kdyby se nic nesplácelo, byla by dlužná částka za rok $(1.005)^{12}H(0)$, tedy roční úrok je $100 \cdot ((1.005)^{12} - 1) \sim 6.17$.

Po dvou letech je dluh $H(24) = 400000 \cdot (1.005)^{24} + 2600000 \sim 3025000$.

(iii): Je potřeba $H - \frac{100S}{r} < 0$ tedy $S > \frac{rH}{100}$.

10b.8: Rozvoj podle prvního sloupce a pak prvního řádku

$$d_{n+1} = \begin{vmatrix} 2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 2 & 1 & \cdots & 0 \\ 0 & 0 & 1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 \end{vmatrix} = 2 \cdot \begin{vmatrix} 2 & 1 & 0 & \cdots & 0 \\ 1 & 2 & 1 & \cdots & 0 \\ 0 & 1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix} - 1 \cdot \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 2 & 1 & \cdots & 0 \\ 0 & 1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{vmatrix} = 2d_n - 1 \cdot 1 \cdot \begin{vmatrix} 2 & 1 & \cdots & 0 \\ 1 & 2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 2 \end{vmatrix},$$

tedy $d_{n+1} = 2d_n - d_{n-1}$.

10b.9: Dosaděte $a_n + \tilde{a}_n$ do levé strany rovnice, rozňasobte, shromážďte k sobě všechna a_n a všechna \tilde{a}_n a pak použijte předpoklad, viz důkaz Věty .

10c. Další rovnice (Master theorem)

Mnohé rekurzivní algoritmy používají metodu „rozděl a panuj“ (divide-and-conquer). Problém velikosti n se rozdělí na a menších problémů velikosti $\frac{n}{b}$ (reálně nejbližší celé číslo k $\frac{n}{b}$, třeba $\lceil \frac{n}{b} \rceil$). Ty se vyřeší a jednotlivé výsledky je pak ještě třeba zpracovat, to taky něco stojí, takže typická rovnice pro náročnost je $a_n = a \cdot a_{n/b} + g(n)$. Bývá tradiční používat v této souvislosti funkce, takže spíš vidíme rovnici $f(n) = a \cdot f(\frac{n}{b}) + g(n)$. Tato rozhodně není lineární.

Příklad 10c.a: Zde se podíváme, kolik porovnávání nás stojí binární vyhledávání v seznamu o délce n , označme to $f(n)$. Binární vyhledávání pracuje s uspořádaným seznamem (dle abecedy, velikosti atd.), takže když hledáme v seznamu o velikosti n , můžeme se podívat na prvek uprostřed a hned zjistíme, zda hledaný objekt je v první nebo druhé polovině seznamu. Na příslušnou polovinu (o velikosti $\frac{n}{2}$) pak zase pošleme binární vyhledávání, pokud ale není prázdná (pak bychom řekli, že hledaný objekt v seznamu není, ověření nás stojí další porovnání). Vidíme, že hledání v seznamu o n položkách vyžaduje kolik porovnávání, kolik hledání v seznamu polovičním plus dvě porovnání, máme tedy $f(n) = f(\frac{n}{2}) + 2$, což je přesně rovnice typu, který v této kapitole budeme zkoumat. Protože jsme o tom ještě nic nevymysleli, zkusíme na to jít selským rozumem.

Nejprve se zbavíme zlomku. Problém, zda zaokrouhlit nahoru či dolů, vyřešíme elegantně tím, že se omezíme (zatím) na n sudá, pak lze rovnici přepsat jako $f(2n) = f(n) + 2$. Kdybychom použili posloupnosti, dostali bychom $a_{2n} = a_n + 2$, což je opravdu rekurentní rovnice, ale rozhodně není lineární. Vidíme, že problémy tohoto typu jsou něčím jiným, než jsme zatím probírali.

Výhoda nového zápisu $f(2n) = f(n) + 2$ je v tom, že jsme se vyhli zlomkům. Nevýhoda je, že nám tato rovnice neumožňuje zjistit hodnoty pro všechna n . Zkusme se intuitivně podívat, co pro neznámou funkci dostáváme.

Začneme počáteční hodnotou pro jednoprvkový seznam $f(1) = 1$ (prostě se podíváme, jestli ten jeden prvek v seznamu je nebo není to, co hledáme). Pak umíme najít $f(2) = f(\frac{2}{2}) + 2 = f(1) + 2 = 1 + 2$, podobně $f(4) = f(\frac{4}{2}) + 2 = f(2) + 2 = (1 + 2) + 2$, $f(8) = f(\frac{8}{2}) + 2 = (1 + 2 + 2) + 2$, $f(16) = (1 + 2 + 2 + 2) + 2$ atd. Naopak třeba $f(6)$ nezískáme, protože $f(6) = f(3) + 2$ a $f(3)$ neznáme. Vidíme, že dotyčný rekurentní vztah nám dává $f(n)$ pro n typu $n = 2^k$, můžeme si tipnout, že pak $f(n) = 2k + 1$. Tento výsledek není uspokojivý, protože se v něm znenadání objevuje k , zatímco proměnná ve funkci je n . Z rovnice $n = 2^k$ si ovšem k dokážeme vyjádřit jako $k = \log_2(n)$ a dostaneme $f(n) = 2\log_2(n) + 1$. To znamená, že binární vyhledávání je dost rychlé, podstatně rychlejší než přímá úměra k délce seznamu.

Jaká je náročnost pro jiná n ? To je něco, co nám dotyčná rovnice přímo neřekne, záleží to na konkrétní interpretaci toho zlomku $\frac{n}{b}$. Zatím to odložíme.

Shrňme si poznatky, které budou hrát významnou roli ve zbytku kapitoly: Jde o zcela nový typ problému, který po vhodném přepisu do tvaru $f(bn) = af(n) + g(bn)$ dává (při troše štěstí) rozumným způsobem hodnoty pro proměnné $n = b^k$. Určení $f(n)$ pro ostatní n není zjevné a později na tom budeme muset zapracovat.

△

Rovnice tohoto typu jsou v teorii algoritmů velice užitečné (viz další příklady či cvičení). Než začneme vytvářet teorii, podíváme se na „homogenní rovnici“, tedy na případ, kdy $g(n) = 0$. Zkušenosť říká, že by měl být nejsnažší.

Příklad 10c.b: Předpokládejme, že funkce f je daná vztahem $f(n) = a \cdot f(\frac{n}{b})$, kde $a > 0$ a $b \in \mathbb{N}$, $b \geq 2$. Zkusíme nalézt f pomocí iterování vztahu z definice.

Jestliže je číslo n ve tvaru $n = b^k$, pak můžeme rekurzí najít

$$\begin{aligned} f(b) &= af\left(\frac{b}{b}\right) = af(1), \quad f(b^2) = af\left(\frac{b^2}{b}\right) = af(b) = a^2f(1), \quad f(b^3) = af(b^2) = a^3f(1), \\ f(b^k) &= af(b^{k-1}) = a^2f(b^{k-2}) = \dots = a^{k-1}f(b) = a^k f(1). \end{aligned}$$

Důkaz správnosti vzorce provedeme indukcí na k , dokazujeme tvrzení $V(k)$: $f(b^k) = a^k f(1)$.

(0) Pro $k = 0$ platí $f(b^0) = f(1) = a^0 f(1)$, to souhlasí.

(1) Předpokládejme, že pro nějaké $k \in \mathbb{N}_0$ platí $f(b^k) = a^k f(1)$. Pak podle definice

$$f(b^{k+1}) = a \cdot f\left(\frac{b^{k+1}}{b}\right) = a \cdot f(b^k) = a \cdot a^k f(1) = a^{k+1} f(1).$$

Dokázali jsme tedy pro všechna $k \in \mathbb{N}_0$ implikaci $V(k) \implies V(k+1)$, címž je důkaz indukcí ukončen a výsledek $f(n) = a^{\log_b(n)} f(1)$ je potvrzen. Opět se zbavíme k ve výsledku: Jestliže $n = b^k$, pak $k = \log_b(n)$, po dosazení máme $f(n) = a^{\log_b(n)} f(1)$.

Kam tato funkce zapadá na naší obvyklé škále (mocniny, geometrické posloupnosti, faktoriály atd.)? To uvidíme, když si tento výraz přepíšeme:

$$a^{\log_b(n)} = (b^{\log_b(a)})^{\log_b(n)} = b^{\log_b(a) \log_b(n)} = (b^{\log_b(n)})^{\log_b(a)} = n^{\log_b(a)}.$$

Proto pro čísla typu $n = b^k$ platí $f(n) = n^{\log_b(a)} f(1)$. Toto je velice výhodný tvar, čísla a, b jsou totiž konstanty známé ze zadání, takže $n^{\log_b(a)} f(1)$ je jistá mocnina n .

Dá se ukázat (uděláme to pro obecnější případy níže), že se tato rychlosť růstu zachová i pro ostatní čísla $n \in \mathbb{N}$. Funkce dané touto nejjednodušší rovnicí tedy rostou polynomiální rychlostí $n^{\log_b(a)}$ neboli $f(n) = \Theta(n^{\log_b(a)})$ (viz kapitola).

Je dobré si všimnout, že pokud $a = 1$, tak dostáváme $\Theta(n^0) = \Theta(1)$, což říká, že funkce by měla být omezená a nikam nerůst. Když použijeme přesné vyjádření výsledku, tak dokonce vidíme, že f je přesně konstantní funkce na

$M: f(n) = a^k f(1) = f(1)$. Je to ostatně vidět i přímo z rovnice, která pak říká $f(bn) = f(n)$ a tedy $f(b) = f(1)$, $f(b^2) = f(b) = f(1)$ atd.

△

Tento příklad byl důležitý, protože ukázal hlavní ingredience naší další práce, zejména ten závěrečný převod na mocninu n budeme opakovaně používat. Raději si jej zvýrazníme:

Fakt 10c.1.

Nechť $b \in \mathbb{N}$ splňuje $b \geq 2$, nechť $a > 0$. Pro všechna $n \in \mathbb{N}$ platí: Jestliže $n = b^k$ pro nějaké $k \in \mathbb{N}_0$, pak $a^k = n^{\log_b(a)}$.

Dalším a hlavním úkolem je zjistit, co se stane, když $g(n)$ není nulové. Postup z příkladu lze zopakovat a dostaneme obecný výsledek.

Lemma 10c.2.

Nechť $b \in \mathbb{N}$, $b \geq 2$. Nechť $f(n)$ je funkce definovaná pro $n \in M = \{b^k; k \in \mathbb{N}_0\}$. Předpokládejme, že existuje $a \in \mathbb{R}$, $a > 0$ a funkce g na M taková, že

$$f(n) = a \cdot f\left(\frac{n}{b}\right) + g(n) \quad \text{pro všechna } n \in M, n \geq 1.$$

Pak pro $n = b^k \in M$ platí

$$f(n) = a^k f(1) + \sum_{i=0}^{k-1} a^i g\left(\frac{n}{b^i}\right).$$

Důkaz (poučný): Zase aplikujeme postup s opakovaným použitím rekurentní rovnice, dokud nedojdeme k $f(1)$.

$$\begin{aligned} f(n) &= af\left(\frac{n}{b}\right) + g(n) = a\left[af\left(\frac{\frac{n}{b}}{b}\right) + g\left(\frac{n}{b}\right)\right] + g(n) = a^2 f\left(\frac{n}{b^2}\right) + ag\left(\frac{n}{b}\right) + g(n) \\ &= a^2 \left[af\left(\frac{\frac{n}{b^2}}{b}\right) + g\left(\frac{n}{b^2}\right)\right] + ag\left(\frac{n}{b}\right) + g(n) = a^3 f\left(\frac{n}{b^3}\right) + a^2 g\left(\frac{n}{b^2}\right) + ag\left(\frac{n}{b}\right) + g(n) = \dots \\ &= a^k f\left(\frac{n}{b^k}\right) + a^{k-1} g\left(\frac{n}{b^{k-1}}\right) + \dots + a^2 g\left(\frac{n}{b^2}\right) + ag\left(\frac{n}{b}\right) + a^0 g\left(\frac{n}{b^0}\right) \quad \text{a } \frac{a}{b^k} = 1. \end{aligned}$$

Ta pasáž se třemi tečkami samozřejmě způsobuje, že toto není žádný důkaz. Poskytlo nám to nicméně kandidáta, správnost vzorce teď dokážeme indukcí na k . Ukážeme, že funkce f daná tímto vzorcem splňuje danou rovnici.

$$(0) \quad k = 0: f(b^0) = a^0 f(1) + \sum_{i=0}^{-1} a^i g\left(\frac{b^0}{b^i}\right) = f(1) + 0 = f(1).$$

V té sumě se sčítá přes prázdnou množinu, což je automaticky 0.

$$(1) \quad \text{Předpokládejme, že platí } f(b^k) = a^k f(1) + \sum_{i=0}^{k-1} a^i g\left(\frac{b^k}{b^i}\right). \quad \text{Pak máme}$$

$$\begin{aligned} f(b^{k+1}) &= a \cdot f\left(\frac{b^{k+1}}{b}\right) + g(b^{k+1}) = a \cdot f(b^k) + g(b^{k+1}) = a \cdot \left(a^k f(1) + \sum_{i=0}^{k-1} a^i g\left(\frac{b^k}{b^i}\right)\right) + g(b^{k+1}) \\ &= a^{k+1} f(1) + \sum_{i=0}^{k-1} a^{i+1} g\left(\frac{b^{k+1}}{b^{i+1}}\right) + g(b^{k+1}) = \left| \begin{array}{l} j = i + 1 \\ i = 0 \implies j = 1 \\ i = k - 1 \implies j = k \end{array} \right| \\ &= a^{k+1} f(1) + \sum_{j=1}^k a^j g\left(\frac{b^{k+1}}{b^j}\right) + a^0 g\left(\frac{b^{k+1}}{b^0}\right) = a^{k+1} f(1) + \sum_{j=0}^{(k+1)-1} a^j g\left(\frac{b^{k+1}}{b^j}\right). \end{aligned}$$

Správnost vzorce je dokázána. □

Tento vzorec nedává hledanou funkci v pěkném tvaru, naštěstí umíme pro nejobvyklejší typy funkce $g(n)$ najít kompaktnější vyjádření.

! **Příklad 10c.c:** Nechť $b \in \mathbb{N}$, $b \geq 2$, uvažujme funkci f danou na množině $M = \{b^k; k \in \mathbb{N}_0\}$ rekurzivním vztahem $f(n) = a \cdot f\left(\frac{n}{b}\right) + g(n)$ pro $n \geq 1$.

1) Nejprve analyzujeme situaci, kdy $g(n) = c$ pro nějaké $c \in \mathbb{R}$. Lemma říká, že pro $n = b^k$ máme $f(n) = a^k f(1) + \sum_{i=0}^{k-1} a^i c$. Teď jsou dvě možnosti.

1a) Pokud $a = 1$, tak dostaneme $f(n) = f(1) + kc = f(1) + c \log_b(n)$. Funkce tedy na M roste logaritmickou rychlostí.

Poznamenejme, že základ logaritmu zde nehráje zásadní roli, protože jej umíme převést na jakýkoliv jiný díky vzorečku $\log_b(n) = c_{b,B} \log_B(n)$, kde $c_{b,B} = \log_b(B)$. V computer science jsou populární logaritmy o základu 2, takže můžeme psát $f(n) = f(1) + c \log_b(2) \log_2(n)$. Vzhledem k tomu, že při zkoumání růstu funkcí nás násobící kladné konstanty nezajímají, nám ten člen $\log_b(2)$ nevadí. Lze tedy konstatovat, že $f(n) = \Theta(\log_2(n))$ pro $n \in M$.

Mimochodem, přesně sem zapadá rovnice z příkladu , výsledky souhlasí.

1b) Pokud $a > 1$, pak pomocí vzorce z Věty a Faktu dostáváme

$$f(n) = a^k f(1) + c \frac{1-a^k}{1-a} = a^k \left(f(1) - \frac{c}{1-a} \right) + \frac{c}{1-a} = n^{\log_b(a)} \left(f(1) - \frac{c}{1-a} \right) + \frac{c}{1-a}.$$

Zatímco tedy v prvním případě $a = 1$ platí $f(n) = \Theta(\log_2(n))$ (logaritmický růst), ve druhém případě máme $f(n) = \Theta(n^{\log_b(a)})$ (polynomiální růst), zatím samozřejmě pouze na M .

Všimněte si, že ve výrazu $n^{\log_b(a)}$ už se základem logaritmu pohybovat beztrestně nelze, protože vzniklá konstanta už nebude nezajímavá, ale podstatně ovlivní rychlosť růstu tím, že nám umocní n : $n^{\log_b(n)} = (n^{\log_b(2)})^{\log_2(n)}$. Takže to necháme pěkně tak, jak to vyšlo.

Ještě bychom měli prozkoumat případ $0 < a < 1$, ale nemá to moc smysl. Pak totiž dostáváme stejný vzorec jako v části 1b) a pro $0 < a < 1$ v něm platí $a^k \rightarrow 0$, pro velká n je tedy funkce f v zásadě konstantní. Najdou se algoritmy, které trvají v zásadě stále stejně, ať je na vstupu cokoliv (třeba algoritmus, který na každý vstup reaguje vypsáním „Dnes nemám náladu“ a skončí), ale na takové algoritmy asi přes rekurentní vzorce stejně nepůjdeme. Většina autorů proto předpokládá automaticky, že ve studovaných rovnicích je $a \geq 1$, začneme to také dělat.

Poslední poznámka: Jak to bude s výsledky, když $c = 0$, neboli máme homogenní případ řešený v příkladě ? Dosazením $c = 0$ do vzorců, které jsme zde odvodili, dostaneme stejné výsledky jako v onom příkladě, takže je vše v pořádku.

2) Teď budeme analyzovat situaci, kdy $g(n) = cn^d$ pro nějaké $c \in \mathbb{R} - \{0\}$, $d \in \mathbb{N}$. Lemma pak pro $n = b^k$ dává

$$f(n) = a^k f(1) + \sum_{i=0}^{k-1} a^i c \left(\frac{n}{b^i} \right)^d = a^k f(1) + cn^d \sum_{i=0}^{k-1} \left(\frac{a}{b^d} \right)^i.$$

Opět scítáme geometrickou posloupnost, takže musíme rozebrat dva případy.

2a) Jestliže $a = b^d$, pak dostaneme

$$f(n) = a^k f(1) + cn^d \sum_{i=0}^{k-1} 1 = a^k f(1) + cn^d k = n^{\log_b(a)} f(1) + cn^d \log_b(n).$$

Všimněte si, že $a = b^d$ znamená $\log_b(a) = d$, takže dostáváme $f(n) = n^d f(1) + cn^d \log_b(n)$.

2b) Druhý případ je, že $a \neq b^d$, tedy $\frac{a}{b^d} \neq 1$ a geometrický součet získáme pomocí vzorce z Věty . V následných úpravách se nám pak kromě Faktu bude hodit pozorování, že $(b^d)^k = (b^k)^d = n^d$.

$$\begin{aligned} f(n) &= a^k f(1) + cn^d \cdot \frac{1 - (\frac{a}{b^d})^k}{1 - \frac{a}{b^d}} = a^k f(1) + cn^d \cdot \frac{b^d - b^d \frac{a^k}{(b^d)^k}}{b^d - a} = a^k f(1) + cn^d \cdot \frac{b^d - b^d \frac{a^k}{n^d}}{b^d - a} \\ &= a^k f(1) - n^d \cdot \frac{a^k}{n^d} \frac{cb^d}{b^d - a} + n^d \cdot \frac{cb^d}{b^d - a} = n^{\log_b(a)} \left(f(1) - \frac{cb^d}{b^d - a} \right) + n^d \cdot \frac{cb^d}{b^d - a}. \end{aligned}$$

3) První část byla dobrá jako příprava na těžší část 2), nicméně se nabízí otázka, zda to nebylo zbytečné. Když ve výsledku části 2) použijeme $d = 0$, dostaneme stejné vzorce jako v části 1)? Ano. Je tedy možné udělat obecný závěr.

Výsledek: Nechť $g(n) = cn^d$ pro $c \in \mathbb{R} - \{0\}$ a $d \in \mathbb{N}_0$.

Jestliže $a = b^d$, tak $f(n) = n^d f(1) + cn^d \log_b(n)$.

Jestliže $a \neq b^d$, tak $f(n) = n^{\log_b(a)} \left(f(1) - \frac{cb^d}{b^d - a} \right) + n^d \frac{cb^d}{b^d - a}$.

Všimněte si, že jsme vyloučili možnost $c = 0$. V této verzi výsledku se totiž soustředíme na parametr d , což u homogenní rovnice nemá smysl, třeba proto, že libovolné d bude vyhovovat: Platí $g(n) = 0 = 0n^d$ pro všechna d , címkž by vznikl problém se zařazením do správné varianty v našem výsledku. Všimněte si nicméně, že když dosadíme $c = 0$ do druhého vzorce, tak dostaneme přesně výsledek z příkladu . Je tedy možné do tohoto případu zahrnout i možnost $c = 0$, bude se nám to hodit později.

△

Příklad nám poskytl zajímavé vzorce, zásadní problém ovšem je, že jsme dostali funkci pouze na množině M , ale algoritmy se používají pro n i mimo tu množinu. Funkce náročnosti tam má určitě nějakou hodnotu, ale z

rekurentní rovnice ji nedostaneme, záleží to na tom, jak se v konkrétním algoritmu řeší rozdělování na b částí. Zde si pomůžeme dvěma zjednodušenimi. Za prvé, nás vlastně často nezajímá přesný vzorec pro f , ale odhad rychlosti růstu, což už jsme ostatně dělali v předchozích příkladech nebo ve cvičení. Díky tomu nepotřebujeme úplně přesnou informaci, stačí jen přibližná. Tu získáme z další úvahy, je totiž rozumné předpokládat, že f nikde neklesá (což se u náročnosti algoritmu dá očekávat). Pak už lze výsledky získané v příkladě vztáhnout na všechna n .

! Věta 10c.3. (The Master theorem)

Uvažujme neklesající nezápornou funkci f na \mathbb{N} . Pro nějaké $b \in \mathbb{N}$, $b \geq 2$ označme $M = \{b^k; k \in \mathbb{N}\}$ a předpokládejme, že f splňuje na M rovnici $f(n) = a \cdot f\left(\frac{n}{b}\right) + cn^d$ pro konstanty $a, c \in \mathbb{R}$, $d \in \mathbb{N}_0$ splňující $a \geq 1$ a $c > 0$. Pak platí následující:

- (i) Jestliže $a > b^d$, tak $f(n) = \Theta(n^{\log_b(a)})$.
- (ii) Jestliže $a = b^d$, tak $f(n) = \Theta(n^d \log_2(n))$.
- (iii) Jestliže $a < b^d$, tak $f(n) = \Theta(n^d)$.

Důkaz (poučný): (i): Jestliže $a > b^d$, tak podle příkladu máme pro $n \in M$ vzorec $f(n) = c_1 n^{\log_b(a)} + c_2 n^d$, kde $c_1, c_2 \in \mathbb{R}$ jsou konstanty nezávisející na n . Z $a > b^d$ plyne $\log_b(a) > d$, proto první sčítanec převáží nad druhým a funkce $f(n)$ roste jako $c_1 n^{\log_b(a)}$, pokud tedy toto číslo není nula. Co o c_1 víme? Máme $c_1 = f(1) - \frac{cb^d}{b^d - a}$, kde $f(1) \geq 0$ (f je nezáporná), zlomek je pak díky předpokladům $c > 0$, $b > 0$ a $a > b^d$ záporný a odčítáme jej, tedy $c_1 > 0$. První člen proto nezmizí a je násoben kladným číslem, lze tedy říct, že funkce $f(n)$ roste jako $n^{\log_b(a)}$.

Proto existují konstanty $C_1 < C_2$ takové, že pro $n \in M$ máme $C_1 n^{\log_b(a)} \leq f(n) \leq C_2 n^{\log_b(a)}$. Pro $n = b^k$ to znamená, že $C_1 a^k \leq f(b^k) \leq C_2 a^k$.

Nyní tento odhad rozšíříme i na čísla v mezerách množiny M . Připomeňme, že pro b^k máme $a^k = (b^k)^{\log_b(a)}$.

Vezměme libovolné $n \in \mathbb{N}$, $n \geq b$. Pak existuje $k \in \mathbb{N}$ takové, že $b^k \leq n < b^{k+1}$, a můžeme odhadovat následovně:

$$\begin{aligned} f(n) &\leq f(b^{k+1}) \leq C_2 a^{k+1} = a C_2 a^k = a C_2 (b^k)^{\log_b(a)} \leq a C_2 n^{\log_b(a)}, \\ f(n) &\geq f(b^k) \geq C_1 a^k = \frac{C_1}{a} a^{k+1} = \frac{C_1}{a} (b^{k+1})^{\log_b(a)} \geq \frac{C_1}{a} n^{\log_b(a)}. \end{aligned}$$

Označme $D_1 = \frac{C_1}{a}$ a $D_2 = a C_2$. Právě jsme dokázali, že pro $n \in \mathbb{N}$, $n \geq b$ platí $D_1 n^{\log_b(a)} \leq f(n) \leq D_2 n^{\log_b(a)}$. Odtud už plyne, že podobný odhad (jen možná s jinými konstantami) platí i pro všechna $n \in \mathbb{N}$. Máme $f(n) = \Theta(n^{\log_b(a)})$.

(ii): Jestliže $a = b^d$, tak podle příkladu máme pro $n \in M$ vzorec $f(n) = f(1)n^d + cn^d \log_b(n)$. Druhý člen pro velká n převáží a koeficient c je kladný, proto existují konstanty $C_1 < C_2$ takové, že pro $n \in M$ máme $C_1 n^d \log_b(n) \leq f(n) \leq C_2 n^d \log_b(n)$. Pro $n = b^k$ to tedy znamená, že $C_1 b^{dk} k \leq f(b^k) \leq C_2 b^{dk} k$.

Tento odhad zase rozšíříme. Použijeme přitom s úspěchem ekvivalentní nerovností $k+1 \leq 2k$ a $k \geq \frac{k+1}{2}$, které evidentně platí pro $k \geq 1$.

Vezměme tedy libovolné $n \in \mathbb{N}$, $n \geq b$. Pak existuje $k \in \mathbb{N}$ takové, že $b^k \leq n < b^{k+1}$, a můžeme odhadovat takto:

$$\begin{aligned} f(n) &\leq f(b^{k+1}) \leq C_2 b^{d(k+1)}(k+1) \leq C_2 b^{dk} b^d 2k = 2b^d C_2 (b^k)^d k \leq 2b^d C_2 n^d \log_b(n), \\ f(n) &\geq f(b^k) \geq C_1 b^{dk} k \geq \frac{C_1}{b^d} b^{d(k+1)} \frac{k+1}{2} = \frac{C_1}{2b^d} (b^{k+1})^d (k+1) \geq \frac{C_1}{2b^d} n^d \log_b(n). \end{aligned}$$

Označme $D_1 = \frac{C_1}{2b^d}$ a $D_2 = 2b^d C_2$. Dokázali jsme, že pro $n \in \mathbb{N}$, $n \geq b$ je $D_1 n^d \log_b(n) \leq f(n) \leq D_2 n^d \log_b(n)$, tedy $f(n) = \Theta(n^d \log_b(n))$.

(iii): Jestliže $a < b^d$, tak podle příkladu máme pro $n \in M$ vzorec $f(n) = c_1 n^{\log_b(a)} + c_2 n^d$, kde $c_1, c_2 \in \mathbb{R}$ jsou konstanty nezávisející na n . Z $a < b^d$ máme $\log_b(a) < d$, proto druhý sčítanec převáží nad prvním, pokud tedy není násoben nulou. Máme $c_2 = \frac{cb^d}{b^d - a}$, z předpokladů $c > 0$, $b > 0$ a $a < b^d$ tedy dostáváme $c_2 > 0$ a funkce $f(n)$ opravdu roste jako n^d . Proto existují konstanty $C_1 < C_2$ takové, že pro $n \in M$ máme $C_1 n^d \leq f(n) \leq C_2 n^d$. Pro $n = b^k$ to tedy znamená, že $C_1 b^{kd} \leq f(b^k) \leq C_2 b^{kd}$.

Zase toto rozšíříme. Vezměme libovolné $n \in \mathbb{N}$, $n \geq b$. Pak existuje $k \in \mathbb{N}$ takové, že $b^k \leq n < b^{k+1}$, a můžeme odhadovat následovně:

$$\begin{aligned} f(n) &\leq f(b^{k+1}) \leq C_2 b^{(k+1)d} = b^d C_2 b^{kd} = b^d C_2 (b^k)^d \leq b^d C_2 n^d, \\ f(n) &\geq f(b^k) \geq C_1 b^{kd} = \frac{C_1}{b^d} b^{kd+d} = \frac{C_1}{b^d} (b^{k+1})^d \geq \frac{C_1}{b^d} n^d. \end{aligned}$$

Označíme-li $D_1 = \frac{C_1}{b^d}$ a $D_2 = b^d C_2$, tak jsme právě dokázali, že pro $n \in \mathbb{N}$, $n \geq b$ platí $D_1 n^d \leq f(n) \leq D_2 n^d$ a tedy $f(n) = \Theta(n^d)$. □

Interpretace: Začneme tím, že je-li funkce f dána nejjednodušším homogenním vztahem $f(n) = a \cdot f\left(\frac{n}{b}\right)$, pak má růst $\Theta(n^{\log_b(a)})$, viz příklad . Pokud něco na pravou stranu přidáme, pak nám důsledek tohoto kroku odhaluje právě dokázaná věta. Porovnává přitom a s b^d , což je ekvivalentní s porovnáním $\log_b(a)$ a d . Věta tedy porovnává „přirozený růst“ s tím, co jsme přidali. To je velmi praktický pohled na věc, proto si tak Větu přepíšeme a zároveň do závěru zahrneme i onen homogenní případ, který jsme ve větě předpokladem $c > 0$ vyloučili (viz poznámka na konci příkladu).

! Důsledek 10c.4.

Uvažujme neklesající nezápornou funkci f na \mathbb{N} . Pro nějaké $b \in \mathbb{N}$, $b \geq 2$ označme $M = \{b^k; k \in \mathbb{N}\}$ a předpokládejme, že f splňuje na M rovnici $f(n) = a \cdot f\left(\frac{n}{b}\right) + cn^d$ pro konstanty $a, c \in \mathbb{R}$, $d \in \mathbb{N}_0$ splňující $a \geq 1$ a $c \geq 0$. Pak platí následující:

- (i) Jestliže $d < \log_b(a)$ nebo $c = 0$, tak $f(n)$ je $\Theta(n^{\log_b(a)})$.
- (ii) Jestliže $d = \log_b(a)$, tak $f(n)$ je $\Theta(n^{\log_b(a)} \log_2(n)) = \Theta(n^d \log_2(n))$.
- (iii) Jestliže $d > \log_b(a)$, tak $f(n)$ je $\Theta(n^d)$.

Vidíme tedy, že pokud na pravou stranu nepřidáme moc (malá mocnina), tak se nic nestane, funkce roste stejně rychle, jako kdybychom na pravou stranu nepřidali nic. Převažuje tedy příspěvek od homogenní rovnice. Jakmile ale přesáhneme určitou mez, tak zcela převáží to, co jsme na pravou stranu přidali. Zajimavý je onen okamžik přechodu, výsledný vzorec je rychlejší než jak příspěvek od homogenní rovnice, tak příspěvek od pravé strany (jsou stejné), jako by se navzájem posilovali.

Při zkoumání algoritmů nám tedy po nalezení vhodného rekurentního vztahu věta (či její důsledek) okamžitě a bez jakékoliv další práce dává přesně to, co nás zajímá. Její název je tedy případný.

Příklad 10c.d (pokračování): Vrátíme se k binárnímu vyhledávání. Odvodili jsme rovnici $f(n) = f\left(\frac{n}{2}\right) + 2$. Parametry jsou $a = 1$, $b = 2$ a $d = 0$, neboli $a = 1 = b^d$, máme také $\log_b(a) = \log_2(1) = 0$ takže nám Věta dává, že náročnost tohoto algoritmu je $f(n) = \Theta(n^0 \log_2(n)) = \Theta(\log_2(n))$. Odpovídá to tomu, co jsme si sami předtím intuitivně odvodili. Důsledek nám to samozřejmě dá také, tam bychom použili test $d = 0 = \log_2(1)$.

△

Příklad 10c.e: Zde si představíme rychlé násobení: Mějme dvě čísla, a a b o n cifrách, řekněme v binárním tvaru. Standardní algoritmus pro násobení vyžaduje více než n^2 operací: násobení každého s každým pro jednotlivé bity obou čísel plus věci jako sčítání mezivýsledků, které náročnost dálé zhorší, ale ne natolik, aby to zvýšilo rychlosť růstu n^2 . Mimochodem, toto klasické násobení je zase dosti úsporné na paměť, vyžaduje navíc jen cca $\log(n)$ registrů.

Existuje zajímavá finta: Rozdělíme obě čísla na poloviny (ve smyslu řetězců číslic) o délce $m = \frac{n}{2}$ míst, takže $a = A_1 \cdot 2^m + A_2$, $b = B_1 \cdot 2^m + B_2$. Pak

$$ab = (A_1 2^m + A_2)(B_1 2^m + B_2) = (2^{2m} + 2^m)A_1 B_1 + 2^m(A_1 - A_2)(B_2 - B_1) + (2^m + 1)A_2 B_2$$

(ověřte). Všimněte si, že se tam násobí jen čísla velikosti m , a to třikrát, čili náročnost je $3m^2 = \frac{3}{4}n^2$. Jsou tam i další operace, ale ty jsou ve srovnání s $\frac{3}{4}n^2$ nenáročné. Vypadají sice na první pohled jako násobení, jenže pokud jsou všechna uvažovaná čísla v binárním tvaru, pak jde vlastně jen o posuny doleva, což jsou velice rychlé operace ve srovnání s násobením.

Vidíme, že tímto půlícím trikem jsme se dostali z n^2 na $\frac{3}{4}n^2$, což je pokrok, ovšem nic nám nebrání podobnou fintu rekurzivně aplikovat i na ta tři násobení, dostaneme tak algoritmus pro chytré násobení čísel délky $n = 2^k$. Kolik zabere operací?

Když započítáme doplňující faktory, které jsou všechny lineárně náročné, dostaneme rovnici $f(n) = 3f\left(\frac{n}{2}\right) + cn$. Máme konstanty $a = 3$, $b = 2$ a $d = 1$, zde je d je menší než $\log_b(a) = \log_2(3)$ a podle Důsledku je tedy náročnost algoritmu řádu $\Theta(n^{\log_2(3)})$. Protože $\log_2(3) \sim 1.585\dots$, je to lepší než klasické násobení s jeho n^2 .

Poznámka: Podobná finta existuje pro násobení matic, které standardně „stojí“ n^3 násobení a $n^2(n-1)$ sčítání, čili v zásadě je to algoritmus s náročností n^3 . Chytré násobení pro matice je založeno na rovnosti, pomocí které se místo jednoho násobení dvou $n \times n$ matic sedmkrát násobí dvě matice o rozměru $\frac{n}{2}$ a pak se ještě použije 15 sčítání matic téže velikosti, zase se to dá zrekurzivnit a dostaneme $f(n) = 7f\left(\frac{n}{2}\right) + 15\left(\frac{n}{2}\right)^2$. Máme tedy $a = 7$,

$b = 2, d = 2$, opět je d menší než $\log_b(a) = \log_2(7) \sim 2.8$, proto má toto maticové násobení náročnost $\Theta(n^{\log_2(7)})$, což je poněkud lepší než obvyklých n^3 .

△

Podobně jako u lineárních rekurentních rovnic je také možné pravé strany kombinovat.

!

Věta 10c.5. (o superpozici)

Nechť $a \in \mathbb{R}$, $b \in \mathbb{N}$ splňují $a \geq 1$ a $b \geq 2$, označme $M = \{b^k; k \in \mathbb{N}\}$.

Jestliže funkce f_1 splňuje rovnici $f(n) = af\left(\frac{n}{b}\right) + g_1(n)$, $n \in M$

a funkce f_2 splňuje rovnici $f(n) = af\left(\frac{n}{b}\right) + g_2(n)$, $n \in M$,

pak funkce $f_1 + f_2$ splňuje rovnici $f(n) = af\left(\frac{n}{b}\right) + g_1(n) + g_2(n)$, $n \in M$.

Indukcí se to samozřejmě dá snadno rozšířit na libovolný (konečný) počet sčítaných funkcí g na pravé straně. Co z toho pro nás plyne prakticky? Jestliže zkoumáme funkci danou vztahem $f(n) = af\left(\frac{n}{b}\right) + p(n)$, kde p je polynom, tak pro každou mocninu tohoto polynomu dostaneme funkci známého růstu. Když je sečteme, dostaneme řešení celé dané rovnice, a to roste tak rychle, jak rychle roste nejrychlejší ze sčítanců. Podíváme-li se na jednotlivé možnosti u Důsledku , tak hned vidíme, že když dohromady zamícháme řešení odpovídající několika hodnotám d , tak nejrychleji z nich poroste právě to, které odpovídá největšímu d . Z toho vyplývá následující závěr.

Praktické pravidlo: Jestliže je funkce f určena rovnicí $f(n) = af\left(\frac{n}{b}\right) + p(n)$, kde p je polynom, pak je rychlosť růstu f určena podle Důsledku , kde za d vezmeme stupeň polynomu p .

Jinými slovy, je-li na pravé straně polynom, tak záleží jen na jeho největší mocnině (což je vlastně při úvahách s rychlosťí staré dobré pravidlo, viz kapitola).

Pozorný čtenář ovšem může namítnout, že jsme se počínaje Větou omezili na $c \geq 0$. Co se stane, když budou některé (všechny) koeficienty v polynomu záporné? Z praktického pohledu to není problém. Bud' je příslušná mocnina tak malá, aby výsledek neovlivnila ($d < \log_b(a)$), pak záporný koeficient nevadí, protože skutečný růst funkce udává jiná část vztahu. Nebo je záporný koeficient u mocniny, která je dost velká na to, aby určila růst f , pak se dozvíme třeba to, že f je jako $-n^2$, což je ale u praktických příkladů nemožné, neboť náročnost algoritmů nemůže být záporná. Čili tam, kde by záporné koeficienty mohly vadit, se zase u prakticky motivovaných úloh nemohou objevit.

Na závěr jednu kuriozitu. Rovnice typu $T(n) = aT^2(n/b)$ díky té druhé mocnině neumíme zkoumat pomocí výsledků této kapitoly. Nabízí se trik. Nejprve použijeme substituci $n = b^k$, dostaváme $T(b^k) = aT^2(b^{k-1})$. Teď rovnici zlogaritmujeme: $\ln(T(b^k)) = \ln(a) + 2\ln(T(b^{k-1}))$. Když označíme $a_k = \ln(T(b^k))$, dostaneme rovnici $a_k = 2a_{k-1} + \ln(a)$, což je lineární rekurentní rovnice a snadno ji vyřešíme pomocí algoritmu .

Co to ukazuje? Učebnice se většinou zabývají výkladem metod pro řešení určitých typů rovnic, takže by to mohlo vzbudit dojem, že rovnice umíme nějak plánovitě řešit. Ve skutečnosti umíme takto řešit jen velice malou skupinku rovnic, těch nejpřeknějších. Stačí malá modifikace a objeví se rovnice, na kterou obvyklé metody nelze aplikovat. Pak to začne být zajímavé, znalost metod je jen nezbytný základ, ale pak přichází hledání fint a triků, jak si poradit s něčím, co do našich škatulek nezapadá.

Cvičení

Cvičení 10c.1 (rutinní): (i) Uvažujte funkci danou $f(n) = f\left(\frac{n}{4}\right) + 3$ a $f(1) = 1$. Spočítejte $f(4)$, $f(16)$, $f(256)$.
(ii) Uvažujte funkci danou $f(n) = f\left(\frac{n}{2}\right) + n^2$ a $f(1) = 0$. Spočítejte $f(2)$, $f(4)$ a $f(32)$.

Cvičení 10c.2 (rutinní, poučné, *zkouškové): Pro následující funkce nejprve odhadněte přesný vzorec na množině $M = \{b^k\}$ iterací definičního vztahu a dokažte indukcí jeho správnost (viz příklad , popř. důkaz Lemma), poté aplikujte Master theorem (či jeho důsledek) k ověření asymptotické rychlosti růstu funkce.

- | | |
|--|--|
| (i)* $f(n) = 2f\left(\frac{n}{3}\right)$, $f(1) = 13$; | (iv)* $f(n) = 3f\left(\frac{n}{2}\right) + 1$, $f(1) = 13$; |
| (ii)* $f(n) = f\left(\frac{n}{3}\right) + 1$, $f(1) = 13$; | (v) $f(n) = f\left(\frac{n}{2}\right) + \frac{n}{2}$, $f(1) = 13$; |
| (iii)* $f(n) = 2f\left(\frac{n}{2}\right) + 1$, $f(1) = 13$; | (vi) $f(n) = f\left(\frac{n}{2}\right) + 3n^2$, $f(1) = 13$. |

Cvičení 10c.3 (rutinní): Pro následující funkce určete rychlosť jejich asymptotickou růstu pomocí Master theorem, popřípadě Důsledku :

- | | | |
|---|---|--|
| (i) $f(n) = 2f\left(\frac{n}{3}\right)$; | (iv) $f(n) = f\left(\frac{n}{3}\right) + 2n$; | (vii) $f(n) = 8f\left(\frac{n}{2}\right) + 13n$; |
| (ii) $f(n) = f\left(\frac{n}{3}\right)$; | (v) $f(n) = 3f\left(\frac{n}{3}\right) + 2n$; | (viii) $f(n) = 4f\left(\frac{n}{2}\right) + n^2$; |
| (iii) $f(n) = 4f\left(\frac{n}{2}\right)$; | (vi) $f(n) = 2f\left(\frac{n}{4}\right) + 27$; | (ix) $f(n) = 4f\left(\frac{n}{2}\right) + n^3$. |

Cvičení 10c.4 (poučné): V tomto cvičení bude stručně popsáno několik užitečných algoritmů. Pro každý z nich sestavte rekurentní rovnici popisující náročnost algoritmu, poté odhadněte asymptotickou rychlosť růstu doryčné funkce pomocí Věty či Důsledku.

(i) Chceme-li počítat mocninu x^n přímo, vyžaduje $n - 1$ násobení.

Rychlé mocnění: Jestliže máme umocňovat na sudou mocninu, můžeme použít $x^{2m} = (x^m)^2$. K výpočtu je tedy potřeba 1) spočítat x^m 2) vynásobit $x^m \cdot x^m$.

Pokud je i m sudé, můžeme v rozkladu rekurzivně pokračovat, ideální je používat tento algoritmus na mocniny typu x^{2^k} . Určete, kolik je pak třeba násobení.

Například pro výpočet x^8 stačí násobit $x^2 = x \cdot x$, $x^4 = x^2 \cdot x^2$ a $x^8 = x^4 \cdot x^4$, tedy celkem třikrát.

Bonus: Takto lze zjednodušit výpočet libovolné mocniny. Například x^{13} si napíšeme jako $x^{1+4+8} = x \cdot x^4 \cdot x^8$. Tři násobení nám dala všechny nutné mocniny typu x^{2^k} , další dvě násobení nám dají x^{13} . Stačí tedy pět násobení namísto dvanácti. Odhadněte počet násobení nutný k výpočtu obecného x^m .

(ii): Máme seznam o n položkách a chceme jej seřadit podle velikosti/abecedy. Pokud bychom použili metodu „najdi největší, pak najdi největší ze zbytku, pak najdi největší ze zbytku …“, bylo by třeba řádově n^2 porovnání.

Merge sort: Nechť má seznam sudý počet položek $n = 2m$. Rozdělíme jej na poloviny, uspořádáme každou z nich a dva uspořádané seznamy o délce m spojíme do jednoho uspořádaného seznamu, což lze udělat s pouhými $2m = n$ srovnáními. Pro $n = 2^k$ lze postup snadno iterovat. Odhadněte počet srovnání potřebný k urovnání celého seznamu.

(iii) Máme seznam o n položkách a chceme najít největší a nejmenší položku v seznamu. To stojí přinejhorším $2n$ srovnání, pokud použijeme přímočarý útok: První číslo si schováme jako min a max, každé další číslo porovnáme s dočasným minimem a maximem a pokud je extrémnější, nahradíme jím příslušnou hodnotu.

Pokus o nápadobu merge sortu: Rozdělíme seznam na polovinu, najdeme u každé poloviny její maximum a minimum, pak stačí porovnat obě maxima a obě minima a máme globální extrémy. Kolik porovnání vyžaduje algoritmus, když toto půlení iterujeme?

Cvičení 10c.5 (dobré, poučné): Předpokládejme, že funkce f splňuje rekurentní vztah $f(n) = 3f(\sqrt{n}) + 13$ pro n ve tvaru k^2 , $k \in \mathbb{N}$, $k \geq 2$ a platí $f(2) = 1$.

(i) Najděte $f(16)$

(ii) Uvažujte funkci $F(m) = f(2^m)$. Pomocí původní rekurentní rovnice najděte novou rekurentní rovnici pro F , určete rychlosť růstu F a pak i f .

Cvičení 10c.6 (dobré, poučné): Nechť $a \in \mathbb{R}$, $b \in \mathbb{N}$ splňují $a \geq 1$ a $b \geq 2$, označme $M = \{b^k; k \in \mathbb{N}\}$. Dokažte následující:

Jestliže funkce f_1 splňuje rovnici $f(n) = af\left(\frac{n}{b}\right) + g_1(n)$, $n \in M$

a funkce f_2 splňuje rovnici $f(n) = af\left(\frac{n}{b}\right) + g_2(n)$, $n \in M$,

pak funkce $f_1 + f_2$ splňuje rovnici $f(n) = af\left(\frac{n}{b}\right) + g_1(n) + g_2(n)$, $n \in M$.

Cvičení 10c.7 (dobré, poučné): Nechť $a \in \mathbb{R}$, $b \in \mathbb{N}$ splňují $a \geq 1$ a $b \geq 2$, označme $M = \{b^k; k \in \mathbb{N}\}$. Nechť funkce f_p splňuje rovnici $f(n) = af\left(\frac{n}{b}\right) + g(n)$, $n \in M$.

Dokažte, že funkce f splňuje rovnici $f(n) = af\left(\frac{n}{b}\right) + g(n)$ pro všechna $n \in M$ právě tehdy, když $f = f_p + f_h$, kde f_h je nějaká funkce splňující rovnici $f(n) = af\left(\frac{n}{b}\right)$, $n \in M$.

Cvičení 10c.8 (dobré, poučné): Nechť $a \in \mathbb{R}$, $b \in \mathbb{N}$ splňují $a \geq 1$ a $b \geq 2$, označme $M = \{b^k; k \in \mathbb{N}\}$. Nechť N je množina všech funkcí na M splňujících rovnici $f(n) = af\left(\frac{n}{b}\right)$, $n \in M$.

Dokažte, že N je jednodimenzionální vektorový prostor.

Řešení:

10c.1: (i): $f(4) = f(1) + 3 = 1 + 3 = 4$, $f(16) = f(4) + 3 = 4 + 3 = 7$, pracovní mezivýsledek $f(64) = f(16) + 3 = 7 + 3 = 10$ a tedy $f(256) = f(64) + 3 = 10 + 3 = 13$.

(ii): $f(2) = f(1) + 2^2 = 0 + 4 = 4$, $f(4) = f(2) + 4^2 = 4 + 16 = 20$, pracovní mezivýsledek $f(8) = f(4) + 8^2 = 20 + 64 = 84$, $f(16) = f(8) + 16^2 = 84 + 256 = 340$ a tedy $f(32) = f(16) + 32^2 = 340 + 1024 = 1364$.

10c.2: (i): Přepis: $f(3n) = 2f(n)$; $f(3) = 2f(1) = 2 \cdot 13$, $f(3^2) = 2f(3) = 2 \cdot (2 \cdot 13) = 2^2 \cdot 13$, $f(3^3) = 2f(3^2) = 2 \cdot (2^2 \cdot 13) = 2^3 \cdot 13$, odhad $f(3^k) = 13 \cdot 2^k$.

(0) $k = 0$: $f(3^0) = 13 \cdot 2^0 = 13 = f(1)$.

(1) $f(3^k) = 13 \cdot 2^k \implies f(3^{k+1}) = 2f(3^k) = 13 \cdot 2^{k+1}$ souhlasí.

Přepis: $f(n) = 13 \cdot 2^{\log_3(n)} = 13 \cdot n^{\log_3(2)}$ na M .

Důsledek: $a = 2$, $b = 3$, $c = 0$, proto $f(n) = \Theta(n^{\log_3(2)})$.

(ii): Přepis: $f(3n) = f(n) + 1$; $f(3) = f(1) + 1 = 13 + 1$, $f(3^2) = f(3) + 1 = (13 + 1) + 1 = 13 + 2$, $f(3^3) = f(3^2) + 1 = (13 + 2) + 1 = 13 + 3$, odhad $f(3^k) = 13 + k$.

(0) $k = 0$: $f(3^0) = 13 + 0 = 13 = f(1)$.

(1) $f(3^k) = 13 + k \implies f(3^{k+1}) = f(3^k) + 1 = (13 + k) + 1 = 13 + (k + 1)$ souhlasí.

Přepis: $f(n) = 13 + \log_3(n) = 13 + \log_3(2) \log_2(n)$ na M .

Důsledek: $a = 1$, $b = 3$, $d = 0 = \log_3(1)$, proto $f(n) = \Theta(n^0 \log_2(n)) = \Theta(\log_2(n))$.

(iii): Přepis: $f(2n) = 2f(n) + 1$; $f(2) = 2f(1) + 1 = 2 \cdot 13 + 1$, $f(2^2) = 2f(2) + 1 = 2 \cdot (2 \cdot 13 + 1) + 1 = 2^2 \cdot 13 + 1 + 2$,

$f(2^3) = 2f(2^2) + 1 = 2 \cdot (2^2 \cdot 13 + 1 + 2) + 1 = 2^3 \cdot 13 + 1 + 2 + 4$,

$f(2^4) = 2f(2^3) + 1 = 2 \cdot (2^3 \cdot 13 + 1 + 2 + 4) + 1 = 2^4 \cdot 13 + 1 + 2 + 4 + 8$,

odhad $f(2^k) = 13 \cdot 2^k + 1 + 2 + 4 + \dots + 2^{k-1} = 13 \cdot 2^k + \frac{1-2^k}{1-2} = 13 \cdot 2^k + 2^k - 1 = 14 \cdot 2^k - 1$.

(0) $k = 0$: $f(2^0) = 14 \cdot 2^0 - 1 = 13 = f(1)$.

(1) $f(2^k) = 14 \cdot 2^k - 1 \implies f(2^{k+1}) = 2f(2^k) + 1 = 14 \cdot 2^{k+1} - 2 + 1 = 14 \cdot 2^{k+1} - 1$ souhlasí.

Přepis: $f(n) = 14n - 1$ na M .

Důsledek: $a = 2$, $b = 2$, $d = 0 < \log_2(2) = 1$, proto $f(n) = \Theta(n^{\log_2(2)}) = \Theta(n)$.

(iv): Přepis: $f(2n) = 3f(n) + 1$; $f(2) = 3f(1) + 1 = 3 \cdot 13 + 1$, $f(2^2) = 3f(2) + 1 = 3 \cdot (3 \cdot 13 + 1) + 1 = 3^2 \cdot 13 + 1 + 3$,

$f(2^3) = 3f(2^2) + 1 = 3 \cdot (3^2 \cdot 13 + 1 + 3) + 1 = 3^3 \cdot 13 + 1 + 3 + 9$,

$f(2^4) = 3f(2^3) + 1 = 3 \cdot (3^3 \cdot 13 + 1 + 3 + 9) + 1 = 3^4 \cdot 13 + 1 + 3 + 9 + 27$,

odhad $f(2^k) = 13 \cdot 3^k + 1 + 3 + 9 + \dots + 3^{k-1} = 13 \cdot 3^k + \frac{1-3^k}{1-3} = 13 \cdot 3^k + \frac{1}{2} \cdot 3^k - \frac{1}{2} = (13 + \frac{1}{2}) \cdot 3^k - \frac{1}{2}$.

(0) $k = 0$: $f(2^0) = (13 + \frac{1}{2}) \cdot 2^0 - \frac{1}{2} = 13 = f(1)$.

(1) $f(2^k) = (13 + \frac{1}{2}) \cdot 3^k - \frac{1}{2} \implies f(2^{k+1}) = 3f(2^k) + 1 = (13 + \frac{1}{2}) \cdot 3^{k+1} - \frac{3}{2} + 1 = (13 + \frac{1}{2}) \cdot 3^{k+1} - \frac{1}{2}$ souhlasí.

Přepis: $f(n) = (13 + \frac{1}{2}) \cdot 3^{\log_2(n)} - \frac{1}{2} = (13 + \frac{1}{2})n^{\log_2(3)} - \frac{1}{2}$ na M .

Důsledek: $a = 3$, $b = 2$, $d = 0 < \log_2(3)$, proto $f(n) = \Theta(n^{\log_2(3)})$.

(v): Přepis: $f(2n) = f(n) + n$; $f(2) = f(1) + 1 = 13 + 1$, $f(2^2) = f(2) + 2 = 13 + 1 + 2$,

$f(2^3) = f(2^2) + 2^2 = 13 + 1 + 2 + 2^2$, $f(2^4) = f(2^3) + 2^3 = 13 + 1 + 2 + 2^2 + 2^3$,

odhad $f(2^k) = 13 + 1 + 2 + 4 + \dots + 2^{k-1} = 13 + \frac{1-2^k}{1-2} = 13 + 2^k - 1 = 2^k + 12$.

(0) $k = 0$: $f(2^0) = 2^0 + 12 = 13 = f(1)$.

(1) $f(2^k) = 2^k + 12 \implies f(2^{k+1}) = f(2^k) + 2^k = 2^k + 12 + 2^k = 2 \cdot 2^k + 12 = 2^{k+1} + 12$ souhlasí.

Přepis: $f(n) = n + 12$ na M .

Důsledek: $a = 1$, $b = 2$, $d = 1 > \log_2(1) = 0$, proto $f(n) = \Theta(n)$.

(vi): Přepis: $f(2n) = f(n) + 3(2n)^2 = f(n) + 12n^2$; $f(2) = f(1) + 12 \cdot 1^2 = 13 + 12 \cdot 1$,

$f(2^2) = f(2) + 12 \cdot 2^2 = 13 + 12 \cdot 1 + 12 \cdot 2^2$, $f(2^3) = f(2^2) + 12 \cdot (2^2)^2 = 13 + 12 \cdot 1 + 12 \cdot 2^2 + 12 \cdot (2^2)^2$,

$f(2^4) = f(2^3) + 12 \cdot (2^3)^2 = 13 + 12 \cdot 1 + 12 \cdot 2^2 + 12 \cdot (2^2)^2 + 12 \cdot (2^2)^3$,

odhad $f(2^k) = 13 + 12(1 + 4 + 4^2 + \dots + 4^{k-1}) = 13 + 12 \frac{1-4^k}{1-4} = 13 + 4 \cdot 4^k - 4 = 4^{k+1} + 9$.

(0) $k = 0$: $f(2^0) = 4^1 + 9 = 13 = f(1)$.

(1) $f(2^k) = 4^{k+1} + 9 \implies f(2^{k+1}) = f(2^k) + 12(2^k)^2 = 4^{k+1} + 9 + 12(2^2)^k = 4^{k+1} + 3 \cdot 4^{k+1} + 9 = 4 \cdot 4^{k+1} + 9 = 4^{k+2} + 9$ souhlasí.

Přepis: $f(n) = 4^{\log_2(n)+2} + 9 = (2^2)^{\log_2(n)} \cdot 4^2 + 9 = (2^{\log_2(n)})^2 \cdot 16 + 9 = 16n^2 + 9$ na M .

Důsledek: $a = 1$, $b = 2$, $d = 2 > \log_2(1) = 0$, proto $f(n) = \Theta(n^2)$.

10c.3: (i): $a = 2$, $b = 3$, $c = 0$, tedy $f(n) = \Theta(n^{\log_3(2)})$;

(ii): $a = 1$, $b = 3$, $c = 0$, tedy $f(n) = \Theta(n^{\log_3(1)}) = \Theta(1)$;

(iii): $a = 4$, $b = 2$, $c = 0$, tedy $f(n) = \Theta(n^{\log_2(4)}) = \Theta(n^2)$;

(iv): $a = 1$, $b = 3$, $d = 1 > \log_3(1) = 0$, tedy $f(n) = \Theta(n)$;

(v): $a = 3$, $b = 3$, $d = 1 = \log_3(3)$, tedy $f(n) = \Theta(n \log_2(n))$;

(vi): $a = 2$, $b = 4$, $d = 0 < \log_4(2) = \frac{1}{2}$, tedy $f(n) = \Theta(n^{\log_4(2)}) = \Theta(\sqrt{n})$;

(vii): $a = 8$, $b = 2$, $d = 1 < \log_2(8) = 3$, tedy $f(n) = \Theta(n^{\log_2(8)}) = \Theta(n^3)$;

(viii): $a = 4$, $b = 2$, $d = 2 = \log_2(4)$, tedy $f(n) = \Theta(n^{\log_2(4)} \log_2(n)) = \Theta(n^2 \log(N))$;

(ix): $a = 4$, $b = 2$, $d = 3 > \log_2(4)$, tedy $f(n) = \Theta(n^3)$.

10c.4: (i): $f(n) = f(\frac{n}{2}) + 1$, $a = 1$, $b = 2$, $d = 0 = \log_2(1)$, proto $f(n) = \Theta(n^0 \log_2(n)) = \Theta(\log_2(n))$.

Poznámka: Je to rozhodně lepší než n při umocňování podle definice.

Bonus: Každé $m \in \mathbb{N}$ lze napsat jako součet mocnin typu 2^i (neboli zapsat ve dvojkové soustavě), nejvyšší použitá mocina 2^k je dána jako $k = \lfloor \log_2(m) \rfloor$. Mocninu x^m pak získáme vynásobením těch z mocnin $x^1, x^2, x^4, \dots, x^{2^k}$, které se objeví v rozkladu. Výpočet nejvyšší mocniny nás dle předchozího výpočtu stojí $\log_2(2^k) = k$ násobení, těch mocnin je celkem $k+1$ a v nejhorsím násobíme všechny, což je dalších k násobení. Nejhorší scénář pro výpočet x^m tímto způsobem tedy dává $2k = 2\lfloor \log_2(m) \rfloor$ násobení, tedy náročnost $\Theta(\log_2(m))$ násobení. To je velmi pěkné.

(ii): $f(n) = 2f(\frac{n}{2}) + n$, $a = 2$, $b = 2$, $d = 1 = \log_2(2)$, proto $f(n) = \Theta(n^1 \log_2(n)) = \Theta(n \log_2(n))$.

Poznámka: Je to dost lepší než oněch n^2 .

(iii): $f(n) = 2f(\frac{n}{2}) + 2$, $a = 2$, $b = 2$, $d = 0 < \log_2(2) = 1$, proto $f(n) = \Theta(n^{\log_2(2)}) = \Theta(n)$.

Poznámka: Je to tedy řádově stejně rychlé jako přímý útok, moc jsme si nepomohli.

10c.5: (i): $f(4) = 3f(2) + 13 = 3 \cdot 1 + 13 = 16$, $f(16) = 3f(4) + 13 = 3 \cdot 16 + 13 = 61$.

(ii): $F(m) = f(2^m) = 3f(2^{m/2}) + 13 = 3F\left(\frac{m}{2}\right) + 13$. Důsledek: $a = 3$, $b = 2$, $d = 0 < \log_2(3)$, proto $F(m) = \Theta(m^{\log_2(3)})$, to dává $f(2^m) = \Theta(m^{\log_2(3)})$ a proto $f(n) = \Theta([\log_2(n)]^{\log_2(3)})$.

10c.6: Dosazení: $(f_1 + f_2)(n) = f_1(n) + f_2(n) = af_1\left(\frac{n}{b}\right) + g_1(n) + af_2\left(\frac{n}{b}\right) + g_2(n) = a(f_1 + f_2)\left(\frac{n}{b}\right) + (g_1(n) + g_2(n))$.

10c.7: 1) Nechť f je řešení. Definujte $f_h = f - f_p$, dosadit:

$$f_h(n) = f(n) - f_p(n) = af\left(\frac{n}{b}\right) + g(n) - af_p\left(\frac{n}{b}\right) - g(n) = a(f - f_p)\left(\frac{n}{b}\right) = af_h\left(\frac{n}{b}\right).$$

2) Nechť f_h je homogenní řešení. Pak podle věty o superpozici $f_p + f_h$ řeší rovnici $f(n) = af\left(\frac{n}{b}\right) + g(n) + 0$.

10c.8: Podle příkladu je $N = \{f(b^k) = a^k f(1)\} = \{f(b^k) = c \cdot a^k; c \in \mathbb{R}\}$, jde tedy o jednorozměrný vektorový prostor s bází danou funkcí $f(b^k) = a^k$.

10d. Bonus: Generující funkce

Zde si představíme zajímavou alternativní metodu řešení rekurentních rovnic. Bude založena na materiálu vyloženém v kapitole , kde jsme se seznámili s řadami a mocninnými řadami. Začneme tím, že se na ty věci podíváme trochu z jiné strany.

Je jasné, že existuje vzájemně jednoznačně vztah mezi posloupnostmi a mocninnými řadami. Každá posloupnost $\{a_k\}_{k=0}^{\infty}$ dává vzniknout odpovídající mocninné řadě $\sum_{k=0}^{\infty} a_k x^k$ a naopak, pokud si z mocninné řady vytáhneme její koeficienty a_k , vytvoří nám posloupnost. Nás budou následně zajímat jen řady, které se chovají rozumně vzhledem ke konvergenci. Uvažujme proto množinu M všech posloupností $\{a_k\}_{k=0}^{\infty}$ takových, že pro ně odpovídající řady $\sum_{k=0}^{\infty} a_k x^k$ konvergují na nějakém nedegenerovaném intervalu, jinými slovy, mají poloměr konvergence $\varrho > 0$.

Pokud má nějaká řada netriviální obor konvergence, tak na tomto intervalu definuje jistou funkci. To znamená, že vlastně ke každé posloupnosti z M dostáváme určitou funkci f , která je definovaná na nějakém intervalu $(-\varrho, \varrho)$ pro $\varrho > 0$. Vzniká nám tím přiřazení neboli zobrazení. Označme jako N množinu všech funkcí, které dostaneme pomocí posloupností z M , nechť T je příslušné zobrazení $M \mapsto N$, které ke každé posloupnosti $\{a_k\}_{k=0}^{\infty}$ přiřadí funkci danou jako součet mocninné řady $\sum_{k=0}^{\infty} a_k x^k$.

Příklad 10d.a: Víme, že $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ na $(-1, 1)$. Tato řada se dá napsat jako $\sum_{k=0}^{\infty} 1 \cdot x^k$ neboli odpovídá posloupnosti $\{1\}_{k=0}^{\infty} = (1, 1, 1, \dots)$.

Můžeme tedy tvrdit, že $\{1\}_{k=0}^{\infty} \in M$ a $T(\{1\}) = \frac{1}{1-x}$.

Další známý rozvoj je pro exponenciálu, $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. Vidíme tedy, že $\{\frac{1}{k!}\} \in M$ a $T(\{\frac{1}{k!}\}) = e^x$.

Naopak se pomocí analytických metod snadno ukáže, že řada $\sum_{k=0}^{\infty} k! x^k$ konverguje jen pro $x = 0$, tedy má poloměr konvergence $\varrho = 0$. Proto posloupnost $\{k!\}_{k=0}^{\infty}$ neleží v M a v této teorii na ni tedy nedosáhneme.
△

Jaké vlastnosti můžeme od našich nových pojmu čekat? Hodně příjemné a občas také zajímavé.

Věta 10d.1.

Množina M je vektorový prostor.

Zobrazení T je lineární.

Důkaz (poučný): Vezměme dvě posloupnosti $\{a_k\}_{k=0}^{\infty}, \{b_k\}_{k=0}^{\infty} \in M$. Odpovídající řady pak konvergují na netriviálním intervalu, označme funkce odpovídající příslušným řadám jako $f(x) = \sum_{k=0}^{\infty} a_k x^k$, $g(x) = \sum_{k=0}^{\infty} b_k x^k$.

V našem novém značení to znamená, že $f(x) = T(\{a_k\})$ a $g(x) = T(\{b_k\})$.

Součet posloupností je definován jako $\{a_k\}_{k=0}^{\infty} + \{b_k\}_{k=0}^{\infty} = \{a_k + b_k\}_{k=0}^{\infty}$ a této posloupnosti odpovídá řada $\sum_{k=0}^{\infty} (a_k + b_k)x^k$. Věta říká, že tato řada konverguje a její součet je roven $\sum_{k=0}^{\infty} a_k x^k + \sum_{k=0}^{\infty} b_k x^k = f(x) + g(x)$. Tím se dozvídáme dvě věci. Za prvé, součet $\{a_k\}_{k=0}^{\infty} + \{b_k\}_{k=0}^{\infty}$ leží v M , tedy tato množina je uzavřená na sčítání. Za druhé, tomuto součtu odpovídá funkce $f(x) + g(x)$, což se dá zapsat jako $T(\{a_k\} + \{b_k\}) = T(\{a_k\}) + T(\{b_k\})$ a jedna podmínka linearity pro T je splněna.

Nechť $c \in \mathbb{R}$. Posloupnost $c\{a_k\}_{k=0}^{\infty} = \{ca_k\}_{k=0}^{\infty}$ odpovídá řada $\sum_{k=0}^{\infty} ca_k x^k$. Podle Věty tato řada konverguje a její součet je $\sum_{k=0}^{\infty} ca_k x^k = cf(x)$, proto $c\{a_k\}_{k=0}^{\infty} \in M$ a $T(\{ca_k\}) = cf(x) = cT(\{a_k\})$, čímž se potvrdila druhá podmínka linearity T .

Dokázali jsme, že množina M je coby podmnožina vektorového prostoru všech posloupností uzavřená na sčítání a násobek konstantou, tudíž je M také vektorový prostor. \square

Linearita bude velice užitečná, díky ní už například hravě odvodíme, kam se posírají konstantní posloupnosti $\{a\}_{k=0}^{\infty} = (a, a, a, \dots)$:

$$T(\{a\}_{k=0}^{\infty}) = aT(\{1\}_{k=0}^{\infty}) = \frac{a}{1-x}.$$

Vztah mezi posloupností a funkcí je v některých aplikacích tak užitečný, že si zaslouží své jméno.

Definice.

Řekneme, že funkce f je **generující funkcií (generating function)** pro posloupnost $\{a_k\}_{k=0}^{\infty}$, jestliže na nějakém intervalu $(-\varrho, \varrho)$ pro $\varrho > 0$ platí $f(x) = \sum_{k=0}^{\infty} a_k x^k$.

Někdy se tomu také říká **vytvořující funkce**.

My teď potřebujeme udělat dvě věci. Jednak si vytvořit nějaký slovníček výrazů, které umíme transformovat, a za druhé rozšířit vlastnosti, protože linearita nám nebude stačit, budeme chtít dělat i jiné triky.

Abychom si udělali rozumný slovníček, zamyslíme se nejprve nad vlastnostmi T z trochu jiné strany. Již z definice je jasné, že T je na. Otázka, zda je prosté, je ovšem značně těžká a vyžaduje to tvrdou analytickou práci. Nakonec se ale ukáže, že když máme řady s rozdílnými koeficienty, tak už nutně musí dávat různé funkce. To znamená, že T je bijekce a tudíž můžeme pracovat i s její inverzí T^{-1} . Víme už například, že $T^{-1}\left(\frac{1}{1-x}\right) = \{1\}_{k=0}^{\infty}$. Tento oboustranný vztah budeme značit $\{1\} \leftrightarrow \frac{1}{1-x}$.

V kapitole jsme ukázali ještě jeden součet řady, ze kterého dostaneme další obousměrný vztah.

Fakt 10d.2. (slovník)

$$\begin{aligned} T(\{1\}_{k=0}^{\infty}) &= \frac{1}{1-x} \text{ neboli } (1, 1, 1, 1, \dots) \leftrightarrow \frac{1}{1-x}; \\ T(\{k+1\}_{k=0}^{\infty}) &= \frac{1}{(1-x)^2} \text{ neboli } (1, 2, 3, 4, \dots) \leftrightarrow \frac{1}{(1-x)^2}. \end{aligned}$$

Není to zrovna nejbohatší slovník, ale ve spojení s triky, které uvidíme vzápětí, nám to postačí. Dalším krokem jsou pravidla, která nám umožní pomocí slovníčku pracovat i s příbuznými výrazy. V zásadě jde jen o standardní manipulace s mocninnými řadami vyjádřené v našem novém jazyce. Nejprve si ukážeme jeden příklad, který snad ozajemní, že opravdu najde o nic jiného než dobrou práci s řadami, při které se často vyplatí si je napsat v dlouhém tvaru. To se nám bude hodit v důkazech, které přijdou.

Příklad 10d.b: Co víme o posloupnosti $(1, 0, 1, 0, 1, 0, 1, \dots)$? Odpovídá jí řada

$$x^0 + x^2 + x^4 + x^6 + \dots = \sum_{k=0}^{\infty} x^{2k} = \sum_{k=0}^{\infty} (x^2)^k = \frac{1}{1-x^2}.$$

Takže $(1, 0, 1, 0, 1, \dots) \in M$ a $T(1, 0, 1, 0, 1, \dots) = \frac{1}{1-x^2}$ neboli $(1, 0, 1, 0, 1, \dots) \leftrightarrow \frac{1}{1-x^2}$.

Dá se tato posloupnost zapsat nějak přesně? Nejjednodušší je specifikovat její členy coby $a_k = \begin{cases} 1, & k \text{ sudé}; \\ 0, & k \text{ liché}. \end{cases}$

Pokud chceme pěkný vzoreček, nabízí se trik, rozmyslete si, že to je vlastně posloupnost $\{\frac{1}{2}[1 + (-1)^k]\}_{k=0}^{\infty}$.

\triangle

Teď už se podívejme na pravidla, připomeňme si, že už jsme dokázali linearitu. Co ještě můžeme chtít s posloupností udělat? Kromě násobení celé posloupnosti konstantou $\{ca_k\}$ je možné také tuto konstantu umocňovat, tedy z posloupnosti $\{a_k\}$ vyrobit $\{c^k a_k\}$. Občas se hodí umět přejít k posloupnosti $\{ka_k\}$. Poslední významnou skupinou operací je posun v posloupnosti. Pokud posouváme členy doleva, tak dostáváme posloupnosti (a_1, a_2, a_3, \dots) , (a_2, a_3, a_4, \dots) , (a_3, a_4, a_5, \dots) a tak dále, rozmyslete si, že se toto dá zapsat vzorcem $\{a_{k+N}\}_{k=0}^{\infty}$, kde N udává, o kolik jsme členy posunuli.

Trochu obtížnější je posun doprava. Myslíme tím posloupnosti $(0, a_0, a_1, a_2, \dots)$, $(0, 0, a_0, a_1, \dots)$, $(0, 0, 0, a_0, \dots)$ a tak dále. Nabízí se zápis $\{a_{k-N}\}_{k=0}^{\infty}$, ale má to podstatný zádrhlel. Představme si posun od dva, naivní pokus by byl $\{a_{k-2}\}_{k=0}^{\infty}$, ale jak vypadá první člen této posloupnosti? Pro $k = 0$ dostáváme a_{-2} , kteréžto číslo vůbec neexistuje. Toto se řeší zavedením Heavisideovy funkce $H(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0, \end{cases}$ a pak uvažujeme posloupnost $\{a_{k-N}H(k-N)\}_{k=0}^{\infty}$. Pro indexy $k \geq N$ je $H(k-N) = 1$ a tudíž členy neovlivní, naopak pro $k < N$ je $H(k-N) = 0$ a to se bere tak, že první členy jsou automaticky nulové a na a_{k-N} už se ani nedíváme, čímž se elegantně vyhneme problémům.

Teď se podíváme, jak se tyto operace odrazí na našem přiřazení. V zásadě jde jen o aplikaci Věty na naší situaci.

Věta 10d.3. (gramatika)

Nechť $\{a_k\}_{k=0}^{\infty} \in M$. Pak platí následující:

- (i) Pro $c \in \mathbb{R}$ platí $T(\{c^k a_k\}_{k=0}^{\infty}) = T(\{a_k\}_{k=0}^{\infty})(cx)$;
- (ii) Pro $N \in \mathbb{N}$ platí $T(\{a_{k+N}\}_{k=0}^{\infty}) = \frac{1}{x^N} [T(\{a_k\}_{k=0}^{\infty}) - \sum_{k=0}^{N-1} a_k x^k]$;
- (iii) Pro $N \in \mathbb{N}$ platí $T(\{a_{k-N}H(n-N)\}_{k=0}^{\infty}) = x^N T(\{a_k\}_{k=0}^{\infty})$;
- (iv) $T(\{ka_k\}_{k=0}^{\infty}) = x[T(\{a_k\}_{k=0}^{\infty})]'$;
- (v) $T(\{(k+1)a_k\}_{k=0}^{\infty}) = [x T(\{a_k\}_{k=0}^{\infty})]'$.

Než to dokážeme, ukážeme si tatáž pravidla v jiném zápisu, který je pro mnoho lidí (ale ne všechny) uživatelsky přítlulnější. Zahrneme i linearitu, ať to máme všechno pohromadě.

Důsledek 10d.4. (pravidla)

Nechť $\{a_k\}_{k=0}^{\infty}, \{b_k\}_{k=0}^{\infty} \in M$, označme $f(x) = T(\{a_k\})$ a $g(x) = T(\{b_k\})$. Nechť $c \in \mathbb{R}$ a $N \in \mathbb{N}$. Pak platí následující.

- (1) $(a_0, a_1, a_2, \dots) + (b_0, b_1, b_2, \dots) \leftrightarrow f(x) + g(x)$;
- (2) $\{ca_k\} = (ca_0, ca_1, ca_2, \dots) \leftrightarrow cf(x)$;
- (3) $\{c^k a_k\} = (c^0 a_0, c^1 a_1, c^2 a_2, \dots) \leftrightarrow f(cx)$;
- (4) $(a_N, a_{N+1}, a_{N+2}, \dots) \leftrightarrow \frac{1}{x^N} [f(x) - a_0 - a_1 x - \dots - a_{N-1} x^{N-1}]$;
- (5) $(0, 0, \dots, 0, a_0, a_1, a_2, \dots) \leftrightarrow x^N f(x)$;
- (6) $\{ka_k\} = (0 \cdot a_0, 1 \cdot a_1, 2a_2, 3a_3, \dots) \leftrightarrow x[f(x)]'$;
- (7) $\{(k+1)a_k\} = (1 \cdot a_0, 2a_1, 3a_2, 4a_3, \dots) \leftrightarrow [x f(x)]'$.

Všimněte si, že druhý vzoreček ve slovníku (Fakt) vznikne z prvního aplikováním pravidla (7) z této věty.

Teď tato pravidla dokážeme, vždy si napíšeme řadu odpovídající upravené posloupnosti nalevo a pak z ní nějak zkusíme vyrobit řadu odpovídající původní posloupnosti. Použijeme značení $f(x)$ pro $T(\{a_k\})$, protože nám to ulehčí život. Pokud vám některé úpravy přijdou jako černá magie, zkuste si zvolit konkrétní hodnotu (třeba $N = 4$) a napsat si ty vzorečky v dlouhé formě namísto sum.

Důkaz (poučný): (i): Posloupnosti $\{c^k a_k\}$ odpovídá řada

$$\sum_{k=0}^{\infty} (c^k a_k) x^k = \sum_{k=0}^{\infty} a_k (c^k x^k) = \sum_{k=0}^{\infty} a_k (cx)^k = f(cx).$$

(ii): Posloupnosti $\{a_{k+N}\} = (a_N, a_{N+1}, a_{N+2}, \dots)$ odpovídá řada

$$\begin{aligned} \sum_{k=0}^{\infty} a_{k+N} x^k &= \sum_{k=0}^{\infty} a_{k+N} x^{k+N-N} = \sum_{k=0}^{\infty} a_{k+N} x^{k+N} \frac{1}{x^N} = \frac{1}{x^N} \sum_{k=0}^{\infty} a_{k+N} x^{k+N} \\ &= \left| \begin{array}{l} m = k + N \\ k \geq 0 \implies m \geq N \end{array} \right| = \frac{1}{x^N} \sum_{m=N}^{\infty} a_m x^m = \frac{1}{x^N} \left[\sum_{m=N}^{\infty} a_m x^m + \sum_{m=0}^{N-1} a_m x^m - \sum_{m=0}^{N-1} a_m x^m \right] \\ &= \frac{1}{x^N} \left[\sum_{m=0}^{\infty} a_m x^m - \sum_{m=0}^{N-1} a_m x^m \right] = \frac{1}{x^N} \left[f(x) - \sum_{m=0}^{N-1} a_m x^m \right]. \end{aligned}$$

(iii): Posloupnosti $\{a_{k-N}H(k-N)\} = (0, \dots, 0, a_0, a_1, a_2, \dots)$ odpovídá řada

$$\begin{aligned}\sum_{k=N}^{\infty} a_{k-N}x^k &= \left| \begin{array}{c} m = k - N \\ k = m + N \\ k \geq N \implies m \geq 0 \end{array} \right| = \sum_{m=0}^{\infty} a_m x^{m+N} = \sum_{m=0}^{\infty} a_m x^m x^N \\ &= x^N \sum_{m=0}^{\infty} a_m x^m = x^N f(x).\end{aligned}$$

(iv): Posloupnosti $\{ka_k\} = (0, a_1, 2a_2, 3a_3, \dots)$ odpovídá řada

$$\begin{aligned}\sum_{k=0}^{\infty} ka_k x^k &= x \sum_{k=0}^{\infty} a_k k x^{k-1} = x \sum_{k=0}^{\infty} a_k [x^k]' = x \sum_{k=0}^{\infty} [a_k x^k]' \\ &= x \left[\sum_{k=0}^{\infty} a_k x^k \right]' = x[f(x)]' .\end{aligned}$$

(v): Posloupnosti $\{(k+1)a_k\} = (a_0, 2a_1, 3a_2, 4a_3, \dots)$ odpovídá řada

$$\begin{aligned}\sum_{k=0}^{\infty} (k+1)a_k x^k &= \sum_{k=0}^{\infty} a_k (k+1)x^k = \sum_{k=0}^{\infty} a_k [x^{k+1}]' = \left[\sum_{k=0}^{\infty} a_k x^{k+1} \right]' \\ &= \left[x \sum_{k=0}^{\infty} a_k x^k \right]' = [x f(x)]' .\end{aligned}$$

□

Jsou dva možné přístupy k práci s těmito pravidly a generujícími funkcemi vůbec. Je možné ignorovat „vnitřnosti“ a při transformování mechanicky aplikovat slovník a gramatiku, kterou se uživatel prostě naučí nazepamět. Tento přístup funguje vcelku uspokojivě u rutinních příkladů a má malé nároky na přemýšlení takového uživatele. Selhává ovšem při setkání s příkladem, který nějak vybočuje, takových je samozřejmě spousta, možná většina.

Proto je perspektivnější také věci rozumět, chápát podstatu transformace i způsob, jak se k oněm pravidlům přichází. Jednak to dodá sebejistoty při práci s pravidly, druhak to představuje záchrannou kotvu pro případ, že by paměť nefungovala úplně spolehlivě, a hlavně je to jediný způsob, jak přistupovat k nerutinným příkladům.

Jako příklad použití gramatiky si dokážeme pár dalších zajímavých vzorečků, které se občas hodí.

Fakt 10d.5.

- (i) $(1, -1, 1, -1, 1, \dots) \leftrightarrow \frac{1}{1+x};$
- (ii) $(0, 1, 1, 1, 1, \dots) \leftrightarrow \frac{x}{1-x};$
- (iii) $\{a^k\} \leftrightarrow \frac{1}{1-ax};$
- (iv) $\left\{-\frac{1}{a^{k+1}}\right\} \leftrightarrow \frac{1}{x-a}.$

U (iv) není jasné, proč by někdo chtěl pracovat s tak ošklivou posloupností, ale na ten vzoreček je nutné se podívat z opačné strany: Bude se nám silně hodit umět transformovat funkce typu $\frac{1}{x-a}$.

Důkaz (poučný): (i) Použijeme pravidlo (3) s volbou $c = -1$ na vzorec (i) z Faktu . Nebo počkáme na důkaz (iii) a pak tam použijeme $a = -1$.

(ii) Pravidlo (5) na vzorec (i) z Faktu .

(iii) Pravidlo (3) na vzorec (i) z Faktu .

(iv) $\frac{1}{x-a} = \frac{1}{a} \frac{1}{\frac{x}{a}-1} = -\frac{1}{a} \frac{1}{1-\frac{1}{a}x} \leftrightarrow -\frac{1}{a} \left\{ \left(\frac{1}{a}\right)^k \right\}$ pomocí (iii), dál už se to snadno upraví.

□

Tím máme všechny nástroje pohromadě a je načase si ukázat, k čemu generující funkce mohou být.

Příklad 10d.c: Uvažujme rekurentní rovnici $a_{n+2} - a_n = 3 \cdot 2^n - 12$, $n \geq 0$ s počátečními podmínkami $a_0 = 1$, $a_1 = 0$.

Víme, že daná rovnice vlastně reprezentuje mnoho rovnic:

$$a_2 - a_0 = 3 \cdot 2^0 - 12$$

$$a_3 - a_1 = 3 \cdot 2^1 - 12$$

$$a_4 - a_2 = 3 \cdot 2^2 - 12$$

⋮

Vlastně tedy porovnáváme nekonečně mnoho čísel neboli postupně členy jistých posloupností. Na pravé straně vidíme posloupnost $\{3 \cdot 2^n - 12\}_{n=0}^{\infty}$. Na levé straně bude lepší si to rozložit, odečítá se tam neznámá posloupnost $\{a_n\}_{n=0}^{\infty}$ a přičítá posunutá posloupnost $\{a_{n+2}\}_{n=0}^{\infty}$. Danou rovnici lze tedy interpretovat jako rovnost mezi posloupnostmi:

$$\{a_{n+2}\}_{n=0}^{\infty} - \{a_n\}_{n=0}^{\infty} = \{3 \cdot 2^n - 12\}_{n=0}^{\infty}.$$

Když se posloupnosti rovnají, musí se rovnat i jejich obrazy vzhledem k zobrazení T , tedy díky linearitě a dalším pravidlům máme

$$\begin{aligned} T[\{a_{n+2}\}_{n=0}^{\infty} - \{a_n\}_{n=0}^{\infty}] &= T[\{3 \cdot 2^n - 12\}_{n=0}^{\infty}] \\ T[\{a_{n+2}\}_{n=0}^{\infty}] - T[\{a_n\}_{n=0}^{\infty}] &= T[\{3 \cdot 2^n\}_{n=0}^{\infty}] - T[\{12\}_{n=0}^{\infty}] \\ \frac{1}{x^2} [T(\{a_n\}_{n=0}^{\infty}) - a_0 - a_1 x] - T(\{a_n\}_{n=0}^{\infty}) &= 3T(\{2^n\}_{n=0}^{\infty}) - 12T(\{1\}_{n=0}^{\infty}). \end{aligned}$$

Bude se nám lépe pracovat, když si označíme $f(x) = T(\{a_n\})$. Díky počátečním podmínkám také umíme dosadit za a_0 a a_1 , to je ale náhodička, ono to tak musí vyjít už z principu (počet těch členů v prvním vzorci je roven stupni rovnice, což se rovná počtu počátečních podmínek). Také umíme vyhodnotit výrazy napravo. Dostáváme

$$\frac{1}{x^2} [f(x) - 1 - 0] - f(x) = 3 \frac{1}{1-2x} - 12 \frac{1}{1-x}.$$

Tuto rovnici teď vyřešíme pro neznámou funkci $f(x)$.

$$\begin{aligned} \frac{1}{x^2} [f(x) - 1] - f(x) &= 3 \frac{1}{1-2x} - 12 \frac{1}{1-x} \\ f(x) - 1 - x^2 f(x) &= \frac{3x^2}{1-2x} - \frac{12x^2}{1-x} \\ (1-x^2)f(x) &= 1 + \frac{3x^2}{1-2x} - \frac{12x^2}{1-x} \\ f(x) &= \frac{1}{1-x^2} + \frac{3x^2}{(1-x^2)(1-2x)} - \frac{12x^2}{(1-x^2)(1-x)}. \end{aligned}$$

Máme funkci, teď bychom ji potřebovali pomocí T^{-1} převézt zpět na posloupnost. Výrazy na pravé straně ale nemáme ve slovníku (ani rozšířeném). Pomůže algebra, dané výrazy lze rozložit na parciální zlomky. Detaily necháme do kursu analýzy, dostáváme

$$\begin{aligned} f(x) &= \frac{1}{(1-x)(1+x)} + \frac{3x^2}{(1-x)(1+x)(1-2x)} - \frac{12x^2}{(1-x)^2(1+x)} \\ &= \left(\frac{\frac{1}{2}}{1-x} + \frac{\frac{1}{2}}{1+x} \right) + \left(\frac{-\frac{3}{2}}{1-x} + \frac{\frac{1}{2}}{1+x} + \frac{1}{1-2x} \right) - \left(-\frac{9}{1-x} + \frac{6}{(1-x)^2} + \frac{3}{1+x} \right) \\ &= \frac{8}{1-x} - \frac{6}{(1-x)^2} - \frac{2}{1+x} + \frac{1}{1-2x}. \end{aligned}$$

Tento výraz již dokážeme převést na posloupnosti. Pomocí T^{-1} neboli čtení zprava doleva $u \leftrightarrow$ dostáváme

$$\{a_n\}_{n=0}^{\infty} = 8 - 6(k+1) - 2 \cdot (-1)^k + 2^k = 2^k - 2 \cdot (-1)^k - 6k + 2.$$

Přesně tento výsledek jsme dostali, když jsme tento problém řešili jako příklad pomocí algoritmu přes charakteristická čísla, přidruženou homogenní rovnici a podobně.

△

Příklad 10d.d: Najdeme obecné řešení rovnice $a_{n+2} - 3a_{n+1} + 2a_n = 0$.

Naše metoda potřebuje znalost počátečních podmínek, tak si tam dáme parametry $a_0 = p$, $a_1 = q$. Teď aplikujeme transformaci na rovnici, kterou vnímáme jako vztah o posloupnostech:

$$\{a_{n+2}\}_{n=0}^{\infty} - 3\{a_{n+1}\}_{n=0}^{\infty} + 2\{a_n\}_{n=0}^{\infty} = \{0\}_{n=0}^{\infty}.$$

Když si obraz $\{a_n\}_{n=0}^{\infty}$ označíme jako $f(x)$, dostáváme

$$\frac{1}{x^2} [f(x) - a_0 - a_1 x] - 3 \frac{1}{x} [f(x) - a_0] + 2f(x) = 0$$

$$[f(x) - p - qx] - 3x[f(x) - p] + 2x^2 f(x) = 0$$

$$f(x)[1 - 3x + 2x^2] = p + qx - 3px$$

$$f(x) = \frac{p + qx - 3px}{1 - 3x + 2x^2} = \frac{p + qx - 3px}{(2x-1)(x-1)}$$

$$f(x) = \frac{q-2p}{x-1} + \frac{p-q}{2x-1} = (2p-q) \frac{1}{1-x} + (q-p) \frac{1}{1-2x}.$$

Zpětná transformace dává $\{a_n\}_{n=0}^{\infty} = \{(2p - q) + (q - p) \cdot 2^n\}_{n=0}^{\infty}$.

Není to úplně nejtradičnější tvar. Vzhledem k tomu, že p, q jsou libovolné, můžeme si označit $v = q - p$, čímž dostáváme $a_n = (p - v) + v2^n$, a díky svobodě volby p pak ještě máme i $p - v$ s libovolnou hodnotou, můžeme jej označit u . Dostáváme tak řešení $a_n = u + v2^n$, což opět odpovídá řešení, které bychom dostali klasickým způsobem. U tohoto příkladu by se díky charakteristickým číslům získalo na jednom řádku.

△

Když porovnáme řešení Algoritmem a přes generující funkce, vidíme následující rozdíly:

- Máme-li najít obecné řešení, bývá klasická metoda přes vlastní čísla často výrazně rychlejší.
- Potřebujeme-li čistě partikulární řešení, pak je metoda pomocí transformace přinejmenším rovnocenná, může být i snažší.

Kritické body řešení přes generující funkce jsou dva. Jednak musíme být schopni vyřešit explicitně rovnici, kterou jsme odvodili pro funkci f , a pak ještě musíme umět tuto funkci transformovat zpět (parciální zlomky jsou možná pracné, ale zaručeně fungují, může být hůř). Výhodou této metody je, že je docela flexibilní. Na této úrovni to ovšem moc neuvidíme. Když si představíme, jaké pravé strany bychom uměli s našimi znalostmi transformovat, tak nám v zásadě zase vyjdou kvazipolynomy, na levé straně pak dostáváme rozumné rovnice v případech, kdy jde o lineární rekurentní rovnici. Jinými slovy, s tím, co zatím umíme, se transformací dají řešit v zásadě stejně rovnice jako pomocí metod z přecházejících sekcí. Potvrď to i následující příklad.

Příklad 10d.e: Vyřešíme rovnici $a_{n+1} - (n+1)a_n = 0$, $n \geq 0$ s počáteční podmínkou $a_0 = 1$.

Po transformaci pomocí T a s označením $T(\{a_n\}) = f(x)$ dostáváme

$$\begin{aligned} \frac{1}{x}[f(x) - 1] - [xf(x)]' &= 0 \\ \frac{1}{x}[f(x) - 1] - f(x) - xf'(x) &= 0 \\ f(x) - 1 - xf(x) - x^2f'(x) &= 0 \\ x^2f'(x) + (x-1)f(x) + 1 &= 0 \end{aligned}$$

Toto je dosud drsná diferenciální rovnice, takže se dál nedostaneme. Ani tato metoda nepomůže.

Zajímavé je, že transformace je užitečná v obou směrech. Již jsme mluvili o převodu rekurentní rovnice na rovnici s funkcemi, ale velice užitečný může být i převod nepříjemné diferenciální rovnice na rekurentní, jejíž řešení se pak dá třeba nějak odhadnout.

Zrovna náš příklad to ale neukáže. Pokud někdo potřebuje vyřešit diferenciální rovnici výše a převede si ji transformací na rekurentní rovnici, může relativně snadno uhodnout řešení $a_n = n!$. Bohužel, již jsme diskutovali, že tato posloupnost nepatří do M , tudíž ji pomocí T neumíme převést zpět na funkci a řešení diferenciální rovnice tak nedostaneme.

△

Výhody metody vyniknou ve chvíli, kdy si uživatel rozšíří slovníček a zásobu rozličných transformačních triků, navíc lze upravit i koncept generující funkce, aby obsáhl více posloupnosti. Tato rozšíření jsou mimo rozsah této kapitoly, jejím cílem je čtenáře dovést k pochopení, jakým mechanismem transformace fungují, bude pak pro něj snažší pracovat i s jinými. Obecný mechanismus je vždy stejný: 1. Daná rovnice se přetransformuje do zcela jiného jazyka, bonusem bývá, pokud se nepříjemné operace díky pravidlům převedou na příjemnější. 2. Nová rovnice se vyřeší. 3. Získané řešení se zase převede zpět.

Na závěr ukážeme páár zajímavých rozšíření slovníčku pro naši transformaci.

Fakt 10d.6.

- (i) Pro $n \in \mathbb{N}$ platí $\left\{ \binom{n}{k} \right\} = \left(\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \dots, \binom{n}{n}, 0, 0, \dots \right) \leftrightarrow (1+x)^n$.
- (ii) Pro $\alpha \in \mathbb{R}$ platí $\left\{ \binom{\alpha}{k} \right\} \leftrightarrow (1+x)^\alpha$.
- (iii) Pro $n \in \mathbb{N}$ platí $\left\{ \binom{n+k}{n} \right\} = (1, n+1, \binom{n+2}{n}, \binom{n+3}{n}, \dots) \leftrightarrow \frac{1}{(1+x)^{n+1}}$.

Jak tyto vzorečky naznačují, generující funkce jsou užitečným nástrojem i v kombinatorice.

Cvičení

Cvičení 10d.1: Najděte $T(\{k^2\})$.

Řešení:

10d.1: $\frac{x(x+1)}{(1-x)^3}$.

11. Kombinatorika (Počítání)

Kombinatorika je nauka o uspořádání věcí, její důležitou součástí je schopnost věci spočítat. Dnešní podobu lze vystopovat do 17. století a vzhledem k tomu, že počítání hraje zásadní roli v pravděpodobnosti, tak asi nepřekvapí, že významným impulsem tehdy byly záležitosti ohledně sázení. Čtenář se pravděpodobně již s některými kombinatorickými věcmi setkal, možná se dokonce učil rozeznávat variace od kombinací, s opakováním či bez. Tyto znalosti jsou bezesporu užitečné, ale mají jedno významné omezení: Mnoho situací se do těchto jednoduchých škatulek nevejde. Dlouhodobě výhodnější je tedy rozumět základním principům a umět si pomocí nich rozmyslet i komplikované situace. Jinými slovy, tato oblast rozhodně není algoritmická, není to ten typ příkladů, kde se stačí naučit dostatečný počet vzorečků a úspěch je zaručen. Spíše je to umění, kde hodně záleží na dobrém pochopení základů, zkušenostech a invenci.

Zde do této oblasti spíš jen nahlédneme. Nejprve se podíváme na jednodušší situace, které v zásadě odpovídají oném permutacím/variacím/kombinacím probíraným často na střední škole. V další kapitole zajdeme trochu dále. Protože u kombinatoriky záleží více než obvykle na zkušenostech, ukážeme víc než obvykle příkladů a cvičení.

11a. Základní principy

Začneme třemi základními principy, jejichž aplikací se dá v zásadě vyřešit většina běžných situací. Nebudeme je formulovat jako věty, už proto, že v nich nebudeme vždy používat přesnou matematickou terminologii. Vždy uvedeme dvě verze, jednu používající jazyka počítání, druhou používající jazyka množin.

! 11a.1. Sčítací princip

- Jestliže je možné jistý proces rozdělit na dva disjunktní případy, kdy si proces vždy vybere právě jeden z těchto případů, první případ je možno provést n_1 způsoby a druhý n_2 způsoby, pak je proces možno provést $n_1 + n_2$ způsoby.

Zobecnění: Jestliže je možno jistý proces rozdělit na N případů, kdy si proces vždy vybere právě jeden z těchto případů, a i -tý případ je možno provést n_i způsoby, pak je proces možno provést $\sum_{i=1}^N n_i$ způsoby.

- Uvažujme množinu M objektů. Jestliže existuje rozklad $M = \bigcup_{i=1}^N M_i$ (tedy M_i jsou navzájem disjunktní), pak $|M| = \sum_{i=1}^N |M_i|$.

Toto asi nevyžaduje bližšího komentáře, už od dětství víme, že když si hromádku rozdělíme na více menších, tak je stačí posčítat zvlášť a pak výsledky sečít. V kapitole 11b se podíváme na situaci, kdy množiny M_i nejsou disjunktní.

! 11a.2. Násobící princip

- Předpokládejme, že jistý proces lze rozložit do dvou po sobě následujících fází. Jestliže je první fázi možné udělat vždy n_1 způsoby a druhou vždy (nezávisle na výsledku první fáze) n_2 způsoby, pak je celý proces možno udělat $n_1 \cdot n_2$ způsoby.

Zobecnění: Je-li N fází, každá vždy n_i způsoby, pak je celý proces možno provést $\prod_{i=1}^N n_i$ způsoby.

- Uvažujme množinu M počítaných objektů. Jestliže je $M = M_1 \times M_2$, pak $|M| = |M_1| \cdot |M_2|$.

Všimněte si, že druhé vyjádření není tak univerzální jako to první, nutí nás totiž vybírat do druhé fáze stále tytéž objekty, zatímco první vyjádření připouští také možnost, že si v závislosti na výsledku první fáze měníme množinu voleb v druhé fázi; jediná podmínka je, že musí mít vždy stejnou velikost. I to by se dalo vyjádřit matematicky, ale bylo by to komplikovanější, zatímco my se zde snažíme o uchopení základních myšlenek. Přidáme ještě jeden princip, asi nejvíce samozrejmý a možná nejméně používaný z těch tří, ale někdy vysoce užitečný.

! 11a.3. Doplňkový princip

- Předpokládejme, že jistý proces lze provést dvěma způsoby, speciálním a nespeciálním. Pak je počet speciálních způsobů roven počtu všech způsobů provedení sníženým o počet nespeciálních způsobů provedení.
- Uvažujme množinu M počítaných objektů. Pak pro $M_1 \subseteq M$ platí $|M_1| = |M| - |M - M_1|$.

Ted' si všechny tyto principy ukážeme v akci.

Příklad 11a.a: V obchodě mají 6 různých druhů USB flashek. Čtyři kamarádi si je tam jdou kupit, každý jednu. Podíváme se na tuto situaci blíže.

a) Kolika způsoby mohou vejít do obchodu, pokud musí po jednom?

Jde o proces, který lze rozdělit na čtyři fáze. Jako prvního vstupujícího si můžeme vybrat ze čtyř. Tato volba ovlivní, kdo konkrétně může být zvolen jako druhý vstupující, ale neovlivní zásadní parametr: Pro druhého si vždy vybíráme ze tří kandidátů. Opět nezávisle na tom, kdo šel první a druhý, si pro třetího vybíráme ze dvou (i když pokaždé jiných, ale vždy dvou). Na čtvrté místo pak už je jen jedna volba. Je vidět, že jde přesně o situaci z násobícího principu, proto počet způsobů je $4 \cdot 3 \cdot 2 \cdot 1 = 4! = 24$.

Tomuto typu situace, kdy jen měníme pořadí určité množiny objektů, říkáme permutování, a každému jejich konkrétnímu uspořádání říkáme permutace. Obecný vzorec evidentně bude, že existuje $n!$ permutací n různých objektů.

Poznámka: Pokud nemusí po jednom, jde o řádově těžší problém, viz příklad 11a.l.

b) Vešli do obchodu. Kolika různými způsoby si mohou vybrat flashky za předpokladu, že od každého druhu je jich dostatečný počet a zajímá nás, kdo si vybral jakou?

Kamarády můžeme očíslovat a nechat je vybírat jednoho po druhém. Každý z nich má na výběr ze šesti typů, jde tedy zase o násobící princip a odpověď zní $6 \cdot 6 \cdot 6 \cdot 6 = 6^4 = 1296$.

Zde to lze vidět i přes kartézský součin: V okamžiku, kdy kamarády očísloujeme, se vlastně ptáme na množství vektorů o čtyřech souřadnicích, které lze vytvořit, když máme 6 voleb pro každou souřadnici.

Formálně situaci, kdy záleží na tom, kdo si co vybere, říkáme „záleží na pořadí“ výběru. Tomu, že se tentýž druh může vyskytnout vícekrát, říkáme „volba s opakováním“. Naše úvahy tedy vedou ke konstatování, že když vybíráme k -krát z n různých objektů, s opakováním a na pořadí záleží, pak je počet možných výsledků n^k .

c) Kolika různými způsoby si mohou vybrat flashky tak, aby měli každý jinou?

Zase jde o proces, který lze rozložit do fází. První kamarád má na výběr 6 možností. Tím ovšem omezí výběr druhého, ale ať už si první vybere cokoliv, druhý má na výběr vždy pět možností. Podobně pak třetí má jen čtyři a čtvrtý tři, konkrétní možnosti se vždy liší podle toho, co si vybrali ti předtím, ale důležité je, že počty jsou stejné. Je tedy $6 \cdot 5 \cdot 4 \cdot 3 = 360$ možných výběrů dle zadání.

Jak by se takový výsledek dal zapsat kompaktně a ještě tak, aby se v něm objevily parametry 6 a 4? Takto:

$$6 \cdot 5 \cdot 4 \cdot 3 = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = \frac{6!}{2!} = \frac{6!}{(6-4)!}.$$

Obecně když vybíráme k -krát z n různých objektů, na pořadí záleží a bez opakování, pak je to možné provést $\frac{n!}{(n-k)!}$ způsoby.

d) Kolika různými způsoby si mohou vybrat flashky tak, aby se některá opakovala?

Při přímém útoku jde o dosti komplikovanou úlohu, která nezapadá do žádného z principů. Řekněme, že necháme prvního vybrat. Kolik voleb má ten druhý? To záleží na tom, jestli se rozhodne volbu prvního opakovat nebo ne. Tím se situace rozpadne na dva disjunktní případy (sčítací princip), ale nebude to tak jednoduché, protože ti další už se opakovat nemusí, ale také mohou, navíc není vyloučeno, že se některá flashka objeví až třikrát či čtyřikrát. Pokud bychom to tedy chtěli prozkoumat, vzniká mnoho křížovatek a situace se brzy stává nepřehlednou. Budeme proto hledat alternativu, přesto je užitečné poznamenat, že někdy je třeba se s takovouto situací poprat, vrátíme se k tomu blíže v příkladu 11a.l.

Další možný přístup je z pohledu flashky, kdy si řekneme, která se vybere vícekrát, tím se nám situace rozdělí na šest případů. Hlavním problémem zde je, že tyto případy nejsou disjunktní, protože se může stát, že se vyberou dvě různé flashky opakován. Sčítací princip proto nelze aplikovat. Opět je užitečné poznamenat, že pokud bychom nenalezli lepší alternativu a museli tuto situaci dořešit, pak to lze udělat pomocí pokročilých metod z příští kapitoly, jmenovitě by se použil princip inkluze a exkluze.

Tím se dostaváme k optimálnímu řešení. Klíčem je všimnout si, že opakování výběru flashek je přesně opak k situaci, kdy se žádná neopakuje, takže lze použít doplnkový princip a výsledky z částí b) a c). Počet možných výběrů dle zadání je tedy $6^4 - \frac{6!}{2!} = 1296 - 360 = 936$.

Všimněte si, že tato metoda je sice příjemná, ale nelze na ni spoléhat. Co kdybychom měli kamarádů více a zeptali bychom se, kolik výběrů opakuje dvě a více flashek? Opakem by pak byly výběry, kde se opakuje nejvýše jedna flashka, obě úlohy by tedy nezapadal do žádné standardní situace a vyžadovaly by individuální přístup. Jinými slovy, ačkoliv jsme teď první dva postupy zavrhl, jsou situace, kdy se vyplatí umět je dotáhnout do konce, protože už nebude snadná alternativa.

e) Kolika způsoby si mohou vybrat flashky tak, aby se žádná neopakovala, když je nám jedno, kdo má kterou?

Interpretace: Vyberou si flashky, a protože je chtějí platit dohromady, vysypou je před prodavače na jednu hromádku. Kolik různých hromádek, ve kterých se flashky neopakují, může prodavač vidět?

Vymyslet to přímo je poněkud komplikovanější, protože k tomu, abychom mohli použít násobící princip, nám chybí rozlišení na fáze. Proto je asi nejjednodušší si tam pořadí uměle dodat a pak jej zase odebrat. Když se tedy budeme soustředit i na to, v jakém pořadí jsou flashky na tu hromádku pokládány, pak jde o problém z části c) a víme, že takových možností je $\frac{6!}{2!}$. Označme si množinu těchto výběrů pracovně M , jde o množinu uspořádaných čteveřic.

Když si pak pořadí odmyslíme, tak nastane problém, protože mnohé (dokonce všechny) situace jsme započítali vícekrát. Například volby $(1, 3, 5, 2)$ a $(3, 2, 5, 1)$ se v okamžiku, kdy na pořadí výběru nezáleží, smrsknou do jedné možnosti $\{1, 2, 3, 5\}$ (použili jsme množinu, ať zdůrazníme irelevanci pořadí). Je v tom nějaká pravidelnost? Ano, uspořádané výběry se smrskávají na jeden neuspořádaný vždy po stejných počtech. Je to dobře vidět, když se na to podíváme z druhé strany: Každá hromádka čtyř různých flashek nám dá $4! = 24$ permutací neboli 24 uspořádaných výběrů. To znamená, že množina uspořádaných výběrů se přirozeně rozpadá do skupinek (disjunktních) o velikosti $4! = 24$ a každá z těchto skupinek výběrů pak dává jen jednu hromádku. Počet různých hromádek je tedy roven počtu těchto skupin výběrů, což je $\frac{\frac{6!}{2!}}{4!} = \frac{360}{24} = 15$.

Zase to budeme chtít zapsat pomocí vstupních dat 6 a 4, dostaneme $\frac{6!}{4!(6-4)!}$.

Situace, kdy vybíráme k -krát z n různých objektů, bez opakování a také bez pořadí, je v kombinatorice velice častá a vyplatí se ji naučit rozpoznávat. Rovněž se bohatě vyplatí pamatovat si příslušný vzorec $\frac{n!}{k!(n-k)!}$, aby si ho člověk nemusel pořád znova vymýšlet, na rozdíl od vzorců z b) a c) už není zjevný na první pohled. Budeme se mu ještě v této kapitole věnovat.

Užitečné je také spojení mezi situacemi c) a e), které jsme tu odvodili. Funguje totiž v obou směrech, takže pokud si člověk pamatuje situaci z této části, může pomoci ní řešit příklady typu c). Postupuje se pak naopak: Chceme-li rozdělit rozdílné flashky mezi kamarády, pak nejprve rozhodneme, které jim vůbec dáme, to je první fáze s $\frac{n!}{k!(n-k)!}$ možnostmi, a vybrané flashky pak mezi ně v nějakém pořadí rozdáme neboli je permutujeme, to je druhá fáze s $k!$ možnostmi. Podle násobícího principu je celkový počet možností $\frac{n!}{k!(n-k)!} \cdot k! = \frac{n!}{(n-k)!}$, viz c).

f) Kolik různých hromádek může prodavač vidět, když už není žádná podmínka na opakování, takže se mohou i nemusí opakovat?

Jinými slovy, kolika způsoby je možno vybrat čtyři flashky, když je povoleno opakování a na pořadí nezáleží?

Toto je nejtěžší situace z těch základních a selský rozum těžko pomůže, pokud už člověk dopředu neví, co má dělat.

Hlavním problémem tady je, že když zkusíme jednu konkrétní hromádku rozdělit mezi kamarády ve snaze zopakovat postup z části e), tak už to nedopadne vždy stejně. Je pořád pravda, že hromádka $\{1, 2, 3, 5\}$ dává $4! = 24$ různých výběrů pro kamarády, ale hromádka $\{1, 1, 1, 2\}$ už dá jen čtyři různé výběry (kdo dostane dvojku?) a hromádka $\{6, 6, 6, 6\}$ už dokonce jen jeden (každý si vezme šestku). To znamená, že když si množinu M všech výběrů flashek kamarády rozdělíme do skupin podle toho, jaké pak vytvoří hromádky, tak ty skupiny nebudou vždy stejně velké, jinými slovy, počet těchto skupin nezjistíme pomocí velikosti M a dělení jako v části e).

Je to tedy slepá ulička a je na to třeba jít jinak. Tato situace je z těch základních s přehledem nejméně intuitivní, většina lidí nad ní raději moc nepřemýší a rovnou si pamatuje příslušný vzorec, což vám doporučujeme. Pro odvážné přijde jeho odvození.

Klíčová úvaha vypadá následovně: Když na pořadí nezáleží, tak si ty flashky můžeme vždy seřadit podle nějakého kritéria, třeba podle toho, jak jsme si je očíslovali. Dostáváme tak hromádky typu $\{1, 1, 1, 1\}$, $\{1, 1, 3, 6\}$ atd. Každý takovýto výběr je tedy jednoznačně určen rozhodnutím, kolik míst mezi těmi čtyřmi zaberou jedničky, kolik dvojky, kolik trojky a podobně. To se dá realizovat následovně. Vytvoříme si ukazatele, které ukazují, kam až v té čtyřce míst půjdou jedničky, kam až půjdou dvojky a podobně, šestky ukončovat nemusíme, takže je celkem pět ukazatelů změn typu flashky. Pak ještě potřebujeme ukazatele na ta místa k obsazení, tedy celkem $(6-1)+4=9$ ukazatelů. Tvrdíme, že těmito ukazateli se již výběry jednoznačně určí.

Označíme-li písmenem T ukazatel typu a písmenem M ukazatel místa, tak například hromádka $\{1, 1, 1, 1\}$ by se kódovala MMMMMTTTT, neboli první typ končí až po všech čtyřech místech, hromádka $\{1, 1, 3, 6\}$ by se kódovala MMTTMTTTM (po dvou místech ukončíme jedničky a rovnou i dvojky, pak jedno místo trojek a ukončíme trojky, čtyřky i pětky) a hromádka $\{2, 3, 3, 5\}$ by se kódovala TMTMMTTMT. Důležité je, že to funguje i naopak, kdykoliv si vezmeme nějaké pořadí pěti T a čtyř M, tak nám to dá určitou hromádku flashek.

Počet možných hromádek je tedy dán počtem možných uspořádání čtyř M a pěti T, což zjistíme snadno dalším trikem: Podíváme se na to tak, že z devíti míst se vybírají čtyři, nesmíme opakovat a na pořadí nezáleží, čili podle e) víme, že se to dá dělat $\frac{9!}{4!(9-4)!}=126$ způsoby.

Poučení: Když vybíráme k -krát s opakováním z n různých objektů a bez pořadí, tak je to možno udělat $\frac{(n-1+k)!}{k!(n-1)!}$ různými způsoby.

△

Tímto příkladem jsme probrali klasické čtyři základní situace. Než si je shrneme, uděláme si užitečnou definici.

! Definice

Nechť $k \leq n \in \mathbb{N}_0$. Definujeme jejich **kombinační číslo** nebo **binomický koeficient** jako

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Čteme to „ n nad k “.

Let $k \leq n \in \mathbb{N}_0$. We define their **binomial coefficient** or **combinatorial number** as $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. We read it “ n choose k ”.

Ačkoliv to z definice nemusí být zjevné, ze způsobu, kterým jsme k tomuto číslu došli, hned vyplývá, že kombinační čísla jsou vždy přirozená čísla (viz také cvičení 11c.4).

Výraz z definice je nepraktický, protože faktoriály jsou velice drahé na výpočet. Proto bývá lepší nejprve zkrátit jeden z faktoriálů ze jmenovatele se začátkem faktoriálu v čitateli. Máme pak

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+2) \cdot (n-k+1)}{k!} = \frac{n \cdot (n-1) \cdots (k+2) \cdot (k+1)}{(n-k)!}.$$

Samozřejmě vždy volíme tu variantu, která dá méně výsledných činitelů, tedy krátíme ten větší faktoriál ve jmenovateli. Při ručním výpočtu pak můžeme doufat v další krácení.

Příklad 11a.b: Spočítáme nějaké kombinační číslo, třeba

$$\binom{9}{3} = \frac{9!}{3!(9-3)!} = \frac{9 \cdot 8 \cdot 7 \cdot 6!}{3!6!} = \frac{9 \cdot 8 \cdot 7}{3!} = \frac{9 \cdot 8 \cdot 7}{3 \cdot 2} = 3 \cdot 4 \cdot 7 = 84.$$

△

Možná jste si všimli, že jakmile se kombinační číslo přepíše na zlomek, ztrácí se rozdíl mezi $k!$ a $(n-k)!$. Potvrďme si to faktem a přidáme dvě další jednoduchá pozorování.

! Fakt 11a.4.

(i) Pro všechna $n \in \mathbb{N}_0$ platí $\binom{n}{0} = 1$.

(ii) Pro všechna $n \in \mathbb{N}$ platí $\binom{n}{1} = n$.

(iii) Nechť $k \leq n \in \mathbb{N}_0$. Pak platí $\binom{n}{k} = \binom{n}{n-k}$.

Důkazy jsou tak snadné, že je s důvěrou necháme jako cvičení 11a.61. Další vlastnosti kombinačních čísel najde čtenář v kapitole 11c.

Shrneneme si čtyři základní situace, které jsme si rozmysleli v příkladě 11a.a.

!

Věta 11a.5.

Uvažujme množinu o n různých prvcích.

(i) Je $n!$ způsobů, jak je seřadit (neboli je $n!$ permutací).

(ii) Jestliže na pořadí záleží a opakování není povoleno, pak je $\frac{n!}{(n-k)!} = \binom{n}{k} \cdot k!$ různých způsobů, jak vybrat k prvků z této množiny.

(iii) Jestliže na pořadí záleží a opakování je povoleno, pak je n^k různých způsobů, jak vybrat k prvků z této množiny.

(iv) Jestliže na pořadí nezáleží a opakování není povoleno, pak je $\binom{n}{k}$ různých způsobů, jak vybrat k prvků z této množiny.

(v) Jestliže na pořadí nezáleží a opakování je povoleno, pak je $\binom{n+k-1}{k}$ různých způsobů, jak vybrat k prvků z této množiny.

Ony dva parametry, zda se opakuje a zda na pořadí záleží, patří k tomu hlavnímu, co určuje metodu zpracování kombinatorické situace. Je důležité umět situace (ii)–(iv) rozpozнат a přinejmenším pro ty dvě poslední znát příslušné vzorce. Někdo si pamatuje vzorce i pro první dvě z nich, ale jak už jsme viděli, dá se bez toho obejít. Výběrem „uspořádaným“, kde na pořadí záleží, se říká **variace**, zatímco výběrem „neuspořádaným“, kde na pořadí nezáleží, říkáme **kombinace**. Zde to nebudeme příliš používat. Shrňme si to v tabulce.

	bez opakování	s opakováním
s pořadím (variace)	$\frac{n!}{(n-k)!}$	n^k
bez pořadí (kombinace)	$\binom{n}{k}$	$\binom{n+k-1}{k}$

Ted' se podíváme na některé aplikace těchto základních situací.

Příklad 11a.c: Kolik je možno vytvořit osmimístných hesel (password) skládajících se z písmen a číslic?

Každý znak je nezávislý jev, který je možno udělat $26 + 10 = 36$ způsoby, proto je možno vytvořit 36^8 hesel.

△

Příklad 11a.d: Adam, Bára a Cirda chtějí do divadla, hodlají sedět hned v první řadě, kde je 13 sedadel. Kolika způsoby se tam mohou rozesadit?

Tato úloha je neřešitelná, protože nemáme dostatek informací. Jmenovitě, potřebujeme vědět, zda se každý spokojí s jedním sedadlem či jich bude chtít více, popřípadě zda by jim naopak nevadilo sedět jeden druhému na klíně třeba Cirda je ještě malý(á).

a) Pokud přidáme předpoklad, že každý chce své sedadlo (jedno), pak jde o výběr z 13 míst, bez opakování, na pořadí záleží (chceme vědět, kdo kde sedí), tedy $\frac{13!}{10!} = 13 \cdot 12 \cdot 11 = 1716$.

b) Kdyby byli ochotni v nouzi i sdílet sedadlo(a), pak by šlo o výběr s opakováním, tedy $13^3 = 2197$.

c) Co kdyby chtěli sedět vedle sebe? Pak je třeba situaci rozložit na dva kroky. Nejprve vybereme trojici sedadel vedle sebe, kolik je možností? Vybráme z bloků 1–3, 2–4, …, 11–13, těch je 11. Na vybrané trojmísto se pak rozesadí, to jsou permutace tří lidí. Je tedy celkem $11 \cdot 3! = 66$ možností.

d) Co kdyby chtěli sedět vedle sebe a s Cirdou uprostřed? Zase je 11 možností na trojku, ale pak už je dáno, kde sedí Cirda, jediná volba je, kdo sedí na levém a kdo na pravém konci, tedy 2 možnosti. Celkem je tedy $11 \cdot 2 = 22$ možností.

△

Příklad 11a.e: Kolik permutací písmen $ABCDEFGH$ obsahuje slovo $DECH$?

Toto se udělá jednoduchým trikem, prostě se $DECH$ vezme jako jeden celek, který se spolu s ostatními čtyřmi písmenky permutuje, takže celkem permutujeme pět věcí. Možností je tedy $5! = 120$.

△

Příklad 11a.f: Kombinatorika je zásadním nástrojem pro lidi zabývající se seriózně hraním karet.

Pro hráče bridge je základní úvaha, že ze standardního balíčku 52 karet je možno dostat 13 karet přesně

$$\binom{52}{13} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 \cdot 45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40}{13!} = 635013559600 \sim 6.4 \cdot 10^{11}$$

způsoby (vybíráme 13 z 52, bez opakování, na pořadí v ruce nezáleží).

Příznivce hry poker zase zajímá, že dostat z tohoto balíčku pět karet je možno $\binom{52}{5} = 2598960 \sim 2.6 \cdot 10^6$ způsoby.

△

Příklad 11a.g: Kolik různých balíčků bonbónů (ty jsou tam volně ložené) je možné vytvořit, když do balíčku vybíráme 10 bonbónů ze tří druhů, přičemž od každého druhu je k dispozici dostatek kusů?

Vybíráme desetkrát z tříprvkové množiny, výběr můžeme opakovat a na pořadí nezáleží, protože bonbóny se pak stejně budou v balíčku volně míchat. Je to ta nejobtížnější ze čtyř základních situací, proto si vzorec pamatujieme: Je možné udělat $\binom{3+10-1}{10} = \binom{12}{10} = \frac{12!}{10!2!} = \frac{12 \cdot 11}{2} = 66$ různých balíčků.

△

Příklad 11a.h: Uvažujme binární řetězce o délce 8 (bajty).

a) Kolik jich je?

Pro každou pozici vybíráme nezávisle ze dvou hodnot 0 a 1, tedy $2^8 = 256$ řetězců.

b) Kolik z nich obsahuje přesně tři jedničky?

Zde vybíráme, na které pozice jedničky dáme, a na pořadí výběru nezáleží (říct, že jedničky mají být na pozicích 1, 2 a 6, vyjde následně jako říct, že mají být na pozicích 2, 6 a 1). Takže vybíráme z osmi míst, bez opakování a bez pořadí, tedy $\binom{8}{3} = \frac{8!}{3!5!} = \frac{8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3} = 56$ řetězců.

c) Kolik z nich obsahuje nejvýše tři jedničky?

Toto je snadné, stačí posčítat možnosti pro žádnou, jednu, dvě či tři jedničky (jde o navzájem disjunktní situace):

$$\binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \binom{8}{3} = 1 + 8 + 28 + 56 = 93.$$

d) Kolik z nich obsahuje alespoň tři jedničky?

Podobný postup jako v c) vede na $\sum_{k=3}^8 \binom{8}{k}$ řetězců. Zde ale bude jednodušší přejít k opačnému jevu neboli nejvíce dvěma jedničkám, dostaneme

$$2^8 - [\binom{8}{0} + \binom{8}{1} + \binom{8}{2}] = 256 - 1 - 8 - 28 = 219.$$

e) Kolik z nich má stejně jedniček a nul?

Ty, které mají přesně čtyři nuly, $\binom{8}{4} = 70$ řetězců.

f) Kolik z nich má víc jedniček než nul?

Jedna možnost je spočítat $\binom{8}{5} + \binom{8}{6} + \binom{8}{7} + \binom{8}{8} = 93$.

Alternativa: Všech řetězců je 256, z nich 70 má stejný počet, zbývá $256 - 70 = 186$ řetězců, které mají buď víc jedniček nebo víc nul. Mezi těmito situacemi je zjevná symetrie (kdykoliv máme řetězec s větším počtem jedniček, zámenou $0 \leftrightarrow 1$ získáme řetězec s více nulami), proto polovina tohoto počtu má víc jedniček, tedy $\frac{1}{2}186 = 93$.

△

Teď se podíváme na několik teoretických situací.

! **Příklad 11a.i:** Dokážeme, že jestliže je A konečná množina, pak $|P(A)| = 2^{|A|}$.

1) Pro usnadnění označme $|A| = n$. Podmnožiny vytváříme tak, že se u každého prvku z A rozhodujeme, zda jej vezmeme či ne. Takže vlastně vybíráme n -krát z dvouprvkové množiny $\{\text{ano}, \text{ne}\}$, odpovědi se mohou opakovat a na pořadí záleží (chceme vědět, o kterém prvku říkáme ano či ne). Počet možností je tedy 2^n .

Chceme-li to odvodit ze základních principů, pak prostě bereme postupně prvky A a u každého jsou dvě možnosti rozhodnutí, těchto fází je n , celý proces je proto podle násobícího principu možno provést $2 \cdot 2 \cdots 2 = 2^n$ způsoby.

2) Aternativní řešení: Podmnožiny vznikají tak, že si z množiny vybereme několik prvků, na pořadí nezáleží, protože je pak dáváme do množiny, a opakovat nesmíme. Jde tedy o jasnou kombinaci záležitost, chybí už jen zjistit, kolik prvků vlastně máme vybrat. Odpověď zní, že to není dáno, musíme vyzkoušet všechny možnosti. Takže nejprve vybereme nic (prázdná podmnožina), to se dá jedním způsobem, což je mimochodem $\binom{n}{0}$, pak vybíráme jeden prvek, celkem $\binom{n}{1}$ možností, pak dva prvky, celkem $\binom{n}{2}$, a tak dále, až po výběr celé množiny, což se dá jediným způsobem, což je mimochodem $\binom{n}{n}$. Když to sečteme, dostaneme všechny možné způsoby výběru podmnožin: $\sum_{k=0}^n \binom{n}{k}$.

Samozřejmě je to stejný výsledek jako u prvního řešení, viz Důsledek 11c.7.

3) Alternativní řešení: Indukce na $|M|$.

(0) Pro nulaprkvkovou množinu \emptyset existuje jedna podmnožina \emptyset , takže $2^0 = 1$ souhlasí.

(1) Předpokládejme, že vztah platí pro n -prvkové množiny. Mějme množinu M o $n+1$ prvcích, zvolme si jeden z prvků m . Každá podmnožina z M buď v sobě m nemá, pak je to vlastně podmnožina množiny $M - \{m\}$ o n prvcích, těch je podle indukčního předpokladu 2^n , nebo v sobě m má a pak je to vlastně $Y \cup \{m\}$ pro nějakou podmnožinu $Y \subseteq M - \{m\}$, takže podmnožin M obsahujících m je také 2^n . Jde o disjunktní možnosti, proto je celkem $2^n + 2^n = 2 \cdot 2^n = 2^{n+1}$ podmnožin.

△

! Příklad 11a.j: Nechť A, B jsou konečné množiny.

a) Kolik je zobrazení z A do B ?

Každý prvek z A si může zcela svobodně vybrat, kam do B se pošle, což je $|B|$ možností. Vybíráme $|A|$ -krát, celkový počet výběrů je tedy $|B|^{|A|}$.

Závěr: Existuje $|B|^{|A|}$ zobrazení z A do B .

b) Kolik je prostých zobrazení z A do B ?

Teď každý prvek svou volbou omezí volbu prvků následujících. První prvek z A má na výběr $|B|$ možností, druhý už jen $|B| - 1$, třetí už jen $|B| - 2$ atd., ten poslední prvek má $|B| - |A| + 1$ možností. Násobící princip tak dává celkem $|B| \cdot (|B| - 1) \cdots (|B| - |A| + 1)$ možností výběru.

Všimněte si, že když $|A| > |B|$, tak toto číslo dává nulu, což je naprostě správně, pak žádné prosté zobrazení neexistuje. Budeme-li chtít odpověď vyjádřit kompaktně pomocí faktoriálů, tak si na to budeme muset dát pozor.

Závěr: Jestliže $|B| \geq |A|$, pak je $\frac{|B|!}{(|B|-|A|)!}$ různých prostých zobrazení z A do B , jinak není žádné.

c) Počet zobrazení na se určuje obtížně a necháme jej do příští kapitoly, viz Věta 11b.3.

d) Víme, že u konečných množin jsou bijekce možné jen v případě, že $|A| = |B|$. Pak už bijekce souhlasí s prostými zobrazeními a lze použít výsledek z b).

Závěr: Jestliže $|A| = |B|$, tak je $|B|!$ bijekcí z A na B .

Není to žádné překvapení, u bijekce je každý prvek z B napojen na nějaký (jediný) prvek z A , takže se to celé redukuje na otázku, v jakém pořadí se napojí neboli na počet permutací.

△

! Příklad 11a.k: Kolik má rovnice $x_1 + x_2 + x_3 = 13$ řešení splňujících $x_i \in \mathbb{N}_0$?

Pokud se budeme snažit nějak kombinatoricky přidělovat přirozená čísla do x_i , tak budeme mít velké problémy s uhlídáním jejich součtu. V případě malých čísel by ještě šel udělat rozbor všech případů (začít s $(0, 0, 13)$, $(0, 1, 12)$ a postupně se zkusit dopočítat k $(13, 0, 0)$, viz příklad 11a.v (ii) či algoritmus 11d.7), ale pro větší čísla to rozhodně není perspektivní, ono už zde s 13 by to bylo dost drsné.

Nejsnažší řešení spočívá v totální změně zorného úhlu. Nebudeme přidělovat čísla proměnným, ale proměnné číslům. Máme 13 jedniček a každá z nich si vybere, do které x_i půjde. Takže vybíráme třináctkrát ze tří možností x_1, x_2, x_3 , evidentně s opakováním a na pořadí nezáleží, protože je jedno, které konkrétní jedničky jdou třeba do x_1 , nás jen zajímá, kolik jedniček si tu x_1 vybral. Je to tedy zase ten nejméně intuitivní základní případ a pamatuji si, že je celkem $\binom{3+13-1}{13} = \binom{15}{13} = 105$ možných řešení.

Tento trik se změnou přidělování je docela užitečný. Postup je možné také zajímavě modifikovat, třeba takto:

b) Kolik existuje řešení rovnice $x_1 + x_2 + x_3 = 13$ splňujících $x_i \in \mathbb{N}_0$ a také $x_1 \geq 1, x_2 \geq 3, x_3 \geq 2$?

Trik: Nejprve rozdělíme napevno 6 jedniček tak, aby už x_i nabily svých minimálních nutných hodnot. Zbývajících 7 jedniček pak rozdělíme jako předtím (vybíráme pro každou ze tří proměnných), je tedy $\binom{3+7-1}{7} = \binom{9}{7} = 36$ takovýchto řešení.

Pro další varianty tohoto problému viz příklad 11b.d.

△

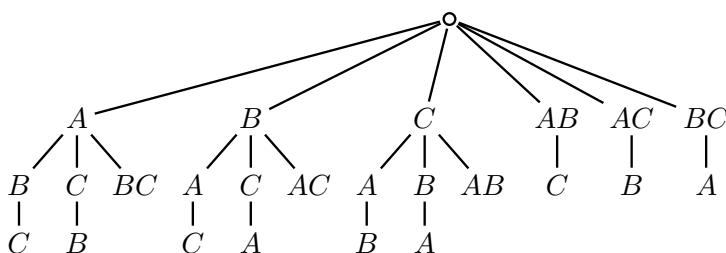
Mnoho kombinatorických situací ale takto přímočarých není, často je potřeba různé postupy kombinovat a nalézt správný přístup není snadné. Jako ilustraci teď ukážeme poněkud zajímavější příklady. Nejprve se ještě vrátíme k našim kamarádům a ukážeme jeden užitečný pohled na věc.

! Příklad 11a.l (pokračování 11a.a): Vrátíme se úplně na začátek, do situace, kdy kamarádi vcházejí do dveří. Kolika způsoby by mohli vejít, kdyby mohli vcházet nejen po jednom, ale také po dvou?

Tato otázka se vymyká základním čtyřem situacím, musíme se tedy vrátit k principům. Zkusme si to rozfázovat. Nejprve necháme vejít buď jednoho kamaráda (4 možnosti) nebo dvojici, což je výběr dvou ze čtyř bez opakování, $\binom{4}{2} = 6$ možností. Je tedy celkem 10 možností, kdo mohl vejít první, ale tím se proces zadrhne, protože nejsme schopni udat jednoznačný počet možností pro další fázi. Rekněme, že by jako druhý vešel jeden kamarád. Počet možností pro jeho výběr závisí na tom, z kolika vybíráme, ale to závisí na tom, co se stalo v prvním kole. Kdyby

na začátku vešel jeden, tak máme tři volby pro toho druhého, ale kdyby jako první vešli rovnou dva, tak zbývají už jen dvě volby pro toho, kdo vejde v druhé fázi. Není tedy splněna základní podmínka z násobícího principu, protože druhou fázi neprovádíme vždy stejným počtem možností.

V takové situaci se obvykle rozbor situace rozdělí na možnosti a každá se zkoumá samostatně, výsledky se pak dávají dohromady podle sčítacího principu. Množinu všech možností vstupu proto rozdělíme na podmnožinu těch, které začaly jedním člověkem, a na pomnožinu těch vstupů, které začaly dvojicí. Když ale v obou podmnožinách postoupíme dál, tak se rozpadnou i tyto podmnožiny na menší kousky, protože i pak máme možnost volby mezi jedním kamarádem či dvojicí. Situace se rychle stává nepřehlednou a v těchto situacích se vyplácí použít strom, který do úvah vnese řád, díky tomu je pak mimo jiné méně snadné nějakou možnost přehlédnout. Zkusme si nejprve udělat strom pro případ, že by kamarádi byli tři, označme si je A, B, C .



Všimněte si, že tři levé části stromu mají stejnou strukturu, stačí tedy spočítat velikost jedné části a vynásobit třemi. Podobně mají i pravé části stejnou strukturu. Stačí tedy nakreslit strom nikoliv se všemi možnostmi, ale se všemi typy možností, a u každého typu pak zjistit, kolik skutečných výběrů reprezentuje. To se dá dělat více způsoby, ukážeme zde dva.

1) První způsob bude asi trochu těžší na vysvětlení, ale v praxi mi přijde rychlejší. Je založen na tom, že využívá násobícího a sčítacího pravidla coby pravidel pro pohyb ve stromu. Zhruba řešeno, kde je ve stromu pravidelnost, tam násobíme, kde je nepravidelnost, tam sčítáme, a to na libovolné úrovni. Ukážeme to na našem původním příkladě, tedy na čtyřech kamarádech vcházejících po jednom či po dvou.

Zakreslíme si možnosti vstupu, ale místo konkrétního kamaráda budeme dávat proměnné, jmenovitě a jako znak pro libovolného kamaráda, b jako znak pro libovolného kamaráda jiného než a , c jako libovolného kamaráda odlišného od a či b atd. V každém okamžiku výběru si zároveň ve stromu uděláme poznámku o tom, kolik možností volby v tom kterém místě máme.

Jak se na to přišlo? Například první kamarád se vybíral ze čtyř možností, proto je nahoře vlevo u a čtyřka. Pokud šel po něm další kamarád, vybíral se ze tří (trojka u b), a když šli jako druzí dva, byly $\binom{3}{2} = 3$ možnosti (trojka u bc). Na druhou stranu pokud šla jako první dvojice, mohlo se to stát $\binom{4}{2} = 6$ způsoby (šestka u ab) a tak dále.

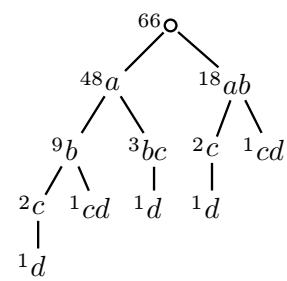
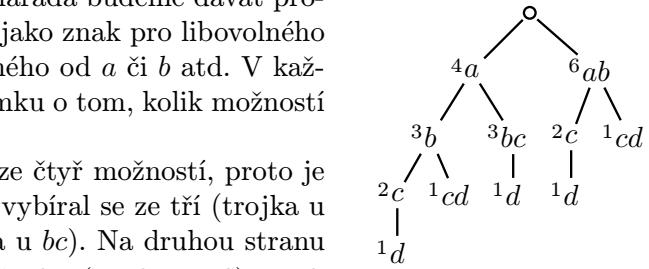
Celkový počet možností teď zjistíme tak, že procházíme strom zdola a aplikujeme násobící či sčítací pravidlo podle toho, jak to na různých místech stromu vypadá (pravidelně či nepravidelně). Vznikající čísla si zapíšeme do nové kopie našeho stromu, teď už čísla u větvě nepředstavují okamžitý počet možností, ale součet za celou část stromu níže. Dostáváme je takto.

Začneme tím d vlevo dole, tam je jedna možnost. Posuneme se nahoru, vidíme c , u něj dvojka, což říká, že ten úsek $c-d$ je opakován dvakrát. Představíme si tedy, jakoby od c vedly dolů dvě stejné části, které obě mají o úroveň níže už jen jednu část. To je pravidelnost, kterou řeší násobící princip, proto celá tato část stromu se může stát $2 \cdot 1$ způsoby.

Poučení: Když od nějakého bodu výběru vede dolů jen jedna cesta, tak to znamená, že je pod ním několikrát opakován tvar, stačí tedy vynásobit velikost tohoto tvaru s číslem u výběrového bodu.

Posuneme se nahoru. Vidíme b , ale také vidíme, že k němu vedou zdola dvě cestičky. Je zde tedy nevyváženosť, kterou řeší sčítací princip. Část stromu s vrcholkem c reprezentuje 2 možnosti a část označená cd má u sebe poznámku, že reprezentuje jednu kopii, tedy jednu možnost. Celkem tedy ty části stromu, které jsou pod zkoumaným b , reprezentují $2 + 1 = 3$ možnosti výběru. A tento (nevyvážený, ale již prozkoumaný) strom je u b zopakován třikrát (to je ta poznámka u b), tudíž zase násobící princip říká, že ten kus stromu, jehož typ začíná běčkem, představuje $3 \cdot 3 = 9$ možností.

Od tohoto b se přesuneme nahoru a jsme u a , ale zase vidíme, že se k němu dostaneme i odjinud. Je tedy čas zapamatovat si, že b -část stromu se může stát 9 způsobů, a podívat se na tu druhou část. Zase začneme zdola, je tam d s jedničkou a nad tím bc s trojkou. Podle násobícího principu se tedy celá část může stát $3 \cdot 1 = 3$

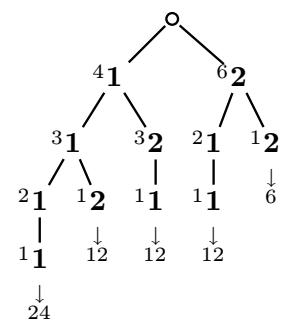


způsoby. Když to sečteme, tak vidíme, že část stromu pod a reprezentuje $9 + 3 = 12$ možností, a tato část stromu je zopakována čtyřikrát (viz poznámka u a). To znamená, že celá áčková část stromu představuje $4 \cdot 12 = 48$ možností. Právě jsme se dozvěděli, že pokud půjde jako první jeden kamarád, tak se to může stát přesně 48 způsoby.

Stejným postupem doplníme počty v druhé části stromu, začneme zdola a skončíme číslem 18 u ab . Dohromady je tedy $48 + 18 = 66$ možností, jak mohou kamarádi vejít po jednom či po dvou.

2) Alternativní způsob funguje tak, že ony počty možností v jednotlivých uzlech nekomplikujeme od listů ke kořeni, abychom tak dostali celkový počet, ale naopak shora dolů, kdy už nemusíme vybírat vhodné principy, ale prostě násobíme všechna čísla na konkrétní větví, čímž se dozvíme, kolika způsoby se daný konkrétní typ vstupu mohl udát. Zde si ukážeme také jinou verzi stromu, kdy si nepřešeme symbolicky výběry, ale jenom to, kolik lidí v jednotlivých fázích vejde, my už si budeme pamatovat, že nesmí dojít k opakování.

Ona čísla u jednotlivých listů lze zjistit i přímo běžnými kombinatorickými metodami, bez pomocí částečných číslíček. Například větev úplně vlevo odpovídá situaci, kdy vstoupí po jednom, což se může stát $4!$ způsoby. Větev o jedno vpravo je situace, kdy vstoupí jeden, jeden a nakonec zbylí dva. Pro výběr prvního jsou 4 možnosti, pro výběr druhého už jen 3 a zbývající dvojice prostě vejde, celkem tedy $4 \cdot 3 = 12$ způsobů tohoto vstupu. Další větev popisuje pořadí jeden-dva-jeden, jsou 4 možnosti pro prvního, další dva se vybírají ze tří a na pořadí nezáleží, což jsou $\binom{3}{2} = 3$ možnosti, a poslední už na výběr nemá, zbyl jeden, celkem $4 \cdot \binom{3}{2} = 12$ způsobů. Takto se projdou všechny větve ve stromu.



Když čísla sečteme, dostáváme závěr, že čtyři kamarádi mohou vcházet po jednom či po dvou celkem 66 způsoby.

Ať už se to zkouší tak či onak, moc efektivně to pořád nevypadá. Stačí si představit, že by lidé bylo 150 a mohli by vcházet po jednom, dvou či třech, strom by se pak musel kreslit na lodní plachtu. Bohužel, pro situace tohoto druhu neexistuje nějaký vzorec, do kterého by se dalo dosadit, jde o jednu z obtížnějších kombinatorických situací. Je vidět, že v kombinatorice stačí malá změna v zadání, aby se úloha změnila ze snadné (vchází po jednom) na velice obtížnou (vchází i po dvou). Situace, pro které chybí vzoreček, jsou běžné, takže vypisování všech možností je legitimní metoda.

U větších problémů pak nezbývá, než použít počítač, namísto kreslení stromů napsat algoritmus, kterým by se všechny možnosti prošly a spočítaly. Pak je samozřejmě třeba dát velký pozor na to, aby program opravdu počítal všechny možnosti a každou jen jednou. Letmo se na to podíváme v kapitole 11d.

△

! Příklad 11a.m: Ve třídě je 150 kluků a 40 holek. Chtějí poslat čtyřčlennou delegaci k přednášejícímu.

a) Kolika způsoby ji mohou vybrat?

Vybírájí čtyři ze $150 + 40$, na pořadí nezáleží a evidentně bez opakování, tedy počet možností je

$$\binom{190}{4} = \frac{190 \cdot 189 \cdot 188 \cdot 187}{1 \cdot 2 \cdot 3 \cdot 4} = 52602165.$$

b) Kolika způsoby ji mohou vybrat, jestliže v ní chtějí dva kluky a dvě holky?

Výběry kluků a holek jsou nezávislé, tvoří dvě fáze výběru, tedy je spojíme násobícím principem, každý z těch výběrů je bez opakování a bez pořadí, proto je odpověď

$$\binom{150}{2} \cdot \binom{40}{2} = \frac{150 \cdot 149}{2} \cdot \frac{40 \cdot 39}{2} = 8716500.$$

c) Kolika různými způsoby ji mohou vybrat, když chtějí dva kluky a dvě holky a navíc chtějí určit toho, kdo za ně bude mluvit?

Metoda 1: Nejprve vybereme delegaci, viz b), pak z ní vybereme mluvčího:

$$\binom{150}{2} \cdot \binom{40}{2} \cdot 4 = 34866000.$$

Metoda 2: Nejprve vybereme mluvčího, pak doplníme delegaci. Mluvčí bude buď kluk nebo holka, což ovlivní počty při dalším výběru, situace se tedy rozpadne na dva disjunktní případy, které spojíme sčítacím principem (bude tam nevyvážený strom). Vyjde

$$150 \cdot \binom{149}{1} \cdot \binom{40}{2} + 40 \cdot \binom{150}{2} \cdot \binom{39}{1} = 34866000.$$

d) Kolika různými způsoby ji mohou vybrat, když chtějí dva kluky a dvě holky a také dopředu určit, v jakém pořadí budou k profesi vcházet?

Nejjednodušší je nejprve vytvořit delegaci a pak řešit pořadí neboli ji permutovat.

$$\binom{150}{2} \cdot \binom{40}{2} \cdot 4! = 209196000.$$

e) Kolika různými způsoby ji mohou vybrat, když chtějí dva kluky a dvě holky a také dopředu určit, v jakém pořadí budou k profesi vcházet, přičemž nechtějí, aby šli dva kluci za sebou?

Zase je nejjednodušší vybrat delegaci a pak ji permutovat, teď ovšem nelze použít všechny permutace. Jak vypočítáme ty povolené? Třeba tak, že nejprve bereme jen obecně kluky a holky a najdeme tři možnosti uspořádání: $hkhk$, $khkh$ a $khhk$. Dvě holky a dva kluci se ovšem na ta místa h a k mohou dát v různých pořadích, čili celkový počet možností je

$$\binom{150}{2} \cdot \binom{40}{2} \cdot 3 \cdot 2! \cdot 2! = 104598000.$$

Poznámka: Jak bychom počet pořadí dělali, kdyby bylo holek 11 a kluků 7? Vypisování možných pořadí by bylo dost dlouhé, existuje nějaká obecná metoda? Jestliže nechceme, aby šli kluci za sebou, tak vždy mezi dvě holky můžeme ale nemusíme postavit kluka, lze také postavit kluka úplně na začátek či konec. To znamená, že z pozic mezi holkami (kterých je 10 plus dvě na konci, tedy dvanáct) vybíráme sedm míst pro kluky, je tedy $\binom{12}{7}$ možností seřazení, pokud nás nezajímají konkrétní osoby, jen pohlaví. Možností seřazení konkrétních osob pak je $\binom{12}{7} \cdot 11! \cdot 7!$, viz také cvičení 11a.39.

f) Kolika způsoby ji mohou vybrat, jestliže chtějí mít alespoň jednoho kluka a alespoň jednu holku?

Zde se řešení rozpadnou na disjunktní případy podle počtu kluků a holek, vyjde

$$\binom{150}{1} \cdot \binom{40}{3} + \binom{150}{2} \cdot \binom{40}{2} + \binom{150}{3} \cdot \binom{40}{1} = 32250500.$$

Alternativa: Co kdybychom rovnou zkusili napevno vybrat holku a kluka a pak zbytek doplnit už bez ohledu na pohlaví? Dostali bychom $\binom{150}{1} \binom{40}{1} \binom{188}{2} = 105468000$. Vyšlo to jinak a znatelně více. To se občas u kombinatorických úloh stává, důležité je rozpoznat správné řešení a najít chybu v chybém. Zde je chyba v alternativním řešení, některé situace totiž počítáme dvakrát. Například pokud jsou ve výboru dva kluci A a B , tak jsme tuto situaci zahrnuli jednou, když jsme nejprve vybírali kluka A a pak doplňovali zbytek, podruhé jsme jako prvního dali B a pak A došel v rámci doplnění. Kdyby ještě k nim byly dvě holky, tak se tatáž situace dokonce započítala čtyřikrát. To je komplikace, na kterou je třeba u kombinatorických úloh dávat velký pozor. Vzhledem k tomu, že násobnost počítání není pořád stejný koeficientem (někdy čtyřikrát, v případě tří kluků či tří holek jen tříkrát), nelze správnou odpověď z té špatné získat dělením. Tento přístup je tedy v tomto příkladě slepá ulička, někdy je ale myšlenka nejprve splnit povinné cenzum a pak libovolně doplnit zbytek užitečná, viz níže či třeba příklad 11a.k.

g) Kolika způsoby ji mohou vybrat, jestliže chtějí dva kluky a dvě holky, určitě má jít Bára ale Adam ne?

Opět vybíráme po fázích. Nejprve vybereme třeba kluky, ale Adam nesmí, vybíráme tedy ze 149. Pak vybereme holky, Báru určitě, pak na zbývající místo dobereme ještě jednu holku z ostatních 39. Celkem je tedy možností

$$\binom{149}{2} \binom{39}{1} = 430014.$$

K tomuto příkladu se vrátíme, viz příklad 11b.a.

△

! Příklad 11a.n: Ve školce je n kluků a n holek. Jdou na vycházku, učitelka je rozřadí do dvojic.

a) Kolika způsoby může děti seřadit?

Protože se blíže nespecifikuje, budeme předpokládat, že záleží na všem, tedy v jaké dvojici kdo jde i kdo je vlevo a vpravo. Jedna možnost řešení je postupně vybírat. Do první dvojice vpravo je $2n$ možností, do druhé dvojice vlevo pak je $2n - 1$ možností, do třetí dvojice vpravo pak je $2n - 2$ atd., možnosti se násobí (jde o fáze výběru) a dostaneme $(2n)!$.

Alternativa: Můžeme děti seřadit vedle sebe a pak dvojice odpočítávat postupně, čímž se celá věc redukuje na počet permutací dětí, což je $(2n)!$.

b) Kolika způsoby je učitelka může děti seřadit, jestliže jí záleží jen na tom, kdo je v které dvojici?

Zde vybíráme bez pořadí do první dvojice, tedy $\binom{2n}{2}$, pak do druhé dvojice, což je další fáze čili násobíme číslem $\binom{2n-2}{2}$, pak násobíme číslem $\binom{2n-4}{2}$ atd., dostáváme

$$\frac{(2n)(2n-1)}{2!} \cdot \frac{(2n-2)(2n-3)}{2!} \cdots \frac{2 \cdot 1}{2!} = \frac{(2n)!}{2^n}.$$

Alternativa: Nejprve vybereme děti do dvojic, jako by na pořadí záleželo, což je $(2n)!$ možností. Pak postupně pořadí ve dvojicích „rušíme“, po zrušení u první dvojice se počet různých možností redukuje na polovinu, u druhé dvojice zrušení pořadí způsobí další redukci na polovinu atd.

c) Kolika způsoby může děti seřadit, jestliže je nám jedno, kdo je v které dvojici a zda vlevo a vpravo, zajímá nás jen, kdo je s kým?

Jedna možnost je nechat prvního vybrat, má $(2n - 1)$ možností. Tím dvě děti odpadnou. Další si pak může vybrat z $(2n - 3)$ možností, další dva odpadnou, další z $(2n - 5)$ atd. Celkem je tedy počet dvojic

$$(2n - 1) \cdot (2n - 3) \cdots 3 \cdot 1 = (2n - 1)(2n - 3) \cdots 3 \cdot 1 \cdot \frac{(2n)(2n-2)(2n-4)\cdots 4 \cdot 2}{(2n)(2n-2)(2n-4)\cdots 4 \cdot 2} = \frac{(2n)!}{2^n \cdot (n-1) \cdot 2 \cdot (n-2) \cdots 2 \cdot 1} = \frac{(2n)!}{2^n n!}.$$

Vzorec jsme zkoušeli upravit tak, aby byl kompaktnější, zajímavé je, že se k té finální verzi dá také přijít přímo. Nejprve vybereme děti do dvojic podle a), pak zrušíme pozici ve dvojicích jako v b), nakonec zrušíme i pořadí

dvojic a máme to.

Mimochodem, číslu $1 \cdot 3 \cdot 5 \cdot (2n - 3) \cdot (2n - 1) = (2n - 1)!!$ se říká dvojný faktoriál. Podobně se pro sudá čísla definuje $2 \cdot 4 \cdot 6 \cdot (2n - 2) \cdot (2n) = (2n)!!$.

d) Kolika způsoby může děti seřadit, jestliže chce mít v každé dvojici kluka a holku a chce mít kluky za sebou a holky za sebou?

Zde je to snadné, nejprve dáme kluky nalevo a holky napravo a pak je nezávisle na sobě můžeme permutovat, to je $(n!)^2$ možností. Je ale také možné řadit holky vlevo a kluky vpravo, celkem je tedy $2(n!)^2$ možností.

e) Kolika způsoby může děti seřadit, jestliže chce mít v každé dvojici kluka a holku, ale neřeší, kdo je vlevo a kdo vpravo?

Zde bude nejjednodušší natvrdo vybrat řekněme chlapce vlevo a holky vpravo, což je $(n!)^2$ možností, oni už se pak seřadí ve dvojicích dle libosti.

f) Kolika způsoby může děti seřadit, jestliže chce mít v každé dvojici vlevo kluka a vpravo holku, ale neřeší, jak jdou dvojice za sebou?

Tady prostě stačí nastoupit kluky do řady a k nim přiřazovat holky, což dá $n!$ možností.

g) Kolika způsoby může děti seřadit, pokud je chce mít v zástupu a aby se střídali kluci s holkami?

Nezávisle srovnáme kluky a holky do zástupů, to dává $(n!)^2$ možností. Pak si jen vybereme, jestli půjdou $hkhk\dots$ nebo $khkh\dots$, tedy celkem $2(n!)^2$ možností.

h) Učitelka děti posadí na kolotoč řetízkáč. Kolika způsoby to lze udělat?

Toto je problém známý jako „problém kulatého stolu“. Jeho zásadním rysem je rotační symetrie, například pokud máme tři děti a sedí na kolotoči způsobem $\begin{smallmatrix} A \\ B & C \end{smallmatrix}$, tak to po otočení kolotoče dá $\begin{smallmatrix} C \\ A & B \end{smallmatrix}$ a také $\begin{smallmatrix} B \\ C & A \end{smallmatrix}$. Jinými slovy, to, co bychom normálně považovali za tři rozdílná seřazení, je jen jedno.

Kulatý stůl se dá v zásadě řešit dvěma přístupy. Jednak je možné jej „rozříznout“, vyřešit problém v jedné řadě a poté tu řadu zase slepit, celkový počet se pak ale musí vydělit číslem $2n$, protože to je přesně počet protočení řetízkáče (stolu). V případě našich dětí je můžeme posadit do řady $(2n)!$ způsoby, takže na řetízkovém kolotoči to bude $\frac{(2n)!}{2n} = (2n - 1)!$ způsoby.

Druhá varianta je, že někam natvrdo posadíme Pepíčka, protože když se kolotoč točí, tak někdy na té pozici určit bude. Zbývá pak rozesadit ostatní, což je $(2n - 1)!$ možností.

i) Kolika způsoby je možné posadit děti na řetízkový kolotoč tak, aby se střídali kluci s holkama?

Posadíme Pepíčka, tím je dáno, kde budou sedět kluci, tak je tam rozmístíme, to je $(n-1)!$ možností. Na zbývající místa dáme holky, celkem $n!(n-1)!$ způsobů.

△

! Příklad 11a.o: 10a.x Uvažujme čísla $\{1, 2, 3, \dots, n\}$, kde $n \geq 11$. Vybereme z nich pět různých. Kolika způsoby je možné to udělat, aby bylo druhé největší číslo nejvýše 10?

Ukážeme dvě řešení, jedno bude snažší (používá standardní přístup), druhé obtížnější (vyžaduje představivost) ale s mnohem lepším tvarem výsledku.

1) Dřevorubecké řešení: Většinou pomůže zaměřit se na omezení, v tomto případě na druhé nejvyšší číslo. Kolika způsoby jej můžeme vybrat? Největší možnost je 10 a nejmenší 4, protože se pod něj ještě musí vejít další tři. Celkem je tedy 7 možností ($10 - 4 + 1$, pozor) pro jeho výběr. Teď vybereme další čísla.

Začneme největším. To musí být nad tím již vybraným, takže máme drobný problém, potřebujeme znát pozici již vybraného čísla. To ukazuje, že je třeba zavést parametr, budeme tedy dále předpokládat, že vybrané druhé největší číslo je k . Největší číslo pak vybíráme z rozsahu $k+1, k+2 \text{ až } n$, celkem tedy $n - (k+1) + 1 = n - k$ možností pro jednu konkrétní volbu k .

Teď tento výběr ještě doplníme o tři nejmenší, ty musí být pod číslem k , tedy vybíráme tři čísla z rozmezí $\{1, 2, \dots, k-1\}$, bez opakování a najednou, tedy na pořadí nezáleží, což je $\binom{k-1}{3}$ možností. Pro jednu konkrétní volbu k pro druhé největší číslo tedy máme celkem $(n-k)\binom{k-1}{3}$ možností výběru. Projinou hodnotu k dostáváme zcela odlišné výběry (liší se přinejmenším pozicí druhého největšího čísla), jde tedy o rozklad na disjunktní množiny. Celkový počet možností tedy dostaneme sečtením možností pro jednotlivá k .

Závěr: Je možno udělat $\sum_{k=4}^{10} (n-k)\binom{k-1}{3}$ výběrů dle zadání.

Komentář: Tento postup nebyl příliš elegantní, ale často je jediný možný, takže se mnohdy smíříme s výsledkem v neuzavřeném tvaru. Zároveň jsme si připomněli několik užitečných triků. Teď už se ale podívejme na druhé řešení.

2) Elegantní řešení. Základem je následující zjednodušující úvaha. Není třeba situaci řešit ve třech krocích, ale ve dvou, nejprve vybereme čtyři nejmenší čísla a pak to největší. Když ta čtyři nejmenší vezmeme 1 až 10, tak máme

zajištěno, že druhé největší nepřekročí desítku. Poznamenejme, že opravdu je třeba to páté brát jako největší, protože když jej budeme brát bez omezení, tak bychom se mohli dostat vícekrát ke stejnemu výběru (například 1, 2, 5, 6 a pak 9 dává stejný výběr jako 1, 2, 5, 9 a pak 6). Tím se ale dostáváme k problému, že když už ty čtyři vybereme najednou, tak neumíme zjistit, jak velké je to největší, abychom tak dostali možnosti pro výběr pátého, ještě většího.

To je problém podstatný a mnohý řešitel by v teď tuto cestu asi vzdal, ale existuje cesta, jak z toho ven. Onen výběr pátého čísla totiž v zásadě funguje dvojím způsobem. Pokud je i páté číslo v rozmezí do deseti, pak prostě stačí vybrat pět čísel z množiny $\{1, \dots, 10\}$ a víc není třeba řešit. Druhá možnost je, že to páté číslo je větší než 10, pak ale zase přesně víme, jak jej můžeme vybrat.

Výběry tedy rozdělíme do dvou možností podle pozice největšího čísla. Pokud je nejvýše 10, pak jde vlastně o výběr pěti čísel z 10 možností. Pokud je největší číslo větší než 10, pak můžeme udělat onen dvoustupňový výběr, nejprve vybereme čtyři čísla z deseti možných a pak dobereme páté z rozmezí 11 až n , celkem $\binom{10}{4}(n-10)$ možností.

Závěr: Je možno udělat $\binom{10}{5} + \binom{10}{4}(n-10)$ výběrů dle zadání.

△

Ted si předvedeme několik teoretičtějších příkladů.

! Příklad 11a.p: Mějme množinu A o n prvcích.

a) Kolik je uspořádaných dvojic, které lze z prvků A vyvoret?

Zde se ptáme na velikost množiny $A \times A$ a tudíž je odpověď jasná, n^2 .

b) Kolik je neuspořádaných dvojic různých prvků, které lze z prvků A vytvořit?

Odpovíme třemi způsoby.

1) Nejprve spočítáme všechny uspořádané dvojice různých prvků. To je snadné, stačí vzít všechny dvojice z části a) a odečíst dvojice stejných prvků, kterých je n . Je jich tedy $n^2 - n = n(n-1)$.

Dvě uspořádané dvojice typu $(a, b), (b, a)$ vždy dají jednu neuspořádanou $\{a, b\}$, proto je počet neuspořádaných dvojic různých prvků roven $\frac{1}{2}n(n-1)$.

2) Další možný pohled na věc je, že neuspořádané dvojice jsou prostě jen dvouprvkové podmnožiny A , jinými slovy vytahujeme dva prvky z n , na pořadí nezáleží a bez opakování, což dává $\binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{1}{2}n(n-1)$.

3) Zkusíme to ještě jinak, podíváme se na to z pohledu jednotlivých prvků A . První prvek se může dostat do dvojice s $n-1$ dalšími prvky. Druhý prvek už v započítané dvojici s prvním prvkem je, takže se může nově družit s dalšími $n-2$ prvky. Třetí prvek už je započítán ve dvojici s prvními dvěma, může tedy přidat dalších $n-3$ dvojic. Předposlední prvek ještě neměl započítanu dvojici s posledním a tím to končí, poslední už má všechny dvojice započítány. Celkem je dvojic

$$(n-1) + (n-2) + \dots + 2 + 1 = \sum_{k=1}^{n-1} k = \frac{1}{2}n(n-1),$$

použili jsme Větu 9c.3 (ii).

c) Kolik je neuspořádaných dvojic prvků, které lze z prvků A vytvořit?

I zde je možné použít několik přístupů. Asi nejjednodušší je využít předchozí práce. Víme už, že je $\frac{1}{2}n(n-1)$ neuspořádaných dvojic s různými členy, dvojic typu $\{a, a\}$ je evidentně n , celkem je tedy $\frac{1}{2}n(n-1) + n = \frac{1}{2}n(n+1)$ neuspořádaných dvojic.

Druhá možnost je aplikovat postup b)3). První prvek A se může družit s n možnými prvky (i se sebou), druhý už je s prvním započítán, zbývá $n-1$ možností atd, poslední má ještě nezapočítanou možnost družit se sám se sebou. Celkem je to

$$n + (n-1) + \dots + 2 + 1 = \sum_{k=1}^n k = \frac{1}{2}n(n+1).$$

Naopak obtížné by bylo zkousit adaptovat postup z b)1). Z části a) víme, kolik je všech uspořádaných dvojic, ale my je neumíme jednoduše převést na neuspořádané. Zádrhel je v tom, že některé uspořádané dvojice se při přechodu na neuspořádané druží po dvou, jmenovitě když jsou jejich složky různé, ale dvojice typu (a, a) už jdou na neuspořádané metodou jeden na jednoho. Pro rozumné zvládnutí by tedy bylo třeba rozdělit situaci na tyto dva případy, ale to jsme zpět u prvního řešení.

△

! Příklad 11a.q: Uvažujme dvě konečné množiny A, B . Kolik je možných relací z A do B ?

Relace jsou podmnožiny $A \times B$, jejich počet je podle příkladu 11a.i roven $2^{|A \times B|} = 2^{|A| \cdot |B|}$.

△

! Příklad 11a.r: Uvažujme množinu A o n prvcích.**a)** Kolik je relací na A ?Podle příkladu 11a.q je jich 2^{n^2} .**b)** Kolik je reflexivních relací na A ?Reflexivní relace automaticky obsahují všechny dvojice typu (a, a) , zde není žádná svoboda volby. Počet reflexivních relací je tedy dán počtem podmnožin množiny všech uspořádaných dvojic různých prvků. Těch je $n(n - 1)$, viz příklad 11a.p b)1), takže počet podmnožin a tím i počet reflexivních relací na A je $2^{n(n-1)}$.**c)** Kolik je symetrických relací na A ?U symetrické relace nezáleží na orientaci, čili se u každé neuspořádané dvojice prvků ptáme, zda ji vezmeme do relace nebo ne. Jinými slovy, zajímá nás počet všech podmnožin množiny neuspořádaných dvojic prvků, která má podle příkladu 11a.p c) velikost $\frac{1}{2}n(n + 1)$. Závěr: Existuje $2^{n(n+1)/2}$ symetrických relací na A .**d)** Kolik je antisymetrických relací na A ?Toto nepříjde najednou, antisymetrická relace se totiž dívá na dvojice různě. Pokud je to dvojice stejných prvků, tak to antisimetrii nezajímá a můžeme si je brát či nebrat dle libosti. Zato u dvojice různých prvků povoluje antisimetrie žádnou či jen jednu spojnici. Antisymetrickou relaci tedy vytvoříme ve dvou nezávislých fázích neboli použijeme násobící princip. Nejprve se rozhodneme, které ze smyček použijeme. Možných smyček je n , my z nich vybíráme podmnožiny, je tedy 2^n možností.V druhé fázi doplníme spojnice mezi nestejnými prvky. Množina neuspořádaných dvojic různých prvků má dle příkladu 11a.p b) velikost $\frac{1}{2}n(n - 1)$, my se u každé dvojice rozhodujeme mezi třemi variantami: nevzít vůbec, vzít od menšího k většímu, vzít od většího k menšímu (vzhledem k pořadí prvků v množině A). Je tedy možno vytvořit $3^{n(n-1)/2}$ výběrů.Závěr: Na množině A je $2^n 3^{n(n-1)/2}$ různých antisymetrických relací.**e)** Vzhledem k tomu, že asymetrické relace jsou vlastně antisymetrické a navíc antireflexivní, znamená to, že u nich nejsou pro dvojice stejných prvků žádné možnosti výběru. Je tedy $3^{n(n-1)/2}$ asymetrických relací na množině A .

Tranzitivita se špatně kombinatoricky uchopuje, ani s tím nebudeme začínat.

△

Příklad 11a.s: **a)** Kolik lze vytvořit různých řetězců ze slova *DUMBO*?To je snadné, jde o permutace čili $5! = 120$ řetězců.**b)** Kolik lze vytvořit různých řetězců ze slova *BAMBI*?Zase jde o permutace, ale máme problém, protože se nám dvě písmena shodují. Zkusme se nejprve odvolut na to, co umíme, a označit si ta béčka, máme tedy B_1AMB_2I . Z tohoto umíme udělat $5!$ řetězců. My si je teď rozdělíme, množinu všech řetězců rozložíme na disjunktní množiny tak, že v každé množině je vždy jeden konkrétní řetězec a také všechny jiné řetězce, které z něj lze získat permutací těch B . Jedna taková množina je třeba $\{B_1B_2AMI, B_2B_1AMI\}$, jiná třeba $\{B_1AB_2IM, B_2AB_1IM\}$. Je snadné si rozmyslet, že každá tato množina má přesně $2! = 2$ prvky.Je také zjevné, že když teď od béček odmažeme ty indexy neboli zrušíme jejich pořadí, tak se každá tato množina zvrkne na jeden řetězec, takže počet různých řetězců vytvoritelných z *BAMBI* je $\frac{5!}{2!} = 60$.Tento postup můžeme snadno aplikovat i na případy, kdy se opakuje více objektů, třeba ze slova *BREKEKE* lze vytvořit $\frac{7!}{2!3!} = 420$ různých řetězců.Ukážeme si ještě jeden způsob, jak na *BAMBI*. Začneme tím, že nejprve mezi pěti pozicemi vybereme ty, kam budeme strkat béčka: $\binom{5}{2}$. V druhé fázi na ostatní místa zpermutujeme *AMI*, celkem dostaneme $\binom{5}{2}3! = 60$ různých řetězců.

△

Tuto myšlenku si nyní zobecníme.

! Věta 11a.6.Mějme n objektů, z toho n_1 je typu 1 (jsou nerozlišitelné), n_2 typu 2 až n_k je typu k , tedy $\sum_{i=0}^k n_i = n$.Pak je $\frac{n!}{n_1! \cdot n_2! \cdots n_k!}$ různých permutací těchto objektů.

Důkaz: Indukcí na k .

(0) $k = 1$: Permutujeme objekty jednoho druhu, máme $n = n_1$, tedy má být $\frac{n_1!}{n_1!} = 1$ možností, což souhlasí, u stejných objektů prohození nepoznáme.

(1) Předpokládejme, že umíme permutovat kolekce k typů objektů. Mějme teď kolekci $k+1$ typů objektů.

Nejprve vybereme umístění pro objekty typu $k+1$ mezi n volnými místy, což se dá udělat $\binom{n}{n_{k+1}}$ způsoby. Zbývá uspořádat $(n - n_{k+1})$ objektů o k typech mezi zbývajícími $n - n_{k+1}$ volnými místy, což podle indukčního předpokladu lze udělat $\frac{(n-n_{k+1})!}{n_1! \cdot n_2! \cdots n_k!}$ způsoby. Celkem je tedy možno vyrobit

$$\binom{n}{n_{k+1}} \cdot \frac{(n - n_{k+1})!}{n_1! \cdot n_2! \cdots n_k!} = \frac{n!}{n_{k+1}!(n - n_{k+1})!} \cdot \frac{(n - n_{k+1})!}{n_1! \cdot n_2! \cdots n_k!} = \frac{n!}{n_1! \cdot n_2! \cdots n_k! \cdot n_{k+1}!}$$

různých permutací. □

Příklad 11a.t: Ze slova *BAOBAB* je možno vytvořit $\frac{6!}{3!2!1!} = 60$ různých slov.

△

Příklad 11a.u: Vráťme se na chvíli ke kartám. Ve hře bridge se všech 52 karet rozdá mezi 4 soutěžící (každý má 13). Záleží však jen na tom, kdo dostane jakou, nikoliv na pořadí, v jakém k nim jednotlivé karty přicházejí. Kolika způsoby to lze udělat?

Jedna možnost je vybrat karty pro prvního, pak pro druhého atd, jde o fáze s počtem možností, který nezávisí na konkrétní volbě předchozí fáze, tedy podle násobícího principu je počet možností

$$\binom{52}{13} \binom{39}{13} \binom{26}{13} \binom{13}{13} = \frac{52!}{13!39!} \cdot \frac{39!}{13!26!} \cdot \frac{26!}{13!13!} \cdot 1 = \frac{52!}{13!13!13!13!} \sim 5.36 \cdot 10^{28}.$$

(Pro sstrandu: Je to 53644737765488792839237440000.)

Ten předposlední výraz vypadá zajímavě, najde se k němu přímá cesta? Ano, položíme karty do řady a rozhodneme, že první hráč dostane prvních 13, druhý druhých 13 atd. Rozdání se tedy realizuje permutacemi karet, ale protože u jednotlivého hráče na pořadí nezáleží, zrušíme příslušné počty permutací tím, že dělíme těmi 13!.

Alternativní řešení: Dá se zkousit i změna úhlu pohledu, nerozdáváme karty hráčům, ale přidělujeme hráče kartám. Vyrobíme za každého hráče 13 žetonků a řadíme je různě podél karet. Jde o permutace s opakujícími se prvky, vychází tedy $\frac{52!}{13!13!13!13!} \sim 5.36 \cdot 10^{28}$. △

Kromě permutací jsme v této kapitole zkoumali také výběry. Jak to vypadá, když máme od každého typu omezený počet kusů a chceme z nich několik vybrat (ať už s pořadím či bez)? Je to velký problém podobný tomu z příkladu 11a.l, protože hned první volbou změníme počty kusů, které jsou k dispozici do dalších kol, takže nelze používat násobící princip a v zásadě nezbyde nic než zase rozebírat možnosti například pomocí stromů. Jinými slovy, žádné rozumné vzorce nejsou a každý příklad je svůj. Jeden takový si ukážeme.

Příklad 11a.v: (i) Máme k dispozici tři bonbóny malinové a sedm bonbónů citrónových. Chceme-li si vzít osm, kolika způsoby je to možné?

a) Na pořadí nezáleží: To je snadná úloha, můžeme mít nejvýše tři malinové, ale alespoň jeden vzít také musíme, jinak by nám citrónové nestačily na doplnění do osmi. Jsou tedy tři možné způsoby (1,2 nebo 3 malinové).

b) Na pořadí záleží: Zde jsou možné dva přístupy. Jeden využívá výběru bez pořadí, který se pak zpermutuje. Nestačí ale vzít výsledek z a) a vynásobit určitou konstantou, protože počet různých permutací záleží na tom, kolikrát se který prvek opakuje. Situace se tedy musí rozložit dle počtů, například malinových bonbónů.

$m = 1$: Zde permutujeme jeden M a sedm C, což dává $\frac{8!}{1!7!} = 8$ možností. Jde to i selským rozumem, prostě vybíráme místo v pořadí pro ten jeden malinový.

$m = 2$: Zde permutujeme dva M a šest C, $\frac{8!}{2!6!} = 28$ možností.

$m = 3$: $\frac{8!}{3!5!} = 56$ možností.

Celkem $8 + 28 + 56 = 92$ možností.

Druhá možnost je udělat si strom, což je ale evidentně vysoce nouzová metoda, protože by měl mít nakonec 92 větví.

(ii) Máme k dispozici pět bonbónů malinových, tři pomerančové a pět citrónových. Chceme-li si vzít devět, kolika způsoby je to možné?

Tady už to začíná být opravdu zajímavé, a to dokonce i v případě, že neuvažujeme pořadí. Je například vidět, že musíme vzít alespoň jeden bonbón malinový a alespoň jeden citrónový, ale co dál? Jedna možnost je rozebrat případy podle toho, kolik vezmeme třeba těch malinových.

$m = 1$: Zbývá osm bonbónů, zde není na výběr, musíme prostě vzít všechny P a všechny C. Jedna možnost.

$m = 2$: Zbývá vzít sedm bonbónů, tedy dva či tři P. Dvě možnosti.

$m = 3$: Zbývá vzít šest bonbónů, tedy jeden až tři P, tři možnosti.

$m = 4$: Zbývá vzít pět bonbónů, tedy nula až tři P, čtyři možnosti.

$m = 5$: Zbývá vzít čtyři bonbóny, tedy nula až tři P, zase čtyři možnosti.

Celkem máme 14 možností.

Pokud bychom chtěli výběr s pořadím, pak pro každou z těchto 14 možností musíme řešit permutace.

Je jasné, že kdyby se vybíralo z řekněme šesti druhů, tak už je tento rozbor úkolem na dlouhé zimní večery. Existují lepší alternativy?

Je zde možnost převodu na rovnici, naše úloha se dá formulovat jako hledání počtu řešení rovnice $m + p + c = 9$, které jsou z \mathbb{N}_0 a splňují $m \leq 5$, $p \leq 3$ a $c \leq 5$. Základem bude přístup z příkladu příkladu 11a.k, ale ještě potřebujeme další nástroje, vrátíme se k tomuto příkladu v další kapitole, viz příklad 11b.e.

Poprvadě řečeno, pro větší počet druhů by i toto bylo dosti pracné. Pak by zase asi bylo nejfektivnějším řešením napsat algoritmus, který by prošel všechny možnosti a spočítal je. Jinými slovy bychom si to rozdělili, my bychom udělali to přemýšlení a počítac by odvedl hrubou manuální práci. Tak to má být.

△

Cvičení

Cvičení 11a.1 (rutinní): Napsali jsme pěti kamarádům e-mail a očekáváme odezvu, předpokládejme, že každý napíše jednou a jiné e-maily už nedorazí.

a) V kolika pořadích se jejich odpovědi mohou v mailboxu objevit?

b) Pokud jsme náš mailserver nechali přecpat a má místo už jen na tři příchozí e-maily, kolik je možností pro to, co v mailboxu najdeme?

Cvičení 11a.2 (rutinní): Kolika způsoby je možné odpovědět na 100 otázek typu ano/ne, je-li možné na otázku také neodpovědět?

Cvičení 11a.3 (rutinní): Kolika způsoby je možno vybrat 6 předmětů z deseti různých věcí, jestliže

a) výběr je uspořádaný a opakování není povoleno?

b) výběr je uspořádaný a opakování je povoleno?

c) výběr je neuspořádaný a opakování není povoleno?

d) výběr je neuspořádaný a opakování je povoleno?

Cvičení 11a.4 (rutinní): Kolik je možno vytvořit poznávacích značek, jestliže se skládá z číslice nerovné 0, pak písmena, pak číslice nerovné 0 a nakonec čtyřmístného čísla?

Cvičení 11a.5 (rutinní): Kolika různými způsoby je možno rozdělit medaile mezi osm běžců za předpokladu, že nebude remíza?

Cvičení 11a.6 (rutinní): Kolika způsoby je možné vybrat 8 mincí z prasátka, ve kterém je 100 stejných korun a 80 stejných dvoukorun?

Cvičení 11a.7 (rutinní): Kolik je osmibitových binárních slov, která obsahují právě pět jedniček?

Cvičení 11a.8 (rutinní): Na škole je 11 profesorů matematiky a 7 profesorů fyziky. Kolika způsoby je možné vybrat výbor skládající se ze tří matiků a tří fyziků?

Cvičení 11a.9 (rutinní): Kolik různých kombinací korun, dvoukorun, pětikorun, desetikorun a dvacetikorun může být v prasátku, jestliže je tam 20 mincí?

Cvičení 11a.10 (rutinní): Uvažujme písmena *abcdef*.

a) Kolik jejich permutací končí na *d*?

b) Kolik jejich permutací končí na *g*?

Cvičení 11a.11 (rutinní): Kolika různými způsoby je možno rozvést 3000 knih do tří různých skladišť?

Cvičení 11a.12 (rutinní): Z 15 otázek na přijímací zkoušky mají mít 4 správnou odpověď *a*, 4 správnou odpověď *b*, 4 správnou odpověď *c* a 3 správnou odpověď *d*. Kolik je třeba vyrobit šablon k opravování, aby ke každé příslušné volbě otázek byla šablona?

Cvičení 11a.13 (rutinní): Kolika způsoby lze rozdělit šest identických balónků do devíti různých koší?

Cvičení 11a.14 (rutinní): V baru nalévají do sklenic tři druhy vína a dva druhy piva. Měli jste za večer pět sklenic. Kolika způsoby to mohlo proběhnout, když jste nechtěli míchat víno s pivem?

Cvičení 11a.15 (rutinní): Učitel se rozhodl vyhodit 40 lidí ze 200 studentů. Kolik se mu nabízí možností? Kolik má možností za situace, kdy těch 40 představuje jen horní mez, ale kolik jich vyhodit nemusí?

Cvičení 11a.16 (rutinní): a) Kolik existuje desetimístných řetězců sestavených ze dvou nul, tří jedniček a pěti dvojek?
b) Kolik existuje desetimístných čísel ve trojkové soustavě, která jsou zapsaná pomocí dvou nul, tří jedniček a pěti dvojek?

Cvičení 11a.17 (rutinní): Kolik je možných Internetových adres podle protokolu IPv4?

Adresy mají 32 bitů a jsou rozděleny do tří tříd.

Class A pro velké sítě: za 0 následuje 7-bitový netID (nesmí být 1111111) a pak 24-bitový hostID (nesmí být samé 0 či samé 1).

Class B pro střední sítě: za 10 následuje 14-bitový netID a pak 16-bitový hostID (nesmí být samé 0 či samé 1).

Class C pro malé sítě: za 110 následuje 21-bitový netID a pak 8-bitový hostID (nesmí být samé 0 či samé 1).

Poznámka: Rezervovány jsou ještě Class D 1110 pro speciální účely (multicasting) a Class E 11110 jako rezerva. IPv6 už má 128 bitové adresy.

Cvičení 11a.18 (rutinní): Kolik má množina se 100 prvky alespoň dvouprvkových podmnožin?

Cvičení 11a.19 (rutinní): Hodí se desetkrát po sobě mincí.

- a) Kolika způsoby to může dopadnout?
- b) Kolik z nich má přesně dvě hlavy?
- c) Kolik z nich má stejně hlav a orlů?
- d) Kolik z nich má více hlav než orlů?
- e) Jak se odpovědi na otázky a) až d) změní, pokud hážeme najednou desíti různými mincemi?
- f) Jak se odpovědi na otázky a) až d) změní, pokud hážeme najednou desíti identickými mincemi?

Cvičení 11a.20 (rutinní): V obchodě mají 5 různých druhů čokolády. Kolika způsoby je možno koupit

- a) tři čokolády?
- b) tři čokolády tak, aby byla každá jiná?
- c) šest čokolád tak, aby byla každá jiná?
- d) šest čokolád tak, aby se každý druh vyskytnul?
- e) šest čokolád tak, aby tam určitě byly alespoň tři mléčné s lískovými oříšky, ale určitě ne víc než jedna hořká?

Cvičení 11a.21 (rutinní): Student si vždycky ráno cestou do školy koupí bagetu.

a) Jestliže je k dispozici šest druhů a nechce si v jednom týdnu kupovat stejnou vícekrát, kolika různými způsoby může během týdne bagety (na pořadí záleží)?

b) Jak dlouho může studovat, pokud nechce bagetové týdny opakovat?

Cvičení 11a.22 (rutinní): Kolik řetězců lze vytvořit permutováním slova *BUBUKUKU*?

Cvičení 11a.23 (rutinní): Kolik permutací řetězce *ABCDEFGH* obsahuje slova *BA* a *FEC*?

Cvičení 11a.24 (rutinní): Kolika způsoby je možno uspořádat písmena *a*, *b*, *c* a *d*, jestliže *b* nesmí přijít hned po *a*?

Cvičení 11a.25 (rutinní): Kolika způsoby je možné uspořádat *AABBCCDDEE* tak, aby nebyly dvě *A* za sebou?

Cvičení 11a.26 (rutinní): Kolik řetězců je možné vytvořit pomocí písmen ze slova *HAHA*?

Cvičení 11a.27 (rutinní): Kolika způsoby je možno z pěti žen a sedmi mužů vybrat čtyřčlenný výbor za předpokladu, že musí obsahovat alespoň jednoho muže a alespoň jednu ženu?

Cvičení 11a.28 (rutinní): V cukrárně mají tři typy oplatku na zmrzlinu (kornoutek, kalíšek, mistička ve tvaru lastury), osm typů zmrzliny a dvě možnosti posypu (čokoláda, oříšky).

- a) Kolik je možné vytvořit zmrzlin ze tří kopečků, jestliže je možné opakovat druh zmrzliny a přidat posyp?
- b) Kolik je možné vytvořit zmrzlin ze tří kopečků, jestliže není možné opakovat druh zmrzliny, ale můžeme přidat až dva různé posypy?

Cvičení 11a.29 (rutinní): Kolika způsoby je možno vybrat pět klobík z 10 možných druhů, jestliže

- a) se nesmí opakovat druh?
- b) musí být všechny stejné?
- c) je to výběr bez omezení?
- d) musí být alespoň dva druhy?
- e) musí být alespoň čtyři jahodové?
- f) nesmí být více než tři pudinkové?

Na pořadí nezáleží.

Cvičení 11a.30 (rutinní): Kolik alespoň osmipísmenných řetězců lze vytvořit z písmen slova *BALAKLAVA*?

Cvičení 11a.31 (rutinní): Kolika způsoby je možné vybrat čtyři studentské zástupce, jestliže tam mají být obsaženy všechny stupně a na škole je 2500 studentů bakalářské etapy, 1200 studentů magisterské etapy a 300 doktorandů?

Cvičení 11a.32 (rutinní): Kolik je možné udělat SPZ, jestliže jsou možné formáty 3 písmena 4 čísla nebo 2 písmena 5 čísel?

Cvičení 11a.33 (rutinní): Kolik je možno utvořit palindromů (řetězců nad 26 písmeny abecedy, které se čtou stejně zleva i zprava) o délce n ?

Cvičení 11a.34 (rutinní): Kolika způsoby je možno vytvořit fotku 6 svatebčanů v řadě, jestliže

- a) nevěsta a ženich musí být vedle sebe;
- b) nevěsta musí být někde nalevo od ženicha;
- c) nevěsta nesmí být vedle ženicha.

Cvičení 11a.35 (rutinní): Jméno proměnné v jazyce C je řetězec o délce 1 až 8 skládající se z malých a velkých písmen, číslic či podtržítka, první znak nesmí být číslice. Kolik proměnných je možno pojmenovat?

Cvičení 11a.36 (rutinní): Uvažujme slova o délce 7 vytvořená z 26 malých písmen abecedy (6 samohlásek, 20 souhlásek).

- a) Kolik z nich má právě dvě samohlásky?
- b) Kolik z nich začíná d nebo t následováno samohláskou?
- c) Jak se to změní, když není možno písmena opakovat?

Cvičení 11a.37 (rutinní): Kolik je čtyřbitových řetězců, které neobsahují tři nuly hned po sobě?

Cvičení 11a.38 (rutinní): Kolik je pětibitových řetězců, které obsahují tři nuly po sobě?

Viz také cvičení 11b.4.

Cvičení 11a.39: (i) Kolika způsoby se může postavit osm mužů a pět žen do řady tak, aby dvě ženy nestály za sebou?

(ii) Kolika způsoby se může postavit m mužů a n žen tak, aby dvě ženy nestály za sebou?

Cvičení 11a.40 (rutinní): Kolika způsoby lze rozdělit 8 dětí do čtyř houpacích lodiček pro dva na pouti, když je nám jedno, jak v každé lodičce sedí?

Cvičení 11a.41 (rutinní): Nástupu v tanečních se účastní n mladíků a n dívek. V kolika různých pořadích mohou (po jednom) nastupovat, mají-li se střídat mladíci s dívками?

Cvičení 11a.42 (rutinní): Zvolme přirozené číslo k menší než 99.

- a) Kolik permutací čísel $1, 2, \dots, 100$ zachovává čísla $k, k+1, k+2$ ve správném pořadí a za sebou?
- b) Kolik permutací čísel $1, 2, \dots, 100$ zachovává čísla $k, k+1, k+2$ ve správném pořadí?

Cvičení 11a.43 (rutinní): Vybíráme čtyři čísla z $1, 2, \dots, 100$, a to bez opakování. Kolika způsoby je to možné udělat tak, aby:

- a) se ve výběru objevila čísla 13, 23, 31 po sobě?
- b) se ve výběru objevila čísla 13, 23, 31 nikoliv nutně hned po sobě, ale v tomto pořadí (tj. může se mezi ně vmačknout i něco jiného)?
- c) se ve výběru objevila nějaká tři čísla po sobě?
- d) se ve výběru objevila tři po sobě jdoucí čísla, ale nemusí být nutně vybrána hned po sobě, jen v tomto pořadí (tj. může se mezi ně vmačknout i něco jiného)?
- e) Vyřešte a) a b), pokud vybíráme ze sedmi čísel.

Cvičení 11a.44 (rutinní): Kolika způsoby lze vybrat tucet jablek z koše, ve kterém je 20 červených, 20 žlutých a 20 zelených, jestliže chceme alespoň tři od každé barvy?

Cvičení 11a.45 (rutinní): Kolik je možno vytvořit binárních řetězců z osmi nul a desíti jedniček, jestliže po každé nule musí přijít jednička?

A dvě jedničky?

Cvičení 11a.46 (rutinní): Kolika způsoby je možno rozesadit n lidí kolem kulatého stolu?

Cvičení 11a.47 (rutinní): Máme n manželských páru.

- a) Kolika způsoby je možno je posadit do řady, aby seděli v pořadí muž-žena?
- b) Kolika způsoby je možno je posadit do řady, aby seděl každý pár vedle sebe?
- c) A co kolem kulatého stolu?

Cvičení 11a.48 (rutinní): Kolik je nejvýše 8-místných binárních slov (neprázdných) splňujících tuto podmínu: Jestliže je první znak 0, pak už může následovat jen nula až sedm znaků 1. Jestliže je první znak 1, pak může následovat jen lichý počet jedniček.

Cvičení 11a.49 (poučné): a) Nechť $m, n \in \mathbb{N}$, uvažujme čtercovou síť o šířce m čtverců a výšce n čtverců. Kolika způsoby je možno se dostat z levého dolního rohu do pravého horního rohu, jestliže je povoleno se pohybovat jen doprava nebo nahoru po této síti?

Interpretace: Chceme se dostat v rovině z bodu $(0, 0)$ do bodu (m, n) , můžeme ale jen postupně zvětšovat jednotlivé souřadnice po 1.

b) Nechť $m, n, k \in \mathbb{N}$. Kolika způsoby je možné dojít v 3D-prostoru z bodu $(0, 0, 0)$ do bodu (m, n, k) za předpokladu, že je možné jít jen kroky o jedno podél některé osy ve směru zvyšující se souřadnice osy?

Cvičení 11a.50: Kód s pojistkou je zvolen tak, aby byl pětimístný, první čtyři místa jsou libovolné číslice a poslední je číslice zvolena tak, aby byl součet dvou posledních cifer dělitelný sedmi. Kolik takových kódů je možné vytvořit?

Cvičení 11a.51: Když vypíšeme všechna přirozená čísla menší či rovna 1000 v desítkové soustavě, kolik se celkem použije číslic

- a) 0?
- b) 1?
- c) 2?
- d) 3?
- e) 9?

Cvičení 11a.52 (poučné): Test má 5 otázek. Jestliže má každá otázka mít váhu alespoň 5 bodů, kolika způsoby se dají body rozvrhnout, aby byl součet 50?

Cvičení 11a.53: Kolik různých celocíselných řešení, které splňují $x_1 \geq 1, x_2 \geq 0, x_3 > 3, x_4 \geq 2$, má rovnice $x_1 + x_2 + x_3 + x_4 = 17$?

Cvičení 11a.54: Kolik podmnožin množiny $\{3, 7, 9, 11, 24\}$ má vlastnost, že mají součet menší než 28?

Cvičení 11a.55: Jakými různými způsoby může proběhnout zápas mezi tenisty A a B hraný na tři vítězné sety?

Cvičení 11a.56: Kolika způsoby může skončit závod čtyř skikrosařů?

Cvičení 11a.57: Kolika způsoby je možno v kině posadit do řady šest chlapců a osm dívek tak, aby žádní chlapci neseděli vedle sebe?

Cvičení 11a.58: U stolu se sešlo n lidí. Když si chtějí připít, kolik se ozve cinknutí?

Cvičení 11a.59: Kolik má konvexní n -úhelník úhlopříček?

Cvičení 11a.60: Nechť M je množina. Kolik lze vytvořit uspořádaných dvojic (A, B) takových, že $A \subseteq B \subseteq M$? Návod: Uvažujme takovou dvojici. Pak každý $x \in M$ patří do právě jedné z $A, A - B, M - B$.

Cvičení 11a.61 (poučné): Dokažte, že (viz Fakt 11a.4)

- (i) pro všechna $n \in \mathbb{N}_0$ platí $\binom{n}{0} = 1$;
- (ii) pro všechna $n \in \mathbb{N}_0$ platí $\binom{n}{1} = n$.
- (iii) pro všechna $k \leq n \in \mathbb{N}_0$ platí $\binom{n}{k} = \binom{n}{n-k}$.

Řešení:**11a.1:** a) $5! = 120$. b) $5 \cdot 4 \cdot 3 = 60$.**11a.2:** 3^{100} .**11a.3:** a) $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 151200$. b) $10^6 = 1000000$. c) $\binom{10}{6} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4!} = 210$.d) $\binom{10+6-1}{6} = \binom{15}{6} = \frac{15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10}{6!} = 5005$.**11a.4:** $9 \cdot 26 \cdot 9 \cdot 10^4 = 21060000$.**11a.5:** $8 \cdot 7 \cdot 6 = 336$.**11a.6:** Počty jsou v zásadě irrelevantní, podstatné je, že od každého je alespoň osm kusů. Odhadneme, že na pořadí nezáleží, $\binom{2+8-1}{8} = \binom{9}{8} = 9$. Logické, vybíráme, kolik vezmeme dvojkorun, možnosti od žádné po osm. Kdyby na pořadí záleželo, tak je to $2^8 = 256$ možností.**11a.7:** $\binom{8}{5} = 56$.**11a.8:** $\binom{11}{3} \cdot \binom{7}{3} = 5775$.**11a.9:** Každá mince si vybírá (s opakováním) z pěti typů, na pořadí nezáleží (jsou volně ložené v prasátku), tedy $\binom{5+20-1}{20} = \binom{24}{20} = \frac{24 \cdot 23 \cdot 22 \cdot 21}{4!} = 10626$.**11a.10:** a) $5! = 120$. b) 0.**11a.11:** Každá kniha si vybírá skladisti, s opakováním, na pořadí nezáleží, proto $\binom{3+3000-1}{3000} = \binom{3002}{3000} = 4504501$.**11a.12:** Je to otázka na seřazení písmen, tedy permutace a některá písmena se opakují: $\frac{15!}{4!4!3!} = 15765750$.**11a.13:** Balónky si vybírají koše s opakováním, $\binom{9+6-1}{6} = \binom{14}{6} = 3003$.**11a.14:** Buď pivo nebo víno, sčítací princip. Pět výběrů ze tří vín s opakováním, na pořadí záleží, tedy 3^5 , podobně pivo, celkem $3^5 + 2^5 = 275$.**11a.15:** $\binom{200}{40}, \sum_{k=0}^{40} \binom{200}{k}$.**11a.16:** a) $\frac{10!}{2!3!5!} = 2520$. b) Nutno vyloučit nulu na prvním místě, tedy dvě možnosti, co tam dát: $\frac{9!}{2!2!5!} + \frac{9!}{2!3!4!} = 2016$. Nebo odečteme od řetězců ty s nulou na začátku: $\frac{10!}{2!3!5!} - \frac{9!}{1!3!5!} = 2016$ **11a.17:** $(2^7 - 1) \cdot (2^{24} - 2) + 2^{14} \cdot (2^{16} - 2) + 2^{21} \cdot (2^8 - 2) = 3737091842$.**11a.18:** $2^{100} - (100 + 1)$.**11a.19:** a) $2^{10} = 1024$. b) $\binom{10}{2} = 45$. c) $\binom{10}{5} = 252$. d) $\frac{1}{2}(1024 - 252) = 386$. e) Nijak. f) Nové a) bude 11 (nula orlů, jeden orel až deset orlů), nové b) a c) jsou 1, nové d) je 5 (šest, sedm až deset hlav).**11a.20:** a) $\binom{5+3-1}{3} = \binom{7}{3} = 35$. b) $\binom{5}{3} = 10$. c) Nejde to.

d) Nejprve vezmeme jednu od každého druhu a šestou čokoládu dobereme, 5 možností.

e) Možnosti se třemi lískovými natvrdo plus zbývající volný výběr, to je $\binom{5+3-1}{3} = 35$ možností, od toho odečteme možnosti se třemi lískovými a dvěma hořkými natvrdo, dobereme tedy šestou 5 možností, odpověď je $35 - 5 = 30$.**11a.21:** a) $6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 = 720$, pokud chodí do školy každý pracovní den.

b) Odečteme-li prázdniny (dva vánoční týdny, v létě 9 týdnů a jarní prázdniny), zbývá 40 týdnů, tedy může si školy užívat 18 let.

11a.22: $\frac{8!}{2!2!4!} = 420$.**11a.23:** Permutujeme celky BA, FEC a tři písmena, tedy $5! = 120$.**11a.24:** Doplňkem, $4! - 3! = 18$.**11a.25:** $\frac{10!}{(2!)^5} - \frac{9!}{(2!)^4} = 90720$.**11a.26:** Nerůká se „všech písmen“, takže se musí brát i podmnožiny písmen. Ze všech čtyř se vytvoří $\frac{4!}{2!2!} = 6$ řetězců. Tři písmena ze čtyř je možno obecně vybrat $\binom{4}{3} = 4$ způsoby, ale tady máme opakování písmena, jsou tedy jen dva způsoby: $\{A, A, H\}$ a $\{A, H, H\}$. Každý z nich vede na $\frac{3!}{2!1!} = 3$ řetězce, celkem tedy $2 \cdot 3 = 6$ třípísmenných řetězců. Podobně výběry dvou písmen jsou jen tři $\{A, A\}$, $\{A, H\}$ a $\{H, H\}$, první a poslední dávají jeden řetězec, prostřední dává dva, celkem tedy $1 + 2 + 1 = 4$ dvoupísmenné řetězce. Jednopísmenné řetězce jsou jen dva, A a H . Celkem: 18 řetězců.**11a.27:** $\binom{5}{3} \binom{7}{1} + \binom{5}{2} \binom{7}{2} + \binom{5}{1} \binom{7}{3} = 455$.**11a.28:** a) Fáze oplatek, zmrzliny, posyp (tam zavedeme třetí možnost „žádný“), $3 \cdot 8^3 \cdot 3 = 4608$ b) Možnosti posypu jsou čtyři (nic, čoko, oříšky, čoko a oříšky), $3 \cdot 8 \cdot 7 \cdot 6 \cdot 4 = 4032$.**11a.29:** Tucet je 12. a) $\binom{10}{6} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{24} = 210$. b) 10; c) $\binom{10+6-1}{6} = \binom{15}{6} = 5005$. d) Doplňkem c) bez a) 4795. e) Vybereme čtyři jahodové a dobereme další dvě bez omezení, $\binom{10+2}{2} = \binom{11}{2} = 55$. f) doplňkem c) bez e) 4950.**11a.30:** Máme A čtyřikrát, L dvakrát a tři písmena po jednom. Ze všech je $\frac{9!}{4!2!}$ permutací.Osmipísmenné řetězce: Vynecháme A , je jich $\frac{8!}{3!2!}$. Vynecháme L , je jich $\frac{8!}{4!}$. Vynecháme jiné písmeno, je jich $\frac{8!}{4!2!}$. Celkem $\frac{9!}{4!2!} + \frac{8!}{3!2!} + \frac{8!}{4!} + 3 \cdot \frac{8!}{4!2!} = 15120$.**11a.31:** Zástupce některé skupiny bude dvakrát, podle toho to rozdělíme na disjunktní případy:

$$\binom{2500}{2} \binom{1200}{1} \binom{300}{1} + \binom{2500}{1} \binom{1200}{2} \binom{300}{1} + \binom{2500}{1} \binom{1200}{1} \binom{300}{2} = \frac{2500 \cdot 2499}{2} 1200 \cdot 300 + 2500 \frac{1200 \cdot 1199}{2} 300 + 2500 \cdot 1200 \frac{300 \cdot 299}{2}$$

$$= \frac{1}{2} 2500 \cdot 1200 \cdot 300(2499 + 1199 + 299) = 1798650000000.$$

11a.32: $26^3 \cdot 10^4 + 26^2 \cdot 10^5 = 243360000$.

11a.33: Stačí vybrat písmena pro první polovinu slova, $26^{n/2}$ pro n sudé, $26^{(n+1)/2}$ pro n liché, dá se to napsat jako $26^{\lceil n/2 \rceil}$.

11a.34: a) Vybereme zda ženich nalevo či napravo, pak usadíme tuto dvojici, pak rozesadíme ostatní: $2 \cdot 5 \cdot 4! = 240$, b) ze symetrie $\frac{1}{2} 6! = 360$; c) doplňkem $6! - 2 \cdot 5! = 480$.

$$\text{11a.35: } 53 + 53 \cdot 63 + 53 \cdot 63^2 + \dots = 53 \sum_{k=0}^7 63^k = 53 \frac{1-63^8}{1-63} = 212133167002880.$$

$$\text{11a.36: a) } \binom{7}{2} 6^2 \cdot 20^5 = 2419200000. \text{ b) } 6 \cdot 26^5 + 6 \cdot 26^5 = 142576512. \text{ c) } \binom{7}{2} 6 \cdot 5 \cdot 20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 = 1172102400 \text{ a} \\ 6 \cdot 24 \cdot 23 \cdot 22 \cdot 21 \cdot 20 + 6 \cdot 24 \cdot 23 \cdot 22 \cdot 21 \cdot 20 = 61205760.$$

11a.37: Všech je 2^4 a tři nuly po sobě mají řetězce 0001, 0000, 1000, tedy odpověď je $2^4 - 3 = 13$.

11a.38: Těch je asi málo, zkusíme výpisem, chce to systém, tak nejprve tři nuly na začátku, pak tři nuly uprostřed, pak tři nuly na konci, ale hlídáme, jestli už to nebylo předtím: 00000, 00001, 00010, 00011, 10000, 10001, 01000, 11000. Je jich 8.

11a.39: (i) Počet různých pozic mužů a žen: Vybíráme pět míst pro ženy z $7+2$ možností mezi muži a na koncích, $\binom{9}{5}$. Pro pevnou pozici mužů a žen zpermutujeme konkrétní muže a ženy, $8! \cdot 5!$. Celkem: $\binom{9}{5} 8! \cdot 5! = 609638400$.

(ii) $\binom{m+1}{n} n! m! = \frac{(m+1)! m!}{(m+1-n)!}$ pokud $m+1 \geq n$, jinak 0.

11a.40: Seřadíme děti a odpočítáme po dvou, pak zrušíme pořadí v lodičkách: $\frac{8!}{(2!)^4} = 2520$. Nebo vybíráme postupně: $\binom{8}{2} \binom{6}{2} \binom{4}{2} \binom{2}{2}$.

11a.41: Jsou dvě situace, jedna začíná kluk-holka a pokračuje stejně, druhá začíná holka-kluk. Pro danou situaci je pak třeba nezávisle naplnit kluky a holky: $2(n!)^2$.

11a.42: a) Držíme je u sebe, takže permutujeme 98 celků, proto $98!$ permutací. b) Nejprve najdeme místa pro ta tři čísla mezi stovkou, pak na zbytek míst zpermutujeme zbytek, proto $\binom{100}{3} 97!$.

11a.43: a) Vybereme další číslo, a to buď před nebo za ty tři, $97 + 97 = 194$. b) Vybereme další číslo a pak ještě na kterou ze čtyř pozic přijde, $97 \cdot 4 = 388$. c) Vybereme číslo mezi 1 a 98, s ním automaticky přijde dvojice za ním, pak dobereme čtvrté číslo a dáme jej za či před, tedy sčítáme možnosti. Pokud by ale šlo o čtyři čísla za sebou, tak jsme to započítali dvakrát, nutno odečíst, $98 \cdot 97 + 98 \cdot 97 - 97 = 18915$. d) Jako v c), ale nutno pro čtvrté číslo vybrat pozici, $98 \cdot 97 \cdot 4 - 97 = 37927$. e) Je pět možností, kam do vybrané sedmice umístit tu trojici, pak dobereme zbývající čísla, $5 \cdot 97 \cdot 96 \cdot 95 \cdot 94 = 415780800$. f) Vybereme místa pro ty tři a dobereme zbývající, $\binom{7}{3} \cdot 97 \cdot 96 \cdot 95 \cdot 94 = 2910465600$.

11a.44: Tucet je 12, tudíž hlavní je, že je od každého alespoň kolik kusů. Odhadneme, že na pořadí výběru nezáleží. Rovnou vezmeme tři od každé barvy, to je celkem devět jablek, zvolíme tři, na pořadí nezáleží a díky velkému počtu můžeme s opakováním, tedy $\binom{3+3-1}{3} = 10$ způsobů.

Co bychom dělali, kdyby na pořadí záleželo? Asi nejsnažší by byl rozbor situací. Rozkládáme 12 na tři čísla, každé alespoň 3.

Možnost 3-3-6: nejprve zvolíme, které jablko bude šestkrát, pak permutujeme, celkem $3 \cdot \frac{12!}{3!3!6!}$ možností.

Možnost 3-4-5: nejprve zvolíme, které jablko bude pětkrát, pak jablko pro čtyři kusy a nakonec permutujeme, celkem $3 \cdot 2 \cdot \frac{12!}{3!4!5!}$ možností.

Možnost 4-4-4: tady jen permutujeme, celkem $\frac{12!}{4!4!4!}$ možností.

Dohromady $\frac{12!}{4!5!6!} [3 \cdot 4 \cdot 4 \cdot 5 + 3 \cdot 2 \cdot 4 \cdot 6 + 5 \cdot 5 \cdot 6] = 256410$.

11a.45: Osm dvojic 01, mezi ně je třeba dát zbývající dvě nuly, místo je 9. Buď se dávají po jednom nebo obě najednou na jedno místo, $\binom{9}{2} + 9 = 45$.

Dvě jedničky nejdou, na to je jich málo.

11a.46: $(n-1)!$, viz příklad 11a.n h).

11a.47: a) Výběr zda začneme muži či ženami, pak obsazení jasné, jen permutujeme muže a nezávisle ženy, $2(n!)^2$. b) Nejprve permutujeme páry jako celky, pak určíme pořadí pro každý pár zvlášť, $n!2^n$. c) Výsledky se vydělí počtem míst $2n$.

11a.48: První znak 0: celkem 8 možností (za 0 nic, jedna jednička, dvě atd.). První znak 1: celkem 4 možnosti (za 1 jedna jednička, tři, pět, sedm). Tedy $8 + 4 = 12$ možností.

11a.49: a) Protože se nevracíme, je naše svoboda silně omezena, musíme se posunout n -krát nahoru a m -krát doprava, možnost volby je jen v tom, v jakém pořadí ty kroky provedeme. Jinými slovy, jde o počet permutací těchto kroků, ještě jinak řešeno, vybíráme, kam v celkem $m+n$ krocích umístíme ty doprava, což je $\binom{m+n}{n}$ možností.

b) Zde je zase dán přesný počet kroků na stranu, na kolmou stranu a nahoru, jde jen o jejich uspořádání. Odpověď zní $\binom{m+n+k}{m,n,k} = \frac{(m+n+k)!}{m!n!k!}$.

11a.50: První tři cifry libovolné, tedy 10^3 možností. Poslední dvě cifry se volí najednou, je to jedna fáze. Kolik je možností pro poslední dvojici? Rozbor podle čtvrté číslice: $0 \mapsto 00, 07, 1 \mapsto 16, 2 \mapsto 25, 3 \mapsto 34, 4 \mapsto 43, 5 \mapsto 52, 59, 6 \mapsto 61, 68, 7 \mapsto 70, 77, 8 \mapsto 86, 9 \mapsto 95$, celkem 14 možností. Je 14000 kódů.

11a.51: Jde o čísla jednociferná, dvouciferná a tříciferná bez omezení a také jedno čtyřciferné 1000, které změní počty pro jedničku a nulu.

c)–e) Uvažujme číslice různou od 0 a 1, čísla si představujeme jako xyz . Na místě x se číslice může vyskytovat stokrát, signalizuje příslušnou stovku. Na pozici y se číslice může vyskytovat desetkrát v každé stovce, stovek je 10, celkem sto výskytů. Na pozici z se číslice vyskytuje jednou v každé desítce, těch je 100. Celkem 300 výskytů.
 b) Jednička má jeden výskyt navíc v tisícovce, je jich 301.

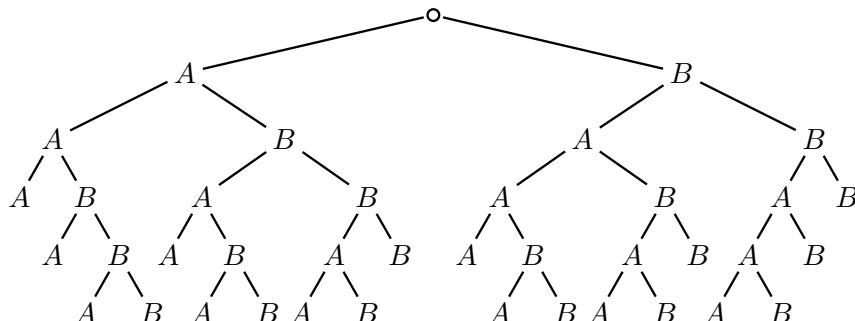
a) V tisícovce jsou tři nuly, dále se díváme jen na čísla do 999. Pak nula nemůže být na pozici x . Na pozici y je nula desetkrát v každé stovce kromě první (dvouciferná čísla), z těch 10 stovek je to tedy jen devět, celkem 90 výskytů. Na pozici z je nula jedenkrát v každé desítce s výjimkou první desítky (jednociferná čísla), celkem je těchto desítek $100 - 1 = 99$. Celkem 192 nul.

11a.52: Dáme každé otázce 5 bodů, zbývá tedy rozdělit 25 mezi pět otázeek, jiný pohled: Každému z 25 bodů přidělit otázku, tedy $\binom{5+25-1}{25} = \binom{29}{25} = 23751$. Jiný pohled: Kolik celočíselných řešení splňujících $a_k \geq 5$ má rovnice $x_1 + x_2 + x_3 + x_4 + x_5 = 50$?

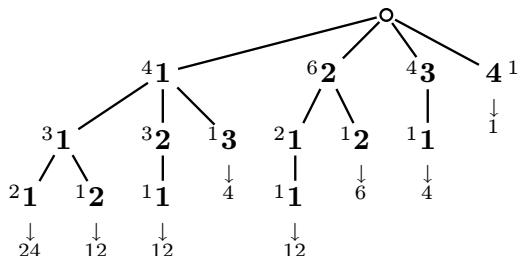
11a.53: Máme 17 jedniček. Nejprve dáme jednu do x_1 , čtyři do x_3 a dvě do x_4 . Zbývajících 10 si vybírá, do které x_i půjde, na pořadí nezáleží a je možné se opakovat. Počet řešení je proto $\binom{4+10-1}{10} = \binom{13}{10} = 286$.

11a.54: Nezbývá než jít na to výčtem. Je jich 16: $\{3\}$, $\{3, 7\}$, $\{3, 9\}$, $\{3, 11\}$, $\{3, 24\}$, $\{3, 7, 9\}$, $\{3, 7, 11\}$, $\{3, 9, 11\}$, $\{7\}$, $\{7, 9\}$, $\{7, 11\}$, $\{7, 9, 11\}$, $\{9\}$, $\{9, 11\}$, $\{11\}$, $\{24\}$.

11a.55:



11a.56: Vzhledem k možnosti remíz se na to musí rozbořem, třeba stromem, kde je vyznačeno, kolik lidí dostane kterou medaili, dole pak je počet možností. Celkem jich je 75.



11a.57: Chlapci vybírají, do kterého místa mezi dívky si který sedne, míst je $8 + 1$ (mohou být i na kraji), $\binom{9}{6} = \frac{9 \cdot 8 \cdot 7}{6} = 84$.

11a.58: Viz příklad 11a.p b). Každý cinkne $(n - 1)$ -krát, ale každé cinknutí započítáno dvakrát, $\frac{1}{2}n(n - 1)$.

11a.59: Každý vrchol má $n - 3$ dalších k propojení, každá úhlopříčka pak je započtena dvakrát, $n(n - 3)/2$.
11a.60: Platí to i naopak, když každému x přidělíme kategorii 0, 1 či 2, dostaneme množiny A (prvky kategorie 0) a B (prvky kategorií 0, 1). Tedy počet dvojic je roven počtu rozhození kategorií 0, 1, 2 mezi prvky M , to je 3^n možností.

11b. Pokročilejší principy

V této kapitole probereme Princip inkluze a exkluze, poté shrneme některé poznatky této a předchozí kapitoly z pohledu rozdělování objektů do krabiček. Dále si představíme Dirichletův šuplíkový princip a nakonec se vrátíme k rekurentním rovnicím jako nástroji pro řešení některých kombinatorických úloh.

Nejprve se vrátíme k jednomu problému, který nám z první kapitoly zůstal nevyřešen.

! Příklad 11b.a (pokračování 11a.m): Ve třídě je 150 kluků a 40 holek. Chtějí poslat čtyřčlennou delegaci k přednášejícímu. Kolika způsoby ji mohou vybrat, když chtějí, aby šel Adam nebo Bára?

Přímý útok: Dáme do delegace Adama a navolíme ostatní: $\binom{189}{3}$ možností. Nebo dáme do delegace Báru a navolíme zbytek. Celkový počet možností ale není $\binom{189}{3} + \binom{189}{3}$, protože některé volby jsou započítány dvakrát, jmenovitě ty s Adamem i Bárou.

V takovéto situaci bychom normálně utekli k jevu opačnému. Pokud nechceme mít vybrány Adama ani Báru, máme $\binom{188}{4} = 50404915$ možností. Celkem je jich $\binom{190}{4}$ (viz příklad 11a.m), takže těch možností výběrů, kde je Adam nebo Bára, je $\binom{190}{4} - \binom{188}{4} = 2197250$.

Nastává čas zkusit dotáhnout do konce přímý útok. Sečtením výběrů s Adamem a výběrů s Bárou jsme přepočítali, takže se nabízí nápad, že bychom jednou odečetli to, co jsme započítali dvakrát. Jde o výběry s Adamem i Bárou, u nich pak dobíráme další dva členy do delegace. Počet možností je tedy $\binom{188}{2}$. Celkový počet možností výběrů s Adamem nebo Bárou by tedy měl být

$$\binom{189}{3} + \binom{189}{3} - \binom{188}{2} = 219750.$$

Dostali jsme stejný výsledek jako při postupu přes doplněk, úvaha tedy byla správná.

△

Princip použitý při přímém řešení lze vyjádřit i jinak. Nechť je A množina všech výběrů, ve kterých je Adam, a B množina všech výběrů, ve kterých je Bára. Nás zajímá $|A \cup B|$ a selským rozumem jsme místo toho počítali $|A| + |B| - |A \cap B|$. Není to náhoda, totéž jsme už viděli jako Fakt 2c.9. Byla tam i verze pro tři množiny a teď vidíme, že jde o něco, co se při kombinatorických úvahách může silně hodit. Zobecníme si to pro libovolný (konečný) počet množin.

! Věta 11b.1. (Princip inkluze a exkluze, principle of inclusion and exclusion)

Jsou-li A_i pro $i = 1, 2, \dots, n$ konečné množiny, pak

$$\begin{aligned} \left| \bigcup_{i=1}^n A_i \right| &= \sum_{i=1}^n |A_i| - \sum_{i < j} |A_i \cap A_j| + \sum_{i < j < k} |A_i \cap A_j \cap A_k| - \dots + (-1)^{n-1} \left| \bigcap_{i=1}^n A_i \right| \\ &= \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} |A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}|. \end{aligned}$$

V důkazu použijeme ten první zápis, sice je o dost delší, ale zase je tam lépe vidět, co se děje.

Důkaz (z povinnosti): Povedeme jej indukcí na n .

(0) $n = 1$: To je triviální, $|A_1| = |A_1|$.

(1) Nechť $n \in \mathbb{N}$. Předpokládejme, že princip inkluze a exkluze platí pro libovolných n množin, a uvažujme nějaké množiny $A_1, A_2, \dots, A_n, A_{n+1}$.

Označme $B = \bigcup_{i=1}^n A_i$, pak $\bigcup_{i=1}^{n+1} A_i = B \cup A_{n+1}$, proto $\left| \bigcup_{i=1}^{n+1} A_i \right| = |B| + |A_{n+1}| - |B \cap A_{n+1}|$ podle Faktu 2c.9.

Tedě použijeme indukční předpoklad na B , což je sjednocení n množin, a také deMorganův zákon (Věta 2a.8) na množinu $B \cap A_{n+1} = (\bigcup A_k) \cap A_{n+1}$:

$$\left| \bigcup_{i=1}^{n+1} A_i \right| = \sum_{i=1}^n |A_i| - \sum_{i < j \leq n} |A_i \cap A_j| + \sum_{i < j < k \leq n} |A_i \cap A_j \cap A_k| - \dots + (-1)^n \left| \bigcap_{i=1}^n A_i \right| + |A_{n+1}| - \left| \bigcup_{k=1}^n (A_k \cap A_{n+1}) \right|.$$

Přesuneme $|A_{n+1}|$ do první sumy a na to poslední sjednocení n množin zase použijeme indukční předpoklad.

$$\begin{aligned} \left| \bigcup_{i=1}^{n+1} A_i \right| &= \sum_{i=1}^{n+1} |A_i| - \sum_{i < j \leq n} |A_i \cap A_j| + \sum_{i < j < k \leq n} |A_i \cap A_j \cap A_k| - \dots + (-1)^n \left| \bigcap_{i=1}^n A_i \right| \\ &\quad - \sum_{i=1}^n |A_i \cap A_{n+1}| + \sum_{i < j \leq n} |A_i \cap A_j \cap A_{n+1}| - \sum_{i < j < k \leq n} |A_i \cap A_j \cap A_k \cap A_{n+1}| + \dots - (-1)^n \left| A_{n+1} \cap \bigcap_{i=1}^n A_i \right|. \end{aligned}$$

Tedě spojíme sumy přes průniky dvou množin (mají stejné znaménko), sumy přes průniky tří množin atd. a dostaneme

$$\left| \bigcup_{i=1}^{n+1} A_i \right| = \sum_{i=1}^{n+1} |A_i| - \sum_{i < j} |A_i \cap A_j| + \sum_{i < j < k} |A_i \cap A_j \cap A_k| - \dots + (-1)^{n+1} \left| \bigcap_{i=1}^{n+1} A_i \right|.$$

□

! Příklad 11b.b: Heslo (password) se skládá ze tří znaků, které mohou být písmena (lower-case) či číslice.

a) Kolik je možné vytvořit hesel, ve kterých se vyskytuje alespoň jedna číslice?

Označíme si množinu takových hesel jako P , potřebujeme najít $|P|$. Jedna možnost je jít na to přes doplněk. Všech tříznakových hesel je $(26 + 10)^3 = 36^3$. Chceme-li hesla bez číslic, pak je tvoříme jen z písmen, je jich 26^3 , proto je hesel dle zadání $36^3 - 26^3 = 29080$.

Alternativa: Rozložíme si $P = Q_1 \cup Q_2 \cup Q_3$, kde Q_i jsou hesla s číslicí na i -tém místě. Jelikož tyto množiny nejsou disjunktní (například heslo 1a3 je v Q_1 i v Q_3), tak nelze jejich mohutnosti jen sčítat, ale musíme použít princip inkluze a exkluze:

$$|P| = |Q_1| + |Q_2| + |Q_3| - |Q_1 \cap Q_2| - |Q_1 \cap Q_3| - |Q_2 \cap Q_3| + |Q_1 \cap Q_2 \cap Q_3|.$$

Jaké jsou příslušné velikosti? Pro libovolné i je $|Q_i| = 10 \cdot 36^2$, volíme jednu číslici a dva znaky. Při $i \neq j$ máme podobně $|Q_i \cap Q_j| = 10^2 \cdot 36$ (volíme dvě číslice a znak) a také $|Q_1 \cap Q_2 \cap Q_3| = 10^3$ (tři číslice). Proto $|P| = 3 \cdot 10 \cdot 36^2 - 3 \cdot 10^2 \cdot 36 + 10^3 = 29080$.

Dopadlo to stejně, ale dalo to víc práce.

b) Kolik je možno vytvořit hesel o třech znacích, které mají alespoň dvě číslice?

Tady už přechod k doplňku nepomůže. Označme jako Q_{ij} hesla, která mají na místech i, j číslice. takové množiny jsou tři, Q_{12}, Q_{13}, Q_{23} , a platí $|Q_{ij}| = 10^2 \cdot 36 = 3600$. Zase nejsou disjunktní, použijeme princip inkluze a exkluze:

$$|Q_{12} \cup Q_{13} \cup Q_{23}| = |Q_{12}| + |Q_{13}| + |Q_{23}| - |Q_{12} \cap Q_{13}| - |Q_{12} \cap Q_{23}| - |Q_{13} \cap Q_{23}| + |Q_{12} \cap Q_{13} \cap Q_{23}|.$$

Všechny průniky jsou zde stejné, jde o hesla, která mají samé číslice, tedy velikosti jsou 10^3 . Počet hesel je proto $3 \cdot 3600 - 3 \cdot 10^3 + 10^3 = 8800$.

△

Situace, kdy při určování velikosti nezáleží na konkrétních množinách, ale na skupině, kterou zkoumáme (průniky dvou, průniky tří atd.), se objevuje velice často. Vyplatí se udělat si na to samostatné tvrzení.

! Důsledek 11b.2.

Nechť jsou A_i pro $i = 1, 2, \dots, n$ konečné množiny. Předpokládejme, že pro libovolné i je $|A_i| = m_1$, pro libovolné $i < j$ je $|A_i \cap A_j| = m_2$, pro libovolné $i_1 < i_2 < i_3$ je $|A_{i_1} \cap A_{i_2} \cap A_{i_3}| = m_3$ atd. až po $\left| \bigcap_{i=1}^n A_i \right| = m_n$.

Pak

$$\left| \bigcup_{i=1}^n A_i \right| = nm_1 - \binom{n}{2}m_2 + \dots + (-1)^{n-2} \binom{n}{n-1}m_{n-1} + (-1)^{n-1}m_n = \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} m_k.$$

Důkaz (rutinní, poučný): vyjdeme z Věty 11b.1. Víme, jak velké jsou množiny v jednotlivých součtech, stačí si tedy rozmyslet, kolikrát se v každé sumě sčítá. V sumě $\sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n}$ se vybírá k indexů i_1, \dots, i_k z množiny $\{1, 2, \dots, n\}$, přičemž se výběr nesmí opakovat a na pořadí nezáleží, protože se pak stejně srovnají podle velikosti. Počet možností je tedy dán kombinačním číslem $\binom{n}{k}$ a žádaný vzorec okamžitě plyne z principu inkluze a exkluze.

□

Ukážeme si teď jednu zajímavou aplikaci tohoto vzorce.

! Fakt 11b.3.

Uvažujme množinu A o m prvcích a množinu B o n prvcích. Tvrdíme, že jestliže je $m \geq n$, pak existuje $\sum_{k=0}^{n-1} (-1)^k \binom{n}{k} (n-k)^m$ zobrazení z A na B .

Důkaz (poučný): Celkem je zobrazení n^m (příklad 11a.j), počet zobrazení „na“ zjistíme doplňkem přes počet zobrazení, u kterých tato vlastnost selhala.

Aby zobrazení nebylo na, musí nějaký prvek v cílové množině vynechat, nechť M_i je množina všech zobrazení, které vynechaly i -tý prvek. Taková zobrazení jsou vlastně zobrazeními z množiny o velikosti m do množiny o velikosti $n-1$, proto je jich $|M_i| = (n-1)^k$.

Tyto množiny evidentně nejsou disjunktní (zobrazení, které vynechá více prvků z B , je i ve více M_i), proto musíme aplikovat princip inkluze a exkluze. Typický člen v onom vzorci je $\sum_{i_1 < i_2 < \dots < i_k} |M_{i_1} \cap M_{i_2} \cap \dots \cap M_{i_k}|$.

Protože jsou i_1, i_2, \dots, i_k navzájem různá pořadová čísla prvků z cílové množiny, pak $M_{i_1} \cap M_{i_2} \cap \dots \cap M_{i_k}$

jsou zobrazení, která určitě vynechala prvky i_1 až i_k , jdou tedy do cílové množiny o velikosti $n - k$ a takových je $(n - k)^m$. Je dobré si rozmyslet, že to platí i pro trochu extrémní případ: Průnik všech množin M_i je možina zobrazení, které vynechají všechny prvky z cílové množiny, taková zobrazení nejsou čili velikost je 0. To souhlasí, vzorec pak dává $0^m = 0$. Takže jsme získali všechny údaje nutné pro vzorec z Důsledku 11b.2. Počet zobrazení, která vynechaly nějaký prvek z cílové množiny, je

$$\sum_{k=1}^n (-1)^{k-1} \binom{n}{k} (n-k)^m.$$

Proto je počet surjektivních zobrazení („na“) roven

$$\begin{aligned} n^m - \sum_{k=1}^n (-1)^{k-1} \binom{n}{k} (n-k)^m &= 1 \cdot (n-0)^m + \sum_{k=1}^n (-1)^k \binom{n}{k} (n-k)^m \\ &= \sum_{k=0}^n (-1)^k \binom{n}{k} (n-k)^m = \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} (n-k)^m. \end{aligned}$$

Vynechali jsme v sumě poslední index $k = n$, protože příslušný sčítanec je stejně nulový. \square

Ne vždy jsou ovšem průniky stejného typu stejně velké.

Příklad 11b.c: Kolik je prvočísel v množině $M = \{1, 2, \dots, 100\}$?

Řešení: Zkusíme to přes doplněk. Která čísla určitě nejsou prvočísla? Protože $\sqrt{100} = 10$, stačí se zeptat, kolik je v M čísel dělitelných prvočísly menšími než 11, což znamená 2, 3, 5 a 7.

Označme jako M_n množinu čísel mezi 1 a 100 dělitelných n . Podle příkladu 6a.b je $|M_n| = \lfloor \frac{100}{n} \rfloor$. Tyto množiny ovšem nejsou disjunktní a je třeba použít princip inkluze a exkluze. Protože pracujeme s různými prvočísly, pak $M_m \cap M_n$ jsou čísla dělitelná číslem mn a tedy $|M_m \cap M_n| = \lfloor \frac{100}{mn} \rfloor$; podobně zpracujeme průniky tří a čtyř množin.

Princip inkluze a exkluze pak dává počet čísel dělitelných 2, 3, 5 či 7 jako

$$\begin{aligned} \lfloor \frac{100}{2} \rfloor + \lfloor \frac{100}{3} \rfloor + \lfloor \frac{100}{5} \rfloor + \lfloor \frac{100}{7} \rfloor - \lfloor \frac{100}{2 \cdot 3} \rfloor - \lfloor \frac{100}{2 \cdot 5} \rfloor - \lfloor \frac{100}{2 \cdot 7} \rfloor - \lfloor \frac{100}{3 \cdot 5} \rfloor - \lfloor \frac{100}{3 \cdot 7} \rfloor - \lfloor \frac{100}{5 \cdot 7} \rfloor \\ + \lfloor \frac{100}{2 \cdot 3 \cdot 5} \rfloor + \lfloor \frac{100}{2 \cdot 3 \cdot 7} \rfloor + \lfloor \frac{100}{2 \cdot 5 \cdot 7} \rfloor + \lfloor \frac{100}{3 \cdot 5 \cdot 7} \rfloor - \lfloor \frac{100}{2 \cdot 3 \cdot 5 \cdot 7} \rfloor \\ = 50 + 33 + 20 + 14 - 16 - 10 - 7 - 6 - 4 - 2 + 3 + 2 + 1 + 0 - 0 = 78. \end{aligned}$$

Toto ovšem nejsou všechno čísla složená, čísla 2, 3, 5 a 7 samotná jsou totiž také svými násobky, aniž by byla složená. Počet čísel složených v množině M je tedy $78 - 4 = 74$. Ta zbývající (je jich 26) jsou buď prvočísla, nebo 1, které prvočíslo není, takže prvočísel je 25. \triangle

! Příklad 11b.d: Kolik řešení $x_1, x_2, x_3 \in \mathbb{N}_0$ rovnice $x_1 + x_2 + x_3 = 11$ splňuje $x_1 \leq 3$, $x_2 \leq 4$ a $x_3 \leq 6$?

Umíme dobře hledat počty řešení s nerovnicí u podmínky v opačném směru, takže to zkusíme doplněkem. Množina M všech řešení z \mathbb{N}_0 (bez omezení) má mohutnost $|M| = \binom{3+11-1}{11} = \binom{13}{11} = 78$, viz příklad 11a.k.

Nechť M_1 je možina řešení splňujících $x_1 \geq 4$ (ta tedy nechceme), máme $|M_1| = \binom{3+7-1}{7} = \binom{9}{7} = 36$ (viz část b) příkladu 11a.k). Podobně pro možinu M_2 řešení splňujících $x_2 \geq 5$ máme $|M_2| = \binom{3+6-1}{6} = \binom{8}{6} = 28$ a pro množinu M_3 řešení splňujících $x_3 \geq 7$ máme $|M_3| = \binom{3+4-1}{4} = \binom{6}{4} = 15$.

Počet řešení, která nás zajímají, je $|M - (M_1 \cup M_2 \cup M_3)| = |M| - |M_1 \cup M_2 \cup M_3|$. Protože množiny M_i nejsou disjunktní, budeme muset použít princip inkluze a exkluze. Nejprve počítáme: Množina $M_1 \cap M_2$ jsou řešení splňující $x_1 \geq 4$ a $x_2 \geq 5$, proto $|M_1 \cap M_2| = \binom{3+2-1}{2} = \binom{4}{2} = 6$, podobně $|M_1 \cap M_3| = \binom{3+0-1}{0} = \binom{2}{0} = 1$, $|M_2 \cap M_3| = 0$ (evidentní, z čísel 5 a 7 či větších není možné získat 11) a také $|M_1 \cap M_2 \cap M_3| = 0$.

Proto počítáme:

$$\begin{aligned} |M - (M_1 \cup M_2 \cup M_3)| &= |M| - |M_1 \cup M_2 \cup M_3| \\ &= |M| - |M_1| - |M_2| - |M_3| + |M_1 \cap M_2| + |M_1 \cap M_3| + |M_2 \cap M_3| - |M_1 \cap M_2 \cap M_3| \\ &= 78 - 36 - 28 - 15 + 6 + 1 + 0 - 0 = 6. \end{aligned}$$

Je tedy šest takových řešení.

Tak málo řešení by mělo jít zjistit rozborem situací, jde o možnosti 1-4-6, 2-3-6, 2-4-5, 3-2-6, 3-3-5, 3-4-4.

U úloh s více proměnnými by asi už stálo za to použít počítač, protože pak ani princip inkluze a exkluze nebude příliš příjemný. Pak je ale třeba vyvinout algoritmus, který opravdu najde všechny možnosti, viz 11d.7.

\triangle

Příklad 11b.e (pokračování 11a.v): Uvažujme rovnici $m + p + c = 9$, kde požadujme $m, p, c \in \mathbb{N}_0$, $m \leq 5$, $p \leq 3$ a $c \leq 5$. Označme množinu M všech řešení z \mathbb{N}_0 bez omezení, těch je $|M| = \binom{3+9-1}{9} = \binom{11}{9} = 55$.

Nechť M_m je možina řešení splňující $m \geq 6$ (ta tedy nechceme), máme $|M_m| = \binom{3+3-1}{3} = \binom{5}{3} = 10$, podobně pro možinu M_p řešení splňující $p \geq 4$ máme $|M_p| = \binom{3+5-1}{5} = \binom{7}{5} = 21$ a pro množinu M_c řešení splňující $c \geq 6$ máme $|M_c| = \binom{5}{3} = 10$. Ještě potřebujeme průniky, ale tam je to snadné, každý z nich je prázdný (není například možné pomocí šesti malinových, čtyř pomerančových a případně dalších citrónových bonbónů dostat celkem devět bonbónů). Proto je počet řešení roven

$$\begin{aligned}|M - (M_1 \cup M_2 \cup M_3)| &= |M| - |M_1| - |M_2| - |M_3| + |M_1 \cup M_2| + |M_1 \cup M_3| + |M_2 \cup M_3| - |M_1 \cup M_2 \cup M_3| \\ &= 55 - 10 - 21 - 10 + 0 + 0 + 0 - 0 = 14.\end{aligned}$$

Souhlasí to s počtem získaným v příkladě 11a.v rozborem situací.

△

Příklad 11b.f: Mějme n objektů. Zajímá nás, kolik je permutací, které nezanechaly ani jeden objekt na svém místě (říká se jim derangements), označme tento počet D_n .

Přímý útok je neperspektivní, protože je těžké zachytit nějak kombinatoricky fakt, že čísla někde nejsou. Mnohem lépe se pracuje s tím, že někde jsou, jinými slovy musí se na to přes doplněk. Nechť M_i jsou permutace, které nechaly i -tý objekt na svém místě, pak jde vlastně jen o permutace ostatních objektů a $|M_i| = (n-1)!$. Tyto množiny ale nejsou disjunktní, je tedy nutno použít princip inkluze a exkluze.

Nechť $1 \leq i_1 < i_2 < \dots < i_k \leq n$. Pak je $M_{i_1} \cap M_{i_2} \cap \dots \cap M_{i_k}$ množina všech permutací, které zachovají na místě těchto k různých objektů, čili permutujeme ty zbyvající, celkem je $(n-k)!$ takových permutací. Permutací, které něco nechají na místě, je tedy

$$\sum_{k=1}^n (-1)^{k-1} \binom{n}{k} (n-k)! = \sum_{k=1}^n (-1)^{k-1} \frac{n!}{k!}.$$

Proto je počet těch ostatních roven

$$D_n = n! - \sum_{k=1}^n (-1)^{k-1} \frac{n!}{k!} = \frac{n!}{0!} + \sum_{k=1}^n (-1)^k \frac{n!}{k!} = n! \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

Alternativní způsob nalezení tohoto vzorce používá rekurzi, viz cvičení 11b.44.

Poznámka: Jde o jeden z klasických pravděpodobnostních problémů, obvykle se presentuje jako otázka, s jakou pravděpodobností odejde každý z n gentlemanů domů v cizím klobouku, když je šatnářka v klubu vydává náhodně. Odpověď zní $\frac{D_n}{n!} = \sum_{k=0}^n \frac{(-1)^k}{k!}$, což je pro větší n asi $\frac{1}{e} \sim 0.368$. Když řekneme, že všichni gentlemani půjdou domů v cizím v průměru v jednom případě ze tří, tak to vystihuje situaci dosti přesně už od $n = 3$ a ukazuje to, že by se možná vyplatila jiná šatnářka.

△

Princip inkluze a exkluze se dá s úspěchem použít ke zjišťování velikosti určitých skupin, při tom často pomůžou i Vennovy diagramy.

Příklad 11b.g: Ve třídě je 250 lidí. Když jsem se zeptal, kdo někdy pracuje na Windows, zvedlo se 235 rukou. Když jsem se zeptal, kdo někdy pracuje na Linuxu, zvedlo se 150 rukou. Za předpokladu, že každý zvedl ruku alespoň jednou a nikdo nezvedl najednou dvě (a více) rukou, kolik lidí pracuje na obou systémech?

Označme jako A množinu lidí pracujících na W a jako B množinu lidí pracujících na L. Dáno: $|A| = 235$, $|B| = 150$, $|A \cup B| = 250$. Z rovnosti $|A \cup B| = |A| + |B| - |A \cap B|$ dostaneme $|A \cap B| = |A| + |B| - |A \cup B| = 135$. △

!**Příklad 11b.h:** Máme 300 lidí. Z nich 240 umí C (myslí se tím také C, mohou umět i něco navíc), 130 umí Pascal, 27 umí Fortran, 117 umí C a Pascal (a možná i Fortran, ale nemusí), 17 umí Pascal a Fortran (a možná i C, ale nemusí), 105 umí C a Pascal ale ne Fortran, 5 umí C a Fortran ale ne Pascal. Kolik z lidí nezná ani jeden z těchto tří jazyků?

Zavedeme množiny C, P, F lidí ovládajících příslušný jazyk, daná data jsou $|C| = 240$, $|P| = 130$, $|F| = 27$, $|C \cap P| = 117$, $|P \cap F| = 17$, $|C \cap P \cap F| = 105$ a $|C \cap F - P| = 5$. Abychom zodpověděli danou otázku, musíme nejprve najít $|C \cup P \cup F|$, evidentně pomocí principu inkluze a exkluze. Na to ale nemáme vhodné údaje, musíme si je tedy vyrobit. Zde se právě bude hodit Vennův diagram, do kterého si zapíšeme, co víme. Je však třeba nějak vyřešit problém, že přímo zakreslit umíme jen ty údaje, které se týkají základních (dále nedělených) oblastí

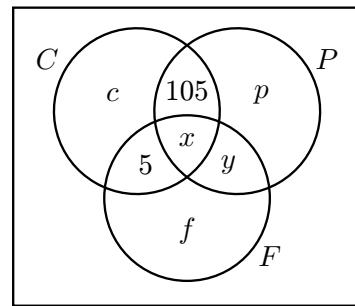
v diagramu. Pak jsou dva možné přístupy. Jeden je mechaničtější, zavedeme si proměnné pro všechny základní oblasti, které v diagramu vidíme a nejsou známy ze zadání.

Pomocí těchto neznámých pak vyjádříme data ze zadání a dostaneme rovnice. V tomto případě jsou to tyto:

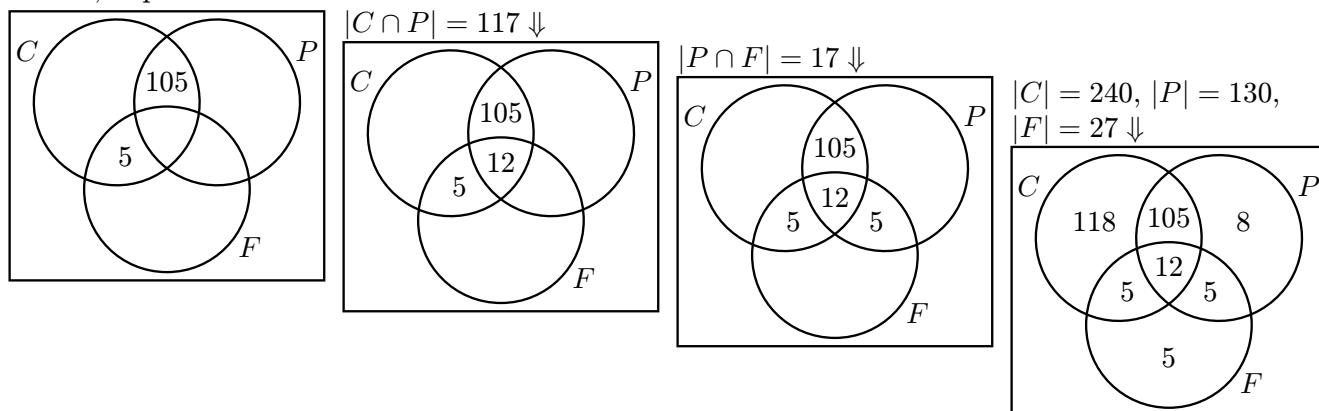
$$\begin{aligned} c + x + 110 &= 240; p + x + y + 105 = 130, \\ f + x + y + 5 &= 27, x + 105 = 117 \text{ a } x + y = 17. \end{aligned}$$

Protože počet neznámých odpovídá počtu rovnic, je úloha řešitelná.

Soustavu teď vyřešíme oblíbenou metodou, například postupná eliminace bude fungovat dobře a dá $x = 12$, $y = 5$, $f = 5$, $p = 8$, $c = 118$. Celkový počet lidí znajících nějaký jazyk se již snadno spočítá z Vennova diagramu: $240 + f + y + p = 240 + 5 + 5 + 8 = 258$. Vzhledem k celkovému počtu lidí to znamená, že 42 z nich neumí C, Pascal ani Fortran.



U méně komplikovaných situací se dá často potřebné údaje dopočítat postupným doplňováním jednotlivých oblastí, například takto:



△

Ted' si představíme další užitečný princip.

! 11b.4. Dirichletův šuplíkový princip (Pigeonhole principle)

Jestliže je alespoň $k+1$ objektů rozděleno do k krabiček, tak musí být krabička obsahující alespoň dva objekty.

Asi jsme tím čtenáře nepřekvapili. Dá se na to dívat i jinak. Přiřazení objektu do krabičky lze považovat za definici zobrazení z množiny objektů do množiny krabiček. Princip tak lze také vyjádřit následovně:

- Nechť A, B jsou konečné množiny. Jestliže $|A| > |B|$, pak pro každé zobrazení $T: A \mapsto B$ existuje $b \in B$ takové, že $|T^{-1}[\{b\}]| > 1$.

Slovny: Jestliže $|A| > |B|$, pak žádné zobrazení nemůže být prosté. To už tu bylo, viz Fakt 2b.12.

Šuplíkový princip se dá zobecnit.

! Fakt 11b.5.

Nechť $c, k \in \mathbb{N}$. Je-li alespoň $ck+1$ objektů umístěno do k krabiček, pak existuje krabička, která má více než c objektů.

Důkaz (rutinní, poučný): Sporem: Kdyby bylo v každé krabičce nejvíše c objektů, pak je celkem jen $c \cdot k$ objektů, což je méně než $ck+1$. □

Opět přepis do jazyka zobrazení:

- Uvažujme konečné množiny A, B . Jestliže existuje $c \in \mathbb{N}$ takové, že $|A| > c|B|$, pak pro libovolné $T: A \mapsto B$ existuje $b \in B$ takové, že $|T^{-1}[\{b\}]| > c$.

A ještě jeden přepis:

Fakt 11b.6.

Je-li N objektů umístěno do k krabiček, pak existuje krabička, která má alespoň $\lceil \frac{N}{k} \rceil$ objektů.

Důkaz: Sporem: Jinak by bylo nejvýše $k(\lceil \frac{N}{k} \rceil - 1)$ objektů. Jenže $\lceil \frac{N}{k} \rceil - 1 < \frac{N}{k}$, viz Fakt 2b.14 (ii), takže bychom dostali méně než $k\frac{N}{k} = N$ objektů.

□

Příklad 11b.i: V libovolné skupině 367 lidí musí mít alespoň dva narozeniny ve stejný den (pozor na přestupní rok, jinak by stačilo 366 lidí), v libovolné skupině 733 lidí musí mít alespoň tři narozeniny ve stejný den.

V libovolné skupině 100 lidí je určitě alespoň $\lceil \frac{100}{12} \rceil = 9$ narozených ve stejný měsíc.

△

Příklad 11b.j: Kolik musí být studentů ve třídě, aby bylo zaručeno, že alespoň 10 dostane stejnou známku?

Je 6 známk A až F, tedy ve třídě stačí $6 \cdot 9 + 1 = 55$ studentů.

△

Příklad 11b.k: Dokážeme, že vybereme-li náhodně $n+1$ různých čísel z množiny $\{1, 2, 3, \dots, 2n\}$, pak se mezi nimi najdou dvě tak, že jedno z nich dělí druhé (viz cvičení).

Napišme všechna ta čísla jako $a_i = 2^{k_i} q_i$, kde q_i je liché. Pak platí $1 \leq q_i \leq 2n$, ale takových lichých čísel je n , zatímco čísel q_i je $n+1$. Proto musí existovat $i \neq j$ takové, že $q_i = q_j = q$. Potom $a_i = 2^{k_i} q$ a $a_j = 2^{k_j} q$, tudíž to s menší mocninou u 2 dělí to druhé.

△

Příklad 11b.l: Dokážeme, že ve skupině o 6 lidech musí být tři lidé, kteří se buď navzájem všichni znají, nebo se žádní dva z nich neznají. Ukážeme také, že ve skupině 5 lidí už tomu tak být nemusí.

Mějme šest lidí. Vezměme si jednoho z nich, nazvěme ho a . Pak je tam 5 dalších lidí, a pokud si uděláme příhrádku nadepsanou „ a se zná“ a příhrádku nadepsanou „ a se nezná“, pak podle šuplíkového principu musí platit, že buď a alespoň tři z lidí zná, nebo alespoň tři z lidí nezná.

V prvním případě vezmeme nějaké tři lidi b, c a d , se kterými se a zná. Pak se buď někteří dva z těch tří také znají a tedy spolu s a tvoří trojici lidí navzájem se znajících. Nebo se mezi nimi nikdo nezná a pak b, c, d tvoří navzájem se neznající trojici.

V druhém případě máme tři lidi b, c a d , se kterými se a nezná, načež zopakujeme zrcadlově předchozí argument. Pro šest lidí je tedy tvrzení dokázáno.

Kdyby bylo pět lidí, tak se mohou navzájem znát $a-b, a-c, b-d, c-e$ a $d-e$ a máme problém. Všimněte si, že každý z pěti se zná s dalšími dvěma lidmi, čili stačí vždy ověřit, že se dotyční dva spolu neznají, a je vyloučen vznik trojice znajících se lidí (například a se zná jen s b a s c , ale ti dva se navzájem neznají). Podobně každý člověk nezná dva lidi a ověří se, že v tomto případě se ti dva naopak znají a nevznikají neznající se trojice.

Jde o jednoduchý případ obecnějšího kombinatorického problému. Zvolme čísla $m, n \in \mathbb{N}$ a označme jako $R(m, n)$ nejmenší nutný počet lidí, aby již bylo zaručeno, že existuje skupina m navzájem se znajících lidí nebo n navzájem se neznajících lidí, tomuto se říká Ramseyho čísla. Ukázali jsme, že $R(3, 3) = 6$. O těchto číslech se dá dokázat spousta zajímavého, třeba že $R(m, n) = R(n, m)$, ale asi nejzajímavější je, že se velice obtížně určuje.

Jednoduchý případ je, když je jedno z čísel dvojka, pak $R(2, n) = R(n, 2) = n$ pro $n \geq 2$. Když totiž máme n lidí, tak se buď najdou dva, kteří se znají (jedna podmínka z definice $R(2, n)$ splněna), nebo se nikdo s nikým nezná a máme n navzájem se neznajících lidí (druhá podmínka splněna).

Jakmile ale vyžadujeme, aby $n, m \geq 3$, tak to začne být obecně neřešitelné a zatím (2009) je známo pouze devět takových Ramseyho čísel. Pro mnohá další jsou známa omezení, třeba že $43 \leq R(5, 5) \leq 49$. Je to tedy jeden z těch opravdu dobrých kombinatorických problémů. Zajímavé je, že toto téma má širší souvislosti, existuje tzv. Ramseyho teorie, která má aplikace i v překvapivě vzdálených oblastech matematiky, třeba ve funkcionální analýze.

△

Tedě se dostáváme k poslední pokročilejší metodě a není to vlastně metoda nová, využijeme znalostí z kapitoly 9.

! Příklad 11b.m: Kolik existuje binárních řetězců délky n takových, že neobsahují žádné po sobě jdoucí nuly?

Stačí chtít, aby se v řetězci nevyskytovalo 00. Jak se to zajistí kombinací? Dost obtížně. Což takhle zkuste opak, kolik je binárních řetězců délky n , které obsahují 00? Vezmeme-li jako množinu R_i řetězce, kde je 00 na pozicích i a $i+1$, pak dozajista $|R_i| = 2^{n-2}$ (volíme z možností 0 a 1 na zbývajících $n-2$ míst) a $S = \bigcup_{i=1}^{n-1} R_i$ jsou všechny řetězce obsahující 00. Množiny R_i ale nejsou disjunktní, takže je třeba použít princip inkluze a exkluze.

Jak velká je množina $|R_i \cap R_j|$? Zde je problém, jsou dvě různé hodnoty, podle toho, jestli $|i - j| = 1$ (pak se ty dvojice nul překrývají a jde vlastně o tři nuly za sebou, čili 2^{n-3} řetězců), nebo $|i - j| > 1$ a jsou to rozdílné dvojice (2^{n-4} řetězců). Máme tedy první komplikaci a to ještě nejsme u průniků tří množin, navíc bychom to pak měli dělat obecně pro k průniků. Jinými slovy, v tomto řešení budeme pokračovat jen v případě, že opravdu nebude jiného zbytí.

Zkusíme tedy najít lepší variantu. Jak vlastně ty řetězce vypadají? Mají docela zajímavou definici rekurzí:

$$(0) \lambda \in M, 0 \in M$$

$$(1) s \in M \implies w1 \in M, w10 \in M.$$

Induktivní podmínka (1) ukazuje, jak se delší řetězce budují z kratších. Jinak řečeno, každý řetězec o délce n bez 00 buď končí na 10, pak vznikl z nějakého řetězce bez 00 délky $n - 2$, nebo končí 1 a vznikl z nějakého řetězce bez 00 délky $n - 1$, přičemž se tyto možnosti vylučují, jinými slovy vzniká disjunktní rozklad a počty se sčítají. Označíme-li jako a_n počet řetězců délky n bez nul za sebou, pak jsme právě ukázali, že $a_n = a_{n-1} + a_{n-2}$.

Přepíšeme si to na rovnici $a_{n+2} - a_{n+1} - a_n = 0$ pro $n \geq 1$, je to lineární rekurentní rovnice s konstantními koeficienty a z kapitoly 10b víme, jak je řešit. Z charakteristické rovnice $\lambda^2 - \lambda - 1 = 0$ dostaneme charakteristická čísla $\lambda = \frac{1 \pm \sqrt{5}}{2}$, takže obecné řešení naší rovnice je $a_n = u\left(\frac{1+\sqrt{5}}{2}\right)^n + v\left(\frac{1-\sqrt{5}}{2}\right)^n$. Teď ještě najdeme počáteční podmínky, potřebujeme dvě:

Jsou dva binární řetězce délky 1, ve kterých se nevyskytují po sobě jdoucí nuly, jmenovitě řetězce 0 a 1. Takže $a_1 = 2$. Podobně máme tři „správné“ řetězce délky dva, jmenovitě 01, 10, 11, takže $a_2 = 3$.

Z těchto počátečních podmínek dostáváme rovnice $u\frac{1+\sqrt{5}}{2} + v\frac{1-\sqrt{5}}{2} = 2$ a $u\left(\frac{1+\sqrt{5}}{2}\right)^2 + v\left(\frac{1-\sqrt{5}}{2}\right)^2 = 3$. Přepis: $u(1 + \sqrt{5}) + v(1 - \sqrt{5}) = 4$, $u(3 + \sqrt{5}) + v(3 - \sqrt{5}) = 6$. Odečteme: $u + v = 1$, odtud $v = 1 - u$, dosadíme do první rovnice a dostaneme $2\sqrt{5}u = 3 + \sqrt{5}$, tedy $u = \frac{3+\sqrt{5}}{2\sqrt{5}}$ a $v = \frac{\sqrt{5}-3}{2\sqrt{5}}$.

Závěr: Počet binárních řetězců délky n , které neobsahují po sobě jdoucí nuly, je dán vzorcem

$$a_n = \frac{3 + \sqrt{5}}{2\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2}\right)^n + \frac{\sqrt{5} - 3}{2\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2}\right)^n.$$

Je to vlastně posunutá Fibonacciho posloupnost, což vyplývá už z toho rekurentního vztahu, a počáteční podmínka ukáže, jak: $a_n = F_{n+2}$. Takže dostáváme čísla 2, 3, 5, 8, 13, 21, 34, ...

Poznámka: Příklad by šlo stejným způsobem řešit zrcadlově, připojováním 1 či 01 zleva.

△

Tento postup někdy bývá velice efektivní.

Příklad 11b.n: Nechť $n \in \mathbb{N}$. Kolik je n -ciferných čísel, které mají sudý počet nul?

Zkusíme to pomocí rekurentního vztahu, budeme čísla vytvářet připojováním číslic zprava. Označme jako a_n počet „správných“ n -ciferných čísel (se sudým počtem nul). Správné n -ciferné číslo může vzniknout dvěma způsoby, které se vylučují, vzniká tedy disjunktní rozklad. Jedna možnost je, že připojíme nenulovou číslici zprava ke správnému $(n-1)$ -cifernému číslu, těchto možností je proto a_{n-1} . Druhá možnost je, že pravá číslice je nula, taková čísla dostaneme připojením nuly zprava k „nesprávnému“ $(n-1)$ -cifernému číslu (lichý počet nul), takových je (přes doplněk) $10^{n-1} - a_{n-1}$. Když tyto možnosti sečteme, dostáváme vztah $a_n = 9a_{n-1} + (10^{n-1} - a_{n-1})$, tedy $a_n = 8a_{n-1} + 10^{n-1}$.

Vznikla lineární rekurentní rovnice $a_{n+1} - 8a_n - 10^n = 0$ s počáteční podmínkou $a_1 = 9$ (všechna jednociferná čísla kromě 0 jsou korektní). Přidružená homogenní rovnice $a_{n+1} - 8a_n = 0$ má charakteristické číslo $\lambda = 8$ a obecné řešení $a_{h,n} = u \cdot 8^n$. Pravá strana 10^n je kvazipolynom s $P(n) = 1$ a $\lambda = 10$, což se neshoduje s charakteristickým číslem levé strany, tedy není třeba korekce a $m = 0$. Proto uhádneme řešení $a_n = A \cdot 10^n$. Dosadíme do dané rovnice:

$$A \cdot 10^{n+1} - 8A \cdot 10^n - 10^n = 0 \implies 10A - 8A = 1 \implies A = \frac{1}{2}.$$

Dostáváme obecné řešení rovnice dané vzorcem $a_n = \frac{1}{2}10^n + u8^n$.

Počáteční podmínka dává $a_1 = 5 + 8u = 9$, proto $u = -\frac{1}{2}$ a řešení je $a_n = \frac{1}{2}10^n + \frac{1}{2}8^n$, což vypadá lépe ve tvaru $a_n = 5 \cdot 10^{n-1} + 4 \cdot 8^{n-1}$ pro $n \geq 1$.

Poznámka: Na rozdíl od příkladu 11b.m zde nemáme symetrii, nelze připojovat zleva. Důvod je ten, že připojováním nul zleva nevzniká nové číslo, pokud pak zase nedáme nenulu, což ale provedeným postupem nelze zaručit.

△

11b.7 Krabičky.

Zajímavým pohledem na kombinatoriku je problém rozdělování objektů do krabiček. Zase existují čtyři verze podle toho, zda jsou krabičky rozlišitelné či identické a zda jsou objekty rozlišitelné či ne, přičemž obvykle bývá

úplně jedno, v jakém pořadí předměty do určité konkrétní krabičky přišly. Je užitečné v tom mít trochu pořádek, protože mnoho různých úloh se dá na rozdělování do krabiček převést. Většinou jde jen o jiný pohled na již popsané principy, uděláme si přehled.

a) Jestliže máme k **různých** krabiček a chceme do nich rozdělit n **stejných** objektů, pak je možné to udělat celkem $\binom{n+k-1}{n}$ způsoby.

Proč? Stačí si vyrobit žetonky s čísly $1, \dots, k$, každý objekt si pak jeden vybere (s opakováním) a tím určíme, do které krabičky přijde.

Viz také cvičení 11b.20.

b) Jestliže máme k **různých** krabiček a chceme do nich rozdělit n **různých** objektů, pak je možné to udělat celkem k^n způsoby.

Proč? Každý objekt si vybere svou krabičku, vybírá se s opakováním.

c) Jestliže máme k **různých** krabiček a chceme mezi ně rozdělit n **různých** objektů tak, aby v i -té krabičce bylo n_i objektů, kde $\sum_{i=1}^k n_i = n$, pak je možné to udělat $\frac{n!}{n_1!n_2!\cdots n_k!}$ různými způsoby.

Proč? Srovnáme objekty do řady, pak si vyrobíme žetonky očíslované $1, \dots, k$ tak, aby bylo n_i žetonů s číslem i . Rozdělení objektů do krabiček odpovídá tomu, že vyrobíme permutaci žetonků podél těch objektů a objekt s žetonkem číslo i jde do krabice i .

Pokud bychom nevyžadovali použití všech předmětů, tedy pokud by platilo $\sum_{i=1}^k n_i < n$, tak to snadno vyřešíme zavedením imaginární krabičky navíc, kterou po naplnění zahodíme. Proto je počet možností, jak objekty rozdělit, roven $\frac{n!}{n_1!n_2!\cdots n_k!(n-\sum n_i)!}$.

d) Jestliže máme k **různých** krabiček a chceme do nich rozdělit $n \geq k$ **různých** objektů tak, aby žádná krabička nebyla prázdná, pak je možné to udělat celkem $\sum_{i=0}^{k-1} (-1)^i \binom{k}{i} (k-i)^n$ způsoby.

Proč? Každé takové rozdelení totiž definuje zobrazení, které je na (což je dáno tím, že na každou krabičku dojde), výsledek tedy plyne z Faktu . Je to situace, která se nedá převést na základní principy, vzoreček si člověk musí pamatovat či někde najít (nebo znova na místě vymyslet).

e) Jestliže máme k **stejných** krabiček a chceme do nich rozdělit $n \geq k$ **různých** objektů tak, aby žádná nezůstala prázdná, pak je to možné udělat celkem $S(n, k) = \frac{1}{k!} \sum_{i=0}^{k-1} (-1)^i \binom{k}{i} (k-i)^n$ způsoby. Ěíká se tomu Stirlingova čísla druhého rádu.

Přijde se na to tak, že se nejprve krabičky očíslují, použije se vzorec z d) a pak se číslíčka z krabiček smažou, jinými slovy se počet možností vydělí počtem možných pořadí krabiček, což je $k!$.

Vlastně to říká, kolik způsobů lze rozdělit n -prvkovou množinu na k neprázdných podmnožin.

f) Jestliže máme k **stejných** krabiček a chceme do nich rozdělit n **různých** objektů, pak je to možné udělat celkem $\sum_{j=1}^K S(n, j) = \sum_{j=1}^K \frac{1}{j!} \sum_{i=0}^{j-1} (-1)^i \binom{j}{i} (j-i)^n$ způsoby, kde $K = \min(k, n)$.

Přijde se na to tak, že při rozdělování může zůstat $0, 1, 2$ až $k-1$ krabiček prázdných, tedy vlastně rozdělujeme do $k, k-1$ až 1 krabiček tak, aby byly neprázdné, což je v případě $n < k$ omezeno počtem objektů.

Všimněte si, že pro poslední čtyři situace nemáme vzorec v uzavřeném tvaru, jde o sumy, což pro větší počet objektů a krabiček znamená někdy dost náročné výpočty. V jednodušších případech (zvlášť pokud člověk nemá po ruce tyto vzorce) bývá jednodušší takové situace řešit stromem nebo výčtem.

Jestliže máme k **stejných** krabiček a chceme do nich rozdělit n **stejných** objektů, pak je to ještě komplikovanější situace než všechny předchozí a nebudeme pro ni hledat nějaký vzorec, řešíme ji vždy individuálně. Jde o jednu z těžších kombinatorických situací, viz také příklad 11b.r a část 11d.8.

Příklad 11b.o: Ve standardním pokeru se čtyřem hráčům rozdá po pěti kartách ze standardního balíčku o 52 kartách.

Jde o různé hráče a různé karty, můžeme si představit zbylý balíček jako pátou krabičku a dostáváme, že počet různých rozdání je $\frac{52!}{5!5!5!5!32!} = 19069457194788$.

△

Příklad 11b.p: Máme rozdat 9 bonbónů stejného druhu čtyřem dětem. Děti jsou asi různé, proto je možné to udělat $\binom{4+9-1}{9} = \binom{12}{9} = 220$ způsoby.

Na jednu ze čtyř základních situací to převedeme tak, že vyrobíme (velký) počet lístečků s čísly 1, 2, 3 a 4 a každý bonbón si jeden lístek vytáhne, tedy vybíráme devětkrát ze čtyř možností, s opakováním a na pořadí záleží, což je ta nejméně intuitivní ze čtyř základních situací a je nejlepší si dotyčný vzorec pamatovat.

Pokud bychom chtěli, ať má každé dítě alespoň jeden bonbón, tak prostě rovnou každému jeden dáme a zbylých 5 rozdělíme běžným způsobem, což je tedy možno $\binom{4+5-1}{5} = \binom{8}{5} = 56$ způsoby.

Co kdyby byly 3 stejné bonbóny na čtyři děti? Žádný problém, vzorec pořád funguje, je $\binom{4+3-1}{3} = \binom{6}{3} = 20$ způsobů. Jak bychom na to přišli výčtem? Nejlépe se to dělá postupně. Začne se situacemi čistě podle počtu bonbónů, bez ohledu na děti, protože je jich méně a lépe vidíme, zda máme všechny, například si je můžeme srovnat podle velikosti. Pak pro každou takovou možnost přiřadíme konkrétní děti.

Možnost 0-1-1-1: Čtyři možnosti, vybereme, které dítě nedostane bonbón.

Možnost 0-0-1-2: 12 možností, vybereme, které dítě dostane dva bonbóny a které jeden.

Možnost 0-0-0-3: Čtyři možnosti, vybereme, které dítě dostane tři bonbóny.

Celkem je to 20, přesně jako ze vzorce.

△

Příklad 11b.q: Máme 4 různé bonbóny. Kolika způsoby je možno vyrobit tři dárkové balíčky, když jsou krabičky stejné?

Je to problém, který není převoditelný na jednu ze čtyř základních situací, takže to bud' řešíme výčtem, nebo někde najdeme příslušné vzorce. To zkusíme nejdřív.

a) Pokud budeme chtít, aby žádná krabička nebyla prázdná, pak je odpověď dána jako

$$S(4, 3) = \frac{1}{3!} \sum_{i=0}^2 (-1)^i \binom{3}{i} (3-i)^4 = \frac{1}{6} (3^4 - 3 \cdot 2^4 + 3 \cdot 1^4) = 6.$$

b) Co kdyby nám prázdné krabičky nevadily? Pak ještě musíme uvažovat rozdělení do dvou a jedné krabičky. Evidentně $S(4, 1) = 1$, dále

$$S(4, 2) = \frac{1}{2!} \sum_{i=0}^1 (-1)^i \binom{2}{i} (2-i)^4 = \frac{1}{2} (2^4 - 2 \cdot 1^4) = 7.$$

Celkem je to tedy $6 + 7 + 1 = 14$ možností.

Zkusme to bez vzorců. Protože jde o krabičky identické, můžeme je vždy srovnat podle množství bonbónů, řekněme od největšího, takže dostaváme (připomínáme, že na pořadí v krabičce nezáleží)

$\{\{A, B, C, D\}, \{\}, \{\}\};$
 $\{\{A, B, C\}, \{D\}, \{\}\}; \{\{A, B, D\}, \{C\}, \{\}\}; \{\{A, C, D\}, \{B\}, \{\}\}; \{\{B, C, D\}, \{A\}, \{\}\};$
 $\{\{A, B\}, \{C, D\}, \{\}\}; \{\{A, C\}, \{B, D\}, \{\}\}; \{\{A, D\}, \{B, C\}, \{\}\};$
 $\{\{A, B\}, \{C\}, \{D\}\}; \{\{A, C\}, \{B\}, \{D\}\}; \{\{A, D\}, \{B\}, \{C\}\}; \{\{B, C\}, \{A\}, \{D\}\}; \{\{B, D\}, \{A\}, \{C\}\};$
 $\{\{C, D\}, \{A\}, \{B\}\}.$

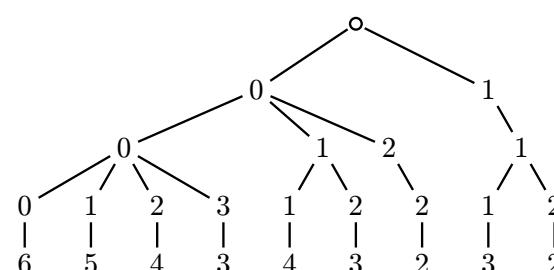
Opravdu celkem 14 možností.

△

Příklad 11b.r: Profesor má v kanclu hromádku 6 kopí domácích úkolů a čtyři studenty. Když studenti odchází, prof je poprosí, aby s sebou ty úkoly vzali a donesli je do třídy k nakopírování ostatním. Z pohledu profesora jsou studenti identičtí :). Student, který odchází poslední, se většinou cítí blbě, aby něco na stole nechal, takže předpokládejme, že opravdu je těch 6 kopí odneseno. Kolika různými způsoby se to může stát?

Je to přesně ten nejhorší problém, nerozlišitelné krabičky i objekty, jde jen o to, kolik kopií který student vezme, přičemž je můžeme srovnávat podle toho, kolik nesou, třeba od nejméně po nejvíce. Nezbývá, než to zkusit manuálně výčtem možností.

Celkem je tedy 9 možností.



Tento těžký kombinatorický úkol je ekvivalentní ještě jiným, neméně zajímavým úlohám. Nechť je dáno $k, n \in \mathbb{N}$.

- Kolik řešení má rovnice $x_1 + x_2 + \dots + x_k = n$ takových, aby $0 \leq x_1 \leq x_2 \leq \dots \leq x_k$?

- Kolik řešení má rovnice $x_1 + x_2 + \dots + x_m = n$ takových, aby $0 < x_1 \leq x_2 \leq \dots \leq x_k$ a $m \leq k$?

O to více zamrzí, že pro to nemáme rozumný vzorec, ještě že jsou počítáče, viz část 11d.8.

△

Cvičení

Cvičení 11b.1 (rutinní): Napište vzorec vyplývající z principu inkluze a exkluze pro sjednocení následujících množin. U všech výrazů ze vzorce určete interpretaci.

- a) Nechť A je množina studentů, kteří absolvovali kurs analýzy, a B množina studentů, kteří absolvovali kurs lineární algebry.
- b) Nechť A je množina nákladních aut s dvěma koly a B množina nákladních aut s šesti koly.
- c) Nechť A je množina knížek psaných česky, B množina knížek psaných anglicky a C množina knížek psaných německy.

Cvičení 11b.2 (rutinní): Máme čtyři množiny obsahující po řadě 50, 60, 70 a 80 prvků. Každá dvojice z nich má 5 společných prvků, každá trojice má jeden společný prvek a všechny čtyři nemají žádný společný prvek. Kolik prvků má sjednocení těchto čtyř množin?

Cvičení 11b.3 (rutinní): Sto vstupenek očíslovaných 1 až 100 jde do tomboly, losují se čtyři výhry. Kolika způsoby to může dopadout, jestliže

- a) lístek 13 má vyhrát první cenu?
- b) lístek 13 má vyhrát cenu?
- c) lístek 13 nemá vyhrát cenu?
- d) lístky 13 a 23 mají vyhrát cenu?
- e) vyhrát první cenu má buď lístek 13 nebo lístek 23?
- f) lístek 13 nebo lístek 23 mají vyhrát cenu?
- g) lístky 13, 23, 31, 33 mají vyhrát cenu?
- h) první cenu má vyhrát jeden z lístků 13, 23, 31, 33?
- i) lístky 13 a 23 mají vyhrát, ale lístky 7 a 2 vyhrát nemají?

Cvičení 11b.4 (rutinní): Kolik osmibitových řetězců neobsahuje šest nul po sobě?

Cvičení 11b.5 (rutinní): Kolik permutací standardních 10 číslic začíná 987 nebo končí 123 nebo má 45 na páté a šesté pozici?

Cvičení 11b.6 (rutinní): Kolik permutací řetězce $ABCDEFGH$ obsahuje slova BA nebo FEC ?

Cvičení 11b.7 (rutinní): Nechť $A = \{1, 2, \dots, n\}$ a B je množina o k prvcích, kde $k, n \geq 2$. Nechť $a \neq b$ jsou jisté prvky z B .

- a) Kolik je zobrazení $T: A \rightarrow B$, pro které platí $T(1) = a$ a $T(2) = b$?
- b) Kolik je zobrazení $T: A \rightarrow B$, pro které platí $T(1) = a$ nebo $T(2) = b$?

Cvičení 11b.8 (rutinní): Uvažujme čtyřpísmenná slova (z 26 malých písmen abecedy).

- a) Kolik jich obsahuje znak x ?
- b) Kolik jich obsahuje přesně tři x ?
- c) Kolik jich obsahuje alespoň tři x ?

Cvičení 11b.9 (poučné): Kolik je přirozených čísel menších než 1000, které

- a) jsou dělitelné 7?
- b) jsou dělitelné 7 a 11?
- c) jsou dělitelné 7 ale ne 11?
- d) jsou dělitelné 7 nebo 11?
- e) nejsou dělitelné ani 7 ani 11?
- f) neobsahují stejné číslice?
- g) neobsahují stejné číslice a jsou sudé?

Cvičení 11b.10 (poučné): Uvažujme přirozená čísla mezi mezi 23 a 131313 včetně.

- a) Kolik jich je dělitelných 5?
- b) Kolik jich je dělitelných 7?
- c) Kolik jich je dělitelných 5 nebo 7?
- d) Kolik jich není dělitelných 7?
- e) Kolik jich má stejné číslice?
- f) Kolik jich je lichých?
- g) Kolik jich je dělitelných 4 nebo 6?

Cvičení 11b.11: Heslo (password) se skládá ze čtyř znaků, které mohou být písmena (lower-case) či číslice, přičemž se v heslu musí vyskytovat alespoň dvě číslice. Kolik je možné vytvořit takových passwordů?

Cvičení 11b.12: Svědek dopravní nehody si pamatuje, že auto, které z nehody ujelo, mělo SPZ ze tří písmen a čtyř číslic, začínala AS (Pražák!) a určitě v ní byla jednička a dvojka. Kolik je takových SPZ možných?

Cvičení 11b.13 (poučné): Uvažujme řešení $x_i \in \mathbb{N}_0$ rovnice $x_1 + x_2 + x_3 = 13$.

- a) Kolik z nich splňuje podmínu $x_1 \leq 6, x_2 \leq 6, x_3 \leq 6$?
- b) Kolik z nich splňuje podmínu $x_1 < 6, x_2 < 6, x_3 < 6$?

Cvičení 11b.14 (poučné): Kolik řešení $x_i \in \mathbb{N}_0$ rovnice $x_1 + x_2 + x_3 + x_4 = 23$ splňuje podmínky $1 < x_1 \leq 13, 7 \leq x_2, 0 \leq x_3, 6 \leq x_4 < 9$?

Cvičení 11b.15 (poučné): Kolik řešení $x_i \in \mathbb{N}_0$ rovnice $x_1 + x_2 + x_3 = 23$ splňuje podmínky

- a) $1 < x_1 \leq 9, 6 \leq x_2 \leq 10, 0 \leq x_3 < 7$?
- b) $1 \leq x_1 < 6, 6 \leq x_2 \leq 10, 0 \leq x_3 \leq 6$?
- c) $3 \leq x_1 < 7, 6 < x_2 \leq 10, 2 < x_3 \leq 9$?

Cvičení 11b.16 (poučné): Kolik celých čísel menších než 1000 má ciferný součet 13?

Cvičení 11b.17: Kolika způsoby se dá rozdělit šest různých hraček mezi tři (různé) děti tak, aby každé mělo alespoň jednu?

Cvičení 11b.18 (dobré): Kolika způsoby je možno rozdělit sedm úkolů mezi čtyři zaměstnance tak, aby každý měl alespoň jeden úkol a nejtěžší úkol byl přidělen nejlepšímu zaměstnanci?

Cvičení 11b.19 (rutinní, poučné): Kolika způsoby je možno roztrídit šest věcí do pěti krabic, jestliže

- a) věci i krabice jsou různé?
- b) věci jsou různé, krabice stejné?
- c) věci jsou stejné, krabice různé?
- d) věci i krabice jsou stejné?

Cvičení 11b.20 (poučné): Kolika způsoby je možno umístit n identických věcí do m různých krabic za předpokladu, že žádná není prázdná?

Cvičení 11b.21 (rutinní): Kolik je třeba vybrat náhodně karet ze standardního balíčku o 52 kartách, aby bylo zaručeno, že

- a) se najdou alespoň tři stejné barvy?
- b) se najdou alespoň tři srdce?

Cvičení 11b.22 (rutinní): V šuplíku je 10 černých a 10 šedých ponožek. Ráno ještě v polospánku taháme naslepo ponožky.

- a) Kolik jich musíme vzít, aby bylo jisté, že se mezi nimi najde pár?
- b) Kolik jich musíme vzít, aby se mezi nimi našel černý pár?

Cvičení 11b.23 (rutinní): Ve hře Magic je pět základních barev. Kolik náhodně namíchaných karet typu land je třeba mít, aby měl člověk jistotu, že od alespoň jedné barvy bude mít 10 karet?

Cvičení 11b.24 (rutinní): V lepších čínských restauracích dávají po obědě hostům „štěstíčko“ („fortune cookie“, kousek těsta se zapečenou věštboou). Jestliže jich mají 50 druhů, jaký je nejvyšší možný počet návštěv, kdy člověk nedostane stejně štěstíčko více než třikrát?

Cvičení 11b.25 (rutinní, poučné): Dokažte, že se mezi libovolnými $d+1$ čísly najdou dvě, která mají při dělení d stejný zbytek.

Cvičení 11b.26 (dobré): Dokažte, že vybereme-li z množiny $\{1, 2, \dots, 50\}$ celkem 10 různých čísel, pak z nich lze dálé vybrat dvěma různými způsoby pět čísel tak, aby měly stejný součet.

Cvičení 11b.27 (dobré): Ukažte, že v libovolné uspořádané n -tici různých přirozených čísel se najdou alespoň dvě po sobě následující, jejichž součet je dělitelný n .

Cvičení 11b.28 (poučné): Dokažte, že když se vezme libovolných pět bodů v rovině, které mají celočíselné souřadnice, tak se mezi nimi dá vybrat dvojice tak, aby měl střed jejich spojnice také celočíselné souřadnice. Ukažte totéž pro 9 bodů v prostoru \mathbb{R}^3 .

Cvičení 11b.29 (poučné): Dokažte, že když se vybere 5 různých čísel z množiny $\{1, 2, 3, \dots, 8\}$, pak se mezi nimi najde dvojice se součtem 9.

Cvičení 11b.30 (poučné): Dokažte, že když se vybere 7 různých čísel z množiny $\{1, 2, 3, \dots, 10\}$, pak se mezi nimi najdou dvě dvojice se součtem 11.

Cvičení 11b.31 (rutinní): Dokažte, že jestliže je ve třídě 9 studentů, tak mezi nimi musí být alespoň 5 mužů nebo alespoň 5 žen.

Cvičení 11b.32 (rutinní): Dokažte, že jestliže si loni alespoň 3000000 lidí vydělalo méně než 30000 měsíčně, tak se mezi nimi museli najít alespoň dva, kteří si vydělali na haliště stejně.

Cvičení 11b.33 (dobré, poučné): Ve třídě se sešlo 17 lidí narozených ve znamení Štíra. Dokažte, že se musí najít dva, kteří mají narozeniny ve stejný den nebo hned den po sobě.

Cvičení 11b.34 (poučné): Ukažte, že když vezmete libovolných pět bodů ze čtverce o straně 2, tak se tam najde dvojice, která je od sebe nejvýše $\sqrt{2}$ daleko.

Cvičení 11b.35 (poučné): Kolika způsoby je možno vyrobit z n dlaždic chodník (o šířce jedné dlaždice, tj. dáváme je za sebe), pokud máme na výběr dlaždice tří barev a nechceme, aby někdy šly hned za sebou stejné dlaždice?

Cvičení 11b.36 (poučné): Bankomat akceptuje koruny a dvoukoruny. Kolika způsoby je mu možné zaplatit n korun, pokud na pořadí záleží?

Cvičení 11b.37 (poučné): Označme a_n počet binárních řetězců délky n , které obsahují dvě po sobě jdoucí nuly. Najděte pro tuto posloupnost rekurentní vztah a počáteční podmínky, pak určete a_5 .

Cvičení 11b.38 (poučné): Kolika způsoby je možno vyjít schodiště o n schodech, jestliže se dá jít po jednom ale také po dvou schodech? Kolik to dělá pro tradičních osm schodů?

Cvičení 11b.39 (dobré, poučné): Na kolik nejvíce oblastí lze rozdělit rovinu pomocí n přímek?

Návod: Označte toto číslo R_n a najděte rekurentní vztah.

Cvičení 11b.40 (dobré, poučné): Kolik je $S(n)$, maximální počet částí v 3D-prostoru, na který je možno prostor rozdělit pomocí n rovin?

Cvičení 11b.41 (poučné): Kolik je řetězů ze znaků 0, 1, 2 délky n , které neobsahují 00?

Cvičení 11b.42 (poučné): Kolik derrangements množiny $\{1, 2, 3, 4, 5, 6\}$ začíná

- a) 456?
- b) 345?
- c) 234?
- d) 321?
- e) 312?

Cvičení 11b.43 (dobré, poučné): Kolika způsoby je možno přerovnat číslo 1234567890 tak, aby číslice na sudých pozicích nezůstaly na svých místech?

Cvičení 11b.44 (dobré, poučné): Dokažte kombinatoricky, že počet derrangements splňuje

$$a) D_n = (n-1)(D_{n-1} + D_{n-2}) \text{ pro } n \geq 2;$$

$$b) D_n = nD_{n-1} + (-1)^n \text{ pro } n \geq 1.$$

Poznámka: V b) získáváme pro počet derrangements rekurentní rovnici $D_{n-1} - (n+1)D_n = (-1)^{n+1}$, kterou lze vyřešit pomocí postupu v poslední poznámce kapitoly 10b: Máme $f(n) = 1$, $g(n) = -(n+1)$, $h(n) = (-1)^{n+1}$, proto $Q(n) = \frac{1}{(n+1)!}$, po substituci vyjde rovnice $b_{n+1} - b_n = \frac{(-1)^{n+1}}{(n+1)!}$, také $b_1 = 0$, odtud $b_n = \sum_{k=0}^n \frac{(-1)^k}{k!}$ a

$$D_n = \frac{1}{n!} \sum_{k=0}^n \frac{(-1)^k}{k!}, \text{ přesně jako jsme odvodili v příkladu 11b.f.}$$

Řešení:

11b.1: a) $|A \cup B| = |A| + |B| - |A \cap B|$; $A \cap B$ jsou studenti, kteří absolvovali oba kurzy.

b) $|A \cup B| = |A| + |B| - |A \cap B|$; $A \cap B$ jsou nákladní auta s čtyřmi a šesti koly, prázdná množina.

c) $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$; $A \cap B$ jsou knihy psané česky a anglicky (dtřeba slovníky), $A \cap C$ jsou knihy psané česky a německy, $B \cap C$ jsou knihy psané anglicky a německy, $A \cap B \cap C$ jsou knihy psané česky, anglicky a německy.

11b.2: $50 + 60 + 70 + 80 - \binom{4}{2}5 + \binom{4}{3}1 - 0 = 260 - 30 + 4 = 234$.

11b.3: a) Jde tedy jen o ostatní tři ceny, $99 \cdot 98 \cdot 97 = 941094$. b) Nejprve čtyři možnosti ceny pro lístek 13, pak výběr na ostatní tři ceny, $4 \cdot 99 \cdot 98 \cdot 97 = 3764376$. c) Vybíráme čtyři ceny z 99 lístků, $99 \cdot 98 \cdot 97 \cdot 96 =$

90345024. d) Vybíráme cenu pro lístek 13, ze zbývajících tří cen pro lístek 23, na další dvě ceny ze zbývajících 98, $4 \cdot 3 \cdot 98 \cdot 97 = 114072$. e) Pokud vyhraje první cenu 13, dobereme lístky na ostatní: $99 \cdot 98 \cdot 97$. Pokud vyhraje první cenu 23, dobereme ostatní. Tyto možnosti jsou disjunktní, proto stačí sečít, $2 \cdot 99 \cdot 98 \cdot 97 = 1882188$. f) Vybereme cenu pro lístek 13 (4 možnosti), pak dobereme ostatní ceny: $99 \cdot 98 \cdot 97$. Ditto pro 23. Jenže tyto možnosti nejsou disjunktní, pokud mají cenu 13 i 23, započítali jsme je dvakrát, nutno použít princip inkluze a exkluze. Možnosti kdy mají 13 i 23 cenu: Nejprve pro ně vybereme ceny, $4 \cdot 3$, na další dobereme ze zbytku lístků. Celkově máme $2 \cdot 99 \cdot 98 \cdot 97 - 4 \cdot 3 \cdot 98 \cdot 97 = 1768116$. g) Stačí je srovnat u cen, $4! = 24$. h) Viz e), $4 \cdot 99 \cdot 98 \cdot 97 = 3764376$. i) Viz d), $4 \cdot 3 \cdot 96 \cdot 95 = 109440$.

11b.4: M_i řetězce s šesti nulami počínaje místem i , $|M_1| = |M_2| = |M_3| = 2^2 = 4$, průniky: $|M_1 \cap M_2| = |M_2 \cap M_3| = 2$ (řetězce se sedmi nulami), $|M_1 \cap M_3| = |M_1 \cap M_2 \cap M_3| = 1$ (osm nul), tedy celkem $3 \cdot 4 - 2 - 2 - 1 + 1 = 8$, proto výsledek je $2^8 - 8 = 248$.

11b.5: Množiny M_1 (permutace začínající 987), M_2 (permutace končící 123), M_3 (permutace s 45 na pozicích 5, 6). Pak $|M_1| = |M_2| = 7!$, $|M_3| = 8!$, $|M_1 \cap M_2| = 4!$, $|M_1 \cap M_3| = |M_2 \cap M_3| = 5!$, $|M_1 \cap M_2 \cap M_3| = 2!$. Odpověď: $7! + 7! + 8! - 4! - 5! + 2! = 50138$.

11b.6: 7! permutací má BA; 6! permutací má FEC; obě najednou má 5! permutací (dva celky a tři zbývající písmena). Sjednocení má velikost $7! + 6! - 5! = 5640$.

11b.7: a) Volíme jen $T(3), \dots, T(n)$, takže k^{n-2} . b) $M_1 = \{T: A \mapsto B; T(1) = a\}$ a $M_2 = \{T: A \mapsto B; T(2) = b\}$. Zajímá nás $|T_1 \cup T_2|$. Máme $|T_1 \cup T_2| = |T_1| + |T_2| - |T_1 \cap T_2| = 2 \cdot k^{n-1} - k^{n-2}$.

11b.8: a) Doplňkem, bez x je jich 25^4 , tedy odpověď $26^4 - 25^4 = 66351$.

b) Vybereme místo pro jiný znak (4 možnosti) a tam vybereme, $4 \cdot 25 = 100$ slov.

c) K b) přidáme slovo ze čtyř x , celkem 101.

11b.9: a) $|M_7| = \lfloor \frac{999}{7} \rfloor = 142$. b) $|M_7 \cap M_{11}| = \lfloor \frac{999}{7 \cdot 11} \rfloor = 12$. c) $|M_7| - |M_7 \cap M_{11}| = 142 - 12 = 130$.

d) $|M_7| + |M_{11}| - |M_7 \cap M_{11}| = 142 + 90 - 12 = 220$. e) $999 - 220 = 779$. f) Rozdělit podle počtu cifer (jedna až tři), první cifra nesmí být 0, proto s ní musíme začít, na dalších pozicích už nula být může: $9 + 9 \cdot 9 + 9 \cdot 9 \cdot 8 = 738$.

g) Zase podle počtu cifer. Nejprve vybíráme sudé číslo na poslední cifru, tím se ale změní možnosti na první cifru, podle toho, zda jako poslední dáme nulu (pak to první cifre nevadí) nebo ne (pak jsme první cifre vzali jednu možnost), čili další dělení na případu. Čísla končící na nulu: $0 + 9 + 9 \cdot 8 = 81$. Sudá čísla nekončící na nulu omezí výběr první cifry, $4 + 4 \cdot 8 + 4 \cdot 8 \cdot 8 = 4 + 32 + 256 = 292$. Celkem: 373.

11b.10: a) $|M_5| = \lfloor \frac{131313}{5} \rfloor - \lfloor \frac{22}{5} \rfloor = 26262 - 4 = 26258$. b) $|M_7| = \lfloor \frac{131313}{7} \rfloor - \lfloor \frac{22}{7} \rfloor = 18759 - 3 = 18756$.

c) $|M_5 \cap M_7| = \lfloor \frac{131313}{35} \rfloor - \lfloor \frac{22}{35} \rfloor = 3751 - 0 = 3751$, proto $|M_5 \cup M_7| = 26258 + 18756 - 3751 = 41263$.

d) $(131313 - 23 + 1) - 18756 = 112534$. e) Vybereme nenulovou číslici a rozhodneme se, kolik jich může být. Jedniček může být tři až šest (čísla 111 až 111111) neboli 4 čísla, dvojek může být tři až pět neboli 3 čísla, trojek až devítek může být dvě až pět neboli čtyři čísla. Celkem $4 + 3 + 7 \cdot 4 = 35$.

f) Čísel je $131313 - 23 + 1 = 131291$, z toho dělitelných dvěma je $\lfloor \frac{131313}{2} \rfloor - \lfloor \frac{22}{2} \rfloor = 65656 - 11 = 65645$. Lichých je 65646. Pozor, počet lichých čísel se nedá určit jen z rozdílu, třeba mezi čísky 6 a 8 je jedno liché, ale mezi 7 a 9 jsou dvě lichá, ačkoliv je vzdálenost vždy stejná, $8 - 6 = 9 - 7 = 2$.

g) $|M_4| = \lfloor \frac{131313}{4} \rfloor - \lfloor \frac{22}{4} \rfloor = 32828 - 5 = 32823$; $|M_6| = \lfloor \frac{131313}{6} \rfloor - \lfloor \frac{22}{6} \rfloor = 21885 - 3 = 21882$;

Pozor, $|M_4 \cap M_6| = \lfloor \frac{131313}{12} \rfloor - \lfloor \frac{22}{12} \rfloor = 10942 - 1 = 10941$, proto $|M_4 \cup M_6| = 32823 + 21882 - 10941 = 43764$.

11b.11: Rozborem možností, c je číslo, p je písmeno (tedy nečíslo), pak jsou možnosti $ccpp, cpcp, cppc, pccp, pcpc, ppcc, cccp, ccpc, cpcc, pccc, cccc$, tudíž počet hesel je $6 \cdot 10^2 \cdot 26^2 + 4 \cdot 10^3 \cdot 26 + 10^4 = 519600$.

Nebo stručněji: Pro přesně dvě čísla vybereme pozice a pak znaky: $\binom{4}{2} 10^2 26^2$. Pro přesně tři čísla vybereme pozice a pak znaky: $\binom{4}{3} 10^3 26$. Přesně čtyři čísla: 10^4 .

11b.12: Přímý útok: M_{ij} budíž množina značek, kde je 1 na i -tém místě a 2 na j -tém místě či naopak. Chceme velikost sjednocení, to je princip inkluze a exkluze se šesti výchozími množinami, nic pěkného. Zkusme doplněk. Celkem SPZ začínajících AS: $|M| = 26 \cdot 10^4 = 260000$. M_1 jsou SPZ bez jedničky, $|M_1| = 26 \cdot 9^4$, podobně pro M_2 neboli SPZ bez dvojky. Průnik $|M_1 \cap M_2| = 26 \cdot 8^4$. Pomocí principu doplňku a inkluze-exkluze dostaneme $26 \cdot 10^4 - [2 \cdot 26 \cdot 9^4 - 26 \cdot 8^4] = 25324$.

11b.13: Všech řešení M je $\binom{3+13-1}{13} = 105$.

a) M_i jsou řešení s $x_i \geq 7$, $|M_i| = \binom{3+6-1}{6} = 28$. Jsou navzájem disjunktní, neboť v $M_i \cap M_j$ jsou řešení s $x_i, x_j \geq 7$, zároveň $x_i + x_j \leq 13$, nelze. Proto $|M_1 \cup M_2 \cup M_3| = 3 \cdot 28 = 84$. Hledaných řešení je $105 - 84 = 21$.

b) M_i jsou řešení s $x_i \geq 6$, $|M_i| = \binom{3+7-1}{7} = 36$. $M_i \cap M_j$ jsou řešení s $x_i, x_j \geq 6$, těch je $\binom{3+1-1}{1} = 3$. $M_1 \cap M_2 \cap M_3 = \emptyset$. Hledaných řešení je $105 - 3 \cdot 36 + 3 \cdot 3 = 6$.

11b.14: Množina M řešení splňujících dolní odhad: Dáme 2 do x_1 , 7 do x_2 , 6 do x_4 , zbytek rozdělíme, $\binom{4+8-1}{8} = 165$.

Odebereme řešení nevyhovující horním mezím. M_1 řešení s $x_1 \geq 14$ a ostatními dolními odhady: rozdělíme $14 + 7 + 0 + 6 = 27$, nejde, $M_1 = \emptyset$. M_2 řešení s $x_4 \geq 10$ a ostatními dolními odhady: rozdělíme $1 + 7 + 10 = 18$ napevno, zbytek: $\binom{4+5-1}{5} = 56$. Také $M_1 \cap M_2 = \emptyset$, počet řešení je tedy $165 - 56 = 109$.

11b.15: a) Množina M řešení splňujících dolní odhad: Dáme 2 do x_1 , 6 do x_2 , zbytek rozdělíme, $\binom{3+15-1}{15} = 136$. Odebereme řešení nevyhovující horním mezím. M_1 řešení s $x_1 \geq 10$ a ostatními dolními odhady: rozdělíme $10 + 6 + 0 = 16$ napevno, další $\binom{3+7-1}{7} = 36$ možností. M_2 řešení s $x_2 \geq 11$ a ostatními dolními odhady: rozdělíme $2 + 11 + 0 = 13$ napevno, další $\binom{3+10-1}{10} = 66$ možností. M_3 řešení s $x_3 \geq 7$ a ostatními dolními odhady: rozdělíme $2 + 6 + 7 = 15$ napevno, další $\binom{3+8-1}{8} = 45$ možností.

Průniky: $M_1 \cap M_2$: rozdělíme $10 + 11 + 0 = 21$ napevno, další $\binom{3+2-1}{2} = 6$ možností. $M_1 \cap M_3$: rozdělíme $10 + 6 + 8 = 24$ nejde. $M_2 \cap M_3$: rozdělíme $2 + 11 + 7 = 19$ napevno, další $\binom{3+4-1}{4} = 15$ možností. $M_1 \cap M_2 \cap M_3 = \emptyset$. Špatných řešení je $36 + 66 + 45 - 6 - 0 - 15 + 0 = 126$. Počet řešení je tedy $136 - 126 = 10$.

b) Žádné řešení: I když dáme maximální povolené počty, dostaneme $x_1 + x_2 + x_3 = 5 + 10 + 6 = 21$.

c) $3 \leq x_1 < 7$, $6 < x_2 \leq 10$, $2 < x_3 \leq 9$?

c) Množina M řešení splňující dolní odhad: Dáme 3 do x_1 , 7 do x_2 , 3 do x_3 , zbytek rozdělíme, $\binom{3+10-1}{10} = 66$.

Odebereme řešení nevyhovující horním mezím: $7-7-3 \Rightarrow \binom{3+6-1}{6} = 28$, $3-11-3 \Rightarrow \binom{3+6-1}{6} = 28$,

$3-7-10 \Rightarrow \binom{3+3-1}{3} = 10$, $7-11-3 \Rightarrow \binom{3+2-1}{2} = 6$, $7-7-10 \Rightarrow \emptyset$, $3-11-10 \Rightarrow \emptyset$.

Špatná řešení: $28 + 28 + 10 - 6 - 0 - 0 + 0 = 60$. Počet řešení je tedy $66 - 60 = 6$.

To by možná šlo rychleji výpisem. Protože $x_2 + x_3 \leq 19$, musí být x_1 alespoň 4. Možnosti: (4, 10, 9), (5, 9, 9), (5, 10, 8), (6, 8, 9), (6, 9, 8), (6, 10, 7).

11b.16: Jde o jedno až tříciferná čísla, takže hledáme počet řešení rovnic $a_1 = 13$, $a_1 + a_2 = 13$ a $a_1 + a_2 + a_3 = 13$, kde $a_i \in \mathbb{N}_0$, $a_i < 10$ a $a_1 \geq 1$. První rovnice řešení nemá, druhá: 6 možností, třetí 85 možností.

11b.17: Jde o problém ekvivalentní počtu surjektivních zobrazení z 6-prvkové množiny na tříprvkovou, podle vzorce je to $\sum_{i=0}^2 (-1)^i \binom{3}{i} (3-i)^6 = \binom{3}{0} 3^6 - \binom{3}{1} 2^6 + \binom{3}{2} 1^6 = 3^6 - 3 \cdot 2^6 + 3 = 540$.

11b.18: Nejtěžší přidělíme hned, zbývá rozdělit šest úkolů čtyřem lidem tak, aby tři (nikoliv nejlepší) zaměstnanci měli alespoň jeden. Rozložit na dvě disjunktní množiny podle toho, zda nejlepší ještě něco dostane či ne. Pokud už ne, tak stačí rozdělit všechn 6 úkolů mezi tři zaměstnance, aby měli každý alespoň jeden, to je celkem

$$\sum_{i=0}^2 (-1)^i \binom{3}{i} (3-i)^6 = \binom{3}{0} 3^6 - \binom{3}{1} 2^6 + \binom{3}{2} 1^6 = 3^6 - 3 \cdot 2^6 + 3 = 540 \text{ možností.}$$

Nebo i ten nejlepší alespoň jeden navíc dostane, rozdělujeme 6 mezi čtyři tak, aby měl každý alespoň jeden, to je $\sum_{i=0}^3 (-1)^i \binom{4}{i} (4-i)^6 = \binom{4}{0} 4^6 - \binom{4}{1} 3^6 + \binom{4}{2} 2^6 - \binom{4}{3} 1^6 = 4^6 - 4 \cdot 3^6 + 6 \cdot 2^6 - 4 = 1560$ možností. Celkem 2100.

11b.19: a) Každá věc vybírá číslo krabice, záleží na pořadí (věci jsou různé), s opakováním, $5^6 = 15625$.

$$\text{b) Toto je ten těžší případ, } \sum_{j=1}^5 S(6, j) = \sum_{j=1}^5 \frac{1}{j!} \sum_{i=0}^{j-1} (-1)^i \binom{j}{i} (j-i)^6 = 1^6 + \frac{1}{2!} [2^6 - 2 \cdot 1^6] + \frac{1}{3!} [3^6 - 3 \cdot 2^6 + 3 \cdot 1^6] + \frac{1}{4!} [4^6 - 4 \cdot 3^6 + 6 \cdot 2^6 - 4 \cdot 1^6] + \frac{1}{5!} [5^6 - 5 \cdot 4^6 + 10 \cdot 3^6 - 10 \cdot 2^6 + 5 \cdot 1^6] = 1 + 31 + 90 + 65 + 15 = 202.$$

c) Každá věc vybírá číslo krabice, nezáleží na pořadí (věci jsou stejné), s opakováním, $\binom{5+6-1}{6} = 210$.

d) Toto jedině výčtem či jiným podobným způsobem, třeba tak, že srovnáme krabice podle počtu prvků od nejmenšího, možnosti jsou 0-0-0-0-6, 0-0-0-1-5, 0-0-0-2-4, 0-0-0-3-3, 0-0-1-1-4, 0-0-1-2-3, 0-0-2-2-2, 0-1-1-1-3, 0-1-1-2-2, 1-1-1-1-2. Celkem 10.

11b.20: 0 pro $n < m$, jinak nejprve natvrdo m věcí po jedné do krabic a pro zbývajících $n-m$ věcí losujeme čísla krabic s opakováním, pořadí nezáleží, tedy $\binom{m+(n-m)-1}{n-m} = \binom{n-1}{n-m}$.

11b.21: a) $2 \cdot 4 + 1 = 9$; b) $13 \cdot 3 + 3 = 42$.

11b.22: a) $2 \cdot 1 + 1 = 3$; b) $10 + 2 = 12$.

11b.23: $5 \cdot 9 + 1 = 46$.

11b.24: $3 \cdot 50 = 150$.

11b.25: Rozdělíme je do krabiček podle zbytků, možných zbytků je d , tudíž dle šuplíkového principu musí nějaká krabička obsahovat alespoň dvě čísla.

11b.26: Max. možná suma je $50 + 49 + 48 + 47 + 46 = 240$, minimální je $1 + 2 + 3 + 4 + 5 = 15$. Je tedy možné vytvořit 226 možných sum, ale existuje 252 pětiprvkových podmnožin množiny o 10 prvcích, takže se dvě musí v součtu shodnout.

11b.27: Nechť jsou to čísla a_1, \dots, a_n , pro $k = 1, \dots, n$ označme $d_k = \sum_{i=1}^k a_i$. (Tedy uvažujeme $a_1, a_1 + a_2, \dots$)

Jestliže pro nějaké k platí $d_k \bmod n = 0$, pak je hotovo. Jinak máme n čísel d_k , která dávají modn zbytek $1, 2, \dots, n-1$, takže se dvě musí shodnout, tím pádem $d_k - d_m \bmod n = 0$ pro nějaké $m < k$ a tedy $\sum_{i=m+1}^k a_i \bmod n = 0$.

11b.28: Střed se získá jako aritmetický průměr souřadnic x a souřadnic y , tedy $\frac{1}{2}(A+B) = \left(\frac{1}{2}(a_1+b_1), \frac{1}{2}(a_2+b_2)\right)$. Je tedy třeba vybrat body tak, aby součet prvních souřadnic i součet druhých souřadnic byl sudý, což znamená, že souřanice musí mít stejnou paritu.

Bodů je pět, proto po rozdelení do dvou hromádek dle parity první souřadnice musí být alespoň tři body, jejichž první souřadnice má stejnou paritu (tři sudá či tři lichá čísla). Ty tři zase rozdělíme do dvou hromádek podle parity druhé souřadnice a v jedné z hromádek musí zbýt alespoň dva body.

11b.29: Vytvoříme 4 množiny-krabičky $\{1, 8\}$, $\{2, 7\}$, $\{3, 6\}$, $\{4, 5\}$. V každé je součet 9. Rozdělíme do nich 5 čísel, tudíž v nějaké musí být dvě.

11b.30: Pět množin $\{1, 10\}$, $\{2, 9\}$ až $\{5, 6\}$. Rozdělujeme sedm čísel, tudíž alespoň jedna množina má dvě čísla, ale nemůže mít více, tudíž zbývá pět čísel na čtyři krabičky, takže i jiná má dvě čísla.

11b.31: Krabička mužů, krabička žen, v jedné z nich musí být $\lceil \frac{9}{2} \rceil = 5$ lidí.

11b.32: Je 2999999 různých platů.

11b.33: Znamení Štíra reprezentuje maximálně 31 po sobě jdoucích dní (existují různé verze začátku/konce). Rozdělme je do dvojic, první den s druhým, třetí se čtvrtým atd., je jich 16. Pak musí z těch 17 lidí alespoň dva skončit se dnem narození v jedné dvojici, pak už budou jejich narozeniny souhlasí, nebo jsou den po sobě.

11b.34: Čtverec se rozdělí na čtyři menší o straně 1, dle šuplíkového principu musí do některého z nich padnout alespoň dva body, ty nemohou být od sebe dále než je velikost diagonály neboli $\sqrt{2}$.

11b.35: Cestu délky n dostaneme přiřazením dlaždice k cestě délky $n-1$, jsou dvě možnosti (třetí typ dlaždice by se opakoval). Proto rovnice $a_n = 2a_{n-1}$, $a_1 = 3$. Řešení rovnice $a_{n+1} - 2a_n = 0$ je $a_n = 2^n u$, hledaný počet způsobů je $3 \cdot 2^{n-1}$.

11b.36: Poslední použitá mince je buď koruna nebo dvoukoruna, tedy $a_n = a_{n-1} + a_{n-2}$ pro $n \geq 3$, $a_1 = 1$, $a_2 = 1$, vychází Fibonacciho posloupnost.

11b.37: Rozdělíme situaci podle toho, jak vypadají první dva znaky. Pokud řetězec začíná 00, pak už jsou další znaky libovolné, je tedy 2^{n-2} možností. Další možnost je, že první dvojice je 01, pak musí následovat řetězec délky $n-2$ s dvěma nulami, to je a_{n-1} možností. Poslední varianta je začátek 10 či 11, tedy jde o jedničku a pak řetězec o délce $n-1$ s dvěma nulami. Vzniká rovnice $a_n = a_{n-1} + a_{n-2} + 2^{n-2}$ pro $a_n \geq 2$, podmínky $a_1 = 0$, $a_2 = 1$. Pak $a_3 = 3$, $a_4 = 8$, $a_5 = 19$.

11b.38: Rekurentní vztah je $a_n = a_{n-1} + a_{n-2}$ pro $n \geq 2$, počáteční podmínky $a_1 = 1$, $a_2 = 2$. Jde tedy o posunutou Fibonacciho posloupnost, $a_n = F_{n+1}$. Proto $a_8 = F_9 = 34$.

11b.39: Protože každá přímka má obecně schopnost půlit oblasti a můžeme si představit, že kreslíme jednu za druhou a každá nová může půlit již existující oblasti, bude to určitě nejvíce 2^n . Ve skutečnosti to ale bude znatelně méně, protože jakmile se oblasti rozptýlí, nová přímka je nedokáže zachytit všechny. Zkuste si rozmyslet, že už třetí přímkou nikdy nedostanete osm oblastí.

Rekurentní vztah: Když zakreslíme novou přímku, tak oblasti, které protíná, musí být za sebou ve směru přímky a odděleny od sebe předchozími přímkami. Počet protnutých oblastí tedy nemůže být větší, než počet již nakreslených přímek plus jedna. Počet oblastí po nakreslení n -té přímky je tedy roven počtu původních oblastí plus počtu přeřeřatých, což je nejvíce $(n-1)+1$ (počet předchozích přímek plus jedna). Proto rovnice $R_n = R_{n-1} + n$, počáteční podmínka $R_1 = 2$.

Přepis: $R_{n+1} - R_n = n+1$, pak $a_{h,n} = u$, odhad $a_n = n(An + B) = An^2 + Bn$, po dosazení $A = B = \frac{1}{2}$, poč. podm. dá $u = 1$. Řešení $R_n = \frac{n(n+1)+2}{2}$.

11b.40: Zakreslíme n -tou rovinu a nyní se podívejme jen na ni jako na dvojrozměrný útvar. Vidíme tam přímky, což jsou průniky s předchozími rovinami, a oblasti. Každá tato dvojrozměrná oblast na n -té rovině odpovídá jedné 3D části, kterou jsme touto rovinou přeřali a tím vytvořili novou. Počet nově vytvořených částí je tedy roven počtu oblastí, které na n -té rovině vytvořilo $n-1$ přímek (průniků s předchozími rovinami), což je podle předchozího cvičení nejvíce $\frac{(n-1)n+2}{2}$. Dostáváme rovnici $S_n = S_{n-1} + \frac{n^2-n+2}{2}$ a podmínu $S_1 = 2$, odtud $S_n = \frac{n^3+5n+6}{6}$.

11b.41: Vytváříme zleva, přidáváme nenulu ke „správnému“ řetězci délky $n-1$ či 10 nebo 20 ke správnému řetězci délky $n-2$. Rovnice $a_n = 2a_{n-1} + 2a_{n-2}$ pro $n \geq 2$, podmínky $a_0 = 1$, $a_1 = 3$. Vyjde $\lambda = 1 \pm \sqrt{3}$, obecné řešení $a_n = (1 + \sqrt{3})^n u + (1 - \sqrt{3})^n v$, hledané řešení $a_n = \left(\frac{1}{2} + \frac{1}{3}\sqrt{3}\right)(1 + \sqrt{3})^n + \left(\frac{1}{2} - \frac{1}{3}\sqrt{3}\right)(1 - \sqrt{3})^n$.

11b.42: a) Prvky 1, 2, 3 se dávají na místa 4, 5, 6, nikdy tedy nemůže dojít ke shodě, libovolná permutace zabere: $3! = 6$.

b) Prvky 1, 2, 6 se permutují na místa 4, 5, 6, musíme si tedy pohlídat, aby 6 nešlo na 6. Všech permutací je $3!$, těch dřívajících 6 na 6 je $2! = 2$, proto odpověď zní $6 - 2 = 4$.

c) Prvky 1, 5, 6 se permutují na místa 4, 5, 6, musíme si tedy pohlídat, aby 5 nešlo na 5 ani 6 nešlo na 6. To je tak restriktivní, že to půjde spočítat. 6 může jít na místo 5, pak libovolné pozice ostatních nebudou mít shodu, tedy $2! = 2$ možností. Nebo 6 může jít na místo 4, pak rozdělujeme 1, 5 na místa 5, 6 a nechceme shodu, jediná možnost. Celkem 3.

d) Takové derangements nejsou, protože dvojka už je na původní pozici a zkazila to.

e) Prvky 4, 5, 6 se permutují na místa 4, 5, 6, jde tedy o otázku, kolik je derrangements tříprvkových množin. Podle vzorce je odpověď $3! \sum_{k=0}^3 \frac{(-1)^k}{k!} = 6\left(\frac{1}{1} - \frac{1}{1} + \frac{1}{2} - \frac{1}{6}\right) = 2$.

Jde to také rozborem situací, jsou to permutace 645 a 564.

11b.43: Jde o derrangements, ale nelze přímo použít hotový výsledek, protože se sudé a liché číslice mohou mezi sebou míchat. Proto se použije jen postup z příkladu o derrangements. M_i množina permutací, které nechají i -tou cifru na svém místě, počítáme permutace, které nechají některou sudou cifru na svém místě, tedy množinu $M_2 \cup M_4 \cup M_6 \cup M_8 \cup M_{10}$.

$|M_i| = 9!$, $|M_i \cap M_j| = 8!$, $|M_i \cap M_j \cap M_k| = 7!$ atd, celkem je jich

$$5 \cdot 9! - \binom{5}{2} 8! + \binom{5}{3} 7! - \binom{5}{4} 6! + 5! = 5 \cdot 9! - 10 \cdot 8! + 10 \cdot 7! - 5 \cdot 6! + 5! = 1458120.$$

Odečteme od všech permutací $10!$, dostaneme 2170680.

11b.44: a) Nejprve přesuneme 1, máme $n-1$ kandidátů. Je třeba přesunout ostatní. Nechť $1 \mapsto k$. Vezmeme tedy čísla $\{2, 3, 4, \dots, n\}$, přesuneme jejich pořadí takto: $\{k, 2, 3, 4, \dots, k-1, k+1, \dots, n\}$. Každá jejich derrangement pak dává derrangement původní množiny, když prvních $k-1$ dáme na místa 1 až $k-1$ a zbytek za tu jedničku na místě k (nakreslete si to). Je tedy D_{n-1} možností.

Nezískáme tím ale všechny možné derrangements, protože tímto způsobem nelze umístit k na pozici 1. Tyto možnosti je třeba doplnit: Pošleme $1 \mapsto k$ a $k \mapsto 1$, pak zbývajících $n-2$ prostě permutujeme a jejich derrangements dávají derrangements původní množiny. Celkem tedy $D_{n-1} + D_{n-2}$ pro situaci, kdy $1 \mapsto k$. Násobící princip pak dá výsledek.

b) Podle a) je $D_n - nD_{n-1} = -[D_{n-1} - (n-1)D_{n-2}] = -[-(D_{n-2} - (n-2)D_{n-3})] = D_{n-2} - (n-2)D_{n-3} = \dots = (-1)^n(D_2 - 2D_1) = (-1)^n(1 - 2 \cdot 0)$.

11c. Binomická věta, kombinační čísla

Než se k binomické větě dostaneme, budeme potřebovat znát jednu důležitou vlastnost kombinačních čísel.

!**Fakt 11c.1.** (Pascalova identita, Pascal's identity)

Pro všechna $k \leq n \in \mathbb{N}_0$ platí

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}.$$

Důkaz (poučný): Jedna možnost je použít algebru:

$$\begin{aligned} \binom{n}{k-1} + \binom{n}{k} &= \frac{n!}{(k-1)!(n-k+1)!} + \frac{n!}{k!(n-k)!} = \frac{n!k + n!(n-k+1)}{k!(n-k+1)!} = \frac{n!k + n!(n+1) - n!k}{k!(n-k+1)!} \\ &= \frac{(n+1)!}{k!((n+1)-k)!} = \binom{n+1}{k}. \end{aligned}$$

Alternativa: Číslo $\binom{n+1}{k}$ udává, kolik má množina M s $|M| = n+1$ podmnožin o k prvcích. Teď totto číslo dostaneme ještě jinak. Vyberme prvek $m \in M$ a rozdělme podmnožiny M podle toho, zda m obsahuje nebo ne. Ty, které jej neobsahují, jsou vlastně k -prvkovými podmnožinami množiny $M - \{m\}$ o n prvcích, je jich tedy $\binom{n}{k}$. Ty, které m obsahují, jsou pak určeny ostatními prvky, kterých je $k-1$ a jsou z $M - \{m\}$, počet těchto podmnožin je tedy stejný jako počet $(k-1)$ -prvkových podmnožin množiny $M - \{m\}$. Alternativní počítání proto dalo $\binom{n}{k} + \binom{n}{k-1}$ podmnožin. Protože jsme pokaždé počítali totéž, musí platit Pascalova identita. \square

Tomuto alternativnímu důkazu se říká „kombinatorický důkaz“, spočívá v tom, že se tatáž věc spočítá dvěma různými způsoby, výsledky se pak musejí rovnat.

!**Vzorec z Faktu** je velice užitečný z hlediska výpočetního, protože jej lze využít k rekurzivní definici:

$$(0) \quad \binom{n}{0} = 1, \quad \binom{n}{n} = 1.$$

$$(1) \quad \binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k} \text{ pro } k \leq n \in \mathbb{N}_0.$$

Výhoda tohoto vzorce je, že používá jen sčítání, proto je výpočet tímto způsobem často rychlejší, zejména pokud potřebujeme spočítat více kombinačních čísel najednou. Používá se to občas i při ručním výpočtu a pak má Pascalova nerovnost i velice intuitivní geometrickou podobu, které se říká Pascalův trojúhelník. Srovnáme si kombinační čísla pod sebe do řádků podle n , jednou symbolicky a podruhé skutečné hodnoty. V rámci úspory místa nenapíšeme všech nekonečně mnoho čísel, ale skončíme s $n = 5$.

$\binom{0}{0}$	$\binom{1}{0}$	$\binom{1}{1}$	$\binom{2}{0}$	$\binom{2}{1}$	$\binom{2}{2}$	$\binom{3}{0}$	$\binom{3}{1}$	$\binom{3}{2}$	$\binom{3}{3}$	$\binom{4}{0}$	$\binom{4}{1}$	$\binom{4}{2}$	$\binom{4}{3}$	$\binom{4}{4}$	$\binom{5}{0}$	$\binom{5}{1}$	$\binom{5}{2}$	$\binom{5}{3}$	$\binom{5}{4}$	$\binom{5}{5}$
1						1				1	2	1			1	3	3	1		
										1	4	6	4	1						
										1	5	10	10	5	1					

Podívejte se teď na ten pravý trojúhelník. Na hranách má jedničky, to plyne z Faktu 11a.4. Fakt 11c.1 tam vidíme tak, že libovolné číslo mimo hranu získáme součtem dvou čísel, které jsou nad ním a bezprostředně doleva a doprava. Jinak řečeno, když vezmeme dvě sousední čísla v Pascalově trojúhelníku a sečteme, dostaneme číslo pod jejich středem.

V trojúhelníku vidíme i další věci, třeba symetrii, která vyplývá z Faktu 11a.4. Vidíme také, že v každém řádku čísla nejprve rostou a pak klesají. To si potvrďme obecně:

Fakt 11c.2.

Nechť $n \in \mathbb{N}$. Pak platí:

$$1 = \binom{n}{0} < \binom{n}{1} < \binom{n}{2} < \dots < \binom{n}{\lfloor \frac{n}{2} \rfloor} = \binom{n}{\lceil \frac{n}{2} \rceil} > \dots > \binom{n}{n-1} > \binom{n}{n}.$$

Důkaz (poučný): 1) Nejprve dokážeme tu rovnost uprostřed. Pokud je n sudé, pak ta rovnost vlastně říká $\binom{n}{\frac{n}{2}} = \binom{n}{\frac{n}{2}}$, což je určitě pravda.

Pokud je n liché, pak $\lfloor \frac{n}{2} \rfloor = \frac{n-1}{2}$, $\lceil \frac{n}{2} \rceil = \frac{n+1}{2}$ a máme

$$\binom{\frac{n-1}{2}}{\frac{n-1}{2}} = \frac{n!}{(\frac{n-1}{2})!(n-\frac{n-1}{2})!} = \frac{n!}{(\frac{n-1}{2})!(\frac{n+1}{2})!} = \frac{n!}{(n-\frac{n+1}{2})!(\frac{n+1}{2})!} = \binom{\frac{n+1}{2}}{\frac{n+1}{2}}.$$

2) Teď ukážeme nerovnosti, díky symetrii je stačí ukázat pro levou polovinu. Mějme tedy k splňující $k < \lfloor \frac{n}{2} \rfloor$. Prozkoumáme kýzenou nerovnost:

$$\begin{aligned} \binom{n}{k} < \binom{n}{k+1} &\iff \frac{n!}{k!(n-k)!} < \frac{n!}{(k+1)!(n-k-1)!} \iff k+1 < n-k \\ &\iff 2k < n-1 \iff k < \frac{n-1}{2}. \end{aligned}$$

Poslední nerovnost pro k splňující $k < \lfloor \frac{n}{2} \rfloor$ platí, což je vidět například rozbořením pro n sudé a n liché. □

Ke kombinačním číslům se ještě vrátíme.

Binomickou větu asi každý čtenář zná alespoň v té nejjednodušší formě pro $(x+y)^2$, právě výrazu $x+y$ říkáme binom (neboli dvojčlen). Některí čtenáři patrně umí rozvinout i $(x+y)^3$, jak je to dál?

! Věta 11c.3. (binomická věta, binomial theorem)

Pro každé $n \in \mathbb{N}$ a všechna $x, y \in \mathbb{R}$ platí

$$\begin{aligned} (x+y)^n &= \sum_{k=0}^n \binom{n}{k} x^k y^{n-k} = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k \\ &= x^n + nx^{n-1}y + \frac{n(n-1)}{2}x^{n-2}y^2 + \dots + \frac{n(n-1)}{2}x^2y^{n-2} + nxy^{n-1} + y^n. \end{aligned}$$

Všimněte si, že díky symetrii kombinačních čísel si můžeme vybrat, jestli se bude dolní číslo k u $\binom{n}{k}$ shodovat s mocninou u x nebo s mocninou u y , proto jsme to napsali oběma způsoby.

Důkaz (poučný): Důkazů existuje povíce, například indukcí na n . Mějme čísla $x, y \in \mathbb{R}$.

(0) Pro $n = 1$ jistě platí $(x+y)^1 = x+y = \binom{1}{0}x + \binom{1}{1}y$.

(1) Předpokládejme, že vzorec pro $(x+y)^n$ platí. Pak pomocí běžné algebry a Faktu 11c.1 máme

$$\begin{aligned}
 (x+y)^{n+1} &= (x+y)(x+y)^n = (x+y) \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k = x \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k + y \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k \\
 &= \sum_{k=0}^n \binom{n}{k} x^{n-k+1} y^k + \sum_{k=0}^n \binom{n}{k} x^{n-k} y^{k+1} = \left| \begin{array}{l} j = k+1 \\ k = 0 \mapsto j = 1 \end{array} \right| \\
 &= \sum_{k=0}^n \binom{n}{k} x^{n-k+1} y^k + \sum_{j=1}^{n+1} \binom{n}{j-1} x^{n-(j-1)} y^j \\
 &= x^{n+1} + \sum_{k=1}^n \binom{n}{k} x^{n-k+1} y^k + \sum_{j=1}^n \binom{n}{j-1} x^{n-j+1} y^j + y^{n+1} \\
 &= x^{n+1} + \sum_{k=1}^n \left(\binom{n}{k} + \binom{n}{k-1} \right) x^{n-k+1} y^k + y^{n+1} \\
 &= x^{n+1} + \sum_{k=1}^n \binom{n+1}{k} x^{(n+1)-k} y^k + y^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} x^{(n+1)-k} y^k.
 \end{aligned}$$

Lze také zkousit kombinatorický důkaz. Při roznásobování $(x+y)(x+y) \cdots (x+y)$ se bere člen z každé závorky. Abychom dostali $x^k y^{n-k}$, musíme vybrat, z kterých k závorek vybereme x , to je možné $\binom{n}{k}$ způsoby.

□

Takže například

$$(x+y)^5 = x^5 + 5x^4y + \binom{5}{2}x^3y^2 + \binom{5}{3}x^2y^3 + 5xy^4 + y^5 = x^5 + 5x^4y + 10x^3y^2 + 10x^2y^3 + 5xy^4 + y^5.$$

Porovnejte poslední řádek Pascalova trojúhelníka výše s koeficienty tohoto rozkladu. Je to příjemné, na druhou stranu kreslit trojúhelník o dvaceti řádcích kvůli rozkladu $(x+y)^{20}$ asi není nejlepší metoda. Na to se hodí následující zajímavý trik:

Fakt 11c.4.

Nechť $k < n \in \mathbb{N}_0$. Pak platí:

$$\begin{aligned}
 \text{(i)} \quad &\binom{n}{k} \cdot \frac{n-k}{k+1} = \binom{n}{k+1}, \\
 \text{(ii)} \quad &\binom{n}{k} x^{n-k} y^k \cdot \frac{n-k}{k+1} \frac{y}{x} = \binom{n}{k+1} x^{n-(k+1)} y^{k+1}.
 \end{aligned}$$

Důkaz (rutinní): (i):

$$\binom{n}{k} \cdot \frac{n-k}{k+1} = \frac{n!}{k!(n-k)!} \cdot \frac{n-k}{k+1} = \frac{n!}{(k+1)!(n-k-1)!} = \binom{n}{k+1}$$

Z toho hned plynne (ii).

□

Takže například v rozkladu $(x+y)^{20}$ máme postupně členy x^{20} , $x^{20} \frac{20}{1} \frac{y}{x} = 20x^{19}y$, $20x^{19}y \frac{19}{2} \frac{y}{x} = 190x^{18}y^2$, $190x^{18}y^2 \frac{18}{3} \frac{y}{x} = 1140x^{17}y^3$, $1140x^{17}y^3 \frac{17}{4} \frac{y}{x} = 4845x^{16}y^4$ atd.

Binomická věta je silně užitečná, tím spíš, že se dá dále zobecnit, a to dokonce dvěma směry. Jedna možnost je umocňovat mnohočleny neboli multinomy. Jako inspiraci si nejprve uvedeme ještě jinou verzi binomické identity:

$$(x+y)^n = \sum_{i+j=n} \frac{n!}{i!j!} x^i y^j.$$

Zobecnění na mnohočlen pak funguje podobně.

Věta 11c.5. (multinomická věta, multinomial theorem)

Pro každé $n \in \mathbb{N}$ a všechna $x_1, \dots, x_m \in \mathbb{R}$ platí

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{n_1+n_2+\cdots+n_m=n} \frac{n!}{n_1!n_2!\cdots n_m!} x_1^{n_1} x_2^{n_2} \cdots x_m^{n_m}.$$

Někteří autoři zavádějí značení $\binom{n}{n_1, n_2, \dots, n_m} = \frac{n!}{n_1!n_2!\cdots n_m!}$, jde o zobecněná kombinační čísla. Mají také kombinatorický význam, říkají nám, kolika způsoby lze (bez ohledu na pořadí výběru) vybrat n_1 objektů do první krabičky, n_2 objektů do druhé atd, viz 11b.7 c).

Například máme

$$\begin{aligned}(x+y+z)^2 &= \sum_{i+j+k=2} \binom{2}{i,j,k} x^i y^j z^k \\ &= \binom{2}{2,0,0} x^2 y^0 z^0 + \binom{2}{0,2,0} x^0 y^2 z^0 + \binom{2}{0,0,2} x^0 y^0 z^2 \\ &\quad + \binom{2}{1,1,0} x^1 y^1 z^0 + \binom{2}{1,0,1} x^1 y^0 z^1 + \binom{2}{0,1,1} x^0 y^1 z^1 \\ &= x^2 + y^2 + z^2 + 2xy + 2xz + 2yz.\end{aligned}$$

Další možnost je zobecnit binomickou větu pro necelé mocniny. Nejprve ale potřebujeme rozšířit definici kombinačního čísla. Nejprve připomeňme, že pro $k \leq n \in \mathbb{N}$ máme

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+2) \cdot (n-k+1)}{k!}.$$

To nás inspiruje k následující definici.

Definice

Nechť $\alpha \in \mathbb{R}$ a $k \in \mathbb{N}_0$. Pak definujeme

$$\binom{\alpha}{k} = \prod_{i=1}^k \frac{\alpha - i + 1}{i}.$$

U takto zobecněných kombinačních čísel již nemusí platit, že výsledkem je přirozené číslo, například

$$\binom{\frac{1}{2}}{3} = \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)}{1 \cdot 2 \cdot 3} = \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})}{2 \cdot 3} = \frac{1}{16}.$$

Rozmyslete si, že pro $k \geq 2$ z definice dostáváme

$$\binom{\alpha}{k} = \frac{\alpha \cdot (\alpha-1) \cdots (\alpha-k+2) \cdot (\alpha-k+1)}{k!}.$$

Máme také $\binom{\alpha}{1} = \alpha$ a $\binom{\alpha}{0} = 1$, protože pak se v součinu násobí přes prázdnou množinu, což je podle definice rovno jedné.

Všimněte si ještě, že pokud $\alpha \in \mathbb{N}$, tak je v případě $k \leq \alpha$ výraz $\binom{\alpha}{k}$ roven kombinačnímu číslu dle původní definice, a v případě $k > \alpha$ je $\binom{\alpha}{k} = 0$, protože se mezi činiteli v čitateli objeví nula.

Ted už jsme připraveni.

! Věta 11c.6. (Newtonův binomický rozvoj, Newton binomial expansion)

Pro každé $\alpha > 0$ a všechna $x, y \in \mathbb{R}$ splňující $|x| < |y|$ platí

$$\begin{aligned}(x+y)^\alpha &= \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k y^{\alpha-k} \\ &= y^\alpha + \alpha xy^{\alpha-1} + \frac{\alpha(\alpha-1)}{2} x^2 y^{\alpha-2} + \frac{\alpha(\alpha-1)(\alpha-2)}{3!} x^3 y^{\alpha-3} + \dots\end{aligned}$$

Suma jde do nekonečna, což je problém, který se řeší v analýze, kde se dozvímme, že některé nekonečné součty smysl mají a jiné ne. V tomto případě nám úspěch zaručuje právě podmínka $|x| < |y|$. Jako zajímavou aplikaci si ukažme následující vzorec platný pro $|x| < 1$:

$$\sqrt{1+x} = (x+1)^{1/2} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \dots$$

Vynecháním vyšších mocnin dostáváme odhad $\sqrt{1+x} \sim 1 + \frac{1}{2}x$, který je pro malá x dost dobrý (například pro $|x| < 0.5$ už má relativní chybu menší než 2%).

Všimněte si, že pokud vezmeme $\alpha \in \mathbb{N}$, pak budou skoro všechna kombinační čísla nulová a vznikne z toho standardní suma jako ve Větě 11c.3.

Teď se podíváme na několik zajímavých důsledků binomické věty.

Důsledek 11c.7.

Nechť $n \in \mathbb{N}$. Pak platí

$$(i) \sum_{k=0}^n \binom{n}{k} = 2^n;$$

$$(ii) \sum_{k=0}^n (-1)^k \binom{n}{k} = 0.$$

Důkaz (poučný): (i): Stačí rozvinout $2^n = (1+1)^n$.

Alternativa: Lze provést kombinatorický důkaz spočítáním počtu podmnožin, viz příklad 11a.i.

(ii): Stačí rozvinout $0^n = (1-1)^n$. □

Z části (ii) plyne, že $\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \dots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \dots$. O kombinačních číslech se dá dokázat doslova stovky vztahů, jsou to velice zajímavé objekty. Několik takových si ukážeme.

Fakt 11c.8.

$$\text{Nechť } k \leq n \in \mathbb{N}_0. \text{ Pak } \sum_{j=k}^n \binom{j}{k} = \binom{n+1}{k+1}.$$

Důkaz (poučný): Kombinatorický důkaz: Kolik je binárních řetězců o délce $n+1$, které obsahují přesně $k+1$ jedniček? Jedna možnost je místa pro jedničky vybrat, to je to číslo na pravé straně.

Druhá možnost je rozdělit situaci podle toho, kde je poslední jednička. Protože je jedniček $k+1$, tak možnosti pro poslední jedničku jsou $i = k+1, k+2, \dots, n+1$. Pro každou takovou pozici je pak třeba vybrat místo pro předchozích k jedniček a na výběr je z $j = i-1$ pozic, tedy $\binom{j}{k}$ možností. □

Fakt 11c.9. (Vandermondeho identita)

$$\text{Pro všechna } k, m, n \in \mathbb{N}_0 \text{ taková, že } k \leq m \text{ a } k \leq n, \text{ platí } \sum_{i=0}^k \binom{n}{i} \binom{m}{k-i} = \binom{m+n}{k}.$$

Důkaz (poučný): Zase provedeme kombinatorický důkaz. Nechť A, B jsou libovolné disjunktní množiny takové, že $|A| = m$ a $|B| = n$. Kolik je podmnožin $A \cup B$ o k prvcích?

Přímý výběr dá číslo na pravé straně. Další možnost je rozdělit tento počet podle toho, kolik z prvků se vybírá z množiny A . Je zjevné, že když vybereme i prvků z množiny A , pak musíme zbývajících $k-i$ dobrat z B a jde o nezávislé fáze výběru, tudíž se použije násobící princip. Z toho hned dostáváme vzorec na levé straně. □

Důsledek 11c.10.

$$\text{Pro } n \in \mathbb{N}_0 \text{ platí } \sum_{i=0}^n \binom{n}{i}^2 = \binom{2n}{n}.$$

Další identity čtenář najde ve cvičeních.

Kombinační čísla se pro větší hodnoty k, n obtížně počítají, takže podobně jako u faktoriálu někdy raději použijeme odhad. Pro dobré odhady je třeba udělat hodně práce, jako ukázkou si uvedeme několik lehčích.

Fakt 11c.11.

Nechť $n \in \mathbb{N}$.

$$(i) \text{ Pro } k \in \mathbb{N}, k \leq n \text{ platí } \binom{n}{k} \leq 2^n.$$

$$(ii) \text{ Platí } \binom{n}{\lfloor \frac{n}{2} \rfloor} \geq \frac{2^n}{n}.$$

Důkaz (poučný): (i): Toto okamžitě vyplývá z Důsledku 11c.7 (i), sčítáme tam kladná čísla, takže každé z nich musí být nejvýše rovno jejich součtu.

(ii): Pro $n = 0$ a $n = 1$ se to ověří hned, pro $n \geq 2$ to dokážeme sporem. Předpokládejme tedy, že $n \geq 2$ a $\binom{n}{\lfloor \frac{n}{2} \rfloor} < \frac{2^n}{n}$. Pak podle Faktu 11c.2 bude pro všechna k platit $\binom{n}{k} \leq \frac{2^n}{n} - 1$ a tudíž $\sum_{k=1}^n \binom{n}{k} \leq 2^n - n$. Proto

$$\sum_{k=0}^n \binom{n}{k} = 1 + \sum_{k=1}^n \binom{n}{k} \leq 1 + 2^n - n < 2^n,$$

což je ve sporu s Důsledkem 11c.7 (i). □

Další odhad ukážeme ve cvičení 11c.12.

Cvičení

Cvičení 11c.1 (rutinní, dobré): Najděte n , jestliže

- a) $\binom{n}{2} = 45$;
- b) $\binom{n}{8} = \binom{n}{5}$.

Cvičení 11c.2 (rutinní): Najděte rozvoj $(1+x)^7$.

Cvičení 11c.3 (rutinní, *dobré): Jaký je koeficient u x^k v rozkladu

- (i) $(x+1)^{100}$;
- (ii) $(x+\frac{1}{2})^{100}$;
- (iii) $(x+y)^{100}$;
- (iv)* $(x+x^2)^{100}$;
- (v)* $(x+1/x)^{100}$;
- (vi)* $(x^2-1/x)^{100}$.

Cvičení 11c.4 (rutinní): Dokažte indukcí na n , že pro každé $k \leq n \in \mathbb{N}_0$ platí $\binom{n}{k} \in \mathbb{N}$.

Nápověda: Pascalova identita.

Cvičení 11c.5 (poučné): Dokažte, že jestliže je p prvočíslo a $k \in \mathbb{N}$, $k < p$, pak p dělí $\binom{p}{k}$.

Cvičení 11c.6: Dokažte, že pro $k \leq n \in \mathbb{N}$ platí $k\binom{n}{k} = n\binom{n-1}{k-1}$

Nápověda: Kombinatorický důkaz, viz příklad 11a.m.

Cvičení 11c.7: Dokažte, že pro $m \leq k \leq n \in \mathbb{N}_0$ platí $\binom{n}{k}\binom{k}{m} = \binom{n}{m}\binom{n-m}{k-m}$.

Nápověda: Kombinatorický důkaz, viz příklad 11a.m.

Cvičení 11c.8: Dokažte, že pro každé $n \in \mathbb{N}_0$ platí $\sum_{k=0}^n 2^k \binom{n}{k} = 3^n$.

Nápověda: Binomická věta.

Cvičení 11c.9: Dokažte pomocí matematické indukce, že pro $n \in \mathbb{N}$, $n \geq 2$ platí $\sum_{k=2}^n \binom{k}{2} = \binom{n+1}{3}$.

Cvičení 11c.10: Dokažte, že pro $n \in \mathbb{N}$ platí $\sum_{k=1}^n k\binom{n}{k} = n2^{n-1}$

Cvičení 11c.11: Dokažte, že pro $n, m \in \mathbb{N}$ platí $\sum_{k=0}^m \binom{n+k}{k} = \binom{n+m+1}{m}$

Cvičení 11c.12 (poučné): Dokažte, že pro $k \leq n \in \mathbb{N}$ platí $\binom{n}{k} \leq \frac{n^k}{2^{k-1}}$.

Řešení:

11c.1: a) $\frac{n(n-1)}{2} = 45 \implies n^2 - n - 90 = 0 \implies n = \frac{1}{2} \pm \frac{1}{2}\sqrt{19}$, $n = 10$.

b) Každý řádek v Pascalově trojúhelníku je symetrický podle středu, tudíž musí platit $n = 8 + 5 = 13$.

11c.2: $1 + 7x + 21x^2 + 35x^3 + 35x^4 + 21x^5 + 7x^6 + x^7$.

11c.3: (i): Pro $k > 100$ nebo $k < 0$ je to 0, jinak $\binom{100}{k}$.

(ii): Pro $k > 100$ nebo $k < 0$ je to 0, jinak $\binom{100}{k}\left(\frac{1}{2}\right)^{100-k} = \binom{100}{k}2^{k-100}$.

(iii): Pro $k > 100$ nebo $k < 0$ je to 0, jinak $\binom{100}{k}y^{100-k}$.

(iv): j -tý člen rozvoje je $\binom{n}{j}x^{100-j}x^{2j} = \binom{n}{j}x^{100+j}$. Pro $k > 200$ nebo $k < 100$ je koeficient 0, jinak je $\binom{n}{100-k}$. (v): j -tý člen rozvoje je $\binom{n}{j}x^{100-j}x^{-j} = \binom{n}{j}x^{100-2j}$. Pro $k > 100$ nebo $k < -100$ nebo k liché je koeficient 0, jinak je $\binom{n}{(100-k)/2}$. (vi): j -tý člen rozvoje je $\binom{n}{j}(-1)^jx^{200-2j}x^{-j} = \binom{n}{j}(-1)^jx^{100-3j}$. Pro $k > 100$ nebo $k < -200$ nebo k splňující $k \bmod 3 = 1$ či $k \bmod 3 = 0$ je koeficient 0, jinak je $(-1)^{(200-k)/3}\binom{100}{(200-k)/3}$.

11c.4: $V(n)$: Pro každé $k \in \{0, 1, \dots, n\}$ platí $\binom{n}{k} \in \mathbb{N}$.

(0) Pro $n = 0$ to ověříme dosazením.

(1) Předpokládejme, že pro nějaké (libovolné) $n \in \mathbb{N}$ platí $V(n)$. Ověříme teď platnost $V(n+1)$.

Nechť $k \in \{0, \dots, n+1\}$. Pokud $k = 0$ nebo $k = n+1$, pak $\binom{n+1}{k} = 1 \in \mathbb{N}$. Jinak platí $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n-1}{k}$ a ona dvě čísla napravo jsou z \mathbb{N} dle indukčního předpokladu.

11c.5: $\binom{p}{k} = \frac{p \cdot (p-1) \cdots (n-p+1)}{2 \cdot 3 \cdots k}$. Výsledek je celé číslo, proto se $k!$, musí zkrátit, ale čísla $2, \dots, k$ nemohou krátit p (je to prvočíslo a větší než ta čísla), tudíž se musí zkrátit s čísly $(p-1)(p-2) \cdots 2$.

11c.6: Jev: vybrat delegaci o k lidech a v ní určit mluvčího. Buď nejprve delegaci a z ní mluvčího, nebo nejprve mluvčího a k němu dobrat zbytek delegace.

11c.7: Kombinatorický důkaz: Vybrat delegaci o k členech a v ní podvýbor o m členech.

11c.8: $3^n = (1+2)^n = \dots$

11c.9: (0) $n = 2$: $1 = 1$ O.K. (1) $V(n) \implies V(n+1)$: $\sum_{k=2}^{n+1} \binom{k}{2} = \sum_{k=2}^n \binom{k}{2} + \binom{n+1}{2} = \binom{n+1}{3} + \binom{n+1}{2} = \binom{n+2}{3}$ podle Pascalovy identity.

11c.10: Kombinatorický důkaz: Kolika způsoby je z n lidí možno vybrat nějakou delegaci a jejího mluvčího? Buď vybíráme delegace různých velikostí a z nich mluvčí (nalevo), nebo vybereme mluvčího a k němu doplníme zbytek delegace, což jsou podmnožiny z $n-1$ lidí. Lze také důkaz indukcí, mnohem horší.

11c.11: Kombinatorický důkaz: Kolik je binárních řetězců s m nulami a $n+1$ jedničkami? Buď vybereme nuly hned, nebo to rozdělíme na případy podle pozice poslední nuly. Pak se použije $\sum_{i=n}^{n+m} \binom{i}{i-n} = \sum_{i=0}^m \binom{n+i}{i}$. Nebo indukcí na m .

11c.12: $\binom{n}{k} = \frac{n \cdot (n-1) \cdots (n-k+1)}{2 \cdot 3 \cdots k} \leq \dots$

11d. Bonus: Generování výběrů

Mnohé kombinatorické situace se zkoumají výpisem možností, často pomocí počítače, ale i rukou na papíře. Pak je důležité umět je generovat nějak plánovitě, abychom některou možnost nevynechali ani nepočítali dvakrát. Jinými slovy, při tvoření situací (výběrů, pořadí) je potřebujeme umět uspořádat a pak jít od jedné k další a další. Osvědčuje se tu lexikografické zobrazení pro uspořádané k -tice čísel: $(a_1, a_2, \dots, a_k) \prec (b_1, b_2, \dots, b_k)$ právě tehdy, pokud pro nějaké $m \in \{1, 2, \dots, k\}$ platí $a_m < b_m$ a $a_i = b_i$ pro všechna $1 \leq i < m$ (viz řazení dle abecedy, podrobnosti v části 4b.17).

Dvojímu použití odpovídají i dva přístupy, o které se zde budeme pokoušet. Při použití počítače nám stačí znát algoritmus, který vysype všechny možnosti. To je většinou relativně snadné pomocí vnořených cyklů, případně podmínek. Vnořené cykly ale nejsou příliš pohodlné v situaci, kdy generujeme možnosti ručně. Proto se budeme zabývat také jiným problémem: Jak v situaci, kdy už máme vygenerovanou určitou možnost, dostaneme tu bezprostředně další v lexikografickém uspořádání. To bývá někdy dobrodružnější.

Většinou to není nic složitého a pro čtenáře může být zajímavé si každou situaci nejprve zkusit rozmyslet sám a pak to porovnat s předloženými myšlenkami.

11d.1 Generování všech binárních řetězců délky n .

Algoritmus pro výpis všech řetězců je snadný, ukážeme jej pro obecnější případ v části 11d.3. Teď se zaměříme na postupné generování.

Nejmenší n -bitové binární číslo v lexikografickém uspořádání je 00...000 a největší 11...111. Je-li dáno určité číslo, pak to další získáme velice snadno, binárním přičtením jedničky, například takto: 00000, 00001, 00010, 00011, 00100, ..., 11111. Získání dalšího řetězce je tedy velice jednoduché.

```
procedure NextBinChain(a: binary chain)
output: a + 1;
```

11d.2 Generování všech podmnožin konečné množiny.

Mějme množinu množinu $A = \{a_1, \dots, a_n\}$. Pak stačí vygenerovat všechny binární řetězce délky n a použít reprezentace podmnožin, tj. pro daný binární řetězec $b_1 \dots b_n$ dostaneme podmnožinu $M = \{a_i; b_i = 1\}$.

11d.3 Generování všech výběrů k prvků z n různých objektů, kde na pořadí záleží a s opakováním.

Jinými slovy, jde o generování všech možných vektorů o k souřadnicích, kde se souřadnice berou z n -prvkové množiny. Pokud se prvky očíslují, jde vlastně o nezávislé výběry z množiny $\{1, 2, \dots, n\}$, na to je snadný algoritmus.

```

procedure AllSelections( $n, k$ : integer)
for  $a_1 := 1$  to  $n$ 
  for  $a_2 := 1$  to  $n$ 
    for  $a_3 := 1$  to  $n$ 
      :
      for  $a_k := 1$  to  $n$ 
        output:  $(a_1, a_2, \dots, a_k)$ ;
  
```

Tento algoritmus se snadno upraví na případy, kdy má každá pozice svou vlastní horní mez n_i či dokonce i dolní mez m_i , takže lze třeba vybírat z množiny $\{0, 1\}$ neboli generovat binární řetězce.

Jak se budou dělat postupné výběry? Pokud vybíráme z množiny $\{0, 1, \dots, n - 1\}$, pak stačí použít algoritmus z 11d.1, jen se berou čísla v soustavě o základu n . Například klasická dekadická reprezentace koresponduje výběru z deseti prvků.

Příklad 11d.a: Pokud budeme chtít vybírat ze tří předmětů, tak je zakódujeme 0, 1 a 2 a pomocí přičítání jedničky v trojkové soustavě budeme dostávat řetězce 00...000, 00...001, 00...002, 00...010, 00...011, 00...012, 00...020, 00...021 atd. až po 22...222.

△

Jde vlastně o klasický princip tachometru. Přičítáme k poslední cifre, a pokud na ní dosáhneme maxima, tak se při dalším přičtení přidá jednička k předposlední pozici a ta poslední se vynuluje. Tam ale také může dojít k přetečení, v tom případě se vynuluje a přičte se jednička ještě ještě o pozici předtím a tak dále. Tento princip snadno upravíme i pro výběr z prvků $\{1, 1, \dots, n\}$, kde je nejmenším výběrem (z lexikografického pohledu) výběr $(1, 2, \dots, 1)$ a největší je (n, n, \dots, n) . Při přechodu od jednoho výběru k dalšímu bychom měli podle principu tachometru přičíst jedničku a pak hlídat překročení horní meze, ale postupné přičítání a přelévání je zbytečně pracné. Rozmyslete si, že vlastně stačí najít poslední složku, která není n , přičíst k ní jedničku a „vynulovat“ všechny složky, které jsou za ní.

Algoritmus pro nalezení bezprostředně dalšího výběru v lexikografickém uspořádání, máme-li (a_1, a_2, \dots, a_k) a všechny složky nejsou n :

```

procedure NextSelection( $n$ : integer,  $(a_1, a_2, \dots, a_k)$ )
 $i := k$ ;
while  $a_i = n$  do
   $a_i := 1$ ;
   $i := i - 1$ ;
   $a_i := a_i + 1$ ;
output:  $(a_1, a_2, \dots, a_k)$ ;
  
```

Příklad 11d.b: Vygenerujeme teď (prvním či druhým) algoritmem všechny výběry tří znaků z 0 a 1 neboli všechny třímístné binární řetězce. Vypíšeme je jako sloupce jeden vedle druhého zleva doprava a přidáme čárky, ke kterým se dostaneme v následující poznámce.

0	0	0	0	1	1	1	1			
0	0	1	1	0	0	1	1			
0		1		0		1		0		1

Vidíme, že opravdu jsou trojice generovány v lexikografickém pořadí.

△

Když se na výpis výše podíváme jako na celek, tak vidíme ještě jeden způsob, jak generovat, a to po jednotlivých složkách a v blocích. Nejprve vyplníme první místo, a to n^{k-1} -krát prvním znakem, n^{k-1} -krát druhým znakem atd. až n^{k-1} -krát n -tým znakem, vznikne tím n^k možností rozdělených do n bloků. V každém z těchto bloků pak vyplníme druhou pozici stejným způsobem, jmenovitě n^{k-2} -krát prvním znakem atd. až n^{k-2} -krát n -tým znakem, celkem tedy $n \cdot n^{k-2} = n^{k-1}$ možností, což je přesně velikost jednoho bloku. Původních n^k bloků se teď rozpadá na n^{k-2} bloků vyznačujících se tím, že v každém z nich jsou první dvě pozice stejné. Každý tento menší blok pak na třetí pozici vyplníme ještě menšími bloky o velikosti n^{k-3} atd. až na poslední pozici jen stále opakujeme výpis znaků. Takto se obvykle generují řádky pravdivostních tabulek v logice.

Poznámka: Jak se situace změní, když potřebujeme uspořádané výběry bez opakování? Není to snadné, protože na rozdíl od situací probíraných níže to nelze zařídit nějak snadno kombinatoricky, v zásadě se musí použít generování podobné této části, ale zabudovat do něj test proti opakování. To se dá dělat více či méně elegantně, ale to už je spíš problém pro algoritmizaci než kombinatoriku.

Jeden příklad se kombinatoricky udělat dá, když $n = k$, pak jde totiž o permutace, viz 11d.6.

△

11d.4 Generování všech výběrů k prvků z n různých objektů, kde na pořadí nezáleží a s opakováním.

Pro zjednodušení zase předpokládejme, že jde o objekty $\{1, 2, \dots, n\}$. Protože na pořadí nezáleží, můžeme si každý výběr seřadit podle velikosti. Jde tedy o generování všech uspořádaných k -tic (a_1, a_2, \dots, a_k) splňujících $1 \leq a_1 \leq a_2 \leq \dots \leq a_k \leq n$. Bude to podobné jako předchozí algoritmus, musí se ale modifikovat tak, aby výsledné k -tice byly neklesající.

```
procedure AllCombinations1(k, n: integer)
for a1 := 1 to n
  for a2 := a1 to n
    for a3 := a2 to n
      :
      for ak := ak-1 to n
        output: (a1, a2, ..., ak);
```

Ted' se podíváme na postupné generování. Je zjevné, že nejmenší (vzhledem k lexikografickému uspořádání) je zase $(1, 1, \dots, 1)$ a největší (n, n, \dots, n) . Naše restrikce na výběry se dá snadno zabudovat do algoritmu z části 11d.3. Stačí upravit proces „vynulování“ po zvýšení složky a_i o jedničku, teď hodnoty dalších pozic nevracíme k 1, ale k nejmenšímu možnému číslu (výsledná čísla nesmí klesat), tedy k nové hodnotě a_i .

Algoritmus pro vytvoření následujícího výběru v lexikografickém uspořádání, když už jeden výběr máme a předpokládáme, že není maximální, tj. $a_1 \leq \dots \leq a_k$ a $a_1 < n$:

```
procedure NextCombination1(n: integer, (a1, a2, ..., ak))
i := k;
while ai = n do
  i := i - 1;
  ai := ai + 1;
  for j := i + 1 to k
    aj := ai;
  output: (a1, a2, ..., ak);
```

Příklad 11d.c: Vybíráme tři znaky z množiny $\{1, 2, 3, 4\}$: 111, 112, 113, 114, 122, 123, 124, 133, 134, 144, 222, 223, 224, 233, 234, 244, 333, 334, 344, 444.

Zkuste si je sami vygenerovat pomocí obou algoritmů, ať se přesvědčíte, že to funguje.

△

11d.5 Generování všech výběrů k prvků z n různých objektů, kde $k \leq n$, na pořadí nezáleží a bez opakování.

Pro zjednodušení zase předpokládejme, že jde o objekty $\{1, 2, \dots, n\}$ a budeme je vždy řadit dle velikosti. Jde tedy o generování všech uspořádaných k -tic (a_1, a_2, \dots, a_k) splňujících $1 \leq a_1 < a_2 < \dots < a_k \leq n$. Bude to podobné jako předchozí algoritmus, musí se ale modifikovat tak, aby výsledné k -tice byly rostoucí. Dostáváme tím ale novou situaci. V přechozích dvou situacích jsme mohli hodnoty a_i volit bez omezení, teď to ale není možné. Pokud bychom například zvolili $a_1 = n$, muselo by být $a_2 > n$, což není možné.

Jakých hodnot tedy mohou jednotlivé souřadnice dosahovat? Máme-li a_i , pak musí platit $a_{i+1} \geq a_i + 1$, $a_{i+2} \geq a_{i+1} + 1 \geq a_i + 2$ atd., dojdeme k $a_k \geq a_i + (k - i)$, zároveň musí platit $a_k \leq n$, proto máme obecné omezení $a_i \leq n - k + i$.

```
procedure AllCombinations2(k, n: integer)
for a1 := 1 to n - k + 1
  for a2 := a1 + 1 to n - k + 2
    for a3 := a2 + 1 to n - k + 3
```

```

    :
for  $a_k := a_{k-1} + 1$  to  $n$ 
    output:  $(a_1, a_2, \dots, a_k);$ 
```

Nyní se podíváme na postupné výběry. Nejmenší výběr je $(1, 2, \dots, k-1, k)$, při přechodu od jednoho k dalšímu budeme odzadu zvyšovat. Složka se dá zvýšit, pokud nedosáhla svého možného maxima (viz nerovnost výše), představme si tedy, že jsme složku a_i zvýšili o jedničku. Pak je třeba „vynulovat“ následující složky, v tomto případě postupně přidáváme jedničky k nové hodnotě a_i , abychom zajistili, že výsledný výběr roste, ale zároveň to udělali nejúspornějším možným způsobem.

Algoritmus pro vytvoření následujícího výběru v lexikografickém uspořádání, když máme výběr (a_1, a_2, \dots, a_k) splňující $a_1 < \dots < a_k$ a předpokládáme, že není maximální, tedy $a_1 < n - k + 1$:

```

procedure NextCombination2( $n$ : integer,  $(a_1, a_2, \dots, a_k)$ )
 $i := k;$ 
while  $a_i = n - k + i$  do
     $i := i - 1;$ 
     $a_i := a_i + 1;$ 
    for  $j := i + 1$  to  $k$ 
         $a_j := a_i + j - i;$ 
    output:  $(a_1, a_2, \dots, a_k);$ 
```

Příklad 11d.d: Vybíráme tři znaky z množiny $\{1, 2, 3, 4, 5\}$: 123, 124, 125, 134, 135, 145, 234, 235, 245, 345.

Zase si to zkuste oběma algoritmy.

△

11d.6 Generování všech permutací n různých objektů.

Zde se zaměříme na postupné generování. První krok je jasné, nejmenší permutace je $12\dots n$. Jasně je také to, kde se má podle lexikografického uspořádání skončit: s permutací $n\dots 21$. Kritická otázka je tato: Máme-li permutaci $a_1 a_2 \dots a_n$, jak vyrobíme následující? Je to první netriviální otázka této sekce a čtenář se nemusí cítit špatně, pokud na to sám nepřijde. Tvrdíme, že to udělá následující postup.

1) Hledejme dvojici po sobě jdoucích čísel, která rostou, tj. $a_i < a_{i+1}$. Pokud taková neexistuje, tak již máme největší permutaci $n\dots 21$ a algoritmus skončil.

2) Pokud taková dvojice existuje, tak najdeme největší index s touto vlastností, pak $a_i < a_{i+1} > a_{i+2} > \dots > a_n$. Pokud dáme na místo a_i něco většího a předchozí čísla zachováme, tak dostaneme větší permutaci. My ale chceme hned tu následující, budeme tedy zvyšovat co nejméně.

Určíme, která z čísel $a_{i+1}, a_{i+2}, \dots, a_n$ jsou větší než a_i (taková určitě jsou, třeba a_{i+1}), a mezi nimi najdeme nejmenší, nechť má index j . Při jeho hledání s úspěchem využijeme to, že po a_i už členy klesají, takže j je největší index takový, že $a_j > a_i$. Novou permutaci pak vyrobíme takto: členy $a_1 a_2 \dots a_{i-1}$ ponecháme, na i -té místo dáme a_j a zbývající čísla $a_{i+1}, a_{i+2}, \dots, a_n$ zařadíme tak, aby rostly, čili vlastně převrátíme jejich pořadí, a ještě tam na vhodné místo vsuneme a_i . Jinak řečeno, vyměníme mezi sebou čísla a_i a a_j a pak obrátíme pořadí čísel za i -tou pozicí.

Příklad 11d.e: Najdeme bezprostředního následníka k permutaci 68247531. Pak $i = 4$, protože 47 roste a od 7 dál už čísla klesají. Z čísel 7531 jsou čísla 7, 5 větší než 4, nejmenší mezi nimi je 5, tedy $j = 6$. Toto tedy přijde na místo čtyřky, naopak čtyřka přijde místo něj a čísla 7431 budou seřazena dle velikosti neboli naopak. Dostaneme proto permutaci 68251347.

Proč je lexikograficky větší než původní permutace je jasné, prvních $i-1$ míst je shodných a to i -té má nová permutace větší. Musíme ale ještě ukázat, že neexistuje nějaká permutace mezi nimi. Taková permutace by musela mít stejná první $i-1$ míst, čili od i -tého místa dál by používala stejná čísla jako obě naše permutace (původní i nová). To, že by naše hypotetická permutace byla lexikograficky mezi původní a novou, znamená dvě možnosti. Jedna je, že z čísel, která jsou k dispozici, by se muselo dát vybrat něco mezi $a_i = 4$ a $a_j = 5$, ale takové číslo tam není, jinak bychom jej zvolili při hledání j . Nebo by musela ta jiná permutace mít také na i -tém místě a_j , ale na místě $i+1$ něco menšího. To také nejde, protože jsme následnou část vyrobili jako rostoucí, tj. na místě $i+1$ je nejmenší možné číslo, které je k dispozici. Našli jsme tedy opravdu bezprostředního následníka k dané permutaci.

△

Algoritmus pro vytvoření následující permutace v lexikografickém uspořádání, když máme permutaci $a_1 a_2 \dots a_k$, která není maximální, tedy existuje i takové, že $a_i < a_{i+1}$:

```

procedure NextPermutation(a1a2...ak)
i := n - 1;
while ai > ai+1 do
    i := i - 1;
j := n;
while aj < ai do
    j := j - 1;
output: a1a2...ai-1ajanan-1...aj+1aiaj-1aj-2...ai+1;

```

Příklad 11d.f: Vygenerujeme všechny permutace řetězce 1234 naším algoritmem.

1234: Poslední rostoucí dvojice je 34, tedy $i = 3$. Mezi následujícími čísly 4 vybereme nejmenší z těch, které trojku převyšují, to je čtyřka, a dáme ji místo trojky: 124. Zbývající čísla (trojku) pak přilepíme dozadu dle velikosti: 1243.

1243: Poslední rostoucí dvojice je 24, tedy $i = 2$. Mezi následujícími čísly 43 vybereme nejmenší z těch, které $a_2 = 2$ převyšují, to je trojka, a dáme ji místo dvojky: 13, dvojku dáme zase na místo trojky. Tato zbývající čísla 42 pak přilepíme dozadu dle velikosti: 1324.

Čtenář si jistě rád rozmyslí další postup a dostane 1234, 1243, 1324, 1342, 1423, 1432, 2134, 2143, 2314, 2341, 2413, 2431, 3124, 3142, 3214, 3241, 3421, 4123, 4132, 4213, 4231, 4312, 4321.

△

11d.7 Generování všech řešení rovnice $x_1 + x_2 + \dots + x_k = n$ splňujících podmínky $x_i \in \mathbb{N}_0$ a $a_i \leq x_i \leq b_i$ pro všechna i , kde $a_i \leq b_i$ jsou parametry splňující $\sum a_i \leq n$ a $\sum b_i \geq n$.

Tyto dvě podmínky zaručí, že nějaká řešení budou existovat.

Když vypisujeme všechna řešení vnořenými cykly, tak budujeme vektory po jednotlivých souřadnicích. Představme si, že už jich i máme. Jaká jsou omezení pro volbu x_{i+1} ? Nesmíme vybrat příliš málo, aby to další x_j stačily dorovnat do n v rámci svých omezení shora, tedy musí platit $\sum_{j=1}^{i+1} x_j + \sum_{j=i+2}^k b_j \geq n$, na druhou stranu nesmíme vybrat moc, protože bychom tím nutili další proměnné být menší než povolené dolní meze, tedy musí platit $\sum_{j=1}^{i+1} x_j + \sum_{j=i+2}^k a_j \leq n$. Z těchto dvou nerovností získáme meze pro možné hodnoty x_{i+1} za předpokladu, že známe x_1, \dots, x_i :

$$n - \sum_{j=i+2}^k b_j - \sum_{j=1}^i x_j \leq x_{i+1} \leq n - \sum_{j=i+2}^k a_j - \sum_{j=1}^i x_j.$$

Zároveň ovšem máme meze ze zadání, kterým také musíme vyhovět.

Pokud je ovšem $i+1 = k$, tak žádná volba není a proces končí, musí být $x_k = n - \sum_{j=1}^{k-1} x_j$.

```

procedure AllSolutions1(k, n, aj, bj: integer)
mn := n - ∑j=2k bj; mx := n - ∑j=2k aj;
for x1 := max(a1, mn) to min(b1, mx)
    mn := mn - x1 + b2; mx := mx - x1 + a2;
    for x2 := max(a2, mn) to min(b2, mx)
        mn := mn - x2 + b3; mx := mx - x2 + a3;
        :
        for xk-1 := max(ak-1, mn) to min(bk-1, mx)
            xk := n - ∑j=1k-1 xj;
            output: (x1, x2, ..., xk);

```

teď se podíváme na komplikovanější úlohu postupného generování. Základní představa je, že máme n jednotek a ty budeme porůznu přelévat mezi jednotlivými proměnnými. Zhruba řečeno, čím více jedniček nalijeme co nejvíce doprava, tím menší výsledný vektor bude v lexikografickém uspořádání.

Jak vlastně vypadá minimální vektor (x_1, \dots, x_k) vzhledem k lexikografickému uspořádání? Musí existovat i takové, že pro $j < i$ platí $x_j = a_j$ a pro $j > i$ platí $x_j = b_j$. Proč? Pokud by byl nějaký jiný vektor (y_1, \dots, y_k) lexikograficky menší, pak by se případně shodoval na prvních několika souřadnicích s tímto (x_1, \dots, x_k) a pak je jedna souřadnice, řekněme y_m , která je menší. Evidentně to nemůže být jedna z těch, kde $x_j = a_j$, tam už jsme na minimu. To znamená, že je to buď souřadnice i nebo ještě některá za ní. Protože je celkový součet hodnot stále stejný, tak snížení na souřadnici m znamená, že se některá ze souřadnic y_j pro $j > m$ musel naopak oproti x_j

zvýšit, ale to je nemožné, protože souřadnice následující po x_i už jsou na svých maximálních hodnotách. Menší vektor tedy neexistuje.

Intuitivně tento minimální vektor vznikne tak, že nejprve z daných n jednotek odlijeme do všech souřadnic nutná minima, zbytek pak doléváme postupně odprava do plného. U souřadnice i skončíme. Podobně si rozmyslíme, že největší možný vektor získáme doplňováním odleva, tedy první souřadnice budou rovny b_j až po jistou $i - 1$, od $i + 1$ jsou zase všechny rovny a_j .

Tedě si tedy představme, že máme určité řešení, tedy vektor (x_1, x_2, \dots, x_k) , který není maximální. Větší řešení (v lexikografickém smyslu) vyrobíme tak, že trochu přelijeme doleva, ale snažíme se tuto změnu udělat co nejvíce napravo, abychom tak dostali bezprostředně následující vektor. Přílít se ovšem dá jen tam, kde x_j ještě nedosáhlo horní meze b_j , navíc pak ještě musíme zase někde ubrat, a to logicky napravo od místa, kde přiléváme, abychom si nezkazili začátek vektoru. Tam tedy začneme. Budeme prohlížet vektor odprava a hledat místo, kde se dá trochu ubrat, tedy místo, kde x_j ještě není a_j , a pak se podíváme dál doleva po místě, kde se dá přidat. Tak tam přidáme a pak musíme „vynulovat“ následující souřadnice, a to na nejmenší možné hodnoty podle podmínek odvozených na začátku této části.

Algoritmus pro vytvoření následujícího řešení v lexikografickém uspořádání, když máme řešení (x_1, x_2, \dots, x_k) , která není maximální:

```

procedure NextSolution1(k, n, aj, bj: integer, (x1, x2, ..., xk))
i := k;
while xi = ai do
    i := i - 1;
repeat i := i - 1 until xi < bi ;
xi := xi + 1;
m := n - ∑j=i+2k bj - ∑j=1i xi;
if i < k - 1 then
    for j := i + 1 to k - 1
        xj := max(m, aj);
        m := m - xj + bj+2;
xk := n - ∑j=1k-1 xi;
output: (x1, x2, ..., xk);

```

11d.8 Generování všech řešení rovnice $x_1 + x_2 + \dots + x_k = n$ splňujících podmínky $x_i \in \mathbb{N}_0$ a $x_1 \leq x_2 \leq \dots \leq x_k$, kde k je pevně zvoleno.

Zde nejsou individuální omezení na x_i , ale podobně jako v 11d.6 budou počáteční souřadnice omezeny tím, že ty po nich musí být alespoň stejně velké, přičemž součet nesmí růst. Co o nich tedy víme?

Prvních $k - 1$ souřadnic může klidně začínat na nule (pokud je ty předchozí neomezí jinak), protože poslední souřadnice není shora omezená a tudíž vždy dosáhneme na součet n . Máme ale následující omezení shora: Tvrdíme, že nejvyšší možná hodnota pro x_1 je $\lfloor \frac{n}{k} \rfloor$. Kdyby totiž x_1 byl byť jen o jedno větší, tak už by ostatní x_j musely také splňovat $x_j \geq \lfloor \frac{n}{k} \rfloor + 1$ a jejich součet by převýšil n . Naopak pokud budou všechna x_j rovna $\lfloor \frac{n}{k} \rfloor$, pak je jejich součet nejvýše n a je tedy možné případně x_k doplnit do n a dostat tak řešení.

Podobným postupem odvodíme, že maximální řešení splňuje i $x_2 = \lfloor \frac{n-x_1}{k-1} \rfloor$, $x_3 = \lfloor \frac{n-x_1-x_2}{k-2} \rfloor$ atd. až nakonec

$$x_k = n - \sum_{j=1}^{k-1} x_j.$$

```

procedure AllSolutions(k, n: integer)
for x1 := 0 to ⌊ n/k ⌋
    for x2 := x1 to ⌊ (n-x1)/(k-1) ⌋
        :
        for xk-1 := xk-2 to ⌊ (n-x1-x2-...-xk-2)/2 ⌋
            xk := n - ∑j=1k-1 xj;
            output: (x1, x2, ..., xk);

```

Nakonec se ještě podíváme na postupné generování. Zase rozléváme n jedniček, přičemž tentokráte hladiny při pohledu zleva doprava nesmí klesnout. Je jasné, jak vypadá lexikograficky nejmenší řešení: $(0, 0, \dots, 0, n)$. Největší

řešení (x_1, x_2, \dots, x_k) vypadá trochu komplikovaněji, pokud použijeme horní meze z předchozí úvahy, ale naštěstí se dá rozmyslet, jak opravdu vypadá.

Snažíme se nalít co nejvíce doleva, aniž bychom ale zleva doprava klesali, takže je v zásadě jasné, jak to dopadne. V ideálním případě (pokud k dělí n) by všechny souřadnice byly stejné, jmenovitě $\lfloor \frac{n}{k} \rfloor$, ale v typickém případě jich takových bude jen několik prvních a zbývající budou $\lfloor \frac{n}{k} \rfloor + 1$.

Ted' si ještě rozmyslíme, jak se od jednoho řešení dostat k dalšímu. Podobně jako v části 11d.7, budeme chtít trochu přelít zprava doleva. Abychom ale mohli někde přidat, budeme muset o něco více doprava zase ubrat, a to lze jen tam, kde nemají dvě po sobě jdoucí proměnné stejnou hodnotu. Naše pátrání proto začneme hledáním prvního schodu zprava. Jeho vyšší hodnota se bude o jedničku zmenšovat, takže nikde předtím nebude povoleno jít výš. Proto budeme moci přidávat jen tam, kde je současná hodnota ještě nižší, samozřejmě co nejvíce vpravo.

Algoritmus pro vytvoření následujícího řešení v lexikografickém uspořádání, když máme řešení (x_1, x_2, \dots, x_k) , která není maximální:

```

procedure NextSolution2(k, n: integer, (x1, x2, ..., xk))
i := k - 1;
while xi = xi+1 do
    i := i - 1;
m := ai+1 - 1;
while xi = m do
    i := i - 1;
xi := xi + 1;
if i < k - 1 then
    for j := i + 1 to k - 1
        xj := xi;
xk := n -  $\sum_{j=1}^{k-1} x_j$ ;
output: (x1, x2, ..., xk);

```

12. Grafy

Grafy jsou matematická struktura, která dokáže reprezentovat velké množství situací. Namátkou uvedeme různá dopravní spojení, hierarchie v organizacích, ekonomické analýzy (která společnost vlastní čí akcie, nedávno mimo-chodem vědci odvodili zajímavý graf, který ukazuje, že se největší světové společnosti víceméně vlastní navzájem), analýzy výrobního procesu, zachycení běhu programu a další miliony aplikací.

V této kapitole si pojem grafu představíme. Nejprve si zavedeme základní terminologii a ukážeme, proč je nutné uvažovat více druhů grafů. Pak se postupně podíváme na několik úloh, které hrají v teorii grafů velkou roli. Nepůjdeme příliš do hloubky, tato kapitola je jen stručný úvod a čtenář, kterého to zaujmě (či který to potřebuje), má k dispozici kvalitní literaturu.

Naši (většinou snadnou) práci ovšem zkomplikuje jedno zásadní dilemma. Již jsme v kapitolách narazili na to, že určité věci se dají dělat, nastavit či pojmenovat více způsoby a ne vždy se matematici shodnou na jednom. U teorie grafů je toto dovedeno hodně daleko, celá teorie se totiž dá vystavět dvěma populárními způsoby, každý z nich vede k velmi rozdílné základní terminologii.

Jeden přístup je méně abstraktní, možná trochu přímočařejší, díky čemuž je velice snadný pro začátečníka a mnohdy podává žádanou informaci přímo, bez komplikujících mezistupňů. Asi proto je velice oblíbený, zejména když dojde na základní učebnice či kurzy běžné úrovně. Pokud se ovšem člověk chce posunout k zajímavějším situacím, tak za to zaplatí.

Druhý přístup je abstraktnější a velice elegantní, jedním šmahem zvládne všechny druhy grafů, od nejjednodušších po exotické, takže z pohledu matematika-teoretika je určitě vhodnější, ale jednoduché věci se v něm dělají možná zbytečně složitě.

Protože se čtenář může setkat s oběma přístupy, představíme zde v úvodních sekcích obě varianty, u jednotlivých aplikací pak mezi nimi budeme přeskakovat podle toho, co je zrovna pohodlnější, občas ukážeme v akci obě, snad to nebude totální zmatek. Pro čtenáře to je nicméně užitečné, protože to je dobrý trénink matematického vyjadřování.

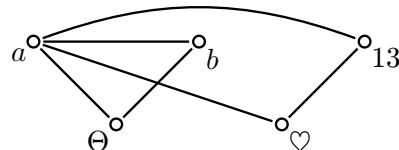
V této přehledové kapitolce se omezíme jen na konečné grafy, což je tradiční.

12a. Co jsou grafy (poprvé)

Ukážeme si typický příklad, jak od reálného problému dospět k matematické struktuře. Každý autoatlas má někde na začátku mapu ČR s vyznačenými hlavními silnicemi. Pokud jsme v situaci, kdy plánujeme, kudy se dostat z bodu A do bodu B , tak je nám většinou jedno, jakou krajinou projízdíme, tuto informaci je tedy z mapy možné vynechat. Také nás nezajímá, jak vlastně města vypadají, nahradíme je kolečky. V této fázi nás dokonce ani nezajímá přesný tvar silnic, takže lze města spojovat úsečkami. Dostaneme hromádku koleček pospojovanou porůznu úsečkami, což je obrázek, který také v mnohých autoatlasech bývá. Zároveň je to nejtypičtější představitel matematického grafu.

Jak bychom toto zachytily matematicky? Evidentně pracujeme s dvěma entitami, máme objekty (města) a vztahy mezi nimi (spojeno či nespojeno). Na zachycení objektů máme nástroj, jmenovitě množinu. Množina uvažovaných objektů tedy tvoří základ grafu, v teorii grafů jim říkáme „vrcholy“ či „uzly“. Vztahy pak zachytíme způsobem, který odkoukáme od relací. Založíme si množinu „hran“ (spojnice v atlase), tedy pokud jsou nějaká města přímo spojena, tak z nich uděláme dvouprvkovou množinu a schováme do té množiny hran. Tím je graf popsán.

Příklad 12a.a: Uvažujme následující obrázek.



Základem grafu jsou vrcholy $V = \{a, b, 13, \Theta, \heartsuit\}$.

Hrany jsou $E = \{\{a, b\}, \{a, 13\}, \{a, \Theta\}, \{a, \heartsuit\}, \{b, 13\}, \{b, \Theta\}, \{b, \heartsuit\}, \{13, \heartsuit\}\}$.

Spojením těchto dvou údajů dostáváme matematický popis grafu (V, E) .

Všimněte si, že z tohoto matematického zápisu vůbec nelze odvodit, jak byl náš graf uspořádán na papíře. To je zcela v pořádku. V teorii grafů se (až na pár výjimek) vůbec nezabýváme „fyzikální“ podstatou situace, soustředíme se čistě na strukturu. Pokud si někdo uspořádá vrcholy na papír jinak, dostane po doplnění příslušných hran jiný obrázek, což nevadí, z hlediska teorie grafů je to totožný graf. Obrázky jsou jenom pomůckou pro přemýšlení o grafu, je na nás, jak si je nakreslíme.

△

!

Definice.

Pod pojmem **graf (graph)** rozumíme libovolnou uspořádanou dvojici $G = (V, E)$, kde V je nějaká konečná množina a E je libovolná podmnožina množiny $\{A \subseteq V; |A| = 2\}$.

Prvkům $v \in V$ se říká **vrcholy** grafu G (**vertex**), prvkům $\{u, v\} \in E$ říkáme **hrany** grafu G (**edge**).

Výhodou tohoto způsobu zápisu je jeho přehlednost. U hran hned vidíme, odkud kam jdou, třeba hrana $\{c, f\}$ spojuje vrcholy c, f , odborně říkáme, že vrchol c či vrchol f je **incidentní** s touto hranou, vzniká tak vztah **incidence**.

Někteří autoři zavádějí značení $\binom{V}{2}$ pro množinu všech dvouprvkových podmnožin množiny V , ale není to moc rozšířené, tak se tomu zde vyhneme. My si z této množiny můžeme vybírat hrany libovolně, extrémy jsou zjevné. Jeden je, že se nevybere nic, pak je $E = \emptyset$, grafy bez hran jsou ovšem nudné.

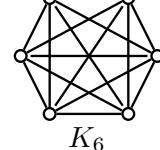
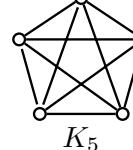
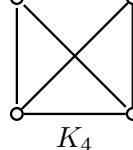
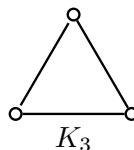
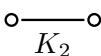
Zajímavější je druhý extrém, kdy jsou v E všechny dvouprvkové podmnožiny, tedy v grafu jsou všechny vrcholy propojeny. Víme, že v případě grafu o n vrcholech je pak počet hran roven $|E| = \binom{n}{2} = \frac{1}{2}n(n-1)$. Takové grafy mají své jméno.

Definice.

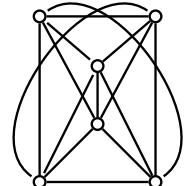
Graf $G = (V, E)$ se nazývá **úplný (complete)**, jestliže $E = \{\{u, v\}; u, v \in V \wedge u \neq v\}$.

Úplný graf o n vrcholech se značí K_n .

Úplné grafy se v teorii občas zajímavě objeví, čtenář si je asi umí představit, ale protože jsou pěkné, neodpustíme si pár obrázků.

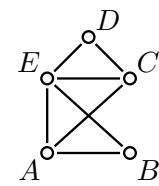
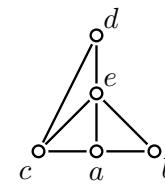
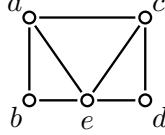


Už jsme zmínili, že jsme ty grafy mohli klidně nakreslit v jiném tvaru a bylo by to také dobré. Pokud se ale pro graf nabízí nějaké pěkně vypadající uspořádání, tak lidé obvykle neodolají. Abychom teď ukázali nezávislost ducha a odvahu nejít s davem, nabídnete ještě jinou podobu K_6 (ověřte na obrázku vpravo, že všech šest vrcholů je navzájem propojeno hranami). V kapitolce 12d se také objeví alternativní verze K_4 .



Teď malá motivace.

Příklad 12a.b: Uvažujme následující grafy:



Čtenář si hradě rozmyslí, že když si tyto grafy zapíšeme matematicky, dostaneme v prvních třech případech vždy totéž: $(V = \{a, b, c, d, e\}, E = \{\{a, b\}, \{a, c\}, \{a, e\}, \{b, e\}, \{c, d\}, \{c, e\}, \{d, e\}\})$. Jde tedy o tentýž graf, jen různě nakreslený.

Zajímavý je čtvrtý graf. Jeho matematický zápis je

$(V' = \{A, B, C, D, E\}, E' = \{\{A, B\}, \{A, C\}, \{A, E\}, \{B, E\}, \{C, D\}, \{C, E\}, \{D, E\}\})$.

Stačí změnit velká písmena na malá a dostáváme zase totéž jako u prvních třech grafů. Dá se proto čekat, že se tento nový graf bude vždy chovat stejně jak první graf. Je to tedy vůbec jiný graf? To je dobrá otázka, zejména když čtenáři prozradíme, že jsme původně jen nakreslili čtyři podoby stejněho grafu a pak u jedné nechali editor přeměnit všechna písmenka na velká.

△

Podíváme se na to blíže. Pro zjednodušení budeme uvažovat trochu menší grafy, G_1 s vrcholy $\{a, b, c\}$ a hranou $\{a, b\}$, dále G_2 s vrcholy $\{1, 2, 3\}$ a hranou $\{2, 3\}$. Je zjevné, že oba grafy lze znázornit přesně stejným obrázkem (až na popisky u vrcholů), je to tedy z praktického hlediska tentýž graf. Proto je poněkud nemilé, že z matematického hlediska jsou to dva zcela rozdílné objekty, protože množiny $V_1 = \{a, b, c\}$ a $V_2 = \{1, 2, 3\}$ prostě nejsou stejné, to neobkrcáme.

Pro situace, kdy máme dvě struktury, které jsou ve skutečnosti vlastně zcela stejné, jen se jinak tváří, se v matematice používá pojem „isomorfismus“. Jak vlastně poznáme, že jsou grafy G_1 a G_2 stejné? Uvidíme to, pokud dokážeme v grafu G_1 nahradit jména vrcholů tak, že po odpovídající změně u hran již dostaváme graf G_2 . Evidentně to funguje i zpětně, je to symetrický vztah mezi grafy. U našich dvou grafů zabere například přejmenování $a \mapsto 2, b \mapsto 3, c \mapsto 1$, fungovalo by i přejmenování $a \mapsto 3, b \mapsto 2, c \mapsto 1$. Vidíme, že vlastně mluvíme o bijekci mezi vrcholy, u které je dále požadavek, aby správně přenesla i hrany.

Definice.

Nechť $G_1 = (V_1, E_1)$ a $G_2 = (V_2, E_2)$ jsou grafy. Řekneme, že jsou navzájem **isomorfní**, jestliže existuje bijekce $T: V_1 \rightarrow V_2$ taková, že

$$\{\{T(u), T(v)\}; \{u, v\} \in E_1\} = E_2.$$

Takovému zobrazení (pokud existuje) říkáme **isomorfismus** grafů.

Neformálně v takovém případě říkáme, že graf G_2 je kopí grafu G_1 , popřípadě naopak.

Je zřejmé, že pokud jsou dva grafy navzájem kopiemi, tak se také musí shodovat ve všech vlastnostech, je to v podstatě porád jeden graf. S tímto porozuměním se pak věci občas dosti zjednoduší.

My už jsme na to ostatně narazili. Podle definice je každý graf o pěti vrcholech a 10 hranách nazvaný K_5 . My si ale můžeme pro vrcholy vymyslet nekonečně mnoho různých značení, takže bráno formálně se spousta různých grafů jmenuje K_5 . To vypadá na pěkný zmatek, ale teď už víme, že všechny tyto grafy jsou jen kopiemi třeba toho pěkného obrázku výše, takže je to vlastně jen jeden objekt a tomu jsme dali jméno. Praktický dopad celého zamýšlení je, že v následujícím textu si občas ušetříme formální komplikace.

Nyní si zavedeme několik základních pojmu. Důležitou věcí u grafů jsou počty různých věcí. Celkový počet vrcholů a hran vidíme hned z mohutnosti V a E , často se také hodí vědět, kolik hran je u zkoumaného vrcholu.

Definice.

Nechť $G = (V, E)$ je graf. Pro vrcholy $v \in V$ definujeme **stupeň vrcholu (degree)** jako

$$\deg(v) = |\{x \in V; \{v, x\} \in E\}|.$$

Neformálně, stupeň vrcholu je počet hran, se kterými má dotyčný vrchol něco do činění. V podmínce množiny jsme mohli psát i $y = v$, v hraně coby množině na pořadí prvků nezáleží. Některí autoři mají alternativní definici

$$\deg(v) = |\{y \in V; \{v, y\} \in E\}|,$$

která počítá vrcholy, do kterých se dá z v jedním krokem dojít, což je evidentně totéž jako naše definice.

Pokud má graf n vrcholů, pak pro všechny vrcholy určitě platí $\deg(v) \leq n - 1$. Snadno si rozmyslíme, že v úplném grafu K_n musí mít každý vrchol maximální možný stupeň $n - 1$, naopak podle toho se takový úplný graf pozná, jsou to právě grafy, kde všechny vrcholy splňují $\deg(v) = |V| - 1$.

Teď něco obecnějšího, ale stále snadného. Každá hrana přispěje dvěma vrcholům jedničkou ke stupni, takže dostaváme následující rovnost.

Fakt 12a.1.

Nechť $G = (V, E)$ je konečný graf. Pak $\sum_{v \in V} \deg(v) = 2|E|$.

Někdy se tomu také říká Princip sudosti, protože z toho vyplývá, že součet stupňů v grafu je sudý, z čehož dále ještě vyplývá, že počet vrcholů s lichým stupněm musí být sudý.

U většiny pojmu z této knihy se nám hodilo mít možnost přejít k menší části zkoumaného objektu (restrikce zobrazení, podrelace, podgrupa). Užitečné je to i u grafů.

Definice.

Nechť $G = (V, E)$ je graf. Dvojice $G' = (V', E')$ je **podgraf (subgraph)** grafu G , jestliže $V' \subseteq V, E' \subseteq E$ a (V', E') je graf.

Poslední podmínka je zásadní, podgrafy dostaneme tak, že si z grafu vybereme nějaké vrcholy a hrany, ale musíme přitom dbát na to, abychom tak nestvořili nějakou příšerku, kde by třeba hrana trčela do nikam jako v případě, že vybereme vrcholy a, b a k nim omylem hrani $\{a, c\}$. Pro případ podgrafu není třeba chodit daleko,

pro $n \leq m$ je K_n podgrafem K_m . Každý graf je podgrafem sebe sama a prázdný graf ($\{\}, \{\}$) je podgrafem všech grafů.

Podgrafy často vznikají konkrétními postupy. Jeden je, že zjistíme, že nás na grafu vlastně nějaké vrcholy vůbec nezajímají, tak je vyřadíme a spolu s nimi přirozeně i ty hrany, které do vyřazených vrcholů šly, ostatní hrany zůstávají. Nahlíženo z opačného konce, z původního grafu si vybereme jen nějaké zajímavé vrcholy a spojíme je hranami všude, kde tomu tak bylo v původním grafu.

Definice.

Nechť $G = (V, E)$ je graf a $M \subseteq V$. Pak definujeme příslušný **podgraf indukovaný** množinou H (**induced subgraph**) grafu G jako dvojici $G' = (M, E')$, kde $E' = \{\{u, v\} \in E; u, v \in M\}$.

Aby byla definice korektní, tak je ještě třeba dokázat, že popsaným vzorcem opravdu vzniká graf, ale to je snadné, již z definice obsahuje E' pouze dvouprvkové podmnožiny M .

Podgrafy také často vznikají postupně zmenšováním původního grafu. Je zjevné, že odebráním hrany zase vznikne graf, ale pokud chceme odebrat vrchol, musíme spolu s ním odebrat všechny hrany, které jsou s ním spojeny, aby tak vznikl graf.

Definice.

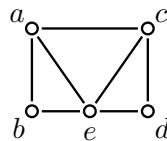
Nechť $G = (V, E)$ je graf.

Řekneme, že graf $G' = (V', E')$ vznikl **odebráním hrany** $e = \{u, v\}$, jestliže $V' = V$ a $E' = E - \{e\}$.

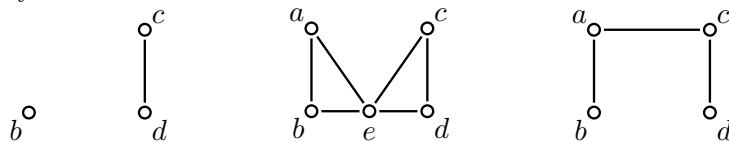
Řekneme, že graf $G' = (V', E')$ vznikl **odebráním vrcholu** v , jestliže $V' = V - \{v\}$ a $E' = E - \{\{v, u\}; u \in V\}$.

Není těžké si rozmyslet, že každý podgraf daného grafu lze z dotyčného grafu vyrobit postupným odebíranám vrcholů a hran, pojednat podgrafu se tak dá definovat.

Příklad 12a.c: Připomeňme si graf $(\{a, b, c, d, e\}, \{\{a, b\}, \{a, c\}, \{a, e\}, \{b, e\}, \{c, d\}, \{c, e\}, \{d, e\}\})$, v příkladě 12a.b jsme pro něj měli třeba totto znázornění:



Následující obrázky ukážou podgraf indukovaný množinou vrcholů $M = \{b, c, d\}$, podgraf vzniklý odebráním hrany $\{a, c\}$ a podgraf vzniklý odebráním vrcholu e .



△

Některé podgrafové jsou speciální.

Definice.

Nechť $G = (V, E)$ je graf. Řekneme, že je to **kružnice** (**circuit**), jestliže lze vrcholy označit jako $V = \{v_1, \dots, v_n\}$ pro $n \geq 3$ tak, aby pak množina hran byla

$$E = \{\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_{n-2}, v_{n-1}\}, \{v_{n-1}, v_n\}, \{v_n, v_1\}\}.$$

Je zřejmé, že v kružnici má každý vrchol stupeň 2, ale nefunguje to naopak. Pokud například sestavíme graf ze dvou disjunktních kružnic, tak v něm bude mít každý vrchol stupeň 2, ale jako celek to kružnice není. Můžeme říct, že graf je kružnicí, pokud je pro nějaké $N \in \mathbb{N}$ kopí grafu s vrcholy $V = \{1, 2, \dots, N\}$ a hranami $\{\{1, 2\}, \{2, 3\}, \dots, \{N-1, N\}, \{N, 1\}\}$. V některých aplikacích hraje zásadní roli, zda se v grafu vyskytují nějaké kružnice coby podgrafové. Pokud ano, tak se takovému grafu říká **cyklický**.

Například K_3 je kružnice, ale K_4 kružnice není. Několik kružnic ale obsahuje, například svůj „obvod“, a tudíž jde o graf cyklický. Příkladem grafu, který cyklický není, může být třeba „činka“ K_2 , blíže na to narázíme v kapitole 12f. Algoritmicky je hledání kružnic v grafu snadná úloha s lineární náročností vzhledem k počtu vrcholů.

Protože se kružnice zřídka používá jako samostatný graf, častěji ji hledáme jako podgraf uvnitř jiného, je mnohem přirozenější ji zavést jako speciální druh sledu, viz kapitola 12c.

Další populární sport je hledat, zda daný graf obsahuje nějaký úplný graf jako svůj podgraf. To je naopak úloha vysoce drsná, jde o NP-úplný problém. Na to narazíme pro změnu v kapitole 12d.

Probrali jsme základní pojmy, bylo to v celku snadné a je čas se zamyslet, jaké struktury vlastně představeným typem grafu dokážeme popsat.

Protože víme, že u množiny na pořadí prvků nezáleží, tak je hned jasné, že není možné u dotyčné hrany udělat jeden z konců nějak speciální, oba jsou rovnocenné. Jinými slovy, není možné pomocí naší definice na takové hraně vytvořit privilegovaný směr. Proto také tomuto typu grafů říkáme **neorientované grafy (undirected graph)**.

Množiny nám také neumožňují skladovat více kopií téhož prvku, což z pohledu konkrétní hrany znamená, že nelze zopakovat tentýž vrchol, jinými slovy nelze v takovém grafu mít **smyčky** (loop) typu $\{u, u\}$. Z pohledu E to pak znamená, že mezi dvěma vrcholy může být nejvýše jedna hrana. Pracovali jsme tedy s grafy, které jsou neorientované a které nemohou mít smyčky a více spojnic mezi dvěma vrcholy. Když chceme zdůraznit, že pracujeme právě s takovýmto grafem, tak řekneme, že jde o **jednoduchý graf (simple graph)**.

Pokud aplikace vyžaduje některý ze „zakázaných“ prvků, je nutné definici grafu modifikovat, což je někdy zjevné a někdy komplikovanější. Většina pojmu a poznatků se pak přenáší i na nové druhy grafů.

Pokud chceme zahrnout smyčky, pak je třeba v definici hran vyžadovat, aby se E skládalo z jedno- a dvou prvkových podmnožin množiny V . Výsledné strukturu říkáme **pseudograf**, je pak nutné upravit i další definice a při práci je nutné být pořád na pozoru, zda je právě zpracovávaná hrana jednoprvková nebo dvouprkvová. Jsou to nudné komplikace a pseudograffy se při zde představeném zápisu grafů moc nezkoumají.

Pokud chceme nějak zahrnout situaci, kdy mezi dvěma vrcholy vede více hran (což třeba v autoatlase opravdu nastává), pak je třeba definovat tzv. **multigraf** a jsou v zásadě dvě možnosti. Jeden přístup je namísto množin pracovat s pojmem „multimnožina“. To je jako množina, ale umožňuje vícenásobný opakovaný výskyt prvků. Lze pak v definici multigrafu říct, že E je multimnožina dvouprkvových podmnožin V . Teorie multimnožin nabízí nástroje na práci s multimnožinami, díky čemuž lze zase zavést pojmy jako stupeň vrcholu a podobně, ukáže se, že multigraffy fungují v zásadě stejně jako grafy jednoduché. Zajímavou alternativou je kódovat vícenásobné výskytty hran pomocí třetí souřadnice (index hrany), ale i to dá práci. S grafy s vícenásobnými hranami se mnohem lépe vypořádáme, pokud použijeme ten konkurenční přístup z příští podkapitoly.

Další dvě varianty jsou natolik významné, že si uděláme nadpis.

12a.2 Grafy orientované a ohodnocené

Naše povídání jsme začali autoatlasem a tam jsou jednosměrky. Obecně řečeno, v mnoha aplikacích potřebujeme u hran jejím koncům přiřadit priority, v jednom vrcholu hrana začíná, v druhém končí. To je vysoko užitečný typ grafu a příslušná definice je velmi snadná.

! Definice.

Pod pojmem **orientovaný graf (directed graph)** rozumíme libovolnou uspořádanou dvojici $G = (V, E)$, kde V je nějaká konečná množina a E je libovolná podmnožina $V \times V$.

Pojem orientovaného grafu je z mnoha pohledů flexibilnější, například nám rovnou umožňuje mít smyčky. Většina pojmu definovaných výše pro neorientované grafy má přirozené verze pro orientované grafy. Některé definice lze převést přímo (třeba pojem podgrafa), někde je třeba udělat modifikaci.

Definice.

Nechť $G = (V, E)$ je orientovaný graf. Pro vrcholy $v \in V$ definujeme

výstupní stupeň (outgoing degree, outdegree) jako $\deg^+(v) = |\{(x, y) \in E; x = v\}|$,

vstupní stupeň (incoming degree, outdegree) jako $\deg^-(v) = |\{(x, y) \in E; y = v\}|$,

stupeň (degree) jako $\deg(v) = \deg^+(v) + \deg^-(v)$.

I zde máme alternativní zápis $\deg^+(v) = |\{x \in V; (v, x) \in E\}|$ a $\deg^-(v) = |\{x \in V; (x, v) \in E\}|$. Pořád platí, že $\sum_{v \in V} \deg(v) = 2|E|$.

Při pohybu v orientovaném grafu je důležité, zda respektujeme orientaci hran či ne. Odráží se to i v terminologii, například lze v orientovaném grafu hledat kružnice, které jsou neorientované, nebo obdobný útvar, ale již s orientací:

Definice.

Orientovaný graf $G = (V, E)$ je **cyklus (cycle)**, jestliže lze vrcholy označit $V = \{v_1, \dots, v_n\}$ pro $n \geq 2$ tak, aby pak

$$E = \{(v_1, v_2), (v_2, v_3), \dots, (v_{n-2}, v_{n-1}), (v_{n-1}, v_n), (v_n, v_1)\}.$$

I u orientovaných grafů nás může zajímat otázka, zda je v daném grafu podgraf, který je cyklus, zkráceně řečeno zda graf obsahuje cykly, pak mu říkáme **cyklický**. Někdy se ovšem pohybujeme v grafu orientovaném a nehledíme přitom na orientaci hran, takže lze mluvit i o kružnici v orientovaném grafu.

V některých aplikacích pak orientovanost vlastně nepotřebujeme, v takovém případě se vyplácí „odmazat“ od hran orientaci a uvažovat takto vzniklý neorientovaný graf.

Definice.

Nechť $G = (V, E)$ je orientovaný graf. Definujeme jeho **symetrizaci** jako neorientovaný graf $G' = (V, E')$, kde

$$E' = \{\{u, v\}; [(u, v) \in E \vee (v, u) \in E] \wedge u \neq v\}.$$

Protože v neorientovaném grafu jsou povoleny jen hrany s rozdílnými konci, museli jsme to v definici ošetřit a smyčky vyloučit. To jsou právě ty komplikace, které s sebou přináší tento jednodušší přístup k teorii grafů.

Při plánování cest v autoatlase nás ale nezajímá jen odkud a kam cesty vedou, ale také jak dlouho nám jednotlivé úseky trvají. Každá hrana (cesta) má své „ohodnocení“ (délku či čas, podle naší preference) a tyto údaje jsou klíčové při naší analýze.

U některých úloh má naopak smysl přiřazovat hodnoty k vrcholům grafu, například pokud graf zachycuje počítacovou síť, je u jednotlivých uzlů velice podstatná jejich propustnost, při návrhu tras pro náš balíček bitů určitě nemá smysl jej posílat skrz uzel, který je již zahlcen.

Definice.

Nechť $G = (V, E)$ je graf (neorientovaný i orientovaný). Řekneme, že je to **ohodnocený graf (weighted graph)**, jestliže k němu existuje **ohodnocovací funkce** $w : E \mapsto \mathbb{R}$, popřípadě $w : V \mapsto \mathbb{R}$.

Někdy dokonce máme u grafu ohodnocovací funkce dvě, jednu pro hrany a jednu pro vrcholy. Ohodnocené grafy jsou další významnou skupinou grafů a řešíme na nich některé zásadní úlohy, viz 12c.

12a.3 Reprezentace grafů

Uvažujme graf, tedy množinu vrcholů a množinu hran, případně nějaká ohodnocení. Pro mnohé účely bývá výhodnější si takový graf reprezentovat jinak než jazykem množin z definice.

S jednou reprezentací jsme se již setkali, graf si můžeme nakreslit, k tomuto tématu se ještě vrátíme v části 12d. Je také možné grafy zadávat různě strukturovanými seznamy, například seznamem vrcholů a ke každému zadat seznam hran s ním incidentních. Další významnou možností je zachycení grafu (konečného) pomocí matice. Je několik oblíbených způsobů.

Uvažujme tedy graf $G = (V, E)$ o n vrcholech, můžeme předpokládat, že vrcholy jsou označeny $V = \{v_1, \dots, v_n\}$ a hrany $E = \{e_1, \dots, e_m\}$.

- **Matrice sousednosti (adjacency matrix)** je čtvercová 01 -matica $n \times n$ definovaná předpisem $a_{ij} = 1$ jestliže $\{v_i, v_j\} \in E$, popřípadě (pro orientované grafy) jestliže $(v_i, v_j) \in E$, jinak $a_{ij} = 0$. Tradičně se tato matice grafu G značí A či A_G .

Je zjevné, že pro neorientované grafy dostáváme symetrickou matici. V takovém případě pro ušetření paměti někdy pracujeme jen s horní či dolní trojúhelníkovou maticí. Touto maticí snadno zachytíme i multigrafy, když definujeme a_{ij} jako počet hran spojujících v_i s v_j .

- **Laplaceova matice** je čtvercová matice $n \times n$ definovaná předpisem $a_{ij} = \begin{cases} \deg(v_i), & i = j; \\ 1, & i \neq j \wedge \{v_i, v_j\} \in E; \\ 0, & i \neq j \wedge \{v_i, v_j\} \notin E, \end{cases}$ popřípadě (pro orientované grafy) $a_{ij} = \begin{cases} \deg(v_i), & i = j; \\ 1, & i \neq j \wedge (v_i, v_j) \in E; \\ 0, & i \neq j \wedge (v_i, v_j) \notin E. \end{cases}$ Tradičně se značí L_G .

- **Matrice incidence** je matice typu $n \times m$ definovaná předpisem $a_{ij} = \begin{cases} 1, & \exists u \in V : e_j = \{u, v_i\} \in E; \\ 0, & \text{jinak,} \end{cases}$

popřípadě (pro orientované grafy) $a_{ij} = \begin{cases} -1, & \exists u \in V : e_j = (v_i, u); \\ 1, & \exists u \in V : e_j = (u, v_i) \in E; \\ 0, & \text{jinak.} \end{cases}$

Slovy, když se podíváme na sloupec matice příslušný hraně e_j , tak vidíme jedničky v řádcích pro vrcholy, kde tato hrana začíná a končí (u neorientovaných grafů). U orientovaných grafů vidíme -1 u vrcholu, kde hrana začíná, a 1 u vrcholu, kde končí.

Matrice incidence grafu G se tradičně značí I_G .

- **Matrice délek** pro grafy s ohodnocenými hranami je čtvercová matice $n \times n$ definovaná předpisem $a_{ij} = \begin{cases} w(\{v_i, v_j\}), & \{v_i, v_j\} \in E; \\ 0, & \text{jinak,} \end{cases}$ popřípadě (pro orientované grafy) $a_{ij} = \begin{cases} w((v_i, v_j)), & (v_i, v_j) \in E; \\ 0, & \text{jinak.} \end{cases}$

Zajímavá modifikace je, když pro dvojice nespojené hranou dáme namísto nuly hodnotu $a_{ij} = \infty$. Tato volba totiž skvěle ladí s některými klíčovými algoritmy.

Každá z těchto matic má v určitých aplikacích výhody a v jiných nevýhody.

12b. Co jsou grafy podruhé

Zde se podíváme na druhý způsob, jak definovat grafy. Vychází v zásadě také z autoatlasní inspirace, významnější silnice totiž mívají svá jména. Je tedy možné si hrany pamatovat jako samostatné objekty, přičemž u každé ještě musíme mít někde schovánu informaci, s kterými vrcholy je incidentní.

!

Definice.

Pod pojmem **neorientovaný graf** rozumíme libovolnou uspořádanou trojici $G = (V, E, \varepsilon)$, kde V je nějaká konečná množina (vrcholy), E je libovolná konečná množina disjunktní s V (hrany) a ε je libovolné zobrazení z E do množiny jedno- a dvouprvkových podmnožin V (vztah indicence).

Pod pojmem **orientovaný graf** rozumíme libovolnou uspořádanou trojici $G = (V, E, \varepsilon)$, kde V je nějaká konečná množina (vrcholy), E je libovolná konečná množina disjunktní s V (hrany) a ε je libovolné zobrazení $E \mapsto V \times V$ (vztah indicence).

Jak poznáme, že je nějaký vrchol v incidentní s jistou hranou $e \in E$? U grafu neorientovaného testujeme podmítku $v \in \varepsilon(e)$. U grafu orientovaného je to jemně složitější, tak se ptáme na pravdivost následujícího výroku:

Existuje $w \in V$ takové, že $\varepsilon(e) = (v, w)$ nebo $\varepsilon(e) = (w, v)$.

Potvrzuje se to, o čem jsme psali v úvodu, tento druhý způsob zavedení grafu dělá jednoduché věci složitěji než přístup první. Na druhou stranu jím lze bezproblémově zpracovat situace, které dělají prvnímu přístupu problémy. Například hned vidíme, že definice umožňuje smyčky a vícenásobná spojení mezi vrcholy. Pokud dvě hrany e_1, e_2 spojují stejné vrcholy, tedy $\varepsilon(e_1) = \varepsilon(e_2)$, pak řekneme, že tyto hrany jsou **paralelní**.

U orientovaných grafů si ulehčíme práci zavedením pomocných funkcí incidence, které nám umožní přímý přístup k počátečnímu a koncovému vrcholu hrany. Jestliže je ε zobrazení incidence z definice orientovaného grafu, pak definujeme zobrazení PV a KV z E do V takto:

- $PV(e) = v$ právě tehdy, když existuje $w \in V$ splňující $\varepsilon(e) = (v, w)$.
- $KV(e) = v$ právě tehdy, když existuje $w \in V$ splňující $\varepsilon(e) = (w, v)$.

Ukažme si, jak se definice z prvního přístupu přepíší do nového jazyka.

Definice.

Nechť $G = (V, E, \varepsilon)$ je neorientovaný graf. Pro vrcholy $v \in V$ definujeme stupeň vrcholu jako

$$\deg(v) = |\{e \in E; v \in e\}|.$$

Nechť $G = (V, E, \varepsilon)$ je orientovaný graf. Pro vrcholy $v \in V$ definujeme

$$\deg^+(v) = |\{e \in E; v = PV(e)\}|,$$

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

Zde šlo o minimální úpravy.

Definice.

Nechť $G = (V, E, \varepsilon)$ je (orientovaný či neorientovaný) graf. Graf $G' = (V', E', \varepsilon')$ je podgraf grafu G , jestliže $V' \subseteq V$, $E' \subseteq E$ a ε' je restrikce zobrazení ε na množinu E' .

Obvykle se pro restrikci zvláštní značení nezavádí, protože na E' dává stejné hodnoty jako původní zobrazení, i zde tedy budeme mluvit o podgrafu (G', E', ε) .

Definice.

Nechť $G_1 = (V_1, E_1, \varepsilon_1)$ a $G_2 = (V_2, E_2, \varepsilon_2)$ jsou grafy. Řekneme, že jsou navzájem **isomorfní**, jestliže existují bijekce $T: V_1 \mapsto V_2$ a $t: E_1 \mapsto E_2$ takové, že pro každou hranu $e \in E_1$ platí

- pro orientované grafy: $\varepsilon_2(t(e)) = (T(PV(e)), T(KV(e)))$;
- pro neorientované grafy: $\varepsilon_2(t(e)) = T[\varepsilon(e)]$.

Zde $T[M]$ je obraz množiny, tedy pro dvojici $\{u, v\}$ je $T[\{u, v\}] = \{T(u), T(v)\}$. Pojem isomorfismu se v nové, abstraktnější definici grafů zdá o něco méně průhledný.

Zkusme si v tom udělat trochu pořádek. Příznivci prvního přístupu říkají „graf“ a myslí tím to, co lze dotyčným přístupem udělat snadno, tedy graf bez smyček a bez paralelních hran. Případu s více hranami říkají „multigraf“.

Příznivci druhého přístupu říkají „graf“ a myslí tím onu obecnou definici výše, která umožňuje všechno. Pokud chtějí graf bez paralelních hran, řeknou „prostý graf“. Prostý graf bez smyček je pro ně „jednoduchý graf“.

Z toho je jasné, že pojem „graf“ samotný je ambivalentní a je třeba se vždy podívat, co tím dotyčný myslí. Dobrá zpráva je, že když řekneme „jednoduchý graf“ či „prostý graf“, tak nám budou rozumět oba tábory, rovněž pojem „multigraf“ je všeobecně srozumitelný, ale příznivci druhého přístupu jej nevidí rádi, protože jim přijde zbytečný.

My se teď podíváme na několik důležitých myšlenek z teorie grafů a budeme mezi těmito dvěma přístupy udatně kličkovat.

12c. Procházení grafem

Mnoho úloh se týká procházení grafem a potřebujeme pojmy, které nějakou konkrétní trasu po grafu zachytí. Začneme abstraktnějším zápisem dle 12a.

!

Definice.

Nechť $G = (V, E, e)$ je graf, uvažujme vrcholy $v_0, v_1, v_2, \dots, v_N \in V$ a hrany (e_1, e_2, \dots, e_N) .

1) Předpokládejme, že G je neorientovaný.

Řekneme, že posloupnost $(v_0, e_1, v_1, e_2, v_2, \dots, e_N, v_N)$ je **sled** (**walk**) z v_0 do v_N , jestliže pro každé $i = 1, \dots, N$ platí, že $v_{i-1} \in e_i$ a $v_i \in e_i$.

2) Předpokládejme, že G je orientovaný.

Řekneme, že posloupnost $(v_0, e_1, v_1, e_2, v_2, \dots, e_N, v_N)$ je **sled** (**walk**) z v_0 do v_N , jestliže pro každé $i = 1, \dots, N$ platí, že $v_{i-1} = PV(e_i)$ a $v_i = KV(e_i)$.

Řekneme, že posloupnost $(v_0, e_1, v_1, e_2, v_2, \dots, e_N, v_N)$ je **neorientovaný sled** (**undirected walk**) z v_0 do v_N , jestliže pro každé $i = 1, \dots, N$ platí, že $v_{i-1} \in \{PV(e_i), KV(e_i)\}$ a $v_i \in \{PV(e_i), KV(e_i)\}$.

3) Řekneme, že sled $(v_0, v_1, v_2, \dots, v_N)$ je **uzavřený** (**closed**), jestliže $v_0 = v_N$. Jinak řekneme, že je **otevřený** (**open**).

Číslo N udává **délku** (**length**) sledu.

Některí autoři berou jako délku počet vrcholů, tedy $N + 1$. Není v tom shoda, my jsme zvolili variantu, která je rozšířená a pohodlná, například se dobře vypořádá s napojováním sledů.

Formálně je tedy sled posloupnost vrcholů, které jsme navštívili, mezi nimi pak informace o hranách, které jsme při přesunu použili. Ne vždy jsou sledy udávány přesně takto. Je například zjevné, že pokud jde o orientovaný graf, tak už je u každé hraně jasné, odkud kam vede, je tedy zbytečné vrcholy zadávat. Některí autoři pak tedy používají jen zápis (e_1, e_2, \dots, e_N) . Toto bude často fungovat i u grafů neorientovaných, ale tam už je třeba být trochu opatrnejší.

Naopak pokud se v grafu nevyskytují paralelní hranы (např. pokud je jednoduchý), pak je zase trasa jasná už ze zastávek, protože při přesunu není na výběr, kterou hranu použít. V takových situacích se sledy často udávají jen výčtem vrcholů (v_0, v_1, \dots, v_N) . Například v grafu z příkladu 12a.b máme sled (a, c, d, e) , zatímco (a, b, c) sled není, protože nemáme hranu z b do c .

Poznamenejme ještě, že možnost chodit v orientovaném grafu neorientovanými sledy je v některých situacích velmi užitečná.

Pokud jsme v situaci, kdy pracujeme výhradně s jednoduchými grafy, tak se může vyplatit ten první přístup k teorii grafů. Zápis se pak poněkud zjednoduší a informace je lépe vidět. Ukážeme, jak by vypadala definice sledu.

! Definice.

Nechť $G = (V, E)$ je graf (neorientovaný či orientovaný). Uvažujme vrcholy $v_0, v_1, v_2, \dots, v_N \in V$.

Řekneme, že posloupnost $(v_0, v_1, v_2, \dots, v_N)$ je sled, jestliže pro každé $i = 1, \dots, N$ platí, že $\{v_{i-1}, v_i\} \in E$, popřípadě (pro orientované grafy) $(v_{i-1}, v_i) \in E$.

Bylo to tedy kratší a dobře se s tím pracuje, jenže v aplikacích o cestování se často setkáváme s paralelními hranami, což právě tento jednodušší přístup nezvládá moc dobře. V této kapitolce proto budeme preferovat spíš tu variantu abstraktnější, s pojmenovanými hranami. Jedna z věcí, kterou nám nabízí, je sled délky nula, což je (v) , tomu říkáme triviální sled a bere se jako orientovaný i neorientovaný, ale nepovažuje se za uzavřený. Jeho existence občas zjednoduší úvahy a důkazy, mimo jiné to znamená, že pro každý vrchol v existuje sled z v do v .

Obvykle nás zajímají jen sledy speciální. Mnoho úloh z teorie grafů se zabývá hledáním efektivních řešení, tedy odstraňování zbytečného opakování. Pokud je například po sněhové kalamitě záhodno zprovoznit alespoň základní dopravní síť, tak by bylo pěkné naplánovat trasu pro sněžný pluh tak, aby nikdy nejel po téže trase dvakrát. Z pohledu sledu to znamená podmínu, že když $i \neq j$, tak $e_i \neq e_j$. Z praktických důvodů (a tradičně, viz definice prostého zobrazení) ji v definici vyjádříme obměnou této implikace.

Další ekonomický požadavek zase může být, abychom nenavštívili totéž místo dvakrát, například pokud máme rozvést zásilky. Podmínka se zdá jasná, zakážeme opakování vrcholů, ale narází to na drobný problém, že bychom si tím zakázali návrat tam, odkud jsme vyrazili. Tím by se nový pojem stal vysoce nepraktický, proto definici poněkud zkomplikujeme tak, aby zakázala opakování v průběhu trasy, ale dovolila nám se vrátit na začátek.

! Definice.

Nechť $(v_0, e_1, v_1, e_2, v_2, \dots, e_N, v_N)$ je sled v nějakém grafu $G = (V, E, e)$.

Řekneme, že je to **tah** (**trail**), jestliže se žádné hrany neopakují, tedy pro každé $i, j = 1, \dots, N$ platí, že když $e_i = e_j$, tak už $i = j$.

Řekneme, že je to **cesta** (**path**), jestliže je to tah a žádné vrcholy se neopakují, tedy pro každé $i, j = 0, \dots, N$ platí, že jestliže $v_i = v_j$, pak už nutně $i = j$ nebo $\{i, j\} = \{0, N\}$.

Zákaz opakování hran je stará známá věc, stejně pravidlo dodržujeme, když se snažíme nakreslit obrázek jedním tahem. U orientovaných grafů „tah“ znamená orientovaný tah, což znamená, že se nesmí opakovat tatáž konkrétní hraná, ale mezi dotyčnými vrcholy je možné jet znovu, pokud tam existuje hrana paralelní. Pokud cestujeme v orientovaném grafu neorientovaným sledem, pak pojem tahu funguje obdobně: Pokud nějakou hranou projedeme, pak ji máme zakázanou, i když jsme ji třeba projeli „opačným“ směrem, ale mezi dotyčnými dvěma vrcholy klidně mohou existovat další hrany, které jsou stále k dispozici (v libovolném směru). Obecná definice tahu toto správně vystihuje.

Pokud se čtenář trochu zamyslel nad definicí cesty, tak jej možná napadlo, jestli není ten požadavek o tahu zbytečný. Přece když nesmí opakově navštívit vrcholy, pak také nemůžu ani zopakovat hrany. Bohužel, u této úvahy existuje výjimka. Pokud se podíváme na neorientovaný graf K_2 , třeba v inkarnaci $V = \{1, 2\}$ a $E = \{\{1, 2\}\}$, tak je možné udělat uzavřený sled $(1, 2, 1)$, který opravdu opakuje jen koncové vrcholy, takže technickou podmínu na cestu splňuje, a přesto opakuje hranu. Právě kvůli tomuto jedinému grafu jsme proto do definice cesty museli přidat podmínu, že jde i o tah, abychom dostali žádanou hierarchii sled–tah–cesta.

V anglické terminologii je docela zmatek. Uvedli jsme termíny, které blízce odpovídají našim, nicméně existuje ještě jiný způsob pojmenování, který je možná dokonce častější. Mnozí autoři mají path jako pojmenování obecné, tedy český sled, naši cestě pak říkají simple path a pro tah nemají dokonce vůbec žádný speciální pojem. A aby to bylo ještě zajímavější, když někteří autoři řeknou path, tak tím myslí uzavřenou cestu/sled. Je v tom guláš.

Cesty známe z plánování tras v autoatlase, nechceme nějakým městem projet dvakrát. Selský rozum říká, že pokud při plánování trasy nějaké město navštívíme podruhé, pak jsme jeli dokola a tuto smyčku navíc lze bez problémů vynechat. Řekneme si to matematicky.

Fakt 12c.1.

Nechť G je graf. Jestliže existuje sled z vrcholu u do vrcholu v , pak už také existuje cesta z vrcholu u do vrcholu v .

Díky tomu se často pracuje jen s cestami.

Sledy, tahy i cesty se dají napojovat, například máme-li sled z a do b a sled z b do c , pak také máme sled z a do c . Při spojování tahů a cest si ale musíme hlídat, zda nedojde k opakování.

Pokud si vrcholy sledu shromáždíme do množiny V' a použité hrany do množiny E' , tak vlastně dostáváme podgraf $G' = (V', E')$. Máme pak k dispozici příslušné pojmy a nástroje. Mnohdy je naopak výhodnější namísto jazyka grafů použít jazyk sledů. Vrátíme se k jednomu pojmu z úvodu.

Definice.

Nechť G je (orientovaný či neorientovaný) graf.

Pojmem kružnice v G rozumíme libovolnou uzavřenou neorientovanou cestu v grafu G .

Je-li graf G neorientovaný a kružnice v něm existuje, nazývá se cyklický.

Je-li graf G orientovaný, pak pod pojmem cyklus rozumíme libovolnou uzavřenou (orientovanou) cestu v grafu G . Pokud taková existuje, graf se nazývá cyklický.

Cesty nám mimo jiné umožní testovat, jak je na tom daný graf s propojeností.

Definice.

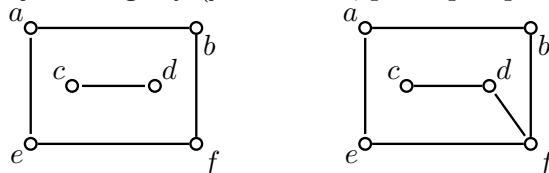
Nechť $G = (V, E, \varepsilon)$ je neorientovaný graf. Řekneme, že je **souvislý (connected)**, jestliže pro libovolné $u, v \in V$ existuje cesta z u do v .

Definice říká, že všechny vrcholy souvislého grafu jsou nějak navzájem propojeny, což naznačuje, že se graf neskládá z více nezávislých částí. Je tomu přesně tak, dokonce by to šlo použít jako definici.

Věta 12c.2.

Nechť $G = (V, E, \varepsilon)$ je neorientovaný graf. Tento graf není souvislý právě tehdy, když existují dva jeho podgrafy $G_1 = (V_1, E_1, \varepsilon)$ a $G_2 = (V_2, E_2, \varepsilon)$ takové, že $V_1 \cap V_2 = \emptyset$, $V = V_1 \cup V_2$ a $E = E_1 \cup E_2$.

Příklad 12c.a: Uvažujme následující dva grafy (jednoduché, proto pak použijeme to snažší značení).



Graf nalevo není souvislý, protože nedokážeme najít cestu z a do c .

Funguje to i podle věty, dokážeme jej vytvořit pomocí dvou disjunktních podgrafů

$G_1 = (V_1 = \{a, b, e, f\}, E_1 = \{\{a, b\}, \{a, e\}, \{b, f\}, \{e, f\}\})$ a $G_2 = (V_2 = \{c, d\}, E_2 = \{\{c, d\}\})$.

Naopak graf napravo souvislý je, což se ověřuje obtížněji, protože musíme prověřit propojenost všech dvojic vrcholů (níže ukážeme, že se toto hledání dá zjednodušit). Z obrázku to nicméně vypadá naprostě jasně. Vychází to i podle věty, jakýkoliv pokus o sestavení ze dvou samostatných grafů selže. Zkusme třeba zvolit

$G_1 = (V_1 = \{a, b, e, f\}, E_1 = \{\{a, b\}, \{a, e\}, \{b, f\}, \{e, f\}\})$ a $G_2 = (V_2 = \{c, d\}, E_2 = \{\{c, d\}\})$.

Pak sice dostáváme dva podgrafy, ale není splněna podmínka $E = E_1 \cup E_2$, protože hrana $\{d, f\}$ není v ani jednom z podgrafů. Když si tu hranu zkusíme k jednomu z nich přidat, třeba k prvnímu, tak to také nepůjde, protože vzniklý útvar není graf (pak E_1 obsahuje $\{d, f\}$, což není podmnožinou V_1).

△

Když máme graf, který souvislý není, tak jej vždycky dokážeme poskládat z částí, které už souvislé jsou. Tyto části mají speciální název.

Definice.

Nechť $G = (V, E, \varepsilon)$ je neorientovaný graf. Řekneme, že podgraf $G' = (V', E', \varepsilon)$ je **komponenta souvislosti (component)** grafu G , jestliže je G' souvislý graf a je to maximální souvislý podgraf.

Obvykle říkáme jen „komponenta“. Maximalita znamená, že pokud k G' přidáme další hrany a vrcholy z G tak, aby zase vznikl podgraf, tak už nemůže být souvislý.

Když jsme u příkladu výše sestavili graf nalevo z podgrafů G_1 a G_2 , tak jsme shodou okolností také našli komponenty.

Uvažujme neorientovaný graf. Snadno si rozmyslíme, cesta z u do v zároveň po přerovnání pořadí dává cestu z v do u , jde tedy o symetrickou situaci. Pokud máme cestu z u do v a cestu z v do w , pak napojením vznikne obecně jen sled, ale vynecháním kružnic dostaneme cestu z u do w , spojení cestou je tedy i tranzitivní. Reflexivita je dána díky té cestě o délce 0, takže vlastnost „být propojen cestou“ je relace ekvivalence na vrcholech grafu. Komponenty jsou samozřejmě právě třídy ekvivalence této relace.

Jeden praktický důsledek je, že souvislost pak můžeme trochu jednodušeji testovat takto:

Fakt 12c.3.

Uvažujme neorientovaný graf $G = (V, E, \varepsilon)$, zvolme si nějaký vrchol $v \in V$.

G je souvislý právě tehdy, když pro každé $u \in V$ existuje cesta z v do u .

Jak upravíme pojem souvislosti pro orientované grafy? Nabízejí dvě různé cesty, a protože se ani jeden ze způsobů neukázal obecně lepším, vznikly dva různé pojmy.

Definice.

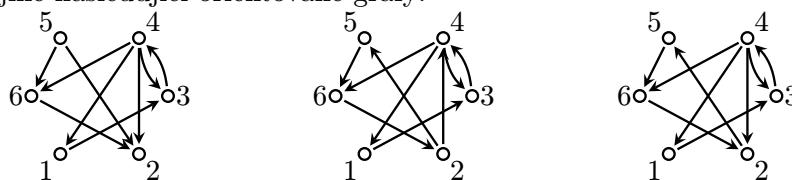
Nechť $G = (V, E)$ je orientovaný graf.

Řekneme, že G je **souvislý** (někdy také **slabě souvislý**), jestliže pro každé dva vrcholy $u, v \in V$ existuje neorientovaná cesta z u do v .

Řekneme, že G je **silně souvislý**, jestliže pro každé dva vrcholy $u, v \in V$ existuje (orientovaná) cesta z u do v .

Druhá definice vypadá velice přirozeně, vždyť jsme jen přepsali definici od neorientovaných grafů, ale pro některé účely je zbytečně náročná a tedy i zbytečně pracná. Stojí za zamýšlení, že definice souvislosti se dá také říct jinak: Orientovaný graf je (slabě) souvislý, jestliže je souvislá jeho symetrizace.

Příklad 12c.b: Uvažujme následující orientované grafy:



Jak je tomu s jejich souvislostí?

Pokud u hran ignorujeme orientaci (tedy provedeme symetrizaci), tak u všech tří dostáváme shodou okolností totéž a snadno si rozmyslíme, že jde o souvislý graf. Dané tři grafy jsou proto všechny slabě souvislé.

Jak to vypadá se silnou souvislostí?

Levý graf není slabě souvislý, protože nedokážeme vytvořit cestu z nějakého vrcholu do vrcholu 5, má vstupní stupeň $\deg^-(5) = 0$. Neexistuje tedy například cesta z 4 do 5 a proto není graf silně souvislý.

Lze také argumentovat, že nelze vytvořit cestu začínající ve vrcholu 2, protože $\deg^+(2) = 0$.

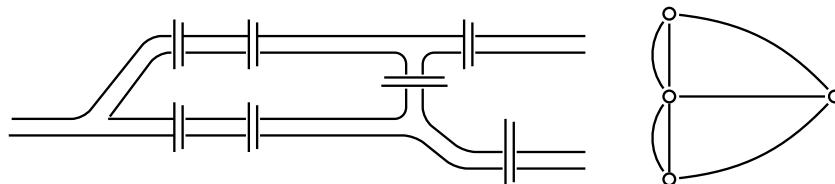
U dalších dvou grafů mají všechny vrcholy nenulové vstupní i výstupní stupně, takže tak snadné to nebude a je třeba trpělivě zkoušet. Nakonec se ukáže, že graf uprostřed silně souvislý je, ale ten napravo není, protože například nelze najít cestu z 5 do 4.

△

Se sledy je svázáno množství úloh, uvedeme si nejznámější. Začneme tam, kde se narodila teorie grafů.

Příklad 12c.c: Bylo jednou v Prusku město Königsberg (neboli Královec na počest Přemysla Otakara II, který financoval výstavbu první tamní pevnosti) a v něm pracoval jistý Leonhard Euler.

Díky zajímavě se chovající řece se město Kaliningrad (to je taky ono) rozpadalo na čtyři oddělené části pospojované porůznou sedmi mosty a Eulera jednou (v roce 1736) přepadla otázka, zda je možné udělat si procházku tak, aby prošel všechny mosty, ale každým jen jednou, a skončil tam, odkud vyšel. Podíval se na mapu a napadlo ho, že tvary částí jsou nepodstatné, tak si místo nich nakreslil kroužky a mezi ně spojnice jako mosty.



Dostal tak graf (podle prvního přístupu multigraf, tady vidíme, že hned u historicky první úlohy o cestování se lépe hodí ta druhá, abstraktnější definice) a v něm chtěl najít uzavřený tah, který by prošel všemi hranami, ale ty pojmy ještě tenkrát neexistovaly, tak si je vymyslel. Chvíli nad situací spekuloval a brzy zjistil, že jeho vysněná procházka je nemožná. Všiml si také, že by to nedokázal, dokonce i kdyby se vzdal myšlenky, že skončí tam, kde začal.

Protože byl matematik, napadlo jej, zda neexistuje obecná metoda, jak v daném grafu takovou otázku zodpovědět. A protože byl geniální matematik, tak ji také našel. Založil tím nový podobor teorie grafů.

△

Definice.

Nechť $G = (V, E, \varepsilon)$ je graf (neorientovaný či orientovaný).

Řekneme, že sled $(v_0, e_1, v_1, \dots, e_N, v_N)$ je **eulerovský tah** (**Eulerian trail/walk**), jestliže je to tah a $\{e_1, e_2, \dots, e_N\} = E$.

Graf G se nazývá **eulerovský** (**Eulerian graph**), pokud v něm existuje uzavřený eulerovský tah.

Připomeňme, že tah v orientovaném grafu znamená orientovaný tah.

My známe eulerovské tahy již z dětství, protože nejde o nic jiného než o profláklý problém, jak ten či onen obrázek nakreslit jedním tahem. Pěticípou hvězdu zmákneme levou zadní (leváci pravou zadní), také známý domeček nakreslit umíme, ale skončíme jinde, než jsme začali, a domeček s komínem už nezvládneme. Schwálňejstli to teorie správně pozná.

Je jasné, že nesouvislé grafy jedním tahem nakreslit nelze, takže se rovnou soustředíme na souvislé a mezi nimi chceme umět rozhodnout. Euler odvodil následující:

Věta 12c.4.

Nechť G je neorientovaný souvislý graf.

(i) G je eulerovský právě tehdy, pokud v něm mají všechny vrcholy sudý stupeň.

(ii) V grafu existuje otevřený eulerovský tah právě tehdy, když v grafu existují právě dva vrcholy s lichými stupni.

Důkaz naznačíme, nejprve ten lehký směr. Uzavřený eulerovský tah musí do každého vrcholu vjet a pak zase vyjet, čímž se stupeň tohoto vrcholu zvýší o dva. Vrchol může takto být projet vícekrát, čímž vznikají sudé stupně. Pokud je to tah otevřený, pak jsou výjimkou vrcholy, kde tento tah začíná a končí, tam vznikají liché stupně.

Těžší je dokázat, že správné parity stupňů již stačí k vytvoření žádaného tahu. Důkaz se nejčastěji dělá popisem konstrukce takového tahu, naznačíme hlavní myšlenku.

Nejprve si rozmyslíme, že pokud má v grafu nějaký vrchol stupeň nula, pak z něj nevedou hrany. Jestliže je takový graf zároveň souvislý, pak už je to nutně jednobodový graf a ten je eulerovský, prázdný sled je hledaným tahem. Pokud tedy pracujeme se souvislým grafem s více vrcholy, tak už má každý vrchol stupeň alespoň jedna.

Mějme teď souvislý graf s více vrcholy a sudými stupni. Zvolme libovolný vrchol v_0 , ten má určitě nenulový stupeň a tudíž z něj vede hrana. Vytvoříme netriviální uzavřený tah, jehož součástí v_0 bude.

Vyberme libovolnou hranu vedoucí z v_0 , její cílový vrchol označme v_1 , pak tuto hranu odebereme z grafu, čímž zároveň klesnou stupně v_0 a v_1 o jedno a jsou liché. Stupeň v_1 byl sudý, tudíž stále není nula, vede z něj proto nějaká jiná hrana, vydáme se po ní do dalšího vrcholu, hranu odebereme a snížíme stupně. Pokud je tím cílem v_0 , našli jsme uzavřený tah, zároveň se stupeň v_0 vrátil na sudou hodnotu.

Pokud to není v_0 , tak jej nazveme v_2 , po snížení stupňů má v_1 zpět sudou hodnotu stupně a lichou teď mají v_0 a v_2 . Kdyby měl náhodou vrchol v_1 teď stupeň nula, tak jej z grafu odebereme také. Takto postupujeme dále, v každém kroku odebereme hranu a případné vrcholy se stupněm spadlým na nulu a budujeme tah z vrcholu v_0 do v_k , přičemž jedině tyto dva vrcholy mají aktuálně (v redukovaném grafu) snížené stupně liché, ostatní vrcholy mají stupně sudé. Klíčem zde je, že pokud nějaký krok vede do vrcholu s aktuálně sudým stupněm, tak z něj zase povede cesta ven, takže jedený způsob, jak tento postup ukončit, je trefit se zase do v_0 . A ukončit musíme, protože v každém kroku umažeme z grafu hranu a těch je jen konečně mnoho.

Proces se tedy ukončí a vznikne uzavřený tah z v_0 do v_0 . Pokud po tomto odmazávání hran a vrcholů už v grafu žádná nezbyla, pak jsme našli kýžený euklidovský tah a máme hotovo.

Uvažujme druhou možnost, a to že v grafu ještě nějaké hrany zbyly. Musely proto také zbýt nějaké vrcholy s nenulovými stupni a tyto stupně jsou sudé, protože naše vytvářecí procedura měla liché stupně vždy jen na začátku a konce budovaného tahu, při uzavření jsme se pak vrátili do v_0 , čímž vznikl sudý stupeň. Dále tvrdíme, že dokonce některý z vrcholů našeho tahu je mezi těmi, kterým ještě zbyl nenulový stupeň.

Vezměme tedy tento vrchol w_0 a přesně jako předtím z něj začneme a nakonec v něm dokončíme netriviální tah. Tento tah „vlepíme“ do našeho původního, formálně to je delší, tak to řekneme neformálně: Nejprve jedeme původním tahem z v_0 do w_0 , pak projedeme tu novou trasu a nakonec dokončíme původní tah do v_0 . Dostali jsme tedy tah nový, který dále zmenšíl počet hran v grafu. Pokud už nic nezbylo, jsme hotovi, pokud ještě něco zbylo, pak to zase musí zahrnovat některý z vrcholů našeho rozrůstajícího se tahu, takže tento proces opakujeme. Protože graf měl konečně mnoho hran, musíme dřív či později zahrnout do tahu všechny a máme euklidovský tah.

Případ s dvěma vrcholy lichého stupně se hravě převede na původní, buď se mezi tyto dva vrcholy vloží extra hrana, začne se s ní hledání tahu a na konci se zase odebere, a v případě, že ty vrcholy už spojeny byly, se ta hrana zase odebere a zpětně doplní (musí se opatrnejí rozebrat případ, že by tím odebráním vznikly dva souvislé podgrafy, ale s pomocí obrázku to jde lehce).

△

Důkaz této věty nám dává algoritmus pro hledání eulerovského tahu, ale dá se vylepšit, dokážeme eulerovský tah najít vždy rovnou, bez nějakého vlepování.

0. Zvolte si libovolný vrchol s lichým stupněm, pokud takový není, tak libovolný vrchol. To je vrchol v_0 hledaného tahu.

1. Jsme ve vrcholu v_k . Pro každou hranu z tohoto vrcholu vedoucí uvažujme následující krok:

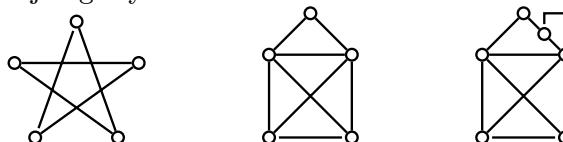
K : Hranu odebereme z grafu, tím se zároveň sníží o jedničku stupeň vrcholu v_k a cílového vrcholu. Pokud $\deg(v_k) = 0$, tak z grafu odebereme i tento vrchol, podobně odebereme cílový vrchol, pokud mu stupeň klesne na nulu.

Z hran vedoucích z v_k si pak vybereme takovou, u které provedení dotyčného kroku K nezpůsobí rozpad grafu na více komponent. Krok provedeme, cílový vrchol označíme v_{k+1} . Pokud byl i tento vrchol odebrán, algoritmus skončil, jinak se vracíme na krok 1.

Je jasné, že klíčový je právě ten výběr hrany. Musíme zaručit, že alespoň jedna z hran vedoucích z v_k po odebrání (a případném odebrání vrcholu) zase vede na souvislý graf. Není na to třeba žádná velká teorie, stačí trocha přemýšlení a hodně kombinatoriky, ale je to delší, odkážeme na nějakou pěknou knížku z teorie grafů.

Pak je ještě třeba ukázat, že po doběhnutí algoritmu se opravdu prošla každá hrana právě jednou. To je ale snadné. Každá použitá hrana se z grafu odebere, není tedy možné ji recyklovat a vznikající procházka je opravdu tah. Pokud by v grafu zbyla nějaká neprojídatá hrana, tak by sousední vrcholy měly nenulové stupně, tedy algoritmus by neskončil.

Příklad 12c.d: Uvažujme následující grafy.



V pěticípé hvězdě má každý vrchol stupeň 2, tudíž ji lze nakreslit jedním (uzavřeným) tahem a je to eulerovský graf. Všimněte si, že když k ní ještě přidáme pentagram, tak vlastně dostáváme K_5 , kde jsou zase všechny stupně sudé a bez problémů kreslíme.

U domečku máme (bráno od střechy) jeden vrchol se stupněm 2, dva vrcholy se stupněm 4 a dva vrcholy se stupněm 3. Nakreslit jedním tahem tedy půjde, ale skončíme jinde, než jsme začali. Věta 12c.4 také napoví, že začít musíme v jednom z těch lichých vrcholů a skončíme v tom druhém.

Po přidělání komínů přibudou další dva vrcholy s lichým stupněm a jedním tahem už to nakreslit nepůjde.

△

Situace v orientovaném grafu je obdobná. Můžeme si to představit tak, že některé mosty jsou jednosměrné, a podmínka z definice tahu říká, že když je nějaký most obousměrný a my po něm jednou přejdeme, tak už podruhé nechceme ani v opačném směru.

Věta 12c.5.

Souvislý orientovaný graf $G = (V, E, \varepsilon)$ je eulerovský právě tehdy, jestliže pro každé $v \in V$ platí $\deg^+(v) = \deg^-(v)$.

U neuzavřených Eulerovských tahů je rovněž myšlenka podobná neorientované verzi, se zjevnou modifikací: Existence eulerovského tahu pozná podle toho, že jeden vrchol má vstupní stupeň o jedničku větší než výstupní, další vrchol to má naopak a ostatní mají oba stupně stejné.

Na závěr se vrátíme za Eulerem so Koenigsbergu. Tam vidíme, že všechny čtyři vrcholy mají lichý stupeň, tudíž je Eulerova procházka neuskutečnitelná.

Příklad 12c.e: Další zajímavý problém má historický název **problém obchodního cestujícího** (travelling salesman problem) a začneme jeho jednodušší verzí. Obchodák má před sebou mapu měst, která potřebuje navštívit, ta jsou různě pospojovaná cestami (což mohou být silnice nebo třeba letecké spoje). Obchodní cestující chce navštívit všechna místa, ale žádné dvakrát (už tam lidem prodal, co šlo), a samozřejmě se chce vrátit, hledá tedy uzavřenou cestu. V situaci, kdy by byla propojena všechna města navzájem, to samozřejmě není problém, ale čím řidší síť, tím náročnější je to na naplánování. Pokud by cestoval živelně, pak se mu může stát, že se najednou ocitne ve městě, ze kterého vedou cesty jen tam, kde už byl.

Lze takovou cestu naplánovat vždy? A pokud to někdy lze, je na nalezení správné cesty metoda? Samozřejmě jsou to otázky pro teorii grafů.

△

Definice.

Nechť $G = (V, E, \varepsilon)$ je graf (neorientovaný či orientovaný).

Hamiltonovská kružnice (Hamiltonian cycle) v tomto grafu je libovolná kružnice, která obsahuje právě jednou každý vrchol tohoto grafu.

Pokud v tomto grafu hamiltonovská kružnice existuje, pak řekneme, že je to **hamiltonovský graf**.

Jinak řečeno, hamiltonovská kružnice je uzavřená cesta, která projde všechny vrcholy grafu. Je zřejmé, že nás mezi dvěma městy zajímá vždy nejvýše jeden spoj, u této úlohy tedy stačí umět pracovat s jednoduchými grafy. Tradičně se pracuje s neorientovanými grafy.

Problematika hamiltonovskosti je výrazně odlišná od euklidovských otázek. Není totiž známa žádná jednoduchá ekvivalentní podmínka, která by nám umožnila rozhodnout, který graf je hamiltonovský. Navíc není znám algoritmus, který by pro hamiltonovské grafy dokázal rychle najít hamiltonovské kružnice. Již z podstaty jsou zjevné dvě nutné podmínky.

Věta 12c.6.

Nechť $G = (V, E, \varepsilon)$ je konečný neorientovaný graf. Jestliže je hamiltonovský, tak musí být souvislý a pro každý vrchol $v \in V$ musí platit $\deg(v) \geq 2$.

Druhá podmínka plyne z toho, že cestující musí do města někudy přijet a pak zase odjet, přičemž se při odjezdu nemůže vrátit, odkud přijel.

Existují také užitečné postačující podmínky, všechny nějakým způsobem nutí graf mít hodně hran. Ukážeme čtyři nejpopulárnější, nejsou zcela nezávislé, některé jsou jen speciálními případy jiných.

Věta 12c.7.

Nechť $G = (V, E)$ je konečný souvislý (neorientovaný) graf takový, že $|V| \geq 3$.

Tento graf je hamiltonovský, jestliže je splněna některá z následujících podmínek:

- (i) Pro všechny vrcholy $v \in V$ je $\deg(v) \geq \frac{1}{2}|V|$; (Diracova podmínka)
- (ii) Pro všechny vrcholy $u, v \in V$ platí, že $\{u, v\} \in E$ nebo $\deg(u) + \deg(v) \geq |V|$; (Oreho podmínka)
- (iii) Předpokládejme, že vrcholy jsou očíslovány $V = \{v_1, v_2, \dots, v_n\}$ tak, aby $\deg(v_1) \leq \deg(v_2) \leq \dots \leq \deg(v_n)$. Pak pro všechna $i < \frac{1}{2}n$ platí $\deg(v_i) \geq i + 1$ nebo $\deg(v_{n-i}) \geq n - i$; (Chvátalova podmínka)
- (iv) Pro každé $k \in \mathbb{N}$, $k < \frac{1}{2}|V|$ platí, že $|\{v \in V; \deg(v) \leq k\}| < k$. (Posova podmínka)

Chvátalova podmínka se dívá na dvojice vrcholů $v_1, v_{n-1}, v_2, v_{n-2}, v_3, v_{n-3}$ atd. a pro každou takovou dvojici v_i, v_j vyžaduje, aby platilo $\deg(v_i) \geq i + 1$ nebo $\deg(v_j) \geq j$.

Pro příklad, že ve všech podmínkách opravdu jde jen o implikace (tedy podmínky nejsou nutné) není třeba chodit daleko. Vezměme kružnici o pěti vrcholech, tedy graf

$$V = \{1, 2, 3, 4, 5\}, E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 5\}, \{5, 1\}\}.$$

Pak všechny vrcholy splňují $\deg(v) = 2 < \frac{1}{2}|V|$, všechny dvojice vrcholů mají $\deg(u) + \deg(v) = 4 < |V|$ a volba $k = 2$ ukazuje, že ani Posova podmínka není splněna, Chvátal také ne, přesto je to evidentně hamiltonovský graf, obchodní cestující prostě kružnici objede dokola.

Příklad 12c.f: Snadno se rozmyslí, že každý úplný graf K_n pro $n \geq 3$ je hamiltonovský, stačí projet cestu $(1, 2, 3, \dots, n, 1)$.

Na druhou stranu toto zajímavé doplnění grafu K_5 zvané Petersenův graf už hamiltonovské není. Jak to víme? Nějaká snadná kritéria na vyrácení hamiltonovskosti nejsou, v tomto případě jsme prostě začali v náhodně vybraném vrcholu a zkoušeli všechny možné cesty, které z něj lze vytvořit, žádná nevedla k cíli.

△

Obecně bohužel nic lepšího než hrubá síla není a procházení všech možných cest je pořádně drahé. Rozhodování, zda je daný graf hamiltonovský, je totiž NP-úplný problém.

Příklad 12c.g: Tradičně se název problém obchodního cestujícího používá pro mírně jinou úlohu. Pracuje se s ohodnoceným úplným grafem a nehledá se jen tak ledajaká hamiltonovská kružnice, ale ta nejkratší. I to je NP-úplná úloha. V praxi se proto používají heuristické algoritmy, které nacházejí řešení blízká optimálnímu (ale neumí se zaručit, jak blízké) v rozumném čase.

Řešení lze nalézt snáze a dokonce rozumně blízké optimálnímu, pokud hodnoty hran splňují trojúhelníkovou nerovnost, což znamená, že přímá cesta mezi dvěma uzly není nikdy ohodnocena více než nějaká cesta oklikou. V mnoha aplikacích je toto splněno, jsou ale i zajímavé aplikace, kde trojúhelníková nerovnost neplatí, například pokud ohodnocení hran odpovídá době jízdy, pak může být přímý spoj dražší než oklika po dálnici.

△

Úlohy řešené v ohodnocených grafech se objevují v mnoha aplikacích.

Příklad 12c.h: Eulerův problém má zajímavou variantu zvanou **problém čínského poštáka** (Chinese postman problem). Odehrává se v ohodnoceném grafu (u ulic města známe jejich délky) a cílem je najít nejkratší (ve smyslu součtu použitých ohodnocení) uzavřený sled, který by prošel všemi hranami. Jinými slovy, hledáme nejkratší trasu, která poštáka povede všemi ulicemi.

Pokud je graf eulerovský, pak je eulerovský uzavřený tah také řešením poštákova problému, protože každou ulicí projít musí a eulerův tah jde každou jen jednou, méně to proto nejde.

Zajímavější je, pokud graf eulerovský není, pak je nutno některými hranami projít vícekrát a není na první pohled zjevné, jak efektivně najít celkově nejlevnější trasu. Jsou algoritmy, které k nalezení řešení vyžadují řádově $|V|^3$ kroků.

△

Existují další varianty poštácké úlohy, některé z nich jsou NP-úplné. Protože se ohodnocení sledů sčítají často, zavádí se pro to jméno. Máme-li graf (neorientovaný či orientovaný) s ohodnocením w pro hranы, pak pro sled (v_0, v_1, \dots, v_N) definujeme jeho **váhu** popř. **cenu** jako $\sum_{i=1}^N w(v_{i-1}, v_i)$. Další verze problému například hledají nejlevnější spojení mezi dvěma zadánými vrcholy, nejlevnější spojení z daného vrcholu do všech ostatních a podobně, na mnohé z těchto úloh existují efektivní algoritmy.

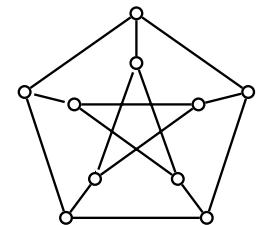
Někdy naopak chceme cenu co největší, což je ale ekvivalentní úloha. Stačí u všech ohodnocení hran změnit znaménko a algoritmus na hledání nejlevnější cesty aplikovaný na tento nový graf nám najde cestu původně nejdražší.

Jiná zajímavá úloha: Ohodnocení představují propustnost (třeba průměr potrubí). Je-li zvolen počáteční a cílový vrchol, jakou největší zásilku dokážeme vcelku poslat? Zde je evidentní, že nestačí jen najít hranu s nejmenší propustností, protože třeba existuje cesta, která se jí dokáže vyhnout. Praktické aplikace si čtenář jistě dokáže živě představit třeba u firmy přepravující rozdílné náklady.

Mnohé posílané věci lze rozdělit, například vodu, elektřinu či byty. V takovém případě se úloha přepouštění co největšího množství z jednoho vrcholu do jiného liší od předchozí. Pro obě verze existují standardní polynomiální algoritmy. Podoblasti teorie grafech, která se zabývá úlohami tohoto typu, se říká **toky v sítích** (flow networks).

12d. Kreslení grafů

Každý konečný graf lze nakreslit, ale ne vždy to k něčemu je. Můžeme například požádat počítač, aby nám nakreslil zadaný graf s milionem vrcholů na lodní plachtu, ale asi ze vzniklé houštiny moc nevyčteme. Přesto je kreslení grafech populární oblastí, protože často své nápady testujeme na menších grafech a u nich nakreslení může výrazně pomoci našemu přemýšlení.



Má to i praktické dopady. Představte si návrháře obvodů. Má hromádku švábů (odpory, kondiky, ledky, tranzistory, integráče, …), které potřebuje porůznu pospojovat, aby dělaly, co potřebuje. Vyrábět houštinku z drátů není nejlepší metoda, bylo by moc pěkné, kdyby prostě mohl vyleptat desku a šváby na ni osadit. Je to možné? Když si každou součástku reprezentujeme vrcholem a nutná propojení pomocí hran, dostáváme graf. To, že součástky někam přimontujeme a pospojujeme vodivými cestičkami, přesně odpovídá nakreslení grafu v rovině. Teď ovšem máme problém, protože na tištěném spoji se vodivé cestičky nesmějí krížit. Pokud tedy navrhne obvod a chceme jej osadit na desku, tak to půjde jedině v případě, že odpovídající graf dokážeme nakreslit tak, aby se hrany nikde nekrížily.

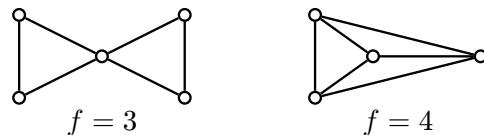
Tento problém evidentně zajímá všechny firmy zabývající se návrhy obvodů, ale matematici se jím zabývali již dávno v dobách, kdy jediná elektřina v domácnosti byl občasný blesk, čistě ze zvědavosti. Ukázalo se, že možnost být takto nakreslen o grafu ledacos důležitého říká, my si zde tuto problematiku jen lehce přiblížíme.

Definice.

Nechť $G = (V, E)$ je graf. Řekneme, že je **rovinný (planar)**, jestliže je možné nakreslit jeho obrázek v rovině tak, aby se hrany neprotínaly.

Pro úplnost je třeba vyjasnit, že když se dvě hrany potkají v jednom vrcholu, tak se to nebore jako protínání. V češtině se někdy říká **planární graf** (mně se to líbí více), ale není to tak běžné.

Když si nakreslíme graf (libovolný), tak nám rozdělí rovinu na oblasti. Pokud použijeme pro hrany úsečky, tak ty oblasti budou mnohoúhelníky, jejichž strany jsou dány hranami. Pokud je graf rovinný a je správně nakreslen, tak u každého takového mnohoúhelníku platí, že má v každém vrcholu nějaký vrchol grafu. Pokud jsme pro hrany nepoužili úsečky, ale rozličné křivky, pak to platí obdobně, vznikají jakési „skoromnohoúhelníky“, podstatné je, že u rovinných nakreslení v každém křížení hran najedeme vrchol. Tyto oblasti (minimální části roviny vymezené hranami grafu) mají své jméno, říkáme jim **stěny**. Z ryze praktických důvodů uvažujeme vždy jako stěnu také „vnějšek“ grafu, protože i to je část roviny, která je vymezená hranami grafu. Počet stěn je jedním z charakteristických rysů grafu, značí se f . Pro jistotu uvedeme dva příklady, čtenář se jistě dopočítá správné hodnoty f .



Mimochodem, opět jde o populární grafy, tomu vlevo se říká motýlek a ten vpravo je samozřejmě K_4 , úplný graf o 4 vrcholech, který jsme ale nakreslili jinak než předtím. Ukazuje se, že K_4 je planární. Připomíná nám to, že když vidíme graf a kříží se v něm hrany, tak to ještě neznamená, že by nešel nakreslit jako rovinný, i když třeba v té poskytnuté podobě vypadal velice pěkně.

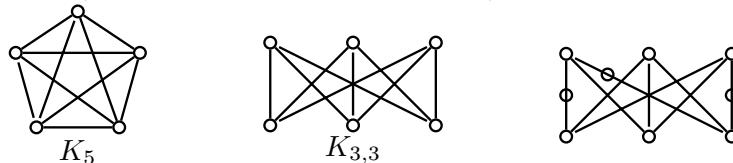
Nyní již můžeme přejít k inzerovanému tvrzení.

Věta 12d.1. (Eulerův vzorec)

Nechť $G = (V, E)$ je souvislý rovinný graf. Uvažujme nějaké jeho rovinné nakreslení a označme jako f počet stěn, které hrany grafu vytvářejí, včetně oblasti vně grafu. Pak $|V| - |E| + f = 2$.

Na důkaz nám chybí některé nástroje, takže to necháme na nějaký pokročilejší kurs teorie grafů.

Dá se u grafu poznat, zda je planární? Přirozeným začátkem je podívat se na populární jednodušší grafy a brzy se zjistilo, že nejjednodušší grafy (s co nejmenším počtem vrcholů), u kterých rovinnost selhává, jsou K_5 a $K_{3,3}$.



Grafy K_n už jsme si představili, pro úplnost dodejme, že grafy $K_{n,m}$ vzniknou tak, že se nakreslí n vrcholů do horní řady, m vrcholů do spodní řady a obě řady se propojí všemi možnými způsoby.

Ke grafu $K_{3,3}$ se váže pěkná pohádka. Představíme si, že tři vrcholy v horní řadě jsou domky a tři vrcholy v dolní řadě jsou vývody elektřiny, vody a plynu. Pokud budou chtít příslušné tři společnosti napojit ony tři domky, pak se jejich vedení budou muset někde překřížit.

Je zjevné, že pokud je nějaký graf G rovinný, pak už musí být i všechny jeho podgrafy rovinné. Z toho mimojiné vyplývá, že jestliže je v nějakém grafu coby podgraf K_5 nebo $K_{3,3}$, tak už tento graf nemůže být rovinný. Zajímavé je, že to v zásadě funguje i naopak, tyto dva grafy jsou hlavním problémem pro rovinnost.

Potřebujeme ještě jedno pozorování, k tomu je ten obrázek vpravo výše. Pokud v grafu přidáme doprostřed nějaké hrany vrchol, tak se tím nemění jeho podstata, pokud byl původní graf rovinný, tak je ten nový také, a naopak.

Ted' už máme vše připraveno.

Věta 12d.2. (Kuratowského věta)

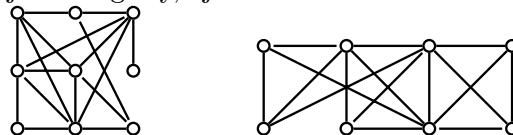
Graf G je planární právě tehdy, pokud neobsahuje podgraf, který je kopí grafů K_5 či $K_{3,3}$, popřípadě kopí grafů, které z K_5 či $K_{3,3}$ vzniknou přidáním vrcholů na hrany.

Podíváme se ted' na rovinnost některých grafů. Při úvahách se nám bude hodit i to, že přidání vrcholu na hranu lze obrátit. Takovouto operaci jsme ještě neměli, tak to řekneme přesně.

Vrchol uprostřed hrany musí mít nutně stupeň dva. Máme-li tedy takový vrchol v v grafu, tak z něj vedou hrany $\{v, u_1\}$ a $\{v, u_2\}$ pro nějaké vrcholy u_1, u_2 . Vymazání vrcholu v z hrany znamená, že v odebereme z množiny vrcholů, hrany $\{v, u_1\}$ a $\{v, u_2\}$ odebereme z množiny hran a místo nich tam přidáme hranu $\{u_1, u_2\}$.

Čtenář si snadno rozmyslí, že všechny vrcholy v nově vzniklém grafu mají stejný stupeň jako předtím, rovněž zůstaly zachovány všechny cesty a tahy a sledy. Vrchol v totiž nebyl křížovatkou, jen odpočívadlem na dálnici. Platí pořád, že vymazáním takového vrcholu neměníme (ne)rovinnost grafu.

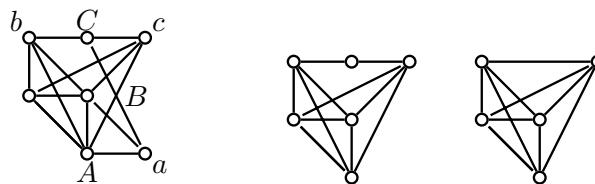
Příklad 12d.a: Uvažujme následující dva grafy, zjevně souvislé.



Začneme přemýšlet nad tím vlevo. Na pravém okraji vidíme „ocásek“, vrchol se stupněm jedna. Ten rovinnost neovlivní, proto jej odstraníme. Vidíme také vlevo dole vrchol se stupněm dva, je uprostřed hrany, proto jej vymažeme a dostáváme následující graf, který dále zjednodušíme odebráním zbytečně zdvojené hrany.



Další redukce není vidět, nezbývá než začít graf hypnotizovat a čekat, jestli se nám vyjeví jeden z těch dvou grafů K_5 a $K_{3,3}$ z Věty. Jsou tam oba. Na obrázku níže vlevo jsou tři vrcholy označeny malými písmeny (to je dolní řada $K_{3,3}$) a tři velkými písmeny. Ověřte, že každý vrchol s malým písmenem je spojen hranou s vrcholy značenými velkými písmeny a naopak, takže pokud přejdeme na podgraf určený touto množinou vrcholů, tak opravdu vznikne $K_{3,3}$.



Prostřední obrázek ukazuje, co vznikne, když vynecháme vrchol vpravo dole. Pak vrchol uprostřed nahoře náhle klesne se stupněm na dva, je uprostřed hrany, takže jej vymažeme a vykoukne na nás K_5 (obrázek vpravo).

Zkoumaný graf tedy obsahuje přímo kopii $K_{3,3}$ a také podgraf, který vznikne přidáním vrcholu na hranu grafu K_5 , takže máme dokonce dvojnásobně potvrzeno, že daný graf není rovinný.

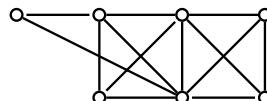
Stojí za zamýšlení, že ve fázi hledání K_5 nebylo možné odebrat z grafu vrcholy se stupněm 3, i když se v žádném K_5 objevit nemohou. Krásným příkladem je vrchol označený C . Má stupeň 3, tudíž se v našem K_5 logicky objevit nemůže. Je ale uprostřed hrany, kterou pro naše K_5 potřebujeme, a o tu bychom odebráním C přišli. Ztratili bychom tak spojku mezi vrcholy vlevo nahoře a vpravo nahoře a už bychom K_5 nenašli, i když tam je. Odebrání vrcholů stupně tří tedy není ekvivalentní úpravou z hlediska přítomnosti K_5 .

Zatímco obecně vrcholy stupně tří odebrat nelze, můžeme to udělat v případě, že za každou takto přerušenou spojovací cestu najdeme náhradu jinde. V našem příkladě umíme přes vrchol C vyrobit tři cesty. Jedna vede z levého horního do prostředního vrcholu a za tu existuje přímá náhrada, totéž platí u cesty mezi pravým horním a středovým vrcholem. Zatím dobré. Pokazí to ale již zmínovaná cesta mezi krajinimi horními vrcholy, ony totiž nejsou přímo spojeny.

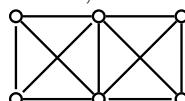
Ted' se podíváme na graf druhý. Nemá přívěsky neboli slepá střeva ani vrcholy se stupněm 2, tudíž si nelze situaci zjednodušit odebráním vrcholů.

Aby se v tom grafu schovávalo K_5 , tak by v něm muselo existovat pět vrcholů se stupněm alespoň čtyři. To není pravda, tudíž daný graf neobsahuje kopii K_5 . Zbývá se zamyslet nad $K_{3,3}$. Na to je třeba šest vrcholů se stupněm alespoň tři, což tam mají všechny, to si nepomůžeme. Uděláme to zkusmo, vyplatí se začínat od vrcholů se stupněm tři, pokud takové jsou.

Mohl by být levý dolní vrchol součástí $K_{3,3}$? Můžeme si jej představit v dolním patře. Je spojen s třemi jinými vrcholy, ty by pak logicky mohly být v patře horním. Aby opravdu vzniklo $K_{3,3}$, tak by se u těch tří vrcholů z horního patra mohly najít další dva společné cíle (vrcholy hypotetického dolního patra), ve kterých se z nich sejdou hrany. Zkoumaný graf ukáže, že takové dva vrcholy nejsou. Levý horní vrchol totiž sice má tři hrany, ale jedna vede ke kolegovi z horního patra, tudíž se nedokáže dostat třikrát dolů. To znamená, že levý dolní vrchol nemůže být součástí $K_{3,3}$ a můžeme jej z grafu odebrat.



Tím ovšem klesl stupeň horního levého vrcholu na dva, takže ani ten nemůže být součástí $K_{3,3}$ a vynecháme jej.



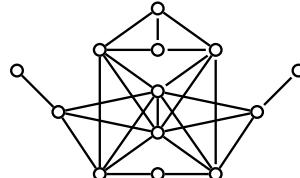
Zbylo šest vrcholů, takže buď je to $K_{3,3}$, nebo zkoumaný graf žádnou příšerku neobsahuje. Můžeme třeba udělat podobný rozbor jako předtím s levým dolním vrcholem a hravě zjistíme, že nemůže být součástí $K_{3,3}$.

Dospěli jsme k závěru, že daný graf je rovinný. Můžete zkoumat najít jeho správný tvar, aby se vám hrany nekřížily. Není špatný nápad začít překreslením toho posledního obrázku a pak přidat odebrané hrany a vrcholy.

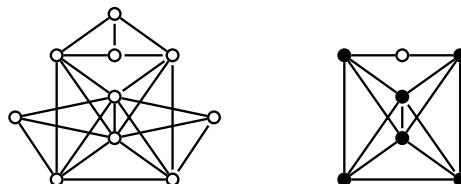
△

Ukázali jsme si několik zajímavých triků, ale je evidentní, že u houštinovitějších grafů bychom se s nimi daleko nedostali.

Příklad 12d.b: Šílený vynálezce dostal nápad, že by měl vzít tři kondenzátory, dvě diody, cívku, tři tranzistory, žárovku, zdroj, chcíplou krysu a rozbušku a propojit je způsobem, který hravě zachytíme jazykem grafů jako množinu hran, ale raději si ukážeme obrázek, součástky neboli vrcholy jsme pro začátek rozmištili v zásadě náhodně.



Chceme vědět, zda toto lze realizovat pomocí tištěného spoje neboli zda je to graf rovinný. Budeme proto hledat K_5 a $K_{3,3}$. Nejprve graf zjednodušíme, odebereme vrcholy se stupněm jedna a vymažeme z hrany vrchol se stupněm dva. Vznikne graf níže vlevo.



Na obrázku vpravo si pak graf připravíme na hledání K_5 . Nejprve si vybarvíme vrcholy se stupněm alespoň čtyři, protože mezi nimi budeme hledat. Dále se pokusíme graf ještě zjednodušit. Horní vrchol má stupeň tři a cesty přes něj spojují vrcholy, které jsou spojeny i jinak, tudíž jej při hledání K_5 lze ignorovat. Stejný argument platí i pro vrcholy, které trčí do stran. Dostáváme výrazně jednodušší graf, ze kterého lze navíc odebrat vrchol se stupněm dva z horní hrany.

Dokážeme z šesti vybarvených vrcholů vybrat pět tak, aby daly K_5 ? Začneme pokusem vyrobit kopii K_5 zahrnující levý dolní vrchol. Z něj vedou přesně čtyři hrany vedoucí k vybarveným vrcholům, takže není na výběr, je jasné, z čeho by se takové K_5 skládalo. Je všech těchto pět vrcholů navzájem propojeno (případně přes prostředníka, u kterého by šlo vyrobit stupeň dva a pak jej odmazat)? Všimneme si, že není přímé spojení pravého dolního a levého horního vrcholu. Jsou spojeny nepřímo mnoha cestami, ale ty, které vedou přes ostatní potencionální vrcholy z K_5 nepomohou, protože dotyčné vrcholy při výrobě kopie nebudeme odmazávat. Zbývá tedy cesta oklikou přes pravý horní vrch. Dokázali bychom z grafu, jak jej máme nakreslen, odebrat některé nevybarvené vrcholy tak, aby pravý horní vybarvený vrchol náhle měl stupeň dva a tím šel odmazat, vznikla by tak hledaná spojinice? Nejde to, protože z pravého horního vrcholu vedou čtyři hrany právě do těch vrcholů, ze kterých chceme vybudovat K_5 .

Závěr je, že pokud budeme uvažovat vrchol vlevo dole a další čtyři vybarvené, se kterými je spojen, tak z toho nevznikne K_5 , neboť u dalších dvou zúčastněných vrcholů nelze vyrobit přímou spojku.

Symetricky to platí i o pravém dolním vrcholu, čímž, odpadají dva, zbývají čtyři kandidáti a z těch K_5 neupečeme. Závěr: Dotyčný graf neobsahuje kopii K_5 (případně doplněnou vrcholy na hranách).

Tedž ještě potřebujeme zjistit, zda se v grafu neschovává $K_{3,3}$. Necháme to na čtenáři, učitě má dost času si takhle hrát. Náš šílený vynálezce tohle vůbec nedělal a místo toho ubastil úžasnou kouli z drátů, ze které do stran trčely rozličné sučástky. I to má svůj půvab.

△

Z příkladu vidíme, že Kuratowského věta je sice pěkná, ale v praxi to zase až tak úžasné není. Matematici samozřejmě nezaháleli a vymýšleli i další kritéria rovinnosti, některá ekvivalentní, jiná alespoň ve formě implikace. Zajímavé je například toto:

Fakt 12d.3.

Jestliže je konečný graf $G = (V, E)$ rovinný a $|V| > 3$, pak $|E| \leq 3|V| - 6$.

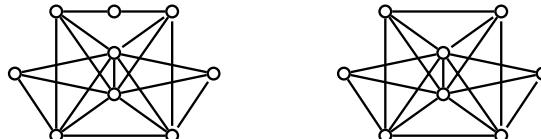
Obměna říká, že když $|E| > 3|V| - 6$, tak už dotyčný graf nutně není rovinný. Jinými slovy, rovinné grafy nemohou mít moc hran. Pokud použijeme princip sudosti, tak se nerovnost z tvrzení dá přepsat jako

$$\frac{1}{|V|} \sum_{v \in V} \deg(v) \leq 6 - \frac{12}{|V|}.$$

Nalevo máme průměrný stupeň vrcholu v grafu. Chceme-li tedy dostat graf rovinný, tak v případě více než 12 vrcholů máme povolen jen průměrný stupeň do pěti, u menších grafů méně (například v rovinném grafu o šesti vrcholech je povolený průměrný stupeň maximálně 4).

Příklad 12d.c (pokračování 12d.b): Jak je na tom nás šílený vynálezce? Tam $|V| = 13$ a $|E| = 26$, podmínka z Faktu je splněna a o rovinnosti nevíme nic. Nápad: Graf jsme si zredukovali vynecháním vrcholů s nízkým stupněm (obrázek výše vlevo), čímž mohl průměrný stupeň vzrůst. Což takhle aplikovat podmínku na něj? U zmenšeného grafu máme $|V| = 10$ a $|E| = 24$, to je sice na hranici, ale pořád nic nevíme.

Lze graf dále zmenšit? O vrcholu nahoře jsme odvodili, že jej lze při hledání K_5 odebrat z grafu. Rozmyslíme si, že jej lze odebrat i při hledání $K_{3,3}$. Již jsme viděli, že pokud jej odebereme, tak nepřerušíme žádná existující spojení mezi ostatními vrcholy. Mohl by být součástí hledaného $K_{3,3}$? Pokud by byl, řekněme v horní řadě, tak by ty tři vrcholy pod ním musely tvořit dolní řadu. To ale není možné, protože prostřední z nich by musel být napojen na horní vrchol a ještě na nějaké dva další, které by také měly přijít do horní řady, ale on není, má spojnice jen s kolegy z dolní řady. Proto můžeme odebrat horní vrchol, aniž bychom změnili rovinnost grafu, následně pak z hrany vymažeme ten nahoře uprostřed.



Vyšlo to, nový graf má $|V| = 8$ a $|E| = 19$, tudíž podmínka rovinnosti není splněna a nejde o rovinný graf. Ale bylo to o chlup, stačilo o hranu méně a zase jsme nic nevěděli.

△

Hlavní problém je, že dotyčné tvrzení je jen implikace, takže rovinnost umí vyvrátit, ale ne potvrdit.

Vzhledem k tomu, že jde jen o přehledovou kapitolu, se v tom dálé nebude vrtat, na závěr čtenáře uklidníme, že existují algoritmy, které dokážou rovinnost grafu rozhodnout, a to dokonce v relativně přijemném lineárním čase, jejich výpočetní náročnost je $O(|V|)$ pro grafy jednoduché, obecně pak $O(|V| + |E|)$, neboť opakováním hran se obtížnost úlohy zjevně zvyšuje. S rovinnými grafy se ještě setkáme v kapitolce o barvení grafů.

Na závěr si neodpustím bonbónek. Podle Fáryho věty lze každý rovinný graf nakreslit tak, že se hrany nejenž neprotínají, ale dokonce jsou to úsečky.

12e. Barvení grafu

Po Eulerově procházkách teorie grafů nějakou dobu spíš odpočívala a nový rozvoj přišel až v druhé polovině 19. století. Jedním z hlavních problémů, které podněcovaly zkoumání v tomto oboru, byla otázka pocházející z kartografie. Autoři map (politických) rádi vybarvují země tak, aby dvě sousední měly vždy jinou barvu, ale z praktických důvodů se snaží udržet počet barev minimální. Kolik barev by stačilo na jakoukoliv mapu? Pokud každou zemi reprezentujeme vrcholem grafu a to, že spolu sousedí, zachytíme doplněním hrany mezi nimi, tak se to převede na úlohu z teorie grafů.

Definice.

Nechť $G = (V, E)$ je konečný rovinný (neorientovaný) graf, $k \in \mathbb{N}$. Řekneme, že graf G lze obarvit k barvami, jestliže existuje funkce $f : V \mapsto \{1, 2, \dots, k\}$ taková, že pro každé $u, v \in V$ platí: Jestliže $\{u, v\} \in E$, pak $f(u) \neq f(v)$.

Klíčová otázka tedy zní: Jaké je nejmenší číslo k takové, že libovolný konečný rovinný graf lze obarvit k barvami? Při pohledu na K_3 vidíme, že alespoň tři barvy potřebujeme vždy. Rovněž K_4 je rovinný graf, jak jsme již viděli výše, což ukazuje, že některé grafy potřebují čtyři barvy. Optimální hodnota k_0 tedy určitě splňuje $k_0 \geq 4$. Nás ale spíš zajímá omezení shora.

Již na konci 19. století bylo dokázáno, že $k = 5$ stačí pro všechny grafy. Naznačíme důkaz, protože jinak by tato kapitola zůstala bez pořádného důkazu a to by čtenáře jistě mrzelo. Navíc mám k tomuto důkazu osobní vztah, při ústní zkoušce z teorie množin jsem dostal za úkol dokázat právě následující větu.

Věta 12e.1. (o pěti barvách)

Pro každý konečný rovinný graf existuje obarvení pěti barvami.

Důkaz (poučný): Důkaz provedeme silnou indukcí na počet vrcholů, dokazujeme tedy pro všechna $n \in \mathbb{N}$ tvrzení $V(n)$: Každý rovinný graf s n vrcholy lze obarvit pěti barvami.

(0) Nechť $n \in \mathbb{N}$, $n \leq 5$. Jestliže má graf nejvýše 5 vrcholů, tak prostě každý obarvíme jinou barvou.

(1) Nechť $n \in \mathbb{N}$, $n \geq 5$, předpokládejme, že všechny rovinné grafy s nejvýše n vrcholy dokážeme obarvit pěti barvami. Potřebujeme ukázat, že totéž platí i pro všechny grafy s $n + 1$ vrcholy.

Vezměme tedy graf $G = (V, E)$ splňující $|V| = n + 1$. Pokud by všechny vrcholy měly stupeň větší než 5, tak $|E| = \frac{1}{2} \sum_{v \in V} \deg(v) \geq \frac{1}{2} 6|V| = 3|V|$, ale pro planární grafy má platit $|E| \leq 3|V| - 6$. Proto musí existovat nějaký vrchol v , jehož stupeň je nejvýše 5. Rozebereme dva případy.

a) Jestliže $\deg(v) \leq 4$, pak uvažujeme graf $G_0 = (V_0, E_0)$, který vznikl z grafu G odebráním vrcholu v . Protože $|V_0| = n$, lze podle indukčního předpokladu tento graf obarvit pěti barvami. Vrchol v ale sousedí bezprostředně s nejvýše 4 vrcholy, proto musí existovat barva, kterou tyto sousední vrcholy nevyužívají. Když touto barvou obarvíme vrchol v , nevznikne tak situace sousedících stejně obarvených vrcholů a celý graf je obarven pěti barvami.

b) Jestliže $\deg(v) = 5$, pak má v pět bezprostředních sousedů. Pokud bychom nyní vrchol v odebrali a použili indukci k obarvení zbytku grafu jako v části a), tak by se mohlo stát, že se u sousedů v objeví všech pět barev a na v bychom tedy potřebovali šestou. Tomu je třeba zabránit.

Podívejme se blíže na množinu M těchto sousedů. Pokud by všechny dvojice prvků z M byly spojeny hranou, tak vlastně dostáváme graf K_5 . Jenže G nemůže mít K_5 jako podgraf, protože by tím pádem nebyl rovinný. Musí tedy existovat dva vrcholy $u_1, u_2 \in M$, které nejsou v grafu G spojeny hranou. V takovém případě je možné začít předstírat, že u_1 a u_2 je vlastně jeden vrchol. Formálně to dá trochu práce, ale není to nic pokročilého jen se musí nadefinovat chytře nový graf, z původního G se odeberou v, u_1, u_2 a pak se tam přidá vrchol u , který zastupuje oba u_1, u_2 , a dále hrany do u jako zástupci všech hran, které původně vedly do u_1 a u_2 . Pak se (snadno ale pracně) ukáže, že vznikl nový rovinný graf.

Ten má $n - 1$ vrcholů, podle indukčního předpokladu jej tedy lze obarvit pěti barvami. Tyto barvy pak přeneseme na graf G_0 , přičemž vrcholy u_1 a u_2 obarvíme stejně jako u . Jsme tedy v situaci, kdy je pět sousedů vrcholu v obarveno, ale dva z nich stejně, tudíž se spotřebovaly jen čtyři barvy a pátá zbývá na v .

Tím končí důkaz kroku (1), ukázali jsme, jak obarvit libovolný graf s $n + 1$ vrcholy. □

Na konci 19. století tedy bylo jasné, že správné odpovědi jsou možné jen dvě, $k_0 = 4$ nebo $k_0 = 5$. Nicméně na základě zkušeností si matematici již od poloviny 19. století sázeli na $k_0 = 4$, ale nebyli schopni to dokázat, říkalo se tomu hypotéza čtyř barev. Zlom přišel v roce 1976.

Věta 12e.2. (o čtyřech barvách)

Každý konečný rovinný graf lze obarvit čtyřmi barvami.

Důkaz neukážeme ani náhodou, dokonce to matematické komunitě trvalo dost dlouho, než jej uznala jako správný. Autoři dotyčné věty totiž použili do té doby nevýdanou kombinaci matematiky a počítačů. Nejprve pomocí různých teoretických úvah zúžili pole působnosti, dokázali, že pokud dokážeme obarvit čtyřmi barvami určitých 1936 grafů, tak už budeme umět takto obarvit všechny rovinné grafy. (Tato matematická část zabraala přes 400 stran, tak si

čtenář jistě umí představit, že trvalo docela dlouho, než matematici zkontovali, že se někde neskrývá chybka. Několik jich našli, ale byla to malá přehlédnutí, která šlo spravit.)

Na pomoc při obarvování těch 1936 grafů pak zavolali počítač, kterému to trvalo přes 1200 hodin, což je mimořadně asi 50 dnů. Bylo to mimo jiné dáním tím, že na hledání obarvení čtyřmi barvami používali algoritmus s náročností řádově n^4 . Od té doby byl vyvinut algoritmus s náročností n^2 a počet grafů k ověření se podařilo snížit na 633, takže samotné provedení důkazu už není otázka týdnů.

Dodejme, že ty 4 barvy jsou nejhorší scénář, ale pro mnohé grafy by šlo vystačit i s méně barvami. To je jiná otázka: Je-li dán rovinný graf, jaký je nejmenší počet barev nutný k jeho obarvení? Pokud to budeme chtít vyřešit algoritmicky, dostaneme další náročnou úlohu, je to NP-úplný problém.

Podobné problémy lze zkoumat i pro jiné grafy než rovinné (grafy obecně, grafy nakreslené bez protnutí na nekonečném válci či toru a podobně) a pak to začne být dobrodružné. Obecně se dá pro každé $k \in \mathbb{N}$ vyrobit graf takový, že vyžaduje alespoň k barev. To není nic těžkého, stačí vzít K_n , zajímavé je, že na to stačí i grafy s relativně méně hranami.

12f. Stromy, kostra grafu

Zatím jsme spíš měli situaci, kdy je graf dán a my jej zkoumáme, ale často také čelíme problému, že máme vytvořit graf, který má splňovat určité požadavky. Představme si třeba plánovače železničního spojení, dálniční sítě či také počítačové sítě (pokud šetříme a nechceme tam mít redundanci pro zvýšení bezpečnosti). Ti všichni mají dané objekty (vrcholy), které chtejí navzájem propojit (tedy má vzniknout souvislý graf), ale z pochopitelných důvodů chtejí těch propojení co nejméně. Jak poznáme, že hranami plýtváme? Když se mezi dvěma body dokážeme dostat dvěma různými cestami. Když tyto dvě cesty spojíme, dostáváme kružnici. To tedy nechceme.

Definice.

Nechť G je graf.

Řekneme, že je to **strom (tree)**, jestliže je souvislý a neobsahuje žádnou kružnici.

Řekneme, že je to **les (forest)**, jestliže neobsahuje žádnou kružnici.

Dá se dokázat, že graf je lesem, jestliže je každá z jeho komponent souvislosti stromem. V zásadě tedy stačí studovat stromy.

Všimněte si, že v definici mluvíme o kružnicích, což jsou neorientované struktury, dokonce i v případě, že graf G je orientovaný. Podobně pojmem souvislosti nebude ohled na orientaci hran, proto u orientovaných grafů rozhoduje jejich neorientovaná verze o tom, zda jsou stromy.

Následující věta potvrdí, že pojmem stromu vystihuje naše přání mít ekonomický graf.

Věta 12f.1.

Nechť $G = (V, E)$ je graf. Následující tvrzení jsou ekvivalentní:

- (i) G je strom;
- (ii) každé dva vrcholy v G jsou spojeny právě jednou cestou;
- (iii) G je souvislý, ale po odebrání libovolné hrany už bude vzniklý graf nesouvislý (strom je tzv. minimální souvislý graf);
- (iv) G neobsahuje kružnici, ale přidáním libovolné hrany v něm již kružnice vznikne (strom je tzv. maximální graf bez kružnic);
- (v) G je souvislý a $|E| = |V| - 1$;
- (vi) G nemá kružnice a $|E| = |V| - 1$.

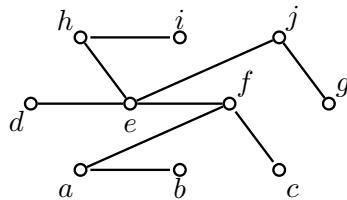
Připomeňme, že zde pracujeme s konečnými grafy, v této souvislosti je zajímavé, že podmínky (i) až (iv) by fungovaly i pro grafy nekonečné.

Věta ukazuje jednu zajímavou věc. Pokud chceme, aby byl graf souvislý, tak jej nutíme mít více hran. Pokud nechceme kružnice, pak by zase hran mělo být relativně málo. Dá se tedy čekat, že existuje jakási zóna optimálního počtu hran. Věta ukazuje, že tato zóna je velice úzká, je to pro daný graf jedno konkrétní číslo.

Ukážeme si to podrobněji, zároveň tím naznačíme důkaz části této věty. Představme si souvislý graf, který nemá kružnice. Navíc si představme, že je to přesně na hranici, tedy vypali jsme jej hranami tak hodně, jak jen to je možné, aniž by vznikla kružnice. Ukážeme, že jsme zároveň na hranici souvislosti, tedy odebráním jediné hrany už by se souvislost ztratila. Uděláme to sporem.

Co kdyby šlo odebrat nějakou hranu $\{u, v\}$ tak, aby graf ještě zůstal souvislý? Pak by nutně musela existovat cesta z u do v , do té když doplníme tu hranu $\{u, v\}$, tak máme kružnici, což je ve sporu s naším předpokladem. Graf je tedy také minimální možný pro souvislost.

Zde máme příklad stromu, rozmyslete si na něm platnost podmínek (ii) až (vi) z věty.



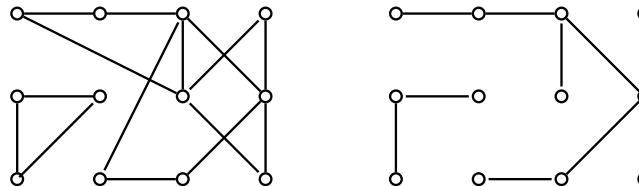
Jednou ze zajímavých úloh je, že nám někdo dá graf, který trochu plýtvá, a my z něj chceme odstraněním hran vytvořit graf, který je ekonomický, aniž by ztratil ze své propojenosti. Takovouto úlohu řeší úřady, když se snaží zrušit tratě, aniž by se snížila dopravní dostupnost jednotlivých zastávek.

Jak tuto úlohu formulovat matematicky? Máme graf G a chceme najít jeho podgraf G' , který by měl stejné vrcholy jako graf původní a přitom by neměl zbytečné hrany, tedy neměl by kružnice. Zároveň by mělo platit, že jsou-li v grafu G nějaké vrcholy propojeny cestou, pak tomu tak musí být i v grafu G' . Obvykle se toto řeší jinak, rovnou se zaměříme na jednotlivé komponenty původního grafu, pak je požadavek, aby z původně souvislého grafu zase vznikl souvislý.

Definice.

Nechť $G = (V, E, \varepsilon)$ je souvislý graf. Řekneme, že jeho podgraf $G' = (V', E', \varepsilon)$ je **kostra (spanning tree)** grafu G , jestliže $V' = V$ a G' je strom.

Příklad grafu a jeho kostr jeho komponent:



Nalezení kostry je snadné, prostě postupně přibíráme hrany z původního grafu, dokud je to možné udělat tak, aby nevznikla kružnice. Je zjevné, že kostra grafu není jezdnoznačně určena, například je-li přímo dána kružnice jako výchozí graf, pak její kostru získáme odebráním libovolné z hran.

Zajímavější problém je, když jsou hrany ohodnoceny, pak se často hledá kostra s co nejmenší cenou (součtem ohodnocení hran). Na to jsou algoritmy, jednoduché nás stojí zhruba $|V|^2$, ale jde to i lépe.

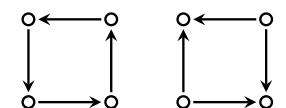
12f.2 Kořenové stromy

Začneme grafem orientovaným.

Definice.

Nechť $G = (V, E, \varepsilon)$ je orientovaný graf. Řekneme, že vrchol $v \in V$ je jeho **kořen (root)**, jestliže pro každý vrchol $w \in V$ existuje orientovaná cesta z v do w .

Protože vždy existuje cesta z v do v , nenutí nás tato definice mít u v smyčku. Je dobré si uvědomit, že kořenů může být v grafu více. Extrémní příklad je cyklus, kde je kořenem každý z vrcholů. Naopak je možné, že žádný kořen neexistuje (viz obr. vpravo).



Je zjevné, že pokud je v grafu kořen, pak již graf musí být souvislý (ale nemusí být silně souvislý). Naplatí to naopak, ne každý souvislý graf má kořen, viz obrázek výše vpravo. My se teď vrátíme k hlavnímu tématu.

Definice.

Nechť $G = (V, E, \varepsilon)$ je orientovaný graf. Řekneme, že je to **kořenový strom (rooted tree)**, jestliže je to strom a existuje v něm kořen.

Připomeňme, že existence kořene již implikuje souvislost, což je jedna z podmínek pro strom. Lze proto také říct, že kořenový strom je libovolný orientovaný graf s kořenem a bez kružnic.

Před chvílí jsme viděli, že obecně může být v souvislém orientovaném grafu i více kořenů. U stromů to ale neplatí, buď je kořen žádný, nebo je jen jeden.

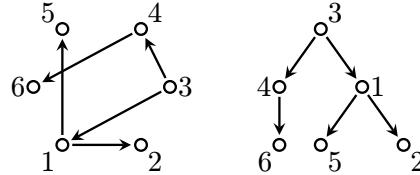
U kořenových stromů se zavádí užitečné názvosloví. U orientované hrany (x, y) říkáme prvku x **rodič (parent)**, zatímco prvku y říkáme **potomek** či **dítě (child)**. Pokud nějaký prvek u nemá potomka, pak mu říkáme **list**.

(leaf). Každé cestě vedoucí od kořene k nějakému listu říkáme **větev (branch)**. Konce takovýchto větví poznáme podle stupňů: Listy se poznají podle toho, že $\deg^+(v) = 0$. Kořen je jediný vrchol, který má $\deg^-(v) = 0$, ostatní vrcholy mají $\deg^-(v) = 1$.

V zakořeněném stromu vzniká relace mezi vrcholy grafu. Řekneme, že vrchol x je **předchůdce** vrcholu y nebo že y je **následovník** x , jestliže ve stromu G existuje cesta z x do y . Je snadné ukázat, že relace „být následovník“ i relace „být předchůdce“ jsou částečná uspořádání.

My už jsme vlastně viděli i opačný proces. Pro částečná uspořádání jsme vytvářeli Hasseovy diagramy. Pokud bychom u hran v takovém diagramu ponechali orientaci, pak by šlo o stromy, dokonce jsem tak vlastně nacházeli kostry grafů daných relací. Jinak řečeno, náš algoritmus pro vytváření Hasseova diagramu lze snadno přepsat na algoritmus sloužící k nalezení kostry daného souvislého orientovaného grafu.

Hasseovy diagramy také ukazují další věc, která je u stromů tradiční, rádi je kreslíme tak, aby byl kořen nejvýše a všechny hrany směrovaly svou orientací dolů, popřípadě naopak (dle konkrétní aplikace).



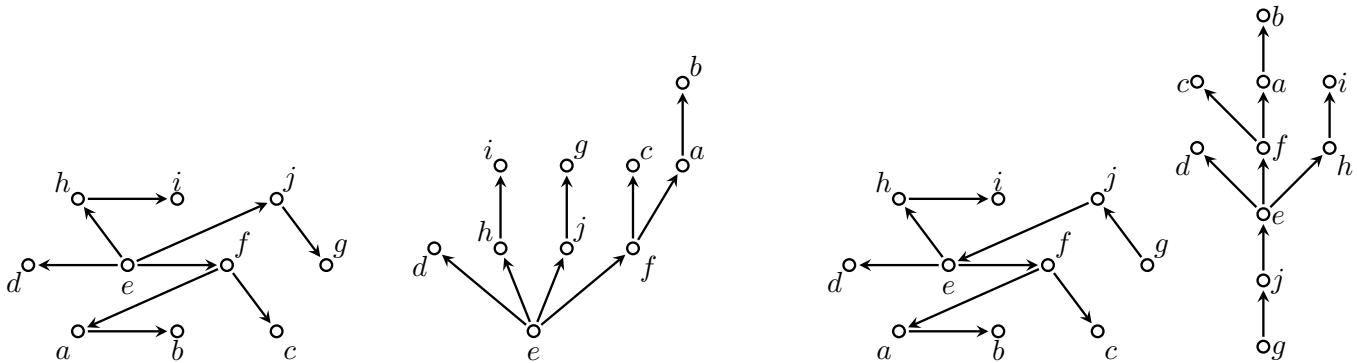
Stromy jsou velice užitečné například jako modely datových struktur, kdy vrcholy nesou informace, pak se zkoumají různé algoritmy pro prohledávání zakořeněného stromu (do šířky, do hloubky).

Zajímavé je, že i neorientované stromy je možné přeměnit v kořenové stromy, přičemž svoboda volby není zase tak velká. Spočívá v tom, že si v takovém stromu zvolíme libovolný vrchol jako kořen. Pak už existuje jen jeden způsob, jak na hranách zavést orientaci, aby vznikl orientovaný strom.

Jak se to dělá? Víme, že u orientovaného stromu z kořene hrany jen vycházejí, tedy pro libovolné $\{v, x\} \in E$ dostaneme orienovanou hranu (v, x) . Pak se podíváme na množinu sousedů M kořene v . Pro souseda $w \in M$ si u všech jeho hran $\{w, x\}$ zavedeme orientaci (w, x) , ovšem s výjimkou hrany, která vede do v a již orientaci má. Následně se podíváme na sousedy prvků, které jsme právě zpracovávali, a dodáme další orientace směrem k novým sousedům a tak dále.

Nejlepší přestava je, že hrany jsou trubky. My do kořene v pustíme vodu. Protože je strom souvislý, voda doteče do každého vrcholu. Protože ale máme strom, tak se do žádného vrcholu nemůže dostat více cestami, následně voda nemá v žádné hraně na výběr a plyne vždy jen jedním směrem, který udává orientaci.

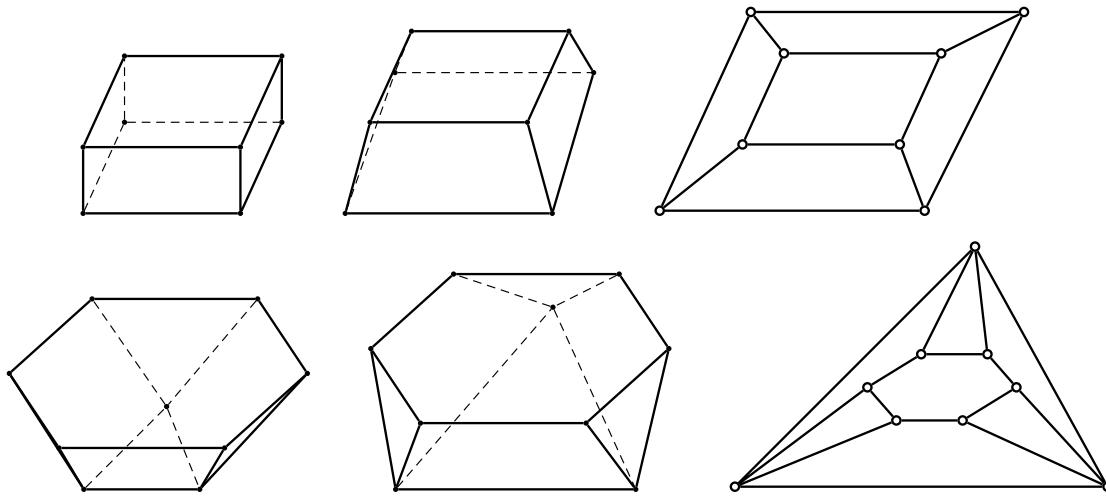
Coby příklad použijeme náš první strom, jako kořen jsme si zvolili e a vidíme jej dvakrát, nejprve v původní pozici a pak kořenem dole. Další dva obrázky ukazují, co se stane, když zvolíme jako kořen vrchol g .



Dá se říct, že neorientované stromy lze zakořenit.

12g. Bonus: Platónovská tělesa

Platónovská tělesa jsou mnohostěny, které jsou pravidelné, jak jen to jde. Jmenovitě, zvolí se konstanty p, q a požaduje se, aby každá stěna byla pravidelný p -úhelník a v každém vrcholu se scházelo q hran. Například klasická krychle má parametry $p = 4$ a $q = 3$. Již staří Řekové znali souvislost mezi tělesy a rovinnými grafy. Když těleso položíme jednou stěnou na podložku a hrany té stěny začneme roztahovat do stran, pak při pohledu shora dříve či později vidíme rovinný graf, jehož stěny odpovídají stěnám příslušného tělesa. Ta stěna, která u tělesa ležela na podložce, se stala tou neohraničenou „vnější“ stěnou v grafu (vidíme teď inspiraci, proč jsme ji u rovinních grafů započítávali). Ukážeme popsáný postup pro dvě tělesa, na která se díváme šikmo shora, nejprve klasický hranol, pak zajímavější objekt.



Trocha aritmetiky a Eulerův vzorec pro rovinné grafy ukáže, že není moc možností, jak Platónovská tělesa vytvářet.

V rovinném grafu pracujeme s počtem vrcholů v , počtem hran e a počtem stěn f . Co víme? Eulerova rovnost dává $v + f = 2 + e$. Každá stěna má p stran, když to sečteme, bereme všechny strany neboli hrany dvakrát, proto $fp = 2e$. Každý vrchol má dle zadání stupeň q , součet stupňů vrcholů je tedy $vq = 2e$. Máme tři rovnice a pět neznámých, tudíž je můžeme například vyřešit takto:

$$\begin{aligned} v &= \frac{4p}{4 - (p - 2)(q - 2)}, \\ e &= \frac{2pq}{4 - (p - 2)(q - 2)}, \\ f &= \frac{4q}{4 - (p - 2)(q - 2)}. \end{aligned}$$

Pak lze udělat rozbor, pro které celočíselné hodnoty p, q dostáváme rozumné výsledky. Zde je dobré si uvědomit, že již ze zadání máme přirozené dolní meze $p \geq 3, q \geq 3, v \geq 4, f \geq 4, e \geq 6$.

Některí autoři dávají přednost pomocí dvou rovností $v = \frac{2e}{p}$, $f = \frac{2e}{q}$ upravit Eulerův vztah na $\frac{1}{p} + \frac{1}{q} = \frac{1}{2} + \frac{1}{e}$ a pak udělat rozbor, které hodnoty $p, q \geq 3$ vedou na celočíselné řešení pro e . Není to těžké, z $e > 0$ dostáváme $\frac{1}{p} + \frac{1}{q} > \frac{1}{2}$, díky čemuž z omezení $p, q \geq 3$ dostáváme pouze možnosti $\{3, 3\}, \{3, 4\}, \{4, 3\}, \{3, 5\}$ a $\{5, 3\}$. Není tedy možné mít více než těchto pět Platónských těles. Zatím jsme nedokázali, že všech pět typů opravdu existuje v reálném světě, ale je to tak, jde o slavnou pětici známou už od antických Řeků.

Staří Řekové algebru moc nepěstovali a volili ještě jiný přístup. Podívali se na jeden vrchol a viděli, že se tam stýká q stěn, každá tam má svůj vrchol s určitým úhlem, tyto úhly jsou díky pravidelnosti stejné a je jich tedy q , dohromady nemohou dát ani 360 stupňů. Na jednu stěnu tedy připadne méně než 120° , což není mnoho. Jediné pravidelné p -úhelníky, které připadají v úvahu, jsou tedy $p = 3, 4, 5$ s úhly $60^\circ, 90^\circ, 108^\circ$. Sestúhelník už má úhel 120, to je moc.

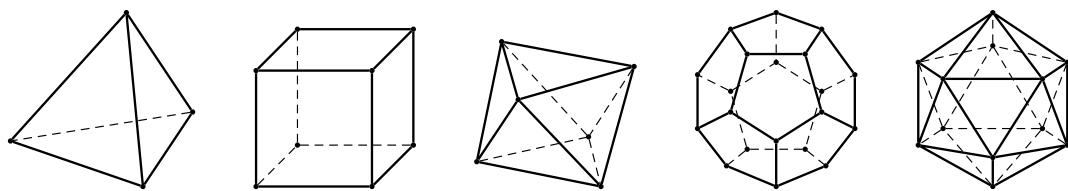
Jestliže $p = 5$, tak úhlů 108° se u jednoho vrcholu tělesa může sejít nejvýše $q = 3$, což je i minimální počet, takže není možné mít více než jedno Platónovské těleso složené z pětiúhelníků, jmenovitě to s $q = 3$, pak dopočítáme $v = 20, e = 30, f = 12$. Tím ještě není zaručeno, že takováto věc existuje, ale my víme, že ano, je to pravidelný dvanáctistěn.

Jestliže $p = 4$, tak čtyři pravé úhly nedají méně než 360, tudíž je zase možnost jen $q = 3$, odtud $v = 8, e = 12, f = 6$, jejen jediná potencionální možnost mít Platónovské těleso ze čtverců. Je to šestistěn neboli naše známá krychle.

Poslední možnost $p = 3$ je nejjednodušší, protože počet úhlů 60° , které se vyměstnají k jednomu vrcholu a tedy dají méně než 360, je $q = 3, 4, 5$. Není tedy možné mít více než tři Platónovská tělesa složená z trojúhelníků. I zde se ukáže, že tato tělesa v reálu existují. Uděláme si přehlednou tabulkou, kde tělesa seřadíme podle počtu stěn.

f	v	e	p	q
4	4	6	3	3
6	8	12	4	3
8	6	12	3	4
12	20	30	5	3
20	12	30	3	5

Obrázky následují. Hráči her na hrdiny je znají coby hrací kostky. Příznivci Rubikových hlavolamů znají více než polovinu tvarů také, sám mám doma čtyřstěn, kostku a dvanáctistěn.



Tím končí náš bonus, pěkná aplikace teorie grafů na geometrii těles.