Obsah

KAPITOLA 11

Algebraické struktury

čím větší abstrakce, tím větší zmatek?
– ne, často to bývá naopak ...

V této kapitole se budeme věnovat zdánlivě velice formálnímu studiu pojmů, které ale ve skutečnosti odráží spoustu skutečných vlastností věcí kolem nás.

Abstrahujeme z nich přitom jen ty nejjednodušší operace a „algebru“ tak lze vnímat jako algoritmické manipulace s písmeny, které zpravidla mají nějaké souvislosti s výpočty nebo popisem procesů. Zároveň si budeme trochu všímat, kde všude jsme takové objekty potkávali v předchozích kapitolách (aniž by ale bylo nutné mít tyto kapitoly předem pročtené). Přímo navážeme víceméně jen na první a šestou část první kapitoly, kde jsme podobně abstraktně pohlíželi na čísla, se kterými počítáme, a obecněji na vztahy mezi objekty, když jsme je abstrahovali do tzv. relací.

V první části této kapitoly se zastavíme u té nejjednodušší situace – budeme se zamýšlet nad případem, kdy máme jen jednu jedinou operaci, která se chová podobně jako násobení čísel. Pak si přídáme druhou operaci, podobně jako jsou u čísel k dispozici společně sčítání a násobení. To nám umožní vysvětlit elementární základy tzv. počítačové algebry, tj. algoritmických postupů, díky kterým počítače umí manipulovat s formálními výrazy a počítat s nimi, včetně řešení systémů polynomiálních rovnic.

V další části se vrátíme k jiné abstrakci situací s jedinou operací a budeme přitom vycházet z uspořádní čísel podle velikosti nebo množinové inkluze. V poslední části kapitoly se pak zastavíme u několika poznámek ohledně využití algebraických nástrojů pro návrhy (samoopravných) kódů využívahých hojně při přenosech dat.

1. Grupy

Naše první úvahy se budou týkat objektů a situací, ve kterých je možné rovnice tvaru $a \cdot x = b$ vždy jednoznačně řešit (tak jako u lineárních rovnic jsou objekty a a b dány, zatímco x hledáme). Půjde o tzv. teorii grup. Všimněme si, že zatím nic nevíme o povaze objektů, ani co znamená ta tečka. Jen předpokládáme, že dvěma objektům a a x umíme přiřadit objekt $a \cdot x$.

Nejprve si oprášíme a rozšíříme náš slovník pojmů ohledně operací, jak jsme jej zavedli již v kapitole první a projdeme přitom příklady čísel a transformací roviny a prostoru, ve kterých se s takovými „grupovými“ objekty potkáváme. Teprve pak se budeme chvíli věnovat základům obecné teorie.

10.1

11.1. Příklady a pojmy. Pro libovolnou množinu A jsme již dříve definovali *binární operace* na A jako libovolné zobrazení $A \times A \to A$. Výsledek takové operace budeme často značit

$$(a, b) \mapsto a \cdot b.$$

Množina s binární operací se nazývá *grupoid*.

Abychom mohli něco podstatného říci, potřebujeme nějaké další vlastnosti operací. Binární operace je *asociativní*, jestliže pro všechny prvky v A platí

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

BINÁRNÍ OPERACE A POLOGRUPY

Grupoid s asociativní binární operací, se nazývá *pologrupa*. Binární operace je *komutativní*, jestliže pro všechny prvky v A platí $a \cdot b = b \cdot a$.

Přirozená čísla $\mathbb{N} = \{0, 1, 2, \dots\}$, spolu s kteroukoliv z operací sčítání a násobení jsou asociativní a komutativní pologrupa. Celá čísla $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ jsou grupoid vůči kterékoliv z operací sčítání, odčítání, násobení. Operace odčítání ale není asociativní. Např.

$$(5 - 3) - 2 = 0 \neq 5 - (3 - 2) = 4,$$

ani komutativní, protože $a - b = -(b - a)$.

JEDNOTKY, INVERZE A GRUPY

Levá jednotka v grupoidu $(A, \cdot)$ je takový prvek $e \in A$, že pro všechny prvky v A platí $e \cdot a = a$; obdobně pro *pravou jednotku* musí platit pro všechny prvky $a \cdot e = a$. *Jednotka* binární operace je prvek e, který je pravou i levou jednotkou zároveň.

Prvek a^{-1} je *levou inverzí* k prvku a v pologrupě $(A, \cdot)$ s jednotkou e, jestliže platí $a^{-1} \cdot a = e$; obdobně je *pravou inverzí* a^{-1} takový prvek, pro který je $a \cdot a^{-1} = e$.

Prvek a^{-1} je *inverzní* k a v pologrupě s jednotkou, jestliže je levou i pravou inverzí zároveň.

Monoid $(M, \cdot)$ je pologrupa s jednotkou. *Grupa* $(G, \cdot)$ je pologrupa s jednotkou, ve které má každý prvek inverzi.

Komutativní grupa, resp. *komutativní pologrupa*, je taková, kde je operace $\cdot$ komutativní.

Komutativní grupy se také často nazývají *abelovské*.[1]

Podívejme se na přímé jednoduché důsledky definic. V monoidu nemohou být pravé a levé inverze různé. Je-li totiž $a \cdot x = x \cdot b = e$, pak také

$$a = a \cdot (x \cdot b) = (a \cdot x) \cdot b = b.$$

Podstatná je zde pouze asociativita operace. Všimněme si, že pro odečítání na celých číslech (tady operace není asociativní) je nula *pravou jednotkou*, tj. $a - 0 = a$ pro všechna

[1] Je to na počest mladého matematika Abela ... V angličtině se používá přídavné jméno „abelian“ a bývá to uváděno jako příklad absolutní pocty, protože se píše s malým písmenem, tzn. je to tak obecně používáno, že se již zapomnělo, že jde o jméno člověka.

celá čísla a, není však levou jednotkou. Dokonce v tomto případě levý neutrální prvek neexistuje.

Celá čísla jsou zjevně pologrupou vůči sčítání i násobní. Grupou jsou přitom jen vůči sčítání, protože pro násobení neexistují inverzní prvky, kromě čísel ± 1.

Je-li $(A, \cdot)$ grupa, pak její podmnožinu $B \subset A$, která je uzavřená vůči zúžení operace $\cdot$ a zároveň je spolu s touto operací grupou, nazýváme *podgrupa*.

Racionální čísla $\mathbb{Q}$ jsou komutativní grupou vzhledem ke sčítání a nenulová racionální čísla jsou také komutativní grupou vůči násobení. Celá čísla spolu se sčítáním jsou jejich podgrupou.

Pro každé kladné přirozené číslo k je množina všech k-tých odmocnin z jedničky, tj. množina $\{z \in \mathbb{C};\ z^k = 1\}$ konečnou grupou vůči násobení komplexních čísel. Např. pro $k = 2$ dostaneme grupu $\{-1, 1\}$ se dvěma prvky, které jsou oba samy sobě inverzí, zatímco pro $k = 4$ dostáváme grupu $G = \{1, i, -1, -i\}$.

Množina Mat_n, $n > 1$, všech čtvercových matic je (nekomutativní) pologrupa vzhledem k násobení matic a komutativní grupa vzhledem ke sčítání matic (viz odstavce 2.2–2.5).

Množina všech lineárních zobrazení $\mathrm{Hom}(V, V)$ na vektorovém prostoru je pologrupa vzhledem ke skládání zobrazení a komutativní grupa vzhledem ke sčítání zobrazení (viz odstavec 2.34).

V obou předchozích příkladech, podmnožina invertibilních objektů uvažované pologrupy tvoří grupu. V prvním případě jde o tzv. grupu invertibilních matic, ve druhém o grupu lineárních transformací vektorového prostoru.

V dřívějších kapitolách jsme již potkali mnoho (polo)grupových struktur, občas asi i docela nečekaně. Vzpomeňme např. různé podgrupy grupy matic nebo grupovou strukturu na eliptických křivkách.

10.3

11.2. Grupy permutací. Velmi často grupy a pologrupy potkáváme jako množiny zobrazení na pevně dané množině M, které jsou uzavřeny vůči skládání zobrazení. Ne vždy si ale tuto skutečnost přímo uvědomujeme, protože vidíme jen některá zobrazení a na všechna ostatní vznikající složeními nemyslíme.

Nejsnáze je tato souvislost vidět na konečných množinách M, kde nám každá podmnožina invertibilních zobrazení vygeneruje pomocí skládání jistou grupu.

Na každé takové množině o $m = |M| \in \mathbb{N}$ prvcích (prázdná množina má 0 prvků) totiž máme k dispozici m^m možných definic zobrazení (každý z m prvků můžeme zobrazit na kterýkoliv v M) a všechna taková zobrazení umíme skládat. Protože skládání zobrazení je samozřejmě asociativní operace, dostáváme grupoid.

Pokud chceme, aby existovala k zobrazení $\alpha : M \to M$ jeho inverze α^{-1}, musí být α bijekcí. Složením dvou bijekcí vznikne opět bijekce a proto podmnožina Σ_m všech bijekcí na

množině M o m prvcích je grupa. Říkáme jí *grupa permutací* (na m prvcích). Je příkladem konečné grupy.[2]

Sám název grupy Σ_m přitom uvádí jinou souvislost, kdy místo bijekcí na konečné množině vnímáme permutace jako přerovnání rozlišitelných prvků. Potkávali jsme se s permutacemi v tomto smyslu např. při studiu determinantů, viz odstavec 2.14 na straně 58.

Promysleme si podrobněji, jak vlastně násobení v takové grupě vypadá. U (malé) konečné grupy si můžeme snadno sestavit úplnou tabulku všech operací. Jestliže v grupě permutací Σ_3 na číslech $\{1, 2, 3\}$ označíme jednotlivá pořadí

$$a = (1, 2, 3),\ b = (2, 3, 1),\ c = (3, 1, 2),$$
$$d = (1, 3, 2),\ e = (3, 2, 1),\ f = (2, 1, 3),$$

pak skládání našich permutací je zadáno tabulkou

$\cdot$	a	b	c	d	e	f
a	a	b	c	d	e	f
b	b	c	a	f	d	e
c	c	a	b	e	f	d
d	d	e	f	a	b	c
e	e	f	d	c	a	b
f	f	d	e	b	c	a

Všimněme si podstatného rozdílu mezi permutacemi a, b a c a dalšími třemi. Ty první tři tvoří tzv. *cyklus* generovaný prvkem b nebo prvkem c:

$$b^2 = c,\ b^3 = a,\ c^2 = b,\ c^3 = a.$$

Samy o sobě jsou tyto tři prvky komutativní podgrupou. V této podgrupě je a jednotka a prvky b s c jsou vzájemně inverzní. Je tedy tato podgrupa stejná jako je grupa $\mathbb{Z}_3$ zbytkových tříd celých čísel modulo 3, resp. jako grupa třetích odmocnin z jedničky.

Další tři prvky jsou samy sobě inverzí a každý z nich je tedy společně s jednotkou a podgrupou stejnou jako je $\mathbb{Z}_2$. Říkáme, že b a c jsou *prvky řádu* 3, zatímco prvky d, e a f jsou řádu 2.

Tabulka ale není symetrická podle diagonály, naše operace $\cdot$ tedy není komutativní.

Obdobně se chovají všechny grupy permutací Σ_m konečných množin o m prvcích. Každá permutace σ rozkládá množinu M na disjunktní sjednocení maximálních invariantních podmnožin, které dostaneme tak, že postupně vybíráme dosud nezpracované prvky $x \in M$ a do třídy rozkladu M_x přidáváme všechny akce iterací $\sigma^k(x)$, $k = 1, 2, \ldots$, dokud není $\sigma^k(x) = x$. Každou permutaci tak dostáváme jako složení jednodušších permutací, tzv. cyklů, které se chovají jako identická permutace vně M_x a tak jako σ na M_x. Pokud přitom očíslujeme prvky v M_x jako pořadí $(1, 2, \ldots, |M_x|)$ tak aby i odpovídalo $\sigma^i(x)$, pak je naše permutace prostým posunutím o jednu pozici v cyklu (tj. poslední prvek je

[2] A lze dokázat, že každá konečná grupa je podgrupou ve vhodné konečné grupě permutací. To si můžeme interpretovat tak, že grupy Σ_m jsou tak nekomutativní a složité, jak to jen jde.

zobrazen zpátky na první). Odtud název *cyklus*. Zjevně přitom tyto cykly komutují, takže je jedno, v jakém pořadí z nich permutaci σ složíme.

Nejjednodušší cykly jsou jednoprvkové pevné body permutace σ a dvouprvkové $(x, \sigma(x))$, kde $\sigma(\sigma(x)) = x$. Těm se říká *transpozice*. Protože každý cyklus zjevně můžeme poskládat z permutací sousedních prvků (necháme „probublat" první prvek nakonec), lze každou permutaci napsat jako složení transpozic sousedních prvků.

Vraťme se k případu Σ_3. Tam máme jednak možnost cyklu, který zahrne všechny tři prvky a v něm dostaneme permutace a, b, c. Kromě toho ještě můžeme mít jeden cyklus o délce 2 a zbývající prvek bude pevným bodem – tak dostaneme zbývající 3 permutace. Více možností není. Z postupu je zřejmé, že u větších počtů prvků bude možností velmi mnoho.

Jednotlivé permutce můžeme obecně vyjádřit pomocí transpozic mnoha způsoby. Přitom ale skutečnost, jestli potřebujeme sudý nebo lichý počet transpozic, je na volbách nezávislá (můžeme tuto skutečnost vyjádřit pomocí počtu tzv. inverzí a poslední tvrzení pak plyne z toho, že každá transpozice mění počet inverzí o lichý počet, viz úvahy v odstavci 2.15 na straně 60).

Máme tedy definováno dobře zobrazení

$$\operatorname{sgn} : \Sigma_m \to \mathbb{Z}_2 = \{\pm 1\},$$

tzv. *paritu*. Dokázali jsme si znovu tvrzení, která jsme již využívali při studiu determinantů (viz 2.14 a dále):

Věta. *Každá permutace konečné množiny je složením cyklů. Cyklus délky ℓ lze vyjádřit jako složení $\ell-1$ transpozic. Parita cyklu délky ℓ je $(-1)^{\ell-1}$.*

Parita složení permutací $\sigma \circ \tau$ je součinem parit σ a τ.

Poslední tvrzení věty říká, že zobrazení sgn převádí složení permutací $\sigma \circ \tau$ na součin $\operatorname{sgn}\sigma \cdot \operatorname{sgn}\tau$ v komutativní grupě $\mathbb{Z}_2$.

Homomorfismy (polo)grup

Obecně říkáme, že zobrazení $f : G_1 \to G_2$ je homomorfismus (polo)grup, jestliže respektuje grupové operace, tzn.

$$f(a \cdot b) = f(a) \cdot f(b).$$

Zejména tedy vidíme, že je naše právě zavedená signatura permutací homomorfismem sgn : $\Sigma_m \to \mathbb{Z}_2$.

10.4

11.3. Symetrie rovinných útvarů. V páté části první kapitoly jsme podrobně a elementárně rozebrali souvislosti invertibilních matic se dvěma řádky a dvěma sloupci a lineárními transformacemi v rovině.

Viděli jsme přitom, že matice v $\mathrm{Mat}_2(\mathbb{R})$ zadávají lineární zobrazení $\mathbb{R}^2 \to \mathbb{R}^2$, které zachovávají standardní vzdálenosti právě, když jsou jejich sloupce ortonormální bazí $\mathbb{R}^2$ (což je jednoduchá podmínka na souřadnice matice, viz odstavec 1.29 na straně 32).

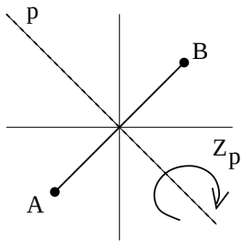
Ve skutečnosti je snadné dokázat, že každé zobrazení roviny do sebe, které zachovává velikosti, je afinní euklidovské, tj. je složením lineárního a vhodné translace.[3]

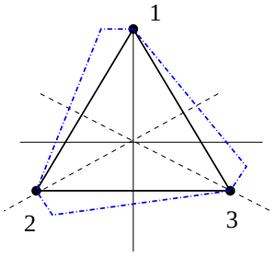
Jak jsme již připomněli, lineární část takového zobrazení přitom musí navíc být ortogonální. Všechna taková zobrazení tedy tvoří grupu všech ortogonálních transformací (nebo také euklidovských transformací) v rovině. Navíc jsme ukazovali, že kromě translací T_a o vektor a jde pouze o rotace R_φ o jakýkoliv úhel φ kolem počátku a zrcadlení Z_ℓ vůči jakékoliv přímce ℓ procházející počátkem (povšimněme si, že středová souměrnost je totéž jako rotace o π).

Zastavíme se teď u ilustrace obecných grupových pojmů

na problému symetrií rovinných obrazců. Budeme přitom uvažovat objekty typu dlaždiček. Nejprve jednotlivě, tj. ve formě ohraničeného obrázku v rovině, později ještě s požadavkem dláždění v rovinném pásu nebo v celé rovině.

Pro začátek uvažme třeba úsečku a rovnostranný trojúhelník. Ptáme se, jak moc jsou symetrické, tzn. vůči kterým trasformacím (zachovávajícím velikost) jsou invariantní. Jinak řečeno, chceme aby obraz našeho obrazce byl od původního k nerozeznání, dokud si nepopíšeme nějaké význačné body, třeba vrcholy trojúhelníka A, B a C a konce úseček. Zároveň je předem jasné, že všechny symetrie pevně zvoleného útvaru budou vždy tvořit grupu (většinou pouze s jediným prvkem, identickým zobrazením).





U úsečky je situace obzvlášť jednoduchá – na první pohled je zřejmé, že jedinými jejími netriviálními symetriemi jsou rotace o π kolem středu úsečky, zrcadlení vůči ose této úsečky a zrcadlení vůči úsečce samotné. Všechny tyto symetrie jsou samy sobě inverzí. Celá grupa symetrií úsečky má tedy čtyři prvky. Její tabulka násobení vypadá takto:

[3] Jestliže totiž má zobrazení $F : \mathbb{R}^2 \to \mathbb{R}^2$ zachovávat velikosti, totéž musí být pravda pro přenášené vektory rychlostí, tj. Jacobiho matice $DF(x, y)$ musí být v každém bodě ortogonální. Rozepsání této podmínky pro dané zobrazení $F = (f(x, y), g(x, y)) : R^2 \to \mathbb{R}^2$ vede na systém diferenciálních rovnic, který má pouze afinní řešení, protože snadno uvidíme, že všechny druhé derivace F musí být nulové (a pak už je naše tvrzení okamžitým důsledkem Taylorovy věty se zbytkem). Zkuste si promyslet detaily! Ve skutečnosti vede stejný postup k výsledku pro euklidovské prostory libovolné dimenze. Všimněte si přitom, že dokazovaná podmínka je nazávislá na volbě afinních souřadnic, proto složením F s lineárním zobrazením výsledek nemění. Můžeme proto pro pevný bod (x, y) složit $(DF)^{-1} \circ F$ a bez újmy na obecnosti rovnou předpokládat, že $DF(x, y)$ je matice identického zobrazení. Derivováním rovnic pak dostáváme důsledky, které přímo říkají požadované tvrzení.

$\cdot$	R_0	R_π	Z_H	Z_V
R_0	R_0	R_π	Z_H	Z_V
R_π	R_π	R_0	Z_V	Z_H
Z_H	Z_H	Z_V	R_0	R_π
Z_V	Z_V	Z_H	R_π	R_0

a je tedy celá tato grupa komutativní.

Pro rovnostranný trojúhelník už symetrií nacházíme víc: můžeme rotovat o $\pi/3$ nebo můžeme zrcadlit vůči osám stran. Abychom dostali grupu celou, musíme přidat všechna složení takovýchto transformací. Už v 1.29 jsme viděli, že složení dvou zrcadlení je vždy otočením. Zároveň je zřejmé, že složení takových zrcadlení v opačném pořadí dá otočení o stejný úhel, ale s opačnou orientací. V našem případě tedy zrcadlení kolem dvou různých os vygenerují postupnou opakovanou aplikací všechny symetrie, který bude dohromady šest. Jestliže si umístíme trojúhelník v souřadnicích jako na obrázku, bude našich šest transformací zadáno maticemi

$$a = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ b = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix},\ c = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$$
$$d = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},\ e = \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix},\ f = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}.$$

Sestavením tabulky pro násobení, tak jak jsme ji udělali pro grupu permutací Σ_3 obdržíme právě stejný výsledek. Pro větší názornost jsou vrcholy označeny čísly, takže jsou příslušné permutace přímo čitelné.

Obdobně umíme nacházet grupy symetrií s k různými rotacemi a k zrcadleními. Stačí si k tomu vzít pravidelný k-úhelník. Takové grupy symetrií se často označují jako grupy D_k a říká se jim *dihedrální grupy* řádu k. Tyto grupy jsou nekomutativní pro všechna $k \geq 3$, zatímco D_2 je komutativní. Název patrně je odvozen od skutečnosti, že D_2 je grupa symetrií molekuly vodíku H_2, ve které jsou dva atomy vodíku a geometricky si ji lze představit jako úsečku.

Stejně tak lze snadno najít obrazce, které mají pouze rotační symetrie a jde tedy o komutativní grupy, které se v chemii značí jako C_k. Říkáme jim *cyklické grupy* řádu k. K tomu postačí např. uvažovat pravidelný mnohoúhelník, u kterého nesymetricky, ale pořád stejně, pozměníme chování hran, viz. čerchované rozšíření trojúhelníku na obrázku. Všimněme si, že grupu C_2 lze realizovat dvěma způsoby – buď jedinou netriviální rotací o π nebo jediným zrcadlením.

Jako první ukázku síly našich abstraktních úvah dokážeme následující větu. Řekneme, že obrazec má *diskrétní grupu symetrií*, jestliže množina obrazů libovolného bodu při působení všemi prvky grupy je diskrétní podmnožinou v rovině (tj. všechny její body mají okolí, ve kterém už žádný další bod množiny není).

Všimněme si, že každá diskrétní grupa symetrií ohraničeného obrazce je nutně konečná.

Věta. *Nechť je M ohraničená množina v rovině $\mathbb{R}^2$ s diskrétní grupou grupou symetrií G. Pak je grupa G buď triviální nebo jedna z grup C_k, D_k, s $k > 1$.*

Důkaz. Kdyby nějaká množina M připouštěla jako svoji symetrii translaci, nemůže být ohraničená. Pokud by M připouštěla rotaci o úhel, jehož žádný celočíselný násobek není 2π (tj. rotaci o iracionální násobek 2π), pak bychom iteracemi této rotace obdželi hustou podmnožinu obrazů na příslušné kružnici. Grupa symetrií by tedy nemohla být diskrétní.

Pokud by M připouštěla netriviální rotace s různými středy, opět nemůže být ohraničená. Napíšeme-li totiž příslušné rotace v komplexní rovině jako

$$R:\ z \mapsto (z-a)\zeta + a, \quad Q:\ z \mapsto \eta z$$

pro komplexní jednotky $\zeta = \mathrm{e}^{2\pi i/k}$, $\eta = \mathrm{e}^{2\pi i/\ell}$ a libovolné $a \neq 0 \in \mathbb{C}$, pak okamžitě vidíme

$$Q \circ R \circ Q^{-1} : z \mapsto z + a\eta(1-\zeta),$$

což je translace o netriviální vektor, pokud není úhel rotace Q nulový. Množina M by tedy nemohla být ohraničená.

Totéž platí pro případ, že by existovala rotační symetrie a zrcadlení podél přímky, která neprochází středem rotace.

Máme tedy k dispozici pouze rotace se společným středem a zrcadlení podél přímek tímto středem procházejících.

Zbývá tedy dokázat, že je celá grupa složena vždy buď pouze z rotací nebo vždy ze stejného počtu rotací a zrcadlení. Připomeňme, že vždy složením dvou různých zrcadlení dostáváme rotaci o úhel rovný polovině úhlu svíraného osami zrcadlení (viz 1.29). Proto je i naopak složením zrcadlení podle přímky p s rotací o úhel $\varphi/2$ zase zrcadlení podél přímky svírající úhel φ s p. Odtud již vcelku snadno plyne požadované tvrzení. □

10.5

11.4. Symetrie rovinných dláždění. Složitější chování lze vypozorovat u rovinných obrazců v pásech nebo v celé rovině (řekněme, že abstraktně zkoumáme možnosti symetrií pro různé dlažby).

Nejprve uvažme množinu tvořenou pásem v rovině uzavřeném mezi dvěma rovnoběžkami a předpokládejme, že je celý tento pás pokryt disjuktními obrazy ohraničené podmnožiny M pomocí nějaké translace. Tato translace bude samozřejmě symetrií zvoleného dláždění rovinného pásu. Grupa symetrií tedy bude vždy nekonečná.

Pro symetrie takové množiny nepřicházejí v úvahu žádné netriviální rotace, kromě R_π, a jediná možná zrcadlení jsou buď horizontální podle osy pásu nebo vertikální podle kterékoliv přímky kolmé na hranice pásu. Zůstavají ještě případné translace zadané vektorem rovnoběžným s osou pásu.

Nepříliš složitá diskuse vede k popisu všech diskrétních grup symetrií pro rovinné pásy. Každá takové grupa je generována některými z následujících možných symetrií: translace T, posunuté zrcadlení G (tj. složení horizontálního zrcadlení

s translací), vertikální zrcadlení V, horizontální zrcadlení H a rotace R o π.

Věta. *Každá diskrétní grupa symetrií dláždění pásu v rovině je jednoho z následujících sedmi typů, tj. je izmorfní s jednou z grup generovaných následujícími symetriemi:*

(1) jedinou translací T
(2) jediným posunutým zrcadlením G
(3) jednou translací T a jedním vertikálním zrcadlením V
(4) jednou translací T a jednou rotací R
(5) jednou posunutou translací G a jednou rotací R
(6) jednou translací T a horizontálním zrcadlením H
(7) jednou translací T, horizontálním zrcadlením H a jedním vertikálním zrcadlením V.

časem by se mohlo dát ověření do série příkladů za kapitolou, případně i sem

Důkaz zde nyní nebudeme uvádět. Příklady vzorů s příslušnými symetriemi jsou aspoň na obrázku.

Složitější je to se symetriemi dláždění, které vyplní celou rovinu. Nemáme zde prostor pro podrobnější zkoumání, nicméně alespoň poznamenejme, že všech takových grup symetrií v rovině je pouze sedmnáct. Říká se jim dvourozměrné krystalografické grupy.

Obdobná úplná diskuse je známa i pro trojrozměrné diskrétní grupy symetrií. Bohatá teorie byla vypracována zejména v 19. století v souvislosti se studiem symetrií krystalů a molekul chemických prvků.

10.6

11.5. Homomorfismy grup. Připomeňme, že zobrazení $f : G \to H$ mezi dvěmi grupami G a H se nazývá homomorfismus grup, jestliže respektuje násobení, tj. pro všechny prvky $a, b \in G$ platí

$$f(a \cdot b) = f(a) \cdot f(b).$$

Povšimněme si, že násobení vlevo je uvnitř grupy G předtím, než zobrazujeme, zatímco vpravo jde o násobení v H poté, co zobrazujeme.

Přímo z definice se snadno ověří následující vlastnosti homomorfismů:

Tvrzení. *Pro každý homomorfismus $f : G \to H$ grup platí*

(1) obraz jednotky $e \in G$ je jednotka v H
(2) obraz inverze k prvku je inverzí obrazu. tj.

$$f(a^{-1}) = f(a)^{-1}.$$

(3) obraz podgrupy $K \subset G$ je podgrupa $f(K) \subset H$.
(4) vzorem $f^{-1}(K) \subset G$ podgrupy $K \subset H$ je podgrupa.
(5) je-li f zároveň bijekcí, pak i inverzní zobrazení f^{-1} je homomorfismus.
(6) f je injektivní zobrazení právě, když $f^{-1}(e) = \{e\}$.

Důkaz. Je-li $K \subset G$ podgrupa, pak pro každé dva prvky $y = f(a)$, $z = f(b)$ v H nutně také $y \cdot z = f(a \cdot b)$ patří do obrazu. Je proto vždy obrazem podgrupy opět podpologrupa (tj. bude to podgrupa, pokud obraz nutně obsahuje i inverze a jednotku).

Specielně, triviální podgrupa $\{e\}$ má za obraz opět podpologrupu. Protože z rovnosti $z \cdot z = z$ v grupě H vynásobením prvkem z^{-1} dostáváme $z = e$, ověřili jsme, že jedinou jednoprvkovou podpologrupou v grupě je triviální podgrupa $\{e\}$. Zejména tedy $f(e) = e$.

Přímo z definice homomorphismu nyní vidíme, že

$$f(a^{-1}) \cdot f(a) = f(e) = e,$$

tj. $f(a)^{-1} = f(a^{-1})$. Dokázali jsme tedy první tři tvrzení.

Stejně postupujeme u vzorů: jestliže $a, b \in G$ splňují $f(a), f(b) \in K \subset H$, potom také $f(a \cdot b) \in K$.

Předpokládejme, že existuje inverzní zobrazení $g = f^{-1}$ a zvolme libovolné $y = f(a), z = f(b) \in H$. Pak $f(a \cdot b) = y \cdot z = f(a) \cdot f(b)$, což je ekvivalentní výrazu $g(y) \cdot g(z) = a \cdot b = g(y \cdot z)$. Je tedy inverze skutečně homomorfismem.

Pokud platí $f(a) = f(b)$, pak $f(a \cdot b^{-1}) = e \in H$. Pokud je tedy jediným vzorem jednotky v H jednotka v G, pak $a \cdot b^{-1} = e$, tj. $a = b$. Opačná implikace je zřejmá. □

Podgrupa $f^{-1}(e)$ jednotkového prvku $e \in H$ se nazývá *jádro* homomorfismu f a značíme ji $\ker f$. Bijektivní homomorfismus grup nazýváme *izomorfismus*.

Z předchozích tvrzení okamžitě vyplývá, že homomorfismus $f : G \to H$ s triviálním jádrem je izomorfismem na obraz $f(G)$.

10.7

11.6. Příklady. Grupy zbytkových tříd $\mathbb{Z}_k$ jsou izomorfní

grupám komplexních k–tých odmocnin z jedničky, což jsou zároveň izomorfní obrazy konečných grup otočení v rovině o celé násobky úhlu $2\pi/k$. Nakreslete si obrázek, počítání s tzv. komplexními jednotkami $e^{2\pi i/k}$ je velmi efektivní.

Zobrazení $\exp : \mathbb{R} \to \mathbb{R}_+$ je izomorfismus aditivní grupy reálných čísel na multiplikativní grupu kladných reálných čísel.

Tento izomorfismus se přirozeně rozšiřuje na morfismus $\exp : \mathbb{C} \to \mathbb{C} \setminus 0$ aditivní grupy komplexních čísel na multiplikativní grupu všech nenulových komplexních čísel. Pro komplexní čísla přitom ale dostáváme netriviální jádro. Viděli jsme totiž, že zúžení exp na ryze imaginární čísla (což je podgrupa izomorfní $\mathbb{R}$) je homomorfismem

$$it \mapsto e^{it} = \cos t + i \sin t,$$

tzn. že čísla $2k\pi i$, $k \in \mathbb{Z}$, jsou v jádru. Snadno se dopočítá, že je to celé jádro. Je-li totiž $e^{s+it} = e^s \cdot e^{it} = 1$ pro reálná s a t, musí být $e^s = 1$, tj. $s = 0$, a pak zbývá pouze $t = 2k\pi$ pro libovolné celé k.

Determinant matice je zobrazením, které každé matici skalárů z $\mathbb{K}$ přiřazuje nějaký skalár v $\mathbb{K}$ (pracovali jsme s $\mathbb{K} = \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$). Cauchyova věta o determinantu součinu čtvercových matic

$$\det(A \cdot B) = (\det A) \cdot (\det B)$$

je tedy tvrzením, že pro grupu $G = GL(n, \mathbb{K})$ invertibilních matic je $\det : G \to \mathbb{K} \setminus 0$ homomorfismem grup.

10.7b

11.7. Součin grup. Jestliže máme k dispozici dvě grupy, můžeme snadno vytvořit složitější grupu následující konstrukcí:

SOUČIN GRUP

Pro každé dvě grupy G, H definujeme *součin grup* $G \times H$ takto: Jako množina je $G \times H$ skutečně součin a násobení definujeme po složkách. tj.

$$(a, x) \cdot (b, y) = (a \cdot b, x \cdot y)$$

kde nalevo vystupuje součin, který definujeme, zatímco napravo používáme tečku k naznačení součinů v jednotlivých grupách G a H.

Projekce na jednotlivé komponenty G a H v součinu,

$$p_G : G \times H \ni (a, b) \mapsto a \in G, \;\; p_H : G \times H \ni (a, b) \mapsto b$$

jsou surjektivní homomorfismy grup s jádry

$$\ker p_G = \{(e_G, b);\; b \in H\} \simeq H$$
$$\ker p_H = \{(a, e_H); a \in G\} \simeq G.$$

Grupa $\mathbb{Z}_6$ je izomorfní součinu $\mathbb{Z}_2 \times \mathbb{Z}_3$. Docela snadno můžeme toto tvrzení vidět při multiplikativní realizaci grup zbytkových tříd $\mathbb{Z}_k$ jakožto komplexních k-tých odmocnin z jedničky. Skutečně tak vidíme, že $\mathbb{Z}_6$ je tvořeno body na jednotkové kružnici v komplexní rovině ve vrcholech pravidleného šestiúhelníku, $\mathbb{Z}_2$ pak odpovídá ± 1, $\mathbb{Z}_3$ pravidelnému trojúhelníku s jedním vrcholem v jedničce. Jestliže budeme ztotožňovat příslušné body s otočeními v rovině, které jedničku převede právě do nich, pak skládání dvou takových otočení bude vždy komutativní a kombinacemi jednoho otočení ze $\mathbb{Z}_2$ a jednoho ze $\mathbb{Z}_3$ dostaneme právě všechna otočení ze $\mathbb{Z}_6$. Nakreslete si obrázek! Takto tedy dostaneme (při obvyklejší aditivní notaci pro zbytkové třídy) izomorfismus:

$$[0]_6 \mapsto ([0]_2, [0]_3)$$
$$[1]_6 \mapsto ([1]_2, [2]_3)$$
$$[2]_6 \mapsto ([0]_2, [1]_3)$$
$$[3]_6 \mapsto ([1]_2, [0]_3)$$
$$[4]_6 \mapsto ([0]_2, [2]_3)$$
$$[5]_6 \mapsto ([1]_2, [1]_3)$$

V zápětí uvidíme, že jsou podobné konstrukce k dispozici pro konečné komutativní grupy úplně obecně.

10.7a

11.8. Komutativní grupy. Libovolný prvek a v grupě G je

obsažen v minimální podgrupě $\{a, a^2, a^3, \dots\}$, která jej obsahuje. Je zjevné, že je tato podgrupa komutativní, a pokud je celá grupa G konečná, nutně musí jednou nastat případ $a^k = e$. Nejmenší k s touto vlastností nazýváme *řád prvku* a v G. Grupa G je *cyklická grupa* je-li celé G generované nějakým

svým prvkem a výše uvedeným způsobem. Pokud je řád k generátoru grupy konečný, jde právě o grupy C_k z naší diskuse symetrií obrazců v rovině.

Z definice přímo vyplývá, že každá cyklická grupa je izomorfní buď grupě celých čísel $\mathbb{Z}$ (pokud je nekonečná) nebo některé grupě zbytkových tříd $\mathbb{Z}_k$ (když je konečná). Ve skutečnosti z takových jednoduchých stavebních kamenů můžeme poskládat všechny konečné komutativní grupy.

Věta. *Každá konečná komutativní grupa G je izomorfní součinu cyklických grup C_k.*

Je-li $n = p_1^{k_1} \cdots p_r^{k_r}$ rozklad přirozeného čísla n na prvočísla, pak je grupa C_n izomorfní součinu

$$C_n = C_{p_1^{k_1}} \times \cdots \times C_{p_r^{k_r}}.$$

Důkaz. Obecné první tvrzení věty zde vůbec nebudeme dokazovat. Kompletní důkaz lze najít např. v [**?**]. Několika poznámkami se ještě k problematice vrátíme v odstavci 11.12.

Důkaz druhého tvrzení začneme jednodušším speciálním případem, kdy $n = pq$ s nesoudělnými p a q. Zvolme generátory a grupy C_n, b grupy C_p a c grupy C_q. Nabízí se definovat zobrazení $f : C_n \to C_p \times C_q$ vztahem

$$f(a^k) = (b^k, c^k).$$

Protože platí $a^k \cdot a^\ell = a^{k+\ell}$ a podobně pro b a c, dostáváme

$$f(a^k \cdot a^\ell) = (b^{k+\ell}, c^{k+\ell}) = (b^k, c^k) \cdot (b^\ell, c^\ell)$$

a jde tedy o homomorfismus. Jestliže je obrazem jednotka, pak k musí být zároveň násobkem p i q. Protože jsou p a q nesoudělné, znamená to, že je k i násobkem n a je proto f injektivní. Přitom zřejmě mají grupy C_n i $C_p \times C_q$ stejně prvků, takže jde o izomorfismus. Odtud již vyplývá tvrzení věty o rozkladu cyklických grup řádu k na součiny menších cyklických grup. □

Všimněme si, že naopak C_{p^2} nikdy není izomorfní součinu $C_p \times C_p$. Zatímco C_{p^2} je totiž generované prvkem řádu p^2, největší dostupný řád prvku v $C_p \times C_p$ je jenom p.

Vzhledem k tomu, že všechny konečné komutativní grupy jsou izomorfní součinu cyklických grup, můžeme pro malé počty prvků najít všechny takové grupy až na izomorfismus. Např. máme jen dvě grupy s 12 prvky

$$C_{12} = C_4 \times C_3, \quad C_2 \times C_2 \times C_3 = C_2 \times C_6.$$

Podobně si můžeme povšimnout, že mají-li všechny prvky v grupě G kromě jednotky řád 2, pak jde o grupu $(C_2)^n$ pro nějaké n, zejména tedy má 2^n prvků. Skutečně, kdybychom totiž v rozkladu G na součin cyklických grup povolili C_p s $p > 2$, pak by tam nutně musely vzniknout prvky vyššího řádu.

10.8

11.9. Rozklady podle podgrup. Volbou libovolné podgrupy H v grupě G dostáváme další informace o struktuře celé grupy. Na množině prvků grupy G nyní definujeme relaci $a \sim_H b$ jestliže $b^{-1} \cdot a \in H$. Snadno ověříme, že je takto definována relace ekvivalence:

Platí $a^{-1} \cdot a = e \in H$, je tedy relace reflexivní. Je-li $b^{-1} \cdot a = h \in H$, potom $a^{-1} \cdot b = (b^{-1} \cdot a)^{-1} = h^{-1} \in H$, je proto relace symetrická. Konečně, je-li $c^{-1} \cdot b \in H$ a zároveň je $b^{-1} \cdot a \in H$, potom $c^{-1} \cdot a = c^{-1} \cdot b \cdot b^{-1} \cdot a \in H$, ověřili jsme tedy i tranzitivitu.

Celá grupa G se proto jako množina rozpadá na tzv. *levé třídy rozkladu* podle podgrupy H vzájemně ekvivalentních prvků. Třídu příslušející prvku a značíme $a \cdot H$ a skutečně platí, že

$$a \cdot H = \{a \cdot h;\ h \in H\},$$

neboť prvek b je ve stejné třídě s a, právě když jde takovýmto způsobem vyjádřit.

Množinu všech levých tříd rozkladu podle podgrupy H označujeme G/H.

Obdobně definujeme pravé třídy rozkladu $H \cdot a$. Příslušná ekvivalence je: $a \sim b$, jestliže $a \cdot b^{-1} \in H$. Proto

$$H \setminus G = \{H \cdot a;\ a \in G\}.$$

Tvrzení. *Pro třídy rozkladu grupy G podle podgrupy H platí:*

(1) Levé a pravé třídy rozkladu podle podgrupy $H \subset G$ splývají právě, když pro každé $a \in G$, $h \in H$ platí

$$a \cdot h \cdot a^{-1} \in H.$$

(2) Všechny třídy (levé i pravé) mají shodnou mohutnost s podgrupou H.

Důkaz. Obě vlastnosti vyplývají bezprostředně z definičních vlastností. V prvém případě chceme, aby pro jakékoliv $a \in G$, $h \in H$ platilo $h \cdot a = a \cdot h'$ pro vhodné $h' \in H$. To ale nastane právě tehdy, když $a^{-1} \cdot h \cdot a = h' \in H$.

Ve druhém případě si stačí uvědomit, že pokud $a \cdot h = a \cdot h'$, pak také vynásobením a^{-1} zleva obdržíme $h = h'$. □

Jako okamžitý důsledek předchozího jednoduchého tvrzení dostáváme

10.9 **11.10. Věta.** *Nechť G je konečná grupa s n prvky, H její podgrupa. Potom*

(1) Mohutnost $n = |G|$ je součinem mohutnosti H a mohutnosti G/H, tj.

$$|G| = |G/H| \cdot |H|$$

(2) Přirozené číslo $|H|$ je dělitelem čísla n.
(3) Je-li $a \in G$ prvek řádu k, pak k dělí n.
(4) Pro každé $a \in G$ je $a^n = e$.
(5) Je-li mohutnost grupy G prvočíslo, pak je G izomorfní cyklické grupě $\mathbb{Z}_n$.

Druhému tvrzení se říkává Lagrangeova věta, předposlednímu malá Fermatova věta.

DŮKAZ. Viděli jsme, že každá třída levého rozkladu má právě $|H|$ prvků. Přitom dvě různé třídy rozkladu musí mít nutně prázdný průnik. Odtud vyplývá první tvrzení.

Druhé tvrzení je okamžitým důsledkem prvního.

Každý prvek $a \in G$ generuje cyklickou podgrupu $\{a, a^2, \dots, a^k = e\}$ a právě počet prvků této podgrupy je řádem prvku a. Proto musí řád dělit počet prvků v G.

Jelikož je řád k prvku a dělitelem čísla n a $a^k = e$, je také $a^n = (a^k)^s = e$ pro nějaké s.

Jestliže je $n > 1$, pak existuje prvek $a \in G$ různý od jednotky e. Jeho řád je přirozené číslo k různé od jedničky a nutně dělí n. Proto musí být k rovno n. Pak ovšem jsou všechny prvky G tvaru a^k pro $k = 1, \dots, n$. □

TADY URČITĚ PŘIJDOU NĚJAKÉ REMINISCENCE Z ELEMENTÁRNÍ TEORIE ČÍSEL!

10.10

11.11. Normální podgrupy a faktorgrupy. Podgrupy H, pro které platí, že $a \cdot h \cdot a^{-1} \in H$ pro všechny $a \in G$, $h \in H$, se nazývají *normální podgrupy*.

Pro normální podgrupy je dobře definováno násobení na G/H vztahem

$$(a \cdot H) \cdot (b \cdot H) = (a \cdot b) \cdot H.$$

Skutečně, volbou jiných reprezentantů $a \cdot h$, $b \cdot h'$ dostaneme opět stejný výsledek

$$(a \cdot h \cdot b \cdot h') \cdot H = ((a \cdot b) \cdot (b^{-1} \cdot h \cdot b) \cdot h') \cdot H = (a \cdot b) \cdot H.$$

Totéž si můžeme odůvodnit tak, že nezáleží na tom jestli pracujeme s pravými nebo levými třídami. Můžeme proto rovnou naše třídy psát jako $H \cdot a \cdot H$ a potom snadno definujeme $(H \cdot a) \cdot (b \cdot H) = H \cdot (a \cdot b) \cdot H$.

Zřejmě jsou splněny pro nové násobení na G/H všechny vlastnosti grupy: jednotkou je sama grupa H jakožto třída $e \cdot H$ jednotky, inverzí k $a \cdot H$ je zřejmě $a^{-1} \cdot H$ a asociativita násobení je zřejmá z definice. Hovoříme o *faktorové grupě* G/H grupy G podle normální podgrupy H.

V komutativních grupách jsou samozřejmě všechny podgrupy normální. Podmnožina

$$n\mathbb{Z} = \{na;\ a \in \mathbb{Z}\} \subset \mathbb{Z}$$

zadává v celých číslech podgrupu a její faktorgrupou je právě (aditivní) grupa zbytkových tříd $\mathbb{Z}_n$.

Z definice je zřejmé, že všechna jádra homomorfismů jsou normální podgrupy. Naopak, jestliže je podgrupa $H \subset G$ normální, pak zobrazení

$$p : G \to G/H, \quad a \mapsto a \cdot H$$

je surjektivní homomorfismus grup s jádrem H. Skutečně, p je dobře definované, přímo z definice násobení na G/H je vidět, že p je homomorfismus, a p je zjevně surjektivní. Je tedy vidět, že normální podgrupy jsou právě všechna jádra homomorfismů.

Dále, pro libovolný homomorfismus grup $f : G \to K$ s jádrem $H = \ker f$ je dobře definován také homomorfismus

$$\tilde{f} : G/\ker f \to K,\ \tilde{f}(a \cdot H) = f(a),$$

který je injektivní.

Zdánlivě paradoxní je příklad homomorfismu grup $\mathbb{C}^* \to \mathbb{C}^*$, který je definovaný na nenulových komplexních číslech vztahem $z \mapsto z^k$ s přirozeným k. Zjevně jde o surjektivní homomorfismus a jeho jádro je množina k–tých odmocnin z jedničky, tj. cyklická podgrupa $\mathbb{Z}_k$. Předchozí úvaha tedy dává pro všechna přirozená k izomorfismus

$$\tilde{f} : \mathbb{C}^*/\mathbb{Z}_k \to \mathbb{C}^*.$$

Tento příklad ukazuje, že u nekonečných grup nejsou počty s mohutnostmi tak přehledné jako tomu bylo u konečných grup ve Větě 11.10.

10.10a

11.12. Exaktní posloupnosti. Kdykoliv zvolíme normální podgrupu H v grupě G, dostáváme tzv. *krátkou exaktní posloupnost grup*

$$e \to H \to G \to G/H \to e,$$

kde šipky postupně znázorňují jediný homomorfismus triviální grupy $\{e\}$ do grupy H, vložení ι podgrupy $H \subset G$, projekci ν na faktorgrupu G/H a, konečně, jediný morfismus grupy G/H na triviální grupu $\{e\}$. Ve všech případech je vidět, že obraz předcházející šipky je přesně jádrem následující. To je definice *exaktnosti* posloupnosti homomorfismů.

Jestliže existuje homomorfismus $\sigma : G/H \to G$, takový že $\nu \circ \sigma = \mathrm{id}_{G/H}$, říkáme, že se naše exaktní posloupnost *štěpí*.

Lemma. *Každá rozštěpená krátká exatkní posloupnost komutativních grup zadává izomorfismus $G = H \times G/H$.*

Důkaz. Definujeme zobrazeni $f : H \times G/H \to G$ vztahem

$$f(a, b) = a \cdot \sigma(b).$$

Protože pracujeme s komutativními grupami, jde zjevně o homomorfismus:

$$f(aa', bb') = aa'\sigma(b)\sigma(b') = (a\sigma(b))(a'\sigma(b')).$$

Jestliže $f(a, b) = e$, pak $\sigma(b) = a^{-1} \in H$, tj. $b = \nu(\sigma(b))$ je tedy jednotkový prvek v G/H. Pak ovšem jeho obraz musí být $\sigma(b) = e$ a je proto i $a = e$. Protože jsou levé a pravé třídy rozkladu u komutativních grup totožné, je zobrazení f zjevně surjektivní. Dokázali jsme tedy, že je f izomorfismus. □

Můžeme nyní naznačit hlavní ideu důkazu Věty 11.8. Kdybychom totiž věděli, že se všechny krátké exaktní posloupnosti vzniklé volbou cyklických podgrup H v konečných komutativních grupách G štěpí, pak bychom snadno důkaz vedli indukcí. V grupě G o mohutnosti n, která není cycklická, bychom totiž zvolili prvek s řádem $p < n$ a našli jím generovanou cyklickou podgrupu H a štěpení příslušné krátké exaktní posloupnosti.

Tím bychom dostali grupu G vyjádřenou jako součin zvolené cyklické podgrupy H a grupy G/H s mohutností n/p.

Hlavním technickým bodem důkazu tedy je ověření, že v každé konečné komutativní grupě najdeme prvky řádu p^r s patřičnými mocninami prvočíselných p a že se skutečně výše uvedené krátké exaktní posloupnosti pro tyto grupy štěpí.

10.11

11.13. Akce grupy. Již jsme viděli, že často potkáváme

grupy jako množiny transformací nějaké pevné množiny. Musí přitom být všechny invertibilní a zároveň musí být naše množina transformací uzavřená na skládání. Často ale také chceme pracovat s pevně zvolenou grupou, jejíž prvky reprezentujeme jako zobrazení na nějaké množině, přitom ale ne nutně jsou zobrazení příslušná různým prvkům grupy různá. Např. všechna otočení roviny kolem počátku o všechny možné úhly odpovídají grupě reálných čísel. Otočení o 2π je ale identické zobrazení.

Formálně si můžeme takovou situaci popsat následovně.

AKCE GRUPY

Levá *akce grupy G na množině S* je homomorfismus grupy G do podgrupy invertibilních prvků v pologrupě S^S všech zobrazení $S \to S$. Takový homomorfismus si také můžeme představit jako zobrazení $\varphi : G \times S \to S$, které splňuje

$$\varphi(a \cdot b, x) = \varphi(a, \varphi(b, x)),$$

odtud název „levá akce“. Často budeme k vyjádření akce prvku grupy na prvku S používat pouze zápis $a \cdot x$ (byť jde o jinou tečku než u násobení uvnitř grup). Definiční vlastnost pak vypadá takto:

$$(a \cdot b) \cdot x = a \cdot (b \cdot x).$$

Obraz prvku $x \in S$ v akci celé grupy G nazýváme *orbita* S_x prvku x, tj.

$$S_x = \{y = \varphi(a, x);\ a \in G\}.$$

Pro každý bod $x \in S$ definujeme *izotropní podgrupu* $G_x \subset G$ akce φ,

$$G_x = \{a \in G;\ \varphi(a, x) = x\}.$$

Jestliže pro každé dva prvky x, $y \in S$ existuje $a \in G$ tak, že $\varphi(a, x) = y$, pak říkáme, že akce φ je *tranzitivní*.

Jestliže zvolíme dva body x, $y \in S$ a prvek $g \in G$ zobrazující x an $y = g \cdot x$, pak je zjevně množina $\{ghg^{-1};\ h \in G_x\}$ izotropní podgrupou G_y. Zobrazení $h \mapsto ghg^{-1}$ je přitom homomorfismem grup $G_x \to G_y$.

Snadno se vidí, že u tranzitivních akcí je celý prostor jedinou orbitou a všechny izotropní podgrupy mají stejnou mohutnost.

Jako příklad tranzitivní akce konečné grupy můžeme uvést např. zjevnou akci grupy permutací pevně zvolené množiny X na samotné množině X. Přirozená akce všech invertibilních lineárních transformací na nenulových prvcích

vektorového prostoru V je také tranzitívní. Pokud vezmeme ale prostor V celý, je nulový vektor zvláštní orbitou.

Výše uváděný příklad akce aditivní grupy reálných čísel prostřednictvím rotací kolem pevného středu O v rovině není tranzitivní. Orbity jednotlivých bodů jsou právě kružnice se středem O procházející těmito body.

Typický příklad tranzitivní akce grupy G je přirozená akce na množině levých tříd G/H pro jakoukoliv podgrupu H. Definujeme ji vztahem

$$g \cdot (aH) = (ga)H.$$

Snadno ukážeme, že takto v podstatě vypadají všechny tranzitivní akce grup. Pro libolnou tranzitivní akci $G \times S \to S$ a pevně zvolený bod $x \in G$ můžeme totiž ztotožnit S s množinou levých tříd G/G_x pomocí vztahu $gG_x \mapsto g \cdot x$. Toto zobrazení je zjevně surjektivní a obrazy $g \cdot x = h \cdot x$ splývají právě když $h^{-1}g \in G_x$ a to je ekvivalentní s $gG_x = hG_x$. Konečně si všimněme, že toto ztotožnění převádí původní akci G na S právě na výše uvedenou akci G na G/G_x.

10.11a **11.14. Věta.** *Pro každou akci konečné grupy G na konečné množině S platí:*

(1) Pro každý prvek $x \in S$ je

$$|G| = |G_x| \cdot |S_x|.$$

(2) (Burnsidovo lemma) Je-li N počet orbit akce G na S pak

$$|G| = \frac{1}{N} \sum_{g \in G} |S_g|,$$

kde $S_g = \{x \in S;\ g \cdot x = x\}$ označuje množinu pevných bodů akce prvku g.

Důkaz. Uvažme libovolný bod $x \in S$ a izotropní podgrupu $G_x \subset G$ tohoto bodu. Stejný argument jako na konci minulého odstavce u tranzitivních grup můžeme uplatnit na každou akci grupy G. Dostáváme zobrazení $G/G_x \to S_x$, $g \cdot G_x \mapsto g \cdot x$. Pokud $(g \cdot S_x) \cdot x = (h \cdot S_x) \cdot x$, pak zjevně $g^{-1}h \in S_x$, je tedy naše zobrazaní injektvní. Zároveň je zjevně surjektivní, proto pro mohutnosti našich konečných množin platí $|G/G_x| = |S_x|$. Odtud již vyplývá první vlastnost z věty, protože $|G| = |G/G_x| \cdot |G_x|$.

Druhé tvrzení dokážeme tak, že dvěma způsoby spočteme mohutnost množiny pevných bodů akce:

$$F = \{(x, g) \in S \times G;\ g(x) = x\} \subset S \times G.$$

Protože jde o konečné množiny, můžeme si představit prvky součinu $S \times G$ jako prvky v matici (sloupce označujeme prvky v S, řádky pak podle prvků v G). Sčítáním po řádcích i sloupcích obdržíme

$$|F| = \sum_{g \in G} |S_g| = \sum_{x \in S} |G_x|.$$

Nyní si pro přehlednost vyberme po jednom reprezentantu $x_1, \ldots, x_N$ z každé orbity v S a připomeňme, že mohutnosti

izotropních grup pro body ve stejné orbitě jsou stejné. Využitím již dokázaného tvrzení (1) nyní vcelku snadno dostáváme

$$|F| = \sum_{g\in G} |S_g| = \sum_{i=1}^{N} \sum_{x\in S_{x_i}} |G_x| = \sum_{i=1}^{N} |S_{x_i}||G_{x_i}| = N\cdot|G|$$

a důkaz je ukončen. □

explicitně zmínit neco z vedlejších sloupců

Doporučujeme si pečlivě promyslet, jak užitečná jsou tvrzení věty pro řešení kombinatorických úloh, viz **??**.

2. Okruhy polynomů

Grupové operace byly podstatné u skalárů i vektorů. Vystupovalo nám tam ovšem několik obdobných struktur zároveň. Zaměříme se teď právě na takové případy. Budeme přitom mít na mysli zejména obvyklé skaláry, tj. celá čísla $\mathbb{Z}$, racionální čísla $\mathbb{Q}$, komplexní čísla $\mathbb{C}$, a množiny polynomů nad takovými skaláry $\mathbb{K}$. Budeme přitom ale pečlivě budovat abstraktní algebraickou teorii.

10.12

11.15. Okruhy a tělesa. Celá čísla mají následující vlastnosti tzv. okruhu:

Okruhy a obory integrity

Definice. Komutativní grupa $(M, +)$ s neutrálním prvkem $0 \in M$, spolu s další operací $\cdot$ splňující

- $(a\cdot b)\cdot c = a\cdot(b\cdot c)$, pro všechny $a, b, c \in M$;
- $a\cdot b = b\cdot a$, pro všechny $a, b \in M$;
- existuje prvek 1 takový, že pro všechny $a \in M$ platí $1\cdot a = a$;
- $a\cdot(b+c) = a\cdot b + a\cdot c$, pro všechny $a, b, c \in M$;

se nazývá *komutativní okruh*.

Jestliže v okruhu $\mathbb{K}$ platí $c\cdot d = 0$ právě, když alespoň jeden z prvků c a d je nulový, pak nazýváme okruh $\mathbb{K}$ *oborem integrity*.

Poslední vlastnosti v našem výčtu axiomů okruhu se říká *distributivita* sčítání vůči násobení. Pokud neplatí vlastnost komutativity operace $\cdot$, hovoříme o *nekomutativním okruhu*. V dalším se ovšem omezíme zpravidla na okruhy komutativní. Operaci $+$ budeme říkat sčítání a operaci $\cdot$ násobení, aniž by to znamenalo, že jde skutečně o tyto operace na některém z našich číselných oborů. Navíc budeme vždy předpokládat existenci jedničky 1 pro operaci násobení. Neutrálnímu prvku pro sčítání říkáme nula.

Tělesa

Netriviální (komutativní) okruh, ve kterém jsou všechny nenulové prvky invertibilní, se nazývá (komutativní) *těleso*. Komutativní těleso se také nazývá *pole*.

Typickým příkladem komutativních těles, tj. polí, jsou číslené obory $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$. Dále pak všechny okruhy zbytkových tříd $\mathbb{Z}_p$ s prvočíselným p. Dobrým příkladem nekomutativního okruhu s jedničkou je množina $\mathrm{Mat}_k(\mathbb{K})$ všech čtvercových matic nad okruhem $\mathbb{K}$ s k řádky a sloupci. Jak jsme viděli dávno, není to ani obor integrity. Jako příklad nekomutativního tělesa uveďme těleso kvaternionů $\mathbb{H}$, viz příklad **??** ve druhém sloupci.

Lemma. *V každém komutativním okruhu $\mathbb{K}$ s jedničkou platí následující vztahy (které nám jistě připadají samozřejmé u skalárů)*

(1) $0 \cdot c = c \cdot 0 = 0$ *pro všechny* $c \in \mathbb{K}$,
(2) $-c = (-1) \cdot c = c \cdot (-1)$ *pro všechny* $c \in \mathbb{K}$,
(3) $-(c \cdot d) = (-c) \cdot d = c \cdot (-d)$ *pro všechny* $c, d \in \mathbb{K}$,
(4) $a \cdot (b - c) = a \cdot b - a \cdot c$,
(5) celý okruh $\mathbb{K}$ *je triviální množinou* $\{0\} = \{1\}$ *právě, když* $0 = 1$.

Důkaz. Všechna tvrzení vyplývají z jednoduché úvahy a definičních axiomů. V prvém případě počítáme pro jakákoliv c, a:

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

a protože jediným neutrálním prvkem vůči sčítání je nula, dostáváme $a \cdot 0 = 0$. Stejně se dokáže i $0 \cdot a = 0$.

Ve druhém případě teď stačí spočíst

$$0 = c \cdot 0 = c \cdot (1 + (-1)) = c + c \cdot (-1),$$

proto je $c \cdot (-1)$ opačný prvek k prvku c, což jsme chtěli dokázat.

Další dvě tvrzení jsou už přímým důsledkem druhého vztahu a základních axiomů. Jestliže je celý okruh tvořen jediným prvkem, je pochopitelně $0 = 1$. Naopak, jestliže platí $1 = 0$, pak pro jakékoliv $c \in \mathbb{K}$ je $c = 1 \cdot c = 0 \cdot c = 0$. □

10.13

11.16. Polynomy nad okruhy. Definice komutativního okruhu s jedničkou abstrahuje právě vlastnosti potřebné k násobení a sčítání. Můžeme je hned využít pro práci s tzv. polynomy. Rozumíme jimi jakýkoliv konečný výraz, který lze poskládat ze známých konstantních prvků $\mathbb{K}$ a jedné neznámé proměnné pomocí operací sčítání a násobení. Formálně můžeme definovat polynomy takto:[4]

Polynomy

Definice. Nechť $\mathbb{K}$ je jakýkoliv komutativní okruh skalárů s jedničkou. Polynomem nad $\mathbb{K}$ rozumíme konečný výraz

$$f(x) = \sum_{i=0}^{k} a_i x^i$$

kde $a_i \in \mathbb{K}$, $i = 0, 1, \ldots, k$, jsou tzv. *koeficienty polynomu*. Je-li $a_k \neq 0$, říkáme, že $f(x)$ má *stupeň* k, píšeme $\deg f = k$. Nulový polynom nemá stupeň, polynomy stupně nula jsou

[4] Ne náhodou je pro okruh použit symbol $\mathbb{K}$ – nadále si pod ním představujte třeba kterýkoliv okruh naších číselných oborů.

právě nenulové prvky v $\mathbb{K}$, kterým říkáme konstantní polynomy.

Polynomy $f(x)$ a $g(x)$ jsou stejné, jestliže mají stejné nenulové koeficienty. Množinu všech polynomů nad okruhem $\mathbb{K}$ budeme značit $\mathbb{K}[x]$.

Každý polynom zadává zobrazení $f : \mathbb{K} \to \mathbb{K}$, jehož hodnota vznikne dosazením hodnoty c za nezávislou proměnnou x, tj.

$$f(c) = a_0 + a_1 c + \cdots + a_k c^k.$$

Všimněme si, že konstantní polynomy odpovídají právě konstantním zobrazením.

Kořen polynomu $f(x)$ je takový prvek $c \in \mathbb{K}$, pro který je $f(c) = 0 \in \mathbb{K}$.

Obecně mohou různé polynomy definovat různá zobrazení. Např. polynom $x^2 + x \in \mathbb{Z}_2[x]$ zadává identicky nulové zobrazení. Obecněji, pro každý konečný okruh $\mathbb{K} = \{a_0, a_1, \ldots, a_k\}$ zadává polynom $f(x) = (x - a_0)(x - a_1) \ldots (x - a_k)$ identicky nulové zobrazení.

Dva polynomy $f(x) = \sum_i a_i x^i$ a $g(x) = \sum_i b_i x^i$ umíme přirozeně také sčítat i násobit:

$$(f + g)(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_k + b_k)x^k$$
$$(f \cdot g)(x) = (a_0 b_0) + (a_0 b_1 + a_1 b_0)x + \ldots$$
$$+ (a_0 b_r + a_1 b_{r-1} + a_r b_0)x^r + \cdots + a_k b \ell x^{k+\ell},$$

kde $k \geq \ell$ jsou stupně polynomů f a g a uvažujeme nulové koeficienty všude tam, kde v původním výrazu pro polynomy nenulové koeficienty nejsou.[5].

Tato definice vskutku odpovídá příslušným operacím sčítání a násobení hodnot zobrazení $f, g : \mathbb{K} \to \mathbb{K}$, díky vlastnostem „skalárů“ v původním okruhu $\mathbb{K}$.

Přímo z definice vyplývá, že množina polynomů $\mathbb{K}[x]$ nad komutativním okruhem s jedničkou je opět komutativním okruhem s jedničkou, přičemž jedničkou v $\mathbb{K}[x]$ je opět jednička 1 v okruhu $\mathbb{K}$ vnímaná jako polynom stupně nula, nulou pro sčítání je nulový polynom.

Lemma. *Okruh polynomů nad oborem integrity je opět obor integrity.*

Důkaz. Máme ukázat, že v $\mathbb{K}[x]$ mohou být netriviální dělitelé nuly pouze, jetliže jsou už v $\mathbb{K}$. To je ale zřejmé z výrazu pro násobení polynomů. Jsou-li $f(x)$ a $g(x)$ polynomy stupně k a ℓ jako výše, pak koeficient u $x^{k+\ell}$ v součinu $f(x) \cdot g(x)$ je součin $a_k \cdot b_\ell$ a ten musí být nenulový, pokud nejsou dělitelé nuly v $\mathbb{K}$. □

10.19

11.17. Polynomy více proměnných. Často se setkáváme s

objekty popsanými pomocí polynomiálních výrazů ale s více proměnnými. Např. kružnici v rovině se středem $S = (x_0, y_0)$ a poloměrem r zapíšeme pomocí rovnice

[5] Formálně bychom mohli naopak za polynom považovat nekonečný výraz pro $i = 0, \ldots, \infty$ s podmínkou, že jen konečně mnoho koeficientů je nenulových.

$$(x - x_0)^2 + (y - y_0)^2 - 1 = 0.$$

Okruhy polynomů v proměnných $x_1, \ldots, x_r$ můžeme zavést úplně podobně jako jsme postupovali s $\mathbb{K}[x]$. Místo mocnin jediné proměnné x^k budeme uvažovat tzv. *monomy*

$$x_1^{k_1} \cdots x_r^{k_r}$$

a jejich formální lineární kombinace s koeficienty $a_{k_1 \cdots k_r} \in \mathbb{K}$.

Formálně i technicky je ale jednodušší je definovat induktivně vztahem

$$\mathbb{K}[x_1, \ldots, x_r] := \big(\mathbb{K}[x_1, \ldots, x_{r-1}]\big)[x_r].$$

Např. $\mathbb{K}[x, y] = \mathbb{K}[x][y]$, tzn. že uvažujeme polynomy v proměnné y nad okruhem $\mathbb{K}[x]$. Snadno si každý ověří (promyslete si to!), že polynomy v proměnných $x_1, \ldots, x_r$ lze i při této definici chápat jako výrazy vzniklé z písmen $x_1, \ldots, x_n$ a prvků okruhu $\mathbb{K}$ konečným počtem (formálního) sčítání a násobení v komutativním okruhu. Například prvky v $\mathbb{K}[x, y]$ jsou tvaru

$$\begin{aligned} f &= a_n(x)y^n + a_{n-1}(x)y^{n-1} + \cdots + a_0(x) \\ &= (a_{mn}x^m + \cdots + a_{0n})y^n + \cdots + (b_{p0}x^p + \cdots + b_{00}) \\ &= c_{00} + c_{10}x + c_{01}y + c_{20}x^2 + c_{11}xy + c_{02}y^2 + \ldots \end{aligned}$$

Pro zjednodušení zápisu si zavedeme tzv. multiindexovou symboliku (kterou jsme používali při diskusi parciálních diferenciálních rovnic vyšších řádů).

Multiindexy

Multiindex α délky r je r-tice nezáporných celých čísel $(\alpha_1, \ldots, \alpha_r)$. Celé číslo $|\alpha| = \alpha_1 + \cdots + \alpha_r$ nazýváme *velikost* multiindexu α.

Stručně zapisujeme monomy x^α místo $x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_r^{\alpha_r}$. Pro polynomy v r proměnných pak máme symbolické vyjádření velice podobné obvyklému značení pro polynomy v jedné proměnné:

$$f = \sum_{|\alpha| \le n} a_\alpha x^\alpha, \; g = \sum_{|\beta| \le m} a_\beta x^\beta \in \mathbb{K}[x_1, \ldots, x_r].$$

Říkáme, že f má celkový stupeň n, je-li alespoň jeden z koeficientů s multiindexem α velikosti n nenulový.

Okamžitě se také nabízejí analogické vzorce pro sčítání a násobení polynomů

$$f + g = \sum_{|\alpha| \le \max(m,n)} (a_\alpha + b_\alpha)x^\alpha$$

$$fg = \sum_{|\gamma|=0}^{m+n} \left(\sum_{\alpha+\beta=\gamma} (a_\alpha b_\beta)x^\gamma \right)$$

kde multiindexy se sčítají po složkách a formálně neexistující koeficienty považujeme za nulové.

Lemma. *Tyto vzorce opravdu popisují sčítání a násobení v induktivně definovaném okruhu polynomů v r proměnných.*

Důkaz. Tvrzení lze snadno dokázat indukcí přes počet proměnných. Předpokládejme, že vztahy platí v $\mathbb{K}[x_1, \dots, x_{r-1}]$ a počítejme součet

$$\begin{aligned}
f &= a_k(x_1, \dots, x_{r-1})x_r^k + \cdots + a_0(x_1, \dots, x_{r-1}) \\
&= \left(\sum_\alpha a_{k,\alpha}x^\alpha\right)x_r^k + \dots \\
g &= b_l(x_1, \dots, x_{r-1})x_r^l + \cdots + b_0(x_1, \dots, x_{r-1}) \\
&= \left(\sum_\beta b_{l,\beta}x^\beta\right)x_r^k + \dots \\
f + g &= \big(a_0(x_1, \dots, x_{r-1}) + b_0(x_1, \dots, x_{r-1})\big) + \\
&+ \big(a_1(x_1, \dots, x_{r-1}) + b_1(x_1, \dots, x_{r-1})\big)x_r + \dots \\
&= \big(\sum_\gamma (a_{k,\gamma} + b_{k,\gamma})(x_1, \dots, x_{r-1})^\gamma\big)x_r^k + \dots \\
&\quad + \big(\sum_\gamma (a_{0,\gamma} + b_{0,\gamma})(x_1, \dots, x_{r-1})^\gamma\big) \\
&= \sum_{(\gamma,j)} (a_{j,\gamma} + b_{j,\gamma})(x_1, \dots, x_{r-1})^\gamma x_r^j.
\end{aligned}$$

Podobně lze vést důkaz pro součin (udělejte samostatně!). □

Jako důsledek naší definice a předchozích výsledků pro polynomy nad obecnými komutativními okruhy dostaneme:

Dusledek. *Jestliže v okruhu $\mathbb{K}$ nejsou dělitelé nuly, pak také v okruhu polynomů $\mathbb{K}[x_1, \dots, x_r]$ nejsou dělitelé nuly.*

Důkaz. Budeme postupovat indukcí přes počet proměnných r.[6] Polynomy v jediné proměnné mají tvar $f = a_n x_1^n + \cdots + a_1 x_1 + a_0$ a $g = b_m x^m + \cdots + b_0$, přičemž $b_m \neq 0$ a $a_n \neq 0$. Vedoucí člen součinu fg je $a_n b_m x^{n+m}$, protože $a_n b_m \neq 0$, zejména tedy je součin nenulových polynomů opět nenulový.

Pokud tvrzení platí pro $r-1$ proměnných, pak použijeme předchozí úvahu pro okruh polynomů v jedné proměnné x_r s koeficienty v $\mathbb{K}[x_1, \dots, x_{r-1}]$. □

10.14

11.18. Dělitelnost a nerozložitelnost. Naším dalším cílem

bude pochopit, jak je to v obecném případě polynomů nad oborem integrity s jejich rozkladem na součin polynomů jednodušších, tj. ve speciálním případě polynomů s jedinou proměnnou budeme diskutovat kořeny polynomů. U polynomů s více proměnnými půjde o rozklad na jednodušší faktory nižších stupňů. Protože již víme, že polynomy ve více proměnných můžeme definovat induktivně, stačí nám nyní uvažovat jen polynomy v jedné proměnné, ovšem nad obecným oborem

[6] Důkaz lze vést také přímo s použitím multiindexových formulí pro součin, když si zavedeme vhodné uspořádání monomů tak, jak to budeme za chvíli stejně dělat.

integrity, a směřujeme ke zobecnění úvah o dělitelnosti, které byly základem našeho počínání v elemntární teorii čísel.

Uvažujme nějaký pevně zvolený obor integrity $\mathbb{K}$. Příkladem nám stále mohou sloužit celá čísla $\mathbb{Z}$ nebo okruh $\mathbb{Z}_p$ s prvočíselným p.

Dělitelnost v okruzích

Obecně říkáme, že $a \in \mathbb{K}$ *dělí* $c \in \mathbb{K}$, jestliže existuje b tak, že $a \cdot b = c$. Skutečnost že $c \in \mathbb{K}$ je dělitelné $a \in \mathbb{K}$ zapisujeme $a|c$.

Dělitelé jedničky, tj. invertibilní prvky v $\mathbb{K}$, se nazývají *jednotky*. Jednotky v komutativním okruhu vždy tvoří komutativní grupu.

V oboru integrity jsou dělitelé určeny jednoznačně. Skutečně je-li $b = a \cdot c$ a $b \neq 0$, pak c je jednoznačně dáno volbou a, b, protože při $b = ac = ac'$ totiž platí $0 = a \cdot (c - c')$ a $a \neq 0$. Z neexistence dělitelů nuly proto vyplývá $c = c'$.

Přimo z definic vyplývají následující tvrzení:

Lemma. *Nechť $a, b, c \in \mathbb{K}$. Potom*

(1) je-li $a|b$ a zároveň $b|c$ pak také $a|c$;
(2) je-li $a|b$ a zároveň $a|c$ pak také $a|(\alpha b + \beta c)$ pro všechny $\alpha, \beta \in \mathbb{K}$;
(3) $a|0$ pro všechny $a \in \mathbb{K}$ (je totiž $a \cdot 0 = 0$);
(4) každý prvek $a \in \mathbb{K}$ je dělitelný všemi jednotkami $e \in \mathbb{K}$ a jejich násobky $a \cdot e$ (jak přímo plyne z existence e^{-1}).

Jednoznačný roklad v oboru integrity

Řekneme, že prvek $a \in \mathbb{K}$ je *nerozložitelný*, jestliže je dělitelný pouze jednotkami $e \in \mathbb{K}$ a jejich násobky $a \cdot e$.

Řekneme, že okruh $\mathbb{K}$ je *obor integrity s jednoznačným rozkladem*, jestliže platí:

- pro každý nenulový prvek $a \in \mathbb{K}$ existují nerozložitelné $a_1, \ldots, a_r \in \mathbb{K}$ takové, že $a = a_1 \cdot a_2 \ldots a_r$
- jsou-li prvky $a_1, \ldots, a_r$ a $b_1, \ldots, b_s$ nerozložitelné, nejsou mezi nimi žádné jednotky a $a = a_1 a_2 \ldots a_r = b_1 b_2 \ldots b_s$, pak je $r = s$ a ve vhodném přeuspořádání platí $a_j = e_j b_j$ pro vhodné jednotky e_j.

Již jsme viděli, že $\mathbb{Z}$ je obor integrity s jednoznačným rozkladem a každé pole (komutativní těleso) je obor integrity s jednoznačným rozkladem (protože každý nenulový prvek v poli je jednotka).

Pro ilustraci si uveďme příklad oboru integrity, který jednoznačný rozklad nemá. Konstrukce je podobná polynomům, jen místo mocnin uvážíme vhodně se skládající odmocniny: Naše $\mathbb{K}$ bude mít prvky tvaru

$$a_0 + \sum_{i=1}^{k} a_i \left(\sqrt[2^{n_i}]{x^{m_i}} \right)$$

kde $a_0, \ldots, a_k \in \mathbb{Z}$, $m_i, n \in \mathbb{Z}_{>0}$. Pak jednotky jsou v $\mathbb{K}$ pouze prvky ± 1, všechny prvky s $a_0 = 0$ jsou rozložitelné, ale např. výraz x nelze vyjádřit jako součin nerozložitelných. Nerozložitelných prvků je v $\mathbb{K}$ prostě příliš málo.

10.15

11.19. Dělení se zbytkem a kořeny polynomu. Základním nástrojem pro diskusi dělitelnosti, společných dělitelů apod. v okruhu celých čísel $\mathbb{Z}$ byla procedura dělení se zbytkem a Euklidův algoritmus pro hledání největších společných dělitelů. Tyto postupy nyní zobecníme.

Lemma (Algoritmus pro dělení se zbytkem)**.** *Nechť $\mathbb{K}$ je komutativní okruh bez dělitelů nuly a $f, g \in \mathbb{K}[x]$ polynomy, $g \neq 0$. Pak existuje $a \in \mathbb{K}$, $a \neq 0$, a polynomy q a r splňující $af = qg + r$, kde $r = 0$ nebo $\deg r < \deg g$. Je-li navíc $\mathbb{K}$ pole, nebo je aspoň vedoucí koeficient polynomu g roven jedné, potom lze volit $a = 1$ a polynomy q a r jsou v tomto případě určeny jednoznačně.*

Důkaz. Tvrzení dokážeme indukcí vzhledem ke stupni f. Je-li $\deg f < \deg g$ nebo $f = 0$, pak volíme $a = 1, q = 0$, $r = f$, což vyhovuje všem našim podmínkám. Pro konstantní polynom g klademe $a = g, q = f, r = 0$.

Předpokládejme tedy, že $\deg f \geq \deg g > 0$ a pišme

$$f = a_0 + \cdots + a_n x^n, \quad g = b_0 + \cdots + b_m x^m.$$

Buď platí $b_m f - a_n x^{n-m} g = 0$ a nebo je $\deg(b_m f - a_n x^{n-m} g) < \deg f$. V prvém případě jsme hotovi, ve druhém pak, podle indukčního předpokladu, existují a', q', r' splňující

$$a'(b_m f - a_n x^{n-m} g) = q' g + r'$$

a buď $r' = 0$ nebo $\deg r' < \deg g$. Tzn.

$$a' b_m f = (g' + a' a_n x^{n-m}) g + r'.$$

Přitom je-li $b_m = 1$ nebo $\mathbb{K}$ je pole, pak podle indukčního předpokladu lze volit $a' = 1$ a q', r' jsou tak určeny jednoznačně. V takovém případě ovšem získáme

$$b_m f = (g' + a_n x^{n-m}) g + r'$$

a je-li $\mathbb{K}$ pole, můžeme rovnost vynásobit b^{-1}.

Předpokládejme, že $f = q_1 g + r_1$ je jiné řešení. Pak $0 = f - f = (q - q_1)g + (r - r_1)$ a buď je $r = r_1$, nebo $\deg(r - r_1) < \deg g$. V prvém případě odtud ovšem plyne i $q = q_1$, protože $\mathbb{K}[x]$ neobsahuje dělitele nuly. Nechť ax^s je člen nejvyššího stupně v $q - q_1 \neq 0$ (určitě existuje). Potom jeho součin se členem nejvyššího stupňe v g musí být nulový (protože nejvyšší stupeň dostaneme tak, že vynásobíme nejvyšší stupně). To ovšem znamená, že $a = 0$. Protože ax^s byl největší nenulový stupeň, nutně dostáváme, že $q - q_1$ žádné nenulové monomy neobsahuje, je tedy určitě nulové. Pak ovšem i $r = r_1$. □

Proceduru dělení se zbytkem můžeme okamžitě využít k diskusi kořenů polynomů. Uvažme tedy polynom $f \in \mathbb{K}[x]$, $\deg f > 0$, a zkusme jej vydělit polynomem $x - b$, $b \in \mathbb{K}$. Protože je vedoucí koeficient jednička, algoritmus pro dělení dává jednoznačný výsledek. Dostáváme tedy jednoznačně zadané polynomy q a r splňující $f = q(x - b) + r$, kde $r = 0$ nebo $\deg r = 0$, tj. $r \in \mathbb{K}$. Tzn., že hodnota polynomu f v $b \in \mathbb{K}$ je rovna právě $f(b) = r$. Z toho plyne, že prvek $b \in \mathbb{K}$ je kořen polynomu f právě, když $(x - b)|f$. Protože po vydělení polynomem stupně jedna vždy klesne stupeň výsledku alespoň o jedničku, dokázali jsme následující tvrzení:

Dusledek. *Každý polynom $f \in \mathbb{K}[x]$ má nejvýše* $\deg f$ *kořenů. Zejména tedy zadávají polynomy nad nekonečným oborem integrity stejná zobrazení $\mathbb{K} \to \mathbb{K}$, právě když jde o stejné polynomy.*

Skutečně, dva polynomy nad nekonečným komutativním okruhem, které zadávají stejné zobrazení $\mathbb{K} \to \mathbb{K}$, mají rozdíl, jehož kořenem je každý prvek v $\mathbb{K}$. To však znamená, že pokud by jejich rozdíl nebyl nulový polynom, pak $\mathbb{K}$ má nejvýše tolik prvků, kolik je maximum ze stupňů uvažovaných polynomů.

Zatímco reálné polynomy mohou být i úplně bez kořenů, každý komplexní polynom naopak takovýto rozklad připouští. To je obsahem tzv. základní věty algebry, kterou pro úplnost uvádíme s (v podstatě) kompletním důkazem. Díky tomuto výsledku víme, že každý polynom v $\mathbb{C}[x]$ má tolik kořenů, včetně násobnosti, jako je jeho stupeň $\deg f = k$. Proto připouští vždy rozklad tvaru

$$f(x) = (x - a_1) \cdot (x - a_2) \dots (x - a_k)$$

s vhodnými komplexními kořeny a_i.

10.18 **11.20. Věta** (Základní věta algebry)**.** *Pole $\mathbb{C}$ je algebraicky uzavřené, tj. každý polynom stupně alespoň 1 má kořen.*

Důkaz. Předpokládejme, že $f \in \mathbb{C}[z]$ je nenulový polynom, který nemá kořen, tj. $f(z) \neq 0$ pro všechny $z \in \mathbb{C}$. Definujme zobrazení

$$\varphi : \mathbb{C} \to \mathbb{C}, \quad z \mapsto \frac{f(z)}{|f(z)|}$$

tj. φ zobrazí celé $\mathbb{C}$ do jednotkové kružnice $K_1 = \{e^{it}, t \in \mathbb{R}\} \subset \mathbb{R}^2 = \mathbb{C}$. Díky našemu předpokladu o nenulovosti $f(z)$ je to skutečně dobře definované zobrazení. Dále definujme zobrazení s hodnotami v kružnici $K_r \subset \mathbb{C}$ se středem v nule a poloměrem $r \geq 0$

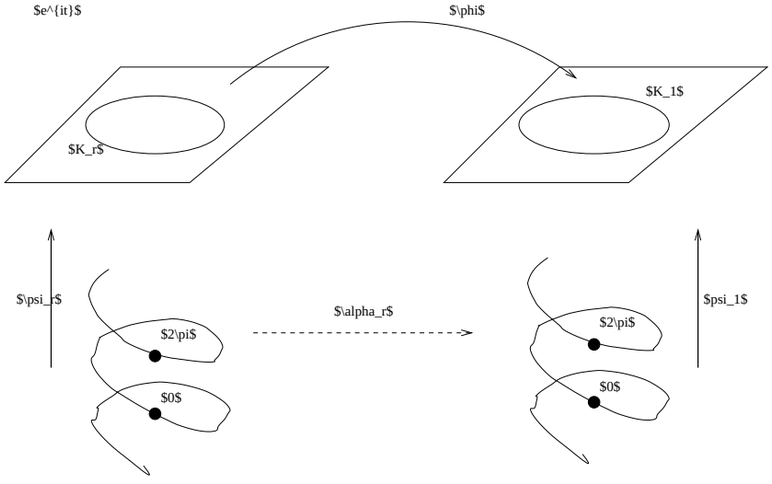
$$\psi_r : \mathbb{R} \to K_r, \quad t \mapsto \psi(t) = re^{it}.$$

Pro každé $r \in \langle 0, \infty)$ máme definováno spojité zobrazení $\kappa_r = \varphi \circ \psi_r : \mathbb{R} \to K_1$. Ze spojité závislosti κ na parametru r navíc vyplývá existence zobrazení $\alpha_r : \mathbb{R} \to \mathbb{R}$ jednoznačně zadaných podmínkami $0 \leq \alpha_r(0) < 2\pi$ a $\kappa_r(t) = e^{i\alpha_r(t)}$.

Získané zobrazení α_r opět spojitě závísí na r. Celkem tedy máme spojité zobrazení

$$\alpha : \mathbb{R} \times \langle 0, \infty) \to \mathbb{R}, \quad (t, r) \mapsto \alpha_r(t)$$

a z jeho konstrukce plyne že pro všechna r je $\frac{1}{2\pi}(\alpha_r(2\pi) - \alpha_r(0)) = n_r \in \mathbb{Z}$. Protože je α spojité, znamená to, že n_r je celočíselná konstanta nezávislá na r. Podívejte se na obrázek, odkud kam jdou jednotlivá zobrazení v naší konstrukci!



Pro dokončení důkazu si stačí uvědomit, že pokud $f = a_0 + \cdots + a_d z^d$ a $a_d \neq 0$, pak pro malá r se bude α_r chovat podobně jako konstantní zobrazení, zatímco pro velká r to vyjde stejně, jako kdyby $f = z^d$. Nejprve si spočtěme, jak tedy n_r dopadne při $f = z^d$, pak toto tvrzení upřesníme a důkaz tím bude ukončen.

Funkce $\mathbb{C} \to \mathbb{C}$, $z \mapsto z^d$, $z \mapsto \frac{z^d}{|z^d|}$ se snadno vyjádří pomocí goniometrického tvaru komplexních čísel $z = r(\cos\alpha + i \sin\alpha)$.

$$z^d = r^d(\cos d\alpha + i \sin d\alpha) = r^d e^{id\alpha}$$

$$\frac{z^d}{|z^d|} = 1(\cos d\alpha + i \sin d\alpha) = e^{id\alpha}$$

zobrazení φ je tedy v tomto případě pouze „zatočení“ na jednotkové kružnici. Pak tedy $\kappa_r(t) = e^{idt}$ a proto $\alpha_r(t) = dt$, nezávisle na r. Odtud pro naši volbu $f = z^d$ vyplývá $n_r = d$. Pokud zvolíme $f = az^d$, $a \neq 0$, nebude to mít na předchozí výsledek žádný vliv (přesvědčte se!).

Zvolme nyní obecný polynom $f = a_0 + \cdots + a_d z^d$, který nemá kořen. Víme tedy, že $a_0 \neq 0$ (pokud by bylo $a = 0$, existoval by kořen). Pro $z \neq 0$ platí

$$\frac{f(z)}{a_d z^d} = 1 + \frac{1}{a_d}(a_0 z^{-d} + \cdots + a_{d-1} z^{-1})$$

a proto $\lim_{|z|\to\infty} \frac{f(z)}{a_d z^d} = 1$. Když tohle víme, můžeme spočítat

$$\lim_{|z|\to\infty} \left| \frac{f(z)}{|f(z)|} - \frac{a_d z^d}{|a_d z^d|} \right| =$$

$$= \lim_{|z|\to\infty} \left| \frac{f(z)}{a_d z^d} \frac{a_d z^d}{|a_d z^d|} \frac{|a_d z^d|}{|f(z)|} - \frac{a_d z^d}{|a_d z^d|} \right| = 0.$$

Proto $n_r = d$ pro velká r.

Podobnou úvahu uděláme i pro malá r. Připomeňme si, že $a_0 \neq 0$.

$$\frac{f(z)}{a_0} = 1 + \frac{1}{a_0}(a_1 z + \cdots + a_d z^d)$$

proto $\lim_{|z|\to 0} \frac{f(z)}{a_0} = 1$. Přitom opět platí $\frac{f(z)}{|f(z)|} = \frac{f(z)}{a_0}\frac{a_0}{|a_0|}\frac{|a_0|}{|f(z)|}$. Odtud $\lim_{|z|\to 0} \frac{f(z)}{|f(z)|} = \lim_{|z|\to 0} \frac{a_0}{|a_0|}$, tj. $n_r = 0$ pro malá r. Celkem vidíme, že stupeň našeho polynomu je $d = 0$. □

10.16

11.21. Největší společný dělitel polynomů. Uvažme okruh polynomů $\mathbb{K}[x]$ nad oborem integrity $\mathbb{K}$. Řekneme, že h je *největší společný dělitel* dvou polynomů a f a $g \in \mathbb{K}[x]$ jestliže:

- $h|f$ a zároveň $h|g$
- jestliže $k|f$ a zároveň $k|g$ pak také $k|h$.

Jako přímý důsledek existence algoritmu pro jednoznačné dělení se zbytkem dostáváme následující důležitou *Bezoutovu rovnost*

Věta. *Nechť $\mathbb{K}$ je pole a nechť $f, g \in \mathbb{K}[x]$. Pak existuje největší společný dělitel h polynomů f a g. Polynom h je určený jednoznačně, až na násobek nenulovým skalárem. Přitom existují polynomy $A, B \in \mathbb{K}[x]$ takové, že $h = Af + Bg$.*

Důkaz. Přímá konstrukce polynomů h, A a B se provede tzv. Euklidovým algoritmem. Provádíme postupně dělení se zbytkem (K je pole, takže to vždy umíme jednoznačně, viz. předchozí lemma):

$$\begin{aligned} f &= q_1 g + r_1 \\ g &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{p-1} &= q_{p+1} r_p + 0. \end{aligned}$$

V tomto postupu neustále klesají stupně r_i, proto jistě nastane rovnost z posledního řádku (pro vhodné p) a ta říká, že $r_p|r_{p-1}$. Z předposledního řádku pak ale plyne $r_p|r_{p-2}$ a postupně dojdeme až nazpět k prvnímu a druhému řádku, které dají $r_p|g$ a $r_p|f$.

Pokud $h|f$ a $h|g$, pak ze stejných rovností postupně plyne, že h dělí všechny r_i, zejména tedy r_p, tzn. získali jsme největšího společného dělitele $h = r_p$ polynomů f a g.

Nyní můžeme postupně dosazovat z poslední do předchozích rovnic.

$$\begin{aligned} h = r_p &= r_{p-2} - q_p r_{p-1} \\ &= r_{p-2} - q_p(r_{p-3} - q_{p-1}r_{p-2}) \\ &= -q_p r_{p-3} + (1 + q_{p-1})r_{p-2} \\ &= -q_p r_{p-3} + (1 + q_{p-1}q_p)r_{p-2} \\ &= -q_p r_{p-3} + (1 + q_p q_{p-1})(r_{p-4} - q_{p-2}r_{p-3}) \\ &\vdots \\ &= Af + Bg. \end{aligned}$$

□

10.20

11.22. Podílová tělesa. Když se potýkáme s celočíselnými

výpočty, je často technicky výhodnější pracovat v číslech racionálních a až na konci postupu ověřit, že výsledek musí ve skutečnosti být celočíselný. Takto jsme už postupovali mnohokrát. Při práci s polynomy nám bude podobný postup užitečný také.

Nechť $\mathbb{K}$ je komutativní okruh (s jedničkou) bez dělitelů nuly. Jeho *podílové těleso* definujeme jako třídy ekvivalence dvojic $(a, b) \in \mathbb{K} \times \mathbb{K}$, $b \neq 0$, které zapisujeme $\frac{a}{b}$, a ekvivalence je dána

$$\frac{a}{b} = \frac{a'}{b'} \quad \Leftrightarrow \quad ab' = a'b.$$

Sčítání a násobení definujeme prostřednictvím reprezentantů tříd

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$$
$$\frac{a}{b}\frac{c}{d} = \frac{ac}{bd}$$

Snadno se ověří korektnost této definice a vąechny axiomy komutativního tělesa. Zejména je $\frac{0}{1}$ neutrální prvek vzhledem ke sčítání, $\frac{1}{1}$ je neutrální prvek vzhledem k násobení a pro $a \neq 0, b \neq 0$ je $\frac{a}{b}\frac{b}{a} = \frac{1}{1}$.

Podílové těleso okruhu $\mathbb{K}[x_1, \ldots, x_r]$ nazýváme *těleso racionálních funkcí* a značíme je $\mathbb{K}(x_1, \ldots, x_r)$. Všechny algebraické operace s polynomy v softwarových systémech jako je Maple nebo Mathematica jsou prováděny ve skutečnosti nad podílovými tělesy, tj. v tělesech raciolnálních funkcí, zpravidla s použitím $\mathbb{K} = \mathbb{Q}$.

Zformulujme si teď velice užitečné (i elegantní) tvrzení, jehož důkaz je docela přímočarý, ale vyžaduje poměrně technické dopracování detailů (a odvíjí se na úrovni podílového tělesa racionálních funkcí). Doporučujeme proto pečlivě pročíst následující odstavec a případně pak při prvním čtení přeskočit další tři lemmata důkazu (a pokračovat na straně **??**).

10.17

11.23. Věta. *Je-li $\mathbb{K}$ obor integrity s jednoznačným rozkladem, pak také okruh polynomů $\mathbb{K}[x]$ je obor integrity s jednoznačným rozkladem.*

Důkaz. Myšlenka důkazu je velice jednoduchá. Uvažujme polynom $f \in \mathbb{K}[x]$. Je-li f rozložitelný, pak je $f = f_1 \cdot f_2$, kde žádný z polynomů $f_1, f_2 \in \mathbb{K}[x]$ není jednotka. Předpokládejme na chvíli navíc, že je-li f dělitelný nerozložitelným polynomem h, pak jistě h dělí f_1 nebo f_2.

Pokud tomu tak vždy bude, docílíme postupnou aplikací předchozí úvahy jednoznačný rozklad. Pokud je totiž f_1 dále rozložitelné, opět $f_1 = g_1 \cdot g_2$, kde g_1, g_2 nejsou jednotky, a přitom vždy buď oba polynomy g_1 a g_2 mají menší stupeň než f, nebo se sníží počet nerozložitelných faktorů ve vedoucích členech g_1 a g_2 (např. nad celými čísly $\mathbb{Z}$ je $2x^2 + 2x + 2 = 2(x^2 + x + 1)$). Proto po konečném počtu kroků dojdeme k rozkladu $f = f_1 \dots f_r$ na nerozložitelné polynomy $f_1, \dots, f_r$.

Z našeho dodatečného předpokladu také plyne, že každý nerozložitelný polynom h dělící f, dělí některý z $f_1, \dots, f_r$. Proto pro každý další rozklad $f = f'_1 f'_2 \dots f'_s$ nutně každý z faktorů f_i dělí některý z f'_j a v takovém případě musí být $f'_j = e f_i$ pro vhodnou jednotku e. Postupným krácením takových dvojic odvodíme, že $r = s$ a jednotlivé faktory se liší pouze o násobky jednotek. □

Zbývá tedy dokázat, že je-li $f = f_1 f_2$ dělitelný nerozložitelným polynomem h, pak jistě h dělí f_1 nebo f_2. Toto tvrzení odvodíme v několika následujících odstavcích.

Dusledek. *Je-li $\mathbb{K}$ obor integrity s jednoznačným rozkladem, pak také okruh polynomů $\mathbb{K}[x_1, \dots, x_r]$ je obor integrity s jednoznačným rozkladem.*

Vidíme tedy, že každý polynom nad oborem integrity s jednoznačným rozkladem můžeme rozložit tak, jak to známe s polynomy s reálnými nebo komplexními koeficienty.

Zejména je tomu tedy tak pro polynomy nad jakýmkoliv polem skalárů.

10.17a **11.24. Lemma.** *Nechť $\mathbb{K}$ je obor integrity s jednoznačným rozkladem. Pak platí:*

(1) Jsou-li $a, b, c \in \mathbb{K}$, a je nerozložitelné a $a|bc$, pak buď $a|b$ nebo $a|c$.

(2) Jestliže konstantní polynom $a \in \mathbb{K}[x]$ dělí $f \in \mathbb{K}[x]$ pak a dělí všechny koeficienty polynomu f.

(3) Je-li a nerozložitelný konstantní polynom v $\mathbb{K}[x]$ a $a|fg$, $f, g \in \mathbb{K}[x]$, pak $a|f$ nebo $a|g$.

Důkaz. 1. Podle předpokladu $bc = ad$ pro vhodné $d \in \mathbb{K}$ a nechť $d = d_1 \dots d_r$, $b = b_1 \dots b_s$, $c = c_1 \dots c_q$ jsou rozklady na nerozložitelné faktory. To znamená

$$ad_1 \dots d_r = b_1 \dots b_s c_1 \dots c_q.$$

Z jednoznačnosti rozkladu ad plyne $a = eb_j$ nebo $a = ec_i$ pro vhodnou jednotku e.

2. Nechť $f = b_0 + b_1 x + \dots + b_n x^n$. Protože $a|f$, jistě existuje polynom $g = c_0 + c_1 x + \dots c_k x^k$ takový, že $f = ag$. Odtud okamžitě plyne $k = n$, $ac_0 = b_0, \dots, ac_n = b_n$.

3. Uvažujme $f, g \in \mathbb{K}[x]$ jako výše a předpokládejme, že a nedělí ani f ani g. Pak podle předchozího bodu existuje

nějaké i tak, že a nedělí b_i a nějaké j tak, že a nedělí c_j. Zvolme taková i, j nejmenší možná. Koeficient u x^{i+j} v polynomu fg je $b_0c_{i+j}+b_1c_{i+j-1}+\cdots+b_{i+j}c_0$. Podle naší volby a dělí všechny $b_0c_{i+j}, \ldots, b_{i-1}c_{j+1}, b_{i+1}c_{j-1}, \ldots, b_{i+j}c_0$. Zároveň ale nedělí b_ic_j. Proto nemůže dělit celý koeficient. □

10.17b **11.25. Lemma.** *Uvažme podílové těleso $\mathbb{L}$ oboru integrity $\mathbb{K}$ s jednoznačným rozkladem. Je-li polynom f nerozložitelný v $\mathbb{K}[x]$ je nerozložitelný také v $\mathbb{L}[x]$.*

Důkaz. Každý koeficient $a \in \mathbb{K}$ můžeme považovat za prvek $\frac{a}{1} \in \mathbb{L}$. Proto každý nenulový polynom $f \in \mathbb{K}[x]$ můžeme uvažovat jako polynom v $\mathbb{L}[x]$.

Předpokládejme, že $f = g'h'$ pro vhodné $g', h' \in \mathbb{L}[x]$, kde polynomy g', h' nejsou jednotky v $\mathbb{L}[x]$ (tzn. nejsou to konstantní polynomy, neboť $\mathbb{L}$ je pole). Nechť a je společný násobek jmenovatelů koeficientů v g' a b je společný násobek jmenovatelů koeficientů v h'. Pak $bh', ag' \in \mathbb{K}[x]$ a platí $abf = (bh')(ag')$. Nechť c je nerozložitelný faktor v rozkladu ab. Pak c dělí $(bh')(ag')$ a proto c dělí polynom bh' nebo polynom ag' (podle předchozího lemmatu). To ale znamená, že c můžeme vykrátit. Po konečném počtu takových krácení zjistíme, že $f = gh$ pro polynomy $g, h \in \mathbb{K}[x]$. Přitom stupeň polynomů se neměnil, proto i g a h nejsou konstantní.

Tím jsme dokázali, že když je f rozložitelné v $\mathbb{L}[x]$, je rozložitelné i v $\mathbb{K}[x]$ a odtud negací vyplývá i požadovaná implikace. □

10.17c **11.26. Lemma.** *Nechť $\mathbb{K}$ je obor integrity s jednoznačným rozkladem a $f, g, h \in \mathbb{K}[x]$. Předpokládejme, že f je nerozložitelné a $f|gh$. Pak buď $f|g$ nebo $f|h$.*

Důkaz. Je-li f konstantní polynom (tj. prvek v $\mathbb{K}$), pak jsme tvrzení již dokázali, viz. jedno z předchozích lemmat.

Předpokládejme, že $\deg f > 0$. Již víme, že f je nerozložitelný také v $\mathbb{L}[x]$, kde $\mathbb{L}$ je podílové těleso okruhu $\mathbb{K}$. Předpokladejme tedy nejdříve, že $\mathbb{K}$ je pole (a je tedy rovno svému podílovému tělesu). Předpokládejme dále, že $f|gh$ a zároveň f nedělí g. Ukážeme, že pak jistě $f|h$. Největší společný dělitel polynomů g a f musí být konstantní polynom v $\mathbb{L}$, proto existují $A, B \in \mathbb{L}[x]$ takové, že $1 = Af + Bg$. Odtud $h = Afh + Bgh$ a protože $f|gh$ musí platit i $f|h$.

Vraťme se nyní k obecnému případu. Podle předchozího vyplývá z našich předpokladů, že $f|g$ nebo $f|h$ v okruhu polynomů $\mathbb{L}[x]$ nad podílovým tělesem $\mathbb{L}$ okruhu $\mathbb{K}$. Nechť např. $h = kf$ v $\mathbb{L}[x]$ a zvolme $a \in \mathbb{K}$ tak, aby $ak \in \mathbb{K}[x]$. Pak $ah = akf$ a pro každý nerozložitelný faktor $e \in a$ musí platit $e|ak$, protože f je nerozložitelný a nekonstantní. Můžeme proto e krátit. Po konečném počtu takových krácení je z a jednotka, tzn. $h = k'f$ pro vhodné $k' \in \mathbb{K}[x]$. □

Důkaz tohoto lemmatu ukončil celý důkaz věty 11.23.

3. Systémy polynomiálních rovnic

V praktických úlohách se často setkáváme s objekty nebo ději popsanými polynomy, resp. systémy polynomiálních rovnic.

Může jít o hledání příslušnosti bodu k nějakému tělesu, hledání extrémů na algebraicky popsaných podmnožinách mnohorozměných prostorů, analýzu pohybů součástí nějakého stroje atd.

10.51

11.27. Afinní variety. Pro jednoduchost (existence kořenů polynomů) budeme pracovat zejména nad polem komplexních čísel, nicméně některé úvahy rozvineme pro obecné pole $\mathbb{K}$.

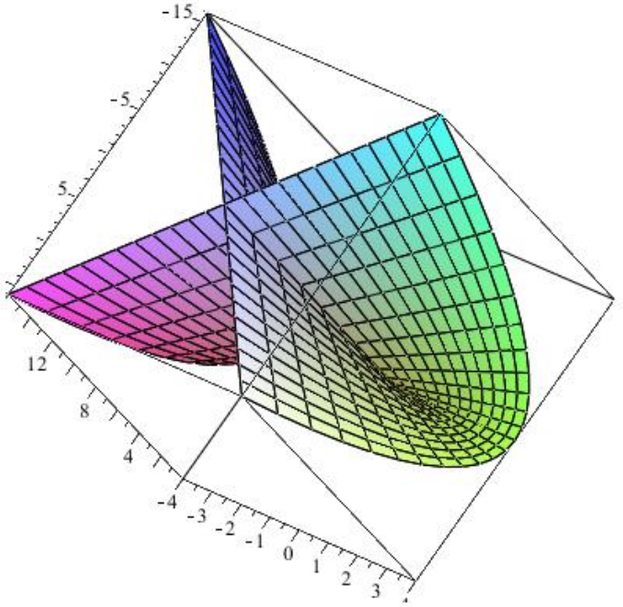
Afinním n-rozměrným prostorem nad polem $\mathbb{K}$ rozumíme $\mathbb{K}^n = \underbrace{\mathbb{K} \times \cdots \times \mathbb{K}}_{n}$ se standardní afinní strukturou, viz začátek čtvrté kapitoly.

Jak jsme již viděli, polynom $f = \sum_\alpha a_\alpha x^\alpha \in \mathbb{K}[x_1 \ldots, x_n]$ lze přirozeným způsobem chápat jako zobrazení $f : \mathbb{K}^n \to \mathbb{K}^n$ definované

$$f(u_1, \ldots, u_n) := \sum_\alpha a_\alpha u^\alpha \quad \text{kde } u^\alpha = u_1^{\alpha_1} \cdots u_n^{\alpha_n}$$

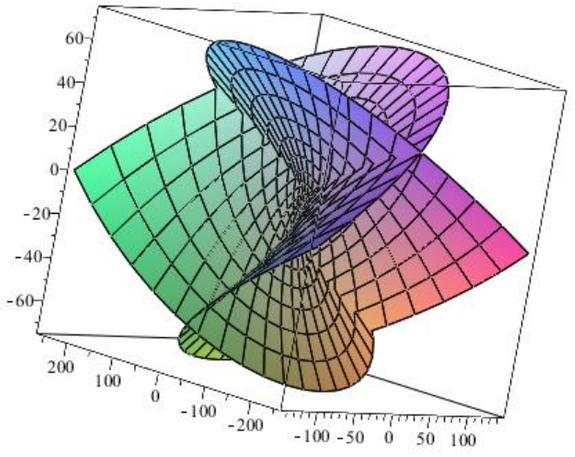
V dimenzi $n = 1$ popisuje rovnost $f(x) = 0$ jen nejvýše konečně mnoho bodů v $\mathbb{K}$. Ve vyšší dimenzi bude rovnost $f(x_1, \ldots, x_n)$ popisovat podmnožiny podobné, jako jsou křivky v rovině nebo plochy v trojrozměrném prostoru, mohou ale mít docela složité a samoprotínající se tvary.

Např. množina zadaná rovnicí $(x^2 + y^2)^3 - 4x^2y^2 = 0$ vypadá jako čtyřlístek. Pěkný obrázek dvourozměrné plochy dává tzv. Whitneyho deštník $x^2 - y^2z = 0$, který kromě znázorněné části na obrázku obsahuje také celou přímku $\{x = 0, y = 0\}$.



Obrázek byl vykreslen s pomocí parametrického popisu $x = uv$, $y = v$, $z = u^2$, ze kterého nejspíš snadno uhádneme i implicitní popis $x^2 - y^2 z = 0$.

Další obrázek ukazuje tzv. Enneperovu plochu s parametrizací $x = 3u+3uv^2-u^3$, $y = 3v+3u^2v-v^3$, $z = 3u^2-3v^2$.



Těžko si představit, jak z této parametrizace dopočítat ručně implicitní popis, přesto to budeme umět algoritmicky zvládnout eliminací proměnných u a v z těchto tří rovnic.

Budeme k tomu ale muset vybudovat docela složitou teorii. Začneme jako obvykle formalizací objektů.

Afinní variety

Nechť $f_1, \ldots, f_s \in \mathbb{K}[x_1, \ldots, x_n]$. *Afinní varietou* v $\mathbb{K}^n$ určenou polynomy $f_1, \ldots, f_n$ nazveme množinu

$$\mathfrak{V}(f_1, \ldots, f_s) = \big\{(a_1, \ldots, a_n) \in \mathbb{K}^n, \\ f_i(a_1, \ldots, a_n) = 0;\ i = 1, \ldots, s\big\}$$

Afinní variety jsou například všechny kuželosečky, kvadriky a nadkvadriky singulární i regulární. Mnoho pěkných křivek či ploch můžeme snadno popsat jako afinní variety.

Varieta určená více polynomy je pak průnik variet zadaných jednotlivými polynomy. Tedy například $\mathfrak{V}(x^2+y^2-1, z)$ je kružnice se středem $(0, 0, 0)$ a poloměrem jedna, ležící v rovině xy.

Podobně $\mathfrak{V}(xz, yz)$ je sjednocení přímky $x = 0$, $y = 0$ a roviny $z = 0$, protože pro právě pro body těchto dvou útvarů jsou oba polynomy xz, yz nulové.

Vidíme na těchto příkladech, že není lehké s vypořádat s pojmem dimenze. Stačí zmíněná přímka navíc k rovině, aby naše varieta byla třírozměrná, nebo ji ještě budeme považovat za dvojrozměrnou s jistou anomálií?

Následující přímočaré tvrzení si ověřte samostatně:

Věta. *Nechť $V = \mathfrak{V}(f_1, \ldots, f_s)$, $W = \mathfrak{V}(g_1, \ldots, g_t) \subseteq \mathbb{K}^n$ jsou afinní variety. Potom i $V \cup W$ a $V \cap W$ jsou afinní variety a platí*

$$V \cap W = \mathfrak{V}(f_1, \ldots, f_s, g_1, \ldots, g_t),$$
$$V \cup W = \mathfrak{V}(f_i g_j) \quad \textit{pro } 1 \leq i \leq s,\ 1 \leq j \leq t.$$

V následujících odstavcích se mimo jiné pokusíme zodpovědět otázky, které se v souvislosti s varietami bezprostředně nabízejí.

A. Platí $\mathfrak{V}(f_1, \ldots, f_s) = \emptyset$?
B. Je $\mathfrak{V}(f_1, \ldots, f_s)$ konečná množina?
C. Jak lze chápat pojem dimenze v případě variet?

Všechny tyto problémy lze „rozumně“ řešit pro variety v oboru komplexních čísel (resp. pro všechna algebraicky uzavřená pole), pro čísla reálná je to komplikovanější a velmi zlé je to pro obecná pole. Například pro racionální čísla je ověření tvrzení $\mathfrak{V}(x^n + y^n - z^n) = \emptyset$ známo jako tzv. velká Fermatova věta.

10.52

11.28. Parametrizace. Pro některé ryze praktické operace s varietami je vhodné používat implicitní reprezentaci (tedy až dosud používané vyjádření). Např. zjištění, zda daný bod patří do variety, resp. do určité části prostoru jí vymezené, je při implicitním popisu docela snadné. Jindy je naopak daleko užitečnější vyjádření parametrické (např. jsme jej již použili při kreslení obrázků).

Varieta $\mathfrak{V}(x + y + z - 1, x + 2y - z - 3)$ je přímka (průnik dvou rovin). Řešíme-li systém

$$x + y + z - 1 = 0$$
$$x + 2y - z - 3 = 0$$

dostaneme přímo parametrické vyjádření této přímky

$$x = -1 - 3t$$
$$y = 2 - 2t$$
$$z = t$$

RACIONÁLNÍ PARAMETRIZACE

Definice. *Racionální parametrickou reprezentací* variety $\mathfrak{V}(f_1, \ldots, f_r) \subseteq \mathbb{K}^n$ rozumíme racionální funkce $r_1, \ldots, r_n \in \mathbb{K}(t_1, \ldots, t_s)$ splňující následující podmínky

- Je-li $x_i = r_i(t_1, \ldots, t_s)$ pro $i = 1, 2, \ldots, n$ pak $(x_1, \ldots, x_n) \in \mathfrak{V}(f_1, \ldots, f_r)$ pro libovolná $t_1, \ldots, t_s$.
- $\mathfrak{V}(f_1, \ldots, f_r)$ je minimální afinní varieta obsahující takto dané body $(x_1, \ldots, x_n)$.

Všimněme si, že při parametrizaci nepožadujeme popis všech bodů variety. To je podstatné, jak je vidět i na jednoduchém příkladu parametrizace kružnice v rovině,

$$x = \frac{2t}{1 + t^2},\ y = \frac{-1 + t^2}{1 + t^2},$$

kterou obdržíme tzv. stereografickou projekcí. (Ověřte si detailně!) Všimněme si, že skutečně dostaneme parametrizací

všechny body, kromě bodu (0, 1), ze kterého promítáme. Ten totiž není dosažitelný pro žádnou hodnotu parametru t. To není způsobeno naší nešikovností, z rozdílných topologických vlastností přímky a kružnice totiž vyplývá, že globální bijektivní racionální parametrizace existovat nemůže.

V této souvislosti se nabízí další otázky.

D. Existuje parametrizace dané variety, resp. lze ji nalézt?
E. Naopak, umíme k parametricky zadané varietě najít její implicitní popis?

Obecná odpověď na otázku D je záporná. V podstatě lze tvrdit, že většinu afinních variet parametrizovat nelze, respektive neexistuje algoritmus parametrizace implicitního popisu. Ty, u kterých se to podaří, se v angličtině nazývají *unirational*, česky tedy patrně *neiracionální*.

Na první pohled je zřejmé, že pro jednu a tutéž varietu existuje více implicitních, případně i parametrických popisů. Opomeneme-li parametrický popis, nejednoznačnosti implicitního jsou způsobeny reprezentací pomocí několika „generujících“ polynomů a zjevně máme velikou volnost v jejich volbě.

10.53

11.29. Ideály. Abychom se vyhnuli závislosti na jednotlivých zvolených rovnicích zadávajících varietu, budeme chtít uvažovat i všechny důsledky zadaných rovnic. To vede na následující algebraický pojem:

IDEÁLY

Definice. Množinu $I \subseteq \mathbb{K}$, kde $\mathbb{K}$ je komutativní okruh, nazveme *ideálem*, platí-li $0 \in I$ a zároveň

$$f, g \in I \implies f + g \in I$$
$$f \in I, h \in \mathbb{K} \implies f \cdot h \in I$$

Ideály můžeme *generovat* podmnožinami, budeme používat značení $I = \langle a_1, \ldots, a_n \rangle$. Tím máme na mysli

$$I = \Big\{ \sum_i a_i b_i,\ b_i \in \mathbb{K} \Big\}.$$

Množina generátorů může být také nekonečná, je-li generátorů jen konečný počet, říkame, že ideál je *konečně generovaný*.

IDEÁL VARIETY

Pro varietu $V = \mathfrak{V}(f_1, \ldots, f_s)$ klademe

$$\mathfrak{I}(V) := \big\{ f \in \mathbb{K}[x_1, \ldots, x_n],\ f(a_1, \ldots, a_n) = 0, \quad \forall\, (a_1, \ldots, a_n) \in V \big\}$$

Lemma. *Nechť $f_1, \ldots, f_s, g_1, \ldots, g_t \in k[x_1, \ldots, x_n]$ jsou polynomy. Pak platí*

(1) Jestliže $\langle f_1, \ldots, f_s \rangle = \langle g_1, \ldots, g_t \rangle$, pak $\mathfrak{V}(f_1, \ldots, f_s) = \mathfrak{V}(g_1, \ldots, g_t)$.

(2) $\mathfrak{I}(V)$ *je ideál a platí* $\langle f_1, \dots, f_s\rangle \subseteq \mathfrak{I}(V)$*, kde* $V = \mathfrak{V}(f_1, \dots, f_s)$.

Důkaz. Jestliže nějaký bod $(a_1, \dots, a_n)$ patří varietě $\mathfrak{V}(f_1, \dots, f_s)$, s v tomto bodě jistě nuluje i jakýkoliv polynom

$$f = h_1 f_1 + \cdots + h_s f_s,$$

tj. libovolný prvek ideálu $I = \rangle f_1, \dots, f_s \langle$. Proto se v něm dle předpokladu nulují i všechny polynomy g_i. Ověřili jsme tedy

$$\mathfrak{V}(f_1, \dots, f_s) \subseteq \mathfrak{V}(g_1, \dots, g_t).$$

Opačná inkluze se dokáže stejně.

Abychom ověřili druhé tvrzení, zvolme $g, g' \in \mathfrak{I}(V)$, $h \in \mathbb{K}[x_1, \dots, x_n]$. Pak pro každý bod $a \in V$ platí

$$(gh)(a) = 0 \Leftrightarrow gh \in \mathfrak{I}(V)$$

$$(g + g')(a) = 0 \leftrightarrow g + g' \in \mathfrak{I}(V)$$

Je tedy $\mathfrak{I}(V)$ ideál v $\mathbb{K}[x_1, \dots, x_n]$.

Pro libovolný $f = h_1 f_1 + \cdots + h_s f_s \in \rangle f_1, \dots, f_s \langle$ a bod $a \in V$ je samozřejmě také $f(a) = 0$, což ověřuje i dokazovanou inkluzi. □

Nejjednodušší příklady jsou triviální variety – jeden bod a celý afinní prostor:

$$\mathfrak{I}\big(\{(0, 0, \dots, 0)\}\big) = \langle x_1, \dots, x_n\rangle$$

$$\mathfrak{I}(\mathbb{K}^n) = \{\,0\} \quad \text{pro libovolné nekonečné pole } \mathbb{K}$$

Inkluze opačná k druhé části věty obecně neplatí. Například varieta $\mathfrak{V}(x^2, y^2)$ má jediný bod – $(0, 0)$. $\mathfrak{I}(V)$ je potom $\langle x, y\rangle \supset \langle x^2, y^2\rangle$.

Jsou-li $V, W \subseteq \mathbb{K}^n$ variety, pak platí

$$V \subseteq W \implies \mathfrak{I}(V) \supseteq \mathfrak{I}(W)$$

Neboli polynomy, které se nulovaly na nějaké varietě se nutně musí nulovat i na její podmnožině.

Můžeme hned formulovat další přirozené problémy

F. Je každý ideál $I \in \mathbb{K}[x_1, \dots, x_n]$ konečně generovaný?
G. Lze algoritmicky zjistit, zda $f \in \langle f_1, \dots, f_s\rangle$?
H. Jaký je přesný vztah mezi $\langle f_1, \dots, f_s\rangle$ a $\mathfrak{I}\big(\mathfrak{V}(f_1, \dots, f_s)\big)$?

10.54

11.30. Dimenze 1. Podívejme se na polynomy v jedné proměnné x

$$f = a_0 x^n + a_1 x^{n-1} + \cdots + a_n \quad \text{kde } a_0 \neq 0.$$

Vedoucí člen polynomu definujeme jako $LT(f) := a_0 x^n$ (označení pochází z anglického „leading term“). Zřejmě platí

$$\deg f \leq \deg g \iff LT(f) | LT(g)$$

Nechť $\mathbb{K}$ je pole a g nenulový polynomu. Víme, že každý polynom $f \in \mathbb{K}[x]$ lze jednoznačně psát jako

$$f = q \cdot g + r \quad \text{kde } r = 0 \text{ nebo } \deg r < \deg g.$$

Jde ve skutečnosti o algoritmický postup, podíl q a zbytek r počítá následující algoritmus.

(1) $q := 0, r := f$
(2) `while` $r \neq 0 \wedge LT(g) | LT(r)$
  (a) $q := q + LT(r)/LT(g)$
  (b) $r := r - LT(r)/LT(g) \cdot g$

Pro průchod cyklem platí invariant $f = q \cdot g + r$, algoritmus tedy dává správný výsledek, pokud se zastaví. Stupeň r se každým průchodem zmenšuje, proto k zastavení nutně dojde.

Dusledek. *Nechť $\mathbb{K}$ je pole. Pak každý ideál v orkuhu polynomů $\mathbb{K}[x]$ je tvaru $\langle f \rangle$.*

Důkaz. Uvažme jakýkoliv ideál $I \subseteq \mathbb{K}[x]$. Je-li $I = \{0\}$, pak je generován nulovým polynomem. Jestliže I obsahuje nenulový polynom f, pak jistě obsahuje i polynom f minimálního stupně. Jistě je pak $\rangle f \langle \subset I$.

Pro jakýkoliv jiný polynom $g \in I$ spočteme výsledek dělení se zbytkem, tj. $g = qf + r$. Zjevně je tedy $qf \in I$ a proto i $r \in I$. Stupeň f byl ale minimální, takže nutně $r = 0$. Je tedy i $g \in I$ a $I = \rangle f \langle$. □

Ideály generované jediným prvkem se nazývají *hlavní ideály*. Okruhům, které mají vlastnost z posledního lematu říkáme *okruh hlavních ideálů*.

Největší společný dělitel $h = GCD(f, g)$ polynomů f a g lze opět spočítat algoritmicky:

(1) $h := f, s := g$
(2) `while` $s \neq 0$
  (a) $r :=$ zbytek po dělení h/s
  (b) $h := s$
  (c) $s := r$

Nechť $f = q \cdot g + r$ a $h = GCD(f, g)$. Potom $h | r, g$ a zároveň

$$\forall p \in \mathbb{K}[x]\colon\ p | r, g \quad \text{tedy } p | f \text{ a } p | h$$

Odtud h je $GCD(r, g)$. Triviálně $GCD(h, 0) = h$, proto algoritmus počítá správně $GCD(f, g)$. Protože stupně r postupně klesají, algoritmus zastaví.

Největší společný dělitel dvou polynomů tedy existuje. Je určen jednoznačně až na násobek skalárem. Dva různé *GCD* se totiž musí dělit navzájem a to je u polynomů možné právě v tomto případě.

Největšího společného dělitele více než dvou polynomů definujeme takto: Je-li $s > 2$, potom

$$GCD(f_1, \ldots, f_s) := GCD\big(f_1, GCD(f_2, \ldots, f_s)\big)$$

Lemma. *Pro polynomy $f_1, \ldots, f_s$ platí $\langle GCD(f_1, \ldots, f_s) \rangle = \langle f_1, \ldots, f_s \rangle$.*

Důkaz. $GCD(f_1, \ldots, f_s)$ dělí všechny polynomy f_i. Je tedy hlavní ideál $\rangle GCD(f_1, \ldots, f_s) \langle$ obsažen v ideálu $\rangle f_1, \ldots, f_s \langle$. Naopak z Bezoutovy rovnosti okamžitě plyne inkluze opačná. □

Položili jsme několik otázek. Tady jsou odpovědi pro dimenzi 1:

- Protože $\mathfrak{V}(f_1, \ldots, f_s) = \mathfrak{V}(GCD(f_1, \ldots, f_s))$, problém prázdnosti variety se redukuje na problém existence kořene polynomu.
- Ze stejného důvodu je varieta vždy konečnou množinou izolovaných bodů – kořenů $GCD(f_1, \ldots, f_s)$ s jedinou vyjímkou, kdy $GCD(f_1, \ldots, f_s) = 0$; to nastane pouze v případě, že $f_1 = f_2 = \cdots = f_s = 0$. Pak je varietou celá množina $\mathbb{K}$.
- Pojem dimenze v tomto případě postrádá smysl, všechny variety mají coby diskrétní množiny bodů dimenzi nulovou.
- Každý ideál je generovatelný jediným polynomem.
- $f \in \langle f_1, \ldots, f_s \rangle \iff GCD(f_1, \ldots, f_s) | f$.
- Označíme-li $\langle f \rangle := \mathfrak{I}(\mathfrak{V}(f_1, \ldots, f_s))$, pak f a $GCD(f_1, \ldots, f_s)$ se mohou lišit pouze násobností kořenů.

10.55

11.31. Monomiální uspořádání. Abychom mohli zobecnit

dělení polynomů se zbytkem pro polynomy více proměnných, najdeme nejprve dobrý ekvivalent pojmů stupeň polynomu a vedoucí člen polynomu.

Dělením se zbytkem polynomu $f \in \mathbb{K}[x_1, \ldots, x_n]$ polynomy $g_1, \ldots, g_s$ chceme rozumět vyjádření

$$f = a_1 g_1 + \cdots + a_s g_s + r,$$

kde vžádný člen zbytku r nebude dělitelný některým z vedoucích členů $LT\, g_i$.

Zkusme to s $f = x^2y + xy^2 + y^2$, $g_1 = xy - 1$ a $g_2 = y^2 - 1$. Prvním dělením získáme

$$f = (x + y) \cdot g_1 + (x + y^2 + y)$$

$LT(y^2 - 1)$ nedělí x (vedoucí člen zbytku), a tak bychom teoreticky nemohli pokračovat dál.

Přesuneme-li však toto x do zbytku, dostáváme teprve výsledek

$$f = (x + y) \cdot g_1 + g_2 + (x + y + 1)$$

Zde již žádný člen zbytku není dělitelný žádným z $LT(g_1)$, $LT(g_2)$.

Jak jsme ale vlastně určovali vedoucí členy?

Uspořádání monomů

Úplné (lineární) dobré (tj. každá neprázdná podmnožina má nejmenší prvek) uspořádání $<$ na $\mathbb{N}^n$ splňující

$$\forall \alpha, \beta, \gamma \in \mathbb{Z}^n : \alpha < \beta \implies \alpha + \gamma < \beta + \gamma$$

nazveme *monomiálním uspořádáním* na $\mathbb{K}[x_1, \ldots, x_n]$.

Uspořádání na $\mathbb{N}^n$ indukuje uspořádání na monomech.

Každý polynom lze však přeskládat jako klesající posloupnost monomů (na koeficienty teď nehledíme). Uspořádání se na polynomy rozšíříme „lexikograficky“, tedy větší je ten polynom, který má větší první monom, pokud tak nelze rozhodnou, bere se v potaz druhý monom atd.

Následující tři definice zavádějí nejběžněji užívaná monomiální uspořádání. Všechna se opírají o předem dané uspořádání jednotlivých proměnných, standardně $x_1 > x_2 > \cdots > x_n$.

Definice. *Lexikografické uspořádání* je takové $<_{\text{lex}}$, že pro každé $\alpha, \beta \in \mathbb{N}^n$ platí

$\alpha >_{\text{lex}} \beta \iff$ Nejlevější nenulový člen v $\alpha - \beta$ je kladný

Gradované lexikografické uspořádání je takové $<_{\text{grlex}}$, že pro každé $\alpha, \beta \in \mathbb{N}^n$ platí:

$$\begin{aligned} \alpha >_{\text{grlex}} \beta \iff & |\alpha| > |\beta| \quad \text{nebo} \\ & |\alpha| = |\beta| \quad \text{a zároveň } \alpha >_{\text{lex}} \beta \end{aligned}$$

Gradované opačné lexikografické uspořádání je takové $<_{\text{grevlex}}$, že pro každé $\alpha, \beta \in \mathbb{N}^n$ platí:

$$\begin{aligned} \alpha >_{\text{grevlex}} \beta \iff & |\alpha| > |\beta| \quad \text{nebo } |\alpha| = |\beta| \\ & \text{a zároveň nejpravější nenulový člen } (\alpha - \beta) < 0 \end{aligned}$$

Tedy $x_1 >_{\text{grevlex}} x_2 >_{\text{grevlex}} \cdots >_{\text{grevlex}} x_n$, ale pokud $x > y > z$, pak $x^2yz^2 >_{\text{grlex}} xy^3z$, ale $x^2yz^2 <_{\text{grevlex}} xy^3z$.

Ověřte si podrobně, že $>_{\text{lex}}, >_{\text{grlex}}, >_{\text{grevlex}}$ jsou skutečně monomiální uspořádání.

10.56

11.32. Dělení se zbytkem. Nechť $f = \sum_{\alpha \in \mathbb{N}^n} a_\alpha x^\alpha$ je nenulový polynom v $\mathbb{K}[x_1, \ldots, x_n]$ a $<$ monomiální uspořádání. Pak definujeme:

- *Stupeň* multideg $f := \max\{\alpha \in \mathbb{N}^n,\ a_\alpha \neq 0\}$
- *Vedoucí koeficient LC* $f := a_{\text{multideg } f}$
- *Vedoucí monom LM* $f := x^{\text{multideg } f}$
- *Vedoucí člen LT* $f := LC\, f \cdot LM\, f$

Tyto pojmy jsou tedy pro polynomy více proměnných vesměs silně závislé na volbě konkrétního uspořádání.

Lemma. *Nechť $f, g \in \mathbb{K}[x_1, \ldots, x_n]$ a uvažme monomiální uspořádání $<$. Pak*

(1) multideg$(f \cdot g) =$ multideg $f +$ multideg g

(2) $f + g \neq 0 \implies$ multideg$(f + g) \leq \max\{\text{multideg } f, \text{multideg } g\}$

Důkaz. Plyne okamžitě přímo z definic. □

Věta. *Nechť $<$ je monomiální a $F = (f_1, \ldots, f_s)$ s-tice polynomů v $\mathbb{K}[x_1, \ldots, x_n]$. Pak každý $f \in \mathbb{K}[x_1, \ldots, x_n]$ lze vyjádřit jako*

$$f = a_1 f_1 + \cdots + a_s f_s + r$$

kde $a_i, r \in \mathbb{K}[x_1, \ldots, x_n]$ pro všechna $i = 1, 2, \ldots, s$. Navíc $r = 0$ nebo r je lineární kombinací monomů, z nichž žádný není dělitelný kterýmkoli z LT $f_1, \ldots,$ LT f_s a pokud $a_i f_i \neq 0$ pak multideg $f \geq$ multideg $a_i f_i$ *pro každé i.*

Polynom r nazýváme zbytkem po dělení f/F.

Důkaz. Věta neříká nic o jednoznačnosti výsledku. Následující algoritmus dává jedno možné řešení a je tedy důkazem platnosti věty.

Nadále budeme výsledkem dělení se zbytkem chápat právě tento výstup pevně zvoleného algoritmu.

(1) $a_1 := 0, \ldots, a_s := 0, r := 0, p := f$
(2) `while` $p \neq 0$
  (a) $i := 1$
  (b) $d :=$ *false*
  (c) `while` $i \leq s \wedge \text{not}\, d$
    (i) `if` $LT\ f_i | LT\ p$
      $a_i := a_i + LT\ p / LT\ f_i$
      $p := p - (LT\ p / LT\ f_i) \cdot f_i$
      $d :=$ *true*
    (ii) `else` $i := i + 1$
  (d) `if` not d
    (i) $r := r + LT\ p$
    (ii) $p := p - LT\ p$

decdg1

decdg2

Při každém průchodu vnějším cyklem se právě jednou provede právě jeden z příkazů 2(c)i, 2(d)ii, a tedy stupeň p klesne. Proto algoritmus skončí.

Platí invariant $f = a_1 f_1 + \cdots + p + r$ a přitom každý člen každého a_i je podílem $LT\ p / LT\ f_i$ z nějakého okamžiku. Proto stupeň těchto členů je menší než stupeň p v daném okamžiku a ten je nejvýše roven stupni f. Dohromady stupeň každého $a_i f_i$ je menší nebo roven stupni f. □

V okruhu $\mathbb{K}[x_1, \ldots, x_n]$ platí pouze implikace

$$f = a_1 f_1 + \cdots + a_s f_s + 0 \implies f \in \langle f_1, \ldots, f_s \rangle$$

Obrácení obecně pro naše dělení se zbytkem neplatí. Uvažujme $f = xy^2 - x$, $f_1 = xy + 1$, $f_2 = y^2 - 1$. Potom algoritmus dělení dá

$$f = y(xy + 1) + 0(y^2 - 1) + (-x - y)$$

ale přitom evidentně $f = x(y^2 - 1)$, a tedy $f \in \langle f_1, f_2 \rangle$.

10.58

11.33. Monomiální ideály. Ideál $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ nazýváme *monomiální*, jestliže existuje množina multiindexů $\alpha \subseteq \mathbb{N}^n$ taková, že I je generován právě všemi monomy x^α s $\alpha \in A$.

To znamená, že všechny polynomy v I jsou tvaru $\sum_{\alpha \in A} h_\alpha x^\alpha$, kde $h_\alpha \in k[x_1, \ldots, x_n]$.

Zřejmě pro monomiální ideál I platí, že $x^\beta \in I$, právě když existuje $\alpha \in A$ takové, že x^α dělí x^β.

Lemma. *Nechť $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ je monomiální ideál, $f \in \mathbb{K}[x_1, \ldots, x_n]$ polynom. Pak následující tvrzení jsou ekvivalentní*

(1) $f \in I$
(2) Každý člen polynomu f je prvkem I.
(3) Polynom f je lineární kombinací monomů z I s koeficienty z k.

Důkaz. Implikace (3) $\implies$ (2) $\implies$ (1) jsou zřejmé. Zbývá ukázat (1) $\implies$ (3).

Zapišme si polynom $f = \sum_\alpha a_\alpha x^\alpha$, kde $a_\alpha \in \mathbb{K}$. Z předpokladu $f \in I$ vyplývá, že lze také vyjádřit $f = \sum_{\beta \in A} h_\beta x^\beta$, kde $x^\beta \in I$ a $h_\beta \in \mathbb{K}[x_1, \ldots, x_n]$.

Každý člen $a_\alpha x^\alpha$ se musí rovnat některému členu z druhé rovnosti. Jistě tedy každý člen $a_\alpha x^\alpha$ polynomu f můžeme vyjádřit jako součet výrazů $d\, x^{\beta+\delta}$, kde $d \in \mathbb{K}$, $x^\beta \in I$. Pak ale také $x^\alpha \in I$, a tedy platí (3). □

Dusledek. *Dva monomiální ideály splývají právě tehdy, když obsahují stejné monomy.*

10.59 **11.34. Věta** (Dicksonovo lemma)**.** *Každý monomiální ideál* $I = \langle x^\alpha,\ \alpha \in A\rangle \subseteq k[x_1, \ldots, x_n]$ *lze psát ve tvaru* $I = \langle x^{\alpha_1}, \ldots, x^{\alpha_s}\rangle$*, kde* $\alpha_1, \ldots, \alpha_s \in A$.

Důkaz. Důkaz provedeme indukcí podle počtu proměnných. V případě $n = 1$ je $I \subseteq \mathbb{K}[x]$, $I = \langle x^\alpha,\ \alpha \in A \subseteq \mathbb{N}\rangle$. Množina všech exponentů v A má jistě minimum a definujeme $\beta := \min A$. Potom zřejmě x^β dělí všechny monomy x^α s $\alpha \in A$ a tedy také $I = \langle x^\beta\rangle$.

Uvažujme nyní $n > 1$ a předpokládejme, že pro menší počty proměnných tvrzení platí. Pro přehlednost si označíme proměnné jako $x_1, \ldots, x_{n-1}, y$ a monomy budeme psát ve tvaru $x^\alpha y^m$, kde $\alpha \in \mathbb{N}^{n-1}$, $m \in \mathbb{N}$. Množinu monomů x^β s $\beta \in A$ budeme značit I_A. Předpokládejme, že $I \subseteq \mathbb{K}[x_1, \ldots, x_{n-1}, y]$ je monomiální a definujme $J \subseteq \mathbb{K}[x_1, \ldots, x_{n-1}]$ následovně

$$J := \langle x^\alpha,\ \exists m \in \mathbb{N},\ x^\alpha y^m \in I_A\rangle.$$

Zřejmě je J monomiální ideál v $n-1$ proměnných a tedy podle indukčního předpokladu lze psát $J = \langle x^{\alpha_1}, \ldots, x^{\alpha_s}\rangle$. Dále z definice J vyplývá, že existují taková minimální $m_i \in \mathbb{N}$, že $x^{\alpha_i} y^{m_i} \in I_A$. Označme tedy $m := \max\{m_i\}$ a definujme analogicky systém ideálů $J_k \subseteq \mathbb{K}[x_1, \ldots, x_{n-1}]$ pro $0 \le k \le m-1$

$$J_k := \langle x^\beta,\ x^\beta y^k \in I_A\rangle$$

Opět všechny J_k splňují indukční předpoklad a tedy je lze vyjádřit

$$J_k = \langle x^{\alpha_{k,1}}, \ldots, x^{\alpha_{k,s_k}}\rangle.$$

Zbývá ukázat, že I je generovaný právě zkonstruovanou konečnou množinou monomů

$$\begin{gathered} x^{\alpha_1} y^m, \ldots, x^{\alpha_s} y^m \\ x^{\alpha_{0,1}} y^0, \cdots, x^{\alpha_{0,s_0}} y^0 \\ \vdots \\ x^{\alpha_{m-1,1}} y^{m-1}, \ldots, x^{\alpha_{m-1,s_{m-1}}} y^{m-1} \end{gathered}$$

Uvažujme tedy libovolný monom $x^\alpha y^p \in I_A$. Nastane jeden ze dvou případů

- $p \ge m$. Potom jistě $x^\alpha \in J$, a tedy některý z $x^{\alpha_1} y^m, \ldots, x^{\alpha_s} y^m$ dělí $x^\alpha y^p$.
- $p < m$. Potom analogicky $x^\alpha \in J_k$ a některý z $x^{\alpha_{k,1}} y^k, \ldots, x^{\alpha_{k,s_k}} y^k$ dělí $x^\alpha y^p$.

Podle předchozího lemmatu lze každé $f \in I$ vyjádřit jako lineární kombinaci monomů z I_A, ty jsou již dělitelné některým

z našich generátorů, proto f patří do ideálu jimi generovaného. Proto I je jeho podmnožinou. Opačná inkluze je zcela triviální a důkaz Dicsonova lematu je hotov. □

10.60

11.35. Hilbertova věta. Nyní již máme nachystáno vše potřebné pro diskusi pěkných bazí ideálů v okruzích polynomů. Hlavní myšlenkou je maximální využití informací o vedoucích členech prvků v bázi a v celém ideálu.

Je-li $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ nenulový, označíme

$$LT\, I := \{ax^\alpha,\ \exists f \in I\colon\ LT\, f = ax^\alpha\}$$

Zřejmě $\langle LT\, I\rangle$ je monomiální ideál, proto podle Dicksonova lemmatu lze psát $\langle LT\, I\rangle = \langle LT\, g_1, \ldots, LT\, g_s\rangle$ pro nějaká vhodná $g_1, \ldots, g_s \in I$.

Věta. *Každý ideál $I \in \mathbb{K}[x_1, \ldots, x_n]$ je konečně generovaný.*

Důkaz. Pokud je $I = \{0\}$, je tvrzení triviální. Uvažujme tedy $I \neq \{0\}$. Podle Dicksonova lematu a předchozí poznámky existují taková $g_1, \ldots, g_s \in I$, že $\langle LT\, I\rangle = \langle LT\, g_1, \ldots, LT\, g_s\rangle$.

Zřejmě $\langle g_1, \ldots, g_s\rangle \subseteq I$. Vezměme libovolné $f \in I$ a proveďme dělení se zbytkem s-ticí $g_1, \ldots, g_s$. Dostáváme

$$f = a_1 g_1 + \cdots + a_s g_s + r,$$

kde žádný člen r není dělitelný $LT\, g_1, \ldots, LT\, g_s$.

Protože $r = f - a_1 g_1 - \cdots - a_s g_s$, platí $r \in I$, a tedy také $LT\, r \in LT\, I$. Zřejmě tedy $LT\, r \in \langle LT\, I\rangle$. Připusťme, že $r \neq 0$. Protože $\langle LT\, I\rangle$ je monomiální, musí být $LT\, r$ dělitelný některým z jeho generátorů, tj. $LT\, g_1, \ldots, LT\, g_s$. To je ovšem spor s výsledkem algoritmu dělení. Proto $r = 0$ a I je generovaný $g_1, \ldots, g_s$. □

Gröbnerovy báze ideálů

10.61

11.36. Definice. Konečná báze $g_1, \ldots, g_s$ ideálu $I \subseteq k[x_1, \ldots, x_n]$ se nazývá *Gröbnerova,* jestliže platí $\langle LT\, I\rangle = \langle LT\, g_1, \ldots, LT\, g_s\rangle$.

Báze použitá v důkazu Hilbertovy věty byla Gröbnerova.

Dusledek. *Každý ideál $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ má Gröbnerovu bázi. Přitom každá množina polynomů $g_1, \ldots, g_s \in I$ splňující $\langle LT\, I\rangle = \langle LT\, g_1, \ldots, LT\, g_s\rangle$ je Gröbnerovou bází ideálu I.*

Ukažme smysl předchozích obecných výsledků na nejjednodušším případě polynomů stupně jedna s lexikografickým uspořádáním:

Označme generátory $f_i = \sum_j a_{i,j} x_j + a_{i,0}$. Uvažujme matici $A = (a_{i,j})$, kde $i = 1, \ldots, s$ a $j = 0, \ldots, n$ a aplikujme na ni Gausovu eliminaci. Získáme $B = (b_{i,j})$ ve schodovitém tvaru, z ní navíc vypustíme nulové řádky. Máme novou bázi $g_1, \ldots, g_t$, kde $t \leq s$.

Vzhledem k provedeným úpravám je každé f_i vyjádřitelné jako lineární kombinace $g_1, \ldots, g_t$, a tedy

$$\langle f_1, \ldots, f_s \rangle = \langle g_1, \ldots, g_t \rangle$$

Ověříme si, že takto získané polynomy $g_1, \ldots, g_t$ jsou Gröbnerovou bází:

Bez újmy na obecnosti předpokládejme, že proměnné jsou značeny tak, že $LM\, g_i = x_i$ pro $i = 1, \ldots, t$. Libovolný $f \in I$ lze psát

$$f = h_1 f_1 + \cdots + h_s f_s = h'_1 g_1 + \cdots + h'_t g_t$$

Chceme, aby $LT\, f \in \langle LT\, g_1, \ldots, LT\, g_t \rangle$, tj. $LT\, f$ má být dělitelný některým z $x_1, \ldots, x_t$. Předpokládejme, že f je pouze v proměnných $x_{t+1}, \ldots, x_n$. Pak ale $h'_1 = 0$, protože x_1 je vzhedem ke schodovitosti B pouze v g_1. Analogickým postupem získáme $h'_2 = \cdots = h'_t = 0$, a tedy $f = 0$.

Dokázali jsme sice existenci nadějných zvláštních bází, zatím je ale neumíme algoritmicky konstruovat. K tomu se dostaneme v následujících odstavcích.

10.62 **11.37. Věta.** *Nechť $G = \{g_1, \ldots, g_t\}$ je Gröbnerova báze ideálu $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ a f je polynom v $\mathbb{K}[x_1, \ldots, x_n]$. Pak existuje právě jedno $r = \sum_\alpha a_\alpha x^\alpha \in \mathbb{K}[x_1, \ldots, x_n]$ s těmito vlastnostmi*

(1) Žádný člen r není dělitelný žádným z $LT\, g_1, \ldots, LT\, g_t$, tj. $\forall\alpha\forall i : LT\, g_i \not| \, a_\alpha x^\alpha$.

(2) $\exists g \in I : f = g + r$

Důkaz. Algoritmus pro dělení se zbytkem dá

$$f = a_1 g_1 + \cdots + a_t g_t + r,$$

kde r splňuje podmínku (1). Za g si zvolme $a_1 g_1 + \cdots + a_t g_t$, které samozřejmě patří do I.

Zbývá dokázat jednoznačnost. Předpokládejme

$$f = g + r = g' + r',$$

kde $r \neq r'$. Zřejmě platí $r - r' = g' - g \in I$. Protože G je Gröebnerova báze, je $LT(r - r')$ dělitelný některým z $LT\, g_1, \ldots, LT\, g_t$. Máme přitom jen dvě možnosti

- $LM\, r \neq LM\, r'$. Pak ten s vyšším stupněm musí být dělitelný některým z vedoucích členů $LT\, g_1, \ldots, LT\, g_t$, což je spor s podmínkou (1).
- $LM\, r = LM\, r'$ a zároveň $LC\, r \neq LC\, r'$. Potom ale oba mnomy $LM\, r$ a $LM\, r'$ musí být dělitelné některým z $LT\, g_1, \ldots, LT\, g_t$, což je opět spor.

Proto tedy $LT\, r = LT\, r'$ a induktivní úvahou odtud plyne $r = r'$. □

Předchozí věta zobecňuje dělení se zbytkem, kde na místě dělitele vystupuje ideál. V případě jedné proměnné nebylo co zobecňovat, protože každý ideál byl generovaný jedním polynomem. Zajímá-li nás pouze zbytek, věta navíc říká, že nezáleží na pořadí polynomů v Gröbnerově bázi. Proto má smysl zavést značení $\overline{f}^{\,G}$ pro zbytek po dělení f/G, pokud $G = (g_1, \ldots, g_s)$ je Gröbnerova báze.

Dusledek. *Nechť $G = \{g_1, \ldots, g_t\}$ je Gröbnerova báze ideálu $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ a f je polynom v $\mathbb{K}[x_1, \ldots, x_n]$. Pak je libovolný polynom f prvkem ideálu I, právě když je zbytek po dělení f/G nulový.*

10.63

11.38. Syzygy. Dalším krokem bude nalezení dostatečné

„testovací množiny“ polynomů z daného ideálu, které je třeba prověřit dělením se zbytkem, abychom mohli usoudit, že je uvažovaný systém generátorů již Gröbnerovou bazí.

Pro $\alpha = \text{multideg}\, f$ a $\beta = \text{multideg}\, g$ uvažme

$$\gamma := (\gamma_1, \ldots, \gamma_n) \quad \text{kde } \gamma_i = \max\{\alpha_i, \beta_i\}$$

Monom x^γ nazýváme *nejmenším společným násobkem (least common multiple)* monomů $LM\ f$ a $LM\ g$ a zavádíme označení $LCM(LM\ f, LM\ g) := x^\gamma$. Výraz

$$S(f, g) := \frac{x^\gamma}{LT\ f} \cdot f - \frac{x^\gamma}{LT\ g} \cdot g$$

nazýváme S-polynomem (nebo také *syzygy*, neboli spřežení) polynomů f, g.

Jedná se o nástroj k eliminaci vedoucích členů, Gaussova eliminace je speciálním případem tohoto postupu pro polynomy stupně jedna. Narozdíl od ní ale může dojít ke zvýšení stupně, i když původní vedoucí členy odstraní.

Vezměme například $f = x^3y^2 - x^2y^3 + x$, $g = 3x^4y + y^2$, tedy polynomy stupně 5 v $\mathbb{R}[x, y]$ a uspořádání $<_{\text{grlex}}$. Pak $\gamma = (4, 2)$ a

$$S(f, g) = \frac{x^4y^2}{x^3y^2} f - \frac{x^4y^2}{3x^4y} g = xf - \frac{1}{3} yg = -x^3y^3 + x^2 - \frac{1}{3} y^3$$

což je polynom stupně 6.

Věta. *Nechť $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ je ideál. Pak je $G = \{g_j, \ldots, g_t\}$ jeho Gröbnerova báze, právě když pro každé $i \neq j$ je zbytek po dělení $S(g_i, g_j)/G$ nulový.*

Důkaz. Důkaz začneme technickým lemmatem, které popisuje, jakým způsobem mohou nastávat krácení při vyjádření polynomů pomocí generátorů. Přesněji řečeno, že je můžeme vždy vyjádřit pomocí S-polynomů.

Lemma. *Uvažme polynom $f = \sum_{i=1}^{t} c_i x^{\alpha_i} g_i$, kde $c_1, \ldots, c_t \in k$ a $\alpha_i + \text{multideg}\, g_i = \delta$ pro nějaké pevné δ kdykoli $c_i \neq 0$. Pokud* multideg $f < \delta$, *pak existují taková $c_{jk} \in \mathbb{K}$, že*

$$\sum_{i=1}^{t} c_i x^{\alpha_i} g_i = \sum_{j,k=1}^{t} c_{jk} x^{\delta - \gamma_{jk}} S(g_j, g_k),$$

kde $x^{\gamma_{jk}} = LCM(LM\ g_j, LM\ g_k)$ a každý monom $x^{\delta - \gamma_{jk}} S(g_j, g_k)$ má stupeň menší než δ.

Důkaz. Označme $d_i := LC\ g_i$ a $p_i = x^{\alpha_i} g_i / d_i$. Určitě platí $c_i d_i = LC(c_i x^{\alpha_i} g_i)$ a $LC\ p_i = 1$. Protože $\text{multideg}(c_i x^{\alpha_i} g_i) = \delta$ a zároveň $\text{multideg}\, f < \delta$, musí

nutně platit také $\sum_{i=1}^{t} c_i d_i = 0$. Pokusme se teď f vyjádřit jako kombinaci S-polynomů.

$$\begin{aligned} f = \sum_{i=1}^{t} c_i d_i p_i &= c_1 d_1 (p_1 - p_2) + (c_1 d_1 + c_2 d_2)(p_2 - p_3) \\ &\quad + \cdots + (c_1 d_1 + \cdots + c_{t-1} d_{t-1})(p_{t-1} - p_t) \\ &\quad + \underbrace{(c_1 d_1 + \cdots + c_t d_t)}_{0} p_t . \end{aligned}$$

Každý rozdíl $p_j - p_k$ lze vyjádřit v S-polynomech

$$\begin{aligned} \frac{x^\delta}{d_j x^{\delta-\alpha_j}} g_j - \frac{x^\delta}{d_k x^{\delta-\alpha_k}} g_k &= x^{\delta-\gamma_{jk}} \left(\frac{x^{\gamma_{jk}}}{LT\, g_j} g_j - \frac{x^{\gamma_{jk}}}{LT\, g_k} g_k \right) \\ &= x^{\delta-\gamma_{jk}} S(g_j, g_k) \end{aligned}$$

Z obou rovností se už snadno odvodí jednotlivé koeficienty c_{jk}. □

Nyní můžeme přikročit k důkazu věty. Implikace „$\Longrightarrow$“ plyne bezprostředně z důsledku v odstavci 11.37. Musíme dokázat implikaci opačnou.

Uvažme nenulový polynom $f \in I$. Potřebujeme ukázat, že za předpokladu dokazované implikace vždy bude platit $LT\, f \in \langle LT\, g_1, \ldots LT\, g_t \rangle$. Podaří-li se zaručit, že lze náš polynom vyjádřit jako $f = \sum_{i=1}^{t} h_i g_i$ s vlastností

$$\text{multideg}\, f = \max\bigl\{\text{multideg}(h_i g_i)\bigr\}$$

bude $LT\, f$ nutně dělitelný některým $LT\, g_i$, a tedy G bude skutečně Gröbnerova báze.

Označme $m_i := \text{multideg}(h_i g_i)$, $\delta := \max\{m_1, \ldots, m_t\}$. Zřejmě multideg $f \leq \delta$. Nechť jsou polynomy $h_1, \ldots, h_t$ zvoleny tak, že δ je minimální. Protože pracujeme s monomiálním uspořádáním, které je dobré, takové δ existuje.

Dokažme tedy, že multideg $f = \delta$. Můžeme psát

$$\begin{aligned} f &= \sum_{m_i=\delta} h_i g_i + \sum_{m_i<\delta} h_i g_i \\ &= \sum_{m_i=\delta} (LT\, h_i) g_i + \sum_{m_i=\delta} (h_i - LT\, h_i) g_i + \sum_{m_i<\delta} h_i g_i . \end{aligned}$$

Všechny sčítance druhé a třetí sumy mají jistě stupeň menší než δ. Připustíme-li, že multideg $f < \delta$, potom nutně

$$\text{multideg} \left(\sum_{m_i=\delta} (LT\, h_i) g_i \right) < \delta .$$

Označme nyní $c_i x^{\alpha_i} := LT\, h_i$ a aplikujme naše technické lemma.

$$\sum_{m_i=\delta} (LT\, h_i) g_i = \sum_{m_i=\delta} c_i x^{\alpha_i} g_i = \sum_{j,k} c_{jk} x^{\delta-\gamma_{jk}} S(g_j, g_k).$$

Z předpokladu věty a algoritmu o dělení se zbytkem získáváme

$$S(g_j, g_k) = \sum_{i=1}^{t} a_{ijk} g_i$$

a navíc multideg$(a_{ijk}g_i) \leq$ multideg $S(g_j, g_k)$. Označíme-li $b_{ijk} := x^{\delta-\gamma_{jk}}a_{ijk}$, dostáváme

$$x^{\delta-\gamma_{jk}}S(g_j, g_k) = \sum_{i=1}^{t} b_{ijk}g_i$$

Podle druhé části lemmatu platí

$$\text{multideg}(b_{ijk}g_i) \leq \text{multideg}\big(x^{\delta-\gamma_{jk}}S(g_j, g_k)\big) < \delta$$

a dosazením dostáváme

$$\begin{aligned}\sum_{m_i=\delta}(LT\, h_i)g_i &= \sum_{j,k} c_{jk}\left(\sum_{i=1}^{t} b_{ijk}g_i\right)\\ &= \sum_{i=1}^{t}\left(\sum_{j,k} c_{jk}b_{ijk}\right)g_i.\end{aligned}$$

Přitom platí

$$\text{multideg}\left(\sum_{j,k} c_{jk}b_{ijk}g_i\right) < \delta \quad \text{pro } i = 1, \ldots, t.$$

Dosazením do naší původní rovnosti získáváme vyjádření f jako kombinace $g_1, \ldots, g_t$, kde všechny sčítance jsou stupně menšího než δ. To je spor s minimální volbou δ, a tedy multideg $f = \delta$, odkud $LT\, f \in \langle LT\, g_1, \ldots, LT\, g_t\rangle$ a báze G je Gröbnerova. □

10.64

11.39. Naivní algoritmus pro Gröbnerovy báze. Právě dokázaná věta nám již poskytuje účinný prostředek pro zjištění, zda nějaká báze je Gröbnerova. Uvažujme například $I = \langle x + y, y - z\rangle$. Jediný S-polynom, který připadá v úvahu je

$$S(x + y, y - z) = \frac{xy}{x}(x + y) - \frac{xy}{y}(y - z) = xz + y^2$$

Dělením získáme $xz + y^2 = z(x + y) + y(y - z)$, a tedy daná báze je Gröbnerova.

Následující algoritmus využívá přesně tento postup pro nalezení nějaké Gröbnerovy báze ideálu generovaného s–ticí polynomů $F = (f_1, \ldots, f_s)$.

(1) $G := F$, $G' := \emptyset$
(2) `while` $G \neq G'$
  (a) $G' := G$
  (b) $\forall p, q \in G'$: $p \neq q$ `do`
    (i) $s := \overline{S(p, q)}^{G'}$
    (ii) `if` $s \neq 0$
      $G := G \cup \{s\}$

Když se algoritmus zastaví, jistě to bude v G Göbnerova báze. Musíme tedy už jen ověřit, že se skutečně zastaví. V jeho průběhu ovšem při každém běhu vnitřním cyklem (2), tj. když se přidává nějaký netriviální zbytek po dělení, buď monomiální ideál generovaný $LT\, G$ vzroste nebo zůstane stejný. Dostáváme tedy neklesající řetězec (monomiálních) ideálů $I_1 = LT(F) \subseteq I_2 \subseteq \cdots \subseteq I_k \subseteq \cdots$. Označíme-li nyní

$I = \cup_{k=1}^{\infty} I_k$, pak jde jistě o ideál a podle Hilbertovy věty musí být konečně generovaný. To ale znamená, že všechny generátory I jsou již v některém z I_k a proto od tohoto k počínaje bude platit $I_k = I_{k+1} = \ldots$.[7]

Stabilizace tohoto řetezce monomiálních ideálů hlavních členů je ale ekvivalentní zastavení algoritmu.

Tento algoritmus ovšem není zdaleka ideální. Lze vymyslet velmi jednoduše vypadající vstupy, pro něž vrací divoké výsledky. Dále výstupní báze se přímo odvíjí od vstupní, a tedy pro tentýž ideál zadaný různými bázemi dá také různé výsledky.

10.65

11.40. Redukce bazí. Viděli jsme, že k rozpoznání, které

generátory jsou potřebné pro Gröbnerovu bázi, stačí sledovat jejich vedoucí členy. Prvním krokem je prosté vyházení všech prvků, které v tomto smyslu nejsou třeba.

Lemma. *Nechť G je Gröbnerova báze ideálu I a $p \in G$ takový, že $LT\ p \in \langle LT(G - \{p\})\rangle$. Pak $G - \{p\}$ je také Gröbnerova báze I.*

Důkaz. Z definice Gröbnerovy báze platí $\langle LT\ I\rangle = \langle LT\ G\rangle$. Protože $LT\ p \in \langle LT(G - \{p\})\rangle$, platí $\langle LT(G - \{p\})\rangle = \langle LT\ G\rangle$. Odsud již plyne tvrzení. □

Definice. *Minimální Gröbnerovou bází* ideálu I je taková Gröbnerova báze G, že pro všechna $p \in G$ platí $LC\ p = 1$ a zároveň $LT\ p \notin \langle LT(G - \{p\})\rangle$

Například mějme $\mathbb{K}[x, y]$ a $<_{\text{grlex}}$, $I = \langle f_1, f_2\rangle = \langle x^3 - 2xy, x^2y - 2y^2 + x\rangle$. Zmíněný algoritmus dá pět polynomů $F = (f_1, \ldots, f_5)$

$$F = (x^3 - 2xy, x^2y - 2y^2 + x, -x^2, -2xy, -2y^2 + x).$$

Přitom platí $LT\ f_1 = x^3 = -x\ LT\ f_3$ a $LT\ f_2 = -\frac{1}{2}x\ LT\ f_4$ a tedy f_1 a f_2 jsou zbytečné.

Tato redukce nám ale jistě ještě nestačí, protože k redudanci může docházet i na úrovni jednotlivých členů bázových prvků. Např. si můžeme všimnout, že pro každé a je $\{x^2 + axy, xy, y^2 - 1/2x\}$ minimální Gröbnerovou bází uvedeného ideálu.

Proto zavádíme následující pojem:

Redukovaná Gröbnerova báze

Polynom $g \in G$ nazveme *redukovaný* pro bázi G pokud žádný z jeho monomů neleží v $\langle LT(G - \{g\})\rangle$. *Redukovanou Gröbnerovou bází* ideálu I potom nazveme takovou Gröbnerovu bázi G, že pro všechna $p \in G$ platí $LC\ p = 1$ a zároveň p je redukovaný pro G.

[7] V angličtině se podmínce stabilizace každého neklesajícího řetězce ideálů říká ACC, „ascending chain condition“. Okruhy, které splňují ACC se nazývají Noetherovské (na počest Emy Noether). Hilbertovu větu lze tedy ekvivalentně formulovat jako „Okruh polynomů nad noetherovským okruhem je opět noetherovský“.

Zjevně je každá redukovaná Gröbnerova báze je minimální a navíc platí:

Tvrzení. *Je-li polynom g redukovaný pro nějakou minimální Gröbnerovu bázi G ideálu I, pak je také redukovaný pro každou minimální Gröbnerovu bázi G' téhož ideálu, která jej obsahuje.*

Důkaz. Tvrzení dokážeme sporem. Uvažme $G = \{g_1, \ldots, g_s\}$, $G' = \{g'_1, \ldots, g'_t\}$ a $g = \cdots + m + \cdots$ kde $m \in \langle LT(G' - \{g\})\rangle$ (tj. g není redukovaný pro G'). Potom $m = a_1\, LT\, g'_1 + \cdots + a_t\, LT\, g'_t$ pro nějaké vhodné polynomy $a_1, \ldots, a_t$. Protože G i G' jsou Gröbnerovy báze téhož ideálu, platí $\langle LT\, G\rangle = \langle LT\, G'\rangle$, a tedy každé $LT\, g'_i$ lze vyjádřit jako kombinaci $LT\, g_1, \ldots, LT\, g_s$. Odtud už plyne $m \in \langle LT\, G\rangle$ a protože je G' minimální, je $m \in \langle LT(G \setminus \{g\})\rangle$, což je spor s předpokládanou redukovaností g pro G. □

Nyní již máme vše připraveno pro důkaz hlavního výsledku o existenci a jednoznačnosti redukované Gröbnerovy báze.

10.66 **11.41. Věta.** *Nechť $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ je nenulový. Pak pro každé monomiální uspořádání existuje právě jedna redukovaná Gröbnerova báze ideálu I. Navíc každou Gröbnerovu bázi lze algoritmicky redukovat.*

Důkaz. Předpokládejme, že G je Gröbnerova báze ideálu I. S ohledem na lemma z předchozího odstavce lze předpokládat, že G je i minimální. (Algoritmus minimalizace je zřejmý, stačí testovat pouze dělitelnost vedoucích monomů v jakémkoliv pořadí a vypouštět nadbytečné členy báze.)

Předpokládejme, že polynom $g \in G$ není redukovaný. Při dělení $g/(G - \{g\})$ se tedy $LT\, g$ nutně dostane do zbytku, protože nemá čím být dělitelný (báze je minimální). Tedy $LT(\overline{g}^{\,G-\{g\}}) = LT\, g$, protože nic jiného už nemůže být vedoucím členem zbytku. Označme

$$g' := \overline{g}^{\,G-\{g\}} \quad G' := \big(G - \{g\}\big) \cup \{g'\}$$

Tento nový systém generátorů G' je opět minimální Gröbnerovou bází ideálu I, protože $\langle LT\, G'\rangle = \langle LT\, G\rangle$, tjtaké platí $\langle LT\, G'\rangle = \langle LT\, I\rangle$. Polynom g' je zřejmě redukovaný pro G' díky vlastnostem algoritmu pro dělení. Byl-li nějaký $h \neq g$ redukovaný pro G, zůstává podle předchozího tvrzení z předchozho odstavce redukovaný i pro G'.

Při každé provedené redukci některého z prvků dojde ke zmenšení celkového počtu členů polynomů v G. Proto se algoritmus zastaví v okamžiku, kdy už jsou všechny prvky redukované a máme tedy algoritmus konstrukci redukované Gröbnerovy báze.

Zbývá dokázat její jednoznačnost. Předpokládejme dvě redukované Gröbnerovy báze G, $\tilde{G}$ nenulového ideálu I. Platí tedy $\langle LT\, G\rangle = \langle LT\, I\rangle = \langle LT\, \tilde{G}\rangle$. Protože tento ideál je monomiální, lze pro něj aplikovat Dicksonovo lemma. S odvoláním na konstrukci báze v jeho důkazu lze tvrdit, že existuje právě jedna monomiální báze monomiálního ideálu tak, že

koeficienty jejích členů jsou rovny jedné a žádný z členů této báze nedělí jiný.

Podle definice minimality musí být $LT\ G$ i $LT\ \tilde{G}$ právě takovou bází. Tedy $LT\ G = LT\ \tilde{G}$. Ke každému $g \in G$ tedy exisuje právě jedno $\tilde{g} \in \tilde{G}$ takové, že $LT\ g = LT\ \tilde{g}$.

Platí $g - \tilde{g} \in I$. Protože G je Gröbnerova, platí $\overline{g - \tilde{g}}^{\,G} = 0$. Členy $LT\ g$, $LT\ \tilde{g}$ se odečtou už v $g - \tilde{g}$. Protože obě báze jsou redukované, nemůže být žádný ze zbývajících členů $g - \tilde{g}$ dělitelný kterýmkoli z $LT\ G = LT\ \tilde{G}$. Musí se tedy dostat do zbytku. Platí tedy

$$g - \tilde{g} = \overline{g - \tilde{g}}^{\,G} = 0$$

Tím je jednoznačnost dokázána. □

10.66a

11.42. Poznámky. Máme již k dispozici několik odpovědí

na dříve položené otázky. Umíme totiž účinně rozhodnout o příslušnosti polynomu do daného ideálu pomocí dělení se zbytkem prostřednictvím Gröbnerovy báze. A umíme také pomocí redukovaných Gröbnerových bází rozhodnout, zda jsou dva ideály stejné.

Pro náš problém řešení systémů polynomiálních rovnic to znamená, že pro daný systém polynomiálních rovnic umíme rozhodnout, zda nějaká jiná polynomiální rovnice je jejich důsledkem. Umíme také o dvou různých systémech algoritmicky rozhodnout, zda generují stejný ideál svých důsledků.

Při těchto algoritmických konstrukcích bude záležet na zvoleném uspořádání monomů, samotné odpovědi na výše uvedené otázky ale na uspořádání nezávisí.

Jak jsme zmiňovali v úvodu kapitoly, technika Gröbnerových bazí je jedním ze základů počítačové algebry. Samozřejmě jsou při implementacích v programových systémech využita různá zlepšení výše uvedeného algoritmu. Např. je možné využít techniky redukcí již během vytváření Gröbnerovy báze v základním algoritmu z odstavce 11.39 apod.

V literatuře lze dohledat také různé varianty pro nekomu-

tativní algebraické objekty (např. při formálních manipulacích s diferenciálními operátory). Algoritmus pro nalezení Gröbnerovy báze lze také interperetovat jako speciální případ Knuth-Bendixova algoritmu pro přepisovací pravidla řešící problém ekvivalentnosti slov v monoidech zadaných generátory a sadou rovností.

Konečně, v samotné komutativní algebře je technika Gröbnerových bazí použitelná daleko sofistikovaněji. Při průchodu naším algoritmem totiž dostáváme syzygy všech dvojic generátorů konečné báze. Tyto syzygy jsou vlastně bazí tzv. podmodulu všech relací mezi k prvky $g_1, \ldots, g_k$ báze, tj. podmnožiny S v prostoru $(\mathbb{K}[x_1, \ldots, x_n])^k$. Na takové podmnožiny opět můžeme rozšířit samotný algoritmus a najít význačné generátory všech relací mezi mezi generátory. Takto můžeme pokračovat, dokud existují nějaké netriviální relace. Lze dokázat, že nejpozději po n takových krocích už žádné netriviální relace nebudou existovat a počty generátorů

relací v jednotlivých krocích nám dávají velmi podrobnou informaci o topologických vlastnostech příslušné afinní variety $\mathfrak{V}(g_1, \ldots, g_k)$.

10.67

11.43. Eliminace proměnných. Na závěr této části si uvedeme alespoň jednu aplikaci předchozích algoritmů.

Budeme považovat okruh $\mathbb{K}[x_{p+1}, \ldots, x_n]$ za podokruh $\mathbb{K}[x_1, \ldots, x_n]$. Jedná se o polynomy, v nichž se nevyskytují proměnné $x_1, \ldots, x_p$. Je to skutečně podokruh, ale už ne ideál.

ELIMINAČNÍ IDEÁLY

Nechť $I = \langle f_1, \ldots, f_s \rangle \subseteq k[x_1, \ldots, x_n]$. Pro $p = 1, \ldots, n$ definujeme

$$I_p := I \cap \mathbb{K}[x_{p+1}, \ldots, x_n]$$

Tuto množinu nazveme *p-tým eliminačním ideálem.* Všimněme si, že I_p je ideálem pouze v $k[x_{p+1}, \ldots, x_n]$.

Na úrovni polynomiálních rovnic I_p obsahuje všechny rovnice, které jsou důsledky systému $f_1 = 0, \ldots, f_s = 0$ a ve kterých vystupují pouze proměnné $x_{p+1}, \ldots, x_n$.

Věta (Eliminační věta)**.** *Nechť $I \subseteq \mathbb{K}[x_1, \ldots, x_n]$ je ideál, $G = \{g_1, \ldots, g_m\}$ jeho Gröbnerova báze vzhledem k $<_{lex}$. Proměnné nechť jsou uspořádány $x_1 >_{lex} x_2 >_{lex} \cdots$. Potom pro každé $p = 0, \ldots, n$ je $G_p := G \cap \mathbb{K}[x_{p+1}, \ldots, x_n]$ Gröbnerovou bází ideálu I_p.*

Jestliže je G minimální, resp. redukovaná, Gröbnerova báze, pak G_p je opět báze minimální resp. redukovaná.

DŮKAZ. Bez újmy na obecnosti můžeme uvažovat $G_p = \{g_1, \ldots, g_r\}$. Protože $G \subseteq I$, je i $G_p \subseteq I_p$. Inkluze $\langle G_p \rangle \subseteq I_p$ platí triviálně. Dokážeme tedy ikluzi opačnou.

Pro libovolný polynom $f \in I_p$ bychom rádi ověřili, že

$$f = h_1 g_1 + \cdots + h_r g_r.$$

Provedeme za tím účelem dělení se zbytkem původní Gröbnerovou bází G. Protože je také $f \in I$, platí $\overline{f}^G = 0$, a tedy

$$f = h_1 g_1 + \cdots + h_r g_r + h_{r+1} g_{r+1} \cdots + h_m g_m$$

Každý z polynomů $g_{r+1}, \ldots, g_m$ musí obsahovat nějakou z proměnných $x_1, \ldots, x_p$, jinak by byl prvkem G_p. Vzhledem k vlastnostem lexikografického uspořádání takovou proměnnou obsahují i $LT\, g_{r+1}, \ldots, LT\, g_m$. Uvědomíme-li si postup algoritmu pro dělení se zbytkem a skutečnost, že v f není žádný monom obsahující některou z $x_1, \ldots, x_p$, musí být $h_{r+1} = \cdots = h_m = 0$. Ověřili jsem proto $f \in \langle G_p \rangle$.

Dokázali jsme nejen požadovanou inkluzi, ale i fakt, že dělení f/G dopadne na I_p stejně jako f/G_p. Pro $1 \leq i < j \leq r$ uvažujme S-polynomy $S(g_i, g_j)$. Platí

$$\overline{S(g_i, g_j)}^{G_p} = \overline{S(g_i, g_j)}^{G} = 0$$

a tedy G_p je Gröbnerova báze ideálu I_p.

Tvrzení o minimalitě nebo redukovanosti báze je zřejmé z definic těchto pojmů. □

Jediná vlastnost lexikografického uspořádání, kterou jsme použili v důkazu, je tvrzení, že pokud se některé proměnné objevují v polynomu f, pak se objevují v jeho vedoucím členu. To je ovšem podstatně slabší požadavek, než definice lexikografického uspořádání. Proto lze při skutečných implementacích používat jakékoliv uspořádání, které bude zajišťovat tuto vlastnost. Dosáhne se tak většinou efektivnějších výpočtů, protože čisté lexikografické uspořádání zpravidla vede k nepříjemnému nárůstu stupňů polynomů.

10.68

11.44. Implicitní popis parametrizovaných variet. Z předchozí věty lze docela snadno odvodit algoritmus pro nalezení implicitního popisu variet zadaných pomocí polynomiální parametrizace. Nebudeme se věnovat detailní diskusi, protože nemáme k dispozici všechny nástroje pro práci s nejménšími varietami obsahujícími body zadané parametrizací. Zůstaneme proto na úrovni poznámek.

Jestliže je naše parametrizace variety dána polynomiálními vztahy

$$x_1 = f_1(u_1, \ldots, u_k), \ldots, x_n = f_n(u_1, \ldots, u_k),$$

spočteme redukovanou Gröbnerovu bázi ideálu

$$\langle x_1 - f_1, \ldots, x_n - f_n \rangle$$

v lexikografickém uspořádání, kde $u_i > x_j$ pro všechna i, j. Z této báze dostaneme redukovanou Gröbnerovu bázi eliminačního ideálu I_k a to je přesně hledaný ideál a jeho implicitní popis.

Ve skutečnosti nám pro výpočet stačí takové uspořádání, které zaručí převahu všech u_i nad x_j, aby se algoritmem pro výpočet Gröbnerovy báze eliminovala u_i, jinak může být uspořádání libovolné. Máme tak naději dosáhnout efektivnějšího výpočtu, než s čistým lexikografickým uspořádáním.

Jako příklad si uveďme už dříve zobrazenou varietu v $\mathbb{R}^3$ nazývnou Enneperova plocha. Její parametrický popis byl $x = 3u + 3uv^2 - u^3$, $y = 3v + 3u^2v - v^3$, $z = 3u^2 - 3v^2$. Aplikace eliminační procedury (např. v systému MAPLE za použití `gbasis` s uspořádáním `plex`) dá odpovídající implicitní popis, tj. rovnici s jediným polynomem devátého stupně: $-59049z - 104976z^2 - 6561y^2 - 72900z^3 - 18954y^2z - 23328z^4 + 32805z^2x^2 + 14580z^3x^2 + 3645z^4x^2 - 1296y^4z - 16767y^2z^2 - 6156y^2z^3 - 783y^2z^4 + 39366zx^2 + 19683x^2 - 1296y^4 - 2430z^5 + 432z^6 + 108z^7 + 486z^5x^2 - 432y^4z^2 + 54y^2z^5 + 27z^6x^2 - 48y^4z^3 + 15y^2z^6 - 64y^6 - z^9$.

Když je naše parametrizace racionální, tj.

$$x_i = \frac{f_i(t_1, \ldots, t_m)}{g_i(t_1, \ldots, t_m)},$$

asi nás hned napadne dosadit do předchozí věty ideál $\langle x_1 g_1 - f_1, \ldots, x_n g_n - f_n \rangle$. To ale většinou nefunguje dobře. Například uvažujme

$$x = \frac{u^2}{v} \quad y = \frac{v^2}{u} \quad z = u.$$

Dostali bychom $I = \langle vx - u^2, uy - v^2, z - u\rangle$ a po eliminaci $I_2 = \langle z(x^2y - z^3)\rangle$. Správný výsledek je ale jenom $\mathfrak{V}(x^2y - z^3)$, tedy náš postup přidal navíc celou rovinu.

Problém je v tom, že zahrnujeme i celou varietu nulových bodů jmenovatelů v parametrizacích jednotlivých proměnných, $W = \mathfrak{V}(g_1, \ldots, g_n)$. Raději tedy parametrizaci F chápejme jako zobrazení $F : (\mathbb{K}^k - W) \to \mathbb{K}^n$. Pro implicitizaci pak použijeme ideál

$$\begin{aligned} I &= \langle g_1x_1 - f_j, \ldots, g_nx_n - f_n, 1 - g_1 \cdots g_n y\rangle \\ &\subseteq \mathbb{K}[y, t_1, \ldots, t_m, x_1, \ldots, x_n], \end{aligned}$$

kde si navíc pomáháme dodatečnou proměnnou y. Potom lze ukázat, že $V(I_{k+1})$ je minimální afinní varieta obsahující $F(\mathbb{K}^m - W)$.

4. Uspořádané množiny a Booleovská algebra

Tak jako jsme z vlastností čísel nebo symetrií objektů abstrahovali podstatné axiomy a dostali jsme daleko šířeji použitelné nástroje pro úvahy v linerární algebře, při diskusi grup symetrií a jejich akcí, studium okruhů polynomů atd.

Nyní budeme postupovat obdobně a okamžitě uvidíme, že jen docela drobnou změnou základních vlastností dostaneme na první pohled úplně jiné objekty. To co zůstane podobné je algebraická práce se symboly zastupujícími velice rozmanité objekty a tím pádem i docela univerzální použitelnost výsledků.

Za východisko si vezmeme základní operace s množinami, tj. jejich sjednocení, průnik a vztahy inkluze a naším prvním cílem bude uvést tyto operace do souvislosti s výrokovou logikou (tj. formalizovanými postupy pro vyjadřování výroků a vyhodnocování jejich pravdivosti).

10.21

11.45. Množinová algebra. S každou množinou M máme k dispozici také množinu $K = 2^M$ všech jejích podmnožin a na ní operace $\vee : K \times K \to K$ sjednocení množin a $\wedge : K \times K \to K$ průniku množin. To jsou dvě *binární operace*, které jsme dosud značili $\cup$ a $\cap$.

Dále máme ke každé množině $A \in K$ také její množinu doplňkovou $A' = K \setminus A$, což je další *unární operace*.

Konečně máme „největší objekt“, tj. celou množinu M, který je neutrální vůči operaci $\wedge$ a který proto budeme v této souvislosti označovat jako 1, a obdobně se chová prázdná množina $\emptyset \in K$ vůči operaci $\vee$. Tu budeme v této souvislosti značit jako 0.

Na množině K všech podmnožin v M můžeme velmi snadno ověřit pro všechny prvky A, B, C následující vlastnosti (již jsme definovali význačné prvky $0 = \emptyset$ a $1 = M$ a

unární operaci vzetí doplňku A' k podmnožině A):

puvodni

(1) $A \wedge (B \wedge C) = (A \wedge B) \wedge C$,

(2) $A \vee (B \vee C) = (A \vee B) \vee C$,

(3) $A \wedge B = B \wedge A,\ A \vee B = B \vee A$,

(4) $A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C)$,

(5) $A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C)$,

(6) existuje $0 \in K$ tak, že $A \vee 0 = A$,

(7) existuje $1 \in K$ tak, že $A \wedge 1 = A$,

(8) $A \wedge A' = 0, \quad A \vee A' = 1$.

Porovnejme si tyto vlastnosti s vlastnostmi okruhů:

První dvě z nich, tj. (1) a (2) říkají, že obě operace $\wedge$ a $\vee$ jsou asociativní. Vlastnost (3) konstatuje komutativitu obou operací.

Až potud je tedy vše jako u číslených oborů a operací sčítání a násobení. Zásadní změnou jsou ale další dvě vlastnosti (4) a (5), protože ty vyžadují jak distributivitu operace $\vee$ vůči průniku $\wedge$, tak naopak. To pochopitelně u sčítání a násobení čísel nejde — máme tam pouze distributivitu sčítání vůči násobení, ale ne naopak.

Poslední tři vlastnosti (6) – (8) konstatují existenci neutrálních prvků vůči oběma operacím, ale také existenci obdoby k „inverzím“ ke všem prvkům (ale všimněme si, že průnikem s komplementem chceme dostat neutrální prvek ke sjednocení a naopak, tedy odlišně od vzetí inverzí v okruzích). Jistě nás nepřekvapí, když zachvíli uvidíme, že takto silně provázané vlastnosti dvou různých operací nemůže mít příliš mnoho objektů.

Booleovské algebry

Definice. Množině K spolu s dvěmi binárními operacemi $\wedge$ a $\vee$ a jednou unární operací $'$ splňující vlastnosti (1)–(8) říkáme *Booleovská algebra*. Operaci $\wedge$ budeme říkat *infimum* (případně *sjednocení*, anglicky často také *meet*), operaci $\vee$ budeme říkat *supremum* (případně *průnik*, anglicky také *join*). Prvku A' se říká *doplněk* k prvku A.

Všimněme si, že axiomy Booleovské algebry jsou zcela symetrické vůči záměně operací $\wedge$ a $\vee$, společně se záměnou prvků 0 a 1. Důsledkem tohoto faktu je, že jakékoliv tvrzení, které odvodíme z axiomů, má také platné *duální tvrzení*, které vznikne z prvého právě záměnou všech výskytů $\wedge$ za $\vee$ a naopak a stejně tak všech výskytů 0 a 1. Hovoříme o *principu duality*.

10.21a

11.46. Vlastnosti Booleovských algeber. Jako obvykle si hned odvodíme několik elementárních důsledků axiomů. Zejména si povšimněme, že tak dokážeme u speciálního případu Booleovské algebry všech podmnožin v dané množině M elementární vlastnosti známé z množinové algebry. Např. je doplněk k $A \in K$ určen svými vlastnostmi jednoznačně. V obecném pohledy však toto pozorování říká,

že máme-li dáno $(K, \wedge, \vee)$, může existovat nejvýše jedna unární operace $'$, se kterou dostaneme Booleovskou algebru $(K, \wedge, \vee, ')$.

Skutečně, pokud B a $C \in K$ splňují vlastnosti A' (tj. poslední axiom (8) v definici výše), platí

$$\begin{aligned} B &= B \vee 0 = B \vee (A \wedge C) \\ &= (B \vee A) \wedge (B \vee C) = 1 \wedge (B \vee C) = B \vee C \end{aligned}$$

a stejně také spočteme

$$C = C \vee B.$$

Je tedy nutně $B = C$. Všimněme si, že použitím tohoto výsledku na prvky 1 a 0, společně s jejich definicí, okamžitě dostáváme jednoznačnost pro tyto výjimečné prvky v libovolné Booleovské algebře (promyslete si podrobně!).

Vlastnosti v následujícím tvrzení mají svá zavedená jména v množinové algebře: vlatnosti (2) se říká *absorpční zákony*, vlastnosti (3) popisují *idempotentnost* operací $\wedge$ a $\vee$ a rovnosti (4) jsou známy jako *De Morganova pravidla*.

Tvrzení. *V každé Booleovské algebře* $(K, \wedge, \vee, ')$ *platí pro všechny prvky v* K

(1) $A \wedge 0 = 0, \quad A \vee 1 = 1$
(2) $A \wedge (A \vee B) = A, \quad A \vee (A \wedge B) = A$
(3) $A \wedge A = A, \quad A \vee A = A$
(4) $(A \wedge B)' = A' \vee B', \quad (A \vee B)' = A' \wedge B'$
(5) $(A')' = A$.

Důkaz. Podle principu duality potřebujeme z každého z duálních tvrzení na jednotlivých řádcích dokázat pouze jedno. Počítejme s využitím axiomů:

Začneme s vlastností (3)

$$A = A \wedge 1 = A \wedge (A \vee A') = (A \wedge A) \vee 0 = A \wedge A.$$

Nyní už umíme snadno (1)

$$A \wedge 0 = A \wedge (A \wedge A') = (A \wedge A) \wedge A' = A \wedge A' = 0$$

a pak je snadné i (2)

$$\begin{aligned} A \wedge (A \vee B) &= (A \vee 0) \wedge (A \vee B) \\ &= A \vee (0 \wedge B) = A \vee 0 = A. \end{aligned}$$

K důkazu De Morganových pravidel stačí ověřit, že $A' \vee B'$ má vlastnosti doplňku k $A \wedge B$, pak to totiž bude doplněk dle úvahy výše. S využitím (1) spočteme

$$\begin{aligned} (A \wedge B) \wedge (A' \vee B') &= ((A \wedge B) \wedge A') \vee ((A \wedge B) \wedge B') \\ &= (0 \wedge B) \vee (A \wedge 0) = 0. \end{aligned}$$

Obdobně, s použitím (2) dostáváme

$$\begin{aligned} (A \wedge B) \vee (A' \wedge B') &= (A \vee (A' \vee B')) \vee (B \vee (A' \vee B')) \\ &= (1 \vee B') \wedge (1 \vee A') = 1. \end{aligned}$$

Konečně, přímo z definice je $A' \wedge A = 0$ a $A' \vee A = 1$, má proto A požadované vlastnosti doplňku k A' a je tedy $A = (A')'$. □

10.21b

11.47. Příklady Booleovských algeber. Nejmenší zajímavá algebra je množina všech podmnožin jednoprvkové množiny M. Ta má právě dva prvky $0 = \emptyset$ a $1 = M$. Operace $\wedge$ a $\vee$ v tomto případě splývají s násobením a sčítáním v okruhu abytkových tříd $\mathbb{Z}_2$, proto budeme nadále hovořit o Booleovské algebře $\mathbb{Z}_2$.

Podobně jako u vektorových prostorů a okruhů můžeme algebraickou strukturu Booleovské algebry přenášet na prostory funkcí, jejichž obor hodnot Booleovskou algebrou je. Skutečně, pro množinu všech funkcí $S = \{f : M \to K\}$ zmnožiny M do Booleovské algebry $(K, \wedge, \vee, ')$ definujeme potřebné operace a vybrané prvky 0 a 1 na S jako funkce v argumentu $x \in M$ takto:

$$\begin{aligned}(f_1 \wedge f_2)(x) &= (f_1(x)) \wedge (f_2(x)) \in K\\ (f_1 \vee f_2)(x) &= (f_1(x)) \vee (f_2(x)) \in K\\ (1)(x) = 1 \in K,\ &(0)(x) = 0 \in K\\ (f)'(x) &= (f(x))' \in K.\end{aligned}$$

Ověření, že tyto nové operace skutečně zadávají Booleovskou algebru je zcela přímočaré a jednoduché.

Připoměňme, že všechny podmnožiny dané množiny M lze interpretovat jako zobrazení $M \to \mathbb{Z}_2$, když na jedničku zobrazíme právě body vybrané podmnožiny. Pak skutečně můžeme sjednocení a průnik definovat výše uvedeným způsobem — např. o každém bodu $x \in M$ rozhodujeme u $(A \wedge B)(x)$ zda patří do A a zda patří do B a vezmeme sjednocení výsledků v $\mathbb{Z}_2$, tj. výsledek bude 1, právě když x padne do sjednocení.

10.22

11.48. Výroková logika. V předchozích odstavcích jsme použili symboliku, kterou je často rozumné interpretovat tak, že z prvků $A, B, \ldots \in K$ tvoříme „slova“ pomocí operací $\vee$, $\wedge$, $'$ a závorek vyjasňujících v jakém pořadí a na jaké argumenty jsou operace aplikovány. Samotné axiomy Booleovských algeber a jejich důsledky pak říkají, že velice často různá slova dávají stejnou hodnotu výsledku v K.

V případě množiny všech podmnožin $K = 2^M$ je to zřejmé – prostě jde o rovnost podmnožin. Nyní uvedeme stručně jinou podobnou souvislost.

Budeme pracovat opět se slovy jako výše, interpretujeme je ale jako tvrzení složené z elementárních výroků $A, B, \ldots$ a logických operací AND (binární operace $\wedge$), OR (binární operace $\vee$) a negace NOT (unární operace $'$). Taková slova nazýváme *výroky* a přiřazujeme jim pravdivostní hodnotu v závislosti na pravdivostní hodnotě jednotlivých elementárních argumentů. Pravdivostní hodnotu přitom bereme jako prvek z triviální Booleovy algebry $\mathbb{Z}_2$, tedy buď 0 nebo 1. Pravdivostní hodnota výroku je plně určena přiřazením hodnot pro nejjednoduší výroky $A \wedge B$, $A \vee B$ a A', tj. $A \wedge B$ je pravdivé pouze, když jsou oba výroky A a B pravdivé, $A \vee B$

je nepravdivé pouze. když jsou oba výroky nepravdivé a A' má opačnou hodnotu než A.

Výrok obsahující n elementárních výroků tedy představuje funkci $(\mathbb{Z}_2)^n \to \mathbb{Z}_2$ a dva výroky nazýváme logicky ekvivalentní, jestliže zadávají stejnou funkci. V předchozím příkladu jsme již ověřili, že na množině tříd logicky ekvivalentních výroků je dána struktura Booleovy algebry. Nutně tedy pro výrokovou logiku bude v tomto smyslu platné vše, co dokážeme pro obecné Booleovy algebry.

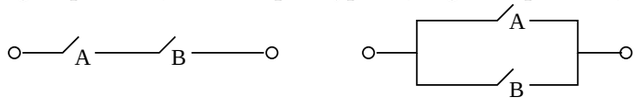
Stručně si proberme, jak vypadají obvyklé další jednoduché výroky ve výrokové logice jakožto prvky Booleovy algebry (tj. reprezentujeme vždy naším výrazem třídu výroků ekvivalentních):

Implikaci $A \Rightarrow B$ dostaneme jako $A' \vee B$, ekvivalenci $A \Leftrightarrow B$ odpovídá $(A \wedge B) \vee (A' \wedge B')$. Dále vylučovací OR, neboli XOR, je dáno jako $(A \wedge B') \vee (A' \wedge B)$, negace NOR operace OR je vyjádřena jako $A' \wedge B'$ a negace NAND operace AND je dána jako $A' \vee B'$. Konečně tautologie je dána pomocí libovolného elementárního výroku jako $A \vee A'$.

Všimněme si také, že XOR odpovídá v množinové algebře symetrickému rozdílu množin.

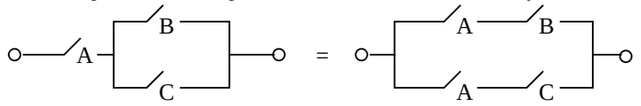
10.23

11.49. Přepínače jako Booleova algebra. Přepínač je pro nás černá skříňka, která má jen dva stavy, buď je zapnut (a signál prochází) nebo naopak vypnut (a signál neprochází).



Jeden nebo více přepínačů zapojujeme do sítě sériově nebo paralelně. Sériové zapojení je popsáno pomocí binární operace $\wedge$, paralelní je naopak $\vee$. Unární operace A' zadává přepínač, který je vždy v opačné poloze než A. Každé konečné slovo vytvořené pomocí přepínačů $A, B, \ldots$ a operací $\wedge$, $\vee$ a $'$ umíme převést na obrázek, který bude představovat systém přepínačů propojených dráty a zcela obdobně jako v minulém odstavci nám každá volba poloh jednotlivých přepínačů zadá hodnotu „zapnuto/vypnuto“ pro celý systém.

Opět se snadno krok po kroku ověří platnost základních axiomů Booleových algeber pro náš systém. Na následujícím obrázku je ilustrován jeden z axiomů distributivity.



Propojení bez přepínače odpovídá prvku 1, koncové body bez propojení (nebo sériové zapojení A a A') dává prvek 0. Nakreslete si obrázky pro všechny axiomy Booleovské algebry a ověřte si je!

10.24

11.50. Dělitelé. Dalším přirozeným příkladem Booleovské algebry může být systém dělitelů přirozeného čísla nebo polynomu.

Začněme trochu obecněji a zvolme pevně takové přirozené číslo $p \in \mathbb{N}$ nebo polynom $p \in \mathbb{K}[x_1, \ldots, x_s]$ nad oborem integrity $\mathbb{K}$ s jednoznačným rozkladem. Za nosnou množinu D_p bereme množinu všech dělitelů q našeho p. Pro dva takové dělitele definujeme $q \wedge r$ jako největší společný dělitel prvků q a r, $q \vee r$ je nejmenší společný násobek. Dále definujeme význačný prvek $1 \in D_p$ jako naše číslo nebo polynom p a neutrálním prvkem 0 vůči supremu na D_p je jednička v $\mathbb{N}$, resp. $1 \in \mathbb{K} \subset \mathbb{K}[x_1, \ldots, x_s]$. Unární operaci $'$ definujeme pomocí dělení: $q' = p/q$.

Lemma. *Množina D_p spolu s výše uvedenými operacemi $\wedge$, $\vee$ a $'$ je Booleovská algebra, právě když rozklad p na nerozložitelné faktory neobsahuje žádné kvadráty (tj. v jednoznačném rozkladu $p = q_1 \ldots q_n$ na nerozložitelné faktory jsou všechna q_i po dvou různá).*

Důkaz. Ověření axiomů je vcelku snadné, projdeme jeden po druhém a budeme zkoumat, kdy je zapotřebí nešeho požadavku na nepřítomnost kvadrátů v rozkladu.

Největší společný dělitel konečného počtu čísel nebo polynomů nezávisí na pořadí, ve kterém jej počítáme. Stejně tak pro nejmenší společný násobek. To odpovídá axiomům (1) a (2) v 11.45. Komutativita, tj. axiom (3) je zcela zřejmá.

Pro tři libovolné prvky a, b, a c můžeme bez újmy na obecnosti psát jejich rozklad ve tvaru $a = q_1^{p_1} \ldots q_s^{p_s}$, $b = q_1^{m_1} \ldots q_s^{m_s}$ a $c = q_1^{k_1} \ldots q_s^{k_s}$, kde připouštíme i mocniny 0 a všechny prvky q_j jsou po dvou nesoudělné. Prvek $a \wedge b \in D_p$ je tedy prvek s rozkladem, ve kterém se objeví všechna společná q_i v mocnině, která bude minimem z mocnin v a a b. Naopak $a \vee b$ bude mít rozklad, ve kterém se objeví všechny členy z rozkladů a a b a to s mocninou, která bude tou větší z mocnin příslušného faktoru v a a b. Z této úvahy nyní snadno plynou distributivní zákony (4) a (5) z 11.45.

Problém nemáme ani s existencí prvků 0 a 1, které jsme přímo definovali a zjevně splňují axiomy (6) a (7). Případná existecne kvadrátů v rozkladech ale znemožní určení doplňku. Např. v $D_{12} = \{1, 2, 3, 4, 6, 12\}$ nelze $6 \wedge 6' = 1$ dosáhnout, protože má 6 netriviálního společného dělitele se všemi ostatními prvky v D_{12} mimo jedničku, ta ovšem nesplňuje $6 \vee 1 = 12$.

Pokud ovšem nejsou v rozkladu čísla nebo polynomu p kvadráty, definujeme doplněk pomocí dělení jako $q' = p/q$ a snadno ověříme potřebné vlastnosti z axiomů (6)–(8). □

10.25

11.51. Částečná uspořádání. K Booleovským algebrám teď půjdeme z jiné strany. Základní strukturou pro nás bude pojem *uspořádání*. Vzpomeňme na definici uspořádání jakožto reflexivní, antisymetrické a tranzitivní relace $\leq$ na množině K. Taková relace obecně neříká o každé dvojici $a, b \in K$ jestli je $a \leq b$ nebo $b \leq a$ (takové uspořádání se nazývá *úplné uspořádání* nebo dobré uspořádání). Často v našem případě obecného uspořádání proto hovoříme také o *částečném uspořádání* a množina $(K, \leq)$ vybavená

částečným uspořádáním se nazývá *poset* (z anglického „partially ordered set“).

Takové uspořádání je zejména vždy na množině $K = 2^M$ všech podmnožin množiny M prostřednictvím inkluze podmnožin. Pomocí naší relace infima na K je můžeme definovat jako $A \subset B$ právě, když $A \wedge B = A$. Ekvivalentně, $A \subset B$ právě, když $A \vee B = B$.

Lemma. *Je-li $(K, \wedge, \vee, ')$ Booleova algebra, pak relace $\leq$ definovaná vytahem $A \leq B$ právě, když $A \wedge B = A$, je částečné uspořádání. Navíc platí pro všechny prvky $A, B, C \in K$*

(1) $A \wedge B \leq A$
(2) $A \leq A \vee B$
(3) jestliže $A \leq C$ a zároveň $B \leq C$, pak také $A \vee B \leq C$
(4) $A \leq B$*, právě když* $A \wedge B' = 0$
(5) $0 \leq A$ *a* $A \leq 1$.

Důkaz. Všechny dokazované vlastnosti a vztahy jsou výsledkem jednoduchého výpočtu v Booleovské algebře K. Začněme s vlastnostmi uspořádání pro $\leq$. Reflexivita je přímým důsledkem idempotence: $A \wedge A = A$, tj. $A \leq A$. Podobně komutativita pro $\wedge$ zaručuje antisymetrii $\leq$, protože z $A \wedge B = A$ a zároveň $B \wedge A = B$ vyplývá

$$A = A \wedge B = B \wedge A = B.$$

Konečně z platnosti $A \wedge B = A$ a $B \wedge C = B$ vyvodíme

$$A \wedge C = (A \wedge B) \wedge C = A \wedge (B \wedge C) = A \wedge B = A,$$

což ověřuje tranzitivitu relace $\leq$.

Dále počítáme $(A \wedge B) \wedge A = (A \wedge A) \wedge B = A \wedge B$, takže $A \wedge B \leq A$.

Ze vztahu $A \wedge (A \vee B) = A$, viz 11.46(2), plyne $A \leq A \vee B$, což dokazuje tvrzení (2).

Distributivita společně s předpokladem (3) dává

$$(A \vee B) \wedge C = (A \wedge C) \vee (B \wedge C) = A \vee B,$$

takže skutečně platí (3).

Tvrzení (5) plyne přímo z axiomů pro význačné prvky 1 a 0.

Zbývá nám tvrzení (4). Jestliže $A \leq B$, pak $A \wedge B' = A \wedge B \wedge B' = 0$. Naopak je-li $A \wedge B' = 0$, pak $A = A \wedge 1 = A \wedge (B \vee B') = (A \wedge B) \vee (A \wedge B') = (A \wedge B) \vee 0 = A \wedge B$. Odtud $A \leq B$ a důkaz je ukončen. □

Všimněme si, že stejně jako v případě algebry podmnožin je v Booleovských algebrách $A \wedge B = A$ právě, když je $A \vee B = B$. Skutečně, je-li $A \wedge B = A$, pak z absorpčních zákonů plyne $A \vee B = (A \wedge B) \vee B = B$, a naopak. Můžeme proto pro definici částečného uspořádání stejně dobře používat také operaci $\vee$.

10.26

11.52. Svazy. 
Viděli jsme, že každá Booleova algebra zadává poset $(K, \leq)$. Zdaleka ne každý poset ovšem vzniká takovýmto způsobem. Např. triviální částečné uspořádání, kdy $A \leq A$ pro všechny A a všechny dvojice různých prvků jsou nesrovnatelné, samozřejmě z Booleovy algebry vzniknout nemůže, pokud je v K více než jeden prvek (viděli jsme, že největší a nejmenší prvek v Booleově algebře je totiž srovnatelný s každým prvkem). Zkusme se zamyslet, do jaké míry lze z uspořádání budovat operace $\wedge$ a $\vee$.

Pracujme s pevně zvoleným posetem $(K, \leq)$. O prvku $C \in K$ řekneme, že je *dolní závorou* pro nějakou množinu prvků $L \subset K$, je-li $C \leq A$ pro všechny $A \in L$. Prvek $C \in K$ je *infimem množiny* $L \subset K$, jestliže je dolní závorou a pro každou jinou dolní závoru D téže množiny platí $D \leq C$.

Obdobně definujeme *horní závory* a *supremum* podmnožiny L záměnou $\leq$ za $\geq$ v posledním odstavci.

Konečné posety se přehledně zobrazují pomocí orientovaných grafů. Prvky K jsou představovány uzly a hranou jsou spojeny právě prvky v relaci s orientací od většího k menšímu. *Hasseho diagram* posetu je zakreslení takového grafu v rovině tak, že větší prvky jsou zobrazeny vždy výš než menší (a orientace hran je tedy dána takto implicitně). Zvláště u malého počtu prvků množiny K je to velmi přehledný způsob, jak diskutovat různé příklady, viz obrázek níže.

SVAZY

Definice. *Svaz* je poset $(K, \leq)$, ve kterém každá dvouprvková množina $\{A, B\}$ má supremum $A \vee B$ a infimum $A \wedge B$. Hovoříme přitom o *úplném svazu*, jestliže existuje supremum a infimum každé podmnožiny v K.

Na svazu $(K, \leq)$ tedy máme definovány binární operace $\wedge$ a $\vee$ a přímo z definice je zjevná asociativita a komutativita těchto operací (dokažte si podrobně!).

Všimněme si také že jakýkoliv prvek v K je horní závorou pro prázdnou množinu, proto supremum prázdné množiny musí být menší než všechny prvky v K. Obdobně infimum prázdné množiny musí být větší než jakýkoliv prvek v K. Zejména tedy úplný svaz má vždy *největší* a *nejmenší prvek*.

Protože jsou binární operace $\wedge$ a $\vee$ asociativní a komutativní, jistě existují v každém svazu suprema a infima všech konečných neprázdných množin. V případě konečných posetů $(K, \leq)$ jde proto o úplný svaz tehdy a jen tehdy, když v něm existuje jediný největší prvek $1 \in K$ a jediný nejmenší prvek $0 \in K$.

O svazu říkáme, že je *distributivní*, jestliže operace $\wedge$ a $\vee$ splňují axiomy distributivity (4) a (5) z odstavce 11.46 na straně 513. Snadno lze ale nakreslit Hasseho diagram svazu, který není distributivní, viz obrázek níže.

Nyní už můžeme snadno definovat Booleovskou algebru v jazyce svazů: Booleovská algebra je distributivní svaz s největším prvkem 1 a nejmenším prvkem 0 takový, že v něm

existují ke všem prvkům komplementy (tj. prvky splňující vlastnost 11.45(8).

Ověřili jsme již, že v takovém případě jsou komplementy definovány jednoznačně (viz úvahy na začátku odstavce 11.46), takže je naše alternativní definice Booleovské algebry korektní.

Všimněme si také, při diskusi dělitelů daného čísla nebo polynomu p jsme narazili na distributivní svazy D_p, které jsou Booleovskou algebrou právě tehdy, když rozklad p neobsahuje kvadráty, viz Lemma 11.50.

10.28

11.53. Homomorfismy. Jak jsme již viděli u mnoha matematických struktur, o objektech se dozvídáme informace pomocí tzv. homomorfismů, tj. zobrazení, které zachovávají příslušné operace. Obzvlášť jednoduché je to u posetů:

Isotonní zobrazení

Homomorfismem posetů $(K, \leq_K)$ a $(L, \leq_L)$ rozumíme takové zobrazení $f : K \to L$, že z $A \leq_K B$ vždy vyplývá také $f(A) \leq_L f(B)$. Hovoříme přitom také o *izotonních zobrazeních*.

V případě svazů a Booleovských algeber definujeme homomorfismy podobně jako u okruhů:

Homomorfismy svazů a algeber

Zobrazení $f : (K, \wedge, \vee, ') \to (L, \wedge, \vee, ')$ se nazývá *homomorfismus Booleovských algeber*, jestliže pro všechny $A, B \in K$ platí

(1) $f(A \wedge B) = f(A) \wedge f(B)$
(2) $f(A \vee B) = f(A) \vee f(B)$
(3) $f(A') = f(A)'$.

Homomorfismus f je izomorfismus Booleovských algeber, jestliže je f bijektivní.

Podobně definujeme homomorfismy svazů jako zobrazení, která splňují vlastnosti (1) a (2).

Snadno se ověří, že bijektivnost f již zaručí, že f^{-1} je opět homomorfismem.

Z definice uspořádání na Booleových algebrách nebo svazech je zřejmé, že každý homomorfismus $f : K \to L$ bude také splňovat $f(A) \leq f(B)$ pro všechny prvky $A \leq B$ v K, půjde tedy vždy o izotonní zobrazení.

Jakkoliv umíme rekonstruovat operace suprema a infima z uspořádání, pokud toto vzniklo z Booleovy algebry, není pravda, že by každý homomorfismus posetů byl automaticky homomorfismem příslušných algeber nebo svazů, viz obrázek níže.

10.26a

11.54. Věty o pevném bodě. Mnoho praktických úloh spočívá v diskusi pevných bodů zobrazení $f : K \to K$ na nějaké množině K, tj. prvků $x \in K$ s vlastností $f(x) = x$. Naše úvahy o infimech a supremech umožňují překvapivě snadno odvodit velice silná tvrzení tohoto typu. Dokážeme

si jednu takovou klasickou větu, kterou odvodili Knaster a Tarski (ve speciálním případě Booleovské algebry podmnožin dané množiny již koncem dvacátých let 20. století, obecné tvrzení pak publikoval Tarski v r. 1955):

Věta (Tarského věta)**.** *Uvažujme libovolný úplný svaz* $(K, \wedge, \vee)$ *a libovolné isotonní zobrazení* $f : K \to K$. *Pak* f *má pevný bod a množina všech pevných bodů* f *je (s uspořádáním zděděným z* K*) opět úplný svaz.*

Důkaz. Označme $M = \{x \in K; x \leq f(x)\}$. Protože v K existuje minimální prvek, je jistě M neprázdná a protože f zachovává uspořádání, je $f(M) \subset M$. Označme dále $z_1 = \sup M$. Pak jistě pro $x \in M$ platí $x \leq z_1$, tedy také $f(x) \leq f(z_1)$. Přitom zároveň víme $x \leq f(x)$, takže $f(z_1)$ je horní závorou pro M. Pak ovšem nutně $z_1 \leq f(z_1)$, takže i $z_1 \in M$ a proto $f(z_1) \leq z_1$. Dokázali jsme tedy $f(z_1) = z_1$ a pevný bod je nalezen.

Trochu složitější je dokázat dovětek, že množina $Z \subset K$ všech pevných bodů zobrazení f je úplný svaz. Zřejmě jsme již našli její největší prvek $z_1 = \max Z$ a úplně stejným postupem s použitím infima a vlastnosti $f(x) \leq x$, místo definice M a jejího suprema, bychom našli také nejmenší bod $z_0 = \min Z$.

Uvažme nyní libovolnou neprázdnou množinu $Q \subset Z$ a označme $y = \sup Q$. Toto supremum sice nemusí ležet v Q, ukážeme ale, že bude přesto v Z existovat supremum i ve zděděném uspořádání $\leq_Z$ z uspořádání v K. Za tím účelem si označme $R = \{x \in K; y \leq x\}$, tj. množinu všech prvků v K větších než naše y. Přímo z definic je zřejmé, že tato množina R je, spolu s uspořádáním zděděným z K, opět úplný svaz a zúžení zobrazení f na R bude opět izotonní zobrazení $f_{|R} : R \to R$. Podle výše dokázaného tedy bude mít $f_{|R}$ nejmenší pevný bod $\bar{y}$. Samozřejmě je $\bar{y} \in Z$ a snadno nahlédneme, že ve skutečnosti je $\bar{y}$ supremem námi zvolené množiny Q vůči zděděnému uspořádání na Z. Přitom je možné, že $\bar{y} > y$. Obdobným postupem se zaměněnými relacemi a volbou infima najdeme i infimum libovolné neprázdné podmnožiny v Z. Největší a nejmenší prvek jsme již našli dříve a důkaz je ukončen. □

Poznámka. V literatuře lze najít mnoho variant vět o pevných bodech v různých souvislostech. Jednou z velmi užitečných je tzv. *Kleeneho věta*, jejíž tvrzení můžeme vyčíst z právě dokázané věty následujícím způsobem.

Jestliže (ve značení Tarského věty) uvážíme spočetnou podmnožinu v K tvořenou tzv. *Kleeneho řetězcem*

$$0 \leq f(0) \leq f(f(0)) \leq \dots,$$

pak supremum z této podmnožiny zjevně nemůže být větší, než kterýkoliv pevný bod zobrazení f. Skutečně, pokud je y pevný bod zobrazení f pak ze vztahu $0 \leq y$ dostaneme $f(0) \leq f(y) = y$ atd. Pokud má f navíc vlastnost, že „dostatečně“ zachovává suprema, můžeme dovodit že $f(z)$ bude opět supremem téhož řetězce a tedy pevný bod. Musí to

proto být nejmenší pevný bod. Toto tvrzení se nazývá *Kleeneho věta o pevném bodě* a má četná použití v teorii rekurzí, při diskusi zastavení algoritmů atd.

Nebudeme zde zacházet do podrobností kolem tzv. *spojitosti* zobrazení mezi posety. Přesné formulace Kleeneho věty a dalších souvislostí lze nalézt např. v **??**.

10.27

11.55. Normální tvary výrazů. Vrátíme se závěrem zpět k diskusi Booleovských algeber s konečným počtem prvků a ukážeme si jejich úplnou klasifikaci.

Při diskusi výrokové logiky jsme se potýkali s problémem, co vlastně jsou prvky příslušné Booleovy algebry. Formálně vzato jsme je definovali jako třídy ekvivalentních výroků. Jinak řečeno, pracovali jsme s hodnotovými funkcemi pro výroky s daným počtem argumentů. Vůbec jsme přitom neřešili obtížný problém, jak rozpoznat stejné výroky v tomto smyslu. Také jsme neřešili, jestli všechny formálně možné hodnotové funkce $(\mathbb{Z}_2)^n \to \mathbb{Z}_2$ lze zadat pomocí základních logických operací.

Zcela obdobně se můžeme tázat, jak poznat, zda dva systémy přepínačů mají stejnou funkci. Obdobně jako u výroků zde pro systém s n přepínači pracujeme s funkcemi $(\mathbb{Z}_2)^n \to \mathbb{Z}_2$ a zjevně existuje právě 2^{2^n} různých takových přepínacích funkcí. Na těchto funkcích umíme přirozeným způsobem zadat strukturu Booleovy algebry (využíváme, že hodnoty, tj. $\mathbb{Z}_2$ jsou Booleovou algebrou).

Odpovíme nyní na výše uvedené otázky tak, že pro libovolný prvek obecné Booleovy algebry sestrojíme jeho tzv. normální tvar, tj. napíšeme jej pomocí dobře vybrané skupiny nejjednodušších prvků a operace $\vee$. Porovnáním normálních tvarů dvou prvků pak už snadno poznáme, zda jsou stejné či nikoliv. Nejprve si tedy vybereme obzvlášť jednoduché prvky Booleovských algeber:

ATOMY V BOOLEOVSKÉ ALGEBŘE

Prvek $A \in K$ nazveme *atom* v Booleově algebře K, jestliže pro všechny $B \in K$ platí $A \wedge B = A$ nebo $A \wedge B = 0$.

Jinak řečeno, A je atom, když pro všechny ostatní prvky $B \leq A$ implikuje $B = 0$ nebo $B = A$.

Velmi jednoduché je to v Booleovské algebře všech podmnožin dané konečné množiny M. Zjevně budou atomy právě všechny jednoprvkové podmnožiny $A = \{x\}$ v množině M. Skutečně, pro každou podmnožinu B budeme mít buď $A \wedge B = A$, pokud $x \in B$, nebo $A \wedge B = 0$, pokud $x \notin B$.

Podívejme se ještě, jak vypadají atomy v Booleově algebře funkcí přepínačového systému s n přepínači $A_1, \ldots, A_n$. Snadno ověříme, že zde je 2^n atomů, které jsou tvaru $A_1^{\sigma_1} \wedge \cdots \wedge A_n^{\sigma_n}$, kde buď $A_i^{\sigma_i} = A_i$ nebo $A_i^{\sigma_i} = A_i'$.

Skutečně, pro dvě funkce φ a ψ je jejich infimem funkce $\varphi \wedge \psi$, jejíž hodnoty jsou dány součinem jejich hodnot v $\mathbb{Z}_2$. Platí tedy $\varphi \leq \psi$ jestliže φ má hodnotu $1 \in \mathbb{Z}_2$ všude tam, kde má ψ hodnotu $1 \in \mathbb{Z}_2$. Odtud už plyne, že v naší Booleově algebře hodnotových funkcí je funkce φ atomem

právě, když z 2^n hodnot φ na jednotlivých možnostech hodnot jednotlivých argumentů má právě jednou hodnotu $1 \in \mathbb{Z}_2$. Všechny takové funkce ovšem lze vytvořit právě uvedeným způsobem.

A nyní můžeme sformulovat slibovanou větu o *normálním disjunktivním tvaru*

Věta. *Každý prvek B v konečné Booleově algebře $(K, \wedge, \vee, ')$ lze zapsat jako supremum atomů*

$$B = A_1 \vee \cdots \vee A_k.$$

Tato formule je navíc jednoznačná až na pořadí atomů.

Důkaz nám zabere několik odstavců, ale základní idea je docela jednoduchá:

Uvažme všechny atomy $A_1, A_2, \ldots, A_k$ v K, které jsou menší nebo rovny B. Z vlastností uspořádání na množině K (viz 11.51(3)) je okamžitě vidět, že také

$$Y = A_1 \vee \cdots \vee A_k \leq B.$$

Hlavním naším krokem v důkazu bude ověřit, že $B \wedge Y' = 0$, což podle 11.51(4) zaručuje $B \leq Y$. Tím bude dokázána rovnost $B = Y$.

11.56. Tři pomocná tvrzení. Postupně si odvodíme několik technických vlastností atomů a pak teprve dokončíme důkaz věty o normálním disjunktivním tvaru. Pokračujeme v symbolice používané v minulém odstavci.

Tvrzení. *(1) Jestliže jsou $Y, X_1, \ldots, X_\ell$ atomy v K, pak $Y \leq X_1 \vee \cdots \vee X_\ell$ tehdy a jen tehdy, když $Y = X_i$ pro nějaké $i = 1, \ldots, \ell$.*
(2) Pro každý prvek $Y \neq 0$ v K existuje atom X, pro který je $X \leq Y$.
(3) Jestliže jsou $X_1, \ldots, X_r$ všechny atomy v K, pak $Y = 0$ právě, když $Y \wedge X_i = 0$ pro všechny $i = 1, \ldots, r$.

Důkaz. (1) Jestliže platí nerovnost v tvrzení, pak

$$Y \wedge (X_1 \vee \cdots \vee X_\ell) = Y.$$

Díky distributivitě můžeme rovnost přepsat jako

$$(Y \wedge X_1) \vee \ldots (Y \wedge X_\ell) = Y,$$

přitom ale je pro všechna i buď $Y \wedge X_i = 0$ nebo $Y \wedge X_i = X_i$. Pokud by tedy byly všechny tyto průniku 0, bylo by i $Y = 0$. Musí být tedy nějaké i, pro které je $Y \wedge X_i = X_i$. Prvek Y je přitom také atom, takže jsme dokázali požadovanou rovnost $Y = X_i$.

Opačná implikace je zřejmá.

(2) Pokud je Y samo atomem, pak stačí zvolit $X = Y$. Jestliže Y není atom, pak z definice vyplývá, že musí existovat nenulový prvek Z_1, pro který je $Z_1 \leq Y$. Jestliže ani Z_1 není atom, pak ze stejných důvodů najdeme $Z_2 \leq Z_1$ a postupně tak sestrojíme posloupnost různých prvků

$$\ldots Z_k \leq Z_{k-1} \leq \cdots \leq Z_1 \leq Y,$$

která nemůže být nekonečná, protože celá Booleovská algebra K je konečná. Proto musí skončit nějakým atomem Z_k.

(3) Předpokládejme nejprve, že $Y \wedge X_i = 0$ pro všechny indexy i. Pokud by ale bylo $Y \neq 0$, pak podle předchozího bodu musí existovat atom X_j, pro který $X_j \wedge Y = X_j$, což je spor.

Opačná implikace je triviální. □

11.57. Důkaz věty o normálním tvaru. Pokračujme v naší úvaze o přepsání prvku B pomocí výrazu

$$Y = A_1 \vee \cdots \vee A_k \leq B,$$

kde A_i jsou všechny atomy v K menší nebo rovny B. Spočteme

$$B \wedge Y' = B \wedge (A_1 \vee \cdots \vee A_k)' = B \wedge A_1' \wedge \cdots \wedge A_k'.$$

Jestliže je $A = A_i$ atom obsažený ve sjednocení Y, pak tedy $B \wedge Y' \wedge A = 0$. Pokud ale je A atom, který ve výrazu Y nevystupuje, dostáváme také $B \wedge Y' \wedge A = 0$, neboť Y obsahuje právě všechny atomy menší než B a proto $B \wedge A = 0$.

Dokázali jsme tedy, že $B \wedge Y'$ má nulový průnik se všemi atomy a proto je to nulový prvek podle našeho druhého pomocného tvrzení výše. Proto tedy $B \leq Y$. Z definice Y ale víme $Y \leq B$ a antisymetrie uspořádání tedy zaručuje $B = Y$, jak jsme chtěli dokázat.

Zbývá jednoznačnost výrazu, až na pořadí. Předpokládejme tedy, že jsme zapsali B dvěma způsoby v požadovaném tvaru

$$B = A_1 \wedge \ldots A_k = \tilde{A}_1 \wedge \ldots \tilde{A}_\ell.$$

Nyní každé A_i splňuje $A_i \leq B$ a proto je podle prvního pomocného tvrzení výše nutně rovno jednomu z $\tilde{A}_j$. Opakováním tohoto argumentu dostáváme požadovanou jednoznačnost a důkaz je ukončen.

11.58. Klasifikace. Na závěr našich úvah ještě dokážeme, že ve skutečnosti byly všechny naše příklady konečných Booleovských algeber izomorfní. Zejména tedy uvidíme, že každou z 2^{2^n} hodnotových funkcí pro n elementárních výroků umíme zapsat vhodným výrokem, stejně jako každou z 2^{2^n} různých přepínacích funkcí umíme zadat pomocí vhodně sestavených n přepínačů. Ve obou případech se bude diskutovaná algebra chovat stejně jako Booleovská algebra všech podmnožin v množině s 2^n prvky.

Navíc jsme se naučili každý takový výraz napsat v jednoznačném normalizovaném tvaru, takže umíme algoritmicky určit, zda budou např. dva přepínačové systémy vykazovat stejné chování, aniž bychom porovnali hodnoty při všech 2^n možných vstupech.

10.27c **Věta.** *Každá konečná Booleova algebra je izomorfní Booleovské algebře $K = 2^M$, kde M je množina atomů v K.*

Důkaz. Myšlenka důkazu je zcela přímočará. Při každém izomorfismu konečné Booleovské algebry $(K, \wedge, \vee, ')$ musí atomy být zobrazeny na atomy. Najděme si tedy množinu

M všech atomů v K a uvažme Booleovu algebru $(2^M, \cap, \cup, ')$ všech podmnožin v M. Tím máme i zadánu přirozenou korespondenci mezi atomy v K a 2^M.

Použijeme nyní disjunktivní normální formu k rozšíření zobrazení na celé K. Každý prvek $X \in K$ lze psát jednoznačně, až na pořadí atomů A_i, ve tvaru

$$X = A_1 \vee \cdots \vee A_k$$

a definujeme tedy funkci $f : K \to 2^M$ vztahem

$$f(X) = f(A_1) \cup \cdots \cup f(A_k),$$

tj. jako sjednocení jednoprvkových podmnožin $A_i \subset M$ obsažených ve výrazu.

Z jednoznačnosti normální formy vyplývá, že f je nutně bijekcí. Zbývá dokázat, že jde o morfismus Booleovských algeber.

Jsou-li X a Y dva prvky v K, pak v normální formě jejich sjednocení jsou právě atomy, které vystupují v X nebo v Y, zatímco u průniku jsou to atomy vystupující v obou výrazech současně. To ale právě ověřuje, že f zachovává operace $\wedge$ a $\vee$. Pro doplňky si všimněme, že atom A vystupuje v normální formě X', právě když $X \wedge A = 0$. Odtud již vidíme, že i komplementy f zachovává a důkaz je ukončen. □

Pro nekonečné Booleovské algebry obecně neplatí, že by byly izomorfní Booleovské algebře všech podmnožin nějké vhodné množiny M. Platí však, že je izomorfní Booleově podalgebře vhodné podmnožiny všech množin nějaké množiny M. Tomuto výsledku se říká *Stoneova věta o reprezentaci*, důkaz lze najít např. v **??**.

5. Kódování

Často potřebujeme přenášet informace a přitom zajišťovat jejich správnost. Někdy stačí zajistit, abychom poznali, zda je informace nezměněná, a při chybě si vyžádáme informaci znovu, jindy potřebujeme zajistit, aby chyby byly i opraveny bez nového přenášení zprávy. To vše je úkol kódování a v dalších odstavcích se tomuto úkolu budeme věnovat.

Pokud navíc chceme, aby zprávu mohl číst pouze adresát, potřebujeme i tzv. šifrování. Tomu jsme se krátce věnovali na konci minulé kapitoly.

10.29

11.59. Kódy. Při přenosu informace zpravidla dochází k její deformaci. Budeme pro jednoduchost pracovat s modelem, kdy jednotlivé částečky informace jsou buď nuly nebo jedničky (tj. prvky v $\mathbb{Z}_2$), říkáme jim *bity*, a přenášíme konečná slova o k bitech, pro nějaké pevně zvolené $k \in \mathbb{N}$. Obdobné postupy jsou možné nad libovolnými konečnými poli, my ale zůstaneme u nejjednoduššího případu $\mathbb{Z}_2$.

Přenosové chyby chceme rozpoznávat, případně i opravovat, a za tím účelem přidáváme ke k–bitovému slovu dodatečných $n - k$ bitů informace pro pevně zvolené $n > k$. Hovoříme o (n, k)–*kódech*.

Všech slov o k bitech je 2^k a každé z nich má jednoznačně určovat jedno *kódové slovo* z 2^n možných. Máme tedy u (n, k)-kódů ještě

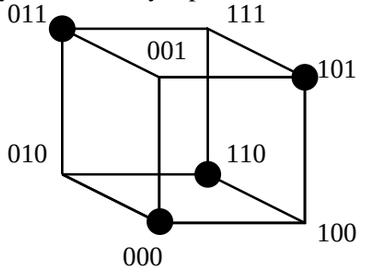
$$2^n - 2^k = 2^k(2^{n-k} - 1)$$

slov, které jsou chybové. Lze tedy tušit, že pro veliké k nám i malý počet přidaných bitů dává hodně redundantní informace.

Úplně jednoduchým příkladem je *kód kontrolující paritu*. Kódové slovo o $k + 1$ bitech je určené tak, aby přidáním prvního bitu ke k–bitovému slovu byl zaručen sudý počet jedniček ve slově. Jde tedy o $(k + 1, k)$–kód.

Pokud při přenosu dojde k lichému počtu chyb, s použitím tohoto jednoduchého kódu na to přijdeme. Dvě různá kódová slova se při tomto kódu vždy liší alespoň ve dvou pozicích, chybové slovo se ale od alespoň dvou různých kódových slov liší pouze v pozici jedné. Nemůžeme proto umět chyby opravovat, ani kdybychom věděli, že při přenosu došlo k právě jedné chybě.

Přehledně jsou všechna možná slova o dvou bitech s jedním přidaným paritním bitem vidět na obrázku níže. Kódová slova jsou zvýrazněna tučným puntíkem.



Navíc kódem kontrolujícím pouze paritu neumíme detekovat tak obvyklé chyby, jako je záměna dvou sousedních hodnot ve slově.

10.30

11.60. Vzdálenost slov. Na obrázku ilustrujícím $(3, 2)$–kód kontrolující paritu je vidět, že ve skutečnosti každé chybné slovo je „stejně“ daleko od tří kódových slov — jsou to ta, která se liší v právě jednom bitu. Ostatní jsou dál. Abstraktně můžeme takové pozorování zachytit definicí vzdálenosti:

VZDÁLENOST SLOV

Hammingova vzdálenost dvou slov je rovna počtu bitů, ve kterých se liší.

Pokud uvažujeme slova x, y, z a první dvě se liší v r bitech, zatímco y a z se liší v s bitech, pak se nutně x a z liší v nejvýše $r + s$ bitech, je tedy splněna trojúhelníková nerovnost pro vzdálenosti.

Aby kód mohl odhalovat chyby v r bitech, musí být minimální vzdálenost mezi kódovými slovy alespoň $r + 1$. Pokud budeme chtít i opravit nepřesně přenesené slovo s r chybami, pak nutně musí existovat jen jediné kódové slovo, které má od přijatého chybného slova vzdálenost nejvýše r. Ověřili jsme tedy jednoduchá tvrzení:

Věta. *(1) Kód spolehlivě odhaluje nejvýše r chyb ve slově, právě když je minimální Hammingova vzdálenost kódových slov $r+1$.*

(2) Kód spolehlivě odhaluje i opravuje nejvýše r chyb, právě když je minimální Hammingova vzdálenost kódových slov $2r+1$.

10.31

11.61. Konstrukce polynomiálních kódů. K praktickému použití potřebujeme efektivně konstruovat kódová slova tak, abychom je mezi všemi slovy sladno rozpoznali. Kontrolu parity jsme už viděli, další triviální možnost je prosté opakování bitů. Např. (3, 1)–kód bere jednotlivé bity a posílá je třikrát po sobě.

Docela systematickou cestou ke konstrukci kódů je využití dělitelnosti polynomů. Zpráva $b_0b_1\dots b_{k-1}$ je reprezentována jako polynom nad polem $\mathbb{Z}_2$

$$m(x) = b_0 + b_1x + \cdots + b_{k-1}x^{k-1}.$$

Polynomiální kód

Nechť $p(x) = a_0 + \cdots + a_{n-k}x^{n-k} \in \mathbb{Z}_2[x]$ je polynom s koeficienty $a_0 = 1$, $a_{n-k} = 1$. *Polynomiální kód generovaný polynomem* $p(x)$ je (n,k)–kód jehož slova jsou polynomy stupně menšího než n dělitelné $p(x)$.

Zpráva $m(x)$ je zakódována jako

$$v(x) = r(x) + x^{n-k}m(x),$$

kde $r(x)$ je zbytek po dělení polynomu $x^{n-k}m(x)$ polynomem $p(x)$.

Z definice kódového slova $v(x)$ pro přenášené slovo $m(x)$ čteme:

$$\begin{aligned} v(x) &= r(x) + x^{n-k}m(x) \\ &= r(r) + q(x)p(x) + r(x) = q(x)p(x), \end{aligned}$$

protože nad $\mathbb{Z}_2$ je součet dvou stejných polynomů vždy nulový. Budou tedy skutečně všechna kódová slova dělitelná $p(x)$.

Naopak, je-li $v(x)$ dělitelné $p(x)$, můžeme číst poslední výpočet z opačné strany a vidíme, že jde skutečně o kódové slovo vzniklé uvedeným postupem.

Z definice je také vidět, že kódové slovo vznikne přidáním $n-k$ bitů na začátek slova. Původní zpráva je tedy obsažena přímo v polynomu $v(x)$, takže dekódování správného slova je velmi snadné.

Uveďme si dva jednoduché příklady, které už známe. Všimněme si nejprve, že $1+x$ dělí polynom $v(x)$ tehdy a jen tehdy, když $v(1) = 0$. To nastane právě tehdy, když je ve $v(x)$ sudý počet nenulových koeficientů. Polynom $p(x) = 1+x$ proto generuje $(n, n-1)$–kód kontroly parity pro všechna $n \geq 3$.

Obdobně se snadno ověří, že polynom

$$p(x) = 1 + x + \cdots + x^{n-1}$$

generuje $(n, 1)$–kód n–násobného opakování bitů. Skutečně, dělením polynomu $b_0 x^{n-1}$ polynomem p dostaneme zbytek $b_0(1 + \cdots + x^{n-2})$ a tedy příslušné kódové slovo je $b_0 p(x)$.

10.32

11.62. Detekce chyb. Označme si $e(x)$ *vektor chyb*, které

vzniknou při přenosu. Místo posílaného slova $v \in (\mathbb{Z}_2)^n$ tedy příjmeme polynom

$$u(x) = v(x) + e(x).$$

Chyba je rozpoznatelná pouze tehdy, když generátor kódu $p(x)$ nedělí $e(x)$. Máme proto zájem o polynomy $p(x)$ v $\mathbb{Z}_2[x]$, které nevystupují jako dělitelé zbytečně často.

Definice. Ireducibilní polynom $p(x) \in \mathbb{Z}_2[x]$ stupně m se nazývá *primitivní*, jestliže $p(x)$ dělí polynom $(1 + x^k)$ pro $k = 2^m - 1$, ale nedělí jej pro žádná menší k.

Věta. *Je-li $p(x)$ primitivní polynom stupně m, pak pro všechna $n \leq 2^m - 1$ odhaluje příslušný $(n, n - m)$–kód všechny jednoduché a dvojité chyby.*

Důkaz. Jestliže nastane právě jedna chyba, pak $e(x) = x^i$ pro vhodné $0 \leq i < n$. Protože je $p(x)$ ireducibilní polynom, nemůže mít kořen v $\mathbb{Z}_2$. Zejména tedy nemůže dělit beze zbytku x^i, protože rozklad x^i je jednoznačný. Tedy je každá jednotlivá chyba rozpoznatelná.

Jestliže nastanou chyby právě dvě, pak

$$e(x) = x^i + x^j = x^i(1 + x^{j-i})$$

pro jistá $0 \leq i < j < n$. Již víme, že $p(x)$ nedělí beze zbytku žádné x^i a protože je primitivní, nedělí beze zbytku ani $1 + x^{j-i}$, pokud je $j - i < 2^m - 1$. Zároveň je $p(x)$ ireducibilní, nedělí proto ani součin $e(x) = x^i(1 + x^{j-i})$ a důkaz je ukončen. □

10.32a

11.63. Dusledek. *Je-li $q(x)$ primitivní polynom stupně m, pak pro všechna $n \leq 2^m - 1$ rozpoznává $(n, n - m - 1)$–kód generovaný polynomem $p(x) = q(x)(1 + x)$ všechny dvojité chyby a všechna slova s lichým počtem chyb.*

Důkaz. Kódová slova generovaná zvoleným polynomem $p(x)$ jsou dělitelná jak $x + 1$, tak primitivním polynomem $p_1(x)$. Jak jsme již ověřili, faktor $x + 1$ má za důsledek kontrolu parity, tj. všechna kódová mají sudý počet nenulových komponent. Tím umíme odhalit výskyt lichého počtu chyb. Jak jsem již také viděli v předchozí větě, druhý faktor umí odhalit dvojnásobné chyby. □

Následující tabulka ilustruje sílu výsledků předchozích dvou tvrzení pro několik primitivních polynomů v nízkých stupních. Např. poslední řádek nám říká, že přidáním pouhých 11 kontrolních bitů ke slovu o délce 1012 bitů budeme umět pomocí polynomu $(x + 1)p(x)$ odhalit jednotlivé, dvojité, trojité a všechny liché počty výskytů chyb v přenosu. Jde přitom o přenášení dosti velkých čísel, v desítkové soustavě by měly přes třista cifer.

primitivní polynom $p(x)$	kontrolní bity	délka slova
$1+x$	1	1
$1+x+x^2$	2	3
$1+x+x^3$	3	7
$1+x+x^4$	4	15
$1+x^2+x^5$	5	31
$1+x+x^6$	6	63
$1+x^3+x^7$	7	127
$1+x^2+x^3+x^4+x^8$	8	255
$1+x^4+x^9$	9	511
$1+x^3+x^{10}$	10	1023

Nástroje pro konstrukci primitivních polynomů dává teorie konečných polí. Souvisí s tzv. primitivními prvky v Galoisových polích $G(2^m)$. Ze stejné teorie lze také dovodit příjemnou realizaci dělení se zbytkem, tj. ověřování, zda je přijaté slovo kódové, pomocí zpožďovacích registrů. Jde o jednoduchý obvod s tolika prvky, kolik je stupeň polynomu.[8]

10.33

11.64. Lineární kódy. Polynomiální kódy lze efektivně popisovat také pomocí elementárního maticového počtu. Budeme přitom pracovat s vektorovými prostory nad $\mathbb{Z}_2$, takže musíme být opatrní při využívání výsledků elementární lineární algebry, protože jsme v ní často využívali vlastnost že $v = -v$ zaručuje $v = 0$. To nyní samozřejmě neplatí.

Základní definice vektorových prostorů, existence bazí a popis lineárních zobrazení pomocí matic ale zůstavají v platnosti. Bude užitečné připomenout si při čtení následujících odstavců obecnou teorii a ujistit se o její použitelnosti.

Vyjdeme z obecnější definice kódů, která požaduje lineární závislost kódového slova na původní informaci:

LINEÁRNÍ KÓDY

Injektivní lineární zobrazení $g : (\mathbb{Z}_2)^k \to (\mathbb{Z}_2)^n$ je *lineární kód*. Matice G typu n/k reprezentující toto zobrazení ve standardních bazích se nazývá generující *matice kódu*.

Pro každé slovo u je příslušným kódovým slovem delší vektor

$$v = G \cdot u.$$

Věta. *Každý polynomiální (n, k)–kód je lineární kód.*

DŮKAZ. Použijeme elementární vlastnosti dělení polynomů se zbytkem. Použijme naše přiřazení polynomu $v(x) = r(x) + x^{n-k}m(x)$ původní polynomiální zprávě $m(x)$ na součet dvou různých zpráv $m(x) = m_1(x) + m_2(x)$. Zbytek po dělení $x^{n-k}(m_1(x) + m_2(x))$ je díky jednoznačnosti dělení dán jako součet zbytků $r_1(x) + r_2(x)$ pro jednotlivé zprávy. Dostaneme tedy

$$v(x) = r_1(x) + r_2(x) + x^{n-k}(m_1(x) + m_2(x)),$$

[8] Více o této krásné teorii a jejích souvislostech s kódy se lze dočíst např. knize Gilbert, W., Nicholson, K., Modern Algebra and its applications, John Wiley & Sons, 2nd edition, 2003, 330+xvii pp., ISBN 0-471-41451-4.

což je požadovaná aditivita.

Protože jediným nenulovým skalárem je v $\mathbb{Z}_2$ jednička, dokázali jsme požadovanou lineritu zobrazení slova $m(x)$ na delší slovo $v(x)$.

Toto zobrazení je navíc injektivní, protože původní slovo $m(x)$ je prostě zkopírováno za přidané bity. □

Např. uvažujme polynomiální $(7,4)$–kód využívající primitivního polynomu $p(x) = 1 + x + x^3$ pro kódování slov se čtyřmi bity (se třemi kontrolními bity). Vyčíslením na jednotlivých bázových prvcích $m_i(x) = x^{i-1}$, $i = 1,2,3,4$ dostáváme

$$\begin{aligned} v_1 &= (1+x) + x^3 \\ v_2 &= (x+x^2) + x^4 \\ v_3 &= (1+x+x^2) + x^5 \\ v_4 &= (1+x^2) + x^6 \end{aligned}$$

a tedy matice odpovídající tomuto $(7,4)$–kódu je

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Tato matice ilustruje obecné vlasnosti polynomiálních kódů. Protože je u nich vždy původní slovo zkopírováno za přidané kontrolní bity, musí mít každý lineární kód vzniklý z polynomiálního matici s jednotkovým blokem $\mathbb{I}_k$ řádu k zabírajícím posledních k řádků matice, doplněným maticí P s $n-k$ řádky a k sloupci.

10.34 **11.65. Věta.** *Je-li $g : (\mathbb{Z}_2)^k \to (\mathbb{Z}_2)^n$ lineární kód s (blokově zapsanou) maticí*

$$G = \begin{pmatrix} P \\ \mathbb{I}_k \end{pmatrix},$$

potom zobrazení $h : (\mathbb{Z}_2)^n \to (\mathbb{Z}_2)^{n-k}$ s maticí

$$H = \begin{pmatrix} \mathbb{I}_{n-k} & P \end{pmatrix}$$

má následující vlastnosti

(1) $\operatorname{Ker} h = \operatorname{Im} g$

(2) Přijaté slovo $v = G \cdot u$ je kódové slovo právě, když je $H \cdot v = 0$.

Důkaz. Složení $h \circ f : (\mathbb{Z}_2)^k \to (\mathbb{Z}_2)^{n-k}$ je dáno součinem matic (počítáme nad $\mathbb{Z}_2$)

$$H \cdot G = \begin{pmatrix} \mathbb{I}_{n-k} & P \end{pmatrix} \cdot \begin{pmatrix} P \\ \mathbb{I}_k \end{pmatrix} = P + P = 0.$$

Dokázali jsme tedy $\operatorname{Im} g \subset \operatorname{Ker} h$. Protože je prvních $n-k$ sloupců v H tvořeno bázovými vektory v $(\mathbb{Z}_2)^{n-k}$, má obraz $\operatorname{Im} h$ maximální dimenzi $n-k$ a tedy má tento obraz 2^{n-k} různých vektorů. Vektorové prostory nad $\mathbb{Z}_2$ jsou konečné

komutativní grupy, proto můžeme použít vztah mezi mohutnostmi pogrup a faktorgrup z odstavce 11.10 a dostáváme

$$|\operatorname{Ker} h| \cdot |\operatorname{Im} h| = |(\mathbb{Z}_2)^n| = 2^n.$$

Proto je počet vektorů v $\operatorname{Ker} h$ roven $2^n \cdot 2^{k-n} = 2^k$. K dokončení důkazu prvního tvrzení si nyní stačí povšimnout, že obraz $\operatorname{Im} f$ má také 2^k prvků.

Druhé tvrzení je samozřejmým důsledkem prvního tvrzení. □

Matici H z věty se říká *matice kontroly parity* přílušného (n, k)–kódu.

Např. matice $H = (1\ 1\ 1)$ je zjevně takovou maticí pro $(3, 2)$ kód přidávající jeden paritní bit k slovu o dvou bitech. Skutečně ji snadno dostaneme z matice

$$G = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

zadávající tento kód.

Pro výše uvedený $(7, 4)$–kód to bude matice

$$H = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

10.35

11.66. Samoopravné kódy. Jak jsme viděli, přenos zprávy u dává výsledek

$$v = u + e.$$

To je ale nad $\mathbb{Z}_2$ ekvivalentní $e = u + v$.

Pokud tedy známe vektorový podprostor $V \subset (\mathbb{Z}_2)^n$ správných kódových slov, víme u každého výsledku, že správné slovo se něj liší o případnou chybu, která je ve třídě rozkladu $v + V$ ve faktorovém vektorovém prostoru $(\mathbb{Z}_2)^n / V$.

Zobrazení $h : (\mathbb{Z}_2)^n \to (\mathbb{Z}_2)^{n-k}$ zadané maticí kontroly parity má V za jádro, proto indukuje injektivní lineární zobrazení $h : (\mathbb{Z}_2)^n / V \to (\mathbb{Z}_2)^{n-k}$. Jeho hodnoty jsou jednoznačně určeny hodnotami $H \cdot u$.

SYNDROMY SLOV

Hodnota $H \cdot u$, kde H je matice kntroly parity pro lineární kód, se nazývá *syndrom* slova u v tomto kódu.

Samozřejmým důsledkem této konstrukce je následující tvrzení.

Věta. *Dvě slova jsou ve stejné třídě rozkladu $v + V$ právě, když sdílí syndrom.*

Samoopravné kódy nyní můžeme konstruovat tak, že pro každý syndrom určíme prvek v příslušné třídě, který je nejvhodnějším výběrem pro chybu. Budeme přitom vycházet z představy, že nejpravděpodobněji nastal nejmenší možný počet chyb.

Ukážeme si postup na jednoduchém polynomiálním $(6, 3)$ kódu zadaném polynomem $1 + x + x^3$. Příslušné matice G a H obdržíme z těch z odstavce 11.65 tak, že prostě

odebereme poslední sloupec a řádek u G a poslední sloupec u H.

Sestavíme si nyní tabulku všech syndromů a jim odpovídajících kódových slov.

Syndrom 000 mají právě všechna kódová slova. Všechna možná slova s daným syndromem pak dostaneme přičtením syndromu (doplněného nulami na délku kódového slova) ke všem kódovým slovům.

V následujících dvou tabulkách jsou v prvním řádku příslušné syndromy, na dalším řádku pak uvádíme ten z vektorů v příslušné třídě, který má nejméně jedniček. Skoro ve všech případech jde o jedinou jedničku, jen v posledním řádku máme jedničky dvě a zvolili jsme si jako význačný prvek ten, který má jedničky vedle sebe (třeba protože věříme, že násobné chyby s větší pravděpodobností nastávají těsně po sobě)

000	100	010	001
000000	**100000**	**010000**	**001000**
110100	010100	100100	111100
011010	111010	001010	010010
111001	011001	101001	110001
101110	001110	111110	100110
001101	101101	011101	000101
100011	000011	110011	101011
010111	110111	000111	011111

110	011	111	101
000100	**000010**	**000001**	**000110**
110000	110110	110101	110010
011110	011000	011011	011100
111101	111011	111000	111111
101010	101100	101111	101000
001001	001111	001100	001011
100111	100001	100010	100101
010011	010101	010110	010001

Počínaje druhým sloupcem první tabulky, je každý sloupec afinním podprostorem v $(\mathbb{Z}_2)^6$, jehož zaměřením je vektorový prostor daný prvním sloupcem první tabulky. Je tomu tak, protože je daný kód lieární, všechna kódová slova tedy tvoří vektorový prostor a jednotlivé třídy ve faktorovém prostoru jsou afinní podprostory.

Zejména je tedy rozdíl každých dvou slov ve stejném sloupci nějakým kódovým slovem. Tučně vyznačená slova představují tzv. vedoucí representanty třídy (afinního prostoru) odpovídajícího danému syndromu. Jsou to slova s nejmenším počtem jedniček v řádku. Udávají tak nejmenší počet bitových změn, které musíme v libovolném slovu v sloupci provést, abychom dostali kódové slovo.

Např., pokud dostaneme kódové slovo $v = 111101$, má syndrom $H \cdot v = 110$. Vedoucím representantem ve třídě tohoto syndromu je slovo 000100 a jeho odečtením od obdrženého kódového slova (což je nad $\mathbb{Z}_2$ totéž jako přičtení) dostaneme platné kódové slovo 111001. Je to platné kódové slovo s nejmenší možnou Hammingovou vzdáleností od obdrženého slova. Odeslaná zpráva tedy patrně byla 001.