Kapitola 1

Algebraický úvod

1.1 Pologrupa, monoid, neutrální prvek

Binární operací na množině M se rozumí každé zobrazení $M \times M \to M$. Binární operace se často zapisují jako násobení či sčítání, a to i když operace nijak nesouvisí s číselnými strukturami. Binární operace $\cdot$ na M se nazývá *asociativní*, pokud

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z \text{ pro všechna } x, y, z \in M.$$

Množině s asociativní operací se říká *pologrupa*. Prvek e pologrupy S nazveme *neutrální*, splňuje-li

$$xe = x = ex \text{ pro všechna } x \in M.$$

Lemma. *V pologrupě existuje nanejvýš jeden neutrální prvek.*

Důkaz. Ať e a f jsou neutrální prvky. Pak $e = ef = f$. □

Pologrupě s neutrálním prvkem se říká *monoid*. Při použití *multiplikativní notace* (operace násobení) se neutrálnímu prvku říká též *jednotkový*. V *aditivní notaci* (operace sčítání) hovoříme o *nulovém* prvku.

1.2 Grupa, inverzní prvek, krácení

Ať $M = M(\cdot, 1)$ je monoid. Prvek $x \in M$ může mít *levý inverzní* a *pravý inverzní* prvek. Levý je dán vztahem $yx = 1$ a pravý vztahem $xz = 1$.

Lemma. *Má-li prvek monoidu jak levý, tak pravý inverzní prvek, jsou tyto prvky shodné.*

Důkaz. Je-li $yx = 1 = xz$, je $y = y \cdot 1 = y(xz) = (yx)z = 1 \cdot z = z$. □

Hovoříme-li o *inverzním* prvku, míníme tím, že je inverzním jak zleva, tak zprava. Z lemmatu plyne, že inverzní prvek je určen jednoznačně. V multiplikativní notaci se značí x^{-1}, v aditivní $-x$.

Grupa je monoid, ve kterém ke každému prvku existuje prvek inverzní. Je-li $G = G(\cdot, ^{-1}, 1)$ grupa a pro její prvky platí $xy = xz$, máme $y = (x^{-1}x)y = x^{-1}(xy) = x^{-1}(xz) = (x^{-1}x)z = z$. Podobně z $yx = zx$ plyne $y = z$. V grupě lze tedy *krátit* zleva i zprava.

1.3 Podobjekty. Invertibilní prvky

Podpologrupou pologrupy S se rozumí každá podmnožina P taková, že z $x, y \in P$ plyne $xy \in P$ (říkáme, že P je *uzavřená* na násobení). Podpologrupa monoidu $M = M(\cdot, 1)$, která obsahuje 1, se nazývá *podmonoid*. *Podgrupou* je pak každý podmonoid, který je uzavřený na inverzní prvky.

Prvek x monoidu M se nazývá *invertibilní*, pokud v M existuje prvek vůči x inverzní.

Tvrzení. *Invertibilní prvky monoidu tvoří podgrupu. Jsou-li x a y invertibilní, platí $(xy)^{-1} = y^{-1}x^{-1}$.*

Důkaz. Stačí ověřit poslední vztah, neboť z něj plyne, že invertibilní prvky jsou uzavřené na násobení. Ovšem z $xx^{-1} = 1 = yy^{-1}$ máme $(xy)(y^{-1}x^{-1}) = x(yy^{-1})x^{-1} = xx^{-1} = 1$. Podobně $(y^{-1}x^{-1})(xy) = 1$, a zbytek je snadný. □

Mezi podobjekty bývá největší a nejmenší. S tím je spojeno určité názvosloví. Uvedeme ho pro grupy, jinde je obdobné. Největší podgrupa grupy G je ona sama. Nejmenší podgrupa je jednoprvková množina $\{1\}$. Ta se obvykle značí též 1. Podgrupy 1 a G se nazývají podgrupy *nevlastní*. Ostatní podgrupy jsou *vlastní*.

1.4 Homomorfismus a izomorfismus grup

Ať $G = G(\cdot, ^{-1}, 1)$ a $H = H(\cdot, ^{-1}, 1)$ jsou grupy. Zobrazení $\varphi : G \to H$ nazveme *homomorfismus*, pokud $\varphi(x \cdot y) = \varphi(x) \cdot \varphi(y)$ pro všechna $x, y \in G$.

V předchozí definici může zarazit, že operace v G i H se značí stejně. V obecných definicích bývá zvykem takto postupovat, i když o homomorfismu lze mluvit i v případě, že jsou obě operace značeny různě. Někdy se pro přehlednost může připojit G nebo H jako index, aby se zdůraznilo, o kterou grupu jde. Identitu homomorfismu lze tedy psát též jako $\varphi(x \cdot_G y) = \varphi(x) \cdot_H \varphi(y)$.

Lemma. *Ať $\varphi : G \to G$ je homomorfismus grup. Potom $\varphi(1_G) = 1_H$ a $\varphi(x^{-1}) = (\varphi(x))^{-1}$, pro každé $x \in G$.*

Důkaz. Máme $\varphi(1_G) \cdot 1_H = \varphi(1_G) = \varphi(1_G \cdot 1_G) = \varphi(1_G) \cdot \varphi(1_G)$, takže krácením $1_H = \varphi(1_G)$. Zbývá ukázat, že $\varphi(x^{-1})$ je vůči $\varphi(x)$ inverzní. Ovšem pro $\varphi(x^{-1})\varphi(x) = \varphi(x^{-1}x) = \varphi(1_G) = 1_H$. □

Pokud je φ bijektivní zobrazení, hovoříme o *izomorfismu*. Je-li φ izomorfismus, tak se místo $\varphi : G \to H$ často píše $\varphi : G \cong H$. Pro $a, b \in H$ v takovém případě platí $\varphi(\varphi^{-1}(a) \cdot \varphi^{-1}(b)) = \varphi\varphi^{-1}(a) \cdot \varphi\varphi^{-1}(b) = ab = \varphi(\varphi^{-1}(ab))$, odkud $\varphi^{-1}(a) \cdot \varphi^{-1}(b) = \varphi^{-1}(ab)$. Vidíme, že z $\varphi : G \cong H$ plyne $\varphi^{-1} : H \cong G$.

Jsou-li $\varphi : G \to H$ a $\psi : H \mapsto K$ homomorfismy grup, je $\psi \cdot \varphi : G \to K$ také homomorfismus. Vskutku, $\psi\varphi(xy) = \psi(\varphi(xy)) = \psi(\varphi(x) \cdot \varphi(y)) = (\psi\varphi(x)) \cdot (\psi\varphi(y))$, pro všechna $x, y \in G$.

Homomorfismus $\varphi : G \to G$ se nazývá *endomorfismus* a izomorfismu $\varphi : G \cong G$ se říká *automorfismus*. Protože endomorfismy lze skládat a protože skládání zobrazení je asociativní, je množina $\mathrm{End}\,(G)$ všech endomorfismů grupy G monoid s neutrálním prvkem id_G (identické zobrazení $x \mapsto x$ grupy G). Z tvrzení 1.3 plyne, že $\mathrm{Aut}\,(G)$ (množina všech automorfismů grupy G) tvoří podgrupu $\mathrm{End}\,(G)$.

1.5 Abelova grupa. Ekvivalence modulo podgrupa.

Grupa se nazývá *komutativní*, platí-li $xy = yx$. Komutativním grupám se často říká Abelovy (nebo abelovské). Budeme s nimi většinou pracovat v aditivní notaci (neutrální prvek je nula, inverzním prvkům se říká opačné).

Ať N je podgrupa Abelovy grupy G. Definujeme na G relaci $\equiv \mod N$ tak, že $x \equiv y \mod N$ právě když $x - y \in N$. (Čteme, x je *kongruentní* s y *modulo* N; zápis $x - y$ je zkrácením $x + (-y)$).

Lemma. *Relace* $\equiv \mod N$ *je ekvivalence na* G. *Blok této ekvivalence, který obsahuje prvek* x, *je roven množině* $x+N = \{x + a; a \in N\}$. *Zobrazení* $a \mapsto x+a$ *je bijekcí* N *a* $x + N$.

Důkaz. Z $x - y \in N$ plyne $-(x - y) = y - x \in N$, takže relace $\equiv \mod N$ je symetrická (a také zřejmě reflexivní). Z $x \equiv y \mod N$ a $y \equiv z \mod N$ máme $x \equiv z \mod N$, neboť $x - z = (x - y) + (y - z)$, takže jde skutečně o ekvivalenci. Zbytek je jasný, neboť $y = x - (x - y)$ pro všechna $x, y \in G$. □

Množinám $x + N$ se říká *rozkladové třídy* modulo N. Lemma tedy mimo jiné praví, že všechny rozkladové třídy modulo N jsou stejně mohutné (tj. mají stejný počet prvků) jako N.

Mohutnost grupy G se značí $|G|$ a říká se jí *řád* grupy. Počet (mohutnost) všech rozkladových tříd modulo N se nazývá *index* podgrupy N a značí se $|G : N|$. Ze stejných velikostí rozkladových tříd plyne

Lagrangeova věta $|G| = |N| \cdot |G : N|$ *pro každou podgrupu* N *(komutativní grupy)* G. □

Lagrangeova věta platí i pro nekomutativní grupy. Proto je v jejím znění slovo komutativní v závorkách. Zde se nekomutativními grupami zabývat nebudeme.

1.6 Faktorgrupa

Buď opět G Abelova grupa s podgrupou N. Uvažme $x, y \in G$ a $a, b \in N$. Pak $(x + a) + (y + b) = (x + y) + (a + b) \in x + y + N$, takže

$$(x + N) + (y + N) = (x + y) + N$$

se definuje operace na $G/N = \{x + N; x \in G\}$. Výsledek operace závisí na třídách $x + N$ a $y + N$, nikoliv na reprezentantech x a y těchto tříd. Je zřejmé, že $((x+N)+(y+N))+(z+N) = (x+y+z)+N = (x+N)+((y+N)+(z+N))$, takže definovaná operace je asociativní. Jistě je i komutativní, a z $(x + N) + ((-x) + N) = 0 + N = N$ plyne, že ke každému prvku lze najít prvek opačný, přičemž $0 + N = N$ je prvek neutrální. Vidíme, že G/N je Abelova grupa (říká se jí grupa *kvocientní* nebo *faktorgrupa* modulo N). Vidíme, že $|G/N| = |G : N|$.

1.7 Iterované sčítání a násobení. Exponent

Buď G Abelova grupa. Pro $n \geq 1$ a $x \in G$ definujeme nx jako součet $x + \cdots + x$, který má n sčítanců. (Formálně lze definovat $1 \cdot x = x$, $(n + 1)x = nx + x$.) Položíme také $0 \cdot x = 0$ a $(-n) \cdot x = -(nx)$. Z počtu sčítanců je patrné, že

$$(n + m)x = nx + mx \quad \text{a} \quad n(mx) = (nm)x$$

pro $n, m \geq 0$. Použitím vztahu $(-n)x = -(nx) = n(-x)$ lze snadno dokázat, že vzorce pro iterovaný součat a násobení platí pro všechna n, m celá. (Formální důkaz lze provést indukcí. Vzhledem k názornosti obou vzorců však od něj upustíme).

Nejmenší $m \geq 1$ takové, že $mx = 0$, se nazývá (pokud existuje) *řád* prvku $x \in G$. (Neexistuje-li, jde o prvek nekonečného řádu). Každému $m \geq 1$ takovému, že $mx = 0$ pro všechna $x \in G$, se říká *exponent* grupy G. Pokud je m minimální možné, hovoříme o *minimálním exponentu*.

V multiplikativní notaci se iterované násobení značí jako mocniny. Je tedy $x^n = x \cdot \cdots \cdot x$, kde součin má n činitelů, pro každé $n \geq 1$. Dále $x^0 = 1$ a $x^{-n} = (x^{-1})^n$. (Opět lze formálně položit $x^0 = 1$, $x^{n+1} = x^n x$.) Uvedené vztahy mají v multiplikativní notaci tvar

$$x^n x^m = x^{n+m} \quad \text{a} \quad (x^n)^m = x^{nm}.$$

1.8 Okruh

Algebraický systém R spolu s binárními operacemi $+$ a $\cdot$, unární operací $-$ a konstantami 0 a 1 se nazývá *okruh*, pokud

$R(+, -, 0)$ je Abelova grupa

$R(\cdot, 1)$ je monoid a

všechna $x, y, z \in R$ splňují $x \cdot (y+z) = (x \cdot y)+(x \cdot z)$ a $(y+z) \cdot x = (y \cdot x)+(z \cdot x)$.

Okruh je *komutativní*, je-li operace násobení komutativní. V komutativním okruhu stačí ověřit pouze jeden z obou distributivních zákonů.

Při práci s oběma operacemi se používají obvyklé precendence operátorů, takže například levý distributivní zákon píšeme jako $x(y + z) = xy + xz$.

Okruh s jediným prvkem nazýváme triviální.

Lemma. *Ať R je okruh. Pak $x \cdot 0 = 0 = 0 \cdot x$ a $x(-y) = -xy = (-x)y$, pro všechna $x, y \in R$. Pokud R není triviální, je $0 \neq 1$.*

Důkaz. Máme $x = x \cdot 1 = x \cdot (1 + 0) = x \cdot 1 + x \cdot 0$, takže $x \cdot 0 = 0$ dostaneme odečtením $x = x \cdot 1$. Z $0 = x(y + (-y)) = xy + x(-y)$ plyne $x(-y) = -xy$. Je-li $0 = 1$, platí $x = x \cdot 1 = x \cdot 0 = 0$ pro každé $x \in R$. □

1.9 Ideál a faktorokruh. Dělitelnost

Kvocientní struktura okruhu R musí vyjadřovat absorpční vlastnosti nulového prvku. Proto je přirozená definice *ideálu* jakožto každé podmnožiny $I \subseteq R$ takové, že $I(+, -, 0)$ je podgrupa $R(+, -, 0)$ a pro každé $a \in I$, $r \in R$, platí jak $ar \in I$, tak $ra \in I$.

Jsou-li $r, s \in R$ a $a, b \in I$, tak $(r+a)\cdot(s+b) = ab+t$, kde $t = rb+as+rs \in I$.

Proto je možné na $R/I = \{r + I; r \in R\}$ definovat jak operaci sčítání (viz oddíl 1.6), tak operaci násobení (a to tak, že $(r + I) \cdot (s + I) = rs + I$).

Při takové definici je R/I okruhem. Říkáme mu *kvocientní okruh* nebo *faktorokruh* modulo I.

Nevlastní ideály jsou 0 a R. Ideál I se nazývá *maximální*, pokud $I \subsetneq R$ a $I \subsetneq J \subsetneq R$ neplatí pro žádný ideál J okruhu R.

Dělitelností se budeme zabývat pouze v komutativních okruzích. Je-li R komutativní okruh, $a \in R$, je $aR = Ra = \{ra; r \in R\}$ zjevně ideál. Jde o *hlavní ideál* prvku a. Prvku a se též říká *generátor* daného hlavního ideálu.

Řekneme, že $a \in R$ dělí $b \in R$, pokud existuje $c \in R$, že $b = ac$. Zapisujeme jako $a|b$.

Lemma. *Buď R komutativní okruh. Ať $a, b \in R$. Pak a dělí $b \Leftrightarrow Ra \supseteq Rb$.*

Důkaz. Je-li $b = ca$, je každé rb rovno $(rc)a \in Ra$. Je-li $Rb \subseteq Ra$, je $a = rb$ pro nějaké $r \in R$. □

1.10 Invertibilní prvky. Obory

V okruhu se invertibilita vztahuje samozřejmě k operaci násobení. Množina všech invertibilních prvků okruhu R se značí R^*. Z tvrzení 1.3 plyne, že R^* je grupa.

Invertibilní prvek komutativního okruhu R lze zjevně charakterizovat také tak, že jeho hlavní ideál je roven R.

Dělitelem nuly rozumíme každý nenulový prvek a, který splňuje $ac = 0$ pro nějaké $c \neq 0$. Komutativní okruh bez dělitelů nuly, který je netriviální, se nazývá *obor integrity* (krátce též pouze *obor*).

V oboru integrity z $a_1c = a_2c$, $c \neq 0$, plyne $(a_1 - a_2)c = 0$, odkud $a_1 - a_2 = 0$, takže $a_1 = a_2$. Obory integrity patří tudíž mezi okruhy, ve kterých lze krátit nenulovými prvky.

Lemma. *Buď R obor integrity. Ať $a, b \in R$. Je ekvivalentní:*

(1) Současně a dělí b, b dělí a,

(2) $a = bx$ pro nějaké $x \in R^*$,

(3) $Ra = Rb$.

Důkaz. Je-li $x \in R^*$, tak $xy = yx = 1$ pro nějaké $y \in R^*$. Proto z lemmatu 1.9 plyne jak (2)⇔(3), tak (2)⇒(1). Předpokládejme, že $a = bx$, $b = ay$. Pak $a \cdot 1 = a = bx = a \cdot yx$, takže krácením dostaneme $1 = yx$, a tedy $x \in R^*$. □

Skutečnost, že R je obor integrity, jsme v předchozím důkaze potřebovali pouze v posledním kroku; naším cílem zde není největší možná obecnost. Teorii dělitelnosti nebudeme také rozvíjet pro obory integrity obecně, ale pouze pro *obory hlavních ideálů*, což jsou ty obory integrity, ve kterých je každý ideál hlavní.

1.11 Řetězce ideálů. Noetherovské okruhy

Okruh R se nazývá *noetherovský*, pokud v něm nelze nalézt ostře rostoucí posloupnost ideálů $I_1 \subsetneq I_2 \subsetneq I_3 \subsetneq \cdots$. Jinými slovy, v noetherovském okruhu mají všechny řetězce ideálů $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ jen konečně mnoho prvků (od jistého indexu se stabilizují).

Lemma. *Ať $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$ je řetězec ideálů okruhu R. Potom $I = \bigcup (I_j; j \geq 1)$ je také ideál okruhu R.*

Důkaz. Pro $a, b \in I$ existuje $j \geq 1$, že a i b padnou do I_j. Pak ale $a+b \in I_j \subseteq I$ a $ra \in I_j \subseteq I$ pro každé $r \in R$. □

Důsledek. *Každý obor hlavních ideálů je noetherovský.*

Důkaz. Uvažme posloupnost ideálů $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$. Potom $I = \bigcup (I_j; j \geq 1)$ je rovno nějakému aR, přičemž a padne do nějakého I_k. To ale znamená $I_k = I_j$ pro každé $k \leq j$. □

1.12 Součiny a podíly ideálů

Jsou-li I a J ideál okruhu R bude jejich průnik $I \cap J$ zjevně také ideálem. Definujeme též součet $I + J$, součin IJ a podíl $I : J$, a to tak, že

$$I + J = \{x + y; x \in I \text{ a } y \in J\},$$

$$IJ = \left\{\sum_{i=1}^{k} x_i y_i; x_i \in I, y_i \in J, \text{ kde } k \geq 1\right\} \text{ a}$$

$$I : J = \{x \in R; xJ \subseteq I\}$$

Ověřit, že jde ve všech třech případech o ideály, je snadné. Učiňme tak detailně pouze v posledním případě. Je-li $r \in R$ a $xJ \subseteq I$, tak $(rx)J = r(xJ) \subseteq rI \subseteq I$ a $(xr)J = x(rJ) \subseteq xJ \subseteq I$. Je-li ještě $yJ \subseteq I$, je $(x + y)J = xJ + yJ \subseteq J + J = J$.

Lemma. *Ať R je obor integrity s hlavními ideály $I = aR$ a $J = bR$. Pak $IJ = (ab)R$ a $I \subseteq I : J$. Je-li $I \subseteq J$, tak $a = bc$ pro nějaké $c \in R$, a $I : J = cR$. Přitom*

(1) $I : J = R \Leftrightarrow I = J$ a

(2) $I : J = I \Leftrightarrow J = R$.

Důkaz. Podle definice se IJ rovná všem možným hodnotám tvaru $\sum(r_i a)(s_i b)$, $1 \leq i \leq k$, které lze samozřejmě zapsat jako $(\sum r_i s_i)\,(ab)$. Proto $IJ = (ab)R$. Pro $a \in I$ je $aR \subseteq I$, takže i $aJ \subseteq I$. Odsud $I \subseteq I : J$.

Předpokládejme $I \subseteq J$. Pak $a = bc$ dle lemmatu 1.9. Přitom $(cR)(bR) = aR$ dle prvé části důkazu, takže $cR \subseteq I : J$. Je-li $x \in I : J$, máme $xb \in I$, tedy pro nějaké $r \in R$ je $xb = ra = rcb$. Odsud krácením $x = rc \in cR$.

Rovnost $I = J$ je podle lemmatu 1.9 ekvivalentní invertibilitě c, což nastává právě když $cR = R$.

Rovnost $I : J = I$ je podle téhož lemmatu ekvivalentní invertibilitě prvku b. □

Pro teorii dělitelnosti má zásadní význam následující vlastnost maximálních ideálů.

Tvrzení. *Buď M maximální ideál okruhu R. Je-li součin ideálů $I_1 \dots I_k$, $k \geq 1$, obsažen v M, je $I_j \subseteq M$ pro nějaké j, $1 \leq j \leq k$.*

Důkaz. Ideál $I_j + M$ obsahuje M, a proto je buď roven M, nebo R. Prvý případ implikuje $I_j \subseteq M$; předpokládejme, že tedy pro každé j platí druhý, $1 \leq j \leq k$.

Pak lze 1 vyjádřit jako $a_j + m_j$, kde $a_j \in I_j$, $m_j \in M$. Roznásobením součinu $1 = (a_1 + m_1) \dots (a_k + m_k)$ dostaneme výraz, který lze zapsat jako $a_1 \dots a_k + m$, kde $m \in M$. Protože předpokládáme $a_1 \dots a_k \in M$, musí M obsahovat prvek 1, a to je spor. □

Důsledek. *Ať $M_1 \dots M_k \subseteq M$, kde $M_1, \dots, M_k$ i Mjsou maximální ideály okruhu R. Pak $M_j = M$ pro nějaké j, $1 \leq j \leq k$.*

Důkaz. Podle tvrzení je $M_j \subseteq M$ pro nějaké j. Ovšem maximální ideál může být obsažen toliko ve shodném maximálním ideálu. □

1.13 Rozklady v oborech hlavních ideálů

Lemma. *Každý vlastní ideál oboru hlavních ideálů lze vyjádřit jako součin maximálních ideálů.*

Důkaz. Daný ideál označíme $I_0 = I$ a budeme indukcí konstruovat maximální ideály $M_1 \dots M_k$ a ideál I_k tak, aby $I = M_1 \dots M_k I_k$. Pokud I_k je maximální ideál, položíme $M_{k+1} = I_k$, a konstrukci ukončíme. Předpokládejme, že I_k je vlastní ideál, který je vlastní podmnožinou v nějakém maximálním ideálu M_{k+1} (jeho existence plyne například z toho, že R je podle důsledku 1.11 okruh noetherovský). Položíme $I_{k+1} = I_k : M_{k+1}$. Podle lemmatu 1.12 máme $I_k = M_{k+1} I_{k+1}$. Navíc z $I_k \neq M_{k+1}$ plyne, že I_{k+1} je vlastní ideál, a z $M_{k+1} \neq R$ plyne, že je $I_k \subsetneq I_{k+1}$. Kdyby bylo takto možné sestrojit nekonečnou posloupnost, dostali bychom takto spor s noetherovskostí daného okruhu. □

Tvrzení. *Každy vlastní ideál oboru hlavních ideálů lze až na pořadí jednoznačně vyjádřit jako součin maximálních ideálů*

Důkaz. Vzhledem k lemmatu stačí dokázat, že z $I = M_1 \dots M_k = N_1 \dots N_h$, kde M_i i N_j jsou maximální ideály, plyne že obě vyjádření jsou až na pořadí shodná. Budeme postupovat indukcí dle k a předpokládat $k \leq h$. Každý z ideálů M_i obsahuje ideál I, takže $M_1 = N_j$ pro nějaké j, $1 \leq j \leq h$, dle důsledku 1.12. Proto můžeme předpokládat $M_1 = N_1$. Je-li $h = 1$, není co řešit. Předpokládejme $h \geq 2$.

Z $M_1 = aR$ a $N_2 \dots N_j = bR$ plyne $I = (ab)R$ a $I : M_1 = N_2 \dots N_h$, dle lemmatu 1.12. Proto je $I : M_1$ vlastní ideál R. Odsud $I \neq M_1$ a $k \geq 2$. Tudíž $I : M_1 = M_2 \dots M_k = N_2 \dots N_h$ a lze použít indukční předpoklad. □

Nevlastní ideál R lze považovat za součin nulového počtu maximálních ideálů.

Důsledek. *Ať I a J jsou dva nenulové ideály oboru hlavních ideálů. Pak $I \subseteq J$ právě když existují maximální ideály $M_1, \dots, M_h$ takové, že $I = M_1 \dots M_h$ a $J = M_1 \dots M_k$, kde $1 \leq k \leq h$.*

Důkaz. Z lemmatu 1.12 plyne, že $I = J(I : J)$. Proto skutečně jde o přímý důsledek předchozího tvrzení. □

1.14 Dělitelnost prvků v oborech hlavních ideálů

Buď nejprve R obor integrity. Předpokládejme, že pR je maximální ideál v R. Pokud p dělí součin prvků $a_1 \dots a_k$, tak $pR \supseteq (a_1R) \dots (a_kR)$, takže podle tvrzení 1.12 je $pR \supseteq a_jR$ pro nějaké j, $1 \leq j \leq k$, a tedy $p|a_j$.

Obecně se prvek $p \neq 0$ oboru integrity R nazývá *prvočinitel*, pokud není invertibilní a pokud pro všechna $a_1, \dots, a_k \in R$ platí implikace $p|a_1 \dots a_k \Rightarrow p|a_i$ pro některé i, $1 \leq i \leq k$.

(Je snadné nahlédnout ,že p je prvočinitel, pokud je uvedená vlastnost splněna pro $k = 2$.)

Viděli jsme, že je-li pR maximální ideál, je p nutně prvočinitel.

Lemma. *Buď R obor hlavních ideálů. Prvek $p \neq 0$ je v R prvočinitel právě když pR je maximální ideál.*

Důkaz. Zbývá dokázat přímou implikaci. Ať je tedy p prvočinitel. Podle tvrzení 1.13 máme $pR = (a_1R) \dots (a_kR)$, kde a_iR jsou maximální ideály. Tudíž $p|a_1 \dots a_k$, takže $p|a_j$ pro nějaké j, $1 \leq j \leq k$. Odsud $pR \supseteq a_jR$ a tedy $pR = a_jR$ (p není invertibilní, takže $pR \neq R$). □

Tvrzení 1.13 tedy říká, že každé $aR \neq R$ lze vyjádřit jako $(p_1R) \dots (p_kR)$, kde p_j jsou prvočinitele. Toto vyjádření je až na pořadí jednoznačné, ovšem prvky $p_1, \dots, p_k$ jsou podle lemmatu 1.10 určeny jednoznačně svými ideály až na násobky invertibilními prvky. Podle lemmatu 1.12 je $(p_1R) \dots (p_kR) = (p_1 \dots p_k)R$, takže $a = p_1 \dots p_kc$ pro nějaké c invertibilní. Nahradíme-li prvočinitel p_k prvočinitelem p_kc vidíme, že každé neinvertibilní $a \in R$ lze vyjádřit jako součin prvočinitelů. Z tvrzení 1.13 víme, že tento součin je určen jednoznačně až na pořadí a až na modifikace invertibilními prvky. Vidíme, že tvrzení a důsledek 1.13 lze též zapsat jako

Tvrzení. *Buď R obor hlavních ideálů. Předpokladejme, že z každého maximálního ideálu M je vybrán právě jeden prvočinitel p, a označme P množinu všech takto vybraných prvočinitelů P. Pak každé nenulové $a \in R$ lze až na pořadí jednoznačně vyjádřit ve tvaru*

$$a = up_1^{e_1} \dots p_k^{e_k}, \text{ kde } u \in R^*, p_i \in P \text{ a } e_i \geq 1, 1 \leq i \leq k,$$

přičemž se předpokládá, že $p_1, \dots, p_k$ jsou po dvou různé.

Je-li $a = up_1^{e_1} \dots p_k^{e_k}$ a $b = vp_1^{f_1} \dots p_k^{f_k}$, kde $p_i \in P$ jsou po dvou různé, $e_i \geq 0$, $f_i \geq 0$, $1 \leq i \leq k$, a kde $u, v \in R^*$, tak a dělí b právě když $e_i \leq f_i$ pro všechna i, $1 \leq i \leq k$. □

1.15 Největší společný dělitel. Nejmenší společný násobek

Buď R obor integrity s nenulovými prvky a a b. Prvek d nazveme *největším společným dělitelem* prvků a a b, pokud splňuje

$$d|a, d|b \text{ a } t|d, \text{ kdykoliv } t|a \text{ a } t|b.$$

Podobně $n \in R$ nazveme *nejmenším společným násobkem* prvků a a b, pokud

$$a|n, b|n \text{ a } n|m, \text{ kdykoliv } a|m \text{ a } b|m.$$

Jsou-li d_1 a d_2 dva největší společní dělitelé, je $d_1|d_2$ a $d_2|d_1$, takže $d_1R = d_2R$. Podobně nahlédneme, že i nejmenší společné násobky jsou určeny jednoznačně až na násobek invertibilním prvkem (viz oddíl 1.10).

Ať R je obor hlavních ideálů, přičemž $a = up_1^{e_1} \dots p_k^{e_k}$ a $b = vp_1^{f_1} \dots p_k^{f_k}$ jsou rozklady prvků a a b na prvočinitele ve smyslu tvrzení 1.14 (předpokládáme, že e_i a f_i jsou nezáporné, $1 \leq i \leq k$). Z tohoto tvrzení okamžitě plyne

Lemma. *Největší společný dělitel prvků a a b je roven $p_1^{\min(e_1,f_1)} \dots p_k^{\min(e_k,f_k)}$. Jejich nejmenší společný násobek je roven $p_1^{\max(e_1,f_1)} \dots p_k^{\max(e_k,f_k)}$.* □

Tvrzení. *Buď R obor hlavních ideálů a ať a a b jsou jeho nenulové prvky. Označme d jejich největší společný dělitel a n jejich nejmenší společný násobek. Pak $dR = (aR) + (bR)$ a $nR = (aR) \cap (bR)$.*

Důkaz. Víme, že $(aR) + (bR)$ je rovno nějakému tR. Z $aR \subseteq tR$ a $bR \subset tR$ plyne, že t dělí jak a, tak b. Proto t dělí d, čili $dR \subseteq tR$. Současně máme $dR \supseteq aR$ a $dR \supseteq bR$, takže $dR \supseteq aR + bR = tR$.

Podobně je $(aR) \cap (bR)$ rovno nějakému mR. Z $aR \supseteq mR$ a $bR \supseteq mR$ plyne, že a i b dělí m. Proto n dělí m, čili $mR \subseteq nR$. Současně máme $nR \subseteq aR$ a $nR \subseteq bR$, takže $nR \subseteq aR \cap bR = mR$. □

Pojem největšího společného dělitele (i nejmenšího společného násobku) lze přímočaře zobecnit i na situace více prvků než dvou. Podle předchozího tvrzení je pak d největší společný dělitel prvků $a_1, \dots, a_k$ právě když $dR = a_1R + \dots + a_kR$. Tento fakt vyjádříme ještě jinou formou (jde o důsledek předchozího tvrzení).

Důsledek. *Ať R je obor hlavních ideálů a ať $a_1, \dots, a_k \in R$ jsou nenulové prvky, $k \geq 2$. Buď d největší společný dělitel těchto prvků. Pak d dělí $a_1r_1 +$*

$\cdots + a_k r_k$ *pro všechna* $r_1, \ldots, r_k \in R$. *Přitom* $r_1, \ldots, r_k \in R$ *lze zvolit tak, aby platilo* $d = a_1 r_1 + \cdots + a_k r_k$.

1.16 Eukleidovská zobrazení

Buď R obor integrity. Ať pro každé $a \in R$ je $f(a)$ nezáporné číslo, přičemž $f(a) = 0 \Leftrightarrow a = 0$. Zobrazení f nazveme *eukleidovským*, pokud navíc platí, že

(1) pro všechna $a, b \in R$, kde $b \neq 0$, lze najít $q, r \in R$ taková, že $a = bq + r$ a $f(r) < f(b)$; a

(2) $f(b) \leq f(a)$, kdykoliv b dělí a, $a \neq 0$

Pokud pro obor integrity R existuje alespoň jedno eukleidovské zobrazení, nazývá se tento obor *eukleidovským*.

Tvrzení. *Každý eukleidovský obor je oborem hlavních ideálů.*

Důkaz. Ať I je ideál R. Uvažme $b \in I$, $b \neq 0$, takové, že $f(b)$ nejmenší možné. Jistě $bR \subseteq I$. Ukážeme, že platí i opačná inkluze. Zvolme $a \in I$ a uvažme $q, r \in R$ taková, že $a = bq + r$, kde $f(r) < f(b)$. Prvek $r = a - bq$ leží v I. Musí tedy být $f(r) = 0$, odkud $r = 0$. Dokázali jsme, že b dělí každé $a \in I$. □

V důkazu jsme využili ze dvou vlastností eukleidovského zobrazení pouze vlastnost (1). Je tedy vlastnost (2) zbytečná? Ne zcela. Jednak lze její pomocí obdržet další vlastnosti (viz lemma níže), jednak v přirozeně se vyskytujících příkladech, které splňují (1), vždy platí. Navíc, jak nyní naznačíme, z vlastnosti (1) lze modifikací získat zobrazení, které splňuje i (2). Stačí položit $g(a) = \min\{f(ax); x \in R^*\}$. Snadnou úvahou lze ověřit, že g splňuje (1). Z důkazu tvrzení výše pak lze odvodit, že g splňuje i (2).

Lemma. *Ať* R *je obor integrity a* f *jeho eukleidovské zobrazení. Položme* $m = \min\{f(a); a \in R, a \neq 0\}$. *Pak* $u \in R$ *je invertibilní právě když* $f(u) = m$.

Důkaz. Je-li u invertibilní, tak u dělí každý prvek $a \in R$. Je-li $a \neq 0$, máme $f(u) \leq f(a)$, dle vlastnosti (2). Proto $f(u) = m$. Není-li $b \neq 0$ invertibilní, tak b nedělí prvek 1. Tudíž $1 = bq + r$, kde $r \neq 0$ a $f(r) < f(b)$. Proto $f(b) > m$. □

1.17 Okruhy polynomů

Buď R okruh. Formální součty tvaru $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, kde $a_0, \ldots, a_n \in R$, se považují za polynomy. Prvkům $a_0, \ldots, a_n$ se říká *koeficienty*. V zápise polynomu se nulové koeficienty nemusí uvádět. Rovněž lze v zápise polynomu měnit pořadí členů $a_i x^i$. Dva polynomy jsou shodné, právě když se shodují ve všech členech s nenulovými koeficienty. (Proto definujeme polynomy jako formální součty. Polynom je pro nás určen svým zápisem a ne funkčními hodnotami. Nad některými okruhy se totiž mohou různé polynomy shodovat při dosazení každé hodnoty z R.)

Polynom $a = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ nazveme *nulový* právě když $0 = a_0 = a_1 = \cdots = a_n$. Je-li a nenulový, definujeme jeho *stupeň* $\deg(a)$ jako největší $k \leq n$ takové, že $a_k \neq 0$. Pak se a_k nazývá *vedoucí koeficient*. Polynom

je *monický* právě když je nenulový a jeho vedoucí koeficient je roven 1. Stupeň nulového polynomu klademe definitoricky jako -1.

Množinu všech polynomů nad okruhem R budeme značit $R[x]$. Součet polynomů $a = \sum_{i\le n} a_i x^i$ a $b = \sum_{i\le n} b_i x^i$ se definuje jako $\sum_{i\le n}(a_i + b_i)x^i$. Je zřejmé, že platí $\deg(a+b) \le \max(\deg(a), \deg(b))$.

Součin polynomů $a = \sum_{i\le n} a_i x^i$ a $b = \sum_{j\le m} b_j x^j$ se definuje jako $\sum_{k\le n+m} c_k x^k$, kde $c_k = \sum_{i+j=k} a_i b_j$ (a tedy $c_k = \sum_{i\le k} a_i b_{k-i} = \sum_{j\le k} a_{k-j} b_j$). Koeficient c_{n+m} je roven $a_n b_m$. Je-li $n = \deg(a)$ a $m = \deg(b)$, je $c_{n+m} \ne 0$, pokud v R nejsou dělitelé nuly. Můžeme tedy například říci, že $\deg(ab) = \deg(a) + \deg(b)$, je-li R obor integrity a polynomy a, b jsou nenulové.

Pro polynomy $a = \sum a_i x^i$, $b = \sum b_j x^j$ a $c = \sum c_k x^k$ máme

$$a(bc) = a\Big(\Big(\sum_{r=j+k} b_j c_k\Big)x^r\Big) = \Big(\sum_{t=i+r} a_i\Big(\sum_{r=j+k} b_j c_k\Big)\Big)x^t = \Big(\sum_{t=i+j+k} a_i b_j c_k\Big)x^t.$$

Stejný výraz lze obdržet při výpočtu $(ab)c$, takže vidíme, že násobení polynomů je asociativní.

Sčítání v $R[x]$ zjevně poskytuje Abelovu grupu, kde nulový polynom je neutrálním prvkem. Násobení dává monooid s neutrálním prvkem $1 = 1.x^0$. Ověřit distributivitu násobení ke sčítání je snadné, takže vidíme, že $R[x]$ je okruh.

Prvky $\alpha \in R$ se obvykle ztotožňují s polynomy αx^0. Tímto způsobem lze R chápat jako podokruh $R[x]$. Přitom $R[x]$ je komutativní právě když je R komutativní, a ze vztahu $\deg(ab) = \deg(a) + \deg(b)$ plyne, že $R[x]$ je obor integrity právě když R je obor integrity.

Lemma. *Ať a,b jsou polynomy nad oborem integrity R. Předpokládejme, že b je nenulový a že jeho vedoucí koeficient je invertibilní prvek R. Pak existují polynomy $q, r \in R[x]$ takové, že $a = bq + r$, přičemž* $\deg(r) < \deg(b)$.

Důkaz. Je-li $\deg(a) < \deg(b)$, lze položit $q = 0$, $r = a$. Postupujeme dále indukcí dle $n = \deg(a) \ge k = \deg(b)$. Ať $a = \sum a_i x^i$ a $b = \sum b_j x^j$. Položme $\overline{a} = a - (a_n b_k^{-1})x^{n-k} b$. Oba polynomy rozdílu jsou stupně n a mají shodné vedoucí koeficienty. Proto je $\deg(\overline{a}) < n$, takže podle indukčního přepokladu existují $q, r \in R[x]$ takové, že $\overline{a} = bq + r$, $\deg(r) < k$, Odsud $a = \overline{a} + (a_n b_k^{-1})x^{n-k} b = b(a_n b_k^{-1} x^{n-k} + q) + r$. □

1.18 Komutativní tělesa a polynomy

Okruh R se nazývá *tělesem* právě když je netriviální a každý nenulový prvek je invertibilní. Jinými slovy, R je těleso právě když $R \setminus \{0\}$ je vzhledem k násobení grupa (značíme ji R^*).

Připomeňme, že v komutativním okruhu R každý vlastní ideál obsahuje vlastní hlavní ideál (je-i $a \in I$, $a \ne 0$, je $0 \subsetneq aR \subseteq I$). V komutativním tělese tedy vlastní ideály nejsou. Jinak řečeno, komutativní okruh R je tělesem právě když má přesně dva ideály, a to 0 a R.

Buď nyní T komutativní těleso. Pro $a \in T[x]$ položme $f(a) = \deg(a) + 1$. Pokud $b \in T[x]$ je nenulový polynom, tak podle lemmatu 1.17 existují $q, r \in T[x]$

taková, že $a = bq + r$, $f(r) < f(b)$. Vidíme, že f je eukleidovské zobrazení ve smyslu oddílu 1.16. $T[x]$ je tudíž eukleidovský obor, a speciálně obor hlavních ideálů (viz tvrzení 1.16).

Invertibilní prvky okruhu $T[x]$ jsou právě všechny polynomy stupně nula (tedy všechny nenulové prvky T). To lze nahlédnout mnoha způsoby, například pomocí lemmatu 1.16. V každém nenulovém hlavním ideálu I lze proto nalézt právě jeden generátor, který je monický (viz lemma 1.10). Všechny nenulové ideály $T[x]$ jsou tedy tvaru $aT[x]$, kde a je jednoznačně určený monický polynom.

Polynom $a \in T[x]$ takový, že a je v okruhu $T[x]$ prvočinitel, se nazývá *ireducibilní*. Podle lemmatu 1.14 je a ireducibilní právě tehdy když ideál $aT[x]$ je maximální. Tvrzení 1.14 znamená, že každý polynom $a \in T[x]$ lze jednoznačně až na pořadí vyjádřit ve tvaru $tp_1^{e_1} \dots p_k^{e_k}$, kde $t \in T$, $p_i \in T[x]$ monické ireducibilní, $e_i \geq 1$, přičemž $1 \leq i \leq k$ a polynomy p_i jsou po dvou různé. Vidíme, že polynom $a \in T[x]$ je ireducibilní právě když je nenulový a nemá dělitele b takového, že $0 < \deg(b) < \deg(a)$ (jinými slovy: nemá vlastního dělitele).

Dosazením prvku $\alpha \in T$ do polynomu $a = \sum a_i x^i$ se rozumí hodnota $a(\alpha) = \sum a_i \alpha^i$. Prvek $\alpha \in T$ se nazývá *kořenem* polynomu a právě tehdy když $a(\alpha) = 0$.

Lemma. *Buď $a \in T[x]$ nenulový polynom. Pak α je kořenem polynomu a právě když polynom $x - \alpha$ dělí a.*

Důkaz. Vyjádřeme a jako $(x - \alpha)b + \varrho$, kde $\deg(\varrho) < \deg(x - \alpha) = 1$. Je tedy $\varrho \in T$. Přitom $a(\alpha) = (\alpha - \alpha)b + \varrho = \varrho$. □

Je-li α kořenem polynomu $a \neq 0$, tak nejvyšší $e \geq 1$ takové, že $(x - \alpha)^e$ dělí a, se nazývá *násobnost* kořene α.

Tvrzení. *Buď $a \in T[x]$ nenulový polynom, kde T je komutativní těleso. Ať $\alpha_1, \dots, \alpha_k$ jsou všechny jeho kořeny, s násobnostmi po řadě $e_1, \dots, e_k$. Potom $\sum e_i \leq \deg(a)$.*

Důkaz. Víme, že $(x - \alpha_i)^{e_i}$ dělí a pro každé i, $1 \leq i \leq k$. Polynomy $x - \alpha_i$ jsou ireducibilní, takže z tvrzení 1.14 plyne, že rozklad a na prvočinitele má tvar $(x - \alpha_1)^{e_1} \dots (x - \alpha_k)^{e_k} q$, kde q je buď ireducibilní stupně alespoň 2, nebo $q \in T^*$. Tím pádem $\deg(a) = e_1 + \dots + e_k + \deg(q)$. □

Důsledek. *Buď T komutativní těleso a ať $a \in T[x]$ je stupně $n \geq 0$. Pak má a nanejvýš n kořenů.*

1.19 Součiny a homomorfismy

Ať $R_1, \dots, R_k$ jsou okruhy. Na množině $R_1 \times \dots \times R_k$ definujeme strukturu okruhu tak, že jednotlivé operace jsou definovány „po složkách“. Je tedy

$$(a_1, \dots, a_k) + (b_1, \dots, b_k) = (a_1 + b_1, \dots, a_k + b_k),$$

$$(a_1, \dots, a_k) \cdot (b_1, \dots, b_k) = (a_1 b_1, \dots, a_k b_k),$$

$$-(a_1, \dots, a_k) = (-a_1, \dots, -a_k)$$

a neutrální prvky jsou $0 = (0, \dots, 0)$ a $1 = (1, \dots, 1)$.

Vzhledem k tomu, že identity, které okruh definují, se vztahují k jednotlivým souřadnicím („složkám“), je $R = R_1 \times \cdots \times R_k$ skutečně okruhem.

Podobně lze definovat součin grup $G_1 \times \cdots \times G_k$. Výsledná grupa je komutativní právě když každá z grup G_i je komutativní.

Je-li $R_1 \times \cdots \times R_k$ součinem okruhů (nebo grup) je projekce $\pi_i : (a_1, \ldots, a_k) \mapsto a_i$ jistě homomorfismus $R_1 \times \cdots \times R_k \to R_i$.

Lemma. *Ať $R = R_1 \times \cdots \times R_k$ je součin okruhů (grup) a ať S je okruh (grupa). Zobrazení $f : S \to R$ je homomorfismus právě když $\pi_i f : S \to R_i$ je homomorfismus pro každé i, $1 \leq i \leq k$.*

Důkaz. Je-li f homomorfismus, je $\pi_i f$ homomorfismus pro každé i, $1 \leq i \leq k$. Pro důkaz opačným směrem uvažme ,že pro $a, b \in S$ máme $\pi_i f(a + b)$ rovno $\pi_i f(a) + \pi_i f(b)$, takže z $f(a) = (a_1, \ldots, a_k)$, $f(b) = (b_1, \ldots, b_k)$ vyplývá $f(a + b) = (a_1 + b_1, \ldots, a_k + b_k)$. Podobně lze postupovat i pro operaci násobení. □

1.20 Eukleidův algoritmus

Buď R eukleidovský obor integrity a ať $f : R \to \mathbb{Z}$ je eukleidovské zobrazení. Pro všechna $a, b \in R$, $b \neq 0$, existují tedy $q, r \in R$ taková, že $a = bq + r$, přičemž $f(r) < f(b)$. Podle tvrzení 1.16 je R oborem hlavních ideálů. V oddíle 1.16 jsme nahlédli, že prvky oboru hlavních ideálů mají vždy nějaký největší společný dělitel. Z existence největšího společného dělitele ovšem neplyne existence algoritmu, který ho nalezne. Nyní popíšeme stařičký algoritmus, jenž takové hledání umožňuje za předpokladu, že pro všechna $a, b \in R$, $b \neq 0$, je nalezení nějaké dvojice (q, r), $a = bq + r$ a $f(r) < f(b)$, rovněž algoritmicky možné.

Položme $a_0 = a$, $a_1 = b$ a konstruujme posloupnost $a_0, a_1, \ldots$ tak, že pro $i \geq 1$ odvodíme a_{i+1} z a_i a a_{i-1} právě tehdy, když a_i nedělí a_{i-1}. V takovém případě nalezneme q a r taková, že $a_{i-1} = a_i q + r$, $f(r) < f(a_i)$, a položíme $a_{i+1} = r$. Máme $r \neq 0$, a tedy $0 < f(r) = f(a_{i+1}) < f(a_i)$. Posloupnost a_i proto nelze konstruovat neomezeně, takže a_k dělí a_{k-1} pro nějaké $k \geq 1$.

Lemma. *a_k je největším společným dělitelem a a b.*

Důkaz. Protože a_k dělí a_{k-1}, tak je a_k také rovno největšímu společnému děliteli a_k a a_{k-1}. Stačí tedy ukázat, že každé $c \in R$ je společným dělitelem a_{i-1} a a_i právě tehdy, když je společným dělitelem a_i a a_{i+1}. To však plyne ze vztahu $a_{i-1} = a_i q + a_{i+1}$ okamžitě. □

Podle podle důsledku 1.15 pro největší společný dělitel d hodnot $a, b \in R$ existují $x, y \in R$ taková, že $xa + yb = d$. Stojí za povšimnutí, že eukleidovský algoritmus dovoluje i nalezení prvků x a y.

Vskutku, označme q_i tu hodnotu, pro kterou je $a_{i-1} = a_i q_i + a_{i+1}$. Každé a_i, $i \geq 1$, je možné vyjádřit jako $x_i a + y_i b = x_i a_0 + y_i a_1$,$1 \leq i \leq k$. To je víceméně zřejmé přímočarou rekurzivní úvahou. Jejím zpřesněním dostáváme i vzorce pro x_i a y_i:

Máme $x_1 = 0$ a $y_1 = 1$. Dále $a_{i+1} = a_{i-1} - a_i q_i = x_{i-1} a + y_{i-1} b - q_i (x_i a + y_i b) = (x_{i-1} - q_i x_i) a + (y_{i-1} - q_i y_i) b$, takže lze klást

$$x_{i+1} = x_{i-1} - q_i x_i \text{ a } y_{i+1} = y_{i-1} - q_i y_i.$$

Pak $a_k = x_k a + y_k b$, což je požadované vyjádření největšího společného dělitele. Můžeme tedy uzavřít

Tvrzení. *Buď R obor integrity s eukleidovským zobrazením f, a ať $a, b \in R$, $b \neq 0$. Konstruujme posloupnosti a_i $i \geq 0$, a x_i a y_i, $i \geq 1$, tak, že $a_0 = a$, $a_1 = b$, $x_1 = 0$, $y_1 = 1$, přičemž v případě, kdy a_i nedělí a_{i-1}, nalezneme $q, r \in R$ taková, že $a_i = a_{i-1}q + r$, $f(r) < f(a_{i-1})$, a položíme*

$$a_{i+1} = r, x_{i+1} = x_{i-1} - qx_i, y_{i+1} = y_{i-1} - qy_i.$$

Pokud a_i dělí a_{i-1}, položíme $k = i$, a posloupnost ukončíme.

Posloupnost je vždy konečná, platí $a_i = x_i a + y_i b$ pro každé i, $1 \leq i \leq k$, a a_k je největší společný dělitel a a b.

Důsledek. *Ať $T \subseteq U$ jsou do sebe vřazená komutativní tělesa a ať $a, b \in T[x]$. Pak existují $u, v, c \in T[x]$ takové, že $c = ua + vb$ je největším společným dělitelem a a b v $U[x]$.*

Důkaz. Hledejme eukleidovským algoritmem největší společný dělitel polynomů a a b, $b \neq 0$. Pracujeme uvnitř $U[x]$, polynomy takové, že $a = bq + r$, kde $\deg(r) < \deg(b)$ lze ovšem volit tak, aby q i r padlo do $T[x]$. Podobně lze i v $T[x]$ volit všechny další členy posloupnosti eukleidova algoritmu. Proto i výsledné polynomy leží v $T[x]$. □

1.21 Charakteristika těles

Ať R je okruh s jednotkou $e = 1$. Pak $e + e = 2e$, $e + e + e = 3e$, atd., kde násobky celým číslem znamenají iterované sčítání v abelově grupě $R(+, -, 0)$ ve smyslu oddílu 1.7. Je tedy $(n + m)e = ne + me$ a $n(me) = (nm)e$ pro všechna $n, m \in \mathbb{Z}$. Máme ale také $ne \cdot me = (nm)e$ což plyne z distributivního zákona a skutečnosti, že $e^2 = e$, například $(e+e)(e+e+e) = e(e+e+e)+e(e+e+e) = e^2 + e^2 + e^2 + e^2 + e^2 + e^2 = e + e + e + e + e + e = 6e$.

Je-li $ne = me$, tak musí být $(n - m)e = 0$. Pokud $ne \neq me$ pro všechna $n, m \in \mathbb{Z}$, $n \neq m$, říkáme, že R je *okruh charakteristiky nula*. Je-li $ke = 0$ pro nějaké $k \neq 0$, tak zvolíme nejmenší možné takové kladné k, a to nazveme *charakteristikou okruhu R*.

Je-li k číslo složené, $k = mn$, kde $0 < m < k$, tak máme $ne \neq 0$, $me \neq 0$, ale $ke = 0 = ne \cdot me$. Okruh složené charakteristiky má tedy dělitele nuly, a tudíž nemůže být oborem integrity, natož tělesem.

Není-li charakteristika okruhu rovna nule, hovoříme o okruzích kladné charakteristiky. Ukázali jsme, že *pokud je charakteristika tělesa kladná, je rovna nějakému prvočíslu.*

Tvrzení. *Buď R komutativní okruh prvočíselné charakteristiky p. Zobrazení $x \mapsto x^p$ je endomorfismem toho okruhu.*

Důkaz. Protože $(xy)^p = x^p y^p$ platí pro všechna $x, y \in R$, je třeba ukázat, že pro ně platí i $(x+y)^p = x^p + y^p$. Pro komutativní okruhy lze použít binomickou větu (což lze snadno ověřit indukcí), takže máme

$$(x + y)^p = x^p + \binom{p}{1} x^{p-1}y + \binom{p}{2} x^{p-2}y^2 + \cdots + \binom{p}{p-1} xy^{p-1} + y^p.$$

Pro každé $a \in R$ je $pa = (pe)a = 0 \cdot a = 0$. Proto stačí nahlédnout, že p dělí $\binom{p}{i}$, je-li $1 \leq i \leq p-1$. Protože $\binom{p}{i}$ je celé číslo reprezentováno zlomkem $(pu)/(i!)$, kde $u = (p-1)\dots(p-i+1)$, musí $i!$ dělit pu. Protože p nedělí $i!$, musí být $u/(i!)$ celé číslo. □

Endomorfismu $x \mapsto x^p$ se říkává *Frobeniův*. Buď nyní $f : R \to S$ homomorfismus komutativních okruhů. Uvažme zobrazení $f_x : R[x] \to S[x]$, které každému $a = \sum a_i x^i \in R[x]$ přiřazuje polynom $\sum f(a_i)x^i$. Slovně vyjádřeno, f_x nahrazuje koeficienty polynomu s využitím zobrazení f.

Lemma. *Buď $f : R \to S$ homomorfismus komutativních okruhů. Potom je i $f_x : R[x] \to S[x]$ okruhový homomorfismus.*

Důkaz. Ať $a = \sum a_i x^i$ a $b = \sum b_i x^i$ jsou polynomy z $R[x]$. Pak $f_x(a + b) = \sum f(a_i + b_i)x^i = \sum (f(a_i) + f(b_i))\, x^i = \left(\sum f(a_i)x^i\right) + \left(\sum f(b_i)x^i\right) = f_x(a) + f_x(b)$. Nechť $c = ab = \sum c_k x^k$. Máme $c_k = \sum_{i+j=k} a_i b_j$, takže $f_x(c)$ má koeficienty $f(c_k) = \sum_{i+j=k} f(a_i)f(b_i)$, odkud je rovnost $f_x(c) = f_x(a)f_x(b)$ okamžitě patrná. □

Důsledek. *Ať T je komutativní těleso kladné charakteristiky p, a ať $a \in T[x]$, $a = \sum a_i x^i$, je takový polynom, že $a_i^p = a_i$ pro každé $i \geq 0$. Jestliže α je kořen a, tak α^p je rovněž kořen a.*

Důkaz. Ať f označuje Frobeniův endomorfismus. Máme $f_x(a) = a$, a také $a = (x-\alpha)b$ pro nějaké $b \in T[x]$. Tudíž $a = f_x(a) = f_x(x-\alpha)f_x(b) = (x-\alpha^p)c$, kde $c = f_x(b)$. □

1.22 Rozšíření těles

Lemma. *Ať R je komutativní okruh a ať M je maximální ideál tohoto okruhu. Potom je R/M komutativní těleso. Je-li přitom $T \subseteq R$ podokruh, který je tělesem, je $t \mapsto t + M$ injektivním homomorfismem (takovým homomorfismům se často říká vnoření) tělesa T do tělesa R/M.*

Důkaz. Prvky okruhu R/M jsou rozkladové třídy modulo M. Přitom $r + M$ je nenulový prvek okruhu R/M právě když $r \notin M$. Součet hlavního ideálu rR a ideálu M (viz oddíl 1.12) je ideálem, který obsahuje r i M, a proto musí být roven R. Je tedy $1 \in rR + M$, odkud $1 = rs + m$ pro nějaké $m \in M$ a $s \in R$. Tudíž $(r + M)(s + M) = (1 - m) + M = 1 + M$, což je jednotka okruhu R/M. Prvek $r + M$ je invertibilní pro každé $r \in R \setminus M$, takže R/M je těleso.

Zobrazení $t \mapsto t+M$ je jistě homomorfismus $T \to R/M$ ($(s+M)+(t+M) = (s+t)+M$) a $(sM)\cdot(tM) = (st)M$), přičemž $t+M$, $t \in T\setminus\{0\}$, není nikdy rovno $0_{R/M} = 0+M$, neboť t je invertibilní prvek, a tedy neleží v M. Z $t+M = s+M$ ale plyne $t - s \in M$. Proto je zobrazení $t \mapsto t + M$ injektivní. □

Konstrukce *Buď T komutativní těleso a ať $a = \sum a_i x^i \in T[x]$ je ireducibilní polynom. Položme $U = T[x]/aT[x]$ a označme f vnoření $T \to U$, $t \mapsto t + aT[x]$. Pak $x + aT[x]$ je kořenem polynomu $f_x(a)$.*

Důkaz. Ideál $aT[x]$ je maximální (viz oddíl 1.18) takže z předcházejícího lem-

matu plyne, že U vskutku je komutativní těleso a že f je korektně definovaný homomorfismus, který je injektivní. Definice f_x je v oddíle 1.21. Máme $f_x(a)(x+T[x]) = \sum(a_i + aT[x])(x+aT[x])^i = \sum(a_i+aT[x])(x^i+aT[x]) = \left(\sum a_i x^i\right) + aT[x] = a + aT[x] = aT[x] = 0_{T[x]/aT[x]}$. □

Každý polynom stupně alespoň jedna je součin ireducibilních. Ztotožníme-li tedy v předchozí konstrukci každé $\alpha \in T$ s jeho obrazem $\alpha + aT[x]$, stane se U rozšíření T, ve kterém má a alespoň jeden kořen. Proto platí

Důsledek. *Buď $a \in T[x]$ polynom stupně alespoň jedna, T komutativní těleso. Pak existuje komutativní těleso $U \supseteq T$, ve kterém má a alespoň jeden kořen.*

Zmíněný polynom $a \in T[x]$ lze tedy vyjádřit jako $(x-\alpha)b$, kde $b \in U[x]$, $\alpha \in U$. Je-li b stupně alespoň 1, můžeme najít nadtěleso U, ve kterém má b alespoň jeden kořen.

Pokračováním tohoto procesu pak můžeme nalézt $V \supseteq U$ takové, že V je komutativní těleso a $a = (x-\alpha_1)^{e_1} \dots (x-\alpha_k)^{e_k}$ pro nějaké $\alpha_1, \dots, \alpha_k \in V$, a nějaká $e_1 \geq 1, \dots, e_k \geq 1$. (Říkáme, že a se *rozkládá na kořenové činitele* nad V.) Můžeme proto vyslovit

Tvrzení. *Buď T komutativní těleso a ať $a \in T[x]$, $\deg(a) \geq 1$. Pak existuje komutativní těleso $U \supseteq T$ takové, že a se nad U rozkládá na kořenové činitele.* □

1.23 Násobnost a derivace

Buď $a \in T[x]$, kde T je komutativní těleso, $a = a_n x^n + \dots + a_0 x_0$. Podobně jako v analýze definujeme derivaci a' polynomu a tak, že $a' = na_n x^{n-1} + \dots + a_1$. Jinými slovy, $a' = \sum a'_i x^i$, kde $a'_i = (i+1)a_{i+1}$ pro každé $i \geq 0$.

Je třeba si uvědomit, že derivace v námi uvedeném smyslu nepožaduje od T žádné topologické vlastnosti a že se jí také nepřisuzuje žádný geometrický význam. Základní vzorce platí ovšem obdobně:

Lemma. *Buď $a, b \in T[x]$ polynomy. Pak $(a+b)' = a' + b'$, $(ab)' = a'b + ab'$ a $(ta)' = ta'$ pro každé $t \in T$.*

Důkaz. Vztahy pro součet a skalární násobek jsou okamžitě zřejmé. Ať $c = ab$, $c = \sum c_k x^k$. Pak $c_k = \sum_{i+j=k} a_i b_j$, kde $a = \sum a_i x^i$ a $b' = \sum b_j x^j$. Nechť $c = \sum c'_k x^k$, $a' = \sum a'_i x^i$ a $b' = \sum b'_j x^j$. Chceme ukázat, že pro každé $k \geq 0$ je $c'_k = \sum_{i+j=k} a'_i b_j + \sum_{i+j=k} a_i b'_j$. Pravá strana je rovna $\sum_{i+j=k}(i+1)a_{i+1}b_j + \sum_{i+j=k}(j+1)a_i b_{j+1} = \sum_{i+j=k+1} i a_i b_j + \sum_{i+j=k+1} j a_i b_j = \sum_{i+j=k+1}(i+j)a_i b_j = (k+1)\sum_{i+j=k+1} a_i b_j = (k+1)c_{k+1} = c'_k$. □

Tvrzení. *Nechť a je nenulový polynom nad komutativním tělesem T. Je-li $\alpha \in T$ kořenem a násobnosti $e \geq 2$, tak $(x-\alpha)^{e-1}$ dělí jak a, tak a'.*

Důkaz. Je-li α násobnosti $e \geq 2$, tak $a = (x-\alpha)^e b$ pro nějaké $b \in T[x]$. Odsud $a' = (e-1)(x-\alpha)^{e-1}b + (x-\alpha)^e b' = (x-\alpha)^{e-1}((e-1)b + (x-\alpha)b')$. □

Důsledek. *Ať $a \in T[x]$, $a \neq 0$, je nesoudělné s a'. Pak nad žádným komutativním tělesem $U \supseteq T$ nemá a vícenásobný kořen. Speciálně tomu tak je, pokud*

$a = x^n - 1$ a *charakteristika* p *tělesa* T *nedělí* n, $n \geq 1$.

Důkaz. Máme $a' \in T[x]$, přičemž největší společný dělitel a a a' lze spočítat eukleidovým algoritmem (viz oddíl 1.20). Proto leží v $T[x]$ a nemění se při přechodu k nadtělesu U. Polynomy $x^n - 1$ a $(x^n - 1)' = nx^{n-1}$ jistě v případě $nx^n - 1 \neq 0$ nesoudělné jsou. Nerovnost vyplývá z předpokladu o nedělitelnosti n charakteristikou tělesa. Jsou-li a a a' nesoudělné, nemůže mít podle tvrzení výše polynom a vícenásobný kořen. □

Kapitola 2

Vlastnosti a použití cyklických grup

2.1 Počítání modulo n

Tvrzení. *Okruh celých čísel $\mathbb{Z}$ je oborem hlavních ideálů. Každý jeho ideál je roven některému z ideálů $n\mathbb{Z}$, $n \geq 0$. Toto n je určeno jednoznačně.*

Důkaz. Pro všechna $a, b \in \mathbb{Z}$, $b \neq 0$, existují $q, r \in \mathbb{Z}$ taková, že $a = bq + r$,$0 \leq r < b$. Vidíme, že $a \to |a|$ je eukleidovskou funkcí ve smyslu oddílu 1.16. Proto je $\mathbb{Z}$ oborem hlavních ideálů a jeho invertibilní prvky jsou 1 a -1 (viz tvrzení a lemma téhož oddílu). Generátor hlavního ideálu je určen jednoznačně až na invertibilní prvek (viz lemma 1.10), odsud jednoznačnost n. □

Místo $a \equiv b \bmod n\mathbb{Z}$ se píše pouze $a \equiv b \bmod n$. Podmínku $a - b \in n\mathbb{Z}$ ze zapsat též jako $n|a - b$. Faktorokruh $\mathbb{Z}/n\mathbb{Z}$ (viz oddíl 1.9) se tedy skládá z rozkladových tříd $0 + n\mathbb{Z}$, $1 + n\mathbb{Z}$,...,$(n-1) + n\mathbb{Z}$.

Ztotožníme-li každé z celých čísel i, $0 \leq i < n$, s třídou $n\mathbb{Z}$, dostaneme okruh izomorfní $\mathbb{Z}/n\mathbb{Z}$. Budeme ho značit $\mathbb{Z}_n$.

Označíme-li na chvíli binární operace $\mathbb{Z}_n$ symboly $\oplus$ a $\odot$, bude platit $0 \leq a \oplus b < n$, $0 \leq a \odot b < n$, $a \oplus b = a + b \mod n$, $a \odot b = a \cdot b \mod n$. Jinak řečeno hodnoty $a \oplus b$ i $a \odot b$ obdržíme tak, že provedeme příslušnou operaci běžným způsobem, a za výsledek vezmeme zbytek po dělení číslem n.

Množina $0, 1, \ldots, n-1$, tvoří *úplnou soustavu zbytků* modulo n (tj. z každé rozkladové třídy modulo n je vybrán právě jeden prvek). Úplných soustav zbytků modulo n existuje samozřejmě nekonečně mnoho. Leckdy je užitečné pracovat místo soustavy $0, 1, \ldots, n-1$ se soustavou tvořenou všemi celými i, jež splňují $-n/2 \leq i < n/2$. Okruh $\mathbb{Z}_n$ však budeme chápat jako definovaný na množině $\{0, 1, \ldots, n-1\}$. Přitom operace $\oplus$ a $\odot$ budeme značit obvykle běžnými symboly sčítání a odčítání.

Budeme-li hovořit o $\mathbb{Z}_n$ jako o grupě, budeme tím rozumět Abelovu grupu $\mathbb{Z}_n(+, -, 0)$.

2.2 Cyklické grupy

Buď G grupa (v multiplikativní notaci). Pak G nazveme *cyklickou*, pokud existuje $a \in G$ takové, že $G = \{a^i; i \in \mathbb{Z}\}$. O a říkáme, že *generuje* grupu G, a že je jejím *generátorem*. Protože $a^i a^j = a^{i+j} = a^j a^i$, pro všechna $i, j \in \mathbb{Z}$, je cyklická grupa nutně abelovská. Můžeme tedy změnit notaci a uvažovat G v notaci aditivní. Pak G je cyklická právě když $G = \{ia; i \in \mathbb{Z}\}$ pro nějaké $a \in G$.

Grupy $\mathbb{Z}(+,-,0)$ i $\mathbb{Z}_n(+,-,0)$ jistě cyklické jsou, za generátor lze vždy volit prvek 1.

Tvrzení. *Ať G je cyklická grupa v aditivní notaci s generátorem a. Je-li G nekonečná, je $i \mapsto ia$ izomorfismus $\mathbb{Z} \cong G$. Je-li G konečného řádu n, je $i \mapsto ia$ izomorfismus $\mathbb{Z}_n \cong G$.*

Důkaz. Budeme používat vztahy pro iterované sčítání a násobení (viz oddíl 1.7). Předpokládejme nejprve, že $na = 0$ pro nějaké $n \neq 0$. Pak $kna = 0$ pro všechna $k \in \mathbb{Z}$, takže lze zvolit $n > 0$. Předpokládejme, že je nejmenší možné. Máme $ia = (i + kn)a$ pro každé i, $0 \leq i < n$ a každé $k \in \mathbb{Z}$. Je-li $0 \leq i < j < n$, tak $ia = ja$ implikuje $(j - i)a = 0$, odkud $j = i$, neboť $j - i < n$. Proto $0, a, \ldots, (n-1)a$ jsou právě všechny prvky grupy G. Je-li $i, j \in \{0, 1, \ldots, n-1\}$ a $i + j = \varepsilon n + h$, kde $h \in \{0, 1, \ldots, n-1\}$ a $\varepsilon \in 0, 1$, pak $(i + j)a = ha$. Proto je zobrazení $i \mapsto ia$ izomorfismem grup (viz oddíl 1.4).

Jestliže $na \neq 0$ pro každé $n \neq 0$, $n \in \mathbb{Z}$, je i $na \neq ma$ pro všechna $n, m \in \mathbb{Z}$, $n \neq m$ (jinak by bylo $(n - m)a = 0$). Tudíž zobrazení $i \mapsto ia$ je bijektivní homomorfismus $\mathbb{Z} \to G$, a tedy izomorfismus. □

V dalším budeme zkoumat vlastnosti grupy (a okruhu) $\mathbb{Z}_n$. Z předchozího tvrzení plyne, že tím vlastně zkoumáme vlastnosti konečných cyklických grup. Ty se totiž podle tvrzení vhodným označením dají se $\mathbb{Z}_n$ ztotožnit.

2.3 Podgrupy konečných cyklických grup

V okruhu $\mathbb{Z}_n$ lze pro každé $d \in \mathbb{Z}$ uvažovat ideál $d\mathbb{Z}_n$. Jestliže $n = dr$ pro nějaké $r \in \mathbb{Z}$, tak pro všechna $a, i \in \mathbb{Z}$ platí $d(ar + i) \equiv di \mod n$. Proto v takovém případě platí $d\mathbb{Z}_n = \{0, d, 2d, \ldots, (r-1)d\}$.

Tvrzení. *Buď n celé kladné číslo. Pro $A \subseteq \mathbb{Z}_n$ je ekvivalentní:*

(i) A je netriviální podgrupa grupy $\mathbb{Z}_n$,

(ii) A je nenulový ideál okruhu $\mathbb{Z}_n$,

(iii) $A = \{0, d, 2d, \ldots, (r-1)d\}$ pro nějaké $d|n$, $1 \leq d < n$, kde $n = dr$.

Důkaz. Pro $a, b \in \mathbb{Z}_n$ je součin $a \cdot b$ možno obdržet jako součet a sčítanců prvku b. Proto ideály a podgrupy $\mathbb{Z}_n$ splývají. Je-li A nenulový ideál $\mathbb{Z}_n$, zvolíme nejmenší $a \in A$, $a > 0$. Označme m nejmenší násobek a takový, že $m \geq n$. Pokud $m \neq n$, tak $0 < m - n < a$, přičemž $m - n \in A$. To by však byl spor s volbou a, takže a dělí n. □

Důsledek. *Ať G je cyklická grupa řádu n. Pak G obsahuje podgrupu řádu d právě když d dělí n. Pokud d dělí n, tak existuje jediná podgrupa řádu d, a ta je cyklická.*

2.4 Endomorfismy konečných cyklických grup

Endomorfismus okruhu $\mathbb{Z}_n$ musí zobrazovat 1 na 1, a tím pádem $1+1$ na $1+1$, a tak dále. Jediným endomorfismem okruhu $\mathbb{Z}_n$ je proto identita. Endomorfismů grupy $\mathbb{Z}_n$ je však více:

Tvrzení. *Buď $n \geq 2$. Pro každé $a \in \mathbb{Z}_n$ je zobrazení $i \mapsto ia$ endomorfismem grupy $\mathbb{Z}_n$. Tento endomorfismus je automorfismus právě když a a n jsou (jakožto celá čísla) nesoudělná.*

Důkaz. Zobrazení $i \mapsto ia$ je endomorfismem, neboť $(i+j)a = ia + ja$ pro všechna $i, j \in \mathbb{Z}_n$. Je-li φ endomorfismus a $\varphi(1) = a$, tak musí být $\varphi(i) = \varphi(i \cdot 1) = i\varphi(a) = ia$. Uvážili jsme všechny možné obrazy $\varphi(1) = a \in \mathbb{Z}_n$, a proto jsme popsali všechny endomorfismy této grupy. Jestliže $d > 1$ je nějaký společný dělitel čísel a a n, tak $\varphi(i) \in d\mathbb{Z}_n$ pro každé $i \in \mathbb{Z}_n$, takže φ není bijektivní. Jestliže a je nesoudělné s n, tak podle důsledku 1.15 existují $s, r \in \mathbb{Z}_n$, že $sa + rn = 1$. Tudíž $sa \equiv 1 \mod n$, takže $ba = 1$ v $\mathbb{Z}_n$ pro nějaké $b \in \mathbb{Z}_n$. To znamená, že endomorfismy $i \mapsto ia$ a $i \mapsto ib$ jsou navzájem inverzní (ve složení se 1 vždy zobrazí na 1), takže běží o bijektivní zobrazení. □

Je-li $\alpha \in \mathrm{Aut}(\mathbb{Z}_n)$, tak obrazem podgrupy $d\mathbb{Z}_n$, kde d dělí n, $d > 0$, je podgrupa $\alpha(d\mathbb{Z}_n)$. Ta má ovšem stejný počet prvků, jako $d\mathbb{Z}_n$, a proto $\alpha(d\mathbb{Z}_n) = d\mathbb{Z}_n$, podle důsledku 2.3. Každé b, $0 < b < n$, lze jednoznačně vyjádřit jako da, kde d dělí n a a je s n nesoudělné (d je největší společný dělitel b a n). Uvažme automorfismus $\alpha : i \mapsto ai$. Pak $\alpha(di) = adi = bi$.

Vidíme, že jsme obdrželi

Důsledek. *Pro všechna $a \in \mathbb{Z}_n$, $a \neq 0$, je $a\mathbb{Z}_n$ rovno $d\mathbb{Z}_n$, kde d je největší společný dělitel a s n.*

2.5 Zavedení Eulerovy funkce

Tvrzení. *Buď $n \geq 2$, $0 < a < n$. Je ekvivalentní:*

(i) Přirozená čísla a a n jsou nesoudělná;

(ii) ideál $a\mathbb{Z}_n$ je roven $\mathbb{Z}_n$;

(iii) a je jako prvek $\mathbb{Z}_n$ invertibilní;

(iv) zobrazení $i \mapsto ai$ je automorfismus grupy $\mathbb{Z}_n$.

Přitom $\mathbb{Z}_n$ je těleso právě když n je prvočíslo.

Důkaz. Je-li a s n nesoudělné, je $a\mathbb{Z}_n = \mathbb{Z}_n$ dle důsledku 2.4. Pokud $a\mathbb{Z}_n = \mathbb{Z}_n$, tak $ab = 1$ pro nějaké b. Je-li $ab = 1$, jsou endomorfismy $i \mapsto ai$, $i \mapsto bi$ vzájemně inverzní, takže jde o automorfismy. Je-li $i \mapsto ai$ automorfismus, je a nesoudělné

s n, dle tvrzení 2.4. V tělese jsou všechny nenulové prvky invertibilní, což zjevně nastává jedině když n nemá vlastního dělitele. □

Počet všech a splňujících ekvivalentní podmínky tvrzení 2.5 označíme $\varphi(n)$. Zobrazení $n \mapsto \varphi(n)$ se říká *Eulerova funkce*. Definitoricky $\varphi(1) = 1$.

V Abelově grupě G generuje každé $a \in G$ cyklickou podgrupu $\{ia; i \in \mathbb{Z}\}$. Je-li $G = \mathbb{Z}_n$, je tato podgrupa rovna ideálu $a\mathbb{Z}_n$. Tvrzení výše tedy říká, že $\mathbb{Z}_n$ má právě $\varphi(n)$ různých jednoprvkových generátorů. Podgrupa $d\mathbb{Z}_n$, kde d dělí n, $n > d > 0$, je cyklická a izomorfní $\mathbb{Z}_{n/d}$. Proto má $\varphi(n/d)$ generátorů. Podgrupa $0 = 0\mathbb{Z}_n = n\mathbb{Z}_n$ je izomorfní $\mathbb{Z}_{n/n} = \mathbb{Z}_1$, a má $\varphi(1) = 1$ generátorů. Protože $\mathbb{Z}_n$ má n prvků a každý z nich generuje právě jednu podgrupu, dostaneme z tvrzení 2.5 vztah $n = \sum_{d|n} \varphi(n/d)$. Ovšem pokud d probíhá všechny dělitele n, tak n/d probíhá také všechny dělitele. Proto lze obdržený vztah vyjádřit jako

Důsledek. *Pro každé $n \geq 1$ je $n = \sum_{d|n} \varphi(d)$.*

Vzorec pro výpočet φ není obtížné odvodit. Učiníme tak však až poté, co ukážeme aplikaci Eulerovy funkce, která potřebuje pouze zřejmou nerovnost

$$1 \leq \varphi(n) < n.$$

2.6 Cykličnost podgrup tělesa

Nechť T je komutativní těleso a ať U je podgrupa T^* konečného řádu n. Každé $a \in U$ generuje cyklickou podgrupu $\{a^i; i \in \mathbb{Z}\}$. Podle Lagrangeovy věty dělí řád této podgrupy číslo n. Pro každé $d|n$ označme $\tau(d)$ počet $a \in U$, které generují podgrupu řádu d. Protože každý prvek generuje nějakou cyklickou podgrupu, musí platit $\sum_{d|n} \tau(d) = n$.

Předpokládejme, že pro nějaké $d|n$ platí $\tau(d) > \varphi(d)$. Zvolme libovolné $a \in U$ řádu d a položme $A = \{a^i; i \in \mathbb{Z}\}$. Podgrupa A má právě d prvků a každý její prvek je kořenem polynomu $x^d - 1$. Ovšem $A \cong \mathbb{Z}_d$ obsahuje podle oddílu 2.5 právě $\varphi(d)$ prvků řádu d. Protože předpokládáme $\tau(d) > \varphi(d)$, musí existovat ještě nějaké $b \in U \setminus A$ řádu d. I toto b je kořenem polynomu $x^d - 1$. Vidíme, že z $\tau(d) > \varphi(d)$ plyne existence alespoň $d+1$ kořenů polynomu $x^d - 1$. To je ovšem ve sporu s důsledkem 1.18, takže musí být $\tau(d) \leq \varphi(d)$ pro každé $d|n$.

Spojením této nerovnosti s rovností $n = \sum \tau(d) = \sum \varphi(d)$ (viz důsledek 2.5) obdržíme $\tau(d) = \varphi(d)$ pro každé d dělící n. Specielně je tedy $\tau(n) \geq 1$, což znamená, že alespoň jeden prvek U má řád n. To je však jen jiný způsob jak říci, že U je grupa cyklická. Dokázali jsme

Tvrzení. *V každém komutativním tělese je každá konečná multiplikativní podgrupa cyklická.* □

Pro účely teorie čísel je zásadní

Důsledek. *Grupa $\mathbb{Z}_p^*$ je cyklická pro každé prvočíslo p.*

Grupa $\mathbb{Z}_p^*$ má $p-1$ prvků (specielně tedy $\varphi(p) = p-1$). Je-li ξ její generátor, tak $i \mapsto \xi$ dává izomorfismus $\mathbb{Z}_{p-1} \cong \mathbb{Z}_p^*$. Každý takový generátor se nazývá *primitivní prvek* modulo p.

2.7 Součinové rozklady modulo n

Ať $n \geq 1$ je celé. V tomto oddílu použijeme pro $a \in \mathbb{Z}$ značení $(a)_n$ tak, že $b = (a)_n$ právě když $a \equiv b \mod n$, $0 \leq b < n$. Jinými slovy, $(a)_n$ je nezáporný zbytek při dělení čísla a číslem n.

Lemma. *Ať d dělí n. Pak $a \mapsto (a)_d$ je homomorfismus okruhů $\mathbb{Z}_n \to \mathbb{Z}_d$.*

Důkaz. Binární operace v $\mathbb{Z}_n$ lze zapsat jako $(a+b)_n$, $(a \cdot b)_n$. Z $d|n$ plyne, že $((a)_n)_d = (a)_d$, takže $((a)_d + (b)_d)_d = (a+b)_d = ((a+b)_n)_d$ pro všechna $a, b \in \mathbb{Z}$. Stejný vztah platí i když sčítání nahradíme násobením. □

Různé varianty následujícího tvrzení jsou známy jako *Čínská věta o zbytcích*

Tvrzení. *Buď $n = n_1 \cdot \cdots \cdot n_r$ celé číslo takové, že $n_1 \geq 2$ a $n_i, 1 \leq i \leq r$, jsou celá, větší než 1, která jsou po dvou nesoudělná. Položme $m_i = n/n_i$, $1 \leq i \leq r$. Pak existují celá k, $1 \leq i \leq r$, taková, že $\sum k_i m_i = 1$. Zobrazení $\mathbb{Z}_n \to \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_r} : a \mapsto ((a)_{n_1}, \ldots, (a)_{n_r})$ je izomorfismem okruhů. Zobrazení $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_r} \to \mathbb{Z}_n : (a_1, \ldots, a_r) \mapsto (\sum_{i=1}^r a_i k_i m_i)_n$ je izomorfismem inverzním.*

Důkaz. Největší společný dělitel čísel $m_1, \ldots, m_r$ je 1. Proto existují celá čísla k_i taková, že $\sum k_i m_i = 1$ (viz důsledek 1.15). Víme, že každé ze zobrazení $a \mapsto (a)_{n_i}$, je homomorfismus $\mathbb{Z}_n \to \mathbb{Z}_{n_i}$. Podle lemmatu 1.19 je tudíž homomorfismus i zobrazení $\mathbb{Z}_n \to \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_i}$ popsané ve znění tvrzení.

Označme ho α a označme β druhé z popsaných zobrazení. Grupy $\mathbb{Z}_n$ a $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_r}$ mají stejný konečný řád, takže k důkazu, že β a α jsou vzájemně inverzní, stačí ověřit, že $\alpha\beta(a_1, \ldots, a_r) = (a_1, \ldots, a_r)$ pro všechny $(a_1, \ldots, a_r) \in \mathbb{Z}$. Položme $a = \sum a_i k_i m_i$ a zvolme j, $1 \leq i \leq n$. Potřebujeme ukázat $(a)_{n_j} = a_j$. Protože n_j dělí m_i pro $i \neq j$, máme $(a)_{n_j} = (a_j k_j m_j)_{n_j} = (a_j)_{n_j} (k_j m_j)_{n_j}$. Jelikož a_j je prvek $\mathbb{Z}_{n_j}$, tak $(a_j)_{n_j} = a_j$. Protože $\sum k_i m_i = 1$ a protože n_j dělí m_i, pro $i \neq j$, máme $1 = (1)_{n_j} = (\sum k_i m_i)_{n_j} = (k_j m_j)_{n_j}$. □

Důsledek. *Buď $n = p_1^{e_1} \ldots p_k^{e_k}$ prvočíselný rozklad přirozeného čísla $n \geq 2$. Pak je $\mathbb{Z}_n \cong \mathbb{Z}_{p_1^{e_1}} \times \cdots \times \mathbb{Z}_{p_k^{e_k}}$.*

2.8 Výpočet Eulerovy funkce

Obecné lemma *Ať $\varphi : R \cong R_1 \times \cdots \times R_k$ je izomorfismus okruhů. Restrikce φ na R^* dává izomorfismus $R^* \cong R_1^* \times \cdots \times R_k^*$. Speciálně je $R_1^* \times \cdots \times R_k^*$ rovno grupě všech invertibilních prvků okruhu $R_1 \times \cdots \times R_k$.*

Důkaz. Invertibilní prvky skutečně tvoří grupu (viz oddíl 1.10). Izomorfismus okruhů jistě zobrazuje invertibilní prvek na invertibilní prvek. Proto stačí ukázat, že $(a_1, \ldots, a_n) \in (R_1 \times \ldots R_k)^*$ je invertibilní právě když každé z a_i, $1 \leq i \leq n$, je invertibilní. Obě podmínky ovšem značí existenci b_i, $1 \leq i \leq n$, takových, že $a_i b_i = 1 = b_i a_i$. □

Tvrzení. *Buď $n = n_1 \ldots n_r$, kde $r \geq 1$ a čísla $n_i \geq 1$ jsou celá a po dvou nesoudělná. Pak $\mathbb{Z}_n^* \cong \mathbb{Z}_{n_1}^* \times \cdots \times \mathbb{Z}_{n_r}^*$ a $\varphi(n) = \varphi(n_1) \ldots \varphi(n_r)$.*

Důkaz. Izomorfismus plyne z obecného lemmatu a tvrzení 2.7. Obě grupy mají

tedy stejný řád, a ten lze podle tvrzení 2.5 vyjádřit jednak jako $\varphi(n)$, jednak jako $\varphi(n_1)\dots\varphi(n_r)$. □

Je-li p prvočíslo, a $e \geq 1$, pak mezi čísly $\{0, 1, \dots, p^e - 1\}$ je právě p^{e-1} čísel dělitelných p. Ostatní čísla jsou s p^e nesoudělná. Proto je $\varphi(p^e)$ rovno $\varphi(p^e) = p^e - p^{e-1} = p^{e-1}(p-1) = p^e\left(1 - \frac{1}{p}\right)$. Z tvrzení výše tedy plyne

Důsledek. *Ať $n = p_1^{e_1} \dots p_k^{e_k}$ je prvočíselný rozklad. Pak*

$$\varphi(n) = (p_1 - 1)\dots(p_k - 1)p_1^{e_1-1}\dots p_k^{e_k-1}$$

Tuto hodnotu lze též zapsat jako $n\prod_{p|n}\left(1 - \frac{1}{p}\right)$.

Uvedený vzorec platí i pro $n = 1$, přijmeme-li obvyklou konvenci, že součin nulového počtu činitelů je roven 1.

2.9 Valuace a mocniny

V tomto oddíle odvodíme několik vztahů potřebných pro popis struktury grupy $\mathbb{Z}_{p^e}^*$. Pro p prvočíslo a $n \in \mathbb{Z}$, $n \neq 0$, bude $v_p(n)$ značit největší $j \geq 0$ takové, že p^j dělí n. Hovoříme o *p-valuaci* čísla n. Někdy se klade $v_p(0) = \infty$.

Pro $n \neq 0$ tedy máme $n = \prod_{p\in\mathbb{P}} p^{v_p(n)}$, kde $\mathbb{P}$ označuje množinu všech prvočísel.

Lemma. *$v_p(p^s - a) = v_p(a)$ kdykoliv $1 \leq a < p^s$, $s \geq 0$.*

Důkaz. Ať $j = v_p(a)$. Pak $j < s$, takže p^j dělí $p^s - a$. Současně z $p^k | p^s - a$ také plyne $k < s$, takže p^k dělí $a = p^s - (p^s - a)$. □

Připomeňme, že pro *kombinační číslo* $\binom{n}{k}$, $n \geq k \geq 1$, platí

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdot\dots\cdot(n-k+1)}{1\cdot 2\cdot\dots\cdot k}$$

Důsledek. *$v_p\binom{p^s}{k} = s - v_p(k)$ kdykoliv $1 \leq k \leq p^s$, $s \geq 0$.*

Důkaz. Pro každé a, $1 \leq a < k$, máme $v_p(p^s - a) = v_p(a)$, podle lemmatu. Proto p-valuace celého kombinačního čísla závisí jen na p-valuacích čísel p^s a k. □

Tvrzení. *Buď p prvočíslo. Je-li p liché, pak pro každé $e \geq 2$*

$$(1+p)^{p^{e-2}} \equiv 1 + p^{e-1} \mod p^e.$$

Pro $p = 2$ platí obdobný vztah $5^{2^{e-3}} \equiv 1 + 2^{e-1} \mod 2^e$, pokud $e \geq 3$.

Důkaz. Je-li $p = 2$, je $5 = 1 + p^2 = 1 + 4$ (proto hovoříme pro $p = 2$ o vztahu obdobném, nikoliv o vztahu totožném).

Vyjdeme z binomických rozvojů

$$(1+p)^{p^{e-2}} = 1 + p^{e-1} + \binom{p^{e-2}}{2}p^2 + \dots + \binom{p^{e-2}}{p^{e-2}}p^{p^{e-2}} \quad \text{a}$$

$$(1+4)^{2^{e-3}} = 1 + 2^{e-1} + \binom{2^{e-3}}{2}4^2 + \cdots + \binom{2^{e-3}}{2^{e-3}}4^{2^{e-3}}$$

Je třeba ověřit, že p^e dělí každý z členů tohoto rozvoje, s výjimkou prvních dvou. K tomu nám poslouží výše uvedený důsledek. Pro p liché jde o p-valuaci čísla $\binom{p^{e-2}}{i}p^i$, $i \geq 2$, která je podle důsledku rovna $e-2-v_p(i)+i$. Pro $p=2$ dostáváme $\binom{2^{e-3}}{i}2^{2i}$, $i \geq 2$, a 2-valuaci $e-3-v_2(i)+2i$. Potřebujeme tedy dokázat $i \geq v_p(i)+2$, p liché, a $2i \geq v_2(i)+3$, pro každé $i \geq 2$.

Položme $k = v_p(i)$. Pak $i = p^k a$ pro nějaké a nedělitelné p. Je-li $k=0$, pak obě nerovnosti zřejmě platí. Předpokládejme $k \geq 1$ a vyjděme z nerovnosti $i \geq p^k$. Stačí ukázat $p^k \geq k+2$, p liché, a $2^{k+1} \geq k+3$, pro každé $k \geq 1$. To je ovšem již snadné – například indukcí dle k. □

2.10 Násobení modulo mocniny prvočísla

Buď p prvočíslo a $e \geq 1$. Podle oddílu 2.8 má $\mathbb{Z}^*_{p^e}$ právě $(p-1)p^{e-1}$ prvků. Je-li $p^e = 4$, jde o prvky 1 a 3, které tvoří cyklickou dvouprvkovou grupu. Případ $p^e = 4$ v dalším uvažovat nebudeme.

Lemma. *Množina $P = \{1+ap; 0 \leq a < p^{e-1}\}$ tvoří podgrupu $\mathbb{Z}^*_{p^e}$ řádu p^{e-1}. Prvek p^e-1 má v $\mathbb{Z}^*_{p^e}$ vždy řád 2. Je-li $p=2$, tak $2^{e-1}-1$ a $2^{e-1}+1$ jsou také prvky řádu 2.*

Důkaz. Jistě $P \subseteq \mathbb{Z}^*_{p^e}$. Protože $(1+ap)(1+bp) = 1+(a+b+p)p$, tak P je podgrupa $\mathbb{Z}^*_{p^e}$. Zřejmě $(p^e-1)^2 \equiv 1 \mod p^e$ a $\left(2^{e-1}\pm 1\right)^2 = 2^{2(e-1)} \pm 2^e + 1 \equiv 1 \mod 2^e$. □

Tvrzení. *Je-li p liché prvočíslo, je $\mathbb{Z}^*_{p^e}$ cyklická grupa řádu $(p-1)p^{e-1}$ pro každé $e \geq 1$. Pro $e \geq 3$ je $\mathbb{Z}^*_{2^e} \cong \mathbb{Z}_{2^{e-2}} \times \mathbb{Z}_2$.*

Důkaz. Nejprve budeme řešit sudý případ. Podobnou úvahou jako v předchozím případě nahlédneme, že $\{1+4a; 0 \leq a < 2^{e-2}\}$ je podgrupa $\mathbb{Z}^*_{2^e}$ řádu 2^{e-2}. Tato podgrupa neobsahuje prvek 2^e-1 (neboť ten je $\equiv 3 \mod 4$), ale obsahuje prvek 5. Podle tvrzení 1.9 je $5^{2^{e-3}} \not\equiv 1 \mod 2^e$, a proto musí být 5 řádu 2^{e-2}, takže jde o podgrupu cyklickou.

Každý prvek $\mathbb{Z}^*_{2^e}$ je modulo 2^e tvaru $4a+1$ nebo $-(4a+1)$, $0 \leq a < 2^{e-2}$ (Prvky prvého typu jsou $\equiv 1 \mod 4$ a prvky druhého typu jsou $\equiv 3 \mod 4$). Proto lze každý vyjádřit modulo 2^e jednoznačně jako $5^i(2^e-1)^\varepsilon$, kde $\varepsilon \in \{0,1\}$ a $0 \leq i < 2^{e-2}$. Odsud již přímo plyne požadovaný izomorfismus.

Uvažme nyní situaci lichého prvočísla p. Grupa P popsaná v lemmatu je řádu p^{e-1} a obsahuje prvek $1+p$. Z tvrzení 2.9 plyne, že $1+p$ není v $\mathbb{Z}^*_{p^e}$ řádu p^{e-2}, takže musí být řádu p^{e-1}. Vidíme, že P je grupa cyklická. Nyní ozřejmíme, že k dokončení důkazu stačí nalézt $u \in \mathbb{Z}^*_{p^e}$ řádu $p-1$. Předpokládejme, že jsme takový prvek našli. Pak žádná z mocnin u^i, $1 \leq i < p-1$, nepadne do P, neboť není řádu mocniny p, takže každý prvek $\mathbb{Z}^*_{p^e}$ lze jednoznačně vyjádřit jako $(p+1)^j u^i$, kde $0 \leq j < p^{e-1}$ a $0 \leq i < p-1$. Proto $\mathbb{Z}^*_{p^e} \cong \mathbb{Z}_{p^{e-1}} \times \mathbb{Z}_{p-1}$. Podle lemmatu 2.7 je $\mathbb{Z}_{p^{e-1}} \times \mathbb{Z}_{p-1} \cong \mathbb{Z}_{(p-1)p^{e-1}}$.

K nalezení prvku řádu $p-1$ použijeme okruhový homomorfismus $\alpha : \mathbb{Z}_{p^e} \to \mathbb{Z}_p$, $a \mapsto (a)_p$ z lemmatu 2.7. Ať $v \in \mathbb{Z}_p$ je nějaký primitivní prvek (viz oddíl 2.6). Pak v je v $\mathbb{Z}^*_p$ řádu $p-1$. Zvolme $w \in \mathbb{Z}_{p^e}$, aby $\alpha(w) = v$. Protože v je

nesoudělné s p, musí být i w nesoudělné s p, takže $w \in \mathbb{Z}_{p^e}^*$. Je-li k řád w v $\mathbb{Z}_{p^e}^*$, máme $1 = \alpha\left(w^k\right) = (\alpha(w))^k = v^k$. Proto musí $p-1$ dělit k. Vhodná mocnina w je tudíž prvek řádu $p-1$ (to je zřejmé, lze použít například důsledek 2.3). □

Za poznámku stojí, že pro $e \geq 3$ skutečně $\mathbb{Z}_{2^e}^*$ není cyklická grupa. To lze nahlédnout mnoha způsoby. Jedním z nich je pozorování, že cyklická grupa řádu 2^e má jediný prvek řádu 2. Podle lemmatu je však v $\mathbb{Z}_{2^e}^*$ takových prvků více (jsou přesně 3).

2.11 Fermatova a Wilsonova věta. Carmichaelova čísla.

V libovolné grupě G je $\left\{a^i; i \in \mathbb{Z}\right\}$ cyklickou podgrupou generovanou prvkem a. Její řád se nazývá *řádem prvku* a a značí se $|a|$. Je-li G konečná, tak podle Lagrangeovy věty (viz oddíl 1.5) platí, že $|a|$ dělí $|G|$. Je-li $m = |a|$, pak $a^m = 1$. Tedy $a^k = 1$ pro každý násobek m. Specielně $a^{|G|} = 1$.

Grupa $\mathbb{Z}_n^*$ má řád $\varphi(n)$. Proto platí

Lemma. *$a^{\varphi(n)} \equiv 1 \mod n$ pro každé $a \in \mathbb{Z}$ nesoudělné s n.* □

Pro p prvočíslo má lemma tvar $a^{p-1} \equiv 1 \mod p$, pokud p nedělí a. Někdy je výhodné tento vztah psát

$$a^p \equiv a \mod p.$$

To je známé jako *Malá Fermatova věta*. Všimněme si, že tento vztah platí pro všechna $a \in \mathbb{Z}$.

Jestliže máme dáno celé kladné číslo N, o kterém chceme rozhodnout, zda je, či není prvočíslo, tak se nabízí různé metody. Je-li N sudé, je rozhodování snadné, takže budeme předpokládat, že N je liché a větší než 1.

Nalézt největší společný dělitel dvou čísel je výpočetně poměrně snadné (postupuje se Eukleidovým algoritmem – ten je v tomto textu popsán v oddíle 2.10). Stejně tak je výpočetně snadné pro každé $a \in \mathbb{Z}$ spočítat mocniny a^k modulo N. K jejich výpočtu nepotřebujeme vědět, zda N je či není prvočíslo.

Zvolme tedy náhodně nějaké kladné celé $a < N$, $a \neq 1$. Spočítejme a^{N-1} modulo N. Pokud neplatí $a^{N-1} \equiv 1 \mod N$, tak N nemůže být prvočíslo. To je pozorování zásadního významu, neboť nabízí metodu, jak zjistit, že dané číslo není prvočíslo, aniž bychom museli nalézt nějakého jeho dělitele.

Je ovšem otázka, zda metoda výběru náhodného a vede k cíli dostatečně efektivně. Spokojíme se přitom s pravděpodobnostní odpovědí založenou na opakované volbě a. Pokud je N složené a počet $a < N$ takových, že $a^{N-1} \equiv 1 \mod N$ je méně než N/k, tak pravděpodobnost, že po t krocích budeme nacházet pouze taková a, je rovno $1/k^t$, což se (například) pro $k \geq 2$ s rostoucím t blíží k nule velmi rychle. Pokud by existovalo k takové, že ho lze použít pro všechna složená lichá N, vedla by metoda náhodného výběru a k efektivnímu pravděpodobnostnímu testu prvočíselnosti (hovoří se o *Fermatově testu*).

Ovšem takové k neexistuje. Existuje však vylepšená varianta Fermatova testu, tzv. Rabin-Millerův test, který probíhá podobně a kde lze dokázat, že hodnota $k = 4$ má univerzální platnost.

Důvodem, proč pro Fermatův test nelze univerzálně platná k nalézt, je existence tzv. *Carmichaelových čísel*, což jsou lichá složená čísla N taková, že

$a^{N-1} = 1 \mod N$ kdykoliv a je s N nesoudělné. Existuje nekonečně mnoho Carmichaelových čísel, které jsou součinem tří prvočísel p_1, p_2, p_3. Nejmenším Carmichaelovým číslem je $561 = 3 \cdot 11 \cdot 17$.

V tomto textu nebudeme otázky výpočetní složitosti pojednávat rigorózním způsobem. V podstatě platí, že v modulární aritmetice jsou rychlé postupy, které vedle základních aritmetických operací (hlavně sčítání a odčítání) využívají pouze hledání největších společných dělitelů (eukleidův algoritmus) a umocňování modulo dané prvočíslo.

Existují přitom charakterizace složených čísel (resp. prvočísel), které nejsou založeny na nalezení dělitele, ale přesto se z výpočetního hlediska nezdají být použitelné.

Příkladem takové charakterizace je

Wilsonova věta *Pro každé prvočíslo p je $(p-1)! \equiv -1 \mod p$. Je-li n číslo složené, $n \neq 4$, je $(n-1)! \equiv 0 \mod n$.*

Důkaz. Nenulové prvky $\mathbb{Z}_p$ jsou kořenem polynomu $x^{p-1} - 1$. Proto v $\mathbb{Z}_p$ platí $x^{p-1} - 1 = (x-1)(x-2)\cdots(x-(p-1))$. Srovnáním absolutních členů po roznásobení pravé strany dostáváme $(p-1)! \equiv -1 \mod p$.

Jestliže $n = uv$ pro nějaké u a v taková, že $1 < u < v < n$, tak $n = uv$ dělí $(n-1)!$ a $(n-1)! \equiv 0 \mod n$. Uvedená u a v existují pro každé složené číslo jež není tvaru p^2, p prvočíslo. Ať p je liché. Pak $p \neq p^2 - p$ a $p(p^2-p) \equiv 0 \mod p^2$, a proto je též $(p^2-1)! \equiv 0 \mod p^2$. □

2.12 Míjení involucí

Řekneme, že prvek a grupy G *míjí* prvek $e \in G$, jestliže $a^i \neq e$ a $e^i \neq a$, pro všechna $i \in \mathbb{Z}$. Vidíme, že a míjí e právě když e míjí a.

Pozorování. *Ať $G = A \times B$ je konečná grupa s prvkem (e, f). Jestliže počet prvků $a \in A$, které míjí e, je $\alpha \cdot |A|$ (kde α je racionální), tak počet prvků $g \in G$, které míjí (e, f) je alespoň $\alpha \cdot |G|$.*

Důkaz. Jestliže a míjí e, tak (a, b) míjí (e, f) pro každé $b \in B$. □

Involucí se rozumí každý prvek grupy, který je řádu 2. Je-li e involuce, tak a míjí e právě když $a \neq 1$ a $e \neq a^i$ pro všechna $i \in \mathbb{Z}$.

Obecné lemma. *Buď $G = G_1 \times \cdots \times G_k$ součin grup. Řád prvku $a = (a_1, \ldots, a_k)$ je roven nejmenšímu společnému násobku řádů prvků $a_1, \ldots, a_k$.*

Důkaz. Ať $d_i = |a_i|$, $d = |a|$, a ať n je nejmenší společný násobek $d_1, \ldots, d_k$. Máme $a^n = (a_1^n, \ldots, a_k^n) = (1, \ldots, 1)$, odkud $d|n$. Z $a^d = (1, \ldots, 1)$ plyne, že d_i dělí d pro každé i, $1 \leq i \leq n$, takže $n|d$. □

Důsledek. *Ať p je prvočíslo a ať $k_1 \geq \cdots \geq k_r \geq 1$ jsou čísla celá. Prvek $a = (a_1, \ldots, a_k) \in \mathbb{Z}_{p^{k_1}} \times \cdots \times \mathbb{Z}_{p^{k_r}}$ má řád p^s, kde*

$$s = \max\{k_1 - v_p(a_1), \ldots, k_r - v_p(a_r)\}.$$

Důkaz. Případ $r = 1$ plyne z důsledku 2.4. Zbytek plyne z předchozího obecného lemmatu. □

Následující tvrzení se ukáže jako významná pomůcka při důkazu efektivity Rabin-Millerova testu.

Tvrzení. *Ať $k_1 \geq k_2 \geq \cdots \geq k_r \geq 1$ jsou celá čísla, $r \geq 2$. Položme $e = \left(2^{k_1-1}, \ldots, 2^{k_r-1}\right)$. Pak e je involuce a v aditivní grupě $G = \mathbb{Z}_{2^{k_1}} \times \cdots \times \mathbb{Z}_{2^{k_r}}$. Počet a, které míjí e, je roven alespoň $3|G|/4$, pokud $k_1 \neq k_r$ nebo $r \geq 3$. Je-li $r = 2$ a $k_1 = k_2$, je tento počet alespoň $|G|/2$.*

Důkaz. Předpokládejme, že $ma = (ma_1, \ldots, ma_r)$ je rovno e a ať $m = 2^j s$, kde $j = v_2(m)$. Pro každé i, $1 \leq i \leq r$, je $s2^j a_i = 2^{k_i-1}$ stejného řádu jako $2^j a_i$, neboť s je číslo liché (lze použít například tvrzení 2.4). Proto je $2^j a_i$ involuce. Jelikož $\mathbb{Z}_{2^{k_i}}$ obsahuje jedinou involuci, a to 2^{k_i-1}, musí být $2^j a_i = 2^{k_i-1}$. Je tedy $2^j a = e$. Vidíme, že všechny prvky a_i musí mít stejný řád a to 2^{j+1}. Odvodili jsme kritérium

$$a = (a_1, \ldots, a_r) \text{ míjí } e \Leftrightarrow \text{ existují } 1 \leq i < i' \leq r, \text{ že } |a_i| \neq |a'_i|.$$

Nyní budeme sledovat jednotlivé případy.

Ať nejprve $k = k_1 = k_2 = k_3$ a $r = 3$. Prvek $a \in \mathbb{Z}_{2^k}$ je řádu 2^k právě když a je liché. Každému $a = (a_1, a_2, a_3) \in G$ přiřadíme $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ tak, že $\varepsilon_i = (-1)^{a_i}$. Každá z osmi možných hodnot $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ je obrazem přesně $\left(2^{k-1}\right)^3 = |G|/8$ prvků $a \in G$. Je-li $\varepsilon_i \neq \varepsilon_{i'}$ pro nějaká $i, i' \in \{1, 2, 3\}$, pak z odvozeného kritéria plyne, že a míjí e. Našli jsme tím pádem alespoň $6 \cdot |G|/8 = 3|G|/4$ případů míjení.

Je-li $k = k_1 = k_2$ a $r = 2$, lze postupovat obdobně a získá se $2 \cdot |G|/4 = |G|/2$ případů míjení.

Ať je nyní $r = 2$ a $k_1 > k_2$. Podle výše uvedeného důsledku je v G právě $2^{k_1-1} \cdot 2^{k_2} = |G|/2$ prvků $(a_1, a_2) \in G$, které jsou řádu 2^{k_1}. Protože a_2 je řádu nanejvýš $2^{k_2} < 2^{k_1}$, míjí taková (a_1, a_2) prvek e.

Je-li navíc $k_1 > k_2 + 1$, lze k nim připojit ze stejných důvodů dvojice (a_1, a_2), kde a_1 je řádu 2^{k_1-1}. Těchto dvojic je $2^{k_1-2} \cdot 2^{k_2} = |G|/4$, a dohromady získáváme kýžené $3|G|/4$.

Ať je $k_1 = k_2 + 1$. Pak uvážíme dvojice (a_1, a_2), kde jeden z prvků je řádu 2^{k_2} a druhý je řádu menšího. Takových dvojic je $2 \cdot \left(2^{k_2-1} \cdot 2^{k_2-1}\right) = 2^{2k_2-1} = |G|/4$, takže $3|G|/4$ je opět dosaženo.

Zbylé případy lze z odvozených získat ze vstupního pozorování tohoto oddílu. □

2.13 Rabin-Millerův test

Úvodní lemma patří do skupiny přípravných tvrzení potřebných pro důkaz správnosti testu (jeho popis následuje hned za lemmatem).

Lemma. *Ať p_1 a p_2 jsou dvě různá lichá prvočísla taková, že $v_2(p_1 - 1) = v_2(p_2 - 1)$. Označme tuto hodnotu k a definujme $m_1 = (p_1 - 1)2^{-k}$, $m_2 = (p_2 - 1)2^{-k}$, $e = v_2(p_1 p_2 - 1)$, $m = (p_1 p_2 - 1)2^{-e}$. Potom $e > k$ a alespoň jedna z hodnot m_1 a m_2 nedělí m.*

Důkaz. Vyjdeme z rovnosti

$$p_1p_2 - 1 = (p_1 - 1)(p_2 - 1) + (p_1 - 1) + (p_2 - 1).$$

Tuto rovnost je možné zapsat též jako

$$2^e m = 2^k(2^k m_1 m_2 + m_1 + m_2).$$

Odsud ihned plyne $k < e$, neboť číslo $2^k m_1 m_2 + m_1 m_2$ je sudé. Předpokládejme, že $p_1 - 1$ i $p_2 - 1$ dělí $p_1p_2 - 1$. Z prvé rovnosti vidíme, že pak je $p_2 - 1$ dělitelné $p_1 - 1$, a naopak. To však není možné, neboť předpokládáme $p_1 \neq p_2$. Existuje tedy $i \in \{1, 2\}$, že $2^k m_i$ nedělí $2^e m$. Z $k < e$ plyne, že m_i nedělí m. □

Připomeňme, že Fermatův test (viz oddíl 2.10) vychází z poznání chování prvočísel prvočísel a hledání sporu s tímto chováním. Problém je v tom, že existují čísla složená, u kterých je případy onoho sporného chování obtížné nalézt. Rabin-Millerův test vychází z o trochu detailnějšího popisu chování prvočísel. I když jde o drobný rozdíl, je natolik významný, že sporné chování je u všech složených čísel již natolik frekventované, že ho lze odhalit s dostatečně velkou pravděpodobností.

Buď tedy nejprve p liché prvočíslo, $p - 1 = 2^e m$, kde m je liché (takže $e = v_2(p - 1)$). Buď $u \in \mathbb{Z}_p^*$ primitivní prvek (viz oddíl 2.6). Zobrazení $u^i \mapsto i$ je isomorfimus $\mathbb{Z}_p^* \cong \mathbb{Z}_{p-1}$ a $\mathbb{Z}_{p-1} \cong \mathbb{Z}_{2^e} \times \mathbb{Z}_m$ dle tvrzení 2.7. Existuje proto izomorfismus $\alpha : \mathbb{Z}_p^* \cong \mathbb{Z}_m \times \mathbb{Z}_{2^e}$ (jeho konkrétní podoba pro nás nebude důležitá). Ať pro $v \in \mathbb{Z}_p^*$ je $\alpha(v) = (a, b)$. Pak $\alpha(v^m) = (0, c)$, kde $c = mb$, neboť řád $a \in \mathbb{Z}_m$ dělí m. Pokud je prvek c nenulový, tak je řádu 2^j pro nějaké $j \geq 1$. To znamená, že $2^{j-1}c$ je involuce, a ta je v $\mathbb{Z}_{2^e}$ jediná, a to 2^{e-1}. Vidíme dokonce, že $(0, 2^{e-1})$ je jediná involuce v $\mathbb{Z}_m \times \mathbb{Z}_{2^e}$, takže musí být $\alpha(p - 1) = (0, 2^{e-1})$. Je-li $c \neq 0$, je tedy $\alpha(v^{m2^{j-1}}) = p - 1 \equiv -1 \mod p$. Je-li $c = 0$, je $\alpha(v^m) = 0$.

Pro N liché prvočíslo, $N - 1 = 2^e m$, m liché, tedy pro každé kladné $a < N$ platí:

buď $a^m \equiv 1 \mod N$;

nebo $a^{m2^j} \equiv -1 \mod N$ pro nezáporné $j < e$.

Pokud je N složené a kladné $a < N$ splňuje uvedenou podmínku, nazývá se N *silné pseudoprvočíslo v bázi* a. (V tomto oddíle budeme většinou pouze stručně říkat, že a je *báze*. Pro úplnost poznamenejme, že N se nazývá *pseudoprvočíslo v bázi* a, pokud $a^{N-1} \equiv 1 \mod N$.)

Tvrzení. *Buď N liché složené číslo. Pak počet kladných $a < N$ takových, že N je silné pseudoprvočíslo v bázi a, je menší než $N/4$.*

Důkaz. Položme $e = v_2(N - 1)$ a $m = (N - 1)2^{-e}$. Je-li kladné $a < N$ bází, je $a^{2^e m} \equiv 1 \mod N$. Proto $a \in \mathbb{Z}_N$ nemůže být bází, jestliže

(A) a není invertibilní, nebo

(B) $a \in \mathbb{Z}_N^*$ má řád, který nedělí $2^e m$.

Předpokládejme nejprve, že $k = v_p(N) \geq 2$ pro nějaké (nutně liché) prvočíslo p. Položme $s = Np^{-k}$. Připomeňme, že v $\mathbb{Z}_{p^k}$ je p^{k-1} neinvertibilních prvků

a že ze $\mathbb{Z}_{p^k}^* \cong \mathbb{Z}_{p^{k-1}} \times \mathbb{Z}_{p-1}$ (viz tvrzení 2.10 a 2.7) vyplývá existence $(p-1)p^{k-1} - (p-1) = (p-1)(p^{k-1}-1)$ prvků, jejichž řád v $\mathbb{Z}_{p^k}^*$ je dělitelný p. Ze $\mathbb{Z}_{p^k} \times \mathbb{Z}_s \cong \mathbb{Z}_N$ tudíž plyne, že v $\mathbb{Z}_N$ je alespoň $sp^{k-1} + s(p-1)(p^{k-1}-1)$ prvků splňujících podmínku (A) a nebo (B). Na báze tím pádem zbývá pouze $s(p-1)$ možností. Z $p^2 - 4(p-1) = (p-2)^2 > 0$ máme $4(p-1) < p^2 \leq p^k$, takže $s(p-1) < sp^k/4 = N/4$.

Zbývá tedy řešit případ $N = p_1 \dots p_r$ je součin po dvou různých prvočísel p_i, $1 \leq i \leq r$. Budeme dokazovat, že $\mathbb{Z}_N^*$ obsahuje nanejvýš $|\mathbb{Z}_N^*|/4 = \varphi(N)/4$ bází. To stačí, neboť prvky mimo $\mathbb{Z}_N^*$ bázemi nejsou (viz podmínka (A)), takže bází bude nanejvýš $\varphi(N)/4 < N/4$.

Pro každé p_i položíme $k_i = v_2(p_i - 1)$ a $m_i = (p_i - 1)2^{-k_i}$. Z tvrzení 2.7 plyne existence izomorfismu okruhů $\mathbb{Z}_N \cong \mathbb{Z}_{p_1} \times \dots \times \mathbb{Z}_{p_r}$ takového, že obrazem $2^N - 1$ je $(2^{p_1} - 1, \dots, 2^{p_r} - 1)$. Podle tvrzení 2.8 zúžení tohoto izomorfismu na $\mathbb{Z}_N^*$ dává izomorfismus $\mathbb{Z}_N^* \cong \mathbb{Z}_{p_1}^* \times \dots \times \mathbb{Z}_{p_r}^*$. Každé $\mathbb{Z}_{p_i}^*$ je izomorfní $\mathbb{Z}_{p_i - 1} \cong \mathbb{Z}_{2^{k_i}} \times \mathbb{Z}_{m_i}$, podle oddílů 2.6 a 2.7. Vidíme, že $p_i - 1$ je v $\mathbb{Z}_{p_i}^*$ jedinou involucí, a ta odpovídá v $\mathbb{Z}_{2^{k_i}} \times \mathbb{Z}_{m_i}$ hodnotě $(2^{k_i - 1}, 0)$. Položme $M = \mathbb{Z}_{m_1} \times \dots \times \mathbb{Z}_{m_k}$. Z předchozího plyne existence izomorfismu $\alpha : \mathbb{Z}_N^* \cong (\mathbb{Z}_{2^{k_1}} \times \dots \times \mathbb{Z}_{2^{k_r}}) \times M$, kde $\alpha(N-1) = (u, 0)$, $u = (2^{k_1 - 1}, \dots, 2^{k_r - 1})$. Je-li $(v, c) = \alpha(a)$ takové, že v míjí u (ve smyslu oddílu 2.12), tak lze snadno nahlédnout, že a není bází. Vskutku, žádná mocnina a nemůže být rovna -1, neboť žádná mocnina v není rovna u. Rovněž a nemůže být lichého řádu, neboť pak by bylo $v = 1$.

Z tvrzení 2.11 vyplývá, že dvojic (v, c), kde v míjí u, je alespoň $3|G| \cdot |M|/4 = 3|\mathbb{Z}_N^*|/4$, kde $G = \mathbb{Z}_{2^{k_1}} \times \dots \times \mathbb{Z}_{2^{k_r}}$, s výjimkou případu $k = k_1 = k_2$, $r = 2$. Tento případ rozebereme podrobněji.

Máme $\alpha : \mathbb{Z}_N^* \cong \mathbb{Z}_{2^k} \times \mathbb{Z}_{2^k} \times \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$. Je-li $\alpha(a) = (v, c)$, tak a není bází, pokud

(C) v míjí $u = (2^{k-1}, 2^{k-1})$; nebo

(D) řád c v $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$ nedělí m.

Připomeňme, že podmínka je důsledkem toho, že $a^{2^e m} \equiv 1 \mod N$ pro každou bázi a (viz začátek tohoto důkazu).

Podle úvodního lemmatu tohoto oddílu existuje $i \in \{1, 2\}$ takové, že m_i nedělí m. Prvky $c \in \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2}$ exponentu m (tj. takové, že jejich řád dělí m) tvoří tudíž vlastní podgrupu grupy $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} = M$. Jde o grupu lichého řádu, takže index této podgrupy je alespoň 3 a mimo ní leží alespoň $2/3$ prvků. Podmínky (C) a (D) jsou na sobě nezávislé. Podle tvrzení 2.11 nejvýše $|G|/2$ prvků $v \in G$ nesplňuje podmínku (C). Nahlédli jsme, že nejvýše $|M|/3$ prvků $c \in M$ nesplňuje podmínku (D) (tedy řád c dělí m). Žádnou z podmínek (C) a (D) tedy nesplňuje nanejvýš

$$(|G|/2) \cdot (|M|/3) = |\mathbb{Z}_N^*|/6 < |\mathbb{Z}_N^*|/4$$

prvků $\mathbb{Z}_N^*$. □

2.14 Idea metody RSA

Grupa G je exponentu m, pokud $x^m = 1$ pro každé $x \in G$ (viz oddíl 1.7). V grupě exponentu m platí $x^{m+1} = x$ pro každé $x \in G$. Proto se zdá být přirozené

říci, že pologrupa S je exponentu m, jestliže $x^{m+1} = x$ pro každé $x \in S$. Pak $x^{km+1} = x$ pro každé $k \geq 1$.

Předpokládejme, že máme k disposici nějakou pologrupu S exponentu m, jejíž prvky budeme používat pro kódování zprávy. Předpokládejme dále, že existuje jakási Božena, která by ráda umožnila svým ctitelům zasílání zpráv tak, aby je mohla číst pouze ona sama. Božena použije pologrupu S a dvojici kladných čísel (d, e) takovou, že $de \equiv 1 \mod m$. Pologrupu S a číslo e zveřejní. Dvojice (S, e) je tedy *veřejným klíčem*.

Chce-li jí nyní Alois zaslat zprávu vyjádřenou posloupností $x_1 \dots x_n$, pošle posloupnost $x_1^e x_2^e \dots x_k^e$. Božena obdrží posloupnost $y_1 \dots y_n$, ze které chce $x_1 \dots x_k$ dekódovat. Protože $ed = km+1$ pro nějaké $k \geq 1$, je $y_i^d = x_i^{km+1} = x_i$. Boženě tedy stačí každý z prvků x_i umocnit na d. (Písmena e,d jsou volena tak, aby e připomínalo encryption a d decryption). Čísla m a d tvoří *soukromý klíč*.

Všimněte si, že Božena nezveřejňuje číslo m. To zůstává jejím tajemstvím. Aby její metoda mohla mít úspěch, musí být S takovou pologrupou, že je *obtížné zjistit její exponent*.

Níže uvidíme, že teorie čísel takové pologrupy vskutku nabízí. Než je popíšeme, zmíníme několik obecných fakt.

Ne všechny prvky S se musí k enkryptaci používat. Je-li x *idempotent* ($x^i = x$ pro každé $i \geq 1$), tak určitě není k zápisu zpráv vhodný.

Celou proceduru lze také obrátit. Dejme tomu, že Božena chce zprávou $x_1 \dots x_k$ oznámit, se kterým ctitelem půjde do kina, a chce, aby to všichni věděli. Vystavením zprávy $y_1 \dots y_k$, kde $y_i = x_i^d$, Božena nejen Aloisovi umožní, aby ze vztahu $x_i = y_i^e$ zjistil text zprávy $x_1 \dots x_k$, ale také mu poskytuje informaci, že původcem zprávy je ten, kdo zná soukromý klíč d, tedy Božena.

Tento postup je koncepčně shodný s tzv. elektronickým podpisem dokumentů. Ať je text daného dokumentu i jeho otisk $h_1 \dots h_k$ získaný nějakou hashovací funkcí známý obou stranám. Ze znalosti veřejného klíče e lze pak dovodit, že ten, kdo zveřejnil $h_1^d \dots h_k^d$, znal jak dokument, tak soukromý klíč d.

V praxi ovšem může být problémem, jak se ujistit, že ten, kdo vyhlásil veřejný klíč e, skutečně je Božena, a ne někdo, kdo se za ni vydává. K tomu slouží tzv. certifikační autority, jejichž veřejný klíč se již předpokládá být nezpochybnitelný, a se kterými se komunikuje v podstatě stejnými prostředky, jaké jsme popsali.

Z praktických důvodů nebývá vhodné používat tentýž soukromý klíč jak pro ověřování identity (podpisu), tak pro zasílání kódovaných zpráv.

Lemma. *Ať $p_1, \dots, p_r$ jsou po dvou různá lichá prvočísla. Označme S multiplikativní monoid okruhu $\mathbb{Z}_{p_1 \cdots p_r}$. Nejmenší možný exponent monoidu S je roven nejmenšímu společnému násobku čísel $p_1 - 1, \dots, p_r - 1$.*

Důkaz. Označme m uvažovaný nejmenší společný násobek. Podle 2.7 je okruh $\mathbb{Z}_{p_1,\dots,p_r}$ izomorfní okruhu $\mathbb{Z}_{p_1} \times \cdots \times \mathbb{Z}_{p_r}$. Uvažme prvek tohoto okruhu $a = (a_1, \dots, a_r)$. Podle Malé Fermatovy věty (viz 2.11) je $a_i^{p_i} \equiv a_i \mod p_i$ pro každé i, $1 \leq i \leq r$. Současně existují k_i, že $m = k_i(p_i - 1)$, takže také $a_i^{m+1} \equiv a_i \mod p_i$. Proto $a^{m+1} = a$. Minimalita m plyne například z obecného lemmatu 2.12. □

Za S lze tedy zvolit monoid $\mathbb{Z}_{p_1 \cdots p_r}$. O systému (metodě) RSA hovoříme, je-li $r = 2$. Pro $r \geq 3$ se někdy používá označení *multi-RSA*.

Aby bylo možno v S počítat, je třeba zveřejnit hodnotu $n = p_1 p_2$ (tak zvaný modul). Dvojice (n, e) tvoří *veřejný klíč*. Číslu d se říká *tajný exponent*. Hodnotu m (což je minimální hodnota exponentu pologrupy) lze ze znalosti p_1 a p_2 snadno odvodit, dle lemmatu výše. Nelze ho však snadno odvodit z čísla n. To je totiž úkol, který odpovídá nalezení rozkladu n na prvočísla.

Prvočísla p_1 a p_2 není totiž obtížné pomocí Rabin-Millerova testu nacházet, ale pro faktorizaci jejich součinu žádný obdobně účinný algoritmus k dispozici není.

Asymetrie RSA spočívá v tom, že ověřit, zda dané číslo je složené, je algoritmicky daleko snazší, nežli ho skutečně rozložit.

Kapitola 3

Čtverce, charaktery a reciprocita

3.1 Gaussova celá čísla

Čísla tvaru a^2, $a > 1$, nazveme *čtverce*. Čísla $n > 1$ *nedělitelná čtvercem* jsou tedy ta, která mají prvočíselný rozklad $n = p_1 \dots p_k$, kde $p_1 < \cdots < p_k$ jsou prvočísla, $k \geq 1$. Je-li $d > 1$ číslo nedělitelné čtvercem, tak

$$\mathbb{Z}\left[\sqrt{d}\right] = \left\{a + b\sqrt{d};\ a, b \in \mathbb{Z}\right\} \quad \text{i} \quad \mathbb{Z}\left[\sqrt{-d}\right] = \left\{a + ib\sqrt{d};\ a, b \in \mathbb{Z}\right\}$$

tvoří okruhy, které mají v teorii čísel značný význam. Každý prvek takového okruhu určuje dvojici (a, b) jednoznačně. Zobrazení $a + b\sqrt{d} \mapsto |a^2 - db^2|$ a $a + b\sqrt{-d} \mapsto a^2 + db^2$ se nazývá *norma*. V několika případech, kdy d je malé, je tato norma eukleidovským zobrazením ve smyslu oddílu 1.16. Takový je i případ okruhu $\mathbb{Z}\left[\sqrt{-1}\right] = \mathbb{Z}[i]$, kterému se říká okruh *Gaussových celých čísel*. Znalost struktury tohoto okruhu umožňuje rychlý důkaz některých základních fakt teorie čísel. Navíc jde o netriviální příklady využití teorie dělitelnosti.

V případě Gaussových celých čísel je norma shodná s čtvercem absolutní hodnoty příslušného komplexního čísla. Gaussova celá čísla jsou vlastně ty body komplexní roviny, které mají celočíselné souřadnice, a norma je rovna dvojmocí vzdálenosti od počátku. Norma je vždy celočíselná, což o absolutní hodnotě platit nemusí.

Tvrzení. *Norma $a + bi \mapsto a^2 + b^2$ je v okruhu Gaussových celých čísel eukleidovské zobrazení.*

Důkaz. Pro $\alpha = a + bi \in \mathbb{Z}[i]$ položme $N(\alpha) = |\alpha|^2 = a^2 + b^2$. Ať ještě $\gamma = c + di \in \mathbb{Z}[i]$, $\gamma \neq 0$. Jestliže $\alpha = \beta\gamma$ pro nějaké $\beta \in \mathbb{Z}[i]$, tak z $N(\alpha) = N(\beta)N(\gamma)$ plyne $N(\gamma) \leq N(\alpha)$, neboť $N(\beta) \geq 1$. Chceme dokázat, že vždy existují $\beta, \eta \in \mathbb{Z}[i]$, jež splňují $\alpha = \beta\gamma + \eta$, $N(\eta) < N(\gamma)$. Volba η nemusí být jednoznačná, tak, jak tomu není již v případě celých čísel, kde za β volíme celočíselnou aproximaci α/γ. Podobně postupujeme i zde – uvážíme α/γ jako bod komplexní roviny a za β zvolíme jedno z (nejvýše čtyř) Gaussových celých čísel takových, že $|\beta - \alpha/\gamma| < \sqrt{2}/2$. Pak $|\beta\gamma - \alpha| < |\gamma|\sqrt{2}/2$ a $N(\alpha - \beta\gamma) < N(\gamma)/2$. Stačí tedy položit $\eta = \alpha - \beta\gamma$. □

Důsledek. *Okruh $\mathbb{Z}[i]$ je oborem hlavních ideálů a jeho invertibilními prvky jsou 1, -1, i a $-i$.*

Důkaz. Použij tvrzení i lemma oddílu 1.16. □

Přirozenou je nyní otázka, jak vypadají prvočinitelé okruhu $\mathbb{Z}[i]$ (viz oddíl 1.14). Prvky lišící se o násobek invertibilním prvkem generují stejný hlavní ideál, takže $\alpha = a + bi$ je prvočinitel právě když je prvočinitel kterýkoliv z prvků $-a - bi$, $-b + ia$ a $b - ia$. V $\mathbb{Z}[i]$ lze navíc uplatnit konjugaci $\overline{\alpha} = a - bi$. Máme $N(\alpha) = N(\overline{\alpha})$ a $\alpha = \beta\gamma \Leftrightarrow \overline{\alpha} = \overline{\beta}\overline{\gamma}$. Proto je α prvočinitel právě když $\overline{\alpha}$ je prvočinitel. (Je dobré si uvědomit, že $\alpha \mapsto \overline{\alpha}$ je automorfismus okruhu $\mathbb{Z}[i]$.)

Pokud $\alpha \in \mathbb{Z}[i]$ je prvočinitel, tak $N(\alpha) = \alpha\overline{\alpha}$ je v $\mathbb{Z}[i]$ rozkladem $N(\alpha)$ na prvočinitele, a proto $N(\alpha)$ nelze zapsat – až na násobení invertibilními prvky – v $\mathbb{Z}[i]$ jako součin dvou vlastních dělitelů. To znamená, že $N(\alpha)$ nemá dva různé vlastní kladné celočíselné dělitele, takže pro nějaké prvočíslo p je buď $N(\alpha) = p$ nebo $N(\alpha) = p^2$.

Případ $N(\alpha) = p$ značí $p = a^2 + b^2$. Pro $p = c^2 - d^2$ naopak máme $p = (c + di)(c - di)$, přičemž $c + di$ nemůže mít v $\mathbb{Z}[i]$ vlastního dělitele (norma takového dělitele by totiž byla vlastním dělitelem p). Můžeme proto vyslovit

Lemma. *Jestliže $p \in \mathbb{P}$ nelze vyjádřit jako součet dvou čtverců, je $\pm p$ prvočinitelem $\mathbb{Z}[i]$. Jestliže $p = a^2 + b^2 \in \mathbb{P}$ pro $a, b \in \mathbb{Z}$, je $a + bi$ prvočinitelem $\mathbb{Z}[i]$. Všechny prvočinitele $\mathbb{Z}[i]$ jsou jednoho z uvedených dvou tvarů.* □

Protože $2 = 1 + 1 = 1^2 + 1^2$, je možno omezit otázku, která prvočísla jsou součtem dvou čtverců, pouze na prvočísla lichá. Každý čtverec je $\equiv 0, 1 \mod 4$, a proto $p \in \mathbb{P}$ nemůže být součtem dvou čtverců, je-li $p \equiv 3 \mod 4$.

Buď nyní $p \equiv 1 \mod 4$. Ukážeme, že p lze jako součet dvou čtverců vždy vyjádřit. Naše metoda však nebude zcela konstruktivní, protože nalezneme a, b taková, že p dělí $a^2 + b^2$, přičemž a,b jsou nesoudělná kladná. Kdyby p bylo v $\mathbb{Z}[i]$ prvočinitel, tak by bylo $p|a+bi$ nebo $p|a-bi$, což obé implikuje p jako společného dělitele a a b. Zvolíme $b = 1$ a $a = \left(\frac{p-1}{2}\right)!$. Číslo a je sudé a modulo p máme $a^2 = 1(-1)2(-2)\ldots\left(\frac{p-1}{2}\right)\left(-\frac{p-1}{2}\right) \equiv 1(p-1)2(p-2)\ldots\left(\frac{p-1}{2}\right)\left(p-\frac{p-1}{2}\right) \equiv (p-1)!$

Podle Wilsonovy věty je $(p-1)! \equiv -1 \mod p$, takže p vskutku dělí $a^2 + 1$.

Uvědomme si, že rozklad prvočísla $p = a^2 + b^2$, kde a, b jsou kladná celá, je jediný možný. Je-li totiž $p = c^2 + d^2$, tak máme $p = (a + bi)(a - bi) = (c + di)(c - di)$ a stačí použít jednoznačnost rozkladu p na prvočinitele (který platí až na násobky invertibilními prvky).

Závěr. *Buď p liché prvočíslo. Pak je ekvivalentní*

(i) *p je tvaru $4k + 1$;*

(ii) *p není prvočinitel okruhu $\mathbb{Z}[i]$;*

(iii) *$p = a^2 + b^2$ pro nějaká $a, b \in \mathbb{Z}$;*

(iv) *existuje (až na pořadí) jediná dvojice (a, b) přirozených čísel taková, že $p = a^2 + b^2$.*

3.2 Kvadratická rezidua

Uvažme prvočíslo p a číslo a, které je s ním nesoudělné. Řekneme, že a je *kvadratický zbytek (reziduum)* modulo p, jestliže $a \equiv b^2 \mod p$ pro nějaké $b \in \mathbb{Z}$. Tato skutečnost se symbolicky zapisuje vztahem $\left(\frac{a}{p}\right) = 1$. Pokud uvažované b neexistuje, píšeme $\left(\frac{a}{p}\right) = -1$. Je-li a dělitelné p, tak $\left(\frac{a}{p}\right) = 0$.

Označení $\left(\frac{a}{p}\right)$ se říká *Legendreův symbol*. Je zřejmé, že hodnotu $\left(\frac{a}{p}\right)$ určuje hodnota a modulo p, takže při jeho určení lze pracovat v $\mathbb{Z}_p$. Zjevně $\left(\frac{1}{p}\right) = 1$ a $\left(\frac{0}{p}\right) = 0$ pro každé prvočíslo p. Tím jsou dány všechny možné hodnoty pro $p = 2$ a můžeme předpokládat, že p je liché. Víme, že $\mathbb{Z}_p^*$ je cyklická grupa řádu $p-1$, což je podle našeho předpokladu sudé číslo. Proto $\{a^2;\ a \in \mathbb{Z}_p^*\}$ tvoří podgrupu $\mathbb{Z}_p^*$ řádu $\frac{p-1}{2}$ a indexu 2 (viz oddíl 2.3). Prvky této grupy jsou právě ty, jejichž řád dělí $\frac{p-1}{2}$, takže $b \in \mathbb{Z}_p^*$ do ní padne právě když $b^{\frac{p-1}{2}} = 1$. Ovšem $\left\{a^{\frac{p-1}{2}};\ a \in \mathbb{Z}_p^*\right\}$ je také podgrupa $\mathbb{Z}_p^*$, a to řádu 2. Sestává se tedy z prvků 1 a $p-1$. Pro a nesoudělná s p jsme tudíž dokázali

$$a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) \mod p,$$

což zjevně platí i pro $a \equiv 0 \mod p$.

Z tohoto vztahu okamžitě máme i vztah

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \quad \text{pro všechna } a, b \in \mathbb{Z}$$

(vztah je natolik významný, že lze doporučit také samostatné provedení alternativního přímočarého důkazu).

Je-li $\left(\frac{a}{p}\right) \in \{0, 1\}$, říkáme též, že a je *čtverec* modulo p.

Tvrzení. *Buď p liché prvočíslo. Pak*

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} \quad \text{a} \quad \left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$$

Jinými slovy, -1 je čtverec modulo p právě když $p \equiv 1 \mod 4$ a 2 je čtverec modulo p právě když $p \equiv \pm 1 \mod 8$.

Důkaz. Případ -1 je pouze dosazení do vztahu $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \mod p$. V druhém případě budeme pracovat modulo hlavní ideál $p\mathbb{Z}[i]$ okruhu $\mathbb{Z}[i]$. Prvočíslo p dělí každé $\binom{p}{j}$, $1 \le j < p$ (viz důsledek 2.9), takže z binomické věty máme

$$(1+i)^p \equiv 1 + i^p \mod p\mathbb{Z}[i]$$

.

Současně $(1+i)^p = (1+i)((1+i)^2)^{(p-1)/2} = (1+i)(2i)^{(p-1)/2}$. Pro $p \equiv 1 \mod 4$ je $i^p = i$ a $i^{(p-1)/2} = (-1)^{(p-1)/4}$, takže dostáváme $1 + i \equiv (1+i)\left(\frac{2}{p}\right)(-1)^{(p-1)/4} \mod p\mathbb{Z}[i]$, což po vynásobení $1 - i$ dává celočíselný vztah

$2 \equiv 2\left(\frac{p}{2}\right)(-1)^{(p-1)/4} \mod p$, odkud $\left(\frac{2}{p}\right) = (-1)^{(p-1)/4}$. Tím je případ $p \equiv 1 \mod 4$ vyřešen.

Pro $p \equiv -1 \mod 4$ je $i^p = -i$ a $i^{(p-1)/2} = -i \cdot (-1)^{(p+1)/4}$, takže $1 - i \equiv -i(1+i)\left(\frac{2}{p}\right)(-1)^{(p+1)/4} \mod p\mathbb{Z}[i]$. Protože $-i(1+i) = 1-i$, tak vynásobením $1+i$ obdržíme $2 \equiv 2\left(\frac{2}{p}\right)(-1)^{(p+1)/4} \mod p$, odkud $\left(\frac{2}{p}\right) = (-1)^{(p+1)/4}$. □

Vyjádření $\left(\frac{2}{p}\right)$ lze odvodit samozřejmě i jinými, elementárnějšími prostředky. Uvedený způsob se však zdá být nejkratší.

3.3 Diofantické rovnice

Ať $f = f(x_1, \ldots, x_n)$ je polynom s celočíselnými koeficienty. Každá rovnice tvaru $f(x_1, \ldots, x_n) = 0$ se nazývá *diofantická*. Je-li $f(r_1, \ldots, r_n) = 0$ pro nějaká racionální čísla $r_1, \ldots, r_n$, nazývá se $(r_1, \ldots, r_n)$ *racionální řešení* této rovnice. Pokud $r_1, \ldots, r_n \in \mathbb{Z}$, hovoříme o *celočíselném řešení*. Zde budeme řešením diofantické rovnice rozumět pouze řešení celočíselná.

Teorie diofantických rovnic je velmi rozsáhlá a má v matematice a teoretické informatice velký význam. Naším cílem zde je pouze seznámení se s pojmem a ukázka dalšího použití Gaussových celých čísel.

Diofantická rovnice tvaru

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = m$$

se nazývá *lineární*. Všechna čísla tvaru $a_1k_1 + \cdots + a_nk_n$, kde za $k_1, \ldots, k_n$ bereme celá čísla, tvoří ideál $d\mathbb{Z}$ okruhu $\mathbb{Z}$, přičemž d je největším společným dělitelem čísel $a_1, \ldots, a_n$ (viz důsledek 1.15). Lineární diofantická rovnice má tedy (alespoň jedno) řešení právě když d dělí m.

Prvá část následujícího tvrzení se zabývá existencí takzvaných *pythagorejských trojic*, čili trojic přirozených čísel, které mohou být velikostmi stran pravoúhlých trojúhelníků. Důkaz, který uvedeme, má opět delší elementární variantu.

Tvrzení.

(i) Všechna řešení diofantické rovnice $x^2 + y^2 = z^2$ mají (po případné záměně x a y) tvar $x = (a^2 - b^2)c$, $y = 2abc$ a $z = (a^2 + b^2)c$, kde a, b a c jsou libovolná celá čísla.

(ii) Diofantické rovnice $y^2 + 1 = x^3$ má jediné řešení $(x, y) = (1, 0)$.

Důkaz. Pokud $x^2 + y^2 = z^2$ a některé z čísel x, y, z je nulové, lze jistě najít vhodná a, b a c. Můžeme se tedy zabývat pouze případem, kdy x, y a z jsou nenulová čísla. Dále je zřejmé, že pokud d dělí x i y, tak d dělí i z. S ohledem na parametr řešení c se lze tudíž omezit pouze na situaci, kdy x a y jsou nesoudělná. Pokud by x i y bylo liché, máme $x^2 + y^2 \equiv 2 \mod 4$, avšak $z^2 \not\equiv 2 \mod 4$ pro žádné $z \in \mathbb{Z}$. Proto můžeme předpokládat, že y je sudé.

Buď $\alpha \in \mathbb{Z}[i]$ prvočinitel dělicí jak $x + iy$, tak $x - iy$. Pak α dělí $2x$ i $2y$, takže norma α (což je prvočíslo, nebo čtverec prvočísla) dělí $4x^2$ i $4y^2$. Jedinou možností zjevně je $\alpha = 1 + i$ (po případné záměně α jeho invertibilním

násobkem). Pak ovšem $\alpha^2 = 2i$ dělí z^2, což nelze, neboť z je liché. Vidíme, že $x + iy$ je nesoudělné s $x - iy$.

Máme $z > 2$, takže z není v $\mathbb{Z}[i]$ invertibilní. Uvažme vyjádření $z = \alpha_1^{e_1} \dots \alpha_k^{e_k}$ jako součin prvočinitelů $\alpha_1, \dots, \alpha_k \in \mathbb{Z}[i]$. Protože

$$x^2 + y^2 = (x + iy)(x - iy) = z^2 = \alpha_1^{2e_1} \dots \alpha_k^{2e_k},$$

tak z nesoudělnosti $x+iy$ s $x-iy$ vyplývá, že $x+iy = u\beta^2$ pro nějaké $\beta = a+bi$ a pro $u \in \mathbb{Z}[i]$ invertibilní, tedy $u \in \{1, -1, i, -i\}$. Protože $-1 = i^2$, stačí uvažovat pouze případy $u \in \{1, i\}$. Je-li $u = i$, dostaneme $x + iy = i(a + bi)^2 = i(a^2 - b^2 + 2abi) = -2ab + i(a^2 - b^2)$, odkud $x = -2ab$. Předpokládáme, že x je liché, a proto zbývá jedině volba $u = 1$.

Pak $x + iy = (a + bi)^2 = (a^2 - b^2) + 2abi$, $x = a^2 - b^2$, $y = 2ab$ a $z = a^2 + b^2$.

Uvažme nyní diofantickou rovnici $y^2 + 1 = x^3$. Čísla x a y nemohou být současně sudá či současně lichá. Je-li x sudé, musí být $y^2 \equiv 3 \mod 4$, což nelze. Každé řešení tedy má x liché a y sudé.

Je-li $\alpha \in \mathbb{Z}[i]$ prvočinitel dělicí $y + i$ a $y - i$, dělí α také $2i$, takže 2 dělí x^3. Vidíme, že $y + i$ je s $y - i$ nesoudělné. Každý invertibilní prvek $\mathbb{Z}[i]$ lze vyjádřit jako třetí mocninu, například $i = (-i)^3$. Proto existují $\beta = a + bi$ a $\gamma = c + di$ takové, že $y + i = \beta^3$ a $y - i = \gamma^3$. Protože $(a + bi)^3 = (a^3 - 3ab^2) + (3a^2b - b^3)i$, musí být $1 = b(3a^2 - b^2)$ pro nějaká celá $a, b \in \mathbb{Z}$. Jistě $|b| = 1$, a z $b = 1$ máme $3a^2 = 2$, což v celých číslech nemá řešení. Volba $b = -1$ implikuje $-1 = 3a^2 - 1$, odkud $a = 0$, $\beta = -i$, $y + i = i$, $y = 0$ a $x = 1$. □

3.4 Charaktery a Gaussovy součty

V teorii čísel mají velký význam konečné multiplikativní podgrupy tělesa komplexních čísel $\mathbb{C}$. Je-li G taková podgrupa a $\alpha \in G$, tak musí být $|\alpha| = 1$, neboť jinak by mocniny α poskytovaly nekonečně mnoho různých absolutních hodnot. Vidíme, že prvky G jsou rozmístěny na jednotkové kružnici. je-li $\alpha^m = 1$, patří α mezi *m-té odmocniny z jedné*. Ty na jednotkové kružnici vytvářejí pravidelný m-úhelník.

Je-li α řádu m, mluvíme o *primitivní* m-té odmocnině z jedné. Klademe $\zeta_m = \cos(2\pi/m) + i \sin(2\pi/m) = e^{2\pi i/m}$. Multiplikativní cyklická grupa generovaná ζ_m je totožná s grupou všech m-tých odmocnin z jedné.

Multiplikativním (nebo též *Dirichletovým*) *charakterem* modulo n se rozumí každý homomorfismus grup $\chi : \mathbb{Z}_n^* \to \mathbb{C}^*$. Grupa $\mathbb{Z}_n^*$ má řád $\varphi(n)$, může však být exponentu menšího (viz oddíl 2.8). Je-li $\mathbb{Z}_n^*$ exponentu m (tedy $a^m = 1$ pro každé $a \in \mathbb{Z}_n^*$), je $(\chi(a))^m = \chi(a^m) = \chi(1) = 1$, takže χ je m-tou odmocninou z jedné.

Jsou-li χ_1, χ_2 dva charaktery modulo n, pak jejich součin $\chi = \chi_1\chi_2$, $\chi(a) = \chi_1(a)\chi_2(a)$, $a \in \mathbb{Z}_n^*$, je zjevně také charakterem. Ke každému charakteru χ lze definovat *charakter sdružený* $\overline{\chi}$, $\overline{\chi}(a) = \overline{\chi(a)}$ pro každé $a \in \mathbb{Z}_n^*$. Pro $\alpha \in \mathbb{C}$, $|\alpha| = 1$, je $\alpha^{-1} = \overline{\alpha}$. Proto $\chi\overline{\chi} = \varepsilon$, kde ε je *triviální charakter*, $\varepsilon(a) = 1$ pro každé $a \in \mathbb{Z}_n^*$. Vidíme, že množina všech charakterů modulo n tvoří komutativní grupu.

Je-li $n = p$ prvočíslo, má tato grupa právě $p-1$ prvků. Důvodem je cykličnost $\mathbb{Z}_p^*$. Je-li totiž ξ primitivní prvek $\mathbb{Z}_p$, tak obraz ξ určuje i hodnoty obrazu všech mocnin ξ^j. Pro každé $j \in \mathbb{Z}_p^*$ existuje charakter $\chi = \chi_j$, $\chi_j\left(\xi^i\right) = \zeta_{p-1}^{ij}$ a

všechny charaktery modulo p mají nutně uvedený tvar. Přitom $\chi_j\chi_k = \chi_{j+k}$, takže grupa charakterů modulo p je cyklická řádu $p-1$.

Lemma. *Buď $n > 1$ celé. Ať η je netriviální charakter modulo n, ε triviální charakter modulo n a $b \in \mathbb{Z}_n^*$, $b \neq 1$.*

(i) $\sum_{a\in\mathbb{Z}_n^*} \eta(a) = 0$ a $\sum_{a\in\mathbb{Z}_n^*} \varepsilon(a) = \varphi(n)$.

(ii) Je-li $n = p$ prvočíslo a χ probíhá všechny charaktery modulo p, je $\sum \chi(b) = 0$ a $\sum \chi(1) = p-1$.

Důkaz. Protože η je netriviální, musí být $\eta(c) \neq 1$ pro nějaké $c \in \mathbb{Z}_n^*$. Máme $\sum \eta(a) = \sum \eta(ac) = \eta(c)\,(\sum_a \eta(a))$, takže $\sum \eta(a) = 0$. Pro $a \in \mathbb{Z}_n^*$ je $\varepsilon(a) = 1$, a proto $\sum \varepsilon(a) = |\mathbb{Z}_n^*| = \varphi(n)$.

Z předpokladu $b \neq 1$ vyplývá existence charakteru γ modulo p takového, že $\gamma(b) \neq 1$ (stačí položit $\gamma(\xi^i) = \zeta_{p-1}^i$, kde ξ je primitivní prvek modulo p). Tudíž $\sum_\chi \chi(b) = \sum_\chi (\chi\gamma)(b) = \gamma(b)\left(\sum_\chi \chi(b)\right)$, a tedy $\sum \chi(b) = 0$. Již výše jsme odvodili, že charakterů modulo p je $p-1$. Proto $\sum \chi(1) = p-1$. □

Část (ii) lze přiměřeně vyslovit i pro případ n složené. Je k tomu třeba popsat všechny charaktery modulo n. To je možné, neboť oddíl 2.8 popisuje strukturu grupy $\mathbb{Z}_n^*$.

V obou částech předchozího důkazu se objevuje stejný často používaný princip. Vystupují-li v nějakém výrazu všechny hodnoty $g_1, \ldots, g_n$ grupy řádu n, přičemž všechny mají ve výrazu stejnou roli, tak se hodnota výrazu nezmění, pokud g_i nahradíme g_ig_j (pro nějaké pevné $g_j \in G$). Prvky $g_1g_j, \ldots, g_ng_j$ totiž také probíhají všechny hodnoty grupy G.

Jako jednoduchý důsledek tohoto obecného principu uvedeme geometricky názorné tvrzení o nulovém součtu všech n-tých odmocnin z jedné.

Důsledek. *Pro každé $n > 1$ platí* $\sum_{a\in\mathbb{Z}_n} \zeta_n^a = 0$.

Důkaz. Máme $\zeta_n \neq 1$ a $\zeta_n(\sum \zeta_n^a) = \sum \zeta_n^{a+1} = \sum \zeta_n^a$. □

Budiž χ charakter modulo n. *Gaussovým součtem* tohoto charakteru nazveme hodnotu

$$g(\chi) = \sum_{a\in\mathbb{Z}_n^*} \chi(a)\zeta_n^a$$

Prvky a, přes které se sčítá, lze nahrazovat čísly s a kongruentními modulo n, neboť $\zeta_n^n = 1$.

Je-li $p = n$ prvočíslo a ε je triviální charakter modulo p, tak podle důsledku výše máme $g(\varepsilon) = \sum_{a\in\mathbb{Z}_p^*} \zeta_p^a = \left(\sum_{a\in\mathbb{Z}_n} \zeta_p^a\right) - 1 = -1$.

Tvrzení. *Buď χ netriviální charakter modulo prvočíslo p. Potom* $|g(\chi)| = \sqrt{p}$.

Důkaz. Chceme ověřit, že $g(\chi)\overline{g(\chi)} = p$. Pro $y \in \mathbb{Z}_p^*$ je $\overline{\chi(y)} = \chi(y^{-1})$ a $\overline{\zeta_p^y} = \zeta_p^{-y}$. Proto $g(\chi)\overline{g(\chi)} = \left(\sum_x \chi(x)\zeta_p^x\right)\left(\sum_y \chi(y^{-1})\zeta_p^{-y}\right) = \sum_{x,y} \chi(xy^{-1})\zeta_p^{x-y}$. Je-li $z = xy^{-1}$, máme $x = zy$. Probíhá-li (x, y) množinu $\mathbb{Z}_p^* \times \mathbb{Z}_p^*$, probíhá ji i (xy^{-1}, y). Proto $g(\chi)\overline{g(\chi)} = \sum_{z,y} \chi(z)\zeta_p^{y(z-1)} = \sum_z \chi(z)\left(\sum_y \zeta_p^{y(z-1)}\right)$. Pro

$z \neq 1$ je $\sum_y \zeta_p^{y(z-1)} = g(\varepsilon) = -1$, takže $g(\chi)\overline{g(\chi)} = \left(\sum_y \zeta_p^0\right) - \left(\sum_{z\neq 1} \chi(z)\right) = (p-1) - (-1) = p$, neboť $\sum \chi(z) = 0$ dle lemmatu výše a $\chi(1) = 1$. □

Multiplikativní charaktery modulo n se leckdy definují i pro $a \in \mathbb{Z}_n$, jež je s n soudělné, a to tak, že pro netriviální charakter χ je $\chi(a) = 0$, zatímco pro triviální charakter ε je $\varepsilon(a) = 1$. Pak lze Gaussův součet vyjádřit jako $\sum_{a\in\mathbb{Z}_n} \chi(a)\zeta_n^a$. Jeho hodnota pro χ netriviální se nemění, ovšem pro $\chi = \varepsilon$ dostáváme $\sum_{a\in\mathbb{Z}_n} \zeta_n^a = 0$.

3.5 Kvadratické Gaussovy součty

Legendreův symbol $\left(\frac{a}{p}\right)$ poskytuje charakter $\mathbb{Z}_p^* \to \{-1, 1\}$ (p je liché prvočíslo). Z lemmatu 3.4 okamžitě plyne, že $\sum_{a\in\mathbb{Z}_p} \left(\frac{a}{p}\right) = 0$. Tento vztah plyne i z toho, že $a \in \mathbb{Z}_p^*$ je kvadratické reziduum právě když je prvkem (jediné) podgrupy $\mathbb{Z}_p^*$, která má index dva, takže reziduí a nereziduí je stejně.

Položme $S = \sum \left(\frac{a}{p}\right) \xi_p^a$. Jde o Gaussův součet příslušný charakteru $a \mapsto \left(\frac{a}{p}\right)$. Říká se mu *Gaussův kvadratický součet.*

Uvažme pro nějaké $a \in \mathbb{Z}_p^*$ jeho podsoučet $\left(\frac{a}{p}\right) \xi_p^a + \left(\frac{-a}{p}\right) \xi_p^{-a}$. Je-li $\left(\frac{-1}{p}\right) = 1$, je $\left(\frac{a}{p}\right) = \left(\frac{-a}{p}\right)$, takže jde o součet dvou komplexně sdružených čísel, a výsledkem je číslo reálné. Pokud naopak $\left(\frac{-1}{p}\right) = -1$, jde o součet komplexního čísla s číslem opačným než je komplexně sdružené, a výsledek je ryze imaginární číslo. Podle tvrzení 3.2 máme $\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$. Protože S lze rozdělit do $\frac{p-1}{2}$ součtů uvedených podsoučtů, tak jistě platí

Lemma. *Pro každé liché prvočíslo existuje reálné číslo r takové, že $S = i^{\frac{p-1}{2}} r$.*

Důsledek. *Ať p je liché prvočíslo a ať $S = \sum \left(\frac{a}{p}\right) \zeta_p^a$. Pak $S^2 = \left(\frac{-1}{p}\right) p$.*

Důkaz. Z tvrzení 3.4 víme, že $|S| = \sqrt{p}$. Proto musí být $|r| = \sqrt{p}$, a i^{p-1} je jen jiné vyjádření $\left(\frac{-1}{p}\right)$. □

Všimněte si, že jsme našli přesné vyjádření S^2, nikoli však S. Pro $\left(\frac{-1}{p}\right) = 1$ je $S = \sqrt{p}$ a pro $\left(\frac{-1}{p}\right) = -1$ je $= i\sqrt{p}$. Důkaz tohoto faktu (tj. toho, že S nemůže mít opačné znaménko) je však daleko pracnější.

Znalost S^2 nám bude užitečná při důkazu zákona kvadratické reciprocity. Takových důkazů je několik. Námi zvolený patří k nejkratším a získané poznatky o Gaussových součtech jsou potřebné i jinde. Je však třeba připustit, že některé jiné důkazy jsou intuitivnější.

Pro náš důkaz budeme potřebovat ireducibilitu polynomu $1 + x + x^2 + \cdots + x^{p-1} \in \mathbb{Q}[x]$.

3.6 Gaussovo lemma a kruhové polynomy

Gaussovo lemma vyslovíme pro okruh celých čísel. Podobný důkaz lze použít i pro obecnější tvrzení, kdy místo okruhu celých čísel se uvažuje obor hlavních ideálů. Je možné i další zobecnění na obory s jednoznačnými rozklady (tzv. Gaussovy obory).

Tyto zobecněné varianty se označují také jako Gaussovo lemma.

Gaussovo lemma *Ať $a \in \mathbb{Z}[x]$ není nad $\mathbb{Q}[x]$ ireducibilní,* $\deg(a) \geq 1$. *Pak existují $b, c \in \mathbb{Z}[x]$ stupně menšího než a taková, že $a = bc$.*

Důkaz. Vhodný nenulový násobek polynomu s racionálními koeficienty poskytuje polynom s koeficienty celočíselnými. Proto $ha = uv$ pro nějaké kladné $h \in \mathbb{Z}$ a $uv \in \mathbb{Z}[x]$, $1 \leq \deg u < \deg a$. Buď h nejmenší možné. Je-li $h = 1$, je tvrzení dokázáno. Ať prvočíslo p dělí h. Uvažme redukci $\pi_x : \mathbb{Z} \to \mathbb{Z}_p$ modulo p. Pak $0 = \pi_x(ha) = \pi_x(u)\pi_x(v)$ (viz oddíl 1.21). Protože $\mathbb{Z}_p[x]$ je obor integrity, musí být $\pi_x(u) = 0$ nebo $\pi_x(v) = 0$. Ať například $\pi_x(u) = 0$. To znamená, že všechny koeficienty u jsou dělitelné p, takže $u = pw$ pro nějaké $w \in \mathbb{Z}[x]$. Tedy $(hp^{-1})a = wv$, což je spor s minimalitou h, takže nutně $h = 1$. □

Uvažme nyní opět multiplikativní podgrupu $\mathbb{C}^*$ generovanou ζ_m. Ta má m prvků, jež odpovídají všem m-tým odmocninám z jedné. Její podgrupy jsou všechny grupy $\left\{\zeta_m^{d_i};\ 0 \leq i < m/d\right\}$, kde d dělí m. Protože $\zeta_m^d = \zeta_m d$, můžeme říci, že podgrupami jsou všechny h-té odmocniny z jedné, pro každé h, které dělí m. Vidíme, že m-tá odmocnina z jedné je primitivní, není-li h-tou odmocninou z jedné pro nějaké $h|m$, $h \neq m$. Primitivních m-tých odmocnin je tedy $\varphi(m)$, a mají tvar ζ_m^i, kde i je s m nesoudělné. Označme je na chvíli $\alpha_1, \ldots, \alpha_{\varphi(m)}$. Polynom $(x - \alpha_1) \ldots (x - \alpha_{\varphi(m)})$ se nazývá *kruhový (cyklotomický)* a značí se t_m. Součin všech t_h, $h|m$, jistě dělí součin všech $x - \alpha$, kde α je m-tá odmocnina z jedné, tedy $\prod_{0 \leq i < m} \left(x - \zeta_m^i\right)$. Poslední uvedený součin je polynom stupně m, a tedy je roven $x^m - 1$ (každý činitel součinu je kořenem polynomu $x^m - 1$). Polynom $\prod_{h|m} t_h$ tedy dělí $x^m - 1$, a je stupně $\sum_{h|m} \varphi(h)$, což se podle důsledku 2.5 rovná m. Dokázali jsme $x^m - 1 = \prod_{h|m} t_h$. Indukcí dle m okamžitě vidíme, že t_m je polynom celočíselný, neboť je roven podílu dvou monických celočíselných polynomů (viz lemma 1.17). Je-li $m = p$ prvočíslo, je $x^p - 1 = t_1 t_p = (x - 1)t_p$, takže $t_p = 1 + x + \cdots + x^{p-1}$.

Tvrzení. *Pro každé $m \geq 1$ je kruhový polynom $t_m \in \mathbb{Z}[x]$ ireducibilní a platí $x^m - 1 = \prod_{h|m} t_h$.*

Důkaz. Předpokládejme, že t_m není ireducibilní. Podle Gaussova lemmatu ho ze zapsat jako součin ab celočíselných monických polynomů stupně menšího než $\varphi(m)$. Ať ζ_m^i je kořen polynomu a pro každé i nesoudělné s m. Protože t_m je součin všech takových $x - \zeta_m^i$, dojdeme ke sporu, neboť bude $t_m = a$. Zjevně tedy stačí ověřit, že pro každé α kořen a je α^q také kořen a, kde q je prvočíslo nesoudělné s m.

Označme π redukci $\mathbb{Z} \to \mathbb{Z}_q$ modulo q. Protože t_m dělí $x^m - 1$, je $\pi_x(t_m)$ bez vícenásobných kořenů, dle důsledku 1.23. Koeficienty $\pi_x(a)$ leží v $\mathbb{Z}_q$, takže podle důsledku 1.21 je $\pi(\alpha^q) = (\pi(\alpha))^q$ kořenem $\pi_x(a)$ (máme $\pi_x(t_m) = \pi_x(a)\pi_x(b)$, přičemž $\pi(\alpha)$ je kořen $\pi_x(a)$). Pokud by α^q bylo kořenem b, bylo by $\pi(\alpha^q)$ také

kořenem $\pi_x(b)$. To možné není, neboť $\pi_x(t_m)$ je bez vícenásobných kořenů. Proto α^q není kořen b, a tedy je kořenem a. □

3.7 Zákon reciprocity

Buďte p a q lichá prvočísla, $p \neq q$. Budeme pracovat v okruhu

$$R = \left\{a_0 + a_1\zeta_p + \cdots + a_{p-1}\zeta_p^{p-1};\ a_0, \ldots, a_{p-1} \in \mathbb{Z}\right\}.$$

Ověřit, že R je skutečně okruh (jakožto podokruh $\mathbb{C}$) je snadné. Součin dvou prvků z R je totiž jistě lineární kombinace mocnin ζ_p, přičemž z $\zeta_p^i = 1$ plyne, že se lze v součtu omezit na mocniny ζ_p^i, $0 \leq i < p$. Množina

$$I = \left\{a_0 + a_1\zeta_p + \cdots + a_{p-1}\zeta_p^{p-1};\ a_0, \ldots, a_{p-1} \in q\mathbb{Z}\right\}$$

je jistě ideál R. Je to hlavní ideál příslušný prvku q.

Položme opět $S = \sum_{a\in\mathbb{Z}_p} \left(\frac{a}{p}\right) \zeta_p^a$.

Lemma. *Žádný z prvků $S, 2S, S^2$ a p neleží v $I = qR$.*

Důkaz. Protože q je liché, je $2k \equiv 1 \mod q$ pro nějaké $k \in \mathbb{Z}$. Tudíž $2S \in I \Rightarrow k2S \in I \Rightarrow S \in I$. Z $S \in I$ máme $S^2 \in I$, což ale podle důsledku 3.5 značí $p \in I$. Tento předpoklad dovedeme ke sporu. Z $p \in I$ plyne existence $a_0, \ldots, a_{p-1} \in \mathbb{Z}$ dělitelných q takových, že polynom $a = (a_{p-1}x^{p-1} + \cdots + a_1x + a_0) - p$ má kořen ζ_p. Polynom $a \in \mathbb{Z}[x] \subseteq \mathbb{Q}[x]$ je tudíž soudělný s polynomem $b = x^{p-1} + \cdots + x + 1 \in \mathbb{Z}[x] \subseteq \mathbb{Q}[x]$, neboť mají shodný kořen ζ_p. Protože $b = t_r$ je nad $\mathbb{Q}[x]$ podle tvrzení 3.6 ireducibilní, musí být a násobkem b. Odsud $a_{p-1} = a_{p-2} = \cdots = a_1 = a_0 - p$, což je spor, protože q nedělí $a_0 - p$. □

Věta o reciprocitě *Buď p a q různá lichá prvočísla. Pak*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}$$

Důkaz. Připomeňme, že podle tvrzení 3.2 je vztah možno též zapsat jako $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{-1}{q}\right)$. Idea důkazu vychází z dvojího vyjádření S^q modulo I (kde I, S a R mají stejný význam jako výše).

Prvý způsob spočívá ve vyjádření S^q jako $S \cdot S^{q-1}$ a S^{q-1} jako $\left(S^2\right)^{\frac{q-1}{2}} = \left((-1)^{\frac{p-1}{2}}p\right)^{\frac{q-1}{2}} = p^{\frac{q-1}{2}}(-1)^{(p-1)(q-1)/4} \equiv \left(\frac{p}{q}\right)(-1)^{(p-1)(q-1)/4}$ (využili jsme důsledek 3.5 a základní vztahy oddílu 3.2).

Druhý způsob vychází z toho, že po roznásobení výrazu $(x_1 + \cdots + x_k)^q$ jsou všechny koeficienty s výjimkou x_i^q, $1 \leq i < k$, dělitelné q (důkaz lze obdržet například opakovaným použitím binomické věty). Proto modulo I platí

$$S^q \equiv \sum \left(\left(\frac{a}{p}\right)\zeta_p^a\right)^q = \sum \left(\frac{a}{p}\right)\zeta_p^{aq} = \sum \left(\frac{aq^2}{p}\right)\zeta_p^{aq}.$$

Protože $\left(\frac{aq^2}{p}\right) = \left(\frac{q}{p}\right)\left(\frac{aq}{p}\right)$ a protože aq probíhá $\mathbb{Z}_p$, máme

$$S^q \equiv \left(\frac{q}{p}\right)\left(\sum \left(\frac{aq}{p}\right)\zeta_p^{aq}\right) = \left(\frac{q}{p}\right)S.$$

Dokázali jsme, $S \cdot \varepsilon \equiv 0 \mod I$, kde $\varepsilon = \left(\frac{q}{p}\right) - \left(\frac{p}{q}\right)(-1)^{(p-1)(q-1)/4}$. Je-li $\varepsilon \neq 0$, musí být $\varepsilon \in \{2, -2\}$, odkud $2S \in I$. To je však ve sporu s lemmatem výše. Proto $\varepsilon = 0$ a zbytek je jasný. □

3.8 Jacobiho symboly

Výpočetní použití věty o reciprocitě je velké. Dovoluje totiž postupnou redukci řádu čísel při zjišťování, zda daná hodnota a je nebo není kvadratickým zbytkem modulo dané číslo q.

Uvažme prvočíselný rozklad $a = p_1^{e_1} \dots p_k^{e_k}$. Pak $\left(\frac{a}{p}\right) = \left(\frac{p_1}{q}\right)^{e_1} \dots \left(\frac{p_k}{q}\right)^{e_k}$, takže stačí znát pouze hodnoty $\left(\frac{p_i}{q}\right)$. Pro $p_i = 2$ máme explicitní vzorec $(-1)^{(q^2-1)/8}$, a pro p_1 liché lze hodnotu odvodit ze znalosti $\left(\frac{q}{p_i}\right)$. Přitom se q nahradí hodnotou $a_i \equiv q$, kde $0 < a_i < q$ (nebo lépe $|a_i| < |q|/2$), a celý proces se opakuje. Jeho nevýhodou je však potřeba faktorizace čísla a, což pro velká čísla může být problém.

Ukazuje se ale, že zakládní myšlenku převracení lze provést i bez rozkládání – stačí pouze rozklady tvaru $a \cdot 2^e b$, kde b je liché, které jsou výpočetně snadné.

Formálním nástrojem pro modifikovaný postup je *Jacobiho symbol* $\left(\frac{a}{n}\right)$, který se definuje pro každé $a \in \mathbb{Z}$ a každé kladné liché n, a to tak, že $\left(\frac{a}{1}\right) = 1$, $\left(\frac{a}{q_1 \dots q_k}\right) = \left(\frac{a}{q_1}\right) \dots \left(\frac{a}{q_k}\right)$, kde $q_1, \dots, q_k$ jsou lichá prvočísla (ne nutně různá) a $\left(\frac{a}{q_i}\right)$ je Legendreův symbol.

Tvrzení. *Buď $a, b \in \mathbb{Z}$ a ať $n, m \in \mathbb{Z}$ jsou kladná lichá čísla. Pak*

(i) $\left(\frac{a}{nm}\right) = \left(\frac{a}{n}\right)\left(\frac{a}{m}\right)$;

(ii) $\left(\frac{ab}{n}\right) = \left(\frac{a}{n}\right)\left(\frac{b}{n}\right)$;

(iii) $a \equiv b \mod n \Rightarrow \left(\frac{a}{n}\right) = \left(\frac{b}{n}\right)$;

(iv) $\left(\frac{-1}{n}\right) = (-1)^{(n-1)/2}$;

(v) $\left(\frac{2}{n}\right) = (-1)^{(n^2-1)/8}$;

(vi) $\left(\frac{n}{m}\right) = (-1)^{(n-1)(m-1)/4}\left(\frac{m}{n}\right)$.

Body (i), (ii) a (iii) tvzení plynou okamžitě ze základních vlastností Legendreových symbolů zmíněných v oddílu 3.2. Před důkaz zbylých vložíme následující

Lemma. *Ať $a_1, a_2, \dots, a_k$ jsou lichá čísla. Potom*

$$\frac{a_1 - 1}{2} + \frac{a_2 - 1}{2} + \dots + \frac{a_k - 1}{2} \equiv \frac{a_1 a_2 \dots a_k - 1}{2} \mod 2 \quad \text{a}$$

$$\frac{a_1^2 - 1}{8} + \frac{a_2^2 - 1}{8} + \dots + \frac{a_k^2 - 1}{8} \equiv \frac{(a_1 a_2 \dots a_k)^2 - 1}{8} \mod 8.$$

Důkaz. Pro a, b lichá je $\frac{(a-1)(b-1)}{2}$ sudé, takže je sudé i $\frac{ab-1}{2} - \frac{a-1}{2} - \frac{b-1}{2} = \frac{(a-1)(b-1)}{2}$. Podobně je i $\frac{a^2b^2-1}{8} - \frac{a^2-1}{8} - \frac{b^2-1}{8} = \frac{(a^2-1)(b^2-1)}{8}$ dělitelné osmi, neboť $x^2 \equiv 1 \mod 8$ pro každé liché číslo x. Tím jsou dokázány dva vztahy pro $k = 2$. Dále lze postupovat indukcí. □

Důkaz tvrzení. Ať $n = q_1 \dots q_k$, kde q_i, $1 \le i \le k$, jsou (ne nutně různá) prvočísla. Pak $\left(\frac{-1}{n}\right) = \left(\frac{-1}{q_1}\right) \dots \left(\frac{-1}{q_k}\right) = (-1)^s$, kde $s = (q_1 - 1)/2 + \dots + (q_k - 1)/2$. Podle lemmatu je $(-1)^s = (-1)^{(q_1 \dots q_k - 1)/2}$. Bod (v) se dokáže podobně: $\left(\frac{2}{n}\right) = \left(\frac{2}{q_1}\right) \dots \left(\frac{2}{q_k}\right) = (-1)^{(q_1^2-1)/8} \dots (-1)^{(q_k^2-1)/8} = (-1)^{((q_1 \dots q_k)^2-1)/8}$.

Ať $m = p_1 \dots p_k$, kde p_j, $1 \le j \le k$, jsou prvočísla. Jsou-li n a m čísla soudělná, bude $p_i = q_j$ pro nějaké i a j, a v takovém případě $\left(\frac{n}{m}\right) = \left(\frac{m}{n}\right) = 0$. Dále můžeme tedy předpokládat nesoudělnost n a m. Máme $\left(\frac{n}{m}\right) = \prod_{i,j} \left(\frac{p_i}{q_j}\right)$ a $\left(\frac{m}{n}\right) = \prod_{i,j} \left(\frac{q_j}{p_i}\right)$. Protože $\left(\frac{p_i}{q_j}\right)\left(\frac{q_j}{p_i}\right) = (-1)^{(p_i-1)(q_j-1)/4}$, tak $\left(\frac{n}{m}\right)\left(\frac{m}{n}\right) = (-1)^s$, kde $s = \sum_{i,j}(p_i - 1)(q_j - 1)/4$.

Tedy $s = s_1 s_2$, kde $s_1 = \sum_i (p_i - 1)/2$ a $s_2 = \sum_j (q_j - 1)/2$. Protože $s_1 \equiv (m-1)/2 \mod 2$ a $s_2 \equiv (n-1)/2 \mod 2$, je $(-1)^s = (-1)^{s_1 s_2} = (-1)^{(m-1)(n-1)/4}$. □

Je-li n složené , tak vztah $\left(\frac{a}{n}\right) \equiv a^{\frac{n-1}{2}} \mod n$ platit nemusí. Přitom jak $\left(\frac{a}{n}\right)$, tak $a^{\frac{n-1}{2}}$ lze spočítat relativně rychle. Pokud nalezneme nějaké kladné $a < n$ takové, že $\left(\frac{a}{n}\right) \neq a^{(n-1)/2} \mod n$, budeme vědět, že n není prvočíslo. Na tomto principu je založen takzvaný *Solovay-Strassenův* algoritmus. Již jsme ho vlastně popsali: při náhodném výběru celých a, $|a| < n/2$, zjišťuj shodu mezi $\left(\frac{a}{n}\right)$ a $a^{(n-1)/2}$ modulo n. V případě shody pokračuj výběrem dalšího a až do případného vyčerpání předem zadaného počtu kroků.

Tento algoritmus je méně účinný než Rabin-Millerův. Podrobněji se jím zabývat nebudeme.

Kapitola 4

Hustota a existence prvočísel

4.1 Téma a cíle

Kdyby $p_1, \ldots, p_k$ byla všechna prvočísla, tak žádné z nich by nemohlo dělit $p_1 \cdots p_k + 1$. Proto od dob starých Řeků víme, že prvočísel je nekonečně mnoho. Je však mnoho nezodpovězených otázek, které se ptají na existenci nekonečně mnoha prvočísel určitých zvláštních vlastností. Jsou-li p a q prvočísla taková, že $q = p - 2$, hovoříme o *prvočíselných dvojčatech*. Nevíme, zda jich je nekonečně mnoho.

Prvočísla tvaru $2^a - 1$ se nazývají *Mersennova* a prvočíslům tvaru $2^{2^n} + 1$ říkáme *Fermatova*. Nevíme, zda jedněch či druhých je nekonečně mnoho. Je-li $a = a_1 a_2$, je $2^a - 1 = (2^{a_1})^{a_2} - 1$ dělitelné $2^{a_1} - 1$. Proto Mersennovo prvočíslo má vždy tvar $2^p - 1$, p prvočíslo. Čísla tvaru $2^{2^n} + 1$ se označují v teorii čísel F_n. Nejmenší n, že F_n není prvočíslo, je 5. Lze snadno ověřit, že F_5 je dělitelné prvočíslem 641.

Na druhou stranu není obtížné například ukázat, že pro každé prvočíslo q existuje nekonečně mnoho prvočísel p takových, že $p \equiv 1 \mod q$.

Svým způsobem je velmi pozoruhodné, že ačkoliv neumíme nalézt žádný vzorec, který by deterministicky poskytoval nekonečné řady prvočísel, máme velmi přesné odhady pro relativní počet prvočísel. Ten se obvykle měří pomocí funkce $\pi(x)$, která udává *počet prvočísel* $\leq x$ (přitom x může být i reálné).

Zlomek $\pi(x)/x$ tedy udává relativní počet prvočísel $\leq x$. *Věta o hustotě prvočísel* říká, že

$$\lim_{x \to \infty} \log x \frac{\pi(x)}{x} = 1,$$

čili, že hustota prvočísel do x klesá stejně jako funkce $1/\log x$. Důkaz věty o hustotě prvočísel je poměrně obtížný a dělat ho zde nebudeme. Není však těžké ukázat existenci konstant c_1 a c_2 takových, že $0 < c_1 < 1 < c_2$, a

$$\frac{c_1}{\log n} < \frac{\pi(n)}{n} < \frac{c_2}{\log n}$$

pro každé dostatečně velké n (viz tvrzení 4.4).

Vedle funkce π se používá takzvaná „theta“ funkce, kde $\vartheta(n) = \sum_{p\le n} \log p$. Její hodnota je součet přirozených logaritmů všech prvočísel $p \le n$. Tudíž $e^{\vartheta(n)} = p_1 \cdots p_{\pi(n)}$, kde $2 = p_1 < p_2 < \cdots < p_{\pi(n)}$ jsou všechna prvočísla $\le n$.

Funkce π a ϑ se vzájemně ovlivňují a odhady v následujících oddílech jsou na této interakci částečně založeny. Jde ovšem o dosti volnou vazbu, z čehož vyplývá, že na jejím základě se dají získat pouze odhady poměrně hrubé.

Lemma. *Pro každé $n \ge 2$ platí*

$$2ne^{-1} + 2\vartheta(n) > \sqrt{n}\log n + 2\vartheta(n) \ge \pi(n)(\log n)$$

Důkaz. Máme $\vartheta(n) = \sum_{p\le n} \log p \ge \sum_{\sqrt{n}\le p\le n} \log p \ge (\pi(n) - \sqrt{n}) \log\sqrt{n}$, přičemž $\log\sqrt{n} = (\log n)/2$.

Odsud plyne druhá nerovnost, takže zbývá ověřit $2n > e\sqrt{n}\log n$, tedy $(2/e)\,(\sqrt{n}/\log n) > 1$. Ukážeme, že $(2/e)\,(\sqrt{x}/\log x) \ge 1$ pro každé $x > 1$, přičemž rovnost nikdy nenastává pro x celé. Funkce $\sqrt{x}/\log x$ má derivaci rovnou

$$\left(\frac{\log x}{2\sqrt{x}} - \frac{\sqrt{x}}{x}\right)\left(\log^2 x\right)^{-1} = (\log x - 2) \,/\, \left(2\sqrt{x}\log^2 x\right),$$

což je pro $x > 1$ rovno nule právě pro $x = e^2$, takže v tomto bodě má pro $x > 1$ funkce absolutní minimum. Přitom $(2/e)\left(\sqrt{e^2}/\log e^2\right) = (2/e)\,(e/2) = 1$, takže pro $x \ne e^2$ je $(2/e)\,(\sqrt{x}/\log x) > 1$. □

4.2 Čebyševův odhad

Vyjdeme ze známého vztahu $2^n = \sum_{i=0}^{n} \binom{n}{i}$, který lze například obdržet rozvojem $2^n = (1+1)^n$ pomocí binomické věty.

Lemma. *Pro každé celé $k \ge 0$ platí*

$$2^{2k}/(2k+1) \le \binom{2k}{k} \quad \text{a} \quad \binom{2k+1}{k} \le 2^{2k}.$$

Důkaz. Pro $i < n/2$ je $i + 1 \le n - i$, odkud $\binom{n}{i} \le \binom{n}{i}\frac{n-i}{i+1} = \binom{n}{i+1}$. Mezi hodnotami $\binom{2k}{i} = \binom{2k}{2k-i}$ je tudíž $\binom{2k}{k}$ největší. Proto $\sum \binom{2k}{i} \le (2k+1)\binom{2k}{k}$. Druhá nerovnost vyplývá z $\binom{2k+1}{k} + \binom{2k+1}{k+1} = 2\cdot\binom{2k+1}{k} \le 2^{2k+1}$. □

Tvrzení (Čebyšev). *Pro každé $n \ge 2$ je $\vartheta(n) < n\log 4$.*

Důkaz. Jde o důkaz nerovnosti $\prod_{p\le n} p < 4^n$ (viz oddíl 4.1). Pro $n = 1, 2$ je tato nerovnost zřejmá. Platí-li pro $n = 2k+1 > 2$, tak platí i pro $n = 2k+2$, neboť $2k+2$ není prvočíslo. Pro důkaz indukcí tedy stačí uvažovat přechod od $2k$ k $2k+1$.

Prvočíslo $p > k+1$ dělí $\binom{2k+1}{k} = (2k+1)!/(k!(k+1)!)$ právě když $p \le 2k+1$. Proto součin všech takových prvočísel dělí $\binom{2k+1}{k}$, což je $\le 4^k$ dle lemmatu výše. S využitím indukčního předpokladu pro $k+1$ dostáváme

$$\prod_{p\le 2k+1} p = \prod_{p\le k+1} p \prod_{k+2\le p\le 2k+1} p \le 4^{k+1}\binom{2k+1}{k} \le 4^{k+1}4^k = 4^{2k+1}. \quad \square$$

Důsledek. *Existuje $c_2 > 0$, že $\frac{\pi(n)}{n} < \frac{c_2}{\log n}$ pro každé $n \geq 2$. Přitom za c_2 lze zvolit $(\log 4 + 2/e) \approx 3,54$.*

Důkaz. Z Čebyševova odhadu máme podle lemmatu 4.1 nerovnost $2ne^{-1} + 4n \log 2 > \pi(n)(\log n)$, takže lze položit $c_2 = 2e^{-1} + 4 \log 2 \approx 0,74 + 2,80 \approx 3,54$. □

4.3 Odhady a celé části

Pro a reálné značí $[a]$ největší celé n takové, že $n \leq [a]$. Je tedy $n \leq a < n + 1$. Někdy se místo $[a]$ píše $\lfloor a \rfloor$ a $\lceil a \rceil \in \mathbb{Z}$ se definuje tak, že $\lceil a \rceil - 1 < a \leq \lceil a \rceil$. Mluvíme o *celé části* $[a]$ čísla a, případně o *dolní a horní celé části* $\lfloor a \rfloor$ a $\lceil a \rceil$.

Buď p prvočíslo a $n > 1$ kladné celé číslo. Triviálně platí, že $v_p(n)$ je velikost množiny $\{j \geq 1; p^j \text{ dělí } n\}$.

Pokud $n = n_1 \cdots n_k$, tak $v_p(n) = v_p(n_1) + \cdots + v_p(n_k)$, takže $v_p(n)$ lze vyjádřit též jako součet velikostí množin $\{i; 1 \leq i \leq k \text{ a } p^j \text{ dělí } n_i\}$, kde $j = 1, 2, 3, \ldots$. Pokud $n_1 = 1, n_2 = 2, \ldots, n_k = k$, tak $n = k!$ a pro dané j má uvedená množina právě $\left[\frac{k}{p^j}\right]$ prvků.

Lemma. *Uvažme $k \geq 1$. Pak*

(i) $v_p(k!) = \sum_{j \geq 1} \left[\frac{k}{p^j}\right]$ pro každé prvočíslo p;

(ii) $\left[\frac{2k}{p^j}\right] - 2\left[\frac{k}{p^j}\right] \in \{0, 1\}$ pro každé $j \geq 0$.

Důkaz. Bod (i) okamžitě plyne z úvahy před lemmatem. Co se týče bodu (ii), položíme $h = \left[k/p^j\right]$. Pak $hp^j \leq k < (h+1)p^j$, odkud $(2h)p^j \leq 2k < (2h+2)p^j$. Proto $\left[2k/p^j\right]$ je rovno buď $2h$, nebo $2h + 1$. □

Důsledek. *Buď p prvočíslo.*

(i) Pro všechna celá $n \geq m \geq 0$

$$v_p\binom{n}{m} = \sum_{1 \leq j \leq \log n / \log p} \left(\left[\frac{n}{p^j}\right] - \left[\frac{m}{p^j}\right] - \left[\frac{n-m}{p^j}\right]\right).$$

(ii) Pro každé celé $k \geq 1$ je $v_p\binom{2k}{k} \leq \log(2k)/\log p$.

Důkaz. Máme $\binom{n}{m} = n!/(m!(n-m)!)$, takže $v_p\binom{n}{m} = v_p(n!) - v_p(m!) - v_p((n-m)!)$. Proto bod (i) vyplývá přímo z bodu (i) lemmatu (Omezení $p^j \leq n$ zlogaritmováním dává $j \cdot \log p \leq \log n$). Volba $n = 2k, m = k$ vede na

$$v_p\binom{2k}{k} = \sum_{1 \leq j \leq (\log 2k)/(\log p)} \left[\frac{2k}{p^j}\right] - 2\left[\frac{k}{p^j}\right].$$

Podle bodu (ii) lemmatu je každý ze členů tohoto součtu roven 0 nebo 1. Proto $v_p\binom{2k}{k}$ nemůže být větší než počet všech uvažovaných j, což dává důkaz bodu (ii). □

Tvrzení. *Pro každé $n \geq 2$ platí*

$$\binom{2n}{n} \leq (2n)^{\sqrt{2n}} \prod_{\sqrt{2n}<p\leq 2n} p \leq (2n)^{\sqrt{2n}}(2n)^{\pi(2n)}.$$

Přitom prvá nerovnost platí, i když výběr p, $\sqrt{2n} < p \leq 2n$, omezíme na ta prvočísla, pro která $\left[\frac{2n}{p}\right] \neq 2\left[\frac{n}{p}\right]$.

Důkaz. Položme $m = \binom{2n}{n}$. Pokud prvočíslo p dělí m, musí dělit nějaké k, $n < k \leq 2n$. Proto $m = \prod_{p\leq 2n} p^{v_p(m)}$. Je-li $p^2 \geq 2n$, je $2\log p = \log p^2 > \log n$, takže $2 > \log n/\log p$, a $v_p(m) = \left[\frac{2n}{p}\right] - 2\left[\frac{n}{p}\right]$, dle bodu (i) důsledku výše. Ovšem podle bodu (i) lemmatu výše je $\left[\frac{2n}{p}\right] - 2\left[\frac{n}{p}\right] \in \{0,1\}$.

Proto

$$m \leq \prod_{p\leq\sqrt{2n}} p^{v_p(m)} \cdot \prod_{\sqrt{2n}<p\leq 2n} p,$$

přičemž v druhém činiteli lze uvažovat pouze p, která splňují $\left[\frac{2n}{p}\right] \neq 2\left[\frac{n}{p}\right]$.

Druhý činitel nahradíme $\prod_{p\leq 2n} p$. Každé z prvočísel p je $\leq 2n$, takže součin je $\leq (2n)^{\pi(2n)}$.

Pro členy prvého činitele podle bodu (ii) důsledku výše platí $v_p(m) \leq \log(2n)/\log p$, takže $p^{v_p(m)} = (e^{\log p})^{v_p(m)} \leq e^{\log(2n)} = 2n$.

Odhad $\prod_{p\leq\sqrt{2n}} p^{v_p(m)} \leq (2n)^{\sqrt{2n}}$ plyne z toho, že prvočísel $p \leq \sqrt{2n}$ není více než $\sqrt{2n}$.

□

4.4 Existence prvočísel

Bertrandův postulát. *Pro každé $n \geq 2$ lze najít prvočíslo p takové, že $n < p < 2n$.*

Důkaz. Ať $n > 2$ je nejmenší protipříklad. Uvažme nejprve řadu prvočísel 3, 5, 7, 13, 23, 43, 83, 163, 317, 631, 1259 a 2503. Označme její prvky $a_i, 1 \leq i \leq 11$, a všimněme si, že $a_{i+1} < 2a_i$, $1 \leq i \leq 10$. Pokud $a_i \leq k < a_{i+1}$, tak $k < a_{i+1} < 2a_i \leq 2k$. Mezi k a $2k$ tedy leží prvočíslo a_{i+1}, takže protipříklad n není menší než $a_{11} = 2503 > 2048 = 2^{11}$. Tedy $2n > 2^{12}$.

Položme $m = \binom{2n}{n}$. Využijeme odhad tvrzení 4.3. Protože mezi n a $2n$ podle našeho předpokladu žádné prvočíslo neleží, lze interval $\sqrt{2n} \leq p \leq 2n$ nahradit intervalem $\sqrt{2n} \leq p \leq n$. Ukážeme, že ho lze ještě zmenšit, a to na $\sqrt{2n} \leq p \leq 2n/3$. Podle tvrzení 4.3 můžeme vypustit každé p, $2n/3 < p \leq n$, pokud ověříme, že $\left[\frac{2n}{p}\right] = 2\left[\frac{n}{p}\right]$. Jistě $\left[\frac{n}{p}\right] = 1$ a z $2p \leq 2n < 3p$ plyne $\left[\frac{2n}{p}\right] = 2$.

Je tedy $m \le (2n)^{\sqrt{2n}} \prod_{p \le 2n/3} p \le (2n)^{\sqrt{2n}} e^{\vartheta(2n/3)}$. Využijeme nyní Čebyševův odhad $\vartheta(2n/3) \le (2n/3)\log 4$. Z něj plyne $m \le (2n)^{\sqrt{2n}} 4^{2n/3} = (2n)^{\sqrt{2n}} 2^{4n/3}$.

Položme $k = 2n$. Podle lemmatu 4.2 je $2^k/(k+1) \le m$, takže máme nerovnost

$$2^k \le (k+1) k^{\sqrt{k}} 2^{2k/3} < k^{2+\sqrt{k}} 2^{2k/3} \le k^{4\sqrt{k}/3} 2^{2k/3}$$

V úpravách jsme použili $k + 1 < k^2$ a $2 + s \le 4s/3$, což platí pro $s \ge 6$, a tedy i pro $s = \sqrt{k} > 2^6 = 64$.

Nerovnost můžeme přepsat jako $2^{k/3} < k^{4\sqrt{k}/3}$, tedy jako $2^k < k^{4\sqrt{k}}$, odkud $k < 4\sqrt{k}\log_2 k$, a tedy

$$\sqrt{k} < 4 \log_2 k.$$

Definujme reálné $t > 0$ tak, že $k = 2^{2t}$. Z $2k > 4096 = 2^{12}$ plyne $t > 6$. Naše nerovnost dává $2^t < 8t$, ale $2^6 = 64 > 8 \cdot 6 = 48$. Vidíme, že jsme obdrželi nerovnost, která pro $t \ge 6$ neplatí, a to je hledaný spor. □

Důkaz Bertrandova postulátu lze snadno modifikovat tak, aby se ukázalo, že pro každé předem dané $r \ge 1$ lze najít n_0 takové, že pro všechna $n \ge n_0$ leží mezi n a $2n$ alespoň r prvočísel.

Pro důkaz Bertrandova postulátu byly klíčové nerovnosti tvrzení 4.3. Ukážeme, že toto tvrzení lze použít i pro dolní odhad počtu prvočísel, totiž pro důkaz existence $c_1 > 0$ takového, že $c_1/\log n < \pi(n)/n$.

Podle lemmatu 4.2 a tvrzení 4.3 platí

$$2^{2n}/(2n+1) \le \binom{2n}{n} \le (2n)^{\sqrt{2n}} (2n)^{\pi(2n)}.$$

Pro každé $\alpha > 1$ roste $\alpha^{2n}/(2n+1)$ nade všecky meze, a proto pro každé ε, $0 < \varepsilon < 1$, je $2^{2n}/(2n+1) \ge (2-\varepsilon)^{2n}$ pro n dostatečně velké.

Ověříme, že pro $\varepsilon = 1/8$ vztah platí, je-li $n > 210$. Položme $m = 2n$. Nerovnost $2^m/(m+1) \ge (15/8)^m$ lze ekvivalentně vyjádřit jako $2^{4m} \ge (m+1)15^m$ a tedy jako $\left(1 + \frac{1}{15}\right)^m \ge m + 1$. Z binomické věty máme $\left(1 + \frac{1}{15}\right)^m > 1 + \frac{m}{15} + \binom{m}{2}/225$, takže stačí zjistit, že pro $m > 420$ platí $1/15 + (m-1)/450 \ge 1$, což ovšem vyplývá z $30 + 420 = 450$.

Nerovnost $(2-\varepsilon)^{2n} \le (2n)^{\sqrt{2n}} (2n)^{\pi(2n)}$ dává $2n\log(2-\varepsilon) < \sqrt{2n}\log(2n) + \pi(2n)\log(2n)$, odkud

$$\log(2-\varepsilon) - \frac{\log(2n)}{\sqrt{2n}} < \frac{\pi(2n)\log(2n)}{2n}.$$

Protože $\log(2n)/\sqrt{2n}$ je pro dostatečně velká n libovolně blízko nule, vidíme, že pro libovolné $\delta > 0$ lze zvolit n_0 takové, že pro $n > n_0$ je $\log(2-\varepsilon) - \delta < \frac{\pi(n)}{2n}\log(2n)$.

Jelikož $\pi(2n) = \pi(2n-1)$, tak také platí

$$\frac{\pi(2n-1)}{2n-1}\log(2n-1) > \frac{\pi(2n)}{2n}\log(2n) > \log(2-\varepsilon) - \delta.$$

Položíme-li $c_1 = \log(2-\varepsilon) - \delta$ a vezmeme-li v úvahu důsledek 4.2, můžeme vyslovit

Tvrzení. *Existuje celé kladné číslo n_0 a reálná čísla c_1 a c_2, $0 < c_1 < 1$, $c_2 > 1$, taková, že pro $n \geq n_0$ je*

$$\frac{c_1}{\log n} < \frac{\pi(n)}{n} < \frac{c_2}{\log n}.$$

□

Podle důsledku 4.2 lze za c_2 volit $2e^{-1} + 4\log 2 \approx 3,54$, přičemž horní odhad nezávisí na volbě n_0. Podle úvah výše lze c_1 zvolit libovolně blízké $\log 2 \approx 0,7$.

Pro $k = 256$ snadno spočítáme, že $\log(2-\varepsilon) - \frac{\log(2k)}{\sqrt{2k}}$ je pro $\varepsilon = 1/8$ roven $\log 15 - 3\log 2 - (9\log 2)/(16\sqrt{2}) > 0,35$. Vzhledem k tomu, že $2^{2k}/(2k+1) > (7/8)^{2k}$ platí pro $k \geq 256 > 210$, vidíme, že pro $n > 512$ je $\pi(n)\log n/n > 7/20$.

Tento vztah platí dokonce pro všechna $n \geq 2$, ovšem pro $n \leq 512$ je ho třeba přímo ověřit ze znalosti hodnot $\pi(n)$. Dělat to zde nebudeme, protože hodnoty c_1 a c_2, které jsme zde odvodili, mají spíše jen ilustrativní význam. Použitím jemnějších a náročnějších metod je lze výrazně zpřesňovat. Pro čísla n, která mají okolo sta desetinných míst, což je na spodní hranici velikosti prvočísel generovaných pro metodu RSA, je aproximace $\pi(n)/n$ hodnotou $1/\log n$ již velmi přesná.

Kapitola 5

Řetězové zlomky

5.1 Dobré aproximace

Zde a v následujících oddílech budeme vzdálenost reálného čísla ϑ od nejbližšího čísla celého označovat $\|\vartheta\|$. Je tedy

$$\|\vartheta\| = \min\{|\vartheta - n|; n \in \mathbb{Z}\}$$

Budeme též používat označení $\langle\vartheta\rangle$ ve významu $\vartheta - [\vartheta]$.

Celé číslo nejbližší k $\vartheta \in \mathbb{R}$ lze považovat za *celočíselnou aproximaci* ϑ. Vedle celočíselných aproximací můžeme uvažovat i aproximace se jmenovatelem q, kde q je celé kladné. Takovou aproximací je zlomek p/q, pro který platí $|\vartheta - p/q| \leq \frac{1}{2q}$. Je-li ϑ racionální, není p/q nutně určeno jednoznačně, u ϑ iracionálního tomu tak však je.

Nerovnost $|\vartheta - p/q| \leq \frac{1}{2q}$ lze zapsat jako $|\vartheta q - p| \leq 1/2$, takže aproximaci se jmenovatelem q lze odvodit z celočíselné aproximace čísla ϑq. Přitom zjevně $|\vartheta - p/q| = \frac{1}{q}\|q\vartheta\|$.

Ne všechny aproximace jsou stejně dobré. Není těžké najít situace, kdy $\|q'\vartheta\|/q' < \|q\vartheta\|/q$ pro nějaké $q' < q$, čili kdy aproximace s menším jmenovatelem dává menší absolutní aproximační chybu. Nás zde budou zajímat aproximace, kdy žádná aproximace s menším jmenovatelem nedává menší relativní chybu, to jest chybu vztaženou k velikosti škály $1/q$. Relativní aproximační chyba je rovna $(\|q\vartheta\|/q)\,/\,(1/q) = \|q\vartheta\|$. Zmíněná podmínka tudíž říká, že

$$\|q\vartheta\| < \|q'\vartheta\|$$

pro všechna kladná celá $q' < q$.

Pokud q tuto podmínku splňuje a $|q\vartheta - p| = \|q\vartheta\|$, nazveme p/q *dobrou aproximací* čísla ϑ. Vztah $|\vartheta - p/q| \leq 1/(2q)$, který plyne z $\|q\vartheta\| \leq 1/2$, značí, že q určuje p téměř jednoznačně (pro některá racionální ϑ připadají v úvahu dvě volby p).

Lemma. *Ať p/q je dobrou aproximací reálného čísla ϑ. Potom jsou p a q čísla nesoudělná.*

Důkaz. Ať $p = p'd$ a $q = q'd$, kde $d \geq 1$. Pak $\|q\vartheta\| = |q\vartheta - p| = d|q'\vartheta - p'| \geq d\,\|q'\vartheta\|$. Proto nemůže být $\|q\vartheta\| < \|q'\vartheta\|$, takže také není $q' < q$, a tedy nutně $d = 1$. □

Tvrzení. *Ať ϑ a Q jsou reálná čísla, $Q > 1$. Pak existuje $q < Q$ přirozené, jež splňuje $\|q\vartheta\| \leq 1/Q$. Je-li Q necelé nebo ϑ iracionální, platí $\|q\vartheta\| < 1/Q$.*

Důkaz. Zvolme $n \in \mathbb{Z}$ tak, aby $n - 1 < Q \leq n$. Uvažme následujících $n + 1$ hodnot: $0, 1, \langle\vartheta\rangle, \langle 2\vartheta\rangle, \ldots, \langle(n-1)\vartheta\rangle$. Každou z nich lze vyjádřit jako $j\vartheta - r$, kde j a r jsou celá, $0 \leq j < n$. Protože interval $[0, 1]$ je sjednocením n intervalů $[i/n, (i+1)/n]$, $0 \leq i < n$, musí alespoň dvě z uvažovaných hodnot ležet uvnitř jednoho z těchto intervalů. Máme $n \geq 2$, takže takovou dvojicí nemůže být 0 a 1. Proto existují celá j, j', r a r', že $0 \leq j < j' < n$ a $|(j' - j)\vartheta - (r' - r)| = |(j'\vartheta - r') - (j\vartheta - r)| \leq 1/n$. Volba $q = j' - j$ a $p = r' - r$ dává $|q\vartheta - p| = \|q\vartheta\| \leq 1/n$. Přitom $q \leq j' \leq n - 1 < Q$ a z $n > Q$ plyne $\|q\vartheta\| < 1/Q$. □

Předchozí tvrzení dovolují hledat dobré aproximace tak, že za $1/Q$ vezmeme minimum ze všech $\|q'\vartheta\|$, kde q' je menší než předem zadaná mez, a q zvolíme jako nejmenší číslo splňující $\|q\vartheta\| < 1/Q$. V dalších oddílech nahlédneme, že pokud takto postupujeme systematicky, obdržíme všechny dobré aproximace.

5.2 Aproximace racionálních čísel

Dobré aproximace mají užití především tehdy, když je aproximované číslo ϑ iracionální. Pro $\vartheta = \frac{r}{s}$ racionální (kde r a s jsou nesoudělná celá, $s \geq 1$) jde o pojem poměrně průhledný. Seznámíme se s ním zejména pro získání výchozí představy pro iracionální případ.

Pro a celé pišme na chvíli $|a|_s$ pro označení nejmenší možné hodnoty $|b|$, kde b probíhá celá čísla taková, že $a \equiv b \mod s$. Jistě $0 \leq |a|_s \leq s/2$. Pro každé $q \geq 1$ je $\left\|q \cdot \frac{r}{s}\right\|$ rovno nejmenší možné hodnotě $|q \cdot \frac{r}{s} - p|$, kde $p \in \mathbb{Z}$. Jde tedy o minimalizaci $|qr - ps|$, takže $\left\|q \cdot \frac{r}{s}\right\| = \frac{1}{s}|qr|_s$. Požadavek na to, aby k danému $q \geq 1$ bylo možno nalézt p takové, že p/q je dobrá aproximace r/s, lze tedy vyjádřit jako

$$|q'r|_s > |qr|_s \quad \text{pro všechna } q', 1 \leq q' < q.$$

Sestrojme nyní posloupnost $q_1 = 1, \ldots, q_k$ takovou, že pro každé $i \geq 1$, jež má $|q_i r|_s > 0$ položíme

$$q_{i+1} = \min\{q > q_i; |qr|_s < |q_i r|_s\}.$$

Z definice vyplývá $|q_1 r|_s > |q_2 r|_s > \ldots$, takže dostáváme klesající posloupnost nezáporných čísel. Ta je jistě konečná. Čísla r a s jsou nesoudělná, a proto je $|qr|_s > 0$ pro každé celé kladné $q < s$. Jelikož $|sr|_s = 0$, vidíme, že poslední člen posloupnosti $q_1, \ldots, q_k$ je roven s.

Je dobré si uvědomit, že hodnotu q_{i+1} lze také definovat jako minimální $q \geq 1$ takové, že $|qr|_s < |q_i r|_s$. Kdyby tohoto minima bylo dosaženo pro $q \neq q_{i+1}$, muselo by být $q \leq q_i$. Dále by muselo být $q \geq 2$, neboť $|q_i r|_s < |q_1 r|_s = |r|_s$, takže by existovalo celé j, $1 \leq j < i$, že $q_j < q \leq q_{j+1}$. Z $|qr|_s < |q_i r|_s < |q_j r|_s$ plyne $q = q_{j+1}$, odkud $|q_i r|_s \leq |q_{j+1} r|_s = |qr|_s$, což je spor s původním přepokladem.

Z takto pozměněné definice q_i okamžitě vidíme, že pro žádné kladné celé $q < q_i$ nemůže být $|qr|_s = |q_i r|_s$. Je tedy $|qr|_s < |q_i r|_s$, takže podle úvahy výše lze pro každé q_i nalézt celé p takové, že p/q_i je dobrá aproximace čísla $\vartheta = r/s$. Z konstrukce posloupnost q_i vyplývá, že žádné jiné dobré aproximace ϑ neexistují.

Za p zvolíme to celé číslo, jež splňuje $|q_i r - ps| = |q_i r|_s$, čili je to $p = x$, jež dává nejmenší možnou hodnotu $|q_i r - xs|$. Pokud jsou takové hodnoty dvě a za x zvolíme menší z nich, tak $q_i r - xs = (x+1)s - q_i r$, odkud $2q_i r = (2x+1)s$, takže $s = 2t$ je sudé a $|q_i r|_{2t} = t$. Protože t je nejvyšší možná hodnota $|a|_s$, může uvedený stav nastat pouze pro $i = 1$. Požadavek $|r|_{2t} = t$ implikuje soudělnost r a $s = 2t$, pokud $|t| > 1$. Vidíme, že nejednoznačnost volby p nastane jedině když $s = 2$ a r je liché. V ostatních případech určuje q_i hodnotu $p = p_i$ jednoznačně.

Můžeme tedy vyslovit následující tvrzení:

Tvrzení. *Buď r/s racionální číslo, kde r a s jsou celá nesoudělná a $s \geq 1$. Pro $a \in \mathbb{Z}$ ať $|a|_s = \min\{|b|; a \equiv b \mod s\}$. Je-li $s = 2$, jsou mimo číslo r/s jedinými jeho dobrými aproximacemi celá čísla $(r \pm 1)/2$. Je-li $s = 1$, je $r = r/s$ jedinou dobrou aproximací.*

Ať $s > 2$ a definujme $q_1 = 1, \ldots, q_k = s$ tak, že pro $q_i \neq s$ je $q_{i+1} = \min\{q \geq q_i; |qr|_s < |q_i r|_s\}$. Pak existuje jediné p_i, jež splňuje $|q_i r - p_i s| = |q_i r|_s$ a $p_1/q_1, \ldots, p_k/q_k = r/s$ jsou právě všechny dobré aproximace čísla r/s. □

5.3 Aproximace iracionálních čísel

Buď ϑ iracionální číslo. Položme $q_1 = 1$ a konstruujme ostře rostoucí posloupnost q_i tak, že $q_{i+1} = \min\{q > q_i; \|q\vartheta\| < \|q_i\vartheta\|\}$. Množina, ze které se hledá minimum, je podle tvrzení 5.1 neprázdná, neboť $0 < \|q_i\vartheta\| < 1$, takže posloupnost $q_1, q_2, \ldots$ je korektně definovaná. Vztah $\|q_i\vartheta\| = |q_i\vartheta - p_i|$ definuje $p_i \in \mathbb{Z}$ jednoznačně, neboť iracionální číslo $q_i\vartheta$ má jedinou celočíselnou aproximaci.

Tvrzení. *Ať ϑ je iracionální číslo a ať $q_1, q_2, \ldots$ a $p_1, p_2, \ldots$ jsou výše definované posloupnosti celých čísel. Pak p_i a q_i jsou pro každé $i \geq 1$ nesoudělná, p_i/q_i je dobrou aproximací ϑ, $q_i < q_{i+1}$ a*

$$\|q_{i+1}\vartheta\| < \|q_i\vartheta\| < 1/q_{i+1}.$$

Je-li p/q dobrá aproximace ϑ, $q \geq 1$, tak $p/q = p_i/q_i$ pro nějaké $i \geq 1$.

Důkaz. Podle definice je $p_1 = p_1/1 = p_1/q_1$ celočíselnou aproximací ϑ, což je jistě dobrá aproximace. Pro $i \geq 1$ indukcí ukažme, že p_{i+1}/q_{i+1} je také dobrou aproximací. Je-li $1 \leq q \leq q_i$, je $\|q\vartheta\| \geq \|q_i\vartheta\|$ dle definice q_{i+1}. Podle téže definice je ale $\|q\vartheta\| \geq \|q_i\vartheta\| > \|q_{i+1}\vartheta\|$ také pro každé q, $q_i < q < q_{i+1}$, takže q_{i+1} vskutku poskytuje dobrou aproximaci.

Je-li p/q nějaká dobrá aproximace, nalezneme největší i takové, že $q_i < q$. Takové i jistě existuje, neboť posloupnost $q_1, q_2, \ldots$ je ostře rostoucí. Pak ale z $q \leq q_{i+1}$ a $\|q\vartheta\| < \|q_i\vartheta\|$ plyne $q = q_{i+1}$.

Nerovnost $\|q_i\vartheta\| < 1/q_{i+1}$ lze zapsat jako $q_{i+1} < 1/\|q_i\vartheta\|$. Tvrzení 5.1 zaručuje (při volbě $Q = 1/\|q_i\vartheta\|$) existenci $q < 1/\|q_i\vartheta\|$ takového, že $\|q\vartheta\| < \|q_i\vartheta\|$. Každé takové ovšem splňuje $q \geq q_{i+1}$. □

Důsledek. *Ať p_i/q_i, $i \geq 1$, je posloupnost dobrých aproximací iracionálního čísla ϑ. Pak $|p_i/q_i - \vartheta| < 1/(q_i q_{i+1}) < 1/q_i^2$ pro každé $i \geq 1$, takže $\lim_{i\to\infty} p_i/q_i = \vartheta$.*

Důkaz. Máme $|p_i/q_i - \vartheta| = |p_i - q_i\vartheta|/q_i = \|q_i\vartheta\|/q_i < 1/(q_i q_{i+1}) < 1/q_i^2$. Protože $q_1, q_2, \ldots$ je ostře rostoucí posloupnost, kladných čísel, blíží se $1/q_i^2$

limitně k nule. □

Nerovnost $\|q_i\vartheta\| < 1/q_{i+1}$ je pro další pokračování podstatná. Její odvození využívá toho, že tvrzení 5.1 nezaručuje jenom libovolně malou aproximaci ϑ racionálním číslem, ale dává i omezení na velikost jmenovatele. Nerovnost lze použít i v případě racionálního čísla r/s. Je-li $p_1/q_1, \ldots, p_k/q_k$ jemu příslušná posloupnost dobrých aproximací (viz oddíl 5.2), tak $\|(q_i r)/s\| \le 1/q_{i+1}$ je totéž jako

$$q_{i+1}|q_i r|_s \le s \quad \text{pro každé } i, 1 \le i < k.$$

Všimněme si, že na rozdíl od iracionálních čísel zde vystupuje nerovnost neostrá. Je-li $i = k-1$, pak vztah implikuje $|q_{k-1}r|_s = 1$, neboť $q_k = s$, takže zaměnit neostrou nerovnost za ostrou zde nelze.

5.4 Vlastnosti posloupnosti dobrých aproximací

Ať ϑ je iracionální číslo a ať p_i/q_i je posloupnost dobrých aproximací zkonstruovaná v oddíle 5.3. Definujme ještě $q_0 = 0$ a ať

$$p_0 = \begin{cases} 1 & \text{pokud } \vartheta > p_1 \\ -1 & \text{pokud } \vartheta < p_1. \end{cases}$$

Tvrzení. *Za výše uvedených předpokladů platí:*

(i) Čísla $q_{i+1}\vartheta - p_{i+1}$ a $q_i\vartheta - p_i$ mají pro každé $i \ge 0$ opačná znaménka.

(ii) Čísla $q_{i+1}p_i - q_i p_{i+1}$, $i \ge 0$, nabývají střídavě hodnot 1 a -1. Hodnota 1 nastává právě když $q_{i+1}\vartheta - p_{i+1} > 0$.

(iii) Existuje jednoznačně určená posloupnost celých kladných čísel $a_1, a_2, \ldots$ taková, že pro každé $i \ge 1$ je $q_{i+1} = a_i q_i + q_{i-1}$ a $p_{i+1} = a_i p_i + p_{i-1}$. Přitom platí $a_1 = q_2$.

(iv) Pro všechna $i \ge 1$ platí $|q_{i-1}\vartheta - p_{i-1}| = a_i|q_i\vartheta - p_i| + |q_{i+1}\vartheta - p_{i+1}|$.

Důkaz. Bod (i) pro $i = 0$ tvrdí, že $\vartheta - p_1$ a $-p_0$ mají opačná znaménka, tedy že p_0 a $\vartheta - p_1$ mají shodná znaménka (jsou buď obě kladná, nebo obě záporná). Tak tomu podle definice p_0 skutečně je.

Předpokládejme $i \ge 1$. Jsou-li u a v nenulová reálná čísla stejného znaménka, platí $|u - v| = |v - u| < \max(|u|, |v|)$. Pokud $u = q_i\vartheta - p_i$ a $v = q_{i+1}\vartheta - p_{i+1}$ mají stejná znaménka, tak z $|q_{i+1}\vartheta - p_{i+1}| = \|q_{i+1}\vartheta\| < \|q_i\vartheta\| = |q_i\vartheta - p_i|$ plyne $|(q_{i+1} - q_i)\vartheta - (p_{i+1} - p_i)| < \|q_i\vartheta\|$, takže $\|(q_{i+1} - q_i)\vartheta\| < \|q_i\vartheta\|$. To ovšem není možné, neboť $0 < q_{i+1} - q_i < q_{i+1}$ a q_{i+1} je nejmenší přirozené q takové, že $|q\vartheta| < \|q_i\vartheta\|$. Tím je (i) dokázáno.

Pro $i = 0$ je $q_{i+1}p_i - q_i p_{i+1} = p_0 = \pm 1$, přičemž $p_0 = 1$ právě když $q_{i+1}\vartheta - p_{i+1} = \vartheta - p_i$ je kladné. Pro $i \ge 1$ máme

$$q_{i+1}p_i - q_i p_{i+1} = q_i(q_{i+1}\vartheta - p_{i+1}) - q_{i+1}(q_i\vartheta - p_i).$$

Pravá strana této rovnosti je podle bodu (i) rozdílem dvou hodnot opačného znaménka. Lze ji tedy považovat za součet hodnot stejného znaménka jako má $q_{i+1}\vartheta - p_{i+1}$. Absolutní hodnota tohoto součtu je kladné celé číslo, které splňuje:

$$q_i \, \|q_{i+1}\vartheta\| + q_{i+1} \, \|q_i\vartheta\| < (q_i/q_{i+2}) + (q_{i+1}/q_{i+1}) < 2,$$

takže vskutku musí být $q_{i+1}p_i - q_{i+1} = \pm 1$, přičemž znaménka se střídají.

Tím je dokázán bod (ii). Z něj plyne, že

$$q_{i+1}p_i - q_i p_{i+1} = -(q_i p_{i-1} - q_{i-1}p_i) \text{ pro každé } i \geq 1.$$

Proto $p_i(q_{i+1} - q_{i-1}) = q_i(p_{i+1} - p_{i-1})$. Čísla p_i a q_i jsou kladná nesoudělná (viz lemma 5.1), takže q_i dělí $q_{i+1} - q_{i-1}$. Definujme a_i jako $(q_{i+1} - q_{i-1})/q_i$. Vidíme, že a_i je celé kladné a vyhovuje vztahu $q_{i+1} = a_i q_i + q_{i-1}$, přičemž tímto vztahem je zadáno jednoznačně. Pro $i = 1$ je $a_2 = (q_2 - 0)/1 = q_2$. Máme též $p_{i+1} - p_{i-1} = p_i(q_{i+1} - q_{i-1})/q_i = p_i a_i$, čímž je důkaz bodu (iii) u konce.

Podle bodu (iii) vztah

$$q_{i-1}\vartheta - p_{i-1} = -a_i(q_i\vartheta - p_i) + (q_{i+1}\vartheta - p_{i+1})$$

platí pro každé $i \geq 0$. Oba sčítance výrazu vpravo mají podle bodu (i) shodné znaménko, takže rovnost platí i při přechodu k absolutním hodnotám, což je obsahem tvrzení bodu (iv). □

Předchozí tvrzení platí přiměřeně pro $\vartheta = r/s$ racionální, kde r a s jsou nesoudělná a $s \geq 3$. Buď $p_1/q_1, \ldots, p_k/q_k$ příslušná posloupnost dobrých aproximací (viz oddíl 5.3). Dodefinujme p_0 a q_0 stejně jako na počátku tohoto oddílu. Pak lze snadno ověřit, že (i) platí pokud $0 \leq i \leq k-2$, (ii) platí když $0 \leq i \leq k-1$ a (iii) spolu s (iv) jsou splněny, je-li $1 \leq i \leq k-1$.

Vztahy $q_{i+1} = a_i q_i + q_{i-1}$ a $p_{i+1} = a_i p_i + p_{i-1}$ lze vyjádřit maticemi jako rovnost

$$\begin{pmatrix} q_{i+1} & p_{i+1} \\ q_i & p_i \end{pmatrix} = \begin{pmatrix} a_i & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_i & p_i \\ q_{i-1} & p_{i-1} \end{pmatrix}$$

Toto vyjádření nám umožňuje snadno odvodit některé další vlastnosti posloupností p_i a q_i. Budeme přitom postupovat obecněji tak, aby dosažené výsledky nebyly vázány pouze na posloupnosti dobrých aproximací.

Lemma. *Ať $a_1, \ldots, a_n$ jsou kladná reálná čísla, $n \geq 1$, a ať p_0, p_1, q_0 a q_1 jsou rovněž reálná, přičemž $q_0 \geq 0$ a $q_1 > 1$. Položme $\Delta = p_0 q_1 - p_1 q_0$, a definujme $p_i = a_{i-1}p_{i-1} + p_{i-2}$ a $q_i = a_{i-1}q_{i-1} + q_{i-2}$, kde $2 \leq i \leq n+1$. Pak $q_i > 0$, je-li $1 \leq i \leq n+1$, a platí:*

(i) $q_{i+1}p_i - p_{i+1}q_i = (-1)^i\Delta$, *je-li* $0 \leq i \leq n$;

(ii) $\frac{p_i}{q_i} - \frac{p_{i+1}}{q_{i+1}} = \frac{(-1)^i\Delta}{q_i q_{i+1}}$, *je-li* $1 \leq i \leq n$;

(iii) $\frac{p_i}{q_i} - \frac{p_{i+2}}{q_{i+2}} = \frac{(-1)^i a_{i+1}\Delta}{q_i q_{i+2}}$, *je-li* $1 \leq i < n$; *a*

(iv) $C_2 > C_4 > \cdots > C_{2(k+\varepsilon)} > C_{2k+1} > \cdots > C_3 > C_1$, *pokud* $\Delta \neq 0$, $C_i = p_i/(q_i\Delta)$ *a* $n = 2k + \varepsilon$, *kde* $\varepsilon \in \{0, 1\}$ *a* k *je celé.*

Důkaz. Kladnost čísel q_i plyne přímo z definice. Iterací maticového vyjádření obdržíme

$$\begin{pmatrix} q_{i+1} & p_{i+1} \\ q_i & p_i \end{pmatrix} = \begin{pmatrix} a_i & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_i & p_i \\ q_{i-1} & p_{i-1} \end{pmatrix}$$

Matice s hodnotami a_i mají determinant roven -1. Bod (i) tedy plyne z faktu, že determinant součinu matic je roven součinu determinantů. Bod (ii) získáme z bodu (i) krácením hodnotou $q_i q_{i+1}$. Bod (iii) pak dostaneme výpočtem z bodu (ii):

$$\frac{p_i}{q_i} - \frac{p_{i+2}}{q_{i+2}} = \left(\frac{p_i}{q_i} - \frac{p_{i+1}}{q_{i+1}}\right) + \left(\frac{p_{i+1}}{q_{i+1}} - \frac{p_{i+2}}{q_{i+2}}\right) = \\ = \frac{(-1)^i \Delta}{q_i q_{i+1}} + \frac{(-1)^{i+1}\Delta}{q_{i+1} q_{i+2}} = (-1)^i \Delta \frac{q_{i+2} - q_i}{q_i q_{i+1} q_{i+2}},$$

přičemž $q_{i+2} - q_i = a_{i+1} q_{i+1}$.

Předpokládejme nyní $\Delta \neq 0$. Je-li $i < n$ sudé, tak máme $C_i - C_{i+2} = a_{i+1}/(q_i q_{i+2}) > 0$, zatímco pro $i < n$ liché je $C_i - C_{i+2} = -a_{i+1}/(q_i q_{i+2}) < 0$. Odtud plyne $C_2 > C_4 > \cdots$ a $C_1 < C_3 < \cdots$. Pro i sudé je též $C_i > C_{i+1}$, neboť $C_i - C_{i+1} = (q_i q_{i+1})^{-1}$ a pro i liché obdobně $C_{i+1} > C_i$. Je-li n liché, máme $C_{2(k+\varepsilon)} > C_{2k+1}$, neboť $\varepsilon = 1$, $2(k+\varepsilon) = n+1$ a $2k+1 = n$. Pro n sudé stejná nerovnost plyne z $\varepsilon = 0$, $2(k+\varepsilon) = n$ a $2k+1 = n+1$. □

Všimněme si, že $\Delta = \pm 1$, pokud p_i/q_i vyjadřují dobré aproximace čísla ϑ.

5.5 Zavedení řetězových zlomků

Konečným řetězovým zlomkem se rozumí každý výraz tvaru

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots + \cfrac{1}{a_{n-1} + \cfrac{1}{a_n}}}},$$

kde $a_0 \in \mathbb{R}$ a $a_1, \ldots, a_n \in \mathbb{R}^+$.

Výraz může být korektně definován i tehdy, jsou-li některá z čísel $a_1, a_2, \ldots, a_n$ nekladná. Je však zvykem pracovat pouze s kladnými koeficienty, což z hlediska aplikací dostačuje.

Čistým nazýváme takový konečný řetězový zlomek, ve kterém jsou všechna čísla $a_0, a_1, \ldots, a_n$ celá. Takovými řetězovými zlomky se budeme zabývat především. Pro odvození různých rekurzivních vztahů je však účelné připustit i neceločíselné hodnoty.

Uvedený řetězový zlomek zapisujeme $[a_0, \ldots, a_n]$. Zpravidla zapisujeme v hranatých závorkách více argumentů, takže k záměně s označením celé části by nemělo dojít.

Užitečné jsou rekurzivní vztahy jak zprava: $[a_0] = a_0, [a_0, \ldots, a_n, a_{n+1}] = [a_0, \ldots, a_n + a_{n+1}^{-1}]$; tak zleva: $[a_0] = a_0, [a_0, a_1, \ldots, a_n] = a_0 + [a_1, \ldots, a_n]^{-1}$. Oba lze použít pro formální definici konečných řetězových zlomků. Vyjádření

zprava lze zobecnit na vztah $[a_0, \dots, a_n] = [a_0, \dots, a_k, [a_{k+1}, \dots, a_n]]$, pro každé k, $0 \le k < n$.

Tvrzení. *Každé racionální číslo lze vyjádřit čistým konečným řetězovým zlomkem.*

Důkaz. Budeme postupovat indukcí dle $s \ge 1$, kde r/s je racionální číslo, které chceme vyjádřit. Případ $s = 1$ je zřejmý. Nechť $s > 1$, přičemž $r = qs + t$, kde $q \in \mathbb{Z}$ a $0 < t < s$. Podle indukčního předpokladu je $s/t = [a_1, \dots, a_n]$ pro kladná $a_1, \dots, a_s \in \mathbb{Z}$.

Tudíž

$$\frac{r}{s} = q + \left(\frac{s}{t}\right)^{-1} = q + \frac{1}{[a_1, \dots, a_n]} = [q, a_1, \dots, a_n].$$

□

Následující lemma má stejné předpoklady jako lemma 5.4 a ukazuje souvislost zkoumaných posloupností s řetězovými zlomky. Tuto souvislost později použijeme na dobré aproximace.

Lemma. *Ať $a_1, \dots, a_n$ jsou kladná reálná čísla, $n \ge 1$, a ať p_0, p_1, q_0 a q_1 jsou rovněž reálná, přičemž $q_0 \ge 0$ a $q_1 > 1$. Pak pro každé i, $0 \le i \le n$, platí*

$$\frac{p_{i+1}}{q_{i+1}} = \frac{p_0 + p_1[a_1, \dots, a_i]}{q_0 + q_1[a_1, \dots, a_i]}$$

Důkaz. Postupujme indukcí. Pro $i = 1$ dostáváme $p_2/q_2 = (p_0 + a_1p_1)/(q_0 + a_1q_1)$, což odpovídá definici p_2 a q_2. Pro $i \ge 1$ máme

$$\begin{pmatrix} q_{i+1} & p_{i+1} \\ q_i & p_i \end{pmatrix} = \begin{pmatrix} a_i & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & p_1 \\ q_0 & p_0 \end{pmatrix}, \text{ a}$$

$$\begin{pmatrix} q_{i+2} & p_{i+2} \\ q_{i+1} & p_{i+1} \end{pmatrix} = \begin{pmatrix} a_{i+1} & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_2 & p_2 \\ q_1 & p_1 \end{pmatrix},$$

takže z platnosti vztahu pro dané $i < n$ plynou, položíme-li $\alpha = [a_2, \dots, a_{i+1}]$, rovnosti

$$\frac{p_{i+2}}{q_{i+2}} = \frac{p_1 + p_2\alpha}{q_1 + q_2\alpha} = \frac{\alpha p_0 + p_1(1 + \alpha a_1)}{\alpha q_0 + q_1(1 + \alpha a_1)} = \frac{p_0 + p_1(a_1 + \alpha^{-1})}{q_0 + q_1(a_1 + \alpha^{-1})}$$

Ovšem $a_1 + \alpha^{-1} = a_1 + [a_2, \dots, a_{i+1}]^{-1} = [a_1, \dots, a_{i+1}]$. □

Důsledek. *Buď $\vartheta = [a_0, a_1, \dots, a_n]$ konečný čistý řetězový zlomek. Položme $C_i = [a_0, a_1, \dots, a_{i-1}]$ pro každé i, $1 \le i \le n + 1$, a zvolme $k \in \mathbb{Z}$ a $\varepsilon \in \{0, 1\}$ tak, aby $2k + \varepsilon = n$. Potom $C_{n+1} = \vartheta$,*

$$C_2 > C_4 \cdots > C_{2(k+\varepsilon)} > C_{2k+1} > \cdots > C_3 > C_1,$$

a $|C_i - C_{i+1}| < i^{-2}$, *kdykoliv* $1 \le i < n$.

Je-li $n = 1$, je $\vartheta = a_0 + a_1^{-1} > a_0$. Je-li $n \ge 2$, je $a_0 + a_1^{-1} > \vartheta > a_0$.

Důkaz. Položme $q_0 = 0$, $q_1 = 1 = p_0$ a $p_1 = a_0$. Definujme rekurzivně $p_{i+1} = a_i p_i + p_{i-1}$, $q_{i+1} = a_i q_i + q_{i+1}$, kde $1 \leq i \leq n$. Podle lemmatu pro $0 \leq i \leq n$ platí

$$\frac{p_{i+1}}{q_{i+1}} = [a_1, \ldots, a_i]^{-1} + p_1 = [a_0, a_1, \ldots, a_i] = C_{i+1}.$$

Tudíž $C_i = p_i/q_i$, $1 \leq i \leq n+1$, a lze použít bod (iv) lemmatu 5.4. Bod (ii) téhož lemmatu dává $|C_i - C_{i+1}| = (q_i q_{i+1}) < i^{-2}$. Máme totiž $q_i \geq i$ pro každé i, $1 \leq i \leq n$, neboť $1 \leq q_1 < q_2 < \cdots < q_{n+1}$ jsou čísla celá.

Případ $n = 1$ je zřejmý. Je-li $n \geq 2$, tak z nerovnosti plyne $a_0 + a_1^{-1} = C_2 > \vartheta > C_1 = a_0$. □

Je-li $\alpha_1 < \alpha_2 < \alpha_3 \cdots$ rostoucí posloupnost reálných čísel a $H > \alpha_i$ pro každé $i \geq 1$, nazýváme H horní závorou posloupnosti $\alpha_1, \alpha_2, \ldots$. Je známo, že posloupnost $\{\alpha_i\}$ má limitu $\vartheta \leq H$. Obdobné tvrzení platí pro klesající posloupnosti. Obě tato tvrzení využijeme v následující větě. (Její důkaz lze zkrátit využitím vlastností cauchyovských posloupností. Znalost tohoto pojmu však nepředpokládáme.)

Věta. *Buď $a_0, a_1, a_2, \ldots$ nekonečná posloupnost celých čísel, přičemž $a_i \geq 1$ pro $i \geq 1$. Položme $C_i = [a_0, \ldots, a_{i-1}]$ pro každé $i \geq 1$. Pak existuje právě jedno reálné číslo ϑ takové, že*

$$C_2 > C_4 > C_6 > \cdots > \vartheta > \cdots > C_5 > C_3 > C_1.$$

Důkaz. Důsledek výše lze použít na libovolný počáteční úsek $C_1, C_2, \ldots, C_{n+1}$ posloupnosti $\{C_i\}$. Víme tedy, že posloupnost $\{C_{2i}\}$, $i \geq 1$, je klesající a že každé C_{2j+1}, $j \geq 0$, je dolní závorou posloupnosti $\{C_{2i}\}$. Tato posloupnost má tedy limitu ϑ, která je horní závorou posloupnosti $\{C_{2j+1}\}$, $j \geq 0$. Posloupnost $\{C_{2j+1}\}$ má tedy limitu $\vartheta' \leq \vartheta$. Pokud nějaké γ splňuje $C_{2i} > \gamma > C_{2i+1}$ pro každé $i \geq 1$, je γ dolní závorou posloupnosti $\{[C_{2i}\}$, takže $\gamma \leq \vartheta$, a současně horní závorou posloupnosti $\{C_{2j+1}\}$, takže $\gamma \geq \vartheta'$.

Pro každé $i \geq 1$ podle důsledku výše platí $|C_i - C_{i+1}| < 1/i^2$. Je-li i sudé, tak $C_i > \vartheta \geq \vartheta' > C_{i+1}$, takže $|\vartheta - \vartheta'| < 1/i^2$ pro každé sudé $i \geq 1$. To ale znamená $\vartheta = \vartheta'$. □

Každá celočíselná posloupnost $a_0, a_1, \ldots$, kde $a_1, a_2, \ldots$ jsou kladná, tedy určuje nějaké reálné číslo $\vartheta = [a_0, a_1, a_2, \ldots]$. Toto číslo nazýváme *řetězovým zlomkem* posloupnosti $a_0, a_1, a_2, \ldots$. Pro rozlišení také někdy říkáme, že $\vartheta = [a_0, a_1, a_2, \ldots]$ je *nekonečným řetězovým zlomkem.*

5.6 Jednoznačnost řetězových zlomků

Konečný (čistý) řetězový zlomek $[a_0, a_1, \ldots, a_n] = \vartheta$ můžeme pro každé i, $0 \leq i \leq n$, zapsat ve tvaru $[a_0, \ldots, a_{i-1}, \vartheta_i]$, kde $\vartheta_i = [a_i, \ldots, a_n]$. Pokud bychom chtěli odvodit pro hodnoty ϑ_i rekurzivní vzorec, dostaneme

$$\vartheta_0 = \vartheta, \vartheta_{i+1} = \frac{1}{\vartheta_i - a_i}, 0 \leq i < n,$$

neboť $\vartheta_i = a_i + [a_{i+1}, \dots, a_n]^{-1} = a_i + \vartheta_{i+1}^{-1}$. Ověřme, že podobně lze postupovat i v případě nekonečných řetězových zlomků.

Lemma. *Ať $a_0, a_1, \dots$ jsou celá čísla, která jsou pro $i \geq 1$ kladná. Položme $\vartheta_i = [a_i, a_{i+1}, \dots]$ pro každé $i \geq 0$, a ať $\vartheta = \vartheta_0$. Pak $\vartheta = [a_0, \dots, a_{i-1}, \vartheta_i]$ pro každé $i \geq 0$, a jsou splněny rekurzivní vztahy*

$$\vartheta_0 = \vartheta, \vartheta_{i+1} = \frac{1}{\vartheta_i - a_i}, i \geq 0.$$

Důkaz. Pro $i = 0$ není co dokazovat. Ať $i = 1$. Hodnota ϑ_i je limitou racionálních čísel $C_j = [a_1, a_2, \dots, a_j]$, zatímco ϑ_0 je limitou racionálních čísel $[a_0, a_1, \dots, a_j] = a_0 + C_j^{-1}$. Podle věty 5.5 $\vartheta_1 > C_1 = a_1 \geq 1$, takže z $\lim C_j = \vartheta_1$ plyne $\lim C_j^{-1} = \vartheta_1^{-1}$, a proto je $\vartheta_0 = \lim(a_0 + C_j^{-1})$ rovno $a_0 + \vartheta_1^{-1}$, z čehož $\vartheta_1 = (\vartheta_0 - a_0)^{-1}$. Pro $i = 1$ je tvrzení dokázáné.

Tím jsem ovšem dokázali $\vartheta_i = a_i + \vartheta_{i+1}^{-1}$ pro každé $i \geq 0$, neboť $\vartheta_i = [a_i, a_{i+1}, \dots]$ a $\vartheta_{i+1} = [a_{i+1}, a_{i+2}, \dots]$. Indukční krok pak spočívá v rovnosti

$$\vartheta = [a_0, \dots, a_{i-1}, \vartheta_i] = [a_0, \dots, a_{i-1}, a_i + \vartheta_{i+1}^{-1}] = \\ = [a_0, \dots, a_{i-1}, a_i, \vartheta_{i+1}].$$

□

Pro každý konečný řetězový zlomek $[a_0, \dots, a_{n-1}, a_n]$, kde $a_n > 1$, zjevně platí

$$[a_0, \dots, a_{n-1}, a_n] = [a_0, \dots, a_{n-1}, a_n^{-1}, 1].$$

Proto je třeba v úvahách o jednoznačnosti zápisu jisté opatrnosti.

Poznámka. Je zřejmé, že z $[a_0, \dots, a_{n-1}, a_n] = [a_0, \dots, a_{n-1}, b_n]$ plyne $a_n = b_n$. Podobně je zřejmé, že vždy platí

$$[a_0, \dots, a_{n-1}, a_n] \neq [a_0, \dots, a_{n-1}, a_n, b_n],$$

pokud $a_1, \dots, a_n, b_n \in \mathbb{R}^+$ a $a_0 \in \mathbb{R}$.

(Případný formální důkaz lze provést indukcí dle n, vycházeje z $[a_0, \dots, a_{n-1}, a_n] = a_0 + [a_1, \dots, a_n]^{-1}$.)

Tvrzení. *Každé racionální číslo lze jednoznačně vyjádřit čistým konečným řetězovým zlomkem $[a_0, \dots, a_n]$, kde $a_n \neq 1$ nebo $n = 0$. Nekonečný řetězový zlomek udává vždy číslo iracionální. Každé iracionální číslo lze vyjádřit řetězovým zlomkem nejvýše jedním způsobem.*

Důkaz. Buď $a_0, a_1, \dots$ posloupnost celých čísel, která může být konečná nebo nekonečná. Předpokládejme, že hodnoty $a_1, a_2, \dots$ jsou všechny kladné. Je-li posloupnost konečná, označme n její nejvyšší člen, jinak položme $n = \infty$. Hodnotu řetězového zlomku určeného posloupností $a_0, a_1, \dots$ označme ϑ a pro $i < n$ rekurzivně definujme $\vartheta_0 = \vartheta$, $\vartheta_{i+1} = (\vartheta_i - a_i)^{-1}$. Z lemmatu výše (a z textu na začátku oddílu) víme, že $\vartheta = [a_0, a_1, \dots, a_{i-1}, \vartheta_i]$.

Buď nyní $a_0', a_1', \dots$ opět konečná nebo nekonečná posloupnost celých čísel, kde $a_i' \geq 1$ pro $i \geq 1$. Definujme obdobně n', ϑ' a ϑ_i'. Předpokládejme, že posloupnosti $a_0, a_1, \dots$ a $a_0', a_1', \dots$ jsou různé.

Je-li $a_0, a_1, \ldots$ posloupnost konečná a taková, že je počátkem posloupnosti $a'_0, a'_1, \ldots$, tak máme

$$\vartheta = [a_0, \ldots, a_n] \neq [a_0, \ldots, a_n, \vartheta_{n+1}] = \vartheta'.$$

Můžeme tedy předpokládat, že existuje $k \leq \min\{n, n'\}$ takové, že je $a_k \neq a'_k$. Zvolme $k \geq 0$ nejmenší možné. Pak

$$\vartheta = [a_0, \ldots, a_{k-1}, \vartheta_k] \neq [a_0, \ldots, a_{k-1}, \vartheta'_k] = \vartheta'$$

právě když $\vartheta_k = [a_k, a_{k+1}, \ldots] \neq [a'_k, a'_{k+1}, \ldots] = \vartheta'_k$ (viz poznámka před tvrzením). Stačí tedy dokázat $\vartheta \neq \vartheta'$ pro případ $k = 0$. Bez újmy na obecnosti můžeme předpokládat $a'_0 > a_0$. Je-li $n = 0$, je $\vartheta' \geq a'_0 > a_0 = \vartheta$. Je-li $n = 1$, je $a_1 \geq 2$, takže $\vartheta' \geq a'_0 \geq a_0 + 1 > a_0 + a_1^{-1} = \vartheta$. Je-li $n \geq 2$, tak podle důsledku 5.5 máme $\vartheta' \geq a'_0 \geq a_0 + 1 \geq a_0 + a_1^{-1} > \vartheta$.

Dokázali jsme, že různé posloupnosti určují různé hodnoty ϑ. Protože podle tvrzení 5.4 lze každé racionální číslo vyjádřit řetězovým zlomkem, nemůže díky dokázané jednoznačnosti být žádné racionální číslo vyjádřeno nekonečným řetězovým zlomkem. □

5.7 Řetězové zlomky a dobré aproximace

V tomto oddíle navážeme na důsledek 5.3, podle kterého je každé iracionální číslo ϑ rovno $\lim p_i/q_i$, kde p_i/q_i je posloupnost dobrých aproximací čísla ϑ, $i = 1$.

Lemma. *Buď $a_0, \ldots, a_n \in \mathbb{R}$, $n \geq 1$ přičemž $a_1, \ldots, a_n$ jsou kladná a $a_1 > 1$. Pak*

$$(a_0 + 1) - [a_1, a_2, \ldots, a_n]^{-1} = [a_0, 1, a_1 - 1, a_2, \ldots, a_n].$$

Důkaz. Položme $\alpha = [a_1, a_2, \ldots, a_n]$. Levá strana je rovna $a_0 + 1 - \alpha^{-1}$, zatímco pravá strana dává $a_0 + (1 + [a_1 - 1, a_2, \ldots, a_n]^{-1})^{-1} = a_0 + (1 + (\alpha - 1)^{-1})^{-1}$. Ovšem

$$(1 + (\alpha - 1)^{-1})^{-1} = \frac{1}{1 + \dfrac{1}{\alpha - 1}} = \frac{\alpha - 1}{(\alpha - 1) + 1} = \frac{\alpha - 1}{\alpha} = 1 - \alpha^{-1}.$$

□

Věta. *Každé iracionální číslo ϑ lze vyjádřit jediným způsobem jako řetězový zlomek $\vartheta = [a_0, a_1, \ldots]$. Nechť $q_0 = 0$, $q_1 = 1$, a ať $p_0 = 1$, je-li $a_1 > 1$, a $p_0 = -1$, je-li $a_1 = 1$. V prvním případě položme $p_1 = a_0$ a $p_{i+1} = a_i p_i + p_{i-1}$, $q_{i+1} = a_i q_i + q_{i-1}$ pro každé $i \geq 1$. V druhém případě položme $p_1 = a_0 + 1$, $p_2 = a_2 a_0 + a_2 + a_0$, $q_2 = a_2 + 1$, a $p_{i+1} = a_{i+1} p_i + p_{i-1}$, $q_{i+1} = a_{i+1} q_i + q_{i-1}$ pro každé $i \geq 2$. Pak p_i/q_i, $i \geq 1$, jsou všechny dobré aproximace čísla ϑ.*

Důkaz. Vyjděme od posloupnosti dobrých aproximací p_i/q_i čísla ϑ, $i \geq 1$. Definujme $q_0 = 0$ a ať $p_0 = 1$ pro $\vartheta > p_1$ a $p_0 = -1$ pro $\vartheta < p_1$. Naším cílem je nalézt celá čísla $a_0, a_1 \ldots$ tak, aby tato posloupnost vyjadřovala ϑ jako řetězový zlomek, a současně aby byly splněny další uvedené vztahy. Více není

třeba, neboť víme z tvrzení 5.6, že iracionální číslo lze vyjádřit jako řetězový zlomek nejvýše jedním způsobem.

Podle tvrzení 5.4 existují celá kladná čísla $b_1, b_2, \dots$ taková, že $q_{i+1} = b_i q_i + q_{i-1}$ a $p_{i+1} = b_i p_i + p_{i-1}$, pro všechna $i \geq 1$. Podle lemmatu 5.5 máme $b_1 = q_2 \geq 2$ a

$$\frac{p_{i+1}}{q_{i+1}} = p_0[b_1, \dots, b_i]^{-1} + p_1 \text{ pro všechna } i \geq 1.$$

Je-li $p_0 = 1$, položíme $a_0 = p_1$ a $a_i = b_i$ pro $i \geq 1$. Pak $p_{i+1}/q_{i+1} = [a_0, \dots, a_i]$ a $\vartheta = \lim p_i/q_i = [a_0, a_1, a_2, \dots]$.

Zbývá vyřešit případ $p_0 = -1$. Podle lemmatu výše je $p_{i+1}/q_{i+1} = p_1 - [b_1, \dots, b_i]^{-1} = [p_1 - 1, 1, b_1 - 1, b_2, \dots, b_i]$.

Položíme-li $a_0 = p_1 - 1$, $a_1 = 1$, $a_2 = b_1 - 1$ a $a_i = b_{i-1}$ pro $i \geq 3$, dostaneme $\vartheta = [a_0, a_1, a_2, \dots]$, $p_2 = b_1 p_1 + p_0 = (a_2 + 1)(a_0 + 1) - 1 = a_2 a_0 + a_2 + a_0$ a $q_2 = b_1 = a_2 + 1$.

Rekurzivní vzorce pro $i \geq 2$ pak vyjadřují záměnu b_i za a_{i+1}. □

Sloučením dosažených výsledků dostáváme i určitý návod, jak posloupnost a_0, a_1,... počítat.

Tvrzení. *Buď ϑ iracionální číslo. Položme $\vartheta_0 = \vartheta$ a definujeme rekurzivně $a_i = [\vartheta_i]$, $\vartheta_{i+1} = (\vartheta_i - a_i)^{-1}$. Pak $\vartheta = [a_0, a_1, \dots]$.*

Důkaz. Víme, že existují celá čísla $a_0, a_1, \dots$ taková, že $\vartheta = [a_0, a_1, \dots]$. Položme $\vartheta_i = [a_i, a_{i+1}, \dots]$. Podle lemmatu 5.6 je $\vartheta_{i+1} = (\vartheta_i - a_i)^{-1}$. Podle věty 5.4 je $a_i + 1 \geq a_i + a_{i+1}^{-1} > \vartheta_i > a_i$. Proto $a_i = [\vartheta_i]$. □

Příklad. Buď $h \geq 1$ celé. Položme $\vartheta = (h + \sqrt{h^2+4})/2$. Pak $h < \sqrt{h^2+4} < h + 2$ implikuje $h < \vartheta < h + 1$, takže $[\vartheta] = h$. Ukážeme, že $\vartheta_i = \vartheta$ a $a_i = h$ platí pro všechny $i \geq 0$. Je-li to pravda pro $i \geq 0$, dostáváme $\vartheta_{i+1} = (\vartheta - h)^{-1} = 2(\sqrt{h^2+4} - h)^{-1} = 2(\sqrt{h^2+4} + h)(h^2 + 4 - h^2)^{-1} = \vartheta$. Dokázali jsme $\frac{h+\sqrt{h^2+4}}{2} = [h, h, \dots]$, pro každé $h \geq 1$.

Zkoumejme ještě, co lze říci o posloupnostech p_i a q_i, kde $p_0 = 1$, $p_1 = h$, $p_{i+1} = hp_i + p_{i-1}$, zatímco $q_0 = 0$, $q_1 = 1$ a $q_{i+1} = hq_i + q_{i-1}$. Vidíme, že $q_i = p_{i-1}$ pro každé $i \geq 1$, takže platí $\lim \frac{p_i}{p_{i-1}} = (h + \sqrt{h^2+4})/2$.

Speciálně pro $h = 1$ je posloupnost p_i rovna Fibonacciho posloupnosti $1, 1, 2, 3, 5, 8, 13, 21, \dots$, takže poměr dvou následujících členů se limitně blíží $(1 + \sqrt{5}/2)$ (což je poměr tzv. *zlatého řezu*).