STŘEDOŠKOLSKÁ ODBORNÁ ČINNOST

Obor č. 1: Matematika a statistika

Isogenie v kryptografii

Zdeněk Pezlar
Jihomoravský kraj

Brno 2021

STŘEDOŠKOLSKÁ ODBORNÁ ČINNOST

Obor č. 1: Matematika a statistika

Isogenie v kryptografii

Isogeny Based Cryptography

Autor: Zdeněk Pezlar
Škola: Gymnázium Brno, třída Kapitána Jaroše, p. o.
Kraj: Jihomoravský
Konzultant: Mgr. Vojtěch Suchánek

Prohlášení

Prohlašuji, že jsem svou práci SOČ vypracoval samostatně a použil jsem pouze prameny a literaturu uvedené v seznamu bibliografických záznamů. Prohlašuji, že tištěná verze a elektronická verze soutěžní práce SOČ jsou shodné. Nemám závažný důvod proti zpřístupňování této práce v souladu se zákonem č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) v platném znění.

V Brně dne: ..... Podpis: .....

jihomoravský kraj

Poděkování

++Tato práce byla vypracována za finanční podpory JMK.

Abstrakt

abstrakt

Klíčová slova

isogenie;klíčové slovo.

Abstract

abstrakt

Key words

isogenie;klíčové slovo.

Obsah

Úvod

celkem úvod

Použitá značení

$a \mid b$	a dělí b
$\frac{1}{a}$	multiplikativní inverz a, tj. a^{-1}
$\mathcal{D}(a, b)$	největší společný dělitel a, b
$\nu_p(n)$	p-adická valuace n
$\left(\frac{a}{p}\right)$	Legendreův symbol a vzhledem k p
$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$	množina přirozených, celých, racionálních, reálných, komplexních čísel
$\mathbb{Z}_d$	okruh zbytků modulo d
$\mathbb{F}_q$	konečné těleso s q prvky
$\overline{K}$	algebraický uzávěr K
$K^\times$	multiplikativní podrupa K
K^*	$K \setminus \{0\}$
$\mathbb{P}^n(K)$	projektivní prostor nad K o rozměru $n+1$
$E(K)$	množina bodů křivky E nad K
$\#E(K)$	počet bodů na křivce E nad konečným tělesem K
$\mathcal{O}$	bod v nekonečnu křivky E
$[n]_E, [n]$	násobení n na křivce E
π, π_E	Frobeniův morfismus
$\widehat{\phi}$	isogenie duální k ϕ
$\deg \phi$	stupeň isogenie ϕ
$\ker \phi$	jádro isogenie ϕ
$\# \ker \phi$	velikost jádra isogenie ϕ
E/G	obraz E v separabilní isogenii s jádrem G
$E[n]$	n-torze křivky E
$\tilde{E}$	twist křivky E
$\mathrm{End}(E)$	okruh endomorfismů E

$\mathrm{End}^0(E)$	algebra endomorfismů E
$\mathrm{Tr}\,\phi, \mathrm{Tr}\,\alpha$	stopa endomorfismu ϕ, stopa $\alpha \in \mathrm{End}^0(E)$
$\mathrm{N}\,\alpha$	norma $\alpha \in \mathrm{End}^0(E)$
$\widehat{\alpha}$	Rosatiho involuce aplikovaná na α
$j(E)$	j-invariant křivky E
$R[x]$	okruh polynomů s koeficienty nad okruhem R
$K(a_1, \ldots, a_n)$	nejmenší podtěleso L, které obsahuje těleso K i prvky $a_1, \ldots, a_n \in L$
$[K : L]$	stupeň rozšíření tělesa K nad L, tj. dimenze vektorového prostoru K/L
$\alpha(x)$	lineární transoformace $x \mapsto \alpha x$ aktující na $\mathbb{Q}(\theta)$
M_α	matice popisující $\alpha(x)$
TrM	stopa matice M
$\det M$	determinant matice M
$Tr_K(\alpha)$	stopa prvku α v K
$N_K(\alpha)$	norma prvku α v K
$\mathcal{O}_K$	okruh celých algebraických čísel tělesa K
$Cl(\mathcal{O})$	grupa tříd ideálů pořádku $\mathcal{O}$
$h_\mathcal{O}$	řád grupy $Cl(\mathcal{O})$
(a)	hlavní ideál generovaný prvkem a
$\frac{\mathfrak{a}}{m}$	lomený ideál $\frac{\mathfrak{a}}{m}$
$\left(\frac{a}{m}\right)$	hlavní lomený ideál $\frac{(a)}{m}$
$N_\mathcal{O}(\mathfrak{a})$	norma ideálu $\mathfrak{a} \subseteq \mathcal{O}$, tj. $\lvert\mathcal{O}/\mathfrak{a}\rvert$
$\mathfrak{a} + \mathfrak{b}$	součet ideálů $\mathfrak{a}$ a $\mathfrak{b}$
$\mathfrak{a} \cdot \mathfrak{b}$	součin ideálů $\mathfrak{a}$ a $\mathfrak{b}$
$\mathfrak{a} \vert \mathfrak{b}$	ideál $\mathfrak{a}$ dělí ideál $\mathfrak{b}$
G/H	faktorgrupa G podle H
$\deg f$	stupeň polynomu f
f'	derivace f
$f \in O(g)$	f roste asymptoticky nejvýše stejně rychle jako g
$f \in \Theta(g)$	f roste asymtoticky stejně rychle jako g
$f \in \Omega(g)$	f roste asymtoticky alespoň tak rychle jako g

Kapitola 1

Eliptické křivky

V naší první kapitole se budeme procházet světem isogenií eliptických křivek a učit se s nimi pracovat. Kořeny této teorie sahají hluboko do algebraické geometrie, jejíž studium nabízí podrobnější vhled do machinací „pod kapotou“ textu, který následuje. Pro porozumění této kapitoly její znalost ale nevyžadujeme, čtenář si bohatě vystačí se znalostmi abstraktní algebry, viz například [32]. Budeme postupovat volně dle [39], nicméně další vhodný úvodní materiál se nachází na [10]. Ne vždy budeme uvádět důkazy tvrzení, neboť jsou mnohdy příliš pokročilé či technické, v takových případech se odkážeme na relevantní literaturu.

1.1 Základy

Po celou dobu budeme pracovat nad projektivním prostorem nad uzávěrem tělesa K, což je množina bodů v $\overline{K}^n$, kde dva body považujeme za ekvivalentní, pokud leží v přímce s počátkem, můžeme proto místo jednotlivých bodů pracovat s přímkami procházejícími skrz počátek. Chtěli bychom, aby se každé dvě $n-1$ rozměrné roviny protínaly, a s tím máme problém pouze pokud protínáme dvě rovnoběžné. V každém směru si tak můžeme definovat projektivní prostor stupně $n-1$ v nekonečnu, kde se protínají rovnoběžné roviny.

Definice 1.1.1. Buďte K těleso a n přirozené číslo. *Projektivní prostor* $\mathbb{P}^n(\overline{K})$ definujeme jako množinu tříd nenulových vektorů $(a_0, \dots, a_n) \in \overline{K}^{n+1}$ s relací ekvivalence $(a_0, \dots, a_n) \sim (b_0, \dots, b_n)$, pokud existuje $\lambda \in \overline{K}$, že $(a_0, \dots, a_n) = \lambda(b_0, \dots, b_n)$. Tyto třídy ekvivalence budeme značit $(a_0 : \cdots : a_n)$ a nazývat *body*.

Pokud je některé z a_i rovné nule, získáme $n-1$ rozměrný vektorový prostor „v nekonečnu“.

Projektivní prostor $\mathbb{P}^2(\mathbb{R})$ je známý jako projektivní rovina. Každé dvě přímky se protínají v jednom bodě, přičemž rovnoběžné přímky se protínají v bodě v nekonečnu v daném směru. Přímky procházející počátkem tak můžeme ztotožnit s jejich průsečíkem s rovinou neprocházíjí počátkem, tedy každé takové přímce přiřadíme třídu, ve které leží její příslušný

průsečík. Přímky s touto rovnou rovnoběžné, které v ní neleží, ji protínají v nekonečnu, a přiřadíme jim body v nekonečnu v příslušném směru.

Poznámka. Je zajímavé uvážit souvislost projektivních prostorů a barycentrických souřadnic, kde je každý bod vyjádřen jako vážený průměr vrcholů referenčního simplexu. Tyto souřadnice jsou též homogenní a každé dvě přímky se protínají, byť některé v nekonečnu, takové body mají součet vah roven 0. Můžeme o barycentrických souřadnicích tedy přemýšlet jako o projektivním prostoru s jiným základem.

Připomeňme si pak definici eliptické křivky. Často se definuje jako nesingulární projektivní křivka genu 1 v $\overline{K}^3$, tj. jako:

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

kde $a_i \in K$. My si ale pro naše účely definici zúžíme. Konkrétně se budeme pohybovat nad tělesy, jejichž charakteristika není 2 ani 3. Pro tělesa s charakteristikou 2 či 3 se často v jiných kontextech eliptické křivky definovat hodí, nám však jejich vyloučení značně zjednoduší práci. Nejprve totiž můžeme substitucí $Y \mapsto Y - \frac{a_1X+a_3Z}{2}$ zapsat naši křivku jako:

$$\begin{aligned} Y^2Z - \left(\frac{a_1X + a_3Z}{2}\right)^2 Z &= X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3, \\ Y^2Z &= X^3 + \frac{b_2}{4}X^2Z + \frac{b_4}{2}XZ^2 + \frac{b_6}{4}Z^3, \end{aligned}$$

kde $b_2 = a_1^2 + 4a_2, b_4 = a_1a_3 + 2a_4$ a $b_6 = a_3^2 + 4a_6$. Substituce $X \mapsto X - \frac{b_2}{12}Z$ dále zjednodušuje naši křivku:

$$\begin{aligned} Y^2Z &= \left(X - \frac{b_2}{12}Z\right)^3 + \frac{b_2}{4}\left(X - \frac{b_2}{12}Z\right)^2 Z + \frac{b_4}{2}\left(X - \frac{b_2}{12}Z\right)Z^2 + \frac{b_6}{4}Z^3, \\ Y^2Z &= X^3 + \left(\frac{24b_4 - b_2^2}{48}\right)XZ^2 + \left(\frac{b_2^2 + 216b_6 - 36b_2b_4}{864}\right)Z^3. \end{aligned}$$

Naši křivku proto můžeme zapsat ve tvaru:

$$Y^2Z = X^3 + aXZ^2 + bZ^3,$$

kde $a, b \in K$ jsou taková, že diskriminant této křivky, $4a^3 + 27b^2$, je nenulový, protože lineární transformace zachovají (ne)singularitu křivky.

Definice 1.1.2. Mějme K těleso charakteristiky různé od 2 a 3. Pro $a, b \in K$, že $4a^2+27b^3 \neq 0$, definujeme v $\mathbb{P}^2(\overline{K})$ *eliptickou křivku* jako množinu bodů $(X : Y : Z) \in \overline{K}^3$ splňující:

$$Y^2Z = X^3 + aXZ^2 + bZ^3.$$

Definice 1.1.3. Pokud všechny koeficienty eliptické křivky E náleží do tělesa K, značíme ji E/K.

Průsečíky naší křivky s přímkou $Z = 0$ nutně mají i X-ovou souřadnici nulovou, všechny jsou proto reprezentovány třídou $(0 : 1 : 0)$. V opačném případě můžeme přejít na proměnné $x := X/Z, y := Y/Z$, tedy bod $(x : y : 1)$, čímž získáme křivku ve známém *afinním*, v literatuře často uváděneném i jako *Weierstrassově*, tvaru:

$$y^2 = x^3 + ax + b.$$

Množina bodů na naší křivce tedy sestává z bodů $(x, y) \in K^2$ na naší afinní křivce spolu s bodem v nekonečnu $\mathcal{O} = (0 : 1 : 0)$, jenž je exklusivní její projektivní variantě.

Definice 1.1.4. Množinu všech bodů E se souřadnicemi nad K (společně s $\mathcal{O}$) budeme značit $E(K)$ a pokud K je konečné těleso, počet prvků $E(K)$ budeme značit $\#E(K)$.

Počet bodů na E nad konečným tělesem $\mathbb{F}_q$ je shora ohraničen číslem $2q+1$, protože pro každé $x \in \mathbb{F}_q$ existují v $\mathbb{F}_q$ nejvýše 2 odmocniny z x^3+ax+b, a poslední bod do počtu je $\mathcal{O}$. V $\mathbb{F}_q$ leží právě $\frac{q+1}{2}$ čtverců, tudíž za předpokladu, že $x^3 + ax + b$ pokrývá $\mathbb{F}_q$ rovnoměrně, bychom na E očekávali okolo q bodů, společně s bodem v nekonečnu $q + 1$. Roku 1933 tento odhad Helmut Hasse dokázal, tedy skutečně se $\#E(\mathbb{F}_q)$ nepříliš liší od $q + 1$.

Věta 1.1.5. *(Hasse) Nechť $E/\mathbb{F}_q$ je eliptická křivka. Pak:*

$$|q + 1 - \#E(\mathbb{F}_q)| \leqslant 2\sqrt{q}.$$

Důkaz je k nalezení v [34, Thm. V.1.1]. Již zde, ještě na začátku naší poutě, musíme a priori brát jako platný jeden z nejdůležitějších výsledků ohledně eliptických křivek, ne však bez důvodu. Většina učebních textu jej dokáže v průběhu studia algebraické geometrie, v našem případě bychom potřebovali udělat poměrně velkou odbočku. Pokud tuto větu budeme jednoduše předpokládat, mnoho problémů si podstatně zjednodušíme.

Definice 1.1.6. Pod bodem $P \in E$ rozumíme $P = (x, y) \in E(\overline{K})$.

Podívejme se nyní na eliptickou křivku E geometricky, tedy v rovině vyznačme všechny body, které na ní leží. Je zjevné, že E je symetrická podle osy x, definujme proto k $P \in E$ opačný bod $-P \in E$ jako obraz P podle osy x. Pokud bychom na bodech naší křivky definovali součet, přirozeně bychom chtěli, aby součet P a $-P$ byl $\mathcal{O}$.

OBrázky

Řekneme-li, že tečna k E ji protíná ve dvou identických bodech, pak každá přímka protíná E v právě třech bodech včetně multiplicity, průsečíky lineární rovnice s kubickou křivkou budou i s případným bodem v nekonečnu tři. Speciálně tečna v bodě s $y = 0$ tento bod protíná dvakrát a ten třetí je bod v nekonečnu E. Přichází tedy na mysl definice součtu $+$ na E taková, že součet každých tří bodů v přímce je $\mathcal{O}$. Pokud přímka procházející $P, Q \in E$ protíná E potřetí v R, definujeme tedy $P + Q = -R$. Pro součet bodů $P, Q \in E$ můžeme poté odvodit několik klíčových vlastností:

(i) $P + Q = Q + P$,

(ii) $(P + Q) + R = P + (Q + R)$,

(iii) $P + \mathcal{O} = P$,

(iv) $P + (-P) = \mathcal{O}$.

Rovnosti (i),(iii) a (iv) jsou dle naší definice sčítání intuitivně jasné, potíže však nastanou s bodem (ii), který je notoricky obtížné dokázat. Jeho klasický důkaz užívá pokročilejších metod algebraické geometrie, konkrétně Riemann-Rochovu větu, či větu Cayley-Bacharacha, která u dvou kubických křivek protínajících se v 9 bodech zaručuje, že každá jiná kubická křivka procházející osmi z nich obsahuje i ten poslední. Tato poslední věta má aplikace i mimo eliptické křivky, klasické výsledky projektivní geometrie jako Pappova či Pascalova věta z ní totiž snadno plynou. Poměrně elementární, byť výpočetně zdlouhavý důkaz Cayley-Bacharovy věty i jejích zmíněných důsledků se dá najít v [42, Sec. 2.3]

Při takto definovaném součtu můžeme s body na E pracovat jako s abelovskou grupou se sčítáním $+$ a neutrálním prvkem $\mathcal{O}$. Samozřejmě součet dvou bodů dokážeme za pomocí analytické geometrie přímo spočíst. Přímka procházející dvěma různými body $P = (x_1, y_1)$ a $Q = (x_2, y_2)$ v rovině je daná rovnicí $y = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) + y_1$. Známe-li dva průsečíky této přímky s E, tedy P a Q, dosazením do rovnice E jsme schopni spočíst jejich třetí průsečík, bod $-(P + Q)$.

Jediné, co nám chybí ke spokojenosti, je najít dvojnásobek bodu P, omezme se na případ P neležící na ose x. Tečna k E v bodě P je přímka PQ, když se Q limitně blíží k P. Sklon této přímky je tedy dán implicitní derivací $y^2 = x^3 + ax + b$ v bodě $P = (x_1, y_1)$, tedy $2y_1y' = 3x_1^2 + a$. Tečna k E v P je pak určena vztahem $2y_1(y - y_1) = (3x_1^2 + a)(x - x_1)$. Z této rovnost vyjádříme y a dosadíme do rovnice přímky E, kde je x_1 dvojnásobný kořen. Můžeme proto vyfaktorizovat člen $(x - x_1)^2$ a jako třetí lineární člen získat řešení pro $-(P + P)$.

Předchozí úvahy shrnuje následující tvrzení:

Věta 1.1.7. *Buďte $P = (x_1, y_1), Q = (x_2, y_2)$ afinní body na křivce $E : y^2 = x^3 + ax + b$, přičemž $P \neq -Q$. Pak $P + Q = (x_3, y_3)$ je daný:*

$$x_3 = \lambda^2 - x_1 - x_2,$$
$$y_3 = -\lambda x_3 - y_1 + \lambda x_1,$$

kde:

$$\lambda = \begin{cases} \frac{y_2 - y_1}{x_2 - x_1}, & \text{pokud } x_1 \neq x_2, \\ \frac{3x_1^2 + a}{2y_1}, & \text{pokud } x_1 = x_2. \end{cases}$$

Úplný výpočet s dovolením neuvádím. Je možné dokázat asociativitu sčítání i tím, že pro body $P = (x_1, y_1), Q = (x_2, y_2)$ a $R = (x_3, y_3)$ spočteme bod $(P + Q) + R$ a ukážeme, že je symetrický ve dvojicích (x_1, x_3) a (y_1, y_3), případně že je přímo roven $P + (Q + R)$. Tyto výpočty nejsou prakticky proveditelné bez výpočetních přístrojů, nicméně za pomocí například programu Wolfram Mathematica se můžeme přesvědčit, že asociativita platí.

Pro zkrácení zápisu píšeme skalární násobky bodů, jinak řečeno $P + \cdots + P$, následovně:

Definice 1.1.8. Mějme bod $P \in E$. Pak pro n přirozené definujeme jeho n-násobek:

$$[n]_E P = \underbrace{P + \cdots + P}_{n},$$

příčemž definujeme $[0]_E P = \mathcal{O}$ a pro $n < 0 : [n]_E P = [-n]_E(-P)$.

Díky asociativitě sčítání je bod $[n]_E P$ dobře definovaný. Pokud bude z kontextu jasná eliptická křivka, nad kterou pracujeme, budeme značit násobení skalárem pouze $[n]P$. Pojďme se pokusit n-násobek bodu spočíst co nejrychleji, zjevně se stačí omezit na případ $n > 0$.

Naivní postup výpočtu $[n]P$ jímá $n - 1$ sčítání, to jistě dokážeme vylepšit. Analogickým postupem jako při rychlém umocňování využijeme zápis n v binární soustavě. Inicializujeme $Q = \mathcal{O}$ a v k-tém kroku si budeme pamatovat bod $[2^k]P$, který ke Q přičteme jen pokud k-tý bit v binárním zápisu n je 1. Spočteme si pak $[2][2^k]P = [2^{k+1}]P$ a celý proces opakujeme znovu.

Příklad 1.1.9. Spočtěme padesátínásobek nějakého bodu P. Binární zápis 50 je 110010. Počítejme pak:

$$\mathcal{O} \longrightarrow \mathcal{O} \longrightarrow [2]P \longrightarrow [2]P \longrightarrow [2]P \longrightarrow [18]P \longrightarrow [50]P$$

$$Q: \quad \mathcal{O} \longrightarrow P \longrightarrow [2]P \longrightarrow [4]P \longrightarrow [8]P \longrightarrow [16]P \longrightarrow [32]P$$

$$+[2]P \qquad\qquad +[16]P \qquad +[32]P$$

Užijeme tedy pouze 10 operací sčítání.

Dohromady při výpočtu užijeme nejvýše $\lfloor \log_2(n) \rfloor - 1 \leqslant \log_2(n) - 1$ operací sčítání i dvojnásobení. Dvojnásobek prvků spočteme alespoň tak rychle jako součet dvou bodů, tedy tímto postupem spočteme $[n]P$ v nejvýše $2(\log_2(n) - 1)$ sčítáních.

Příklad 1.1.10. Určeme dvojnásobek bodu $P = (x, y)$ na $E : y^2 = x^3 + ax + b$. V duchu značení věty 1.1.7 máme pro $[2]P = (x_1, y_1)$:

$$x_1 = \lambda^2 - 2x = \frac{(3x^2 + a)^2 - 8y^2 x}{4y^2} = \frac{(3x^2 + a)^2 - 8(x^3 + ax + b)x}{4(x^3 + ax + b)} = \frac{x^4 - 2ax^2 - 8bx + a^2}{4(x^3 + ax + b)},$$

$$\begin{aligned} y_1 &= -\lambda x_1 - y + \lambda x = \frac{(3x^2+a)[-(3x^2+a)^2 + 12y^2x] - 8y^4}{8y^4} y \\ &= \frac{(3x^2+a)[-(3x^2+a)^2 + 12(x^3+ax+b)x] - 8(x^3+ax+b)^2}{8(x^3+ax+b)^2} y \\ &= \frac{x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - a^3 - 8b^2}{8(x^3+ax+b)^2} y. \end{aligned}$$

Všimneme si, že pro $P = (x, y)$ na eliptické křivce s $y = 0$ je $[2]P = \mathcal{O}$. Pro bod $Q = (6, 27) := (x_0, y_0)$ na křivce:

$$y^2 = x^3 + 54x + 189$$

nad $\mathbb{Q}$ zase ověříme, že platí:

$$x_0^6 + 5ax_0^4 + 20bx_0^3 - 5a^2x_0^2 - 4abx_0 - a^3 - 8b^2 = 0,$$

tedy $[3]Q = \mathcal{O}$. Obecně by nás mohlo zajímat, které body pošle násobení n do nekonečna.

Definice 1.1.11. Buď n celé číslo. O množině všech $P \in E$, že $[n]P = \mathcal{O}$, řekneme, že tvoří *n-torzi* E, a tuto množinu budeme značit $E[n]$.

Definice 1.1.12. Buď P bod na E. Pokud n je nejmenší kladné číslo, že $[n]P = \mathcal{O}$, nazveme n *řádem* P. Pokud takové n neexistuje tak řekneme, že P má nekonečný řád.

n-torze na eliptické křivce E tvoří podgrupu $E(\overline{K})$, neboť pokud $[n]P = \mathcal{O} = [n]Q$, tak $[n](P+Q) = [n]P + [n]Q = \mathcal{O}$. Torzní grupy nám pomáhají hlouběji studovat eliptické křivky v mnohých směrech. Zprvu si můžeme všimnout, že $E(\overline{\mathbb{F}_q})$ je sjednocením všech torzních grup, tedy že každý bod má konečný řád.

Věta 1.1.13. *Každý bod P na eliptické křivce E nad konečným tělesem má konečný řád.*

Důkaz. Mějme bod $P \in E(\overline{\mathbb{F}_q})$. Bod P leží v konečném rozšíření $E(\mathbb{F}_q)$, neboli pro nějaké přirozené k platí $P \in E\left(\mathbb{F}_{q^k}\right)$. V konečné grupě má každý prvek konečný řád, přičemž neutrální prvek grupy $E\left(\mathbb{F}_{q^k}\right)$ je $\mathcal{O}$, tedy P má na E konečný řád. □

Zatímco $E(\mathbb{F}_q)$ je konečná grupa, množina bodů na racionální křivce $E(\mathbb{Q})$ obecně není a existují na ní i body nekonečného řádu. Příkladem mřížového bodu nekonečného řádu na křivce je bod $(70, 13)$ na křivce:

$$E : y^2 = x^3 - 13,$$

tedy jeho násobením můžeme získat nekonečně mnoho racionálních bodů na E. Body nekonečného řádu jsou obecně těžko spočitatelné, nicméně body s řádem konečným dokážeme všechny najít za pomocí věty Lutz-Nagella [42, Thm. 8.7], dle které všechny takové racionální body (x, y) jsou mřížové a buď 2-torzní, či y^2 dělí diskriminant naší křivky.

1.2 Zobrazení mezi eliptickými křivkami

Když studujeme algebraické struktury, často nás zajímají zobrazení mezi nimi. Násobení bodů na E skalárem určuje homomorfismus grup $E(\overline{K}) \longrightarrow E(\overline{K})$, definuje proto endomorfismus na $E(\overline{K})$ daný lomenou funkcí nad K. Nyní se trochu obecněji podíváme na zobrazení mezi jednotlivými eliptickými křivkami, opět homomorfismy grup $E_1(\overline{K}) \longrightarrow E_2(\overline{K})$.

Nejprve studujme zobrazení invertibilní, tedy lineární změny souřadnic x, y. Pokud zobrazení $(x, y) \mapsto (ax + by + c, dx + ey + f)$ převádí eliptické křivky ve Weierstrassově tvaru, snadno porovnáním koeficientů, například xy na levé straně a x^2y na pravé, dojdeme k nulovosti členů b, c, d i f. Následně, aby členy při y^2 i x^3 byly po krácení oba rovny jedné, musí být $a = u^2, b = u^3$ pro nějaké nenulové číslo $u \in \overline{K}$. Taková zobrazení, $(x, y) \mapsto (u^2x, u^3y)$, převádí křivky:

$$E_1 : y^2 = x^3 + u^4ax + u^6b \quad \longrightarrow \quad E_2 : y^2 = x^3 + ax + b$$

pro nenulové $u \in \overline{K}$. Jako lineární zobrazení mezi $E_1(\overline{K})$ a $E_2(\overline{K})$ jistě naše zobrazení zachovává přímky a tedy i součet bodů na našich křivkách, definuje proto homomorfismus z $E_1(\overline{K})$ do $E_2(\overline{K})$. Díky jeho invertibilitě definuje mezi těmito množinami dokonce isomorfismus.

Isomorfismy nemusí být definované K, ale nad jeho rozšířením. Aby byl nad $\overline{K}$ definovaný, musí být díky předpisu $(x, y) \mapsto (u^2x, u^3y)$ psán nad rozšířením K stupně dělícího 6.

Definice 1.2.1. Buďte E, E' křivky isomorfní nad rozšířením K, ale ne nad K. Pak řekneme, že E' je *twistem* E nad K.

Zobrazení z $E : y^2 = x^3 + ax + b$ dané $(x, y) \mapsto \left(\frac{x}{d}, \frac{y}{\sqrt{d^3}}\right)$ pro $\sqrt{d} \notin K, d \in K$, nám dává isomorfismus do:

$$E_d : y^2 = x^3 + d^2ax + d^3b,$$

avšak ne nad K, ale nad jeho kvadratickým rozšířením $K(\sqrt{d})$. E_d nazveme *kvadratickým twistem* E.

Pro křivky s $a = 0$, resp. $b = 0$, můžeme analogicky najít *kubický* a *sextický*, resp. *kvartický* twist:

$$\begin{aligned} y^2 = x^3 + b \quad &\longrightarrow \quad y^2 = x^3 + d^2b, \\ y^2 = x^3 + b \quad &\longrightarrow \quad y^2 = x^3 + db, \\ y^2 = x^3 + ax \quad &\longrightarrow \quad y^2 = x^3 + dax, \end{aligned}$$

dané po řadě $(x, y) \mapsto \left(\frac{x}{\sqrt[3]{d^2}}, \frac{y}{d}\right)$ a $(x, y) \mapsto \left(\frac{x}{\sqrt[3]{d}}, \frac{y}{\sqrt{d}}\right)$, resp. $(x, y) \mapsto \left(\frac{x}{\sqrt{d}}, \frac{y}{\sqrt[4]{d^3}}\right)$. Vidíme, že poslední dvě zmíněné křivky jsou navíc kvadratickými twisty po řadě kubického a kvadratického twistu E.

Chtěli bychom říci, kdy mezi dvěma eliptickými křivkami existuje isomorfismus, tedy najít nějaký invariant, který isomorfní křivky sdílí. Takovou funkci splňuje právě j-invariant, jehož definice se táhne hluboko do komplexní analýzy.

Definice 1.2.2. Pro eliptickou křivku $E : y^2 = x^3 + ax + b$ definujeme její *j-invariant* jako:

$$j(E) = 1728\frac{4a^3}{4a^3 + 27b^2}.$$

Poznamenejme, že ten je vždy nad K definovaný, neboť eliptické křivky mají nenulový diskriminant.

Věta 1.2.3. *Dvě křivky definované nad K jsou isomorfní nad $\overline{K}$, právě pokud mají stejný j-invariant.*

Důkaz. Nejprve předpokládejme, že křivky $E_1 : y^2 = x^3 + a_1x + b_1$ a $E_2 : y^2 = x^3 + a_2x + b_2$ jsou nad $\overline{K}$ isomorfní. Máme pak $a_2 = u^2a_1$ a $b_2 = u^3b_1$ pro nějaké $b \in \overline{K}$. Spočtěme j-invariant obou křivek:

$$j(E_2) = 1728\frac{4u^6a_1^3}{4u^6a_1^3 + 27u^6b_1^2} = 1728\frac{4a_1^3}{4a_1^3 + 27b_1^2} = j(E_1),$$

j-invarianty isomorfních křivek se proto rovnají.

Nyní předpokládejme, že $j(E_1) = j(E_2)$. Počítejme:

$$\begin{aligned} 1728\frac{4a_1^3}{4a_1^3 + 27b_1^2} &= 1728\frac{4a_2^3}{4a_2^3 + 27b_2^2}, \\ a_1^3(4a_2^3 + 27b_2^2) &= a_2^3(4a_1^3 + 27b_1^2), \\ a_1^3b_2^2 &= a_2^3b_1^2. \end{aligned}$$

Pokud by například a_1 bylo nulové, je z nesingularity E_1 nutně b_1 nenulové, tudíž $a_2 = 0$. Proto ani b_2 není rovno nule, tedy pro $u \in \overline{K}$ s $u^3 = \frac{b_1}{b_2}$ máme $(0, b_1) = (0, u^3b_2)$. Analogicky pokud b_i jsou nulová, máme $(a_1, 0) = (u^2a_2, 0)$ pro u s $u^2 = \frac{a_1}{a_2} \in \overline{K}$.

Konečně v případě, že $a_1a_2b_1b_2 \neq 0$, máme $\frac{a_1^3}{a_2^3} = \frac{b_1^2}{b_2^2}$, což je druhou i třetí mocninou, tedy i šestou mocninou nějakého $u \in \overline{K}$. Toto číslo je tak šestou mocninou i všech šestých odmocnin u^6 v $\overline{K}$, pro tato u je tak $\frac{a_1}{a_2}$ rovno u^2 násobeno třetí odmocninou z 1 (ne nutně primitivní) a $\frac{b_1}{b_2}$ rovno u^3 násobeno odmocninou z 1. Pro nějaké z těchto šesti u se obě odmocniny rovnají 1, čili $a_1 = u^2a_2$ a $b_1 = u^3b_2$. □

Poznámka. Čtenáře by mohla zarazit konstanta $1728 = 12^3$, kterou j-invariant násobíme. Koncept j-invariantu se definuje nejen pro eliptické křivky, ale jako ??. Poznamenejme též, že Weierstrassův tvar není jediný možný vyjadřující eliptickou křivku, existuje rodiny křivek vyjádřitelné v tzv. *Legendreově* či *Edwardsově* tvaru, každá z nich mající svou vlastní formu j-invariantu.

Příklad 1.2.4. Vezměme si následujících pět křivek nad $\mathbb{F}_{101}$:

$$\begin{aligned} E_1 &: y^2 = x^3 + x + 1, \\ E_2 &: y^2 = x^3 + 5x + 23, \\ E_3 &: y^2 = x^3 + x - 1, \\ E_4 &: y^2 = x^3 + 2, \\ E_5 &: y^2 = x^3 + 2x, \end{aligned}$$

a spočtěme si jejich j-invarianty (což jsou čísla v $\mathbb{F}_{101}$):

$$\begin{aligned} j(E_1) &= 1728\frac{4}{31}, \\ j(E_2) &= 1728\frac{4 \cdot 5^3}{4 \cdot 5^3 + 27 \cdot 23^2} = 1728\frac{4 \cdot 24}{4 \cdot 24 + 27 \cdot 24} = 1728\frac{4}{31}, \\ j(E_3) &= 1728\frac{4}{31}, \\ j(E_4) &= 1728, \\ j(E_5) &= 0. \end{aligned}$$

Vidíme, že j-invarianty E_1 a E_2 se shodují, přičemž v $\mathbb{F}_{101}$ se oba rovnají $1728 \cdot 4 \cdot 88$, nutně mezi nimi nad $\overline{\mathbb{F}}_{101}$ existuje isomorfismus. Snadno ověříme, že zobrazení:

$$(x, y) \longmapsto (3^2x, 3^3y) = (9x, 27y)$$

převádí:

$$\begin{aligned} y^2 = x^3 + x + 1 \qquad \longrightarrow \qquad 27^2y^2 &= 9^3x^3 + 9x + 1, \\ 22y^2 &= 22x^3 + 9x + 1, \\ 22y^2 &= 22x^3 + 110x + 506, \\ y^2 &= x^3 + 5x + 23. \end{aligned}$$

Inverzní isomorfismus $E_2 \longrightarrow E_1$ je pak daný $(x, y) \mapsto (34^2x, 34^3y) = (45x, 15y)$, neboť multiplikativní inverz 3 v $\mathbb{Z}_{101}$ je 34.

Křivka E_3 má stejný j-invariant jako E_1 a E_2, nad $\mathbb{F}_{101}$ mezi nimi a E_3 přesto isomorfismus neexistuje. E_3 je kvadratickým twistem E_1 nad $\mathbb{F}_{101^2} = \mathbb{F}_{101}[i]$, jakožto zobrazení $(x, y) \mapsto \left(\frac{x}{i^2}, \frac{y}{i^3}\right) = (-x, iy)$ převádí:

$$\begin{aligned} y^2 = x^3 + x + 1 \quad \longrightarrow \quad -y^2 &= -x^3 - x + 1, \\ y^2 &= x^3 + x - 1. \end{aligned}$$

Obdobně můžeme najít isomorfismus definovaný nad $\mathbb{F}_{101^2}$ mezi E_1 a E_3.

Dvě speciální hodnoty j-invariantu jsou 0 a 1728, kterých nabývají křivky, které mají po řadě lineární, resp. konstantní člen roven 0. Právě křivky s j-invariantem 0 mají kubický (a sextický) twist, ty s j-invariantem 1728 zase kvartický.

Na propojení twistů křivek a počtu bodů na křivce poukazuje následující věta:

Věta 1.2.5. *Uvažme křivku $E/\mathbb{F}_q : y^2 = x^3 + ax + b$ a $\tilde{E}/\mathbb{F}_q : y^2 = x^3 + g^2ax + g^3b$ její kvadratický twist. Pak $\#E(\mathbb{F}_q) + \#\tilde{E}(\mathbb{F}_q) = 2(q+1)$.*

Důkaz. Protože $0^2 = 0$, je $g \in \mathbb{F}_q^\times$ kvadratický nezbytek. Ukážeme, že každé $x_1 \in \mathbb{F}_q$ dává přispívá právě dvěma body s touto x-ovou souřadnicí na obou křivkách. Pokud platí $x_1^3 + ax_1 + b = 0$, číslo x_1 dává po jednom bodu $(x_1, 0)$ na obou křivkách. Pro zbylé body tvrdíme, že je právě jedno z tvrzení pravdivé:

- Existují dva body na $E(\mathbb{F}_q)$ s x-ovou souřadnicí x_1,
- Existují dva body na $\tilde{E}(\mathbb{F}_q)$ s x-ovou souřadnicí gx_1.

Druhá odrážka je ekvivalentní s faktem, že:

$$(gx_1)^3 + g^2a(gx_1) + g^3b = g \cdot g^2(x_1^3 + ax_1 + b)$$

je nenulový čtverec. Připomeňme, že součin dvou kvadratických nezbytků je kvadratický zbytek a součin kvadratického zbytku a nezbytku je nezbytek. Protože g není čtverec v $\mathbb{F}_q$, je právě jedno z čísel $x_1^3 + ax_1 + b, g(x_1^3 + ax_1 + b)$ (nenulovým) čtvercem, tedy v $\mathbb{F}_q$ má dvě odmocniny. Afinních bodů na obou křivkách je tak dohromady $2q$. Poslední dva jsou příslušné body v nekonečnu. □

Počet různých j-invariantů v K určuje počet tříd isomorfismů křivek nad $\overline{K}$, případně kterých hodnot j-invariant nikdy nenabude. Jak si nyní ukážeme, tento počet je nejvyšší možný.

Věta 1.2.6. *Pro každé $s \in K$ existuje eliptická křivka E nad K s $j(E) = s$.*

Důkaz. Pro $s \in \{0, 1728\}$ poslouží jako příklady po řadě křivky $y^2 = x^3 + x, y^2 = x^3 + 1$. Pro zbylá $s \in K$ uvažme křivku:

$$E : y^2 = x^3 + 3s(1728 - s)x + 2s(1728 - s)^2.$$

Za předpokladu $\mathrm{char}\, K \notin \{2, 3\}$ je E vskutku eliptická, můžeme tedy definovat j-invariant. Ten je roven:

$$\begin{aligned} j(E) &= 1728\frac{4[3s(1728-s)]^3}{4[3s(1728-s)]^3 + 27[2s(1728-s)^2]^2} \\ &= 1728s\frac{4 \cdot 27s^2(1728-s)^3}{4 \cdot 27s^2(1728-s)^3(s+1728-s)} = \frac{1728}{1728}s = s. \end{aligned}$$

Křivka E proto má j-invariant roven s. □

Věta 1.2.7. *Pro každé $s \in \overline{K}$ existuje eliptická křivka E nad $K(s)$, že $j(E) = s$.*

Důkaz. Opět si rozmyslíme, že křivka $y^2 = x^3 + 3s(1728 - s)x + 2s(1728 - s)^2$ je definovaná nad $K(s)$, tedy může posloužit jako řešení. □

Jak násobení bodů E skalárem, tak braní isomorfismu, jsou homomorfismy bodů křivek nad tělesem K, resp. jeho rozšířením. Spadají tak pod rodinu zobrazení eliptických křivek zvaných *isogenie*, o kterých se budeme dále bavit.

1.3 Isogenie

Podívejme se trochu obecněji na zobrazení mezi křivkami. Hlavní vlastnost, kterou bychom chtěli na takových zobrazeních vynutit, by bylo zachování grupové struktury bodů na křivce. Ukáže se, že taková zobrazení mají několik velmi dobrých vlastností.

Definice 1.3.1. Ať E_1, E_2 jsou eliptické křivky nad tělesem K. Surjektivní homorfismus grup $\phi : E_1(\overline{K}) \longrightarrow E_2(\overline{K})$ daný racionální funkcí nad K, který posílá bod v nekonečnu E_1 na bod v nekonečnu E_2, nazveme *isogenií*. Pokud mezi E_1, E_2 existuje isogenie, nazveme je *isogenní*.

Dá se ukázat, viz [17, II.6.8.] a [34, III.4.8.], že nekonstantní racionální funkce mezi hladkými křivkami (diskriminant je nenulový) je surjektivní i homomorfismus mezi grupami $E_1(\overline{K}) \longrightarrow E_2(\overline{K})$, definice výše je tedy příliš silná. Zachycuje nicméně všechny důležité vlastnosti, které v isogeniích hledáme. Pokud naši isogenii uvážíme jako zobrazení:

$$\phi : E_1 \longrightarrow E_2 : (x, y) \longmapsto (u(x, y), v(x, y))$$

pro u, v lomené funkce nad K, tak po substituci $(x, y) \mapsto (x/z, y/z)$, požadujeme, aby $(0 : 1 : 0) \mapsto (0 : 1 : 0)$. Isogenie můžeme zapsat mnohem kompaktněji:

Věta 1.3.2. *Buďte $E_1, E_2 \in K$ eliptické křivky a $\phi : E_1 \longrightarrow E_2$ isogenie. Pak ji můžeme zapsat ve standardním tvaru:*

$$\phi(x, y) = (u(x), v(x)y)$$

pro u, v lomené funkce nad K.

Důkaz. Víme, že isogenii můžeme vyjádřit jako $\phi : (x, y) \mapsto (u(x, y), v(x, y))$ pro u, v lomené funkce nad K. Z rovnice eliptcké křivky $E_1 : y^2 = x^3 + ax + b$ můžeme y v sudé mocnině nahradit polynomem v x, čímž zajistíme, že u i v dokážeme vyjádřit jako funkce r, s, jejichž stupeň v y je nejvýše 1. Speciálně mějme $u(x, y) = \frac{f_1(x)+f_2(x)y}{f_3(x)+f_4(x)y}$ pro $f_i \in K[x]$. Pokud tento zlomek rozšíříme o $f_3(x) - f_4(x)y$, vyruší se nám všechny liché mocniny y ve jmenovateli a sudé dokážeme nahradit polynomem v x. Můžeme proto předpokládat $u(x, y) = \frac{f_1(x)+f_2(x)y}{f_3(x)}$.

Protože ϕ je homomorfismem mezi grupami $E_1(\overline{K}) \longrightarrow E_2(\overline{K})$, platí rovnost $\phi(x,y) = -\phi(x,-y)$, tedy f_2 je identicky nulový polynom a u je lomená funkce v x. Pokud obdobně vyjádříme $v(x,y) = \frac{g_1(x)+g_2(x)y}{g_3(x)}$, získáme $g_1 \equiv 0$ a $v(x,y) = \frac{g_1(x)}{g_2(x)}y$ pro $g_1, g_2 \in K[x]$. □

Díky této charakterizaci můžeme začít s isogeniemi pořádně pracovat. Zprvu hned vidíme, že u, v mají stejné jádro (ve smyslu homomorfismu grup), právě protože bod $\mathcal{O}$ je isogenií zachován.

Definice 1.3.3. Pod *stupněm* isogenie ϕ budeme rozumět jejímu stupni jako lomené funkci v x a značit $\deg\phi$. Definujeme $\deg[0] = 0$.

Všechny vlastnosti stupňů racionálních funkcí jsou u stupňů isogenií zachovány, zejména jeho multiplikativita.

Stejně jako jsme se zabývali torzní podgrupou našich křivek, nebude překvapením, že bude pro studium isogenií důležité, které body zobrazí do nekonečna. Tyto body i v případě isogenií tvoří podgrupu $E(\overline{K})$.

Definice 1.3.4. Pod *jádrem* isogenie ϕ rozumíme jádru ϕ, ve smyslu homomorfismu grup $E_1(\overline{K}) \longrightarrow E_2(\overline{K})$. Značíme $\ker\phi$ a počet jeho prvků $\#\ker\phi$.

Definice 1.3.5. Skládání, resp. sčítání isogenií definujeme následovně: $\phi\circ\psi := \phi(\psi)$, resp. $(\phi+\psi)P := \phi(P) + \psi(P)$. Značme též isogenii opačnou jako: $-\phi = [-1]\circ\phi$.

S isogeniemi jsme se již na naší (prozatím) krátké cestě hned několikrát setkali, jak násobení (nenulovým) skalárem, tak isomorfismy zmíněné v předchozí kapitole, jsou isogeniemi, dokonce jediné invertibilní. Násobení $[n]$ má jádro $E[n]$ a za chvíli si ukážeme, že má coby isogenie stupeň n^2. Zobrazení $[0]$ není surjektivní a proto není isogenií. Isomorfismy jsou isogenie lineární a mají pouze triviální jádro. Zobrazení:

$$\phi : y^2 = x^3 + x \quad \longrightarrow \quad y^2 = x^3 + 11x + 62$$

mezi křivkami nad $\mathbb{F}_{101}$ dané $(x,y) \mapsto \left(\frac{x^2+10x-2}{x+10}, \frac{x^2+20x+1}{x^2+20x-1}y\right)$ je též isogenií, tentokrát stupně dvě. Jádrem ϕ je množina $\{\mathcal{O}, 10\}$, protože $x^2 + 20x - 1 = (x+10)^2$ v $\mathbb{Z}_{101}$.

Jedním z nejdůležitějších zobrazení na $\overline{\mathbb{F}_p}$ je tzv. *Frobeniův morfismus*, pojmenovaný po Ferdinandu Frobeniovi, jemuž diktuje předpis $\pi : x \mapsto x^p$. Pevné body Frobeniova morfismu jsou přesně prvky $\mathbb{F}_p$, tudíž pro lomenou funkci f nad $\mathbb{F}_p$ a $x_i \in \overline{\mathbb{F}_p}$ platí $f(x_1^p, \ldots, x_n^p) = f(x_1, \ldots, x_n)^p$. Speciálně platí vztahy $0^p = 0, 1^p = 1, a^p + b^p = (a+b)^p$ a $a^p \cdot b^p = (ab)^p$ pro libovolné $a, b \in \overline{\mathbb{F}_p}$. Navíc toto zobrazení je nad $\overline{\mathbb{F}_p}$ prosté, pokud $a^p = b^p$:

$$0 = a^p - b^p = (a-b)^p,$$

tedy $a = b$. Frobeniův morfismus je proto nad $\overline{\mathbb{F}_p}$ automorfismem.

Mocninu Frobeniova automorfismu definujeme jako $\pi^n : x \mapsto x^{p^n}$, neboli složení n interací π. Rozkladové těleso polynomu $x^{p^n} - x$ je $\mathbb{F}_{p^n}$, což znamená, že π^n je automorfismem právě nad konečnými tělesy $\mathbb{F}_q$, kde $q = p^k$ s $k \leqslant n$.

Zobrazení s podobným předpisem převádějící eliptické křivky též nese jméno po Frobeniovi.

Definice 1.3.6. Buď $E : y^2 = x^3 + ax + b$ eliptická křivka nad $\mathbb{F}_q$. Zobrazení:

$$\pi_E : y^2 = x^3 + ax + b \quad \longrightarrow \quad y^2 = x^3 + a^q x + b^q,$$

dané:

$$(x, y) \longmapsto (x^q, y^q),$$

se nazývá *Frobeniovým endomorfismem*.

Díky vlastnostem π definuje π_E homomorfismus mezi grupami křivek a zjevně zachovává bod v nekonečnu, tedy je vskutku isogenií. Frobeniův endomorfismus fixuje právě $E(\mathbb{F}_q)$ a má pouze triviální jádro. Dále komutuje s libovolnou lomenou křivkou nad $\mathbb{F}_q$, tj.:

$$\pi_E \circ \phi = \phi \circ \pi_E,$$

speciálně tento vztah platí pro libovolnou isogenii ϕ z E. Mocninu Frobeniova morfismu analogicky definujeme jako ${\pi^n}_E := \underbrace{\pi_E \circ \pi_E \circ \circ \circ \pi_E}_{n}$ a má vlastnosti analogické k π. Pokud bude jasné, kdy mluvíme o isogenii a ne o zobrazení na $\mathbb{F}_q$, zneužitím notace budeme π_E značit pro jednoduchost též π.

Můžeme též definovat p-Frobeniův morfismus $(x, y) \mapsto (x^p, y^p)$ na E nad $\mathbb{F}_q$ pro $q \neq p$, který je opět homomorfismem grup bodů eliptických křivek, ale již ne nutně definuje endomorfismus.

Když již máme solidní představu pojmu isogenie, pojďme se nyní pobavit o několika jejich základních vlastnostech. Jedním z nejdůležitějších výsledků ohledně isogenií mluví o jejich duálu.

Věta 1.3.7. *Buď $\phi : E \longrightarrow E_1$ isogenie stupně n. Pak existuje jediná isogenie $\hat{\phi} : E_1 \longrightarrow E$ splňující $\phi \circ \hat{\phi} = [n]_E$. Tuto isogenie nazýváme k ϕ duální. Definujeme též $\hat{[0]} = [0]$.*

Důkaz existence duální isogenie je poměrně zdlouhavý a vyžaduje rozebírání mnoha případů, zde jej proto vynecháme. Čtenář jej však může najít v [34, Thm. III.6.1.], trochu elementárnější přístup se nachází v [39, Thm. 7.8.].

Duální isogenie konečně opodstatňuje fakt, který na první pohled není vůbec jasný, že „být isogenní“ je relace ekvivalence. Několik základních vlastností duální isogenie stanovuje následující věta:

Věta 1.3.8. *Buď $\phi : E \longrightarrow E_1$ isogenie stupně n. Pak její duální isogenie pro každou jinou isogenii $\psi : E_1 \longrightarrow E_2$ splňuje:*

(i) $\phi \circ \hat{\phi} = [n]_E$,

(ii) $\hat{\phi} \circ \phi = [n]_{E'}$,

(iii) $\widehat{\phi \circ \psi} = \hat{\psi} \circ \hat{\phi}$,

(iv) $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$,

(v) $\hat{\hat{\phi}} = \phi$.

Důkaz. Dokážeme vlastnosti (ii) a (v). Platí:

$$(\hat{\phi} \circ \phi) \circ \hat{\phi} = \hat{\phi} \circ (\phi \circ \hat{\phi}) = \hat{\phi} \circ [n]_E = [n]_{E'} \circ \hat{\phi},$$

kde poslední rovnost platí, právě protože isogenie jsou jsou homomorfismy grup. Protože isogenie jsou surjektivní, musí platit $\hat{\phi} \circ \phi = [n]_{E'}$. Dále bod (v) plyne z (i) a (ii), platí totiž $\hat{\hat{\phi}} \circ \hat{\phi} = [n]_E = \phi \circ \hat{\phi}$, tedy $\hat{\hat{\phi}} = \phi$. Zbytek důkazu je k nalezení v [34, Thm. III.6.1]. □

Lemma 1.3.9. *Platí:*

$$\widehat{[n]} = [n] \qquad a \qquad \deg[n] = n^2.$$

Důkaz. Zjevně $\widehat{[0]} = [0]$ a $\widehat{[1]} = [1]$, dále postupujme indukcí dle n. Za pomocí věty 1.3.8, (iv), máme:

$$\widehat{[n+1]} = \widehat{[n]} + \widehat{[1]} = [n] + [1] = [n+1].$$

Protože $[-1] : P \mapsto -P$ je isogenií stupně 1, je $[-1]$ též duálem sama sebe. Pak díky $[-1] \circ [n] = [-n]$ máme první část hotovou. Z definice sčítání máme $[m] \circ [n] = [mn]$, tudíž $[n] \circ \widehat{[n]} = [n^2]$. Dle věty 1.3.8, (ii), je $[n]$ isogenií stupně n^2. □

Poznámka. V literatuře se vlastnosti duální isogenie dokazují tak, že se elementárnějšími úvahami, například o tzv. *division polynomials*, ukáže $\deg[n] = n^2$, kde pak jednoduše plynou odrážky (ii), (iii) a (v). Čtvrtý bod je obzvláště těžké dokázat a jeho nejvíce přímočarý důkaz užívá *Weilových párování*, kterým se v naší práci nevěnujeme.

Je důležité si uvědomit, co nám předchozí charakterizace vlastně říkají o duální isogenii. Duální isogenie k ϕ je z našeho lemmatu též isogenií stupně n, která má velmi pěkné vlastnosti. Navíc pro libovolnou isogenii ϕ z E stupně n je $\ker \phi \subseteq \ker[n]$, neboť libovolný prvek v jádře ϕ se $\hat{\phi}$ zobrazí do nekonečna E.

Když víme, že „být isogenní“ je relace ekvivalence, dalším krokem je jistě hledat způsob, jak klasifikovat třídy isogenních křivek. V minulé sekci jsme si ukázali, že na kvadratickém

twistu křivky leží pouze určitý počet bodů. I v případě isogenií definovaných nad tělesem $\mathbb{F}_q$ isogenie úzce souvisí s počtem bodů ležících na křivce. Toto krietérium zní překvapivě jednoduše:

Věta 1.3.10. *(Sato-Tate) Buďte E, E' eliptické křivky nad $\mathbb{F}_q$. Pak tyto křivky jsou nad $\mathbb{F}_q$ isogenní, právě pokud platí $\#E(\mathbb{F}_q) = \#E'(\mathbb{F}_q)$.*

Důkaz. Isogenie jsou surjektivní, přičemž isogenie nad $\mathbb{F}_q$ zobrazuje $E(\mathbb{F}_q)$ samu na sebe. Pokud jsou E a E' isogenní, platí pak $\#E(\mathbb{F}_q) \geqslant \#E'(\mathbb{F}_q)$ a $\#E'(\mathbb{F}_q) \geqslant \#E(\mathbb{F}_q)$, což dává jednu polovinu věty. Druhá část již tak jednoduše nepřichází a její důkaz dokonce není ani zdaleko přístupný z pohledu algebraické geometrie. Poprvé byla druhá implikace (resp. tvrzení jí ekvivalentní) zveřejněno v jedné z nejvlivnějších publikací Johna Tate, [40]. □

Body, které se nachází v jádru isogenie tvoří podgrupu $E(\overline{K})$, přičemž její velikost je shora omezena stupněm isogenie. Limitní případ v tomto smyslu má zajímavé vlastnosti.

1.4 Separabilní isogenie

Definice 1.4.1. Mějme E, E' křivky nad K a $\phi : E \longrightarrow E'$ isogenii stupně n. Pokud je $\#\ker\phi = n$, pak o ϕ řekneme, že je *separabilní*. V opačném případě řekneme, že ϕ je *neseparabilní*. V případě, že je $\deg\phi$ roven mocnině $\operatorname{char} K$, mluvíme o ϕ jako o *čistě neseparabilní*.

Pozoruhodné na tomto pojmenování je fakt, že separabilita a čistá neseparabilita se ne nutně vylučují. Každý isomorfismus je isogenií stupně 1 s jádrem velikosti 1, tedy separabilní, přičemž $p^0 = 1$, takže isomorfismy jsou čistě neseparabilní. Naopak Frobeniův automorfismus je isogenie neseparabilní i čistě neseparabilní. Charakterizujme dále separabilní isogenie.

Věta 1.4.2. *Ať E, E' jsou eliptické křivky nad K a $\phi : E \longrightarrow E'$ isogenie, která převádí $x \mapsto \frac{u(x)}{v(x)}$ pro $u, v \in K[x]$ nesoudělné. Pak $\left(\frac{u}{v}\right)' \neq 0$ nastane právě pokud ϕ je separabilní.*

Důkaz. Položme $p = \operatorname{char} K$. Rovnost $0 = \left(\frac{u}{v}\right)' = \frac{u'v - v'u}{v^2}$ v K nastane právě pokud $u'v = v'u$. Protože je ϕ isogenie, jsou u, v nenulové polynomy nad K. Předpokládejme, že u' a tedy i v' nejsou nulové. Z nesoudělnosti polynomů u, v nutně každý kořen u je kořenem u' s nejméně stejnou násobností. Nicméně pro $u' \neq 0$ je $\deg u > \deg u'$, což je spor. Rovnost $u'v = v'u$ proto můžeme relaxovat na $u' = v' = 0$, tedy každý nenulový jednočlen u, v má koeficient dělitelný p a tak $u = f(x^p)$ a $v = g(x^p)$ pro nějaké polynomy $f, g \in K[x]$. Pak ale $\frac{u(x)}{v(x)} = \frac{f(x^p)}{g(x^p)} = \left(\frac{f(x)}{g(x)}\right)^p$ a v jistě jádro velikosti $\deg u/v$, ať už $p > 0$ či ne.

Buď proto $\phi(x, y) = \left(\frac{u(x)}{v(x)}, \frac{r(x)}{s(x)}y\right)$ standardní tvar ϕ, kde $\left(\frac{u}{v}\right)' \neq 0$, a (a, b) bod v obrazu $E(\overline{K})$ ve ϕ takový, že $ab \neq 0$ a a není podílem vedoucích koeficientů u a v. Takový bod jistě existuje, protože obraz $\phi(E(\overline{K}))$ je nekonečná množina. Uvažme nyní množinu

M všech předobrazů (a, b) ve ϕ, neboli bodů $(x, y) \in E$ s $\phi(x, y) = (a, b)$. Protože ϕ je homomorfismus grup, počet prvků M je přesně roven velikosti jádra ϕ.

Pro každé $(x, y) \in \mathsf{M}$ dále platí:

$$\frac{u(x)}{v(x)} = a, \qquad \frac{r(x)}{s(x)} y = b.$$

Díky předpokladu $b \neq 0$ je každé vyhovující y jednoznačně určeno daným x jako $b\frac{s(x)}{r(x)}$, což znamená, že velikost M je rovna počtu x splňujících první naši rovnost, tedy počtu různých kořenů polynomu $h := u - av$, který má díky podmínkám na a stupeň $\deg \phi$. Dejme tomu, že x_0 je vícenásobný kořen h, pak platí:

$$u(x_0) = av(x_0),$$
$$u'(x_0) = av'(x_0).$$

Násobení protějších stran těchto rovností dává $u'(x_0)v(x_0) = u(x_0)v'(x_0)$, x_0 je tedy kořenem (nenulového) polynomu $u'v - uv'$, který má v $\overline{K}$ pouze konečně mnoho kořenů. Protože $\phi(E(\overline{K}))$ je nekonečná a M konečná množina, můžeme si zvolit (a, b) bod takový, že h žádný násobný kořen nemá. Pak $\#\ker\phi = |\mathsf{M}| = \deg h = \deg\phi$. □

Speciálně nad tělesem s nulovou charakteristikou jsou všechny isogenie neseparabilní. Zaměřme se na konečný případ, kde musí pro ϕ ve standardním tvaru platit $(u/v)' = 0$, tedy jak jsme si ukázali v důkazu předchozí věty, u/v je složením racionální funkce nad $\mathbb{F}_q$ a p-Frobeniova morfismu na $\mathbb{F}_q$. Vypadá to tedy, že i y-ová souřadnice se bude chovat podobně, bohužel dokázat tento fakt je poměrně těžší (a ošklivější), než ho konstatovat.

Důsledek 1.4.3. *Buď ϕ isogenie nad $\mathbb{F}_q$. Pak existuje separabilní isogenie ψ a $n \in \mathbb{N}_0$, že:*

$$\phi = \psi \circ \pi^n.$$

Důkaz. Stačí ukázat, že pro neseparabilní isogenii ϕ existuje separabilní ϕ s $\phi = \psi \circ \pi$. Pro x-ovou souřadnici tento výsledek známe, zbytek důkazu se dá najít na [39, Lemma 6.3.]. Iterací tohoto faktu a skutečností, že Frobenius komutuje s libovolnou isogenií nad $\mathbb{F}_q$, pak získáme výsledek. □

Jedna z nejvýznamnějších vlastností separabilní isogenií je úzce spojena s jejich „maximálním“ jádrem.

Věta 1.4.4. *Každá separabilní isogenie ϕ z E je, až na isomorfismus, jednoznačně určena svým jádrem.*

Důkaz tvrzení je uveden v [42, Prop. 12.12], nicméně autor jej zde podává s notnou dávkou Galoisovy teorie, jejíž znalost od čtenáře nepředpokládáme.

Definice 1.4.5. Buď E eliptická křivka, která je doménou separabilní isogenie ϕ. Pokud je $G = \ker \phi$ grupa tvořená jádrem ϕ, budeme značit E/G cílovou křivku ϕ.

Poznámka. Ač E/G je pouze značení pro křivku a nesmí být naivně bráno ve smyslu faktorizace, není zcela nepodložené. Ve zkratce, můžeme každému bodů $P \in G$ přiřadit automorfismus σ_P na E daný $Q \mapsto Q + P$, který až na případ $P = \mathcal{O}$ bod v nekonečnu nezachová, isogenií proto není. Pokud bychom grupu $E(\overline{K})$ vyfaktorizovali grupou všech takových σ_P, získáme homomorfismus $\phi : E \longrightarrow E/G$. Není naprosto vůbec zjevné, že E/G je eliptickou křivkou, ani že ϕ je isogenií, detaily faktorizace E/G též vyžadují náramnou péči. Čtenář obeznámen s teorií tělesových vnoření a obecně Galoisovou teorií nalezne podrobnější náznak důkazu na [39, Thm. 6.10.].

Separabilní isogenie z $E \longrightarrow E'$ je daná lomenou funkcí nad K a známe-li její jádro, dokážeme ji explicitně spočíst, přičemž libovolná podgrupa $E(\overline{K})$ je jádrem separabilní isogenie. Vzorce udávající (až na isomorfismus) přesný tvar separabilní isogenie z $E \longrightarrow E'$ s daným jádrem se nazývají *Véluovy* po Jeanu Véluovy, který je první publikoval roku 1971 ve [41]. Jejich zápis je obecně velice nezáživný a pro nás nepodstatný, stačí nám mít v povědomí, že separabilní isogenie s daným jádrem můžeme explicitně vyjádřit. Jejich přesnou formu a důkaz správnosti jsou k uvedeny v [9, Ch. 8.2]. V Sage 9.0 jsou Véluovy vzorce implementovány pro isogenii z E s jádrem G v $O(\#G)$ příkazem:

`EllipticCurveIsogeny(E,ker G).`

Příklad 1.4.6. Separabilní isogenie s doménou $E/\mathbb{F}_{101} : y^2 = x^3 + 2x + 13$ a jádrem ???

Jistě složením neseparabilní isogenie s libovolnou jinou získáme opět neseparabilní isogenii. Podobné vlastnosti má ale i součet isogenií.

Věta 1.4.7. *Buďte $\phi, \psi : E \longrightarrow E_1$ isogenie, přičemž ϕ je neseparabilní. Pak $\phi + \psi$ je neseparabilní právě pokud ψ je neseparabilní.*

Důkaz. Označme $\pi_p : (x, y) \to (x^p, y^p)$ p-Frobeniův endomorfismus na E, ten komutuje s libovolnou isogenií, a navíc isogenie π je nějakou jeho mocninou. Podle věty 1.4.7 existují separabilní isogenie $\eta, \vartheta : E \longrightarrow E_1$ splňující $\phi = \eta \circ \pi_p^a$ a $\psi = \vartheta \circ \pi_p^b$, kde $a > 0$. Pokud ψ je neseparabilní, je exponent b kladný, tedy součet $\phi + \psi$ je roven:

$$\phi + \psi = \eta \circ \pi_p^a + \vartheta \circ \pi_p^b = (\eta \circ \pi_p^{a-1} + \vartheta \circ \pi_p^{b-1}) \circ \pi_p,$$

neseparabilní isogenii. Naopak je-li isogenie $\phi+\psi$ neseparabilní, je $\psi = (\phi+\psi)-\phi$ součtem neseparabilních isogenií $\phi + \psi$ a $-\phi$, o kterém jsme právě ukázali, že je neseparabilní. □

Poznámka. Tato věta má hned několik důležitých aplikací, jednu z nich si ukážeme hned o dvě sekce vedle. Je ale též jednou z klíčových ingrediencí důkazu Hasseho věty 1.1.5. Konkrétně z ní plyne, že $1-\pi$ je separabilní isogenie, tedy $\deg 1-\pi = \#\ker 1-\pi = \#E(\mathbb{F}_q)$, k tomuto fakt se ještě vrátíme. Stačí si pak všimnout, že příslušné členy v Hasseho větě jsou po řadě $\deg 1-\pi, \deg 1, \deg -\pi$ a užít jednu speciální formu Cauchy-Schwarzovy nerovnosti, na detaily čtenáře odkazujeme na [34, Thm. V.1.1.].

Konečně, ??. Pokud vezmeme podgrupu prvočíselného řádu grupy G, kterou pokládame za jádro separabilní isogenie, můžeme

Věta 1.4.8. *Každou isogenii ϕ složeného stupně můžeme rozložit na kompozici isogenií prvočíselných stupňů.*

Důkaz. Dejme tomu, že ϕ převádí křivky $E \longrightarrow E_1$. Protože π má prvočíselný stupeň charakteristky našeho tělesa, stačí nám díky větě 1.4.3 uvažovat ϕ isogenii separabilní. Postupujme nyní silnou indukcí vzhledem k počtu dělitelů $\deg \phi$. Pokud $G = \ker \phi$ je triviální či má prvočíselný řád, jsme hotovi. V opačném případě dejme tomu, že všechny isogenie s jádrem nižšího počtu dělitelů než $\#\ker \phi$ jsou rozložitelné. Víme, že G obsahuje podgrupu H prvočíselného řádu (tzv. *Sylowova podgrupa*), která určuje separabilní isogenii $\psi : E \longrightarrow E_2 \cong E/H$. Pak obraz G v ψ je konečná podgrupa $E_1(\overline{K})$, která je isomorfní G/H, a definuje isogenii $\chi : E_2 \longrightarrow E_3 \cong E_2/\psi(G)$. Jádro $\chi \circ \psi$ je právě G, tedy podle věty 1.4.4 existuje isomorfismus $\iota : E_3 \longrightarrow E_2$ splňující $\phi = \iota \circ \chi \circ \psi$. Podle předpokladu $\iota \circ \chi$ je buďto isomorfismus, nebo je rozložitelná na kompozici separabilních isogenií prvočíselných stupňů. □

Tato věta na první pohled zní hezky z pohledu čirého studia křivek, má ale kolosální využití. Konkrétně, užijeme ji při počítání separabilní isogenie *hladkého* stupně, tedy dělitelného pouze prvočísly nižšími než daná hranice. Chceme-li nalézt isogenii s jádrem dejme tomu 3^{2021} torzí, počítání jí naivně by zabralo $O(3^{2021+\varepsilon})$ operaci, rozhodně ne efektivní. Pokud si vezmeme její generátor, tedy bod řádu 3^{2021}, můžeme jako v důkazu předchozí věty brát vždy pogrupu řádu 3, tedy při i-tém kroku počítáme separabilní isogenii vycházející ze křivky???????????? $E/E[\ell^{i-1}]$ a s jádrem ℓ^i, viz následující diagram:
(obrázek?)

Za předpokladu, že bereme prvočísla rozumně malá, tedy že jejich příslušnou torzi spočteme v konstantním čase, získáme křivku $E/[p^a]$ v $O(a)$ operacích. Pro nesoudělná čísla m, n platí $(E/[m])/[n] \cong E/[mn]$, tedy celou spočteme alespoň v logaritmickém čase.

1.5 Torzní body

Vraťme se k operaci násobení bodů. Za pomocí vlastností isogenií vyvynutých v předchozích částech budeme konečně schopni přijít na kloub struktuře torzních grup a na základě toho i samotné grupě $E(\mathbb{F}_q)$. Začněme tedy směrem k tomuto cíli dělat první krůčky.

Charakterizovat $E[2]$ je jednoduché. Spolu s bodem v nekonečnu jsou násobením dvěma anihilované právě tři další body, jejich x-ové souřadnice jsou jednotlivými kořeny x^3+ax+b. Protože torze tvoří grupu a na naší 2-torzi má každý afinní bod řád 2, musí nutně být $E[2] \cong \mathbb{Z}_2 \times \mathbb{Z}_2$.

3-torze jsme též schopni diskutovat. Body na ní splňují $[2]P = -P$, speciálně se x-ové souřadnice obou stran rovnají. To znamená, že:

$$\left(\frac{3x^2 + a}{2y}\right)^2 - 2x = x,$$

neboli díky rovnosti $y^2 = x^3 + ax + b$:

$$(3x^2 + a)^2 = 12x(x^3 + ax + b),$$

což je kvartická rovnice, která se snadno ověří jako s nenulovým diskriminantem. Každému ze čtyř různých vyhovujících x přísluší dvě hodnoty y a body (x, y) mají všechny řád 3. Spolu s $\mathcal{O}$ náleží 3-torzi právě 9 bodů. Snadno pak dojdeme k závěru $E[3] \cong \mathbb{Z}_3 \times \mathbb{Z}_3$.

V obou případech implicitně závisíme na faktu, že q není mocnina 2 ani 3, jinak naše eliptická křivka nemá tvar, který jí připisujeme. Tento případ je rozebírán v [42, Ch. 3.1].

Mohli bychom se tedy dovtípit, že n-torze pro n nesoudělné s q je isomorfní $\mathbb{Z}_n \times \mathbb{Z}_n$. Tato skutečnost je díky existenci duální isogenie velmi úzce spjata se separabilitou n-násobící mapy.

Lemma 1.5.1. *Buď E/K eliptická křivka s $p = \operatorname{char} K$ a n celé číslo. Pak $[n]$ je neseparabilní, právě pokud $p \mid n$.*

Důkaz. Dejme tomu, že $[n]$ je neseparabilní, pak díky důsledku 1.4.3 je $[n] = \pi \circ \phi$ pro nějakou isogenii ϕ a tedy $p \mid \deg \pi \cdot \deg \phi = \deg \pi \circ \phi = \deg[n] = n^2$, neboli $p \mid n$. Mějme naopak $p \mid n$, můžeme pak psát $[n] = [p][n/p]$. Víme, že $[p]$ je neseparabilní, protože $\pi \circ \widehat{\pi} = [\deg \pi] = [p]$. Definice separability pomocí velikosti jádra jistě implikuje, že složení neseparabilní isogenie, zde $[p]$, s libovolnou jinou vyprodukuje isogenii neseparabilní, tedy $[n]$ je neseparabilní sama. □

Nejprve se zaměříme na prvočísla a jejich mocniny.

Věta 1.5.2. *Buď E/K eliptická křivka s $p = \operatorname{char} K$ a $\ell \neq p$ prvočíslo. Pak:*

$$E[\ell^e] \cong \mathbb{Z}_{\ell^e} \times \mathbb{Z}_{\ell^e}$$

pro každé $e \geqslant 1$.

Důkaz. Postupujme silnou indukcí podle e. Isogenie $[\ell]$ je pro prvočísla $\ell \neq p$ separabilní, tedy $\#E[\ell] = \#\ker[\ell] = \ell^2$. Každý afinní prvek $E[\ell]$ má řád ℓ, tedy platí $E[\ell] \cong \mathbb{Z}_\ell \times \mathbb{Z}_\ell$. Nyní již uvažme abelovskou grupu $E[\ell^e]$ pro nějaké $e > 1$ a předpokládejme, že věta platí pro všechna kladná $a < e$. Opět víme, že $\#E[\ell^e] = \#\ker[\ell^e] = \ell^{2e}$ a každý afinní prvek $E[\ell^e]$ nemá řád vyšší než ℓ^e. Navíc pro každé $a < e$ existuje na $E[\ell^e]$ právě ℓ^{2a} prvků řádu ℓ^a, tedy $E[\ell^e]$ má shodnou strukturu jako $\mathbb{Z}_{\ell^e} \times \mathbb{Z}_{\ell^e}$. □

Důsledek 1.5.3. *Buď E/K eliptická křivka s $p = \operatorname{char} K$ a $p \nmid m$ přirozené číslo. Pak $E[m] \cong \mathbb{Z}_m \times \mathbb{Z}_m$.*

Důkaz. Pokud m, n jsou nesoudělná čísla, jistě platí $E[m] \times E[n] \cong E[mn]$. Čínská zbytková věta pro taková m, n tvrdí $(\mathbb{Z}_m \times \mathbb{Z}_m) \times (\mathbb{Z}_n \times \mathbb{Z}_n) \cong \mathbb{Z}_{mn} \times \mathbb{Z}_{mn}$, tedy pokud $m = p_1^{a_1} \cdots p_k^{a_k}$ rozložíme na součin prvočíselných mocnin, s pomocí předchozí věty platí:

$$E[m] \cong E[p_1^{a_1}] \times \cdots \times E[p_k^{a_k}] \cong \left(\mathbb{Z}_{p_1^{a_1}} \times \mathbb{Z}_{p_1^{a_1}}\right) \times \cdots \times \left(\mathbb{Z}_{p_k^{a_k}} \times \mathbb{Z}_{p_k^{a_k}}\right) \cong \mathbb{Z}_m \times \mathbb{Z}_m,$$

což jsme chtěli. □

Zásadní rozdíl nastává při násobení mocninou charakteristky našeho tělesa, isogenie $[p]$ je totiž (čistě) neseparabilní. Případ $\operatorname{char} K = 0$ je triviální, podíváme se proto opět pouze na konečný případ.

Věta 1.5.4. *Buď $E/\mathbb{F}_q$ s $q = p^k$ eliptická křivka. Pak platí:*

$$E[p^e] \cong \begin{cases} \{\mathcal{O}\}, & \text{pro každé nezáporné } e, \\ \mathbb{Z}_{p^e}, & \text{pro každé nezáporné } e. \end{cases}$$

Důkaz. Isogenie $[p]$ je neseparabilní a její jádro má tedy řád ostře nižší než $\deg[p] = p^2$. Každý prvek $E[p]$ má ale řád dělící p, platí tedy buď $E[p] \cong \{\mathcal{O}\}$, či $\mathbb{Z}_p$. První případ jistě znamená $E[p^e] \cong \{\mathcal{O}\}$ pro každé $e \geqslant 0$, nyní tedy předpokládeme $E[p] \cong \mathbb{Z}_p$.

Dále postupujme silnou indukcí podle $e \geqslant 1$, $e = 0, 1$ je dáno. Dále ať dané tvrzení platí pro všechna nezáporná čísla nepřevyšující e. Isogenie $[p]$ je surjektivní, tedy pro každé $f \leqslant e$ a P bod řádu p^f existuje bod Q splňující $[p]Q = P$, jehož řád je p^{f+1}. Speciálně existuje bod $P_0 \in E[p^{e+1}]$ řádu p^{e+1}. Takový bod ale existuje díky $E[p^e] \cong \mathbb{Z}_{p^e}$ pouze jeden a $E[p^e] \cong \mathbb{Z}_{p^e}$. □

Předchozí věta ukazuje, že existují dvě rodiny křivek s drasticky odlišnými $[p]$-torzemi. Abychom si je mohli vložit do správných přihrádek, zavedeme nové názvosloví:

Definice 1.5.5. Pokud máme $E[p] \cong \{\mathcal{O}\}$, nazveme E *supersingulární*. Jinak E budeme říkat *obyčejná*.

Jak jsme zmínili před chvíli, pro nesoudělná m, n platí $E[m] \times E[n] \cong E[mn]$, tedy pomocí předchozího páru vět jsme schopni kompletně charakterizovat libovolnou torzní podgrupu E. Speciálně toho můžeme říci mnoho o samotné grupě bodů nad konečným tělesem $E(\mathbb{F}_q)$:

Věta 1.5.6. *Buď $E/\mathbb{F}_q$ eliptická křivka s $q = p^k$. Pak:*

$$E(\mathbb{F}_q) \cong \mathbb{Z}_m \times \mathbb{Z}_n$$

pro $p \nmid m \mid n$ přirozená čísla.

Důkaz. Pokud p nedělí řád $E(\mathbb{F}_q)$, který označme m, pak $E(\mathbb{F}_q) \subseteq E[m] \cong \mathbb{Z}_m \times \mathbb{Z}_m$ je podgrupa řádu nejvýše 2, lze ji proto zapsat jako direktní součin $\mathbb{Z}_m \times \mathbb{Z}_n$ s $m \mid n$ a $p \nmid mn$. Jinak existuje podgrupa $G \subseteq E(\mathbb{F}_q)$ řádu nejvyšší mocniny p, kde $E(\mathbb{F}_q) \cong G \times H$ a $H \cong \mathbb{Z}_m \times \mathbb{Z}_m$ nemá řád dělitelný p. Grupu $E(\mathbb{F}_q)$ tedy můžeme zapsat jako direktní součin nejvýše dvou cyklických grup a pouze jedna z nich má řád dělitelný p. □

Akce isogenie na libovolnou m-torzi na E a tak i na samotné $E(\mathbb{F}_q)$ můžeme jednoznačně určit z její akce na (nejvýše dva) generátory těchto grup a rozšířit lineárně. Docházíme tedy k tomu, že isogenie působí na $E(\mathbb{F}_q)$ i na její torzní podgrupy jako 2×2 celočíselná matice.

1.6 Supersingulární křivky

Slovo supersingulární napovídá, že na křivky takto pojmenované nenarazíme příliš často, že jsou mezi všemi eliptickými křivkami vzácné. Tato malá větev křivek se od obyčejných fundamentálně liší, přičemž jejich nespočetné rozdíly jsou spolu mnohdy těsně provázáné. Ve skutečnosti se mnohé vlastnosti, o kterých se zmíníme, berou jako ekvivalentní definice supersingularity, každá vhodná v jistém úhlu pohledu. Jejich vlastnosti ve všech směrech, které jsme prozatím studovali, do podrobna prozkoumáme, počínaje definicí pomocí torze.

Počítání celé p-torze je pro velká prvočísla výpočetně náročné, chtěli bychom najít vhodnější kritéria supersingularity. Ukáže se, že supersingulární eliptické křivky nesou pouze specifické počty bodů.

Věta 1.6.1. *Nechť E je křivka nad $\mathbb{F}_q$, kde $q = p^r$ je mocnina prvočísla $p > 3$. Pak:*

$$\#E(\mathbb{F}_q) \equiv 1 \pmod{p}$$

nastane právě pokud E je supersingulární.

Důkaz. Věta 1.3.8 říká:

$$[\deg([1] - \pi)] = ([1] - \pi) \circ (\widehat{[1] - \pi}) = ([1] - \pi) \circ (\widehat{[1]} - \widehat{\pi}) = ([1] - \pi) \circ ([1] - \widehat{\pi}),$$

neboli, protože isogenie jsou homomorfismy grup, isogenie:

$$\pi + \widehat{\pi} = [1] - [\deg([1] - \pi)] + \pi \circ \widehat{\pi} = [1] - [\deg([1] - \pi)] + [p]$$

působí jako skalární násobení na E. Isogenie $[1] - \pi = [1] - \pi_p^r$ má jádro $E(\mathbb{F}_q)$, protože tato množina je pod Frobeniovým morfismem invariantní. Navíc $-\pi$ je neseparabilní a $[1]$ zase separabilní, tedy věta 1.4.7 tvrdí, že $[1] - \pi$ je isogenií separabilní se stupněm rovným velikosti jádra, $\#E(\mathbb{F}_q)$. Pak tedy platí:

$$\pi + \widehat{\pi} = [1] - [\deg([1] - \pi)] + [p] = [1 - \deg([1] - \pi) + p] = [p + 1 - \#E(\mathbb{F}_q)].$$

Pokud E je supersingulární, je $\ker \pi \circ \widehat{\pi} = \ker[p] \cong \{\mathcal{O}\}$, neboli $\widehat{\pi}$ má triviální jádro a je neseparabilní. Podle věty 1.4.7 je $\pi + \widehat{\pi}$ neseparabilní, $[p+1-\#E(\mathbb{F}_q)]$ je proto též. Konečně,

díky lemmatu 1.5.1 p dělí $p+1-\#E(\mathbb{F}_q)$.

Naopak pokud platí $E(\mathbb{F}_q) \equiv 1 \pmod{p}$, isogenie:

$$\pi + \widehat{\pi} = [p+1-\#E(\mathbb{F}_q)]$$

je neseparabilní. Víme, že π je neseparabilní isogenie a $\pi + \widehat{\pi}$ taky, opět utilizujeme větu 1.4.7, dle které i $\widehat{\pi}$ není separabilní. Protože stupeň $\widehat{\pi}$ je prvočíselný, $\widehat{\pi}$ má nutně triviální jádro, kompozice $[p] = \widehat{\pi} \circ \pi$ jej proto má též a E je supersingulární. □

Poznámka. Fakt, že $\phi + \widehat{\phi}$ je rovno násobící mapě $[m]_E$ pro nějaké m zřejmě není unikátní pro Frobeniův endomorfismus, stejný postup můžeme replikovat pro každou jinou isogenii. My si však tento fakt „připomeneme“ na vhodnějším místě ve čtvrté kapitole.

Poznámka. Pozorování, že $\pi + \widehat{\pi} = [p+1-\#E(\mathbb{F}_q)]$ a že isogenie působí na torzní grupy jako 2×2 matice nám pomůže podat důkaz Hasseho věty s pomocí znalostí, které nyní máme, spolu s trochu hlubším studiem akce isogenií na torzní grupy. Naznačme jej tu rychle, plný důkaz se nachází na [39, Thm. 8.1, Thm. 7.17]. Pokud M je 2×2 matice udávající akci π na nějakou fixní torzi $E[n]$, pro libovolná celá r, s lze fakt $\deg([r] \circ \pi - [s]) \geqslant 0$ pro dostatečně velké n převést na nezápornost determinantu matice $rM - 1s$, což lze upravit na nezápornost kvadratického polynomu. Konečně se ukáže, že nekladnost jeho diskriminantu je jen jiná forma Hasseho věty.

Důsledek 1.6.2. *Ať E je křivka nad $\mathbb{F}_p$ s $p > 3$. Pak:*

$$\#E(\mathbb{F}_p) = p + 1$$

nastane, právě pokud E je supersingulární.

Důkaz. Pokud $\#E(\mathbb{F}_p) = p+1$, tak dle předchozí věty je E supersingulární. Pro E supersingulární je $\#E(\mathbb{F}_p) \equiv 1 \pmod{p}$, tedy jestli $\#E(\mathbb{F}_p) \neq p+1$, je číslo $p+1-\#E(\mathbb{F}_p)$ v absolutní hodnotě alespoň p. Dle Hasseho věty 1.1.5, kterou a priori bereme za platnou, toto číslo v absolutní hodnotě nepřesahuje $2\sqrt{p}$, neboli:

$$2\sqrt{p} \geqslant |p+1-\#E(\mathbb{F}_p)| \geqslant p,$$

což je spor s $p > 3$. □

Při zkoumání počtu bodů na supersingulárních křivek jsme narazili na číslo $t = q+1-\#E(\mathbb{F}_q)$, které je úzce spojené s Frobeniovým endomorfismem. Tento pár spolu rozhodně nevidíme naposledy, kapitola zaměřena na okruhy endomorfismů jejich pouto prohloubí.

Samotné počítání bodů na eliptické křivce je pro nás zatím obtížný úkon, pro $\mathbb{F}_p$ s malým p můžeme jednoduše projít všechny možné hodnoty x, jak můžeme vidět na následujícím příkladu:

Příklad 1.6.3. Ukažme, že křivka:

$$E : y^2 = x^3 + 10x + 7$$

nad $\mathbb{F}_{13}$ je supersingulární.

Řešení. Mějme $(x, y) \in E(\mathbb{F}_{13})$. Pokud je číslo $x^3 + 10x + 7$ v $\mathbb{F}_{13}$ nenulový čtverec, existují dvě vyhovující y, jedno, pokud je rovno nule, a jinak žádné. Můžeme si proto vypsat hodnoty pravé strany ve všech možných hodnotách a za pomocí Eulerova kritéria snadno určit, zda je výraz čtvercem, viz následující tabulka:

x	$x^3 + 10x + 7$	$\left(\frac{x^3+10x+7}{13}\right)$	počet řešení
0	7	-1	0
1	5	-1	0
2	9	1	2
3	12	1	2
4	7	-1	0
5	0	0	1
6	10	1	2
7	4	1	2
8	1	1	2
9	7	-1	0
10	2	-1	0
11	5	-1	0
12	9	1	2

Spolu s bodem v nekončnu je $\#E(\mathbb{F}_{13}) = 13 + 1 = 14$ a jsme hotovi z důsledku 1.6.2. □

U speciálních případů křivek můžeme rafinovaně využít poznatky z elementární teorie čísel:

Příklad 1.6.4. Ukažme, že křivka:

$$E/\mathbb{F}_p : y^2 = x^3 + kx$$

pro $p \equiv -1 \pmod 4$ je supersingulární.

Řešení. Pro $p \equiv -1 \pmod 4$ je $\left(\frac{-1}{p}\right) = -1$, takže pokud pro a, b platí $p \mid a^2 + b^2$, jsou obě dělitelná p. V opačném případě totiž z $a^2 \equiv -b^2 \pmod p$ vyvodíme:

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

spor. Nenulových čtverců v $\mathbb{F}_p$ je právě $\frac{p-1}{2}$, tudíž každý prvek $\mathbb{F}_p$ je buď čtverec, nebo mínus čtverec. Pro $x = 0$ máme pouze $y = 0$ a pro každé $x \in \mathbb{F}_p^*$ je právě jedno z čísel

$x^3 + kx, (-x)^3 - kx$ nenulovým čtvercem, protože je $x^2 \neq -1$. Pro každou dvojici $(x, -x)$ tak máme právě dvě řešení, dohromady $p - 1$. Spolu s $(0, 0)$ a bodem v nekonečnu je $\#E(\mathbb{F}_p) = p + 1$, díky větě 1.6.2 je E supersingulární. □

Příklad 1.6.5. Ukažme, že křivka:

$$E/\mathbb{F}_p : y^2 = x^3 + k$$

pro $p \equiv -1 \pmod 3$ je supersingulární.

Důkaz. Ukážeme, že třetí mocnina je na $\mathbb{F}_p$ bijekcí. Pokud totiž pro $x \neq y$ platí $x^3 \equiv y^3 \pmod p$, tak:

$$p \mid (x - y)(x^2 + xy + y^2) \Rightarrow p \mid x^2 + xy + y^2$$

Ukážeme, že pak už $p \mid x, y$, v opačném případě p nedělí ani jedno. Poslední rovnost pak vynásobíme čtyřmi a máme:

$$p \mid (x + 2y)^2 + 3x^2 \Rightarrow \left(\frac{x + 2y}{x}\right)^2 \equiv -3 \pmod p.$$

Pro $p \equiv -1 \pmod 3$ je ale -3 kvadratický nezbytek, opět získáváme spor. Pro každé $y \in \mathbb{F}_p$ tedy existuje unikátní třetí odmocnina z $y^2 - k$ dávající bod $(x, y) \in E$. Dohromady máme na E přesně p afinních bodů a ten poslední samozřejmě leží v nekonečnu. □

Protože supersingularita nezávisí na konkrétním rozšíření, křivky výše jsou supersingulární nad libovolným konečným tělesem s charakteristikou po řadě $p \equiv -1 \pmod 4$, resp. $p \equiv -1 \pmod 3$.

Náš první postup počítání počtu bodů na křivce běží nejlépe v $O(p)$ čase, což je pro prvočísla $\log_2(p) > 500$, tedy praktické kryptografické velikosti, jednoduše příliš pomalé. Jedním z nejdřívějších velkých pokroků v oblasti počítání bodů byl *Schoofův algoritmus*, zveřejněn roku 1985 v [35], který $\#E(\mathbb{F}_q)$ jako první dokáže spočíst deterministicky v čase polynomiálním v $\log(q)$. Poskytuje tedy exponenciální zrychlení oproti našemu předchozímu postupu.

Pojďme se podívat na samotnou strukturu bodů na supersingulární E nad konečným tělesem. Ústřední při našem studiu isogenií je fakt, že supersingularita je pod akcí isogenie zachována.

Věta 1.6.6. *Buď $E/\mathbb{F}_q$ eliptická křivka s $q = p^k$ a $\phi : E \longrightarrow E'$ libovolná isogenie vycházející z E. Pak E je supersingulární, právě pokud je E' supersingulární.*

Důkaz. Mějme $\phi : E \longrightarrow E'$ isogenii. Protože isogenie jsou homomorfismy grup bodů na křivkách nad $\mathbb{F}_q$, speciálně zachovají p-násobení:

$$\phi \circ [p]_E = [p]_{E'}$$

a analogická rovnost platí pro duální isogenii. Pokud na p-torzi jedné z křivek existuje netriviální bod, tak leží v p-torzi i druhé křivky, tedy pokud jedna z křivek je obyčejná, obě jsou. Naopak pokud p torze na E triviální, díky $[p]_{E'} = \phi \circ [p]_E$ je i $E'[p] \cong \{\mathcal{O}\}$ a samozřejmě i naopak. □

Speciálně toto tvrzení platí pro isomorfismy, každý j-invariant je proto exklusivní buď obyčejným, či supersingulárním křivkách, můžeme tedy každý nazvat obyčejným nebo supersingulárním podle typu křivek jej sdílejících.

Pokud uvážíme graf všech j-invariantů nad $\overline{\mathbb{F}}_p$ (kterým přiřadíme jejich příslušnou třídu isomorfismů), kde dva vrcholy jsou propojené právě pokud jejich příslušné křivky jsou isogenní, získáme neorientovaný(!) graf rozdělený na obyčejné a supersingulární komponenty.
z každého vrcholu vede $0, 1, 2, p+1$ hran, přičemž supersingulární komponenty jsou $p+1$-regulární, zatímco komponenty obyčejné tvoří zásadně odlišnou strukturu, tzv. *vulkány*, kde „kráter“ je tvořen regulárním grafem stupně nejvýše 2 a každý jiný vrchol je buď listem, či má $p + 1$ sousedů.

Počet supersingulárních j-invariantů je floor(p/12)

Věta 1.6.7. *Buď E supersingulární eliptická křivka nad $\mathbb{F}_q$. Pak $j(E) \in \mathbb{F}_{p^2}$.*

Důkaz. Isogenie $[p]$ na supersingulární křivce E je neseparabilní s triviálním jádrem a stupněm p^2. Podle věty 1.4.3 je pak rovna složení dvou kopii Frobenia s isomorfismem, $[p] = \iota \circ \pi \circ \pi$. Isogenie π^2 zobrazuje:

$$\pi^2 \quad : \quad E : y^2 = x^3 + ax + b \quad \longrightarrow \quad E' : y^2 = x^3 + a^{p^2}x + b^{p^2},$$

tyto dvě křivky jsou proto isomorfní pod ι. Pak díky vlastnostem charakteristiky:

$$j(E) = j(E') = 1728\frac{4a^{3p^2}}{4a^{3p^2} + 27b^{2p^2}} = \left(1728\frac{4a^3}{4a^3 + 27b^2}\right)^{p^2} = j(E)^{p^2},$$

j-invariant naší křivky je tedy fixovaný automorfismem $x^{p^2} = x$ na $\overline{\mathbb{F}}_q$ a leží tak v $\mathbb{F}_{p^2}$. □

Kapitola 2

Moderní kryptografie

Přes Caesarovu šifru až po šifrování za pomocí Enigmy v období druhé světové války, po většinu lidské historie se využívaly kryptografické systémy založené na faktu, že obě komunikující partie si po domluvě vyberou způsob maskování zprávy a ten pro ostatní zůstává skrytý. Příkladem je právě o kolik písmen v Caesarově šifře transponujeme. Tento způsob nutně závisí na faktu, že se obě strany před výměnou mají možnost přes bezpečný kanál na tomto způsobu domluvit. S přibývajícím počtem účastníků a frekvencí komunikace, na příklad našeho každodenního interagování na internetu, kde musí konverzace mezi všemi účastníky být bezpečná, je bohužel na úkor ceny přenosu třeba vyšší počet a velikost klíčů, a příbývá risk kompromitace.

Kvůli takovým obavám přišli Whitfield Diffie a Martin Hellman [13] roku 1976 s revolučním nápadem: asymetrickou kryptografií, kde každý z účastníků má svůj vlastní *privátní klíč*, který s nikým nesdílí. Všechny strany, i potenciální útočník, znají několik informací, které jsou známé jako *veřejné parametry*. Obě komunikující strany za pomocí veřejných informací tajně transformují svůj privátní klíč a výsledek, který budeme nazývat *veřejným klíčem*, publikují. Oba účastnící vezmou veřejný klíč toho druhého a provedou s ním ty samé tajné kroky závisící na jejich privátním klíči. Podstatou takové výměny je, že na jejím konci získají obě původní strany netriviální informaci, tedy informaci takovou, že žádná třetí strana ji nedokáže snadno uhodnout, za pomocí níž poté mohou společnou komunikaci šifrovat a nikdo jiný již jejich zprávy neuvidí. Předpokládá se, že pouze ze znalosti veřejného klíče je pro každou další partii těžké replikovat klíč privátní a že pole možných sdílených informací je obrovské. Vyhneme se tak přímočarým řešením hrubou silou.

Pojďme se podívat na protokol, který Diffie a Hellman navrhli. Budeme o něm dále mluvit jako o *Diffie-Hellmanově výměně*. Je založena na problému *diskrétního logaritmu* prvku $a \in \mathbb{Z}_p^*$. Tento problém po nás ze znalosti primitivního prvku g modulo p žádá najít k splňující $g^k = a$ v $\mathbb{Z}_p$. Obecně můžeme $\mathbb{Z}_p$ nahradit cyklickou grupou G a g je její generátor. Protokol požaduje, aby nebyl diskrétní logaritmus spočitatelný efektivně, tj. v polynomiálním čase vzhledem k velikosti grupy, jinak může útočník jednoduše privátní

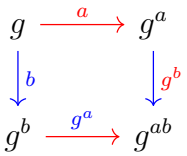
klíče obou stran spočíst, ale mocnění bylo. Umocnit číslo dokážeme v logaritmickém čase, a v konečné grupě nám stačí umocnit pouze na exponent modulo řádu grupy.

Veřejné parametry: Grupa G řádu p, kde p je prvočíslo, s generátorem g.

Alfréd		Blažena
Vstup: $a \in G^*$		**Vstup:** $b \in G^*$
	$\xrightarrow{g^a}$	
	$\xleftarrow{g^b}$	
$G_{AB} := \left(g^b\right)^a = g^{ab}$		$G_{BA} := \left(g^a\right)^b = g^{ba}$
Výstup: G_{AB}		**Výstup:** G_{BA}

Algoritmus 1: Diffie-Hellmanova výměna

Díky předpokladu, že G je cyklická, je i abelovská, tedy $G_{AB} = g^{ab} = g^{ba} = G_{BA}$.



Řád G se prakticky bere prvočíslo $q = 2p + 1$ takové, že p je prvočíslo, pak p nazveme tzv. *Sophie-Germainovým prvočíslem* a q zase *bezpečným prvočíslem*. V takovém případě má G podgrupu (velkého) prvočíselného řádu p, což je z kryptografického hlediska žádané, tuto grupu totiž je o to obtížnější spočíst. Navíc bezpečná prvočísla skýtají i výhody pro inicializování výměny, pro taková prvočísla dokážeme totiž snadno nalezneme primitivní kořen v $\mathbb{Z}_q$. Konkrétně, je-li g primitivní kořen modulo q, má řád $q - 1 = 2p$ modulo q, právě pokud g^p i g^2 nedávají zbytek 1 (mod q). Najít $g^p \pmod q$ nám mohou usnadnit nástroje jako Eulerovo kritérium, díky kterému je postačující mít g kvadratický nezbytek modulo q.

Veřejné klíče g^a, g^b, jsou nicméně, jak jejich název napovídá, veřejné, a má k nim přístup libovolná jiná osoba. Dejme tomu, že Eva, která má přístup pouze k veřejně dostupným informacím G, g, g^a, g^b, by chtěla též znát sdílené tajemství. ??, jak by mohla tajnou informací získat, je pokud by spočítala diskrétní logaritmus $\log_g(g^a) = a$, nicméně předpokládáme, že to je obtížné. Na klasických počítačích jsou nejlepší známé útoky na problémy, jako diskrétní logaritmus a faktorizace čísla, na čemž jsou založené mnohé známé protokoly, subexponenciální, nicméně na počítačích kvantových jsou už od poloviny 90. let známé algoritmy polynomiální. V čem však takto podstatné zrychlení spočívá?

?? složitost

2.1 Kvantové počítače

If computers that you build are quantum,
Then spies of all factions will want 'em.
Our codes will all fail,
And they'll read our email,
Till we've crypto that's quantum, and daunt 'em.

Jennifer a Peter Shorovi

Ve světě kvantových obvodů místo s klasickými bity pracuje s *qubity*. V n bitovém systémů máme 2^n různých stavů, které v n qubitovém systému tvoří generátory našeho prostoru. Podstatou je, že před pozorováním nemá daný qubit jednu z těchto hodnot, leč jejich (komplexní) superpozici. Generátory systému s jedním qubitem jsou stavy $|0\rangle = \begin{bmatrix}1\\0\end{bmatrix}, |1\rangle = \begin{bmatrix}0\\1\end{bmatrix}$, systém je tedy:

$$\alpha|0\rangle + \beta|1\rangle,$$

kde α, β jsou komplexní čísla $|\alpha|^2 + |\beta|^2 = 1$. Zápis $|\psi\rangle$ je tzv. *ket* notace, kde ψ je vektor.

V dvojqubitovém systému máme čtyři báze a stav takového systému je:

$$\alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle,$$

kde $\alpha, \beta, \gamma, \delta$ jsou komplexní čísla s $|\alpha|^2+|\beta|^2+|\gamma|^2+|\delta|^2 = 1$. Qubity jsou značně nestabilní a musí být uchovány v izolované soustavě, nejčastěji v neutrinu. Jakékoli narušení, i pouhé pozorování hodnoty qubitu, ho kolapsuje qubit jedinou hodnotu, kterou už pak zůstane. Při pozorování má qubit pravděpodobnost ukázat stav právě takovou, kolik je druhá mocnina absolutní hodnoty příslušného koeficientu, proto ona normalizační podmínka. Pokud bychom pozorovali náš jedno-qubitový systém, s pravděpodobností $|\alpha|^2$ získáme výstup 0, s pravděpodobností $|\beta|^2$ získáme 1.

Náš systém dvou qubitů můžeme kompaktněji zapsat ve vektorovém tvaru:

$$|\psi\rangle = \begin{bmatrix}\alpha\\\beta\\\gamma\\\delta\end{bmatrix},$$

což samozřejmě zobecníme pro systémy více qubitů. Tento vektor je díky naší podmínce jednotkový. V klasických obvodech jsou bity ovlivňovány branami, které lineární zobrazují naše stavy, příklady takových bran jsou OR a NOT. V kvantových obvodech bereme jako brány právě unitární matice a operaci násobení, takové brány totiž zachovávají normu vektoru jejich aplikací proto obdržíme opět qubit.

Nedá moc práce adaptovat klasické brány do kvantového prostředí, model kvantového počítače jakožto obvodu je tak alespoň stejně silný jako klasický Turingův model (simulovat quantum na klasickém). Tím ale kvantový počítač zdaleka nekončí, nebavili bychom se přece jenom tak o ekvivalentu standardního počítače.

Již v 80. letech minulého století, kdy užití kvantové mechaniky ve výpočetní technice bylo pořád ve svých kojeneckých letech,

Jedním z důvodů, proč se věří, že s veřejně dostupnými kvantovými počítači přijde nová éra výpočetní techniky je, že existují procesy, o kterých se dodnes neví, zda jsou v polynomiálním čase proveditelné na počítači klasickém, jejichž kvantové implementace byly již nalezeny. Klasické násobení (n bitových) čísel, zabere $O(n^2)$ operací, případně až $O(n^2 \log n)$ pro velká čísla. Násobení dvou čísel se dá redukovat na problém násobení dvou polynomů stupně n, přičemž diskrétní Fourierova transformace podá informaci o hodnotách polynomu v n-tých odmocninách z jednotky, což je vše, co potřebujeme k určení celého polynomu. Rychlá Fourierova transformace tento úkon dokáže pouze v $\Theta(n \log n)$ operacích. Díky multiplikativitě, linearitě a funkci inverzní k Fourierově transformaci pak dokážeme zpětně v tomto čase zjistit součin našich čísel.

Kvantová Fourierova transformace, která obdobnou operaci aplikuje na náš vektor, na klasickém počítači s n qubity počítá s 2^n prvky. Nejlepší známé algoritmy ji provádí v $O(n^2 2^n)$ operacích, zatímco na kvantových počítačích pracuje v kvadratickém čase a s jistou přesností i v $\Theta(n \log n)$, viz [3, Ch. 4. a 5.] pro více informací. Důvodem je, že byť výsledek získáme asymptoticky rychleji, pravděpodobnost jeho správnosti je na oplátku snížena.

Vynásobit dvě matice řádu n na klasickém počítači zjevně nedokážeme rychleji než v řádově n^2 operacích, neboť musíme pracovat se všemi prvky matice. Kvantový počítač dokáže dvě matice řádu n vynásobit užitím $O(n^{5/3})$ kvantových operací, což přijde vhod i při aplikaci kvantových bran, dvě po sobě jdoucí brány totiž působí na vektor jako jejich součin. To možná může znít až nemožně, nicméně kvantový svět se mnohdy nechová tak, jak bychom čekali. Pro jedno, Lov Grover [16] navrhl v 90. letech kvantový algoritmus, který s vysokou pravděpodobností nalezne v poli o n prvcích jeden konkrétní v $O(\sqrt{n})$ čase (což je asymptoticky optimální), v klasické mechanice očekávaný čas nalezení bude kvadraticky pomalejších $n/2$ operací.

Další problém notoricky klasicky obtížný je rozklad celého čísla na prvočinitele. Jistě mocniny dvojky můžeme zanedbat, zbyde nám tedy najít lichá a, b se součinem n. Tento problém lze snadno převést na hledání řádu čísla a modulo n. V devadesátých letech minulého století přišel Peter Shor [36] s řešením problému diskrétního logaritmu na kvantovém počítači užívajícím polynomiálního počtu kvantových bran, čímž je řešen i problém rozkladu celého čísla. Mnoho v té době užívaných protokolů na šifrování a podpis dat bylo založeno na jednom z těchto dvou problém, nejprominentější z nich je známý jako RSA [31]. Tento objev pochopitelně způsobil paniku mezi kryptografickou komunitou, všechny

nápady jenom vzdáleně spojené s diskrétním logaritmem musely být smeteny ze stolu.

Poznamenejme, že nejefektivnější známé klasické algoritmy rozkládající velká čísla (dejme tomu $\log_2(n) > 200$) užívají poznatky z teorie eliptických křivek a algebraické teorie čísel, na kterou ještě přijde řeč. Tyto algoritmy nesou názvy *Elliptic-curve factorization* a *General number field sieve*, první z nich je založen. Oba běží v očekávaném subexponenciálním čase.

2.2 ?

???? Zjevnou adaptací Diffie-Hellmanova protokolu je výměna, která nese název ECDH (Elliptic Curve Diffie-Hellman):

Veřejné parametry: Prvočíslo p, eliptická křivka $E/\mathbb{F}_p$ a bodem $G \in E(\mathbb{Z}_p)$ vysokého řádu.

Alfréd		Blažena
Vstup: $a \leqslant \#E(\mathbb{Z}_p) - 1$		**Vstup:** $b \leqslant \#E(\mathbb{Z}_p) - 1$
	$\xrightarrow{[a]G}$	
	$\xleftarrow{[b]G}$	
$G_{AB} := [a]([b]G) = [a][b]G$		$G_{BA} := [b]([a]G) = [b][a]G$
Výstup: G_{AB}		**Výstup:** G_{BA}

Algoritmus 2: Protokol ECDH

Tento protokol je založen na předpokladu, že diskrétní logaritmus na eliptických křivkách, tedy ze znalosti P a $[n]P$ spočíst n, je těžký problém. Není znám žádný algoritmus, který by nezískal společné tajemství výpočtem privátních klíčů obou stran. ???

Kapitola 3

Algebraická teorie čísel

Ve snaze vybudovat teorii k hlubšímu studiu eliptických křivek a isogenií, natož diskuzi prakticky užívaných protokolů, se musíme na tyto objekty podívat v naprosto odlišném světle. Opustíme proto na okamžik eliptické křivky a ponoříme se do říše algebraické teorie čísel.

Na světě se nachází myriáda kvalitních a podrobných materiálů ke studiu této krásné oblasti matematiky, já osobně vřele doporučuji texty [18], [25], [27] či [29]. Jako velmi stručný úvod motivovaný poznatky z elementární teorie čísel může též posloužit má SOČ, [28].

3.1 Moduly nad okruhem

Při definici vektorového prostoru požadujeme, aby byl sestrojen nad tělesem. Objekt mající obdobné vlastnosti můžeme však obecněji sestrojit nad libovolným okruhem.

Definice 3.1.1. Mějme grupu G s a množinu X. Pod *levou akcí* G na X rozumíme zobrazení $\cdot : G \times X \longrightarrow X$, pro které platí $1 \cdot x = x$ a $g \cdot (h \cdot x) = (g \cdot h) \cdot x$ pro $g, h \in G, x \in X$.

Definice 3.1.2. Akci $\cdot : G \times X \longrightarrow X$ nazveme *volnou*, pokud pro libovolná $x \in X$ a $g \in G$ rovnost $g \cdot x = x$ znamená $g = 1$.

Definice 3.1.3. Grupu M s operací $+$ pro okruh R nazveme *levým R-modulem* s akcí $\cdot : R \times M \longrightarrow M$, pokud $\cdot$ je asociativní a na $+$ oboustranně distributivní.

Analogicky definujeme i pravou akci G na X a pravý modul.

Vzpomeňme na definici volné grupy G, jakožto $G \cong \mathbb{Z}^r$ pro nějaké celé r, obdobně definujeme i volný modul.

Definice 3.1.4. Modul M okruhu R nazveme *volným*, pokud má $R-$bázi, tj. pro nějaká $m_i \in M$ lineárně nezávislá nad R je $M = \{r_1 m_1 + \cdots + r_k m_k | r_i \in R\}$. Říkáme, že množina $\{m_1, \ldots, m_k\}$ *generuje* M.

Definice 3.1.5. Buď M volný R-modul. Pokud je k nejmenší přirozené číslo takové, že existuje k prvků M generujících M nad R, řekneme, že *R-rank M* je k.

Pro R těleso je M volným modulem, tedy vektorovým prostorem nad R, protože každý vektorový prostor vyžaduje existenci báze. R-rank M je pak roven stupni rozšíření $[M : R]$.

Nejprve si ukážeme jednoduchý způsob, jak poznat, zda je grupa $\mathbb{Z}$-modulem.

Příklad 3.1.6. Ukažme, že grupa je abelovská, právě pokud je $\mathbb{Z}$-modulem.

Důkaz. Každá abelovská grupa G s operací $+$ je $\mathbb{Z}$-modulem s akcí $n \cdot a$, jakožto součet n čísel $a \in G$, pro záporná čísla $(-n) \cdot a = -(n \cdot a)$. Navíc pro $\mathbb{Z}$-modul s jednotkou 1 s operací $+$ platí:

$$x+y+x+y = 1\cdot(x+y)+1\cdot(x+y) = (1+1)\cdot(x+y) = (1+1)\cdot x+(1+1)\cdot y = x+x+y+y,$$

tedy $y + x = x + y$. □

Každý komutativní okruh R je volným R-modulem, jehož R-rank je 1. Mezi volné $\mathbb{Z}$-moduly patří například okruh zbytkových tříd $\mathbb{Z}_{101}$ ranku 1, či okruh Gaussových celých čísel, jenž má $\mathbb{Z}$-rank 2. Naopak tělesa $\mathbb{Q}, \mathbb{C}$ jsou po řadě $\mathbb{Z}$-modul, resp. $\mathbb{Q}$-modul bez konečné báze, nejsou proto volné.

Poněkud zajímavějším příkladem modulu je grupa nejvýše kvadratických polynomů nad reálnými čísly $\mathbb{R}[x]/x^3\mathbb{R}$, což je volný $\mathbb{R}$-modul ranku 3 s bází $\{1, x, x^2\}$, či grupa $E[n]$ pro křivku nad K s $\operatorname{char} K \nmid n$, což je volný $\mathbb{Z}_n$-modul, který má díky větě **??** rank 2.

Poznámka. Roku 1922 Luis Mordell v [24] dokázal, že pro libovolnou eliptickou křivku E je grupa $E(\mathbb{Q})$ konečně generovaná. Tento výsledek rozšířil André Weil v roce 1928 pro libovolnou projektivní křivku nad číselným tělesem [43], což je pojem, který si za chvíli objasníme. Obecně charakterizovat tuto grupu, či efektivně spočíst její rank, jsou dnes problémy stále velmi obtížné. Clayův institut tuto oblast matematiky považoval za tak důležitou, že roku 2000 mezi problémy tisíciletí (Millenium Prize Problems) zařadil tzv. *Birch-Swinnerton-Dyerovu domněnku*, která se zabývá asymptotickým chováním $E(\mathbb{F}_p)/p$ vzhledem k ranku naší křivky.

Podmnožiny R-modulu uzavřené na sčítání a násobení prvky R jsou též R-moduly. Takový modul pak nazveme podmodulem.

Definice 3.1.7. Nechť M a N jsou R-moduly, přičemž N je podgrupa M. Pak N nazveme *podmodulem* M. *Index* podmodulu N v M definujeme jako počet prvků faktorgrupy M/N.

Věta 3.1.8. *Nechť M je volný $\mathbb{Z}$-modul a N jeho podmodul. Pak rank N je nejvýše tak velký, jako rank M. Speciálně je N volný.*

Hezký důkaz indukcí je podán v [29, Věta 1.3.8]. Pokud bychom však místo $\mathbb{Z}$ uvážili libovolný komutativní okruh R, tvrzení již ne nutně platí!

Příklad 3.1.9. Podmodul $\mathbb{Z}$-modulu $\mathbb{Z}_6$ $2\mathbb{Z}_6 = \{0, 2, 4\}$ není volný. Pokud by totiž modul $2\mathbb{Z}_6$ měl nad $\mathbb{Z}_6$ bázi, musely by její prvky nad $\mathbb{Z}_6$ být lineárně nezávislé. Nicméně $0 \cdot 3 = 2 \cdot 3 = 4 \cdot 3 = 0$, přičemž $3 \neq 0$ v $\mathbb{Z}_6$. Žádná podmnožina $S \subseteq 2\mathbb{Z}_6$ tedy není nad naším okruhem lineárně nezávislá, protože $3S = \{0\}$.

Obdobně vidíme, že pokud R je okruh, který není oborem integrity, obsahující nenulové prvky x, y se součinem 0, a M je jeho volný podmodul, pak xM je podmodul M, který není volný.

Čtenář se mohl setkat s pojmem *tenzorový součin* vektorových prostorů V a W, neboli vektorový prostor U disponující bilineárním zobrazení $V \times W \longrightarrow U$. My tuto definici rozvineme na moduly nad komutativním okruhem, tedy akci $R \times G \longrightarrow G$ rozšíříme na akci $M \times G \longrightarrow G$, kde M je R-modul.

Definice 3.1.10. Buďte R okruh a M a N jeho levý, resp. pravý modul. Uvažme prvky $m \in M, n \in N$ jednotlivých modulů. Pak *tenzorový součin* $m \otimes n$ definujeme jako výraz, který je na sčítání oboustranně distributivní a pro každé $r \in R$ splňuje:

$$(rm) \otimes n = r(m \otimes n) = m \otimes (rn).$$

Definice 3.1.11. Buďte R okruh a M a N jeho levý, resp. pravý modul. Pak *tenzorový součin* M a N je volný R-modul definovaný následovně:

$$M \otimes_R N = \left\{ \sum r_i m_i \otimes n_i \,\middle|\, r_i \in R, m_i \in M, n_i \in N \right\}.$$

Jeho prvky nazveme *tenzory*.

Tenzorový součin dvou modulů je až na isomorfismus unikátní.

Příklad 3.1.12. Buďte m, n nesoudělná celá čísla. Ukažme, že $\mathbb{Z}_m \otimes_{\mathbb{Z}} \mathbb{Z}_n = \{0\}$.

Řešení. Máme:

$$m(1 \otimes 1) = m \otimes 1 = 0 \otimes 1 = 0,$$
$$n(1 \otimes 1) = 1 \otimes n = 1 \otimes 0 = 0.$$

Dle Bezoutovy věty existují $x, y \in \mathbb{Z}$, že $xm + yn = 1$. Pak:

$$1 \otimes 1 = (xm + yn)(1 \otimes 1) = xm(1 \otimes 1) + yn(1 \otimes 1) = 0.$$

Pro každá $x \in \mathbb{Z}_m, y \in \mathbb{Z}_n$ pak platí $x \otimes y = x(1 \otimes y) = xy(1 \otimes 1) = 0$. □

Případ $N = \mathbb{Q}$ a $R = \mathbb{Z}$ je zajímavější:

Věta 3.1.13. *Pokud je M $\mathbb{Z}$-modul, každý prvek $\mathbb{Q} \otimes_{\mathbb{Z}} M$ se dá zapsat ve tvaru $r \otimes m$ pro $r \in \mathbb{Q}, m \in M$.*

Důkaz. Je postačující ukázat, že pro $x, y \in \mathbb{Q}, m, n \in M$ se $x \otimes m + y \otimes n$ dá vyjádřit v takovém tvaru. Zvolme celá a, b, c splňující $x = \frac{a}{c}, y = \frac{b}{c}$. Pak:

$$\frac{a}{c} \otimes m + \frac{b}{c} \otimes n = \frac{1}{c} \otimes am + \frac{1}{c} \otimes bn = \frac{1}{c} \otimes (am + bn),$$

kde $am + bn \in M$, je hledaného tvaru. □

3.2 Číselná tělesa

Za pomocí vlastností modulů můžeme začít studovat konečná rozšíření racionálních čísel, tzv. číselná tělesa.

Definice 3.2.1. Komplexní číslo α, které je kořenem polynomu $P \in \mathbb{Z}[x]$, nazveme *algebraické*. Pokud je navíc α kořenem monického (normovaného) polynomu nad $\mathbb{Z}$, nazveme jej *celým algebraickým* číslem.

Definice 3.2.2. Konečná rozšíření racionálních čísel obsahují pouze čísla algebraická, tato tělesa proto nazveme *algebraická číselná tělesa*, pro jednoduchost je budeme nazývat pouze *číselná tělesa*.

Definice 3.2.3. Pod stupněm číselného tělesa rozumíme stupni jeho rozšíření nad $\mathbb{Q}$ jakožto vektorového prostoru. Číselná tělesa stupně 2 nazveme *kvadratická*.

Jistě obor komplexních čísel s racionální reálnou i imaginární složkou je kvadratickým tělesem, jako je též těleso $\mathbb{Q}(\sqrt{2})$. Obecně každé těleso dáno rozšířením $\mathbb{Q}$ o jednu jedinou odmocninu je kvadratické. Opačná inkluze je též nasnadě:

Věta 3.2.4. *Buď K kvadratické těleso. Pak $K = \mathbb{Q}(\sqrt{m})$ pro nějaké celé bezčtvercové m.*

Důkaz. K je vektorový prostor nad $\mathbb{Q}$ stupně dvě, má tedy nad racionálními čísly bázi $\{1, \theta\}$ a K je rozšířením $\mathbb{Q}(\theta)$ pro algebraické θ. Číslo θ^2 náleží do K, musí proto existovat vyjádření $a + b\theta = \theta^2$ pro a, b racionální čísla, tedy $\theta = \frac{s+t\sqrt{m}}{2}$ pro vhodná racionální s, t. Pak $K = \mathbb{Q}\left(\frac{s+t\sqrt{m}}{2}\right) = \mathbb{Q}(\sqrt{m})$. □

Pro $m > 0$ nazveme K *reálným* kvadratickým tělesem, v opačném případě ($m < 0$) jej nazveme *imaginárním* kvadratickým tělesem. Pokud m je čtvercem celého čísla, je K rovno $\mathbb{Q}$, není tedy kvadratickým tělesem.

Toto tvrzení můžeme rozšířit na všechna číselná tělesa. Konkrétně těleso K je jednoduchým rozšířením $\mathbb{Q}(\theta)$ pro algebraické číslo θ, právě pokud obsahuje pouze algebraická čísla, jak je ukázáno v [32, Věta 11.12]. Dokonce si takové θ můžeme zvolit celé algebraické, viz [27, Lemma 4.3.8]. Báze K jakožto vektorového prostoru je poté $\{1, \theta, \ldots, \theta^{n-1}\}$, kde $n = [K : \mathbb{Q}]$ je stupeň minimálního polynomu prvku θ nad racionálními čísly.

Poznámka. Je zajímavé uvážit případ rozšíření $\mathbb{Q}(\theta)$, kde θ není kořenem žádného polynomu s racionálními koeficienty, takové θ se nazývá *transcendentní*. Pak zobrazení dané $P \mapsto P(\theta)$ pro racionální lomenou funkci P dává bijektivní homomorfismus mezi tělesem racionálních lomených funkcí a $\mathbb{Q}(\theta)$, tato tělesa jsou proto isomorfní.

Pojďme si trochu charakterizovat celá algebraická čísla našeho tělesa.

Věta 3.2.5. *Komplexní číslo α je celé algebraické, právě pokud je $\mathbb{Z}[\alpha]$ volným $\mathbb{Z}$-modulem.*

Důkaz. Je-li α celé algebraické číslo s minimálním polynomem $f \in \mathbb{Z}[x]$ stupně n, pak $\mathbb{Z}[\alpha]$ je volný $\mathbb{Z}$-modul s bází $\{1, \alpha, \ldots, \alpha^{n-1}\}$, číslo α^k pro $k \geqslant n$ totiž dokážeme z $\alpha^{k-n}P(\alpha) = 0$ vyjádřit jako $\mathbb{Z}$-lineárních kombinace mocnin α ostře nižších k, protože je P monický.

Naopak pokud je $\mathbb{Z}[\alpha]$ volný $\mathbb{Z}$-modul, je generovaný prvky $f_i(\alpha) \in \mathbb{Z}[\alpha]$ pro polynomy $f_1, \ldots, f_k \in \mathbb{Z}[x]$. Pro číslo t ostře větší $\max(\deg f_i)$, leží α^t v $\mathbb{Z}[\alpha]$, je proto vyjádřitelné jako $\mathbb{Z}$-lineární kombinace $f_i(\alpha)$. Pro nějaká $a_i \in \mathbb{Z}$:

$$\alpha^t = \sum a_i f_i(\alpha),$$

tedy α je kořenem monického polynomu $x^t - \sum a_i f_i(x)$, dle definice je celé algebraické. □

Díky tomuto tvrzení můžeme jednoduše odůvodnit, proč necelá racionální čísla nejsou celá algebraická.

Příklad 3.2.6. Ukažme, že pro p, q nesoudělná celá s $|q| > 1$ je racionální číslo $\frac{p}{q}$ algebraické číslo, ale již není celé algebraické.

Důkaz. Číslo $\frac{p}{q}$ je kořenem polynomu $qx - p \in \mathbb{Z}[x]$, tedy je algebraické. Dále uvažme pro spor polynom $P = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ s kořenem $\frac{p}{q}$. Rovnost $P\left(\frac{p}{q}\right) = 0$ přenásobíme číslem q^n a získáme:

$$p^n + a_{n-1}p^{n-1}q + \cdots + a_1pq^{n-1} + a_0q^n = 0.$$

Dejme tomu, že $|q| > 1$, a uvažme prvočíslo r dělící q. Pak r dělí číslo $-(a_{n-1}p^{n-1}q + \cdots + a_1pq^{n-1} + a_0q^n) = p^n$, což je spor s faktem, že p a q jsou nesoudělná. Žádné takové prvočíslo proto neexistuje a $b = \pm 1$. □

Poznámka. Na toto tvrzení můžeme nahlížet i jako na problém ukázat, že okruh $\mathbb{Z}\left[\frac{p}{q}\right]$ není volným $\mathbb{Z}$-modulem. Pokud by totiž $\{a_1, \ldots, a_k\}$ byla jeho báze, posloupnost mocnin $\frac{p}{q}, \left(\frac{p}{q}\right)^2, \ldots, \left(\frac{p}{q}\right)^i, \cdots \in \mathbb{Z}\left[\frac{p}{q}\right]$ má pro prvočíslo $r \mid q$ klesající celočíselné hodnoty r-valuací. To je nicméně spor, protože množina $\{\nu_r(a_1), \ldots, \nu_r(a_k)\}$ je zdola omezená a platí $\nu_r(a + b) \geqslant \min\{\nu_r(a), \nu_r(b)\}$.

Důležitým faktem o celých algebraických číslech je, že v číselném tělese tvoří okruh, jak si dále ukážeme.

Věta 3.2.7. *Celá algebraická čísla číselného tělesa K tvoří okruh $\mathcal{O}_K$.*

Důkaz. Ukážeme, že součet a součin dvou algebraických čísel α a β je opět algebraické číslo. Mějme $\mathbb{Z}[\alpha]$ a $\mathbb{Z}[\beta]$ volné moduly a uvažme okruh $\mathbb{Z}[\alpha, \beta]$, jenž je množinou všech polynomů ve dvou proměnných nad celými čísly evaluovaných v bodě (α, β). Ten je abelovskou grupou a díky příkladu 3.1.6 i $\mathbb{Z}$-modulem.

V důkazu věty 3.2.5 jsme si ukázali, že pokud minimální polynom P_α má stupeň n, číslo α^k pro $k \geqslant n$ se dá vyjádřit jako $\mathbb{Z}$-lineární kombinace prvků α s mocninami ostře nižšími n. Víme, že $\mathbb{Z}[\alpha, \beta]$ je množinou $\mathbb{Z}$-lineárních kombinací čísel $\alpha^i \cdot \beta^j$, z čehož plyne, že $\mathbb{Z}[\alpha, \beta]$ je generovaný množinou $\{\alpha^i \beta^j | i \in \{0, 1, \ldots, n-1\}, j \in \{0, 1, \ldots, m-1\}\}$, kde minimální polynom β, P_β, má stupeň m. Okruh $\mathbb{Z}[\alpha, \beta]$ je proto volným $\mathbb{Z}$-modulem ranku mn.

Okruhy $\mathbb{Z}[\alpha+\beta]$ a $\mathbb{Z}[\alpha\beta]$ jsou díky jejich komutativitě $\mathbb{Z}$-moduly a navíc jsou oba zjevně podmoduly $\mathbb{Z}[\alpha, \beta]$. Díky větě 3.1.8 jsou oba volné (ranku nejvýše mn), tedy $\alpha + \beta$ a $\alpha\beta$ jsou celá algebraická čísla. Speciálně pro libovolné α celé algebraické je $-\alpha$ celé algebraické. Množina $\mathcal{O}_K$ celých algebraických čísel tělesa K proto tvoří okruh. □

Poznámka. Okruhy celých algebraických čísel značíme $\mathcal{O}_K$ a později uvedeme *pořádky*, které budeme povětšinou značit $\mathcal{O}$. Shodně jsme značili bod v nekonečnu na křivce, mějme proto na paměti kdy diskutujeme který pojem!

Pokud uvážíme algebraické číslo θ s minimálním polynomem nad celými čísly:

$$P : a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0,$$

pak $a_n\theta$ je kořenem monického polynomu:

$$P^* : x^n + a_n a_{n-1} x^{n-1} + a_n^2 a_{n-2} x^{n-2} + \cdots + a_n^n a_0,$$

toto číslo je tedy celé algebraické. Víme, že podílové těleso okruhu $\mathcal{O}_K$, které označíme L, je nejmenší těleso obsahující $\mathcal{O}_K$, tedy je podtělesem K. Navíc pro libovolné $\alpha \in K$ existuje celé m s $m\alpha \in \mathcal{O}_K$, tedy $\alpha = \frac{m\alpha}{m}$ je podílem dvou prvků $\mathcal{O}_K$, tudíž $K \subseteq L$, což nám stačí k závěru, že K je podílové těleso $\mathcal{O}_K$.

Dle příkladu 3.2.6 je okruhem celých algebraických čísel tělesa $\mathbb{Q}$ množina celých čísel. V libovolném kvadratickém tělese však dokážeme $\mathcal{O}_K$ za pomocí znalosti řešení kvadratické rovnice jednoduše popsat též.

Věta 3.2.8. *Nechť $m \neq 0, 1$ je bezčtvercové celé číslo a $K = \mathbb{Q}(\sqrt{m})$ je algebraické číselné těleso. Pak platí:*

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}[\sqrt{m}], & \text{pokud} \quad m \equiv 2, 3 \pmod 4, \\ \mathbb{Z}\left[\frac{1+\sqrt{m}}{2}\right], & \text{pokud} \quad m \equiv 1 \pmod 4. \end{cases}$$

Důkaz. Jistě $\mathbb{Z}[\sqrt{m}]$, resp. $\mathbb{Z}\left[\frac{1+\sqrt{m}}{2}\right]$, je podmnožinou $\mathcal{O}_K$, neboť minimální polynomy prvků $a+b\sqrt{m}$, resp. $a+b\frac{1+\sqrt{m}}{2}$, jsou po řadě $(x-a)^2 - bm^2$, resp. $(x-a)^2 - bx + ab + b^2\frac{1-m}{4}$.

Ze tvaru řešení kvadratické rovnice plyne, že prvky $\mathcal{O}_K$ jsou ve tvaru $\frac{a+b\sqrt{m}}{2}$ pro $a, b \in \mathbb{Z}$. Zjevně pro $b \neq 0$ sdílí $\frac{a+b\sqrt{m}}{2}$ a $\frac{a-b\sqrt{m}}{2}$ minimální polynom, musí nutně být:

$$\left(x - \frac{a+b\sqrt{m}}{2}\right)\left(x - \frac{a-b\sqrt{m}}{2}\right) = x^2 - ax + \frac{a^2 - b^2 m}{4}.$$

Pokud $\frac{a+b\sqrt{m}}{2} \in \mathcal{O}_K$, je tento monický polynom definovaný nad celými čísly. Proto $a^2 - b^2 m$ je dělitelné čtyřmi. Je-li m je sudé, je a též, tedy a^2 je dělitelné čtyřmi. Za předpokladu, že m je bezčtvercové, je $m \equiv 2 \pmod 4$, tedy i b je sudé.

Nyní již předpokládejme, že m je liché. Pokud je $m \equiv 3 \pmod 4$, platí $4 \mid a^2 + b^2$, což nutně znamená $2 \mid a, b$, protože kvadráty dávají zbytky $0, 1$ po dělení čtyřmi. Pak $\frac{a+b\sqrt{m}}{2} \in \mathbb{Z}[\sqrt{m}]$. Konečně uvažme $m \equiv 1 \pmod 4$. Máme $a^2 \equiv b^2 \pmod 4$, tedy $a \equiv b \pmod 2$. To ale znamená, že $\frac{a+b\sqrt{m}}{2} \in \mathbb{Z}\left[\frac{1+\sqrt{m}}{2}\right]$. □

Poznámka. Je-li čtenář obeznámen s pojmem *diskriminant číselného tělesa*, můžeme okruh $\mathcal{O}_K$ v kvadratickém tělese K vyjádřit kompaktněji jako $\mathbb{Z}\left[\frac{d+\sqrt{d}}{2}\right]$, kde d je diskriminant K.

Každé číselné těleso je jednoduchým rozšířením racionálních čísel, platí však obdobná vlastnost pro okruhy celých algebraických čísel a celá čísla? U kvadratických těles jsme si to právě potvrdili, tělesa vyšších řádů tentokrát tuto vlastnost ne nutně sdílí. Minimální příklad se dokonce nachází již mezi kubickými tělesy.

Ke konci této sekce ještě zběžně definujme *pořádky*, tj. podokruhy číselného tělesa, které mají rank shodný se stupněm tělesa.

Definice 3.2.9. Okruh $\mathcal{O}$ obsažen v číselném tělese K nazveme *pořádkem*, pokud je volným $\mathbb{Z}$-modulem ranku $[K : \mathbb{Q}]$.

Nejprve si všimněme, že $\mathcal{O}_K$ je pořádkem K. Dalším příkladem pořádku je okruh $\mathbb{Z}[3i]$ v $\mathbb{Q}(i)$, který není okruhem Gaussových celých čísel, je v něm ale obsažen.

Věta 3.2.10. *Nechť K je číselné těleso stupně n a $\mathcal{O}$ jeho pořádek. Pak $\mathcal{O}$ je podmodulem $\mathcal{O}_K$.*

Důkaz. Buď $\{a_1, \ldots, a_n\}$ báze $\mathcal{O}$ jakožto $\mathbb{Z}$-modulu. Protože $\mathcal{O} = \mathbb{Z}[a_1, \ldots, a_n]$ je volný modul a $\mathbb{Z}[a_i]$ jsou jeho podmoduly, podle věty 3.1.8 jsou všechny volné. Díky větě 3.2.5 jsou a_i celá algebraická čísla, tedy $a_i \in \mathcal{O}_K$. Protože $\mathbb{Z} \subseteq \mathcal{O}_K$, leží každá $\mathbb{Z}$-lineární kombinace a_i v $\mathcal{O}_K$, jinak řečeno $\mathcal{O} \subseteq \mathcal{O}_K$. □

O $\mathcal{O}_K$ tak můžeme hovořit jako o „maximálním“ pořádku.

Definice 3.2.11. Buď $\mathcal{O}$ pořádek číselného tělesa K. Pak *vodič* $\mathcal{O}$ *v* $\mathcal{O}_K$ definujeme jako index $|\mathcal{O}_K/\mathcal{O}|$.

Kromě faktu, že pořádky jsou $\mathbb{Z}$-moduly ranku n a obsaženy v okruhu $\mathcal{O}_K$, můžeme je přesně vzhledem k maximálnímu pořádku charakterizovat.

Věta 3.2.12. *Nechť $\mathcal{O}$ je pořádek číselného tělesa K stupně $n+1$. Pak existují čísla $a_1, \ldots, a_n \in K$ a celá $k_1, \ldots, k_n$ splňující $k_i \mid k_{i+1}$ a:*

$$\mathcal{O}_K = \mathbb{Z}[a_1, a_2, \ldots, a_n], \qquad \mathcal{O} = \mathbb{Z}[k_1 a_1, k_2 a_2, \ldots, k_n a_n].$$

Důkaz. Buďte $\{1, \alpha_1, \ldots, \alpha_{n-1}\}$, resp. $\{1, \beta_1, \ldots, \beta_{n-1}\}$ báze $\mathcal{O}_K$ a $\mathcal{O}$ jakožto $\mathbb{Z}$-modulů. Zobrazení $\xi : \mathcal{O}_K \longrightarrow \mathcal{O}$ dané $\alpha_i \longmapsto \beta_i$ pro každé i můžeme reprezentovat $n \times n$ maticí M nad celými čísly. Je známé (á la *Smithova normální forma*), že existují $n \times n$ celočíselné matice L, N takové, že LMN je diagonální matice, jejíž hlavní diagonála obsahuje celá k_i splňující $k_i \mid k_{i+1}$. Tato čísla jsou ne všechna nulová, protože index $\mathcal{O}$ v $\mathcal{O}_K$ je konečný. Násobení maticemi pouze mění bázi $\mathcal{O}$, tedy můžeme položit LMN matici udávající $\xi' : \mathcal{O}_K \longrightarrow \mathcal{O}$ a ta definuje k_i ze zadání. □

Jelikož víme, že $\mathcal{O}_K$ obsahuje bázi vektorového prostoru $K/\mathbb{Q}$, každý jeho pořádek ji obsahuje též. O co víc, předchozí věta aplikovaná na pořádky kvadratických těles tvrdí, že můžeme zvolit $\{1, d\}$ a $\{1, fd\}$ báze $\mathcal{O}_K$, resp. $\mathcal{O}$ s f vodičem $\mathcal{O}$ v $\mathcal{O}_K$, a proto symbolicky říci $\mathcal{O} = \mathbb{Z} + f\mathcal{O}_K$. Z toho navíc plyne, že platí inkluze pořádků $\mathcal{O} \subset \mathcal{O}'$ pouze a jenom když vodič $\mathcal{O}'$ dělí vodič $\mathcal{O}$.

Konečně, protože každý pořádek obsahuje racionální bázi pro K, můžeme si uvést ještě jednu ekvivalentní definici pořádku pomocí tenzorového součinu.

Důsledek 3.2.13. *Buď K číselné těleso. Pak podokruh $\mathcal{O} \subseteq K$ je pořádkem, právě pokud je volným $\mathbb{Z}$-modulem splňujícím $\mathbb{Q} \otimes_{\mathbb{Z}} \mathcal{O} \cong K$.*

Pozor, pořádky se od okruhu $\mathcal{O}_K$ obecně liší v několika podstatných oblastech, ke kterým se budeme vracet. Pro jedno, pořádky nejsou vždy celouzavřené nad jejich podílovým tělesem, jak je možné vidět u okruhu $\mathbb{Z}[\sqrt{5}] \subseteq \mathbb{Q}(\sqrt{5})$ a čísla $\frac{1+\sqrt{5}}{2}$ celého nad $\mathbb{Z}[\sqrt{5}]$, zatímco okruh $\mathcal{O}_K$ vždy je.

3.3 Norma, stopa a zkoumání dělitelnosti v okruzích

V této části užijeme pár základních poznatků ze studia lineární algebry ke studiu vlastnosti minimálních polynomů prvků číselného tělesa. Po čtenáři tedy požadujeme, aby se alespoň „stopově“ orientoval v této teorii, pro velmi podrobný úvod do této oblasti matematiky může posloužit [19].

Mějme K číselné těleso a L jeho konečné rozšíření s $[L : K] = n$. Zobrazení na L dané předpisem $a(x) : x \mapsto ax$, tedy násobení prvkem $a \in L$, definuje K-lineární endomorfismus

vektorového prostoru L nad K. Pokud si vybereme bázi $\{\alpha_1, \ldots, \alpha_n\}$ prostoru L nad K, zobrazení $a(x)$ působí na tuto bázi jako matice:

$$\begin{bmatrix} a(\alpha_1) \\ a(\alpha_2) \\ \vdots \\ a(\alpha_n) \end{bmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix},$$

kde $a_{ij} \in K$, a rozšiřuje se K-lineárně na celém L. Pokud vyjádříme $a = t_1\alpha_1 + \cdots + t_n\alpha_n$ jako lineární kombinaci prvků báze, můžeme díky vyjádření $a(\alpha_i) = \sum_j a_{ij}\alpha_j$ jednoznačně určit celou matici.

My se zaměříme na případ $K = \mathbb{Q}$ a L číselné těleso stupně n, který popsat je jednodušší. Víme, že L je jednoduché rozšíření $\mathbb{Q}(\theta)$ s bází $\{1, \theta, \ldots, \theta^{n-1}\}$, vzhledem ke které budeme psát naše mapy $a(x)$. Ukážeme, že toto zobrazení ve své podstatě souvisí s minimálním polynomem prvku a nad racionálními čísly.

Definice 3.3.1. Ať $K = \mathbb{Q}(\theta)$ je číselné těleso a τ je jeho prvek. Pak pod pojmem *charakteristický polynom* τ rozumíme charakteristickému polynomu transformace $\tau(x)$.

Nejprve se podívejme na zobrazení $\theta(x)$, kde θ má minimální polynom nad $\mathbb{Q}$ roven $x^n + b_{n-1}x^{n-1} + \cdots + b_0$. Máme dáno $\theta \cdot \theta^{i-1} = \sum_j a_{ij}\theta^{j-1}$, tedy protože prvky množiny $\{1, \theta, \ldots, \theta^{n-1}\}$ jsou lineárně nezávislé nad $\mathbb{Q}$, můžeme psát $\theta(x)$ jako akci matice M_θ udávající $\theta(x)$:

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -b_0 & -b_1 & \cdots & -b_{n-2} & -b_{n-1} \end{pmatrix}$$

na $\mathbb{Q}(\theta)$. Charakteristický polynom θ je charakteristcký polynom matice M_θ, který je daný $\det(xI - M_\theta)$, tedy:

$$\det \begin{pmatrix} x & -1 & 0 & \cdots & 0 \\ 0 & x & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & x & -1 \\ b_0 & b_1 & \cdots & b_{n-2} & x + b_{n-1} \end{pmatrix},$$

což je $b_0 + b_1x + \cdots + b_{n-1}x^{n-1} + x^n$, minimální polynom prvku θ.

Nyní již uvažme libovolné $\tau \in \mathbb{Q}(\theta)$. Připomeňme známý fakt ze studia tělesových rozšíření: $[\mathbb{Q}(\theta) : \mathbb{Q}(\tau)] \cdot [\mathbb{Q}(\tau) : \mathbb{Q}] = [\mathbb{Q}(\theta) : \mathbb{Q}]$, speciálně stupeň tělesa $\mathbb{Q}(\tau)$ dělí stupeň

$\mathbb{Q}(\theta)$, a totéž proto platí pro stupně minimálních polynomů příslušných τ a θ. Zobrazení $\tau(x)$ působí na $\mathbb{Q}(\theta)$ jako blokově diagonální matice obsahující n/k matic:

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -c_0 & -c_1 & \cdots & -c_{k-2} & -c_{k-1}, \end{pmatrix},$$

kde $k \mid n$ a $x^k + c_{k-1}x^{k-1} + \cdots + c_0$ je minimální polynom τ nad racionálními čísly. Charakteristický polynom τ je pak jeho minimální polynom umocněn na n/k-tou mocninu, což je v souladu s větou Cayley-Hamiltona zaobírající se charakteristickými polynomy matic.

Kvůli této korespondenci zobrazení $\tau(x)$ a minimálního polynomu τ definujeme pojmy stopa a norma, které nám pomohou s prací v okruzích, například při zkoumání dělitelnosti.

Definice 3.3.2. Buď K číselné těleso a τ jeho prvek. Pak jeho definujeme *stopu* $Tr(\tau)$ a *normu* $N(\tau)$ jako stopu, resp. determinant matice udávající $\tau(x)$:

$$\begin{aligned} Tr_K(\tau) &:= TrM_\tau, \\ N_K(\tau) &:= \det M_\tau. \end{aligned}$$

Norma i stopa prvků číselného tělesa jsou tak racionální čísla. Abychom se neudávili notací, pokud bude jasné těleso nad kterým pracujeme, budeme psát jednoduše $Tr(\tau), N(\tau)$.

Pojďme se si spočíst normu a stopu pár prvků v číselným tělesech, abychom získáli intuici, s čím to pracujeme.

V tělese $\mathbb{Q}(\sqrt{-2})$ mějme číslo $a + b\sqrt{2}$. Báze tohoto tělesa jakožto vektorového prostoru nad $\mathbb{Q}$ je $\{1, \sqrt{-2}\}$, pojďme spočíst akci $(a+b\sqrt{-2})(x)$ na tomto tělese, k čemuž nám stačí určit akci na bázi:

$$\begin{aligned} (a + b\sqrt{-2}) \cdot 1 &= a + b\sqrt{-2}, \\ (a + b\sqrt{-2}) \cdot \sqrt{-2} &= -2b + a\sqrt{-2}, \end{aligned}$$

tedy $(1 + 2\sqrt{2})(x)$ působí na $\mathbb{Q}(\sqrt{-2})$ jako matice:

$$\begin{pmatrix} a & b \\ -2b & a \end{pmatrix}.$$

Její stopa je $2a$ a determinant $a^2 + 2b^2$, což souhlasí s tím, že minimální polynom $a + b\sqrt{-2}$ je pro $b \neq 0$ roven $x^2 - 2ax + a^2 + 2b^2$ a pro $b = 0$ jednoduše $x - a$.

Uvažme dále těleso $\mathbb{Q}(\theta)$, kde θ je kořenem polynom x^3-x+3, který je zjevně iracionální. Libovolný jeho prvek τ vyjádřený podle báze $\{1,\theta,\theta^2\}$ jako $a+b\theta+c\theta^2$ působí na bázi jako:

$$\begin{aligned}(a+b\theta+c\theta^2)\cdot 1 &= a+b\theta+c\theta^2,\\ (a+b\theta+c\theta^2)\cdot \theta &= -3c+(a+c)\theta+b\theta^2,\\ (a+b\theta+c\theta^2)\cdot \theta &= -3b+(b-3c)\theta+(a+c)\theta^2,\end{aligned}$$

tedy udává matici:

$$\begin{pmatrix} a & b & c\\ -3c & a+c & b\\ -3b & b-3c & a+c\end{pmatrix}$$

se stopou $3a+2c$ a determinantem $a^3-3b^3+2a^2c+3bc^2+9c^3-ab^2-9abc-ac^2$. Buď je τ racionální číslo, či je jeho minimální polynom roven třemi. V prvním případě je jeho stopa $3a$ a norma a^3, v druhém případě stopa a a norma rovna determinantu M_τ.

Když máme dobrou představu o norma a stopě, pojďme se o těchto funkcích ukázat několik málo důležitých faktů. K tomu nám pomohou klasické výsledky ohledně stop a determinantů matic.

Věta 3.3.3. *Norma je mutiplikativní a stopa je $\mathbb{Q}(\theta)$-lineární funkce.*

Důkaz. Důkaz plyne z faktů, že $\det(A\cdot B)=\det(A)\cdot\det(B)$ a $Tr(kA+\ell B)=Tr(kA)+Tr(\ell B)=kTr(A)+\ell Tr(B)$ pro libovolné čtvercové matice A,B a $k,\ell\in\mathbb{Q}(\theta)$. □

Normu a stopu τ můžeme díky vlastnostem mapy $\tau(x)$ pevněji ukotvit k minimálnímu polynomu τ:

Věta 3.3.4. *Buď K číselné těleso stupně n a τ jeho prvek s minimálním polynomem $x^k+c_{k-1}x^{k-1}+\cdots+c_0$ nad $\mathbb{Q}$. Pak:*

$$\begin{aligned}Tr(\tau) &= -n/k\cdot c_{k-1},\\ N(\tau) &= (-1)^n {c_0}^{n/k}.\end{aligned}$$

Ekvivalentně věta říká, že pokud $\tau,\tau_2,\ldots,\tau_n$ jsou kořeny charakteristického polynomu τ, včetně multiplicity, platí $Tr_K(\tau)=\tau+\tau_2+\cdots+\tau_n$ a $N(\tau)=\tau\cdot\tau_2\cdots\tau_k$. Pokud je tedy τ celé algebraické číslo, jeho norma i stopa jsou celá čísla.

Důkaz. Tvar stopy plyne ihned z faktu, že stopa matice je součtem prvků po hlavní diagonále. Determinant blokové matice je součin determinantů bloků na diagonále, tedy n/k-tá mocnina determinantu matice:

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0 & 1\\ -c_0 & -c_1 & \cdots & -c_{k-2} & -c_{k-1},\end{pmatrix}.$$

což je $(-1)^k c_0$. □

Minimální polynomy prvků α a β nám toho říkají pouze pramálo o minimálních polynomech čísel $\alpha + \beta$ či $\alpha \cdot \beta$, nicméně za pomocí spojení minimálních polynomů s normami a stopami můžeme za pomocí vět výše přesně popsat některé jejich koeficienty.

Protože je norma multiplikativní a na celých algebraických číslech celočíselná, můžeme ji propojit s dělitelností v okruzích. Pokud $b = ac$ pro $a, b, c \in R$ nenulová, máme $N(b) = N(ac) = N(a)N(c)$.

Věta 3.3.5. *Mějme $a, b \in R \subseteq \mathcal{O}_K$ nenulová pro K číselné těleso. Pokud a dělí b, ve smyslu $b = a \cdot c$ pro $c \in R$, tak platí:*

$$N(a) \mid N(b).$$

Pokud a je v okruhu R invertibilní, tedy $a \cdot b = 1$ pro nějaké $b \in R$, nutně platí $N(a)N(b) = N(ab) = N(1) = 1$ a díky celočíselnosti norem je $N(a)$ rovno ± 1.

Definice 3.3.6. Prvek $a \in R$, který je v R invertibilní, nazveme *jednotkou*.

Definice 3.3.7. Pokud je podílem dvou prvků $a, b \in R$ jednotka, nazveme je *asociované*.

Jednotky v okruzích tvoří multiplikativní grupu, přičemž v okruhu celých algebraických čísel kvadratických tělesech jsou určena řešeními kvadratických forem. V okruhu celých čísel tělesa $\mathbb{Q}$ jsou jednotky zjevně pouze ± 1, Gaussova celá čísla připouští multiplikativní inverz prvků ± 1 i $\pm i$. Případ reálných kvadratických těles $\mathbb{Q}(\sqrt{d})$, tedy $0 < d \not\equiv 1 \pmod 4$, je obzvlášťe zajímavý, jednotky $a + b\sqrt{-d} \in \mathbb{Z}[\sqrt{-d}]$ totiž splňují:

$$a^2 - db^2 = \pm 1,$$

tedy rozšířenou Pellovu rovnici.

Studium Pellových rovnic poté poukazuje na fakt, že tato grupa je vesměs cyklická, tedy že všechny jednotky vygenerujeme jako $\pm\omega^n$ pro $n \in \mathbb{Z}$ a ω tzv. *fundamentální jednotku* tohoto okruhu.

Od jednotek se přesuňme na zobecnění prvočísel, tzv. *ireducibilních* prvků, v okruzích.

Definice 3.3.8. Prvek $a \in R$ nazveme *ireducibilním*, nelze-li jej zapsat jako součin dvou prvků R, obou ne jednotek.

Multiplikativita normy tvrdí, že prvky s prvočíselnou normou jsou nad R ireducibilní. Zajímalo by nás tedy, zda dokážeme s ireducibilními prvky operovat podobně jako s prvočísly, tedy rozkládali čísla na ireducibilní prvky. Takový rozklad v obecném okruhu zjevně existuje, bohužel však ne vždy je jednoznačně určený. Koncept dělení v okruzích přivádí na

mysl dělení se zbytkem.

Ze školních lavic víme, že v dokážeme v celých číslech dělit se zbytkem. Tuto vlastnost ale sdílí některé další okruhy, neprominentněji $\mathbb{Z}[i]$. Ukážeme si tedy, jak na to.

Vskutku, ukážeme, že pro libovolná nenulová $a, b \in \mathbb{Z}[i]$ můžeme zvolit Gaussova celá čísla q, r taková, že $a = bq + r$ a $N(r) < N(b)$, kde normu bereme normu komplexního čísla. Norma je multiplikativní, tedy ekvivalentně píšeme $N\left(\frac{r}{b}\right) < 1$ a $\frac{a}{b} = q + \frac{r}{b}$. Existence takových r a b je nicméně zřejmá, pokud se na problém podíváme geometricky. Rovnice $N(z) \leqslant 1$ definuje v komplexní rovině jednotkový kruh se středem v počátku, tedy požadujeme, aby šlo zvolit $q \in \mathbb{Z}[i]$, které je na méně než jednotkovou vzdálenost od libovolného komplexního čísla z. To jistě dokážeme, protože Gaussova celá čísla tvoří v komplexní rovině jednotkovou mřížku.

- obrázek -

Díky tomuto poznatku můžeme v $\mathbb{Z}[i]$ dělit se zbytkem, tudíž existuje pro libovolná $a, b \in \mathbb{Z}[i]$ (až na násobení jednotkou) jednoznačný nejvyšší společný dělitel, a výše uvedená vlastnost efektivně dává vzniku Euklidovu algoritmu v $\mathbb{Z}[i]$. Bezoutova identita proto platí pro dva členy a tedy i pro libovolný počet Gaussových celých čísel.

Definice 3.3.9. Okruhy, ve kterých můžeme obdobně dělit se zbytkem, nazveme *euklidovy*.

Poznámka. Pouze konečně mnoho okruhů celých algebraických čísel imaginárních kvadratických těles $\mathbb{Q}(\sqrt{d})$ pro $d < 0$ je euklidových, vyhovující d se nazývají *Heegnerova*. Z nich v absolutní hodnotě nejvyšší je -163.

V oborech, ve kterých umíme dělit se zbytkem, díky platnosti Bezoutovy rovnosti vykazují ireducibilní prvky podobné vlastnosti jako prvočísla v celých číslech.

Věta 3.3.10. *Buď R euklidův okruh a $p \in R$ ireducibilní. Pokud pro $a, b \in R$ platí $p \mid ab$, pak buď $p \mid a$, či $p \mid b$.*

Důkaz. Nechť naopak platí $p \mid ab$ a p nedělí ani jeden z činitelů. Existuje (až na násobení jednotkou) jednoznačný společný dělitel d prvků p a a, který díky ireducibilitě p je buď s p asociovaný, či je jednotka. Pokud by nastal první případ, pak $p \mid a$, spor. Je tak d jednotkou, vhodným přenásobením a jednotkou uvažujme $d = 1$. Z Bezoutovy rovnost existují $x, y \in R$ splňující $xa+yp = 1$. Analogicky dojdeme k existenci $z, t \in R$ s $zb+tp = 1$. Vynásobením těchto rovnosti získáme:

$$1 = (xa + yp)(zb + tp) = xyab + p(xta + yzb + ytp),$$

díky předpokladu úlohy p dělí pravou stranu a tedy i levou, což je hledaný spor. □

S přechozí větou na mysli se není příliš obtížné dovtípit, že euklidovy okruhy připouští jednoznačný rozklad, protože se ireducibilní prvky opravdu chovají podobně jako prvočísla.

Důsledek 3.3.11. *Buď R euklidův okruh. Pak se každý jeho prvek jednoznačně (až na pořadí prvků a násobení jednotkou) rozkládá na součin ireducibilních prvků a jednotky.*

Důkaz. Pokud prvek $n \in R$ je reducibilní jako $n = ab$ kde a, b nejsou jednotky, platí $1 < |N(a)| < |N(a)N(b)| = |N(n)|$ a obdobně pro b, tedy sestupem dojdeme k tomu, že každý prvek lze rozložit na (nějaký) součin ireducibilních prvků. Dejme tomu, že existují dvě posloupnosti $p_1, \ldots, p_k$ a $q_1, \ldots, q_\ell$ ireducibilních prvků takových, že platí:

$$u_1 p_1 \cdots p_k = n = u_2 q_1 \cdots q_\ell$$

pro nějaké jednotky $u_i \in R$. Platí, že p_1 dělí druhý rozklad, tedy dělí jedno z q_i, bez újmy na obecnosti ať to je q_1, tedy $q_1 = v_1 p_1$. Tento proces opakujeme s číslem $\frac{n}{p_1} = \frac{u_1}{v_1} p_2 \cdots p_k = u_2 q_2 \cdots q_\ell$, čímž docházíme k tomu, že množiny $\{p_1, \ldots, p_k\}$ a $\{q_1, \ldots, q_\ell\}$ jsou až na asociaci shodné, včetně násobnosti, což jsme chtěli. □

V obecném okruhu, dokonce ani $\mathcal{O}_K$, jednoznačnost rozkladu však neplatí. Klasický protipříklad dává okruh $\mathbb{Z}[\sqrt{-5}]$ a dva rozklady čísla $6 = 2 \cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5})$. Ukážeme, že všichni čtyři činitelé jsou v $\mathbb{Z}[\sqrt{-5}]$ ireducibilní.

Norma obecného prvku $a + b\sqrt{-5}$ našeho okruhu je daná $a^2 + 5b^2$, tedy normy našich dělitelů jsou po řadě rovny $4, 9, 6$ a 6. Pokud by nějaký z nich šel rozložit jako součin dvou ireducibilních prvků, s ohledem na multiplikativitu a nezápornost normy v $\mathbb{Z}[\sqrt{-5}]$ by oba měly normu buď 2 či 3. Nicméně 2 a 3 nejsou kvadratické zbytky modulo 5, tedy rovnost $2, 3 = N(a + b\sqrt{-5}) = a^2 + 5b^2$ nemá řešení, taková čísla proto neexistují a všichni čtyři dělitelé jsou ireducibilní. Navíc multiplikativita normy zjevně nepovoluje, aby se libovolný pár dělitelů dělil, takže tyto rozklady jsou různé. V příštích sekcích se k jednoznačnosti rozkladu ještě vrátíme, nicméně prozatím mějme na paměti, že ne vždy nutně platí.

Některé prvky číselných okruhu můžeme tedy vyjádřit jako součin ireducibilních prvků vícero různými způsoby, přinejmenším bychom alespoň očekávali, že počet ireducibilních faktorů je vždy konzistentní. Opět bychom se však mýlili, kvadratický okruh $\mathbb{Z}[\sqrt{-14}]$ poskytuje následující dva rozklady čísla 81:

$$3 \cdot 3 \cdot 3 \cdot 3 = 81 = (5 + 2\sqrt{-14})(5 - 2\sqrt{-14}),$$

rozborem norem a dělitelností opět můžeme dospět k závěru, že všichni tři přítomní dělitelé jsou ireducibilní a dělitelé z jednotlivých rozkladů nejsou asociovaní.

Pojďme se ještě na chvíli pozastavit u okruhu $\mathbb{Z}[i]$, ve kterém jednoznačnost rozkladu platí, a ukázat jednu roztomilou aplikaci předchozí věty. Norma Gaussova celého čísla je $N(a + bi) = a^2 + b^2$, tedy druhá mocnina klasické komplexní absolutní hodnoty. Její vlastnosti pomohou odhalit, přesně která přirozená čísla jsme schopni vyjádřit jako součet dvou čtverců.

???????chci to tu?????????????????

Věta 3.3.12. *Přirozené číslo n lze vyjádřit jako součet dvou čtverců, právě pokud n není dělitelné prvočíslem $p \equiv -1 \pmod 4$ v liché mocnině.*

Důkaz. Odůvodníme, proč ireducibilní prvky v okruhu $\mathbb{Z}[i]$ jsou prvočísla $p \equiv -1 \pmod 4$, prvky normy rovné prvočíslu $p \equiv 1 \pmod 4$ a $\pm 1 \pm i$. Dejme tomu, že jsme schopni zapsat $p \equiv -1 \pmod 4$ jako součin dvou prvků, obou ne jednotek. Rovnost $p = ab$ v $\mathbb{Z}[i]$ díky multiplikativitě normy znamená $p^2 = N(p) = N(ab) = N(a)N(b)$. Protože norma komplexního čísla je nezáporná a a, b nejsou jednotky, platí $N(a) = N(b) = p$, tedy pro $a = x + yi$ platí $x^2 + y^2 = p$. To ale porušuje pravidlo, že čtverce dávají pouze zbytky 0 a 1 modulo čtyřmi. Tato p jsou proto ireducibilní. Dále Gaussova celá čísla s normou 2 jsou jistě pouze $\pm 1 \pm i$, které jsou již ireducibilní.

Pokud je naopak $p \equiv 1 \pmod 4$ prvočíslo, je -1 kvadratický zbytek modulo p. Pro nějaké x proto platí $p \mid x^2 + 1$, což můžeme v rámci $\mathbb{Z}[i]$ zapsat jako $p \mid (x + i)(x - i)$. Pokud by p nešlo rozložit, muselo by dělit právě jednu ze závorek a tak $p \mid i$, což je nemožné. Existuje proto netriviální rozklad $ab = p$ s normou $N(a)N(b) = N(ab) = N(p) = p^2$, tedy $N(a) = N(b) = p$ pro nějaká Gaussova celá a, b. Pro $a = x+yi$ pak platí $p = N(a) = x^2+y^2$. Všechna ostatní Gaussova celá čísla jsou jistě reducibilní.

Jestliže n je dělitelné čtyřmi či prvočíslem $p \equiv -1 \pmod 4$ v liché mocnině, nelze zjevně zapsat jako součet dvou čtverců, protože kvadratické zbytky modulo 4 jsou pouze 0 a 1 a $p \mid a^2 + b^2$ znamená, že buď -1 je kvadratický zbytek modulo p, neboli $p \equiv 1 \pmod 4$, či $p \mid a$ a $p \mid b$. Naopak pokud nenastává ani jeden z těchto případů, můžeme každé prvočíslo $p \equiv 1 \pmod 4$ zapsat jako součet dvou čtverců, tedy díky rovnostem $(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2$ a $q^2a^2 = (qa)^2$ lze n vyjádřit jako součet dvou čtverců. □

Obdobné charakterizace můžeme provést rozkladem v ostatních euklidovských kvadratických okruzích, což staví základy charakterizace vyjádřování celých čísel kvadratickými formami. Podrobněji je toto téma studováno v [26], či v předloze oné práce [8], na které je též založena notná část této kapitoly.

Chtěli bychom tedy hledat strukturu, která poslouží tam, kde nás prvky $\mathcal{O}_K$ selhaly, u jednoznačného rozkladu na ireducibilní prvky. Eduard Kummer v 19.století tento problém vyřešil vložením množiny algebraických celých čísel tělesa K do množiny tzv. *ideálních čísel*, která se jednoznačně rozkládají na součin *ideálních prvočísel*. Tento koncept Richard Dedekind, další z titánů teorie čísel, později nazval *ideály*.

3.4 Ideály

Definice 3.4.1. Neprázdnou aditivní podgrupu $\mathfrak{a}$ okruhu R takovou, že $a \cdot r \in \mathfrak{a}$, resp. $r \cdot a \in \mathfrak{a}$ pro $a \in \mathfrak{a}, r \in R$ označíme jako *pravý*, resp. *levý ideál*. Ideál, který je pravý i levý, nazveme *oboustranným*.

O (levém) ideálu můžeme proto přemýšlet jako o (levém) R-modulu. V případě, že pracujeme nad komutativním okruhem R, pravé a levé ideály nerozlišujeme. Oboustranné ideály R budeme nazývat jednoduše ideály.

Pokud $\mathfrak{a}$ je podgrupa R, tak faktorová grupa $R/\mathfrak{a}$ se stane okruhem, právě pokud $\mathfrak{a}$ je ideálem. Ideály tedy konstruujeme v podobném duchu jako normální podgrupy, kde podgrupa H grupy G je normální, právě když zobrazení $G \longrightarrow G/H$ přiřazující každému prvku G jeho příslušnou třídu v G/H je homomorfismus grup.

Každý ideál $\mathfrak{a} \subseteq R$ tedy definuje podílový okruh $R/\mathfrak{a}$, kde projekce $R \longrightarrow R/\mathfrak{a}$ redukující každé $r \in R$ na jeho příslušnou třídu v $R/\mathfrak{a}$ dává kanonický homomorfismus mezi těmito dvěma okruhy. Navíc homomorfismus jemu inverzní udává bijektivní zobrazení mezi třídami $R/\mathfrak{a}$ a ideály R obsahující $\mathfrak{a}$.

Definice 3.4.2. Pokud $\theta_1, \ldots, \theta_n \in R$ je konečná množina generátorů ideálu (ve smyslu R-modulu) $\mathfrak{a}$, značíme:

$$\mathfrak{a} = (\theta_1, \ldots, \theta_n).$$

Ne každý ideál libovolného okruhu je konečně generovaný, například ideál $(x_1, x_2, \ldots)$ v okruhu $\mathbb{R}[x_1, x_2, \ldots]$ s nekonečně mnoha proměnnými jistě konečně generovaný není, my si však dále odůvodníme, proč v pořádcích tomu tak je.

Ideál generovaný prvkem x^2+1 v $\mathbb{Z}[x]$ má příslušný okruh zbytků $\mathbb{Z}[x]/(x^2+1)$ isomorfní okruhu $\mathbb{Z}[i]$, není tedy konečný a jeho norma není definovaná, jako není v mnoha dalších „divokých“ okruzích. V pořádku $\mathcal{O}$ přesto každý ideál konečný okruh zbytků má.

Věta 3.4.3. *Buď $\mathcal{O}$ pořádek číselného tělesa K stupně n a $\mathfrak{a} \subseteq \mathcal{O}$ ideál. Pak $\mathfrak{a}$ má v $\mathcal{O}$ konečný index.*

Důkaz. Uvažme prvek $x \in \mathfrak{a}$. Pokud $x_2, \ldots, x_k$ jsou kořeny minimálního polynomu x, platí $N(x) = x \cdot x_2 \cdots x_k \in \mathfrak{a}$ je díky větě 3.2.10 celé číslo. Pak $\mathcal{O}/\mathfrak{a}$ je podokruhem $\mathcal{O}/(N(x))$. Konečně, protože $\mathcal{O}$ je konečně generovaný $\mathbb{Z}$-modul, faktorokruh $\mathcal{O}/(N(x))$ je konečný. □

Norma čísla v $\mathbb{Z}[i]$ nám dává představu o jeho vzdálenosti od počátku souřadné soustavy, normu ideálu proto definujeme s podobným účelem.

Definice 3.4.4. Buď $\mathcal{O}$ pořádek číselného tělesa K a $\mathcal{I} \subseteq \mathcal{O}$ ideál. Pod normou $N_{\mathcal{O}}(\mathfrak{a})$ ideálu $\mathfrak{a}$ rozumíme počtu prvků faktorokruhu $\mathcal{O}/\mathfrak{a}$.

Věta 3.4.5. *Každý nekonečný stoupající řetězec inkluzí ideálů pořádku $\mathcal{O}$ je eventuálně konstantní.*

Důkaz. Pokud pro ideály $\mathfrak{b}, \mathfrak{c}$ platí $\mathfrak{b} \subset \mathfrak{c}$, tak jistě i $\mathcal{O}/\mathfrak{c} \subset \mathcal{O}/\mathfrak{b}$, tedy $N(\mathfrak{b}) > N(\mathfrak{c})$. Nekonečný řetězec (ostrých) inkluzí ideálů by znamenal posloupnost norem těchto ideálů

klesající pod všechny meze, speciálně by existoval ideál se zápornou normou, zjevný spor. □

Se znalostí předchozí věty můžeme pak definitivně obhájit definici ideálů $\mathcal{O}$ jako konečně generovaných:

Věta 3.4.6. *Buď $\mathfrak{a} \subseteq \mathcal{O}$ ideál. Pak je konečně generovaný jako $\mathcal{O}$-modul.*

Důkaz. Ideál obsahující pouze 0 je konečně generovaný, dále uvažme nenulový prvek $a_1 \in \mathfrak{a}$. Pokud $\mathfrak{a}$ není generovaný a_1, obsahuje prvek a_2 takový, že $(a_1) \subset (a_1, a_2)$. Pokud $\mathfrak{a}$ není generovaný těmito dvěma prvky, existuje $a_3 \in \mathfrak{a}$ takový, že $(a_1, a_2) \subset (a_1, a_2, a_3)$. V případě, že bychom takovéto prvky mohli hledat do neurčita, získali bychom nekonečný ostře rostoucí řetězec ideálů $(a_1) \subset (a_1, a_2) \subset (a_1, a_2, a_3) \subset \cdots$, spor s předchozí větou. Řetězec se proto musí na nějakém místě rozlomit a zůstane nám konečná množina generátorů. □

V pořádků číselného tělesa K stupně n platí $\mathcal{O} \cong \mathbb{Z}^n$, tedy fundamentální věta konečně generovaných abelovských grup tvrdí, že konečná podgrupa $\mathcal{O}_K$ je buď nulová, či isomorfní direktnímu součinu několika, nejvýše však n, kopií $\mathbb{Z}$. Speciálně každý ideál má nejvýše n generátorů. V další sekci si počet generátorů omezíme dokonce číslem 2.

V euklidově okruhu existuje jednoznačně (až na násobení jednotkou) určený největší společný dělitel čísel θ_i, nějaké d. Jistě pak libovolný prvek $(\theta_1, \ldots, \theta_n)$ náleží do (d). Navíc dle Bezoutovy identity platí opačná inkluze, tedy $(\theta_1, \ldots, \theta_n)$ je ideál generovaný největším společným dělitelem čísel θ_i.

Definice 3.4.7. Ideály generované jediným prvkem označíme jako *hlavní*.

Zajímavé propojení s námi již známou normou prvků $\mathcal{O} \subseteq K$ lze pozorovat právě u ideálů hlavních. Norma hlavního ideálu (α) je dána $[\mathcal{O} : \alpha\mathcal{O}]$, je tedy rovna stupni zobrazení $\alpha(x)$ na K, což je definice čísla $N_K(\alpha)$. Navíc norma α patří do ideálu (α), protože je součinem jeho sdružených čísel. Každý hlavní ideál pořádku obsahuje svou normu, tedy i každý jiný obsahuje celé číslo, konkrétně normu libovolného jeho generátoru.

Pojďme si dále definovat na ideálech pár základních operací.

Definice 3.4.8. Buďte $\mathfrak{a}, \mathfrak{b}$ ideály okruhu R. Pak jejich součet a součin definujeme následovně:

- $\mathfrak{a} + \mathfrak{b} = \{a + b | a \in \mathfrak{a}, b \in \mathfrak{b}\}$,
- $\mathfrak{a} \cdot \mathfrak{b} = \{\sum_{i=1}^{n} a_i b_i |\, a_i \in \mathfrak{a}, b_i \in \mathfrak{b}, n \in \mathbb{N}\}$.

Vidíme, že jak součet, tak součin dvou ideálů je též ideálem, první generovaný sjednocením množin generátorů obou ideálů, druhý součiny po jednom generátoru $\mathcal{I}$ a druhém generátoru $\mathcal{J}$. Sčítání je jistě asociativní a jeho neutrální prvek je nulový ideál $(0) = \{0\}$. Násobení ideálů je taktéž asociativní, neboť:

$$(\mathfrak{a} \cdot \mathfrak{b}) \cdot \mathfrak{c} = \left\{ \sum_{i=1}^{n} a_i b_i c_i \,\middle|\, a_i \in \mathfrak{a}, b_i \in \mathfrak{b}, c_i \in \mathfrak{c}, n \in \mathbb{N} \right\} = \mathfrak{a} \cdot (\mathfrak{b} \cdot \mathfrak{c}),$$

a neutrální prvek je vždy celý okruh R. Ideály okruhu R proto tvoří se sčítáním a násobením monoid, při komutativitě R je monoid komutativní též.

Prostřednictvím násobení si můžeme definovat dělitelnost ideálů:

Definice 3.4.9. Buďte $\mathfrak{a}, \mathfrak{b}$ ideály okruhu R. Pokud pro nějaký ideál $\mathfrak{c} \subseteq R$ platí $\mathfrak{b} = \mathfrak{a} \cdot \mathfrak{c}$, píšeme $\mathfrak{a} \mid \mathfrak{b}$ a říkáme, že $\mathfrak{a}$ *dělí* $\mathfrak{b}$.

Definice 3.4.10. Buďte $\mathfrak{a}, \mathfrak{b}$ ideály okruhu R. Tyto ideály nazveme *nesoudělné*, pokud platí rovnost ideálů:

$$\mathfrak{a} + \mathfrak{b} = (1).$$

Dva ideály jsou tedy nesoudělné, právě pokud součin nějakých dvou jejich prvků obsahuje jednotku R. Nedefinujeme největší společný dělitel, neboť ten ne vždy existuje, alespoň ne v obecném okruhu R. V následující podkapitole dokážeme slíbené tvrzení, že ideály $\mathcal{O}_K$ se rozkládají jednoznačně na součin prvoideálů, a odůvodníme, proč toto tvrzení nedosahuje na zbylé pořádky tělesa K.

3.5 Rozklad na prvoideály

V celých číslech jsou krom samotného okruhu $\mathbb{Z}$ ideály generované prvočísly (p) jediné, které pro libovolná $a, b \in R$ splňující $ab \in (p)$ vynucují alespoň jedno z a či b náležící do (p). Tento koncept si zobecníme do obecných okruhů.

Definice 3.5.1. Neprázdný nenulový ideál $\mathfrak{p} \subset R$ takový, že pro každá $a, b \in R$ splňující $ab \in \mathfrak{p}$ platí buď $a \in \mathfrak{p}$, či $b \in \mathfrak{p}$, nazveme *prvoideálem*.

Prvoideály v pořádcích můžeme ve zkratce charakterizovat v následující větě:

Věta 3.5.2. *Buď $\mathfrak{p} \subseteq \mathcal{O}$ neprázdný nenulový ideál. Pak následující skutečnosti jsou ekvivalentní:*

(i) $\mathfrak{p}$ *je prvoideál,*

(ii) Faktorový okruh $\mathcal{O}/\mathfrak{p}$ *je konečné těleso,*

(iii) $\mathfrak{p}$ *je maximální, neboli neexistuje ideál* $\mathfrak{a}$ *splňující* $\mathfrak{p} \subset \mathfrak{a} \subset \mathcal{O}$*,*

(iv) Rovnost $\mathfrak{p} = \mathfrak{ab}$ *znamená buď* $\mathfrak{a} = \mathfrak{p}$*, či* $\mathfrak{b} = \mathfrak{p}$*.*

Důkaz. Případ, kdy v okruhu zbytků $\mathcal{O}/\mathfrak{p}$ rovnost tříd $(a + \mathfrak{p})(b + \mathfrak{p}) = \mathfrak{p}$ platí jenom pokud jedno z a, b náleží do $\mathfrak{p}$, nastane právě když $\mathfrak{p}$ je prvoideál. Faktorokruh $\mathcal{O}/\mathfrak{p}$ je proto oborem integrity pouze a jenom když $\mathfrak{p}$ je prvoideál. Klasický výsledek abstraktní algebry ale tvrdí, že konečný obor integrity je těleso, což stvrzuje ekvivalenci bodů (i) a (ii).

Dále mějme $\mathfrak{p}$ prvoideál a $\mathfrak{a}$ ideál $\mathcal{O}$ splňující $\mathfrak{p} \subset \mathfrak{a}$. Ukážeme, že $\mathfrak{a}$ je roven samotnému $\mathcal{O}$. Buď $a \in \mathfrak{a} \setminus \mathfrak{p}$, pak a leží v nenulové třídě $\mathcal{O}/\mathfrak{p}$. Tento prvek má v $\mathcal{O}/\mathfrak{p}$ pak multiplikativní inverz b, tedy $ab = 1 + c$ pro nějaké $c \in \mathfrak{p} \subset \mathfrak{a}$, což znamená $1 = ab - c$. Všechna tři čísla a, b, c leží v $\mathfrak{a}$, tedy $1 \in \mathfrak{a}$ a $\mathfrak{a} = \mathcal{O}$. Naopak pokud $\mathfrak{p}$ je maximální, uvažme libovolné $a \in \mathcal{O} \setminus \mathfrak{p}$. Nenulová třída $a + \mathfrak{p} \in \mathcal{O}/\mathfrak{p}$ dává vzniku ideálu $(a) + \mathfrak{p} \subseteq \mathcal{O}$, který obsahuje jak a, tak ideál $\mathfrak{p}$, tedy díky maximalitě $\mathfrak{p}$ i okruh $\mathcal{O}$ samotný. Jednotka náleží do $(a) + \mathfrak{p}$, platí tedy $ra + p = 1$ pro nějaká $r \in \mathcal{O}, p \in \mathfrak{p}$. Platí pak rovnost $(r + \mathfrak{p})(a + \mathfrak{p}) = ra + \mathfrak{p} = 1 + \mathfrak{p}$, čili každá nenulová třída $a + \mathfrak{p}$ má v $\mathcal{O}/\mathfrak{p}$ multiplikativní inverz.

Konečně ať $\mathfrak{p}$ je roven součinu dvou ideálů $\mathfrak{a}$ a $\mathfrak{b}$, speciálně jej oba obsahují. Pokud je $\mathfrak{p}$ prvoideálem, tak je maximální, tedy je jeden z $\mathfrak{a}$ roven $\mathfrak{p}$ a ten druhý okruhu $\mathcal{O}$. Naopak pokud platí bod (iv), tak $\mathcal{O}/\mathfrak{p}$ je oborem integrity, tedy konečným tělesem. □

Tyto výsledky nejsou exlusivní pro pořádky číselných těles, mimo ně však musíme být na pozoru, podmínka (ii) totiž není splněna například pro prvoideál (x) v okruhu $\mathbb{Z}[x]$.

Důsledek předchozí věty, Bezoutovy věty a faktu, že každý ideál obsahuje svoji normu, mluví o normě prvoideálů:

Důsledek 3.5.3. *Pokud ideál $\mathfrak{p} \subset \mathcal{O}$ je prvoideál, pak obsahuje unikátní prvočíslo, jehož některá mocnina je norma $\mathfrak{p}$.*

Nyní jsme konečně připraveni diskutovat jednoznačnost rozklad na prvoideálů.

Vzpomeňme si na náš postup, když jsme dokazovali jednoznačnost rozkladu na ireducibilní prvky v euklidovských doménách. Ten se skládal ze tří kroků, i) ireducibilní prvek dělící součin dvou prvků dělí jeden z nich, ii) každý prvek je součinem několika ireducibilních prvků a iii) rozklad na ireducibilní prvky je (až na násobení jednotkou) jednoznačný.

Tuto proceduru se pokusíme zopakovat a poté odůvodníme, proč v pořádcích zcela zreplikovat nelze, hlavní problém bude činit bod ii). První část přichází bezbolestně:

Věta 3.5.4. *Buďte $\mathfrak{p}, \mathfrak{a}, \mathfrak{b} \subseteq \mathcal{O}$ nenulové ideály. Pak $\mathfrak{p}$ je prvoideál, právě pokud inkluze $\mathfrak{p} \supseteq \mathfrak{ab}$ znamená $\mathfrak{p} \supseteq \mathfrak{a}$, či $\mathfrak{p} \supseteq \mathfrak{b}$.*

Důkaz. Nejprve uvažme $\mathfrak{p}$ prvoideál. Pokud platí $p \supseteq \mathfrak{ab}$ a $\mathfrak{p} \not\supseteq \mathfrak{a}$, uvažme číslo $x \in \mathfrak{a} \setminus \mathfrak{p}$. Pro každé $y \in \mathfrak{b}$ je $xy \in \mathfrak{ab} \subseteq \mathfrak{p}$, tedy $y \in \mathfrak{p}$, neboli platí $\mathfrak{p} \supseteq \mathfrak{b}$. Nyní mějme implikaci ze zadání platnou. Pro libovolná $x, y \in \mathfrak{p}$ platí $(x)(y) = (xy) \subseteq \mathfrak{p}$, tedy jeden ze dvou ideálů generovaných x, y náleží do $\mathfrak{p}$. Jedno z těchto čísel proto leží uvnitř $\mathfrak{p}$ a $\mathfrak{p}$ je prvoideál. □

Pokračujme s naším seznamem, tentokrát ukážeme, že každý ideál $\mathcal{O}$ obsahuje součin prvoideálů.

Věta 3.5.5. *Každý nenulový ideál $\mathfrak{a} \subseteq \mathcal{O}$ splňuje:*

$$\mathfrak{a} \supseteq \mathfrak{p}_1 \mathfrak{p}_2 \cdots \mathfrak{p}_r.$$

pro nějaké prvoideály $\mathfrak{p}_i$.

Důkaz. Dejme tomu, že existují ideály, které toto tvrzení nesplňují, a uvažme mezi nimi exemplář $\mathfrak{a}$ s nejnižší normou. Ten jistě není prvoideálem, existují proto $x, y \notin \mathfrak{a}$, jejichž součinem v $\mathfrak{a}$ leží. Pak ideály $\mathfrak{a} + (x)$ a $\mathfrak{a} + (y)$ oba ostře obsahují samotný ideál $\mathfrak{a}$ a mají tedy vyšší normu, díky našim předpokladům oba obsahují součin nějakých prvoideálů. Díky inkluzi $(\mathfrak{a} + (x)) \cdot (\mathfrak{a} + (y)) \subset \mathfrak{a}$ tak získáváme toužený spor. □

Konečně, na dokončení důkazu budeme do boje muset povolat novou definici:

Definice 3.5.6. Buď $\mathfrak{p} \subset \mathcal{O} \subset K$ prvoideál a definujme jeho inverz jako $\mathcal{O}$-modul:

$$\mathfrak{p}^{-1} := \{x \in K \mid x\mathfrak{p} \subseteq \mathcal{O}\}$$

a definujme násobení $\mathfrak{a}\mathfrak{p}^{-1} := \{\sum a_i p_i | a_i \in \mathfrak{a}, p_i \in \mathfrak{p}^{-1}\} := \mathfrak{p}^{-1}\mathfrak{a}$.

Poslední část definice je díky komutativitě $\mathcal{O}_K$ dobře definovaná a koresponduje s komutativitou násobení ideálů $\mathcal{O}_K$. Důvod zavedení tohoto pojmu závisí na jeho schopnosti *krátit* prvoideály, budeme se ale již muset omezit na maximální pořádek, proč si osvětlíme brzy.

Věta 3.5.7. *Buď* $\mathfrak{p} \subset \mathcal{O}_K$ *prvoideál maximálního pořádku. Pak platí* $\mathcal{O}_K \subset \mathfrak{p}^{-1}$ *a rovnost* $\mathfrak{p}\mathfrak{p}^{-1} = \mathcal{O}_K$.

Důkaz. Protože $\mathfrak{p}$ náleží do $\mathcal{O}_K$, jeho inverz jistě obsahuje celý $\mathcal{O}_K$. Vyberme nyní nenulové $x \in \mathfrak{p}$. Ideál $(x) \subseteq \mathfrak{p}$ podle věty 3.5.5 obsahuje součin prvoideálů $\mathfrak{p}_1 \cdots \mathfrak{p}_k$, kde k je mezi všemi množinami $\{\mathfrak{p}_1, \ldots, \mathfrak{p}_k\}$ nenjnižší možné. Podle věty 3.5.4 $\mathfrak{p}$ je roven jednomu z $\mathfrak{p}_i$, BÚNO ať $\mathfrak{p} = \mathfrak{p}_1$. Díky výběru k ideál (x) neobsahuje $\mathfrak{p}_2 \cdots \mathfrak{p}_k$, uvažme $y \in \mathfrak{p}_2 \cdots \mathfrak{p}_k \setminus (x)$, pak $y/x \notin \mathcal{O}_K$. Platí ale inkluze $y\mathfrak{p} \subseteq \mathfrak{p}\mathfrak{p}_2 \cdots \mathfrak{p}_k \subseteq (x)$, tedy $(y/x)\mathcal{O} \subseteq \mathfrak{p}$ a y/x je proto prvkem $\mathfrak{p}^{-1} \setminus \mathcal{O}_K$.

Nyní již můžeme předpokládat existenci $a \in \mathfrak{p}^{-1} \setminus \mathcal{O}_K$ splňujícího $a\mathfrak{p} \subseteq \mathcal{O}_K$. Platí inkluze $\mathfrak{p} \subseteq \mathfrak{p} + a\mathfrak{p} \subseteq \mathcal{O}_K$, tedy z maximality prvoideálů nastane v jedné z inkluzí rovnost. Dejme tomu, že $\mathfrak{p} = \mathfrak{p} + a\mathfrak{p}$, pak musí platit $a\mathfrak{p} \subseteq \mathfrak{p}$. Protože $\mathfrak{p}$ je konečně generovaný $\mathbb{Z}$-modul, tato inkluze nutně znamená, že a je celý nad $\mathbb{Z}$, spor s $a \notin \mathcal{O}_K$. Platí proto $\mathfrak{p} + a\mathfrak{p} = \mathcal{O}_K$ pro každé $a \in \mathfrak{p}^{-1} \setminus \mathcal{O}_K$, neboli $\mathfrak{p}\mathfrak{p}^{-1} = \mathcal{O}_K$ díky maximalitě prvoideálů. □

Důsledek 3.5.8. *Buďte* $\mathfrak{a}, \mathfrak{b}, \mathfrak{p}$ *ideály* $\mathcal{O}_K$. *Pokud platí* $\mathfrak{p}\mathfrak{a} = \mathfrak{p}\mathfrak{b}$, *tak* $\mathfrak{a} = \mathfrak{b}$.

Důkaz. Vynásobení obou stran $\mathfrak{p}^{-1}$ dává rovnost $\mathfrak{a} = \mathfrak{b}$. □

Nyní již konečně můžeme ukázat jednoznačnost rozkladu prvoideálů v $\mathcal{O}_K$.

Věta 3.5.9. *Každý nenulový ideál* $\mathfrak{a} \subset \mathcal{O}_K$ *lze jednoznačně rozložit na součin prvoideálů.*

Důkaz. Nejprve ukážeme, že každý ideál rozložit lze. Postupujme indukcí vzhledem k t, počtu prvoideálů, jejichž součin $\mathfrak{a}$ obsahuje. Případ $t = 1$ dává $\mathfrak{a}$ prvoideál, dále ať věta platí pro nějaké t a uvažme (ne prvoideál) $\mathfrak{a}$ obsahující $\mathfrak{p}_1 \cdots \mathfrak{p}_{t+1}$. Jistě $\mathfrak{a}$ je obsažen v nějakém maximálním $\mathfrak{p}$, tedy podle věty 3.5.4 je jeden z $\mathfrak{p}_i$ roven $\mathfrak{p}$, ať to je $\mathfrak{p}_1$. Násobením řetězce $\mathfrak{p} \supset \mathfrak{a} \supset \mathfrak{p}_1 \cdots \mathfrak{p}_{t+1}$ modulem $\mathfrak{p}^{-1}$ dává $\mathcal{O}_K \supseteq \mathfrak{p}_2 \cdots \mathfrak{p}_{t+1} \supset \mathfrak{a}\mathfrak{p}^{-1}$, modul $\mathfrak{a}\mathfrak{p}^{-1}$ je tedy ideál $\mathcal{O}_K$ a je rozložitelný, tedy $\mathfrak{a} = \mathfrak{a}\mathfrak{p}^{-1}\mathfrak{p}$ je též.

Nyní se pusťme na jednoznačnost. Dejme tomu naopak, že existují dva rozklady $\mathfrak{p}_1 \cdots \mathfrak{p}_k = \mathfrak{a} = \mathfrak{q}_1 \cdots \mathfrak{q}_\ell$. Ideál $\mathfrak{p}_k$ dělí součin $\mathfrak{q}_i$, je proto roven jednomu z nich. Podle důsledku 3.5.8 a komutativity $\mathcal{O}_K$ je můžeme oba pokrátit (vynásobit $\mathfrak{p}^{-1}$) a pokračovat s ideálem $\mathfrak{a}\mathfrak{p}^{-1}$, čímž ostře snížíme $\max(k, \ell)$, sestupem dojdeme k závěru $k = \ell$ a že množiny prvoideálů na obou stranách musely být, včetně násobnosti, shodné. □

Nyní se již obecněji podívejme na pořádky, kterou část důkazu pro ně nemůžeme adaptovat. Opravdu opět platí, že každý ideál je součinem prvoideálů a prvoideály se chovají podobně jako prvočísla, možnost krátit ale ztrácíme. Než si zmíníme ucelenou větu, ukažme si pár příkladů.

Příklad 3.5.10. Nejprve uvažme pořádek $\mathbb{Z}[2i] \subset \mathbb{Z}[i]$ a jeho ideál (2). Ukážeme, že není prvoideálem. Faktorový okruh $\mathbb{Z}[2i]/(2) \cong \mathbb{Z}[x]/(x^2+4, 2) = \mathbb{Z}[x]/(x^2, 2)$ není oborem integrity, nemůže proto být ani tělesem a podle 3.5.2 není (2) prvoideálem. Platí ale $\mathbb{Z}[x]/(x^2, 2) = \{0, 1, x, x+1\}$, pokud by byl (2) rozložitelný na součin prvoideálů, musí mít oba rank nejvýše 2. Pokud ale $(2) = (a, b)(c, d) = (ac, ad, bc, bd)$¡¡¡¡¡¡¡,

3.6 Grupa tříd ideálů a jednoznačnost rozkladu

Definice 3.6.1. Buď K podílové těleso okruhu R. Pokud $\mathcal{O}$-modul $\mathfrak{a} = m\mathfrak{b}$ je ideál R pro $m \in R$, nazveme $\mathfrak{b}$ lomeným ideálem K. Budeme značit $\mathfrak{b} = \frac{\mathfrak{a}}{m}$.

Definice 3.6.2. Buď K podílové těleso R. Pro $\alpha \in K$ nazveme $(\alpha) = \alpha K$ *hlavním lomeným ideálem* R.

Příkladem hlavního lomeného ideálu okruhu celých čísel tělesa $\mathbb{Q}$ je $\frac{(3)}{2} = \frac{3}{2}\mathbb{Z}$.

Mějme $\mathcal{I}, \mathcal{J}$ lomené ideály pořádku $\mathcal{O}$ a definujme relaci *ekvivalence* $\sim$ s tím, že $\mathcal{I}, \mathcal{J}$ jsou ekvivalentní, pokud existují $a, b \in \mathcal{O}$ taková, že $\mathcal{I} \cdot (a) = \mathcal{J} \cdot (b)$. Relace $\sim$ pak rozkládá množinu lomených ideálů $\mathcal{O}$ na třídy ekvivalence $[\mathcal{I}]$, kde násobení hlavní ideálem ponechá třídu. Snadno nahlédneme, že součin libovolné dvojice ideálů z dané dvojice tříd vždy spadá do stejné třídy, neboli $[\mathcal{I}] \cdot [\mathcal{J}] = [\mathcal{I} \cdot \mathcal{J}]$. Tyto třídy ekvivalence z asociativity násobení ideálů proto tvoří grupu s násobením ideálů.

Definice 3.6.3. Buď $\mathcal{O}$ pořádek číselného tělesa K. Označme $\mathsf{G}(\mathcal{O})$ množinu všech lomených ideálů $\mathcal{O}$ a množinu hlavních lomených ideálů $\mathsf{H}(\mathcal{O})$. *Grupu tříd ideálů* $\mathcal{O}$ pak definujeme jako faktorgrupu $Cl(\mathcal{O}) = \mathsf{G}(\mathcal{O})/\mathsf{H}(\mathcal{O})$.

Z předchozí diskuze plyne, že neutrálním prvkem grupy $Cl(\mathcal{O})$ je právě množina hlavních ideálů $\mathcal{O}$. Ukáže se, že pořádky číselných těles mají pouze konečný počet tříd, v obecném okruhu to neplatí. Můžeme se tak bavit o počtu jejích prvků.

Definice 3.6.4. *Třídové číslo* $h_{\mathcal{O}}$ pořádku $\mathcal{O}$ definujeme jako počet prvků grupy $Cl(\mathcal{O})$.

Každý prvek konečné grupy umocněn na její řád se stává neutrálním. Tento fakt v případě grupy tříd ideálů zní:

Věta 3.6.5. *Buď* $\mathcal{O}$ *pořádek a* $\mathcal{I}$ *jeho lomený ideál. Pak ideál* $\mathcal{I}^{h_{\mathcal{O}}}$ *je hlavním ideálem* $\mathcal{O}$.

Kapitola 4

Okruhy Endomorfismů

Vraťme se k eliptickým křivkám. Jak napovídá název této sekce, endomorfismy na křivce E tvoří okruh. Tento okruh se budeme snažit s pomocí teorie představené v předchozích kapitolách charakterizovat. Omezíme se pro tentokrát na křivky (a tedy i endomorfismy) nad $\mathbb{F}_p$, což nám mnohé věci podstatně usnadní.

Definice 4.0.1. Mějme $E/\mathbb{F}_p$ eliptickou křivku. Označme $\mathrm{End}(E)$ množinu isogenií $\phi : E(\mathbb{F}_p) \longrightarrow E(\mathbb{F}_p)$ spolu s $[0]$. Prvky $\mathrm{End}(E)$ nazvěme *endomorfismy* na E.

Věta 4.0.2. *Množina* $\mathrm{End}(E)$ *tvoří spolu s operacemi* $+$ *a* $\circ$ *okruh.*

Důkaz. Sčítání i skládání endomorfismů jistě zachovává doménu a obor hodnot $E(\mathbb{F}_p)$. Sčítání endomorfismů na E je komutativní i asociativní, přičemž $[0]$ je neutrálním prvkem pro sčítání, a ke každé isogenii ϕ je isogenie $[-1] \circ \phi$ opačnou k ϕ vzhledem ke sčítání. Dále skládání isogenií je asociativní a $[1]$ je jeho neutrálním prvkem. Konečně, skládání je na sčítání oboustranně distributivní, protože endomorfismy na E jsou homomorfismy grup $E(\mathbb{F}_p) \longrightarrow E(\mathbb{F}_p)$. □

V této kapitole se pokusíme přijít na kloub samotné struktuře okruhu endomorfismů a grafům isogenií, které nám pospolu pomohou osvětlit funkčnost několika pár vybraných kryptografických schémat založených na isogeniích.

Než začneme, všimněme si všudypřítomného injektivního homomorfismu $\mathbb{Z} \longrightarrow \mathrm{End}(E)$ daného $m \mapsto [m]$. Protože množina složená ze skalárních násobků na E je isomorfní okruhu $\mathbb{Z}$, můžeme v $\mathrm{End}(E)$ isogenie $[m]$ stotožnit s jejich základem m a považovat inkluzi $\mathbb{Z} \subseteq \mathrm{End}(E)$ za platnou. Kvůli tomuto rozhodnutí bude též přirozenější skládání isogenií zapisovat ve stylu násobení.

Úmluva. V okruhu endomorfismů $\mathrm{End}(E)$ budeme složení isogenií $\phi \circ \psi$ psát jako $\phi\psi$ a isogenii $[m]$ stotožníme s číslem m.

Kvůli multiplikativitě stupnů isogenií (a 0), můžeme o okruhu endomorfismů říci, že je oborem integrity a speciálně má nulovou charakteristiku. Surjektivita isogenií nám též

umožňuje nenulové endomorfismy „krátit“, jak bychom očekávali u okruhu s nenulovou charakteristikou.

4.1 Stopa endomorfismu

Vraťme se nyní na chvíli k první kapitole a duální isogenii. Ta má několik vlastností, které by po seznámení s normou a stopou prvku kvadratického tělesa měly znít povědomě.

Duální isogenie definuje automorfismus na (komutativním) okruhu $\mathrm{End}(E)$, který dokonce prohazuje pořadí skládání násobení, tvoří tedy strukturu známou jako *antihomomorfismus*.

Věta 4.1.1. *Buďte $\phi, \psi \in \mathrm{End}(E)$. Pak $\widehat{\phi\psi} = \widehat{\psi}\widehat{\phi}$.*

Důkaz. Pokud jedna z ϕ, ψ je nulová, věta jistě platí, dále ať ϕ, ψ nulové nejsou. Díky vlastnostem duální isogenie platí v okruhu $\mathrm{End}(E)$:

$$(\widehat{\psi}\widehat{\phi})(\phi\psi) = \widehat{\psi}(\widehat{\phi}\phi)\psi = \widehat{\psi}(\deg\phi)\psi = \widehat{\psi}\psi\deg\phi = \deg\psi\deg\phi = \deg\psi\phi = \deg\phi\psi = (\widehat{\phi\psi})(\phi\psi),$$

tedy díky surjektivitě isogenie $\phi\psi$ platí $\widehat{\psi}\widehat{\phi} = \widehat{\phi\psi}$. □

Nyní nastává čas si vzpomenout na důkaz věty 1.6.1, speciálně že součet $\pi + \widehat{\pi}$ je v $\mathrm{End}(E)$ celé číslo. Naprosto stejně můžeme postupovat u libovolného jiného endomorfismu.

Věta 4.1.2. *Každý endomorfismus ϕ na E splňuje $\phi + \widehat{\phi} \in \mathbb{Z}$.*

Důkaz. Nulová isogenie tvrzení jistě splňuje. Pro ostatní endomorfismy na E si „roznásobme“ výraz $(1-\phi)(1-\widehat{\phi})$:

$$\deg(1-\phi) = (1-\phi)\widehat{(1-\phi)} = (1-\phi)(1-\widehat{\phi}) = 1 - (\phi + \widehat{\phi}) + \phi\widehat{\phi},$$
$$\phi + \widehat{\phi} = 1 + \phi\widehat{\phi} - \deg(1-\phi) = 1 + \deg\phi - \deg(1-\phi) \in \mathbb{Z},$$

což jsme chtěli. □

Stopa prvku v kvadratickém tělese je rovna součtu prvku a jeho konjugátu a je celým číslem, následující definice proto čtenáře nepřekvapí:

Definice 4.1.3. Buď $\phi \in \mathrm{End}(E)$ endomorfismus. Pak definujeme jeho *stopu* jako:

$$\mathrm{Tr}\,\phi := \phi + \widehat{\phi} \in \mathbb{Z}.$$

Můžeme pak ukázat, že $\mathrm{End}(E)$ tvoří kvadratický okruh.

Věta 4.1.4. *Každý endomorfismus ϕ na E je v* $\mathrm{End}(E)$ *kořenem charakteristického polynomu ϕ:*

$$x^2 - \mathrm{Tr}\,\phi + \deg\phi \in \mathbb{Z}[x].$$

Důkaz. Víme, že $\mathrm{Tr}\,\phi = \phi + \widehat{\phi} = \mathrm{Tr}\,\widehat{\phi}$ a $\deg\phi = \phi\widehat{\phi} = \deg\widehat{\phi}$ jsou v $\mathrm{End}(E)$ celá čísla. Viétovy vztahy pak tvrdí, že ϕ a $\widehat{\phi}$ jsou kořeny polynomu výše. □

Endomorfismus, ke kterému se pořád dokola vracíme, je ten pojmenovaný po Frobeniovi. Jeho charakteristický polynom je:

$$x^2 - tx + p = 0,$$

kde $t = p + 1 - \#E(\mathbb{F}_p)$ je stopa Frobenia. V případě supersingulární křivky je lineární člen tohoto polynomu nulový a $\pi = \pm\sqrt{-p}$.

Jako norma v kvadratickém okruhu $\mathrm{End}(E)$ nám působí $\deg\phi$ a ta je zjevně multiplikativní, očekávali bychom proto stopu endomorfismů aditivitní.

Lemma 4.1.5. *Buďte $\phi, \psi \in \mathrm{End}(E)$. Pak platí* $\mathrm{Tr}\,\phi + \psi = \mathrm{Tr}\,\phi + \mathrm{Tr}\,\psi$.

Důkaz. Díky větě 1.3.8 platí:

$$\mathrm{Tr}\,\phi + \psi = \phi + \psi + \widehat{\phi + \psi} = \phi + \psi + \widehat{\phi} + \widehat{\psi} = \mathrm{Tr}\,\phi + \mathrm{Tr}\,\psi.$$

□

4.2 Algebra endomorfismů

V této sekci okruh endomorfismů rozšíříme do vektorového prostoru nad $\mathbb{Q}$ za pomoci tenzorového součinu. Ten skrývá známou strukturu imaginárního kvadratického tělesa a to ne jen tak libovolného, dokonce $\mathbb{Q}(\pi)$. Představme si proto tento modul:

Definice 4.2.1. Buď $E/\mathbb{F}_p$ eliptická křivka. Pak $\mathbb{Z}$-modul definovaný jako:

$$\mathrm{End}^0(E) := \mathbb{Q} \otimes_{\mathbb{Z}} \mathrm{End}(E)$$

nazveme *algebrou endomorfismů* E.

Algebra endomorfismů je generovaná formálními výrazy $r \otimes \phi$, kde $r \in \mathbb{Q}, \phi \in \mathrm{End}(E)$, podle věty 3.1.13 je každý její prvek právě takového tvaru. Zásadní problém s propozicí, že algebra endomorfismů je těleso, je součin tenzorů, který jsme si nedefinovali. Protože ale $\mathrm{End}^0(E)$ má „jednoduché“ prvky, součin dvou tenzorů přichází přímočaře

Definice 4.2.2. Buďte $r \otimes \phi, s \otimes \psi \in \mathrm{End}^0(E)$. Pak definujeme:

$$(r \otimes \phi)(s \otimes \psi) := rs \otimes \phi\psi.$$

Mnoho vlastností okruhu endomorfismů sáha i do $\mathrm{End}^0(E)$, speciálně zřejmě je oborem integrity a tedy má nulovou charakteristiku.

Jistě opět panují injektivní homomorfismy $\mathbb{Q} \longrightarrow \mathrm{End}^0(E)$ a $\mathrm{End}(E) \longrightarrow \mathrm{End}^0(E)$ dané zobrazeními $r \mapsto r \otimes 1$, resp. $\phi \mapsto 1 \otimes \phi$. Ty bychom znovu chtěli uvažovat tyto raději jako inkluze.

Úmluva. Prvky algebry endomorfismů budeme místo $r \otimes \phi$ značit $r\phi$ a považovat inkluze $\mathbb{Q} \subseteq \mathrm{End}^0(E)$ a $\mathrm{End}(E) \subseteq \mathrm{End}^0(E)$ za platné.

Zastavme se nyní, poohlédněme se na předchozí kapitoly, a naplánujme další postup útoku. Prostředek, který připomíná vlastnosti kvadratických těles nejvíce, je duální isogenie jako konjugát prvku. Tento koncept si proto rozšíříme i na algebru endomorfismů:

Definice 4.2.3. Buď $r\phi \in \mathrm{End}^0(E)$. Pak definujeme *Rosatiho involuci* $\widehat{r\phi} = r\widehat{\phi}$.

Jistě položením $\phi = 1$ v definice výše platí $r = \widehat{r}$ a opět snadno dojdeme k tomu, že $\widehat{r\phi}$ je involuce. I ostatní vlastností involuce $\widehat{\phi}$, tedy že je aditivní a antihomomorfismem, samozřejmě platí v $\mathrm{End}^0(E)$. No a kam chodí konjugát, tam se podívají i stopa a norma.

Definice 4.2.4. Normu a stopu prvku $\alpha \in \mathrm{End}^0(E)$ definujeme jako:

$$\mathrm{Tr}\,\alpha = \alpha + \widehat{\alpha},$$
$$\mathrm{N}\,\alpha = \alpha\widehat{\alpha}.$$

Pojďme tedy začít v rychlosti budovat korespondence mezi stopou a normou endomorfismu a těmi příslušící prvku (imaginárního) kvadratického tělesa, kterých je opravdu velká kupa.

Věta 4.2.5. *Norma i stopa $\alpha \in \mathrm{End}^0(E)$ jsou nezáporná racionální čísla a norma je nulová, jen pokud $\alpha = 0$.*

Důkaz. Norma prvku $r\phi$ je rovna $r\phi\widehat{r\phi} = r^2\phi\widehat{\phi} = r^2 \deg\phi$ je nezáporné racionální číslo a jeho stopa je $r\phi + \widehat{r\phi} = r(\phi + \widehat{\phi})$ též. Pokud je norma α nulová, je buď $r = 0$, nebo $\deg\phi = 0$, každopádně $\alpha = 0$. □

Norma i stopa jsou úzce spojeny s konjugáty prvku racionálního čísla, konkrétně ji všechny sdílí. Vlastností, které jsme o normě a stopě odvozovali ve třetí kapitole přichází prakticky zdarma.

Věta 4.2.6. *Pro libovolná $\alpha, \beta \in \mathrm{End}^0(E)$ a $k, \ell \in \mathbb{Q}^+$ platí $\mathrm{Tr}\,k\alpha + \ell\beta = k\,\mathrm{Tr}\,\alpha + \ell\,\mathrm{Tr}\,\beta$, $\mathrm{Tr}\,\alpha = \mathrm{Tr}\,\widehat{\alpha}$, $N\alpha = N\widehat{\alpha}$ a $\mathrm{N}\,\alpha\beta = \mathrm{N}\,\alpha\,\mathrm{N}\,\beta$.*

Důkaz. Aditivita a stopy plyne z aditivity involuce $\widehat{\phi}$ a $\mathbb{Q}$-linearita pak plyne z definice. Protože je Rosatiho involuce involucí, platí:

$$\operatorname{Tr}\widehat{\alpha} = \widehat{\alpha} + \widehat{\widehat{\alpha}} = \widehat{\alpha} + \alpha = \operatorname{Tr}\alpha.$$

Dále:

$$\alpha \operatorname{N}\widehat{\alpha} = \alpha\widehat{\alpha}\alpha = \operatorname{N}\alpha\alpha = \alpha \operatorname{N}\alpha,$$

tedy, protože algebra endomorfismu je oborem integrity, platí $\operatorname{N}\alpha = \operatorname{N}\widehat{\alpha}$. Konečně, podle věty 4.1.1 platí:

$$\operatorname{N}\alpha\beta = \alpha\beta\widehat{\alpha\beta} = \alpha\beta\widehat{\beta}\widehat{\alpha} = \alpha(\operatorname{N}\beta)\widehat{\alpha} = \alpha\widehat{\alpha}\operatorname{N}\beta = \operatorname{N}\alpha\operatorname{N}\beta.$$

□

Důsledek 4.2.7. *Buď $\alpha \in \operatorname{End}^0(E)$ nenulové. Pak má v $\operatorname{End}^0(E)$ multiplikativní inverz.*

Důkaz. Položme $\alpha^{-1} = \widehat{\alpha}/\operatorname{N}\alpha \in \operatorname{End}^0(E)$. Ukážeme, že toto α^{-1} je hledaným inverzem, platí totiž $\alpha\alpha^{-1} = \alpha\widehat{\alpha}/\operatorname{N}\alpha = 1$, je tedy levým inverzem. Analogicky $(\widehat{\alpha}/\deg\alpha)\alpha = 1$, tedy α^{-1} je opravdu hledaným prvkem. □

Předchozí věta nám opodstatní fakt, že algebra endomorfismů tvoří *division ring*, tedy splňuje všechny podmínky na těleso až na nutnost komutativity násobení. Povšimněme si, že pokud bychom endomorfismy i křivky místo nad $\mathbb{F}_p$ doteď definovali nad nějakým jeho rozšířením, pramálo by se změnilo, násobení se ale už bude lišit. Významným výsledkem artibutovaným Maxu Deuringovi [12] je klasifikace algeber endomorfismů nad libovolným konečným tělesem. Ukáže se, že buď jsou isomorfní tělesu racionálních čísel, imaginárnímu kvadratickému tělesu (k čemuž směřujeme), či tzv. *kvaternionové algebře*, tedy rozšíření $\mathbb{Q}(\alpha, \beta)$ s $\alpha\beta = -\beta\alpha$, kde na pořadí zápisu násobení jistě záleží.

Můžeme si pak odůvodnit, proč mají všechny prvky algebry endomorfismů stupeň nejvýše dvě:

Věta 4.2.8. *Každé $\alpha \in \operatorname{End}^0(E)$ je kořenem polynomu:*

$$x^2 - \operatorname{Tr}\alpha + \operatorname{N}\alpha \in \mathbb{Q}[x].$$

Důkaz. Viétovy vztahy říkají, že kořeny tohoto polynomu jsou α a $\widehat{\alpha}$. □

Věta 4.2.9. *Buďte E, E' křivky nad $\mathbb{F}_p$. Pokud jsou tyto dvě křivky jsou nad $\mathbb{F}_p$ isogenní, platí $\operatorname{End}^0(E) \cong \operatorname{End}^0(E')$.*

Důkaz. Uvažme $\phi : E \longrightarrow E'$ isogenii nad $\mathbb{F}_p$. Zobrazení $\operatorname{End}(E) \longrightarrow \operatorname{End}(E')$ (diagram) dané $\psi \mapsto \phi\psi\widehat{\phi}$ je homomorfismem těchto $\mathbb{Z}$-modulů. Můžeme pak definovat injektivní homomorfismus $\operatorname{End}^0(E) \longrightarrow \operatorname{End}^0(E')$ via zobrazení $r\psi \mapsto r\phi\psi\widehat{\phi}$ a duálně definujeme homomorfismus inverzní. □

Endomorphism ring závisí na j invariantu !!!

4.3 Frobeniův endomorfismus

Tvar Hasseho věty 1.1.5 připomíná diskriminant kvadratické rovnice. Vskutku, dá se ukázat, že v $\mathrm{End}(E)$ splňuje Frobeniův endomorfismus kvadratický vztah.

Věta 4.3.1. *Frobeinův endomorfismus π na křivce E nad $\mathbb{F}_q$ v $\mathrm{End}(E)$ splňuje:*

$$\pi^2 - t\pi + q = 0$$

pro nějaké $t \in \mathbb{Z}$, které nazveme stopou Frobenia.

4.4 Obyčejné křivky

4.5 Supersingulární křivky

Konečně se obraťme zpět k supersingulárním křivkám. Víme, že stopa Frobenia je nulová, tedy $\pi^2 = -p$ v $\mathrm{End}(E)$. To znamená, že $\mathbb{Z}[\sqrt{-p}] \subseteq \mathrm{End}(E)$.

Kapitola 5

Kryptosystémy založené na isogeniích

První kryptografické schéma založené na isogeniích obyčejných eliptických křivek navrhl Couveignes [7] již v roce 1997, nicméně nepublikoval jej po dalších deset let. Grafy isogenií byly studovány přes přelom tisíciletí [15], ?. Roku 2006 Rostovtsev a Stolbunov [33] nezávisle na Couveignovi navrhli (prakticky shodný) protokol založen na cestách v grafu obyčejných isogenií.

???
Po celou tuto dobu se supersingulárním křivkám nevěnovalo druhé myšlenky,

Pojďme se nyní znovu podívat na větu 1.4.4. Křivky $\phi(\psi(E)), \psi(\phi(E))$ sdílí j-invariant, neboť jsou isomorfní, což by v potenciálním protokolu založeném na isogeniích mohlo být sdílené tajemství obou stran.

Pokud tak mají obě strany danou počáteční křivku E nad $\overline{\mathbb{F}_q}$ a vyberou si tajné separabilní isogenie ϕ_A, resp. ϕ_B, kter pošlou druhé straně $\phi_A(E)$, resp. $\phi_B(E)$, pouze již s malým množstvím dalších informací obě strany snadno spočtou své tajemství. Takové myšlenky měli De Feo, Jao a Plût v [11], nicméně než se dostaneme přímo k jejich navrhovaném protokolu SIDH, musíme diskutovat několik důležitých detailů, které výměnu umožňují.

$$\begin{array}{ccc} E & \xrightarrow{\phi} & \phi(E) \\ \downarrow{\scriptstyle \psi} & & \downarrow{\scriptstyle \psi} \\ \psi(E) & \xrightarrow{\phi} & \phi\left(\psi(E)\right) \end{array}$$

Jak napovídá název protokolu, budeme pracovat se supersingulárními eliptickými křivkami E nad $\mathbb{F}_{p^2}$ pro prvočíslo $p = \ell_A^{e_A}\ell_B^{e_B} - 1$, kde ℓ_A, ℓ_B jsou (malá) prvočísla. Pokud E má Frobeniův endomorfismus, že $\pi^2 = [-2p]$, díky **??** je $\#E(\mathbb{F}_{p^2}) = (p+1)^2$, přičemž $E[p+1] \cong \mathbb{Z}_{p+1} \times \mathbb{Z}_{p+1}$, tedy

Pak totiž z věty ? je $\#E(\mathbb{F}_{p^2}) = (p+1)^2$ a dle věty 1.6.1: $E(\mathbb{F}_{p^2}) \cong \mathbb{Z}_{p+1} \times \mathbb{Z}_{p+1}$. Pro prvočíslo $p = \ell_A^{e_A}\ell_B^{e_B} - 1$, kde ℓ_A, ℓ_B jsou (malá) prvočísla, proto existují dva body G_1, G_2 řádu $\ell_A^{e_A}\ell_B^{e_B}$, které generují $E(\mathbb{Z}_{p^2})$. Speciálně dvojice $\langle P_A, Q_A\rangle := \langle [\ell_B^{e_B}]G_1, [\ell_B^{e_B}]G_2\rangle$, resp. $\langle P_B, Q_B\rangle := \langle [\ell_A^{e_A}]G_1, [\ell_A^{e_A}]G_2\rangle$, generují po řadě $\ell_A^{e_A}$, $\ell_B^{e_B}$ torzi.

Uvažme bod $P \in E[\ell_A^{e_A}]$ řádu ℓ_A^t a separabilní isogenii $\phi : E \longrightarrow E/\langle P\rangle$. Pokud bychom chtěli $E/\langle P\rangle$ spočíst, stačilo by spočítat celou $\langle P\rangle$ a za pomocí Véluových formulí spočíst výslednou křivku v čase $O(\ell_A^t)$, což zjevně není optimální.

Veřejné parametry: Grupa G řádu p, kde p je prvočíslo, s generátorem g.

Alfréd		Blažena
Vstup: $a \in G$		**Vstup:** $b \in G$
	$\xrightarrow{g^a}$	
	$\xleftarrow{g^b}$	
$G_{AB} := \left(g^b\right)^a = g^{ab}$		$G_{BA} := \left(g^a\right)^b = g^{ba}$
Výstup: G_{AB}		**Výstup:** G_{BA}

Algoritmus 1: Diffie-Hellmanova výměna

Poznámka. Dle věty 1.4.4 jsou křivky, které obě partie na konci protokolu získají, isomorfní. Nicméně pokud na výpočet isogenií užíváme Véluovy formule, tak můžeme heuristicky ověřit, že křivky vždy vycházejí dokonce shodné. Tato domněnka byla dokázána pravdivou v [21]. Místo j-invariantu konečné křivky tak můžeme za sdílené tajmenství považovat přímo koeficienty konečné křivky. Tímto krokem rozlišujeme isomorfní křivky a tedy ? útoky hrubou silou.

!!! příklad !!!!

5.1 Možné útoky na SIDH

Když jsme obeznámeni s vnitřními machinacemi protokolu SIDH, pojďme se pokusit najít způsoby, jak jej rozbít.

Nejprve pozapomeňme na fakt, že známe obrazy generátorů ℓ torzí. Uvažme graf ℓ_A-isogenií ze supersingulární křivky E. Známe počáteční i koncový vrchol E, resp. E_A cesty na našem grafu, která je složená z kroků ϕ_i, což jsou ℓ_A-isogenie, a hledáme posloupnost ϕ_i. Náš úkol tak můžeme přefrázovat čistě jako hledání cesty v grafu ℓ_A isogenií.

Víme, že hledaná cesta má délku e_A, a každý j-invariant sousedí s právě $p+1$ dalšími. Můžeme jednoduše začít prohledávat graf z E, dokud nenarazíme na E_A. Ze všech ? cest končí ? v E_A, očekáváme proto, že na E_A narazíme v $O(\sqrt{p})$ evaluacích ℓ_A isogenie.

Tento postup můžeme vylepšit takzvaným „Meet in The Middle“ útokem, který běží rychleji na úkor prostorové náročnosti. Pod jeho návodem prohlédáváme jak z E, tak i z E_A, přičemž z E hledáme cesty délky $\left\lfloor \frac{e_A}{2} \right\rfloor$ a z E_A cesty délky $\left\lceil \frac{e_A}{2} \right\rceil$. Dle narozeninového paradoxu bychom očekávali shodu v ? evaluacích ℓ_A isogenie.

příklad útkou !!!!

5.2 Následníci protokolu SIDH

Od jeho publikace v roce 2011 pár kolektivů autorů našlo varianty SIDH založené na různých vlastnostech isogenií. Pár z nich zde uveďme.

eSIDH.

BSIDH.

CSIDH.

Závěr

zu ende

Literatura

[1] Čermák, Filip a Matěj Doležálek: *Teorie nejen čísel*. Seriál korespondenčního matematického semináře.

[2] Chen, Evan: *An Infinitely Large Napkin*. Dostupné z: https://venhance.github.io/napkin/Napkin.pdf.

[3] Chuang, Isaac L. a Michael A. Nielsen: *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

[4] Conrad, Keith: *Trace and Norm*. University of Connecticut, Connecticut. Dostupné z: https://kconrad.math.uconn.edu/blurbs/galoistheory/tracenorm.pdf.

[5] Conrad, Keith: *Ideal Factorization*. University of Connecticut, Connecticut. Dostupné z: https://kconrad.math.uconn.edu/blurbs/gradnumthy/idealfactor.pdf.

[6] Costello, Craig: *Supersingular isogeny key exchange for beginners*. Microsoft Research, USA, 2019. Dostupné z: https://eprint.iacr.org/2019/1321.

[7] Couveignes, Jean-Marc: *Hard Homogenous Spaces*. 2006. Dostupné z: https://eprint.iacr.org/2006/291.pdf.

[8] Cox, David: *Primes of the form $x^2 + ny^2$: Fermat, Class Field Theory and Complex Multiplication*. New York, 1989.

[9] De Feo, Luca: *Fast Algorithms for Towers of Finite Fields and Isogenies*. EcolePolytechnique X, 2010.

[10] De Feo, Luca: *Mathematics of Isogeny Based Cryptography*. Université de Versailles & Inria Saclay, 2017. Dostupné z: https://arxiv.org/abs/1711.04062.

[11] De Feo, Luca, David Jao a Jérôme Plût: *Towards quantum-resistant cryptosystems from supersingular elliptic curve isogenies*. Math. Cryptol. 8(3): 209-247, 2014. Dostupné z: https://eprint.iacr.org/2011/506.pdf.

[12] Deuring, Max, *Die typen der multiplikatorenringe elliptischer funktionenkörper*. Abhandlungen aus dem mathematischen Seminar der Universität Hamburg, 14: 197-272, 1941.

[13] Diffie, Whitfield a Martin Hellman: *New Directions in Cryptography.* IEEE Transactions on Information Theory 22, 1976.

[14] Feynman, Richard P.: *Simulating physics with computers.* Int J Theor Phys 21, 467–488, 1982. Dostupné z: https://doi.org/10.1007/BF02650179.

[15] Galbraith, Steven D.: *Constructing Isogenies Between Elliptic Curves Over Finite Fields.* LMS J. Comput. Math., 199, 118-138, 1999. Dostupné z: https://www.math.auckland.ac.nz/~sgal018/iso.pdf.

[16] Grover, Lov K.: *A fast quantum mechanical algorithm for database search.*28th Annual ACM Symposium on the Theory of Computing, 1996. Dostupné z: https://arxiv.org/abs/quant-ph/9605043.

[17] Hartshorne, Robin: *Algebraic Geometry.* Berkley: Springer-Verlag, 1977.

[18] Ireland, Kenneth a Michael Rosen: *A Classical Introduction to Modern Number Theory.* New York, Berlin a Heidelberg: Springer-Verlag, 1982.

[19] Johnson, Lee W., Ronald Dean Riess a Jimmy Thomas Arnold. *Introduction to Linear Algebra.* Fifth edition. Virginia Polytechnic Institute and State University: Addison-Wesley, 2002.

[20] Kuřil, Martin: *Základy teorie grup.*

[21] Leonardi, Christopher: *A Note on the Ending Elliptic Curve in SIDH.* 2020. Dostupné z: https://eprint.iacr.org/2020/262.

[22] Marcus, Daniel A.: *Number fields.* New York: Springer-Verlag, 1977.

[23] Matushak, Andy a Michael A. Nielsen: *Quantum computing for the very curious.* San Francisco, 2019. Dostupné z: https://quantum.country/qcvc.

[24] Mordell, Luis J.: *On the rational solutions of the indeterminate equations of the third and fourth degrees.* Cambridge, 1922.

[25] Neukirch, Jürgen: *Algebraic Number Theory.* New York: Springer-Verlag, 1999.

[26] Perutka, Tomáš: *Vyjadřování prvočísel kvadratickými formami.* Středoškolská odborná činnost. Brno: Masarykova univerzita, 2017.

[27] Perutka, Tomáš: *Užití dekompoziční grupy k důkazu zákona kvadratické reciprocity.* Středoškolská odborná činnost. Brno: Masarykova univerzita, 2018.

[28] Pezlar, Zdeněk: *Zajímavá využití algebraické teorie čísel.* Středoškolská odborná činnost. Brno: Masarykova univerzita, 2020.

[29] Pupík, Petr: *Užití grupy tříd ideálů při řešení některých diofantických rovnic*. Diplomová práce. Brno: Masarykova univerzita, 2009. Dostupné z: https://is.muni.cz/th/v8xsj/.

[30] Raclavský, Marek: *Racionální body na eliptických křivkách*. Bakalářská práce. Praha: Univerzita Karlova, 2014. Dostupné z: https://is.cuni.cz/webapps/zzp/detail/143352/.

[31] Rivest, Ronald L., Adi Shamir a Leonard M. Adleman: *A Method for Obtaining Digital Signatures and Public-Key Cryptosystems*. 1977. Dostupné z: https://people.csail.mit.edu/rivest/Rsapaper.pdf.

[32] Rosický, Jiří: *Algebra*. Brno: Masarykova univerzita, 2002.

[33] Rostovtsev, Alexander a Anton Stolbnov: *Public-key cryptosystem based on isogenies*. 2006. Dostupné z: http://eprint.iacr.org/2006/145/.

[34] Silverman, Joseph H.: *The Arithmetic of Elliptic Curves*. New York: Springer-Verlag, 1992.

[35] Schoof, René: *Elliptic Curves Over Finite Fields and the Computation of Square Roots* mod p. J. Théor. Nombres Bordeaux 7, 1985. Dostupné z: https://www.ams.org/journals/mcom/1985-44-170/S0025-5718-1985-0777280-6/S0025-5718-1985-0777280-6.pdf.

[36] Shor, Peter W.: *Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer*. New York: Springer-Verlag, 1994. Dostupné z: https://arxiv.org/abs/quant-ph/9508027.

[37] Stein, William: *A Brief Introduction to Classical and Adelic Algebraic Number Theory*. 2004. Dostupné z: https://wstein.org/papers/ant/html/node93.html.

[38] Suchánek, Vojtěch: *Vulkány isogenií v kryptografii*. Diplomová práce. Brno: Masarykova univerzita, 2020. Dostupné z: https://is.muni.cz/th/pxawb/.

[39] Sutherland, Andrew V.: *Elliptic Curves*. Massachusetts Institute of Technology, 2017. Dostupné z: https://math.mit.edu/classes/18.783/2017/lectures.html.

[40] Tate, John: *Endomorphisms of Abelian Varieties over Finite Fields*. Inventiones Mathematicae, 2 (2): 134–144, Cambridge, 1966.

[41] Vélu, Jacques: *Isogénies entre courbes elliptiques*. Comptes Rendus de l'Académie des Sci-ences de Paris, 1971. Dostupné z: https://math.mit.edu/classes/18.783/2017/lectures.html.

[42] Washington, Lawrence C.: *Elliptic Curves: Number theory and cryptography*. Maryland, 2008.

[43] Weil, André: *L'arithmétique sur les courbes algébriques*. Acta Mathematica 52, 1929.