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

Za advent užití eliptických křivek k kryptografii lze považovat kryptosystémy (nezávisle) navržené Koblitzem [36] a Millerem [42] v polovině 80. letech minulého století založené na principu Diffie-Hellmanovy [20] výměny. V dnešních dobách jsou autentifikační protokoly pracující s eliptickými křivkami hojně užíváné, jak transakce kryptoměny Bitcoin tak přihlašování do služeb PlayStation jsou chráněné pevnou rukou protokolu ECDSA [32].

Zmíněné protokoly, založené na obtížnosti problému diskrétního logaritmu v konečné grupě, případně na eliptické křivce, stejně jako další prominentní systémy navrženy ke konci minulého století, včetně známého RSA [53] založeného na rozkladu přirozeného čísla, ovšem všechny padají v polynomiálním čase pod rukou algoritmu navrženého v 90. letech Shorem [60], ne však na klasickém počítači, ale ve světě kvantovém. Cílem odborníku pracujících v tzv. post-kvanotvé kryptografii je pak hledat kryptografická primitiva, která pod hrozbou klasickou i kvantovou obstojí.

Přímočará adaptace Diffie-Hellmanovy výměny navržená Koblitzem a Millerem však zdaleka není vše, co eliptické křivky nabízejí. Konkrétně zobrazení mezi nimi, který zachovají jejich grupovou strukuru, tzv. *isogenie*, skýtají bohatou strukturu a nabízejí potenciální těžko prolomitelné protokoly.

Kořeny studia isogenií sahají hluboko do světa algebraické geometrie a studium tohoto oboru poskytuje mnohem lepší pohled „pod kapotu“ teorie s nimi spojené. Naše práce volí jiný směr, nabízíme totiž úvod do studia těchto struktur přístupný pro studenta bakalářského studia matematiky na vysoké škole. „Elementární“ pohled na věc nás donutí jistá klíčová tvrzení předpokládat a priori za platné, práce ale díky tomuto rozhodnutí tvoří dobrý úvod pro čtenáře nezasvěceného do studia eliptických křivek.

Ke konci první kapitoly budeme mít dostatek znalostí k diskuzi protokolu SIDH a na něm založeném kryptosystému SIKE [18], který je jedním z nejprospektivnějších doposud navržených kandidátů na kvantově bezpečný protokol. Tento protokol mimo samotné diskuze implentujeme pro čtenáře k vyzkoušení a podíváme se na některé protokoly na SIDHu založené. Nejnovětší z nich, SITH ??, implementujeme též. Poté se vydáme na cestu ke hlubšímu studiu endomorfismů na eliptické křivky, tedy isogenií zobratujících křivku samu na sebe. K této příležitosti odbočíme do světa algebraické teorie čísel, jejíž poznatky nám budou ve studiu endomorfismů velmi přínosné. Konečne, podíváme se na akci tzv. *grupy tříd ideálů* na množinu isomorfismů našich křivek, která dává vzniku spiritálnímu následníku SIDH pod názvem CSIDH, které zběžně probereme.

Použitá značení

$a \mid b$	a dělí b
$\frac{1}{a}$	multiplikativní inverz a, tj. a^{-1}
$\nu_p(n)$	p-adická valuace n
$\left(\frac{a}{p}\right)$	Legendreův symbol a vzhledem k p
$\bigcap_X$	průnik všech množin $M \in X$
$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$	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
$\langle G \rangle$	podgrupa $E(\overline{K})$ generovaná množinou G
E/G	obraz E v separabilní isogenii s jádrem G
$E/\mathfrak{a}$	obraz E v isogenii generované ideálem $\mathfrak{a}$
$E[n]$	n-torze křivky E

$\mathrm{End}(E)$	okruh endomorfismů E
$M \otimes_R N$	tenzorový součin R-modulů M a N
$\mathrm{End}^0(E)$	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 $\alpha \in \mathrm{End}^0(E)$
$j(E)$	j-invariant křivky E
$G_\ell(\overline{\mathbb{F}}_p)$	graf supersingulárních j-invariantů nad $\overline{F}_p$spojených isogeniemi stupně ℓ
$R[x]$	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
$\mathrm{SL}_2(\mathbb{F})$	množina 2×2 matic nad $\mathbb{F}$
$Tr_K(\alpha)$	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 $\mathfrak{b}$
$\mathfrak{ab}, \mathfrak{a} \cdot \mathfrak{b}$	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, lomené funkce f
f'	derivace f
$f\vert_M$	zúžení f na množinu M
$\phi\vert_\ell$	zúžení isogenie ϕ na ℓ-torzi na E
$f \in O(g)$	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

It is possible to write endlessly on elliptic curves (This is not a threat.)

Serge Lang

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, pro porozumění této kapitoly její znalost ale nevyžadujeme, čtenář si bohatě vystačí se znalostmi abstraktní algebry, viz například [54]. Budeme postupovat volně dle [63], nicméně další vhodný úvodní materiál se nachází na [17]. 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 zjednodušeně řečeno 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.

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

Představme si v $\mathbb{R}^3$ množinu M všech přímek procházejících počátkem a množinu N všech rovin procházejících počátkem. Každé dvě různé přímky z M určují jedinou rovinu z N a naopak každé dvě různé roviny se protínají v jedné přímce z M. Nyní uvažme rovinu například $z = 1$, každá přímka z M, která s ní není rovnoběžná, ji protíná v jednom bodě a každá rovina z N, která s ní není rovnoběžná, ji protíná v jedné přímce.

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. Čtenář je možná obeznámen s *Weirstrassovým tvarem* eliptické křivky $y^2 = x^3 + ax + b$ pro $x, y \in K$, ten však nekreslí celou situaci. Často se eliptické křivky definují jako nesingulární projektivní křivky genu 1 v $\overline{K}^3$, tj. jako množinu bodů $(X : Y : Z)$ splňujících:

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

s koeficienty $a_i \in K$. Pro naše účely si definici zúžíme a práci podstatně zjednodušíme. Konkrétně se budeme pohybovat nad tělesy, jejichž charakteristika není 2 ani 3. Tato tělesa často nabízí praktické výhody, my je však vynecháme. 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 $a, b \in K$. Libovolná taková rovnice udává eliptickou křivku za podmínky, že takzvaný *diskriminant* této křivky, $4a^3 + 27b^2$, je nenulový. Tato skutečnost je ekvivalentní s faktem, že eliptická křivka je nesingulární, přičemž lineární transformace proměnných zachovají (ne)singularitu křivky. Geometricky lze tuto podmínku interpretovat tak, že křivka nemá „hrot“.

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:

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

Značení 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.

Značení 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 $q + 1$.

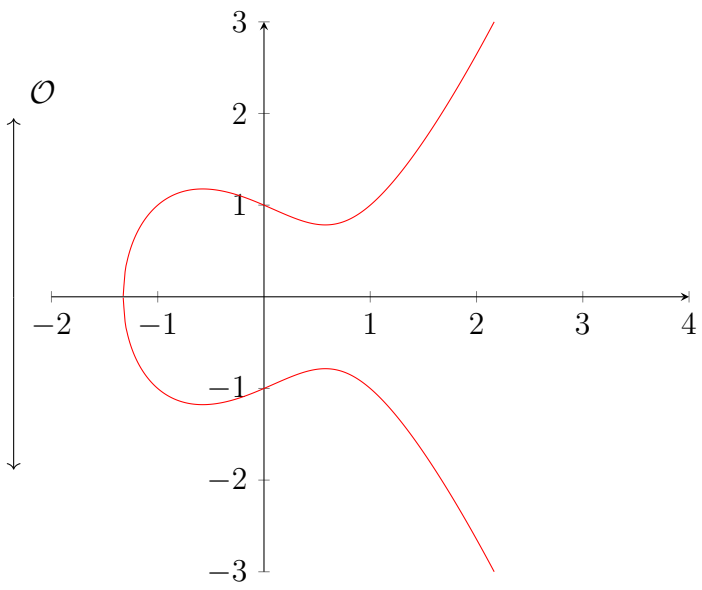
Věta 1.1.5. *(Hasse) Nechť $E/\mathbb{F}_q$ je eliptická křivka. Pak:*

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

Důkaz je k nalezení v [57, 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. Na naší cestě se přesto setkáme s místy, kde uvidíme taký či onaký způsob pohledu na problém poskytující řešení.

Všimněme si, že rozlišujeme body na eliptické křivce definované nad daným tělesem. Jako body na samotné křivce E/K nebudeme, jak by se na první pohled mohlo zdát, brát pouze body definované nad K, nýbrž nad celým uzávěrem, abychom zachytili celou její strukturu.

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



Eliptická křivka $y^2 = x^3 - x + 1$ nad $\mathbb{R}$.

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 $\mathcal{O}$.

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:

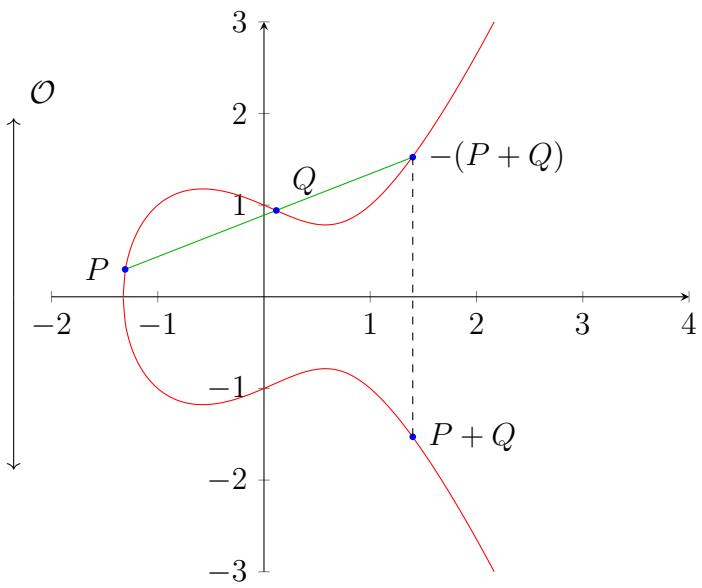
(i) $P + Q = Q + P$,

(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 [67, 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 skálární 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:

$$\begin{array}{lccccccccc}
 & \mathcal{O} & \longrightarrow & P & \longrightarrow & [2]P & \longrightarrow & [4]P & \longrightarrow & [8]P & \longrightarrow & [16]P & \longrightarrow & [32]P \\
Q: & \mathcal{O} & \longrightarrow & \mathcal{O} & \longrightarrow & [2]P & \longrightarrow & [2]P & \longrightarrow & [2]P & \longrightarrow & [18]P & \longrightarrow & [50]P \\
 & & & & & +[2]P & & & & & & +[16]P & & +[32]P
\end{array}$$

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.

Při tomto všem počítání přesto musíme (s případnou počítačovou asistencí) zatnout zuby a ony výrazy přesně spočíst.

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 $[4]Q = \mathcal{O}$. 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* na 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.

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 [67, 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 a x^2y, 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 grupami dokonce isomorfismus.

Isomorfismy nemusí nutně 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 $K(\sqrt{d})$. Křivku 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 i pro tzv. *mřížky* (lattice), viz [63, Def. 16.2]. j-invariant těchto struktur je spojen s tzv. *Laurentovou expanzí*, která je po násobení 1728 vždy celočíselná, viz [15, Ch. 11]. Poznamenejme též, že Weierstrassův tvar není jediný možný vyjadřující eliptickou křivku, existují 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 $\operatorname{char} K \notin \{2, 3\}$ 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 $\phi : E_1(\overline{K}) \longrightarrow E_2(\overline{K})$ tvaru $\phi : (x : y : z) \longmapsto (u(x,y,z) : v(x,y,z) : w(x,y,z))$ pro polynomy $u, v, w \in K[x]$ nazveme *isogenií*.

Dá se ukázat, viz [29, II.6.8.] a [57, III.4.8.], že nekonstantní zobrazení mezi eliptickými křivkami dané polynomy nad K je surjektivní 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. Ekvivalentně naši isogenii můžeme brát jako zobrazení:

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

pro u, v lomené funkce nad K, které navíc zachová bod v nekonečnu. Každý bod (x, y), který je v u či v nedefinovaný, je zobrazen do nekonečna.

Isogenie ale můžeme obecně zapsat mnohem kompaktněji:

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

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

pro u, v lomené funkce nad K a všechny body $(x, y) \in E_1$, které se nezobrazí do nekonečna.

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 $v(x,y) = \frac{g_2(x)}{g_3(x)}y$ pro $g_2, g_3 \in K[x]$. □

Definice 1.3.3. Buď $\phi : E_1 \longrightarrow E_2$ isogenie. Pod *standardním tvarem* ϕ rozumíme vyjádření $\phi(x,y) = \left(\frac{u(x)}{v(x)}, \frac{r(x)}{s(x)}y\right)$, kde $u, v \in K[x]$ a $r, s \in K[x]$ jsou dvojice nesoudělných polynomů, pokud je takový výraz definovaný, a $\phi(x,y) = \mathcal{O}$ jinak.

Díky této charakterizaci můžeme začít s isogeniemi pořádně pracovat. Nyní již nebude překvapením se zabývat otázkou, které body se zobrazí do nekonečna. Zprvu vidíme, že do

nekonečna se zobrazí body s x-ovou souřadnicí kořenem v, s. Tyto polynomy mají navíc stejnou množinu kořenů, právě protože bod $\mathcal{O}$ je isogenií zachován.

Nás zajímají pouze eliptické křivky nad konečnými tělesy a každý polynom nad konečným tělesem má pouze konečně mnoho kořenů, množina bodů zobrazených do nekonečna isogenií je konečná. Tyto body opět tvoří podgrupu $E_1(\overline{K})$, protože isogenie jsou homomorfismy grup bodů na křivkách.

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 $\# \ker \phi$.

Propůjčme si i další terminologii zaobývající se lomenými funkcemi, abychom isogenie mohli přesněji popisovat.

Definice 1.3.5. Pod *stupněm* isogenie ϕ budeme rozumět jejímu stupni jako lomené funkci v x a značit $\deg \phi$. Definujeme $\deg[0] = 0$.

Značení 1.3.6. Skládání, resp. sčítání isogenií definujeme následovně: pro libovolné isogenie $\phi : E \longrightarrow E_1$ a $\psi : E_1 \longrightarrow E_2$ definujeme $\psi \circ \phi := \psi(\phi)$ a pro isogenie $\phi, \psi : E \longrightarrow E_1$ zase $(\phi + \psi)P := \phi(P) + \psi(P)$ pro každý bod $P \in E_1$. Značme též isogenii opačnou jako: $-\phi = [-1] \circ \phi$.

Všimněme si, že složení dvou isogenií je zjevně opět isogenií. Všechny vlastnosti stupňů racionálních funkcí jsou u stupňů isogenií zachovány, zejména jejich multiplikativita.

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, druhý případ dokonce dává 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 $\mathbb{F}_{101}$.

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 1.3.7. 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 isogenií nad $\mathbb{F}_q$, tj.:

$$\pi_{E'} \circ \phi = \phi \circ \pi_E,$$

kde $\phi : E \longrightarrow E'$ je isogenie. 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 $\pi_p : (x, y) \mapsto (x^p, y^p)$ 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 1.3.8. *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 [57, Thm. III.6.1.], trochu elementárnější přístup se nachází v [63, 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 1.3.9. *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 [57, Thm. III.6.1]. □

Lemma 1.3.10. *Platí:*

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

Důkaz. Zjevně $\widehat{[0]} = [0]$ a $\widehat{[1]} = [1]$. Za pomocí věty 1.3.9, *(iv)*, máme pro každé celé n:

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

standardní oboustranný indukční argument pak dokončí první část. Z definice sčítání máme $[m] \circ [n] = [mn]$, tudíž $[n] \circ \widehat{[n]} = [n^2]$. Dle věty 1.3.9, *(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 $\ker \phi \subseteq E[n]$, neboť libovolný prvek v jádře ϕ se skrz $\hat{\phi}$ 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 případ isogenií definovaných nad tělesem $\mathbb{F}_q$ úzce souvisí s počtem bodů ležících na křivce. Samo krietérium zní až překvapivě jednoduše:

Věta 1.3.11. *(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, [65]. □

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 endomorfismus je isogenie neseparabilní i čistě neseparabilní. Charakterizujme dále separabilní isogenie.

Věta 1.4.2. *Ať E, E' jsou eliptické křivky nad K a $\phi : E \longrightarrow E'$ je isogenie daná standardní formou $(x, y) \mapsto \left(\frac{u(x)}{v(x)}, \frac{r(x)}{s(x)}y\right)$. 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 nemá jádro velikosti $\deg u/v$, ať už $p > 0$ či ne.

Uvažme nyní (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 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 každou pro neseparabilní isogenii ϕ existuje separabilní isogenie ψ s $\phi = \psi \circ \pi$. Pro x-ovou souřadnici tento výsledek známe, zbytek důkazu se dá najít na [63, Lemma 6.3.]. Tento důkaz není nijak zvlášť instruktivní, zde ho proto vynecháváme. Iterací tohoto faktu a skutečností, že Frobenius komutuje s libovolnou isogenií nad $\mathbb{F}_q$, pak získáme výsledek. □

Separabilní isogenie, jako takové, zatím nevypadají příliš zajímavě. Mají ale jednu vlastnost úzce spojenou s jejich jádrem, která je pro naši práci natolik stěžejní, že bez jejího zmínění by text byl poloviční.

Věta 1.4.4. *Buďte E eliptická křivka a $\phi : E \longrightarrow E'$ libovolná separabilní isogenie s jádrem $G \subseteq E(\overline{K})$. Pak všechny křivky E' jsou spolu isomorfní.*

Důkaz tvrzení je uveden v [67, Prop. 12.12], nicméně autor jej zde podává s notnou dávkou Galoisovy teorie, jejíž znalost od čtenáře nepředpokládáme.

Značení 1.4.5. Buď $G \subseteq E(\overline{\mathbb{F}}_q)$ konečná grupa. Značme E/G až na isomorfismus unikátní křivku, která pro každou separabilní isogenii $\phi : E \longrightarrow E'$ s jádrem G splňuje $E' \cong E/G$.

Poznámka. Ač E/G je pouze značení pro křivku a nesmí být naivně bráno ve smyslu faktorizace, není zcela nepodložené, ve zkratce zde načrtněme důvod. Každá konečná podgrupa $G \subseteq E(\overline{K})$ definuje surjektivní homomorfismus grup $\phi : E \longrightarrow E/G$ s jádrem G, kde E/G je isomorfní faktorgrupě $E(\overline{K})/G$. 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 [63, Thm. 6.10.].

Jednoznačnost (až na isomorfismus) cílové křivky separabilní isogenie má kolosální dopady na naše pochopení isogenií. Říká nám totiž, že separabilní isogenie můžeme uvažovat ne mezi přímo eliptickými křivkami, ale mezi jejich j-invarianty, což je jedna z klíčových vlastností vedoucí na praktické prokoly užívající isogenií.

Separabilní isogenie z $E \longrightarrow E'$ je daná lomenými funkcemi nad K a známe-li její jádro, dokážeme ji explicitně spočíst, přičemž libovolná konečná 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 [66]. 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 [16, Ch. 8.2]. V Sage 9.0 jsou Véluovy vzorce implementovány pro isogenii z E s jádrem G s časovou složitostí $O(\#G)$ příkazem:

`EllipticCurveIsogeny(E,ker G)`.

Příklad 1.4.6. Spočtěme separabilní isogenii ϕ s doménou eliptickou křivkou $E/\mathbb{F}_{101}$: $y^2 = x^3 + 8x + 23$ a jádrém cyklickou grupou generovanou bodem $P = (68, 9) \in E$. Bod P má řád 4 a grupa $\langle P\rangle = \{P, [2]P, [3]P, \mathcal{O}\} = \{(68, 9), (29, 0), (68, 92), \mathcal{O}\}$ je tedy jádrem ϕ. Příkaz `phi = EllipticCurveIsogeny(E,P)` v Sage 9.0 vygeneruje isogenii ϕ a tu určíme s pomocí Véluových formulí příkazem `phi.rational_maps()`:

$$\phi : (x, y) \longmapsto \left(\frac{x^4 + 37x^3 - 26x^2 - 15x - 21}{x^3 + 37x^2 - 17x + 32}, \frac{x^5 + 41x^4 - 9x^3 + 5x^2 + 3x - 7}{x^5 + 41x^4 - 18x^3 + 6x^2 + 35x - 21} y \right).$$

Příkaz `phi.codomain()` dává cílovou křivku ϕ spočtenou pomocí Vélouvých formulí a je to $E'/\mathbb{F}_{101} : y^2 = x^3 + 53x + 41$, samozřejmě všechny křivky s ní isomorfní jsou doménou eliptické křivky s jádrem $\langle P\rangle$. Kořeny polynomu $x^5 + 41x^4 - 18x^3 + 6x^2 + 35x - 21$ nad $\mathbb{F}_{101}$ jsou pouze 29 a 68, přičemž 29 dvojnásobný a 68 trojnásobný, což odpovídá faktu, že grupa $\langle P\rangle$ se zobrazí do nekonečna. V době psaní této práce je Sage schopen spočíst pouze isogenie s cyklickým jádrem a pomocí Véluových formulí.

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

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 1.4.4 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 [57, Thm. V.1.1.].

Konečně, pojďme se pokusit spočítat separabilní isogenie efektivněji. Je-li velikost jádra této isogenie prvočíselné (a tedy jádro cyklické), nespočteme ji jistě v čase lepším než lineárním vzhledem k velikosti jádra. Pokud ale pracujeme jádrem *hladké* velikosti, tedy dělitelné pouze prvočísly do dané hranice, můžeme postupovat mnohem rychleji.

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 zní hezky z čiře teoretického pohledu studia křivek, je ale hlavní ingrediencí v rychlejším počítání (separabilních) isogenií. Označíme $\langle G \rangle$ podgrupu $E(\overline{K})$ generovanou množinou $G = \{P, Q, R, \dots\}$ a pojďme se pokusit efektivně spočíst isogenii $\phi : E \longrightarrow E/\langle G \rangle$. Postačí nám spočíst separabilní isogenii $\psi : E \longrightarrow E/\langle P \rangle$, kde P má prvočíselný řád, pro podgrupy generované $Q, R, \dots$ spočteme analogicky separabilní isogenie převádějící $E/\langle P \rangle := E' \longrightarrow E'/\langle Q \rangle := E'' \longrightarrow E''/\langle R \rangle \dots$. Věta 1.4.4 nám zaručí, že složení všech takových isogenií bude mít jádro $\langle G \rangle$.

Ale isogenii $E \longrightarrow E/\langle P \rangle$ spočteme jednoduše:

$$E \xrightarrow{\phi_1} E/\langle \ell^{a-1} P \rangle \xrightarrow{\phi_2} E/\langle \ell^{a-2} \phi_1(P) \rangle \xrightarrow{\phi_3} \cdots \xrightarrow{\phi_a} E/\langle \phi_{a-1} \circ \phi_{a-2} \circ \cdots \circ \phi_1(P) \rangle,$$

kde řád P je ℓ^a. Snadno nahlédneme, že jádro $\phi_i \circ \cdots \circ \phi_1$ je $\langle \ell^{a-i} P \rangle$ a tedy separabilní isogenie daná složením všech ϕ_i má jádro přesně $\langle P \rangle$.

Véluovy formule nám umožní spočíst každou ϕ_i spočíst v $O(\ell)$ operacích a tedy celý proces je hotov pouze v $O(\ell a)$ operacích. Celou isogenii $E \longrightarrow E/\langle G \rangle$ takto spočteme v logaritmickém čase vzhledem k velikosti jádra.

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 (různými!)

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 právě dvě hodnoty y (krom $\mathcal{O}$ se 2 a 3-torze neprotínají) 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 argumenty implicitně závisí na faktu, že q není mocnina 2 ani 3, jinak naše eliptická křivka nemá tvar, který jí připisujeme. Tento případ je podrobněji rozebírán v [67, 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 isogenie.

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 dokázat. □

Zásadní rozdíl nastává při násobení mocninou charakteristiky 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*.

Znalost struktury ℓ-torzí pro ℓ prvočíslo nám pomůže spočítat, kolik separabilních isogenií prvočíselného stupně vychází z dané křivky. K tomu si nejprve pochopitelně musíme spočítat podgrupy na E řádu ℓ. Ty musí být generované bodem řádu ℓ, tedy celá pogrupa leží v $E[\ell]$. Přirozeně tedy chceme spočítat podgrupy ℓ-torze řádu ℓ. Případ $\ell = p$ dává buď žádnou či jednu podgrupu, v závislosti na supersingularitě křivky, dále tento případ neuvažujme.

Lemma 1.5.6. *Buď $E/\mathbb{F}_q$ křivka s $q = p^k$ a $\ell \neq p$ prvočíslo. Pak $E[\ell^e]$ obsahuje právě $\ell^{e-1}(\ell + 1)$ podgrup řádu ℓ.*

Důkaz. Díky větě 1.5.2 platí $E[\ell] \cong \mathbb{Z}_\ell \times \mathbb{Z}_\ell$. Každá podgrupa řádu ℓ je generovaná jediným prvkem řádu ℓ a naopak každý prvek $E[\ell]$ řádu ℓ generuje nějakou podgrupu řádu ℓ. Pokud si označíme G_1, G_2 dva generátory této křivky, pak bod $P = [m]G_1 + [n]G_2 \in E[\ell]$ má řád ℓ, právě pokud obě m, n nejsou nulová, neboli je afinní. Takových bodů je tedy $\ell^2 - 1$, přičemž každá podgrupu $G \subseteq E[\ell]$ řádu ℓ je generovaná libovolným jejím afinním prvkem, tj. každou podgrupu G zastupuje $\ell - 1$ bodů. Celkový počet hledaných grup je tudíž $\frac{\ell^2-1}{\ell-1} = \ell + 1$. □

Důsledek 1.5.7. *Buď $E/\mathbb{F}_q$ křivka s $q = p^k$ a $\ell \neq p$ prvočíslo. Pak existuje přesně $\ell + 1$ separabilních isogenií stupně ℓ vycházejících z E definovaných nad $\overline{\mathbb{F}}_q$.*

Důkaz. Podle věty 1.4.4 je počet isogenií vycházejících z E stupně ℓ dán počtem podgrup E řádu ℓ. Všechny takové grupy musí být obsaženy v $E[\ell]$ a předchozí lemma pak tvrdí, že hledaný počet je právě $\ell + 1$. □

Jak jsme zmínili před chvíli, pro nesoudělná m, n platí $E[m] \times E[n] \cong E[mn]$, tedy pomocí před chvílí zmíněného 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 1.5.8. *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. □

Působení isogenie na libovolnou m-torzi či samotnou grupu $E(\mathbb{F}_q)$ je jednoznačně určeno její akcí na (nejvýše dva) generátory těchto grup. Isogenie jsou totiž homomorfismy grup bodů na křivkách, pro příslušné generátory G_1, G_2 a bod $P = [m]G_1 + [n]G_2$ platí:

$$\phi([m]G_1 + [n]G_2) = [m]\phi(G_1) + [n]\phi(G_2).$$

Isogenie tedy působí na $E(\mathbb{F}_q)$ i na její torzní podgrupy jako 2×2 celočíselné matice, v případě m-torzní grupy dokonce jako matice modulo m. Jak se chovají isogenie na torzích budeme podrobněji studovat ve čtvrté kapitole.

Před chvíli jsme ale eliptické křivky rozlišily na dvě třídy podle jejich p-torze. Ty „neobyčejné“ z nich, supersingulární, jsou více než zajímavé.

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 některé 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 1.3.9 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 $E \longrightarrow E$. 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 [63, 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 $rM - Is$, 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 [58], 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 nějaký 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 j-invarianty rozřadit na obyčejné nebo supersingulární 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 křivky jim náležící jsou isogenní, a s isogeniemi prvočíselného stupně ℓ, získáme neorientovaný(!) $\ell + 1$-regulární (díky větě 1.5.7) graf rozdělený na obyčejné a supersingulární komponenty. Supersingulární křivky dokonce tvoří jednu jedinou souvislou komponentu, viz [35, Cor. 78]. Ve čtvrté kapitole

budeme tyto graf studovat trochu podrobněji studovat a ukážeme, že pokud se zaměříme na isogenie definované pouze nad $\mathbb{F}_q$, grafy supersingulárních j-invariantů se zásadně liší od grafů těch obyčejných. Z každého vrcholu totiž vždy vede buď $0, 1, 2$ či $\ell+1$ hran, přičemž supersingulární komponenty jsou pořád $\ell+1$ regulární, zatímco komponenty obyčejné tvoří tzv. *vulkány*, kde regulární graf stupně nejvýše 2 slouží jako „kráter“ a každý jiný vrchol je buď listem, či má $\ell+1$ sousedů. Název vulkány dává mnohem větší smysl při pohledu na konkrétní grafy:

– graf obyčejných a supersingulárních –

– příklad supersingulárního grafu nad F19 –

Na grafu isogenií nad $\mathbb{F}_{19}$ výše se nachází pouze 2 supersingulární vrcholy, což potvrzuje fakt, že tyto křivky se nachází v menšině. Tento trend se drží i pro vyšší rozšíření, nad celým uzávěrem $\mathbb{F}_p$ se nachází relativně málo supersingulárních j-invariantů.

Věta 1.6.7. *Označme S množinu všech supersingulárních j-invariantů nad $\overline{\mathbb{F}}_p$. Pak platí:*

$$\#S = \left\lfloor \frac{p}{12} \right\rfloor + \begin{cases} 0, & \text{pokud} \quad p \equiv 1 \pmod{12}, \\ 1, & \text{pokud} \quad p \equiv 5, 7 \pmod{12}, \\ 2, & \text{pokud} \quad p \equiv 11 \pmod{12}. \end{cases}$$

Důkaz se nachází na [67, Cor. 4.40].

Předchozí věta nám říká, že když se budeme přesouvat do vyšších rozšíření tělesa $\mathbb{F}_p$, nenarazíme na další a další supersingulární j-invarianty. Dokonce se zastavíme už na $\mathbb{F}_{p^2}$:

Věta 1.6.8. *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, $[p] = \iota \circ \pi^2$. 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. 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 $\mathbb{F}_{p^2}$. □

Důsledek 1.6.9. *Buď $E/\mathbb{F}_q$ supersingulární křivka. Pak existuje supersingulární $E'/\mathbb{F}_{p^2}$, která je s E isomorfní.*

Důkaz. Protože $j := j(E)$ leží v $\mathbb{F}_{p^2}$, příklad vyhovující křivky nad $\mathbb{F}_{p^2}$ pro $j \neq 0, 1728$ dává křivka $E' : y^2 = x^3 + 3j(1728 - j)x + 2j(1728 - j)^2$ s $j(E') = j$, viz věta 1.2.6, a případy $j = 0, 1728$ jsou zřejmé. □

Při uvažování grafů supersingulárních j-invariantů se zajímáme vesměs pouze na třídy isomorfismů, postačí nám tedy uvažovat všechny křivky pouze nad $\mathbb{F}_{p^2}$.

Značení 1.6.10. Buďte p, ℓ prvočísla. Graf supersingulárních j-invariantů nad $\overline{\mathbb{F}}_p$ spojené isogeniemi stupně ℓ značme $G_\ell(\overline{\mathbb{F}}_p)$.

Kapitola 2

Uplatnění v kryptografii

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 [20] 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*. Ta je založena na problému *diskrétního logaritmu* prvku $a \in \mathbb{Z}_p^*$, který 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, kde g je její libovolný generátor. Protokol požaduje, aby nebyl diskrétní logaritmus na G 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.

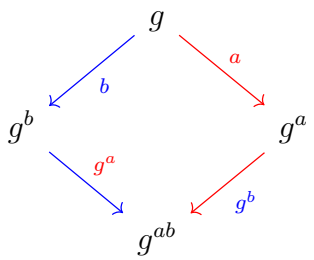
Poznámka. Polynomiální čas je zde pojem mírně zavádějící, nepožadujeme totiž algoritmus polynomiální v $|G|$, ale v $\log|G|$, totiž v jednotce místa, které G zabere při zápisu. Jednoduché procházení všech mocnin g proto běží v (očekávaném) exponenciálním čase.

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. Jeden ze způsobů, jak by mohla tajnou informací získat, spočívá ve výpočtu diskrétního logaritmu $\log_g(g^a) = a$, 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?

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

Klasický počítač operuje se základní jednotkou bit, což je diskrétní logický stav nabývající hodnot 0 a 1. Na n-bitovém počítači tak můžeme mít 2^n různých kombinací těchto stavů a v daný moment je počítač v právě jedné z těchto kombinací.

Ve světě kvantové mechaniky místo s klasickými bity pracujeme s *qubity*. Qubit narozdíl od jeho tradičního protějšku nenabývá pouze hodnot 0 a 1, ale jejich jednoktové lineární kombinace. Přesněji řečeno, stavy 0 a 1 můžeme ztotožnit s vektory $|0\rangle = \begin{bmatrix}1\\0\end{bmatrix}$, resp. $|1\rangle = \begin{bmatrix}0\\1\end{bmatrix}$ ve vektorovém prostoru $\mathbb{C}^2$, qubit je pak libovolná lineární kombinace těchto vektorů s jednotkovou normu, tj. $\alpha|0\rangle + \beta|1\rangle$ s $|\alpha|^2 + |\beta|^2 = 1$. Podobně můžeme uvažovat systém n qubitů jako množinu všech vektorů normy jedna náležících součinu vektorových prostorů $(\mathbb{C} \times \mathbb{C})^n = \mathbb{C}^{2n}$. Systém dvou qubitů tedy můžeme uvážit jako množinu všech lineárních kombinací ortogonálních vektorů $|00\rangle = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}$, $|01\rangle = \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$, $|10\rangle = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}$, $|11\rangle = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}$ v $\mathbb{C}^4$, kde součet druhých mocnin absolutních hodnot koeficientů bude roven jedné, takový stav nazveme tzv. *kvantovou superpozici* vektorů $|00\rangle, |01\rangle, |10\rangle, |11\rangle$. V obecnosti stavy systému s n qubity tvoří tzv. 2^n-dimenzionální *Hilbertův prostor*, jehož bázi tvoři klasické n-bity, viz [27].

Důležitou stránkou práce s kvantovým počítačem je, že nemůžeme jen tak kdykoliv zjistit stav našeho systému. Konkrétně uvážíme-li qubit ve stavu $\alpha|0\rangle + \beta|1\rangle$, nedokážeme zjistit hodnoty α a β, pokud bychom je předtím neznali, jinak řečeno, náš systém nemůžeme v libovolný čas přímo pozorovat. Místo toho musíme provést *pozorování*, které nám podá informaci ohledně systému v podobě klasického bitu. Po pozorování však již nemůžeme dále pracovat s původním stavem, celý se při pozorování totiž kolapsuje do pozorovaného klasického stavu. Jak? V případě pozorování stavu $\alpha|0\rangle + \beta|1\rangle$ získáme s pravděpodobností $|\alpha|^2$ hodnotu 0 a s pravděpodobností $|b|^2$ hodnotu 1. Obdobně koeficienty kvantového stavu určují pravděpodobnost, s jakou při pozorování získáme právě takový klasický stav. To je též důvod k normalizační podmínce $|\alpha|^2 + |\beta|^2 = 1$, součet všech pravděpodobností musí být roven 1. Kvantové algoritmy proto nestudujeme pouze z pohledu časové složitosti, ale i se snahou získat správný výsledek s co nejvyšší pravděpodobností.

Pojďme se nyní pobavit o tom, jak můžeme se systémem qubitů manipulovat. Systém našich qubitů můžeme samozřejmě vyjádřit ve vektorové podobě, přičemž tento vektor

je díky normalizační podmínce jednotkový. V klasických obvodech jsou bity ovlivňovány *branami*, které naše bity (ve vektorové podobě) transformují v jiné bity a zachovají normu příslušných vektorů, tj. jednotkové matice. Ku příkladu brána NOT bere pro vektory nad vektorovým prostorem dimenze 2 nad $\mathbb{Z}^2$ formu $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ a brána OR je zase reprezentována maticí $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. Kvantové brány jsou přirozeným zobecněním těch klasických, tedy v případě dvojqubitového systému jsou to právě jednotkové matice ve vektorovém prostoru stupně 2 nad $\mathbb{C}$. Kvantové obvody proto můžeme přirozeně studovat z pohledu lineární algebry.

Možná ale nejtěžší část práce s kvantovým počítačem je samotná jeho realizace. Běh kvantového počítače je velmi závislý na zdárné dualitě, ve které se každý qubit nachází, což klasická mechanika nepovoluje. Na konstrukci takového stroje se proto musíme ponořit do světa atomů a subatomárních částic, protože u takto malých částic lze pozorovat dualitu částic a vlnění. Ku příkladu elektromagnetické záření lze vnímat jako vlnění elektromagnetického pole, i jako smršť fotonů, která přenáší energii.

Jednou z největších překážek k překonání při praktické implementaci kvantového počítače je často velká nestabilita částic, se kterých pracujeme. V tomto ohledu se zda vhodným kandidátem duální charakter polarizace fotonu, kterou lze vnímat jako superpozici jeho levotočivé a pravotočivé polarizace. Fotony lze stabilně udržovat i při pokojové teplotě, přičemž jako „brány“ se nabízí křemíkové hranoly reflektující světlo. Takovou formu bere jeden z největších moderních kvantových počítačů, známý jako *Jiuzhang*, čítající až 76 qubitů.

Je zjevné, že klasické brány můžeme přirozeně převést do bran kvantových, a tedy libovolnou operaci, kterou provedeme na klasickém počítači provedeme na tom kvantovém, zde zmíněný model kvantového počítače je tedy alespoň stejný silné jako klasický Turingův model. Již v 80. letech minulého století, kdy užití kvantové mechaniky ve výpočetní technice bylo ve svých kojeneckých letech, byly objeveny četné procesy, které lze na kvantovém počítači uskutečnit mnohem rychleji, než na tom klasickém. Zmiňme, že jeden z nich, tzv. *Gaussovo vzorkování bosonů*, byl užit jako důkaz kvantové nádvlády pod návrhem vědců z české FJFI ČVUT [7].

I když jedna libovolná instance nezaručí perfektně přesný výsledek, existuje-li

Jedním z důvodů, proč se věří, že s veřejně dostupnými kvantovými počítači přijde nová éra výpočetní techniky je jejich zdárna superiorita nad klasickým modelem. Již v 80. letech minulého století, kdy užití kvantové mechaniky ve výpočetní technice bylo ve svých kojeneckých letech, byly objeveny procesy o kterých se dodnes neví, zda jsou v polynomiálním čase proveditelné na počítači klasickém, jejichž kvantové implementace běžící v očekávaném polynomiálním čase byly již nalezeny. I když žádná jedna instance nezaručí správný výsledek, v případě, že se bavíme o takovém exponenciálním zrychlením můžeme algoritmus spustit paralelně a porovnat jednotlivé výsledky.

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 je obdobou té klasické, pracující v očekávaném čase $O(\log(n)^2)$, poskytuje tedy exponenciální zrychlení, viz [8, Ch. 4. a 5.] pro více informací. Na oplátku k rychlosti kvantového počítače ale musíme, mimo nějaké speciální případy, zadat vstup jako superpozici všech jeho m qubitů, tj. 2^m čísel. Pro vydatnější kvantové počítače je toto číslo příliš velké a musíme se tedy spokojit pouze s velmi omezenými stavy na vstupu (například klasickými) a následnými operacemi získat kýžený stav. Toto vrozené omezení kvantových počítačů značně uzemňuje výkony, kterých s nimi můžeme dosáhnout.

Další problém notoricky klasicky obtížný je rozklad celého čísla na prvočinitele, který ve své nejčirejší podobě můžeme brát jako rozklad čísla n na součin dvou (velkých) prvočísel. Tento problém lze snadno převést na hledání řádu čísla a modulo n. V devadesátých letech minulého století přišel Peter Shor [60] 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 [53]. Tento objev pochopitelně způsobil paniku mezi kryptografickou komunitou, všechny nápady jenom vzdáleně spojené s diskrétním logaritmem musely být smeteny ze stolu.

Přestože pokrok lidstva v mnohých technologických objevech rostl povětšinou exponenciálně, u kvantových počítačů to tak jednoduché nebude. Problémy fyzické implementace architektury k času psaní této práce vyústili v největší kvantový počítač na světě, tzv. *Jiuzhang*, mající přesně 76 qubitů. I takto malé počítače běží řádově trilionkrát rychleji, nežli standardní počítač, a rekordy rozkládání čísel každých pár let padnou. Pro představu k roku 2012 bylo největší číslo rozložené kvantovým počítačem číslo 15, do konce roku 2019 se povedlo výzkumníkům z IBM [34] toto číslo zvýšit na $1099551473989 = 1048589 \times 1048601$.

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*. Oba běží v očekávaném subexponenciálním čase.

2.2 Vzhůru k eliptickým křivkám

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

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

Protokol ECDH pracuje velmi podobně jako originální Diffie-Hellmanova výměna. Oproti ní nebo RSA má však mnohem menší klíče, což je samozřejmě velmi žádoucí. Podobné

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

Alfréd		Blažena
Vstup: $a \in [1, p]$		**Vstup:** $b \in [1, p]$
	$\xrightarrow{[a]G}$	
	$\xleftarrow{[b]G}$	
$G_{AB} := [a]([b]G) = [a][b]G$		$G_{BA} := [b]([a]G) = [b][a]G$
Výstup: G_{AB}		**Výstup:** G_{BA}

Algoritmus 2: Protokol ECDH

vnitřní machinace obou protokolů přesto ECDH opět vystavují útoku via Schurův algoritmus. Ten totiž v polynomiálním čase nalezne periodu funkce $(a, b) \mapsto [a]P - [b]Q$ a tedy je schopen spočíst diskrétní logaritmus. O co víc, speciální případ diskrétního logaritmu na eliptických křivkách nabízí ještě rychlejší kvantové útoky [50]. Supersingulární křivky jsou v tomto ohledu dokonce o něco slabší, útok navržený v 90.letech [41] redukuje tuto úlohu za pomocí Weilova párování na problém diskrétního logaritmu v samotném konečném tělese.

Kvůli podobným útokům se kryptografie založena na supersingulárních křivkách v této éře (a skoro dvou desetiletích poté) nebrala v praktickém ohledu v potaz. Kdy se ale v této časové ose začalo uvažovat o isogeniích?

První kryptografické schéma založené na isogeniích obyčejných eliptických křivek navrhl Couveignes [14] již v roce 1997, nicméně svůj manuskript nepublikoval po dalších deset let. Grafy isogenií byly studovány přes přelom tisíciletí [23], [24]. Roku 2006 Rostovtsev a Stolbunov [56] nezávisle na Couveignovi navrhli (prakticky totožný) protokol založen na cestách v grafu obyčejných isogenií.

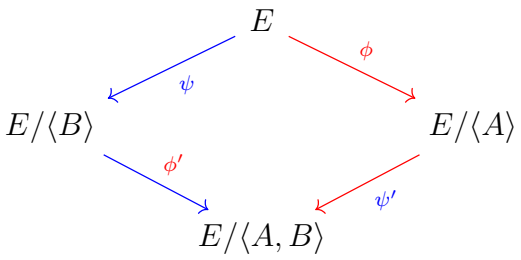
Bez přílišného noření se do detailů, tyto grafy jsou příliš řídké, a tedy existují subexponenciální algoritmy [25] na hledání isogenií mezi nimi, supersingulární grafy jsou ale mnohem hustší. Výměna zmíněna o odstavec výše byla tedy z diskuze vyřazena, ne ale na dlouho, ke konci práce budeme diskutovat jeho spirituálního následnika, protokol CSIDH. Konečně, v roce 2011 výzkumný tým z univerzity ve Waterloo navrhl praktickou výměnu užívající isogenie supersingulárních křivek, nesoucí název SIDH, vysloveno [sajd], - Supersingular Isogeny Diffie-Hellman [18].

2.3 SIDH

Celá výměna SIDH je založena na faktu, že separabilní isogenie je, až na isomorfismus, jednoznačně určena svým jádrem. Dejme tomu, že $\langle A\rangle, \langle B\rangle$ jsou dvě disjunktní grupy bodů ležících na křivce E.

Uvažme proto $\langle A\rangle, \langle B\rangle$ dvě disjunktní grupy bodů ležících na křivce E, složení separabilních isogenií $E \xrightarrow{\phi} E/\langle A\rangle \xrightarrow{\psi} (E/\langle A\rangle)/\langle B\rangle$, kde jádro ϕ je A a jádro ψ je $\phi(B)$, bude

separabilní isogenie $\xi : E \longrightarrow (E/\langle A\rangle)/\langle B\rangle$ s jádrem $\langle A, B\rangle$. Protože separabilní isogenie je (až na isomorfismus) jednoznačně určena, máme následující komutující diagram:



Tento jednoduchý diagram je srdcem samotného protokolu. Samozřejmě isogenie ϕ je separabilní s jádrem $\langle A\rangle$ a ψ' je separabilní isogenie taková, že $\psi' \circ \phi$ má jádro $\langle A, B\rangle$. V případě, že bychom aplikovali obě posloupnosti isogenií, nezískáme nutně na výstupu tu samou křivku $E/\langle A, B\rangle$, křivky se budou lišit až na isomorfismus. Jako sdílené tajemství tak může posloužit třída takových isomorfismů, respektive j-invariant této křivky. Nyní, když máme základní myšlenku principu, na kterém náš protokol pracuje, se pustíme do důležitých detailů, které samotnou výměnu usnadňuji či se vyhýbají známým útokům.

Uvažme výměnu, která ať proběhne mezi Alfrédem a Blaženou. Je dána supersingulární eliptická křivka E, kterou můžeme díky větě 1.6.9 bez ztráty bezpečnosti definovat nad $\mathbb{F}_{p^2}$ s nějakým dále specifikovaným prvočíslem p, protože nás u diagramu výše zajímají pouze třídy isomorfismů a ne nutně samotné křivky. Oba si vyberou tajné disjunktní podgrupy $A, B \subseteq E$ a poté spočtou isogenie zobrazující $E \longrightarrow E/A, E/B$. V případě obyčejných křivek isogenie definované $E \longrightarrow E$ spolu komutují (tedy nezáleží na pořadí jejich aplikace), což poskytuje poněkud přímočarý způsob spočíst isogenie ϕ', ψ' založený na našem diagramu, viz právě [56]. Pro supersingulární křivky tvoří isogenie $E \longrightarrow E$ složitější strukturu, tento problém se ale dá obejít, pokud se nezaměříme na isogenie definované nad celým uzávěrem, ale pouze nad $\mathbb{F}_q$, o tom ale více později.

Jeden z klíčových momentů v historii kryptografie založené na isogeniích nastal na přelomu tisíciletí [23],[24], kdy Galbraith et al. kostruují isogenie mezi obyčejnými křivkami v subexponenciálním čase v $\log p$ i na klasickém počítači, což jistě nevzbuzuje naději pro zabečení před kvantovou hrozbou. Toť tedy důvod, proč volíme E supersingulární. Musíme pak pro výměnu splňující náš diagram předat krom samotné křivky ještě špetku informací, kterou druhá partie užije k počítání příslušné isogenie. Pokud Alféd zveřejní obrazy generátorů G_1, G_2 grupy A v ϕ, Blažena spočte libovolnou podgrupu A generovanou bodem $P = [m]G_1 + [n]G_2$ jednoduše jako:

$$\phi_B(\langle P\rangle) = \phi_A(\langle [m_A]G_{1A}+[n_A]G_{2A}\rangle) = \langle \phi_A([m_A]G_{1A}+[n_A]G_{2A})\rangle = \langle [m_A]\phi_A G_{1A}+[n_A]\phi_A G_{2A}\rangle. \quad (\clubsuit)$$

Zveřejnění obrazů generátorů A v ϕ_A nás umožní spočíst obraz celé grupy A v ϕ_A, což může vést na efektivní útoky, to je ale oběť, kterou musíme podstoupit v uvažování supersingulárních křivek.

Konečně, abychom zaručili dostatečně velké podgrupy pro obě partie, můžeme zvolit $p = f\ell_A^{e_A}\ell_B^{e_A} - 1$ prvočíslo, kde $\ell_A^{e_A} \approx \ell_B^{e_A}$ jsou velké mocniny prvočísel a f je malé. Zvolíme si křivku E s $E(\mathbb{F}_{p^2}) \cong E[p+1] \cong E[f] \times E[\ell_A^{e_A}] \times E[\ell_B^{e_B}]$ obsahující dvě (přibližně stejně) velké podgrupy. To zaručíme například definováním E nad $\mathbb{F}_p$, kde pak díky větě 1.6.2 E nese $p+1$ bodů definovaných nad $\mathbb{F}_p$. Víme, že $E(\mathbb{F}_p)$ je podgrupou $E(\mathbb{F}_{p^2})$, tedy se příslušné počty bodů dělí. Hasseho věta značně omezuje počet bodů na $E(\mathbb{F}_{p^2})$ a pouze jeden z nich je dělitelný $p+1$, konkrétně $(p+1)^2$. Tento případ je tedy ekvivalentní s tím, že stopa Frobenia je nulová.

Alfréd si tedy vybere tajnou podgrupu $E[\ell_A^{e_A}]$ a celý protokol již zapadá do sebe.

Veřejné parametry: Prvočíslo $p = f\ell_A^{e_A}\ell_B^{e_B} - 1$, supersingulární eliptická křivka $E/\mathbb{F}_{p^2}$ splňující $E(\mathbb{F}_{p^2}) = (p+1)^2$, generátory $G_{1A}, G_{2A}, G_{1B}, G_{2B}$ po řadě $E[\ell_A^{e_A}]$, resp. $E[\ell_B^{e_B}]$

Alfréd		Blažena
Vstup: m_A, n_A nedělitelná p		**Vstup:** m_B, n_B nedělitelná p
spočte bod $A = [m_A]G_{1A} + [n_A]G_{2A}$		spočte bod $B = [m_B]G_{1B} + [n_B]G_{2B}$
spočte separabilní isogenii		spočte separabilní isogenii
$\phi_A : E \longrightarrow E_A$ s jádrem $\langle A \rangle$		$\phi_B : E \longrightarrow E_B$ s jádrem $\langle B \rangle$
spočte v ní obrazy G_{1B}, G_{2B}		spočte v ní obrazy G_{1A}, G_{2A}
	$\xrightarrow{E_A, \phi_A(G_{1B}), \phi_A(G_{2B})}$	
	$\xleftarrow{E_B, \phi_B(G_{1A}), \phi_B(G_{2A})}$	
spočte křivku $E_{AB} :=$		spočte křivku $E_{BA} :=$
$:= E_A/\langle m_A\phi_B(G_{1A}) + n_A\phi_B(G_{2A})\rangle$		$:= E_B/\langle m_B\phi_A(G_{1B}) + n_B\phi_A(G_{2B})\rangle$
Výstup: $j(E_{AB})$		**Výstup:** $j(E_{BA})$

Algoritmus 3: Protokol SIDH

Vysvětleme, co se vlastně v každém kroku děje u Alfréda, Blažena postupuje obdobně. Alfréd si na začátku výměny zvolí tajný bod A na $\ell_A^{e_A}$-torzi a spočte separabilní isogenii $\phi_A : E \longrightarrow E_A$ s jádrem rovným grupě $\langle A \rangle$, křivka E_A nese díky větě 1.3.11 $(p+1)^2$ bodů. Poté zveřejní obrazy generátorů Blaženiny $\ell_B^{e_B}$-torze a na oplátku obdrží E_B a body $\phi_B(G_{1A}), \phi_B(G_{2A})$. Obraz bodu A v Blaženině isogenii dokáže spočíst jako:

$$\phi_B(A) = \phi_B([m_A]G_{1A} + [n_A]G_{2A}) = [m_A]\phi_B(G_{1A}) + [n_A]\phi_B(G_{2A}),$$

neboť ϕ_A je homomorfismus grup $E \longrightarrow E_A$. Alfréd je pak schopen spočíst obraz celé grupy $\langle A \rangle$, jak už jsme si před chvíli ukázali u (♣). Separabilní isogenie $\phi_A' : E_A \longrightarrow E_A/\langle [m_A]\phi_B(G_{1A}) + [n_A]\phi_B(G_{2A})\rangle := E_{AB}$ má jádro $\langle B \rangle$, což znamená, že separabilní isogenie $\phi_A' \circ \phi_A : E \longrightarrow E_{AB}$ má jádro $\langle A, B \rangle$. Blažena analogicky ke konci protokolu spočítá separabilní isogenii $E \longrightarrow E_{BA}$ s jádrem $\langle A, B \rangle$. Podle věty 1.4.4 jsou křivky E_{AB} a E_{BA} navzájem isomorfní a tedy sdílí j-invariant, čímž uzavíráme výměnu.

Pojďme nyní vyřešit pár důležitých detailů při seřizování takové výměny ze strany veřejně důvěrné třetí osoby, která nastavuje veřejné parametry, i několik maličkostí ze strany Alfréda a Blaženy.

Zamysleme se nejprve, jak často narazíme na prvočíslo p, které hledáme. Konkrétně, pokud si zvolíme fixní mocniny $\ell_A^{e_A}, \ell_B^{e_B}$, s jakou pravděpodobností nalezneme „malé“ prvočíslo $p \equiv -1 \pmod{\ell_A^{e_A}\ell_B^{e_B}}$. Dirichletova věta o aritmetických posloupnostech, klasický výsledek analytické teorie čísel, nám poví, že takových prvočísel existuje nekonečně mnoho, tu však potřebujeme ještě trochu zesílit. Na pomoc nám přijde Nikolai Chebotarev a jeho věta o hustotě, kterou lze adaptovat na postačující odhad hustoty prvočísel v aritmetické posloupnosti [37]. Takové prvočíslo p jsme pak schopni pro nějaké mocniny prvočísel $\ell_A^{e_A}, \ell_B^{e_B}$ zaručeně najít.

Dále nastává problém najít počáteční křivku tak, aby potenciální útočník neměl při rozbíjení algoritmu přílišnou výhodu. Protokol založený na výměně SIDH, tzv. SIKE - Supersingular Isogeny Key Encapsulation, tento problém řeší jednoduchou volbou křivky $E : y^2 = x^3 + x$ pro $p \equiv -1 \pmod 4$, tedy například s $\ell_A = 2$ či $4 \mid f$. Volba křivky $E : y^2 = x^3 + x$, či obecně křivek s j-invarianty 0 a 1728 poskytuje Předem známá počáteční křivka ale může v některých případech značně ulehčit prolomení protokolu. Ke způsobu, jak se vyhnout možným trablím s fixní počáteční křivkou, se dostaneme o chvíli později u protokolu SITH.

Konečně generátory příslušných torzí spočteme jednoduše, postačí vzít generátory G_1, G_2 naší křivky, jejich násobky $[\ell_B^{e_B}]G_1, [\ell_B^{e_B}]G_2$ generují grupu $E[\ell_A^{e_A}]$.

Nyní již máme vyřešené aranžmá výměny ze strany nestranného prostředníka a můžeme se přesunout na naše účastníky, opět zaostříme na Alfréda. Po výběru bodu $A = [m_A]G_{1A} + [n_A]G_{2A}$ řádu $\ell_A^{t_A}$ si Alfréd chce spočíst isogenii $E \longrightarrow E/\langle A\rangle$ stupně $\ell_A^{t_A}$. Přímočará aplikace Véluových formulí je velmi pomalá a pro volbu $\ell_A^{e_A} \approx \ell_A^{e_B} \approx \sqrt{p}$ bychom průměrně očekávali $\ell_A^{t_A} \approx \sqrt{\ell_A^{e_A}}$ a tedy běžící čas $O(\sqrt[4]{p})$, což je exponenciální. Můžeme ale využít nápady spojené s rozkládáním isogenií na isogenie prvočíselných stupňů, viz věta 1.4.8. Konkrétně můžeme naši isogenii rozložit na t_A isogenií stupně ℓ_A a za pomocí Véluových formulí nyní celý výpočet končí po pouze $O(t_A\ell_A)$ operacích, speciálně se škáluje logaritmicky s rostoucí velikostí p.

Véluovy formule jako volba výpočtu všech isogenií též impaktují bezpečnost protokolu. Heuristicky můžeme ověřit, že křivky E_{AB}, E_{BA} získané na konci protokolu jsou ne jenom isomorfní, ale dokonce tou samou křivkou. Jak bylo podotknuto v článku [38], toto pozorování je při užití Véluových formulí realitou. Tato skutečnost poskytuje protokolu větší bezpečnost, protože místo pouhého j-invariantu můžeme jako sdílené tajemství považovat celou křivku, či nějaký její parametr, který nabývá mezi všemi křivkami dostatečně mnoha různých hodnot a dá se co nejlépe komprimovat.

V rámci projektu SOČ, kterého součást je tato práce, jsme v Sage 9.0 implementovali protokol SIDH, aby si čtenář mohl jeho běh vyzkoušet z první ruky. Na odkazu se nachází mirror naší implementace a samotný kód je k nalezení na https://github.com/zdenekpezlar/isogenie/tree/SIDH-protocol.

2.4 Útoky na SIDH

Pojďme se zběžně pobavit o možnostech, jak nad protokolem SIDH vyzrát. Budeme uvažovat prvočíslo p dostatečně vysoké, aby řešení hrubou silou nebyla možností. Nejprve zmiňme, že i přes pojmenování tohoto protokolu má bezpečnost SIDH pramálo společného s tou Elliptic Curve Diffie-Hellmanovy, oba jsou založené na velmi odlišných principech a proto není žádný důvod očekávat, že bezpečnost jedné ovlivňuje druhou. Dokonce, jak jsme zmínili, diskrétní logaritmus na supersingulárních křivkách je ještě jednodušší, než na obyčejných, u isogenií tedy nastává (z naší momentální znalosti) přesně opačná situace jako u diskrétního logaritmu.

Ve své podstatě je problém prolomení SIDHu ekvivalentní s nalezením isogenie stupně $\ell_A^{e_A}$ mezi známými křivkami E, E_A, spolu s trochu extra informacemi o našich isogeniích. Nemůžeme ale najít jen tak nějakou isogenii mezi křivkami, musí to být Alfrédova isogenie, jinak nejsme schopni užít zveřejněné body k výpočtu křivky $E/\langle A, B\rangle$.

Pozapomeňme na chvíli na obrazy generátorů a zabývejme se pouze problémem hledání isogenie. Pokud si, jak jsme zmínili v poslední sekci minulé kapitoly, vyznačíme supersingulární j-invarianty nad $\mathbb{F}_{p^2}$ a propojíme je via isogenie stupně $\ell \in \{\ell_A, \ell_B\}$ mezi nimi vedoucí, získáme spojitý neorientovaný $\ell+1$-regulární graf $G_\ell(\overline{\mathbb{F}}_p)$ čítající přibližně $p/12$ vrcholů. Každý z účastníků si vybere pseudo-náhodnou procházku a poté podle cesty svého protějšku provede další cestu a oba narazí na křivky se stejným j-invariantem. Pokusíme se ukázat, že je nepravděpodobné, že existuje více cest mezi E, E_A délky $e_A = \log_{\ell_A} \ell_A^{e_A} \approx \log_{\ell_A} \sqrt{p}$. K tomu nám pomůže malé lemma:

Lemma. *Graf $G_\ell(\overline{\mathbb{F}}_p)$ supersingulárních isogenií nad $\mathbb{F}_{p^2}$ má průměr alespoň $O(\log p)$.*

Důkaz. Zvolme fixní vrchol V a uvažme všechny z něj vedoucí cesty délky n. Podle věty 1.5.6 vede v $G_\ell(\overline{\mathbb{F}}_p)$ z V cesta do nejvýše $\ell^{n-1}(\ell+1)$ jiných vrcholů a tento odhad je dokonce ostrý, pokud $G_\ell(\overline{\mathbb{F}}_p)$ neobsahuje cykly. Průměr d tohoto grafu pak musí splňovat nerovnost $\ell^{d-1}(\ell+1) \geqslant \lfloor \frac{p}{12} \rfloor + \varepsilon$, v opačném případě by existoval vrchol vzdálený od V na vzdálenost více než d. To ale znamená $d \in O(\log p)$. □

Pizer [49, Thm. 1.] navíc ukazuje, že $G_\ell(\overline{\mathbb{F}}_p)$ má průměr i shora ohraničen $2\log p + O(1)$. Graf $G_\ell(\overline{\mathbb{F}}_p)$ má tedy poměrně, ale ne příliš, krátký průměr. Pokud mezi vrcholy E, E_A nalezneme cestu délky $e_A < d/2$, s vysokou pravděpodobností je jediná a tedy bude reprezentovat isogenii zvolenou Alfrédem. Jak tedy naleznout cestu mezi E a E_A? Jistě můžeme jednoduše prohlédavat graf z E, případně E_A, do šířky. V nejhorším případě prohledáme všechny křivky E/G s G jádrem velikosti do ℓ^{e_A}, kterých je přibližně $\ell^{e_A} \approx \sqrt{p}$. Tento postup jistě není optimální, pojďme se na něj podívat trochu chytřeji.

Místo toho, abychom prohledávali pouze z jednoho z vrcholů, můžeme prohlédavat z obou a iniciovat tzv. *Meet in The Middle* attack. Při takovém útoky si z E vypíšeme všechny cesty délky $\lfloor \frac{e_A}{2} \rfloor$ a z E_A zase $\lceil \frac{e_A}{2} \rceil$, a budeme hledat v obou listech dvě isomorfní křivky. Oba seznamy pak budou mít přibližně stejnou velikost $\ell_A^{e_A/2} \approx \sqrt[4]{p}$. Problém hledání shody můžeme vyřešit v čase úměrném velikosti obou seznamů, pokud si sestavíme hashovací

tabulky příslušných křivek (resp. jejich j-invariantů) a budeme hledat shodu, tedy skončíme v očekávaném čase $O(\sqrt[4]{p})$.

Obecně tento problém můžeme parafrázovat jako problém hledání shody dvou funkcí $f : A \longrightarrow C, G : B \longrightarrow C$, kde A, B jsou množiny podgrup $E[\ell_A^{e_A}]$,$E[\ell_B^{e_B}]$ a C sestává z j-invariantů supersingulárních křivek nad $\mathbb{F}_{p^2}$, tzv. *claw finding*. Zmíněný postup s hashovací tabulkou problém řeší v $O(|A| + |B|)$, což je na klasickém počítači optimální **??**. Kvantový počítač nepřekvapivě opět nad klasickým triumfuje a problém řeší v očekávaném čase $O(\sqrt[3]{|A| \cdot |B|})$, tedy v případě grafů isogenií $O(\sqrt[6]{p})$, užitím Taniho algoritmu [64]. I tento postup je asymptoticky optimální bez uvažování dalších vlastností grafů isogenií.

Pokud tedy chceme efektivně prolomit SIDH, s útokem v polynomiálním čase, který s netriviální pravděpodobností vyustí ve společný klíč, musíme buď hlouběji studovat grafy isogenií, či uvažovat i obrazy generátorů torzí v jednotlivých isogeniích.

Analýza v [26, Sec. 4.2] ukazuje, že jsme schopni spočíst hledanou isogenii mezi E a E_A, známe-li strukturu *endomorfismů* na křivkách E a E_A, tedy isogeniím $E \longrightarrow E$, resp. $E_A \longrightarrow E_A$, dokážeme spočíst hledanou isogenii. Výpočet této struktury je na obyčejné křivce a priori proveditelný v subexponenciálním čase [2], což je z kryptografického hlediska nežádané, supersingulární křivky tuto strukturu mají velmi různou a útoku se vyhýbají. Přesto je důležité podotknout, že fixní počáteční křivka značně ulehčí práci útočníka, protože mu stačí spočíst množinu endomorfismů jednou.

Konečně, zvolme si na místě Evy trochu jiný plán útoku. Místo útoku zvenku na výměny mezi oběma partiemi, Eva infiltruje výměnu zevnitř. Dejme tomu, že Alfréd chce provést výměnu s Blaženou jako předtím, ale místo tentokrát bere Eva alias Blaženy, Alfréd netuše, že byl ošizen a komunikuje s nečestnou partií. Oba celou výměnu sehrají jako normálně a Eva se dozví nějakou informaci o Alfrédově isogenii ve formě obrazu $\langle B \rangle$ v ní, přičemž tuto grupu si může volit, jak si zamaní. Může se Eva opakováním výměny (Alfréd si ponechá svou tajnou isogenii) a šikovnými volbami svých grup dozvědět Alfrédovu isogenii? Právě takovou otázku si položili Galbraith, Petit, Shani a Ti v [26], a odpověď zní ano. V přibližně $\frac{1}{2}\log_2(p)$ výměnách zaručují výpočet celé isogenie, resp. tajných koeficientů m_A, n_A udávajících Alfrédův bod. Tento útok na SIDH „zevnitř“ je jedním z mála známých pracujících v polynomiálním čase, ať už v počtu výměn, a tvoří pro protokol reálnou hrozbu.

V obou právě zmíněných útocích hraje roli, že užívána nějaká křivka fixní, a je užívána znalost ohledně isogenií na nich. Chceme-li zaručit co nejlepší bezpečnost v duelu s těmito i se zatím nenalezenými útoky, je imperativní, že ze zveřejněné informace jsou nějakým způsobem zašifrovány, případně není dána fixní počáteční křivka.

2.5 Následníci výměny SIDH

Subexponenciální řešení problému hledání isogenie mezi obyčejnými křivkami po nějakou dobu zcela zastavilo snahu nalézt efektivní a bezpečná kryptografická primitiva založená na isogeniích, SIDH tomuto hledání znovu vdechl život v podstatě supersingulárních křivek. Od zveřejnění jeho návrhu uplynula k psaní této práce celá dekáda a za tu dobu mnoho

různých autorů s původním nápadem odeběhlo na všelijaká místa poskytující praktické vylepšení či (větší či menší) variaci na samém principu výměny. Zde dokumentujeme několik pár z těch nejprospektivnějších.

SIKE. [1] Náš první kandidát není tolik co následník SIDHu, jako jeho optimalizace. Oproti křivkám nad $\mathbb{F}_p$ je aritmetika křivek nad $\mathbb{F}_{p^2}$ velmi pomalá a SIKE se ji pokouší co nejvíce zrychlit. Místo křivek ve Weierstrassově tvaru pracuje s křivkami v tzv. *Montgomeryho tvaru*, kde každou křivku vyjádří ve tvaru $by^2 = x^3 + ax^2 + x$, kde $a, b \in \mathbb{F}_{p^2}$. Mimo jiné j-invariant této křivky závisí jen a pouze na a a vzorce pro dvou a třínásobek bodu jsou mnohem stravitelnější a umožňují rychlejší počítání. Z tohoto důvodu jsou též fixována $\ell_A = 2, \ell_B = 3$, což opět neposkytuje útočníkům nějakou podstatnou výhodu. Konečně je implentována komprese posílaných bodů pro ještě menší klíče a tzv. KEM - *Key Exchange Mechanism*, který poskytuje protokolu větší bezpečnost. Excelentní článek detailující tyto specifika je [13].

eSIDH. [5] Pravděpodobně nejvíce přímočará adaptace výměny SIDH spočívá ve změně prvočísla na tvar $p = f2^{e_A}\ell_B^{e_B}\ell_C^{e_C} - 1$, kde ℓ_B, ℓ_C jsou malá prvočísla a $2^{e_A} \approx \ell_B^{e_B}\ell_C^{e_C}$. Alfréd si tentokrát vybírá isogenie stupně dělícího 2^{e_A} a Blažena isogenie stupně dělícího $\ell_B^{e_B}\ell_C^{e_C}$. Tato varianta ???

BSIDH. [12] Tento protokol užívá faktu, že supersingulární křivky s $\#E(\mathbb{F}_p^2) = (p+1)^2$ mají kvadratický twist s počtem bodů $(p-1)^2$, viz 1.2.5. Jeden z účastníků pak pracuje na $(p+1)$-torzi křivky E a druhý na $(p-1)$-torzi jejího kvadratického twistu $\tilde{E}$, přičemž E a $\tilde{E}$ jsou isomorfní nad $\mathbb{F}_{p^4}$.

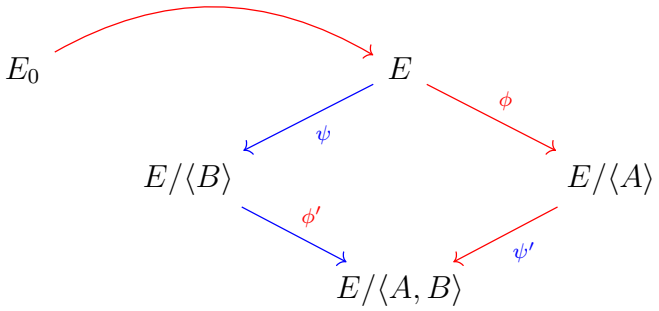
SITH. [3] Zde navržená výměna řeší problém známé počáteční křivky velmi jednoduše, Alfréd provede z křivky E náhodnou procházku v grafu isogenií stupně ℓ_A na E', ze které je poté výměna inicializována podobně jako v SIDHu.

CSIDH. [?] Náš poslední protokol nese název „Commutative Supersingular Isogeny Diffie-Hellman“, přesto není přímou adaptací ani vylepšením SIDHu. Jeho běh užívá hlubších znalostí ohledně eliptických křivek a jejich spojitosti s algebraickou teorií čísel, konkrétně pracuje na principu akce *grupy tříd ideálů* na grafu supersingulárních eliptických křivek. Pro tentokrát bereme pouze křivky definované nad $\mathbb{F}_p$ a zajímavě isogenie též, speciálně jako vrcholy bere třídy křivek isomorfní „nad $\mathbb{F}_p$“. Každá třída isomorfismů je proto reprezentovaná dvěma vrcholy příslušícími křivkám a jejich kvadratických twistům. V dalších kapitolách budeme studovat oblasti matematiky potřebné k pochopení tohoto protokolu.

2.6 SITH

Při statické počáteční křivce E_0 se protokol vystavuje většímu riziku prolomení, proto bychom chtěli při každé výměně zvolit novou křivku. Nestranný prostředník může ale jen zveřejnit přislušnou křivku a její parametry, tato úloha proto padá na samotné účastníky výměny. Protokol SITH - Supersingular Isogeny Two-Party Handsake diktuje, že na začátku výměny například Alfréd zvolí podgrupu $G \subseteq E_0[\ell_A]$ a spočte novou křivku $E = E_0/G$, zbytek výměny postupuje analogicky jako v SIDHu, pouze na křivce E.

Tato jednoduchá volba řeší mnoho útoků, které závisí na známé počáteční křivce. I když



protokol nyní není symetrický pro oba účastníky (například Alfréd musí poslat dohromady dvakrát tolik informací co Blažena), tento problem je snadno řešen během dvou protokolů proti sobě, jeden inicovaný Alfrédem, druhý Blaženou.

Obdobně jako v případě SIDHu jsme v rámci práce SOČ protokol SITH implenetovali v Sage 9.0, viz ?, na adrese https://github.com/zdenekpezlar/isogenie/tree/SITH-protocol se opět nachází kód.

Veřejné parametry: Prvočíslo $p = f\ell_A^{e_A}\ell_B^{e_B} - 1$, supersingulární eliptická křivka $E_0/\mathbb{F}_{p^2}$ splňující $E_0(\mathbb{F}_{p^2}) = (p+1)^2$, generátory $G_{1A}, G_{2A}, G_{1B}, G_{2B}$ po řadě $E_0[\ell_A^{e_A}]$, resp. $E_0[\ell_B^{e_B}]$

Alfréd		Blažena
Vstup: $p \nmid m_A, n_A$, grupa $G \subseteq E_0[\ell_A]$		**Vstup:** $p \nmid m_B, n_B$
spočte křivku $E = E_0/G$		
a její generátory G_1, G_2		
spočte generátory $G_{1A} = [\ell_A^{e_A}]G_1, \ldots$		
	$\xrightarrow{E, G_{1A}, G_{2A}, G_{1B}, G_{2B}}$	
spočte bod $A = [m_A]G_{1A} + [n_A]G_{2A}$		spočte bod $B = [m_B]G_{1B} + [n_B]G_{2B}$
spočte separabilní isogenii		spočte separabilní isogenii
$\phi_A : E \longrightarrow E'_A$ s jádrem $\langle A \rangle$		$\phi_B : E \longrightarrow E_B$ s jádrem $\langle B \rangle$
spočte v ní obrazy G_{1B}, G_{2B}		spočte v ní obrazy G_{1A}, G_{2A}
	$\xrightarrow{E_A, \phi_A(G_{1B}), \phi_A(G_{2B})}$	
	$\xleftarrow{E_B, \phi_B(G_{1A}), \phi_B(G_{2A})}$	
spočte křivku $E_{AB} :=$		spočte křivku $E_{BA} :=$
$:= E_A/\langle m_A\phi_B(G_{1A}) + n_A\phi_B(G_{2A})\rangle$		$:= E_B/\langle m_B\phi_A(G_{1B}) + n_B\phi_A(G_{2B})\rangle$
Výstup: $j(E_{AB})$		**Výstup:** $j(E_{BA})$

Algoritmus 3: Protokol SITH

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 [6, Ch. XIII], [30], [44] či [51]. Jako velmi stručný úvod motivovaný poznatky z elementární teorie čísel může též posloužit má SOČ, [48].

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. Pro jednoduchost se omezíme pouze na obory komutativní.

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

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 $g = 1$. Akci $\cdot$ též nazveme *tranzitivní*, pokud pro každou dvojici $(x, y) \in X^2$ existuje $g \in G$ splňující $g \cdot x = y$.

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

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

Definice 3.1.4. Modul M okruhu R nazveme *volným*, pokud obsahuje $R-$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 prvků $a \in G$, 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 $\mathbb{Z}[i]$, 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 1.5.3 rank 2.

Poznámka. Roku 1922 Luis Mordell v [43] 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 [68], 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, pokud je tato grupa konečná.

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 [51, Věta 1.3.8]. Pokud bychom však místo $\mathbb{Z}$ uvážili libovolný komutativní okruh R, tvrzení již ne nutně platí!

Příklad 3.1.9. Ukážme, že $\mathbb{Z}_6$-modul $2\mathbb{Z}_6 = \{0, 2, 4\}$ není volný. Pokud by totiž modul $2\mathbb{Z}_6$ měl nad $\mathbb{Z}_6$ bází, musely by její prvky nad $\mathbb{Z}_6$ být lineárně nezávislé. Nicméně platí $0 \cdot 3 = 2 \cdot 3 = 4 \cdot 3 = 0$, 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 univerzálním bilineárním zobrazením $V \times W \longrightarrow U$. 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 R-moduly. Uvažme prvky $m \in M, n \in N$ jednotlivých modulů. *Tenzorový součin* $m \otimes n$ definujeme jako výraz, který je na sčítání oboustranně distributivní a pro každé $r \in R$ splňuje:

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

Pak *tenzorový součin* modulů 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*.

Uveďme si jednoduchý příklad tenzorového součinu.

Příklad 3.1.11. Ukažme, že $\mathbb{Z}_m \otimes_{\mathbb{Z}} \mathbb{Z}_n = \{0\}$ pro nesoudělná celá m, n. 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 3.1.12. *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 zobecnit 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 [54, Věta 11.12]. Dokonce si takové θ můžeme zvolit celé algebraické, viz [47, 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 je prosté a dává isomorfismus mezi tělesem racionálních lomených funkcí a $\mathbb{Q}(\theta)$.

Pojďme si trochu charakterizovat celá algebraická čísla v číselném tělese.

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. □

Pro všechna algebraická čísla θ, která nejsou celá algebraická, tedy okruh $\mathbb{Z}[\theta]$ není konečně generovaný jako $\mathbb{Z}$-modul (a je navíc isomorfní s $\mathbb{Z}[x]$), stejně jako v případě θ, které není kořenem žádného polynomu nad racionálními čísly. 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-adických 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 $S = \{\alpha^i \beta^j | i \in \{0, 1, \ldots, n-1\}, j \in \{0, 1, \ldots, m-1\}\}$, kde minimální polynom β, P_β, má stupeň m. Protože K je oborem integrity, největší podmnožina S lineárně nezávislá nad $\mathbb{Q}$ tvoří bázi $\mathbb{Z}[\alpha, \beta]$. Okruh $\mathbb{Z}[\alpha, \beta]$ je proto volným $\mathbb{Z}$-modulem ranku nejvýše 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!

Lemma 3.2.8. *Buď θ libovolné nenulové algebraické číslo. Pak existuje nenulové celé m takové, že $m\theta$ je celé algebraické.*

Důkaz. Pokud minimální polynom θ nad celými čísly je:

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

kde vedoucí člen je nenulový, 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. □

Toto tvrzení je vše, co nám stačí k určení podílového tělesa $\mathcal{O}_K$. Pokud si představíme situaci nad $\mathbb{Z}$ či $\mathbb{Z}[i]$, jistě se dovtípíme, které těleso to bude.

Důsledek 3.2.9. *Číselné těleso K podílovým tělesem okruhu $\mathcal{O}_K$.*

Důkaz. 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 $K \subseteq L$. □

Známe okruh celých algebraických čísel těles $\mathbb{Q}$ a $\mathbb{Q}(i)$. V libovolném kvadratickém tělese však dokážeme $\mathcal{O}_K$ za pomocí znalosti řešení kvadratické rovnice jednoduše popsat též.

Věta 3.2.10. *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 kvadratických rovnic 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, ten je proto roven:

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

Pokud $\frac{a+b\sqrt{m}}{2} \in \mathcal{O}_K$, je tento monický polynom definovaný nad celými čísly. Proto a^2-b^2m 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. Okruh $\mathcal{O}_K$ v kvadratickém tělese $K = \mathbb{Q}(\sqrt{m})$ s m bezčtvercovým můžeme kompatněji vyjádřit jako $\mathbb{Z}\left[\frac{d+\sqrt{d}}{2}\right]$, kde $d = m$, pokud $m \equiv 1 \pmod 4$, a $4m$ jinak. Toto d se nazývá *diskriminant* číselného tělesa 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, konkrétně $\mathbb{Q}(\sqrt[3]{19})$, viz [11, Ex. 2.3.].

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

Definice 3.2.11. Okruh $\mathcal{O}$ obsažen v číselném tělese K nazveme *pořádkem*, pokud je volným $\mathbb{Z}$-modulem ranku $[K : \mathbb{Q}]$.

Nejprve si všimněme, že věta 3.2.10 říká, že okruh celých algebraických čísel kvadratického tělesa je pořádkem. Tuto vlastnost dokonce sdílí všechny okruhy $\mathcal{O}_K$ v číselném tělese K.

Věta 3.2.12. *Okruh $\mathcal{O}_K$ je pořádkem K.*

Důkaz. Ať K má stupeň n a uvažme $\{a_1, \ldots, a_n\}$ jeho bázi jako vektorového prostoru $K/\mathbb{Q}$. Nejprve dejme tomu, že $\mathcal{O}_K$ není konečně generovaný jako $\mathbb{Z}$-modul. Pak tento okruh a speciálně těleso K obsahují alespoň $n+1$ prvků lineárně nezávislých nad $\mathbb{Z}$, tedy i nad $\mathbb{Q}$. To je ale spor, protože K je číselné těleso stupně n. Okruh celých algebraických čísel je proto konečně generovaný a počet jeho generátorů je shora ohraničen číslem n.

Podle lemmatu 3.2.8 existují nenulová celá m_i taková, že každé z čísel $m_i a_i$ je celé algebraické. Protože a_i musela být navzájem lineárně nezávislá nad $\mathbb{Q}$, čísla $m_i a_i$ jsou lineárně

nezávislá nad $\mathbb{Z}$. Volný $\mathbb{Z}$-modul $\mathcal{O}_K$ proto obsahuje volný $\mathbb{Z}$-modul $\mathbb{Z}[m_1a_1, \ldots, m_na_n]$ ranku n a díky větě 3.1.8 sám má tedy rank roven n. □

Mezi pořádky má $\mathcal{O}_K$ speciální postavení, je totiž vzhledem k inkluzi největší.

Věta 3.2.13. *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 3.2.14. 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 3.2.15. *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:*

$$\mathcal{O}_K = \mathbb{Z}[a_1, a_2, \ldots, a_n], \qquad \mathcal{O} = \mathbb{Z}[k_1a_1, k_2a_2, \ldots, k_na_n].$$

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 $\mathcal{O} \subseteq \mathcal{O}'$ 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 3.2.16. *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 zásadních oblastech, ke kterým se budeme vracet. Pro jedno, pořádky nejsou nikdy celouzavřené nad jejich podílovým tělesem, každý prvek $\mathcal{O}_K \setminus \mathcal{O}$ je totiž celý nad $\mathbb{Z}$ a tedy i nad $\mathcal{O}$. Ku příkladu číslo $\frac{1+\sqrt{5}}{2}$ je celé nad pořádkem $\mathbb{Z}[\sqrt{5}] \subseteq \mathbb{Q}(\sqrt{5})$ a neleží v něm.

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 [33].

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 popíšeme jednodušeji. Víme, že L je jednoduché rozšíření $\mathbb{Q}(\theta)$ s bází $\{1, \theta, \ldots, \theta^{n-1}\}$, vzhledem ke které budeme psát $a(x)$. 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 lineárního zobrazení $\tau(x)$.

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 $b_0 + b_1 x + \cdots + b_{n-1}x^{n-1} + x^n$, 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.

Lemma 3.3.2. *Nechť $K = \mathbb{Q}(\theta)$ je číselné těleso stupně mn a $\tau \in K$ má minimální polynom stupně m. Pak charakteristický polynom $\tau(x)$ je n-tou mocninou minimálního polynomu τ nad $\mathbb{Q}$.*

Důkaz. Uvažme $\{1, \theta, \ldots, \theta^{mn-1}\}$ bázi K nad $\mathbb{Q}$ a $\{b_1, \ldots, b_m\}$ bázi K nad $\mathbb{Q}(\tau)$. Množina všech prvků $\theta^i b_j$ je zřejmě nad $\mathbb{Q}$ lineárně nezávislá a tvoří proto bázi prostoru $K/\mathbb{Q}$. Vzhledem k této bázi snadným porovnáním koeficientů zjistíme, že $\tau(x)$ působí na $\mathbb{Q}(\theta)$ jako blokově diagonální matice obsahující n matic:

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

kde $x^m + c_{m-1}x^{m-1} + \cdots + c_0$ je minimální polynom τ nad racionálními čísly. Charakteristický polynom τ je pak jeho minimální polynom umocněn na n-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 3.3.3. Buď K číselné těleso a τ jeho prvek. Pak definujeme jeho *stopu* $Tr(\tau)$ a *normu* $N(\tau)$ jako stopu, resp. determinant matice udávající $\tau(x)$:

$$Tr_K(\tau) := Tr M_\tau,$$
$$N_K(\tau) := \det M_\tau.$$

Norma i stopa prvků číselného tělesa jsou tedy racionální čísla. Podotkněme, že stopa i determinant matice nezávisí na konkrétní volbě báze, definice výše je proto korektní. Abychom se nezadusili 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.

Příklad 3.3.4. 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:

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

$$(a + b\sqrt{-2}) \cdot \sqrt{-2} = -2b + a\sqrt{-2},$$

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 3.3.5. *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 3.3.6. *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 3.3.7. *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 3.3.8. Prvek $a \in R$, který je v R invertibilní, nazveme *jednotkou*.

Definice 3.3.9. 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 [4, 2. díl] 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.

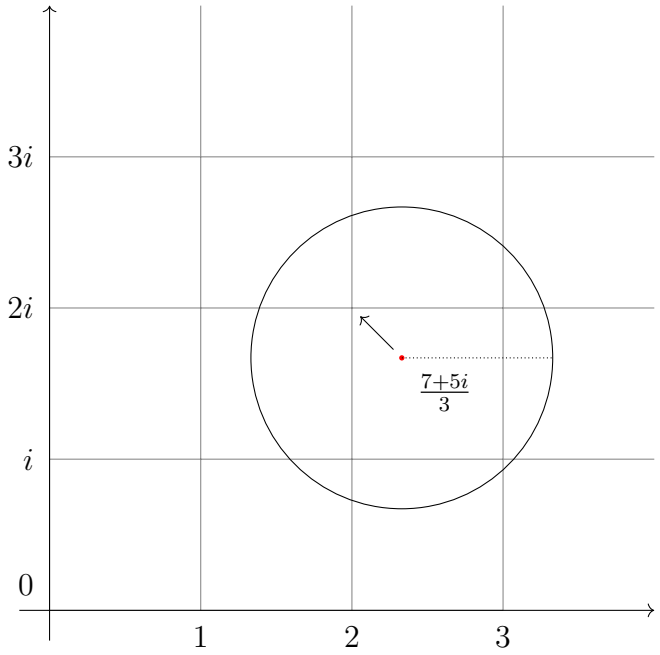
Od jednotek se přesuňme na zobecnění prvočísel, tzv. *ireducibilních* prvků, v okruzích.

Definice 3.3.10. Prvek $a \in R$ nazveme *ireducibilním*, nelze-li jej zapsat jako součin dvou prvků R, ani jeden z nichž není jednotka.

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 rozkládat čísla na ireducibilní prvky. Takový rozklad v obecném okruhu je $a \neq 1$ zjevně existuje, a díky multiplikativitě normy je pro nenulové prvky konečný, 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.



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 3.3.11. Okruh, ve kterém můžeme obdobně dělit se zbytkem, nazveme *euklidovým*.

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 3.3.12. *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 3.3.13. *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. Kvůli multiplikativitě normy je každý nenulový prvek $n \in R \setminus \{1\}$ rozložitelný na konečně mnoho činitelů. 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. Tyto rozklady jsou pak díky uvedeným normám 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, 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.

Věta 3.3.14. *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 [46], či v předloze oné práce [15], 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 $\mathfrak{a}$ komutativního okruhu R takovou, že $a \cdot r \in \mathfrak{a}$ platí pro $a \in \mathfrak{a}, r \in R$ označíme jako ideál.

O ideálech můžeme proto přemýšlet jako o (jediných neprádných) podmodulech R-modulu R.

Pokud $\mathfrak{a}$ je podgrupa R, tak faktorgrupa $R/\mathfrak{a}$ 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 G/H je grupa. Zobrazení $G \longrightarrow G/H$ přiřazující každému prvku G jeho příslušnou třídu v G/H je poté homomorfismus grup.

Každý ideál $\mathfrak{a} \subseteq R$ tedy definuje faktorokruh $R/\mathfrak{a}$, 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) $\mathfrak{a} \subseteq R$, 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. Nejprve se však pozastavíme u okruhů zbytků.

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, 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. Nulový ideál tvrzení zjevně splňuje, dále uvažme opak a nenulový 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 3.2.13 celé číslo. Pak $\mathcal{O}/\mathfrak{a}$ je podokruhem $\mathcal{O}/(N(x))$. Pokud $\{a_1, \ldots, a_n\}$ je báze

pořádku $\mathcal{O}$ jako $\mathbb{Z}$-modulu, libovolné číslo $z \in \mathcal{O}$ můžeme vyjádřit jako $t_1a_1 + \cdots + t_na_n$ pro celá t_i a leží ve stejné třídě jako $z' = t_1'a_1 + \cdots + t_n'a_n$, kde t_i' je zbytek, který t_i dává po dělení $N(x)$. Každé t_i' nabývá jednoho z $N(x)$ těchto zbytků, tedy v $\mathcal{O}/(N(x))$ leží nejvýše $N(x)^n$ prvků a tento okruh 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 $\mathfrak{a} \subseteq \mathcal{O}$ ideál. Pod normou $N_{\mathcal{O}}(\mathfrak{a})$ ideálu $\mathfrak{a}$ rozumíme indexu $\mathfrak{a}$ v $\mathcal{O}$.

Věta 3.4.5. *Každý stoupající řetězec inkluzí ideálů pořádku $\mathcal{O}$ je shora omezený.*

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 $\mathcal{O}$ 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 ideálů $\mathcal{O}_K$ 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. Dokonce všechny ideály

obsahují svou vlastní normu, známá věta připisovaná Lagrangemu říká, že každý řád prvku konečného okruhu $\mathcal{O}/\mathfrak{a}$, dělí jeho velikost, tedy normu $\mathfrak{a}$. Speciálně $N(\mathfrak{a}) \cdot 1 \in \mathfrak{a}$.

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:

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

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

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

a neutrální prvek je vždy celý okruh R. Ideály okruhu R proto tvoří se sčítáním grupu 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 komutativního okruhu R. Pokud pro nějaký ideál $\mathfrak{c} \subseteq R$ platí $\mathfrak{b} = \mathfrak{ac}$, 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 součet nějakých dvou jejich prvků je jednotkou R. Nedefinujeme největší společný dělitel, neboť ten ne vždy existuje, alespoň ne v obecném okruhu R. Nesoudělné ideály mají další zajímavé vlastnosti, jejich součin je totiž shodný s jejich průnikem.

Lemma 3.4.11. *Ať $\mathfrak{a}, \mathfrak{b}$ jsou nesoudělné ideály komutativního okruhu R. Pak platí rovnost $\mathfrak{ab} = \mathfrak{a} \cap \mathfrak{b}$.*

Důkaz. Jistě platí $\mathfrak{ab} \subseteq \mathfrak{a} \cap \mathfrak{b}$. Naopak ale máme:

$$(1)(\mathfrak{a} \cap \mathfrak{b}) = (\mathfrak{a} + \mathfrak{b})(\mathfrak{a} \cap \mathfrak{b}) = \mathfrak{a}(\mathfrak{a} \cap \mathfrak{b}) + \mathfrak{b}(\mathfrak{a} \cap \mathfrak{b}) = \mathfrak{ab} + \mathfrak{ba} \subseteq \mathfrak{ab}$$

díky komutativitě R. □

Nesoudělné ideály v celých číslech jsou generované nesoudělnými celými čísly m, n a podle Čínské zbytkové věty platí $\mathbb{Z}/(m) \times \mathbb{Z}/(n) \cong \mathbb{Z}/(mn)$. Toto tvrzení můžeme pak samozřejmě zobecnit do libovolných okruhu.

Věta 3.4.12. *(Čínská zbytková věta) Ať $\mathfrak{a}, \mathfrak{b}$ jsou nesoudělné ideály komutativního okruhu R. Pak platí:*

$$R/\mathfrak{a} \times R/\mathfrak{b} \cong R/\mathfrak{ab}.$$

Důkaz. Označme $f : R \longrightarrow R/\mathfrak{a} \times R/\mathfrak{b}$ homomorfismus okruhů redukující každý prvek R na příslušné zbytkové třídy v $R/\mathfrak{a}, R/\mathfrak{b}$. Do jádra f spadají právě prvky $\mathfrak{a} \cap \mathfrak{b} = \mathfrak{ab}$, tedy f dává vzniku injektivnímu homomorfismu $g : R/\mathfrak{ab} \longrightarrow R/\mathfrak{a} \times R/\mathfrak{b}$. Navíc pokud $a \in \mathfrak{a}, b \in \mathfrak{b}$ jsou prvky se součtem 1, libovolná jejich lineární kombinace $ax + by$ patří vždy do třídy (x, y) v $R/\mathfrak{a} \times R/\mathfrak{b}$ pro všechna $x, y \in R$, což dokazuje surjektivitu g a tedy hledaný isomorfismus. □

Tato věta mimo jiné znamená, že i norma ideálu je (ne nutně kompletně) multiplikativní funkcí. V příští kapitole ukážeme, že pro okruh $\mathcal{O}_K$ je dokonce kompletně multiplikativní, tj. platí $N(\mathfrak{a})N(\mathfrak{b}) = N(\mathfrak{ab})$ pro libovolné ideály $\mathfrak{a}, \mathfrak{b}$.

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 nenulové, 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. 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 $\mathfrak{p} \subseteq \mathcal{O}$ 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 nenulové 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 nižší normu, díky našim předpokladům oba obsahují součin nějakých prvoideálů. Díky platné inkluzi $(\mathfrak{a}+(x))(\mathfrak{a}+(y)) \subseteq \mathfrak{a}$ tak získáváme toužený spor. □

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

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

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

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

Poznamenejme, že inverz libovolného ideálu je jistě $\mathcal{O}_K$-modulem a navíc platí inkluze $\mathcal{O}_K \subseteq \mathfrak{p}^{-1}$. 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, to je ale (v plné obecnosti) unikátní pro 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 $\mathfrak{p}_i$, bez újmy na obecnosti 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 $\mathcal{O}_K$, $\mathfrak{p}$ prvoideál. Pokud platí $\mathfrak{p}\mathfrak{a} = \mathfrak{p}\mathfrak{b}$, tak $\mathfrak{a} = \mathfrak{b}$.*

S důkazem předchozího tvrzení je dovršen kopec teorie, kterou potřebujeme k důkazu jednoznačnosti rozkladu ideálů $\mathcal{O}_K$ na prvoideály.

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

$\mathfrak{p}_1 \supseteq \mathfrak{a} \supseteq \mathfrak{p}_1 \cdots \mathfrak{p}_{t+1}$ modulem $\mathfrak{p}_1^{-1}$ dává $\mathcal{O}_K \supseteq \mathfrak{a}\mathfrak{p}_1^{-1} \supseteq \mathfrak{p}_2 \cdots \mathfrak{p}_{t+1}$, modul $\mathfrak{a}\mathfrak{p}_1^{-1}$ je tedy ideál $\mathcal{O}_K$ a je rozložitelný, tedy $\mathfrak{a} = \mathfrak{a}\mathfrak{p}_1^{-1}\mathfrak{p}_1$ je též.

Nyní se pusťme na jednoznačnost. Dejme tomu, 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 $\mathfrak{p}_k^{-1}$) a pokračovat s ideálem $\mathfrak{a}\mathfrak{p}_k^{-1}$, čímž ostře snížíme $\max(k, \ell)$, sestupem dojdeme k závěru $k = \ell$ a že množiny prvoideálů na obou stranách musely být, včetně násobnosti, shodné. □

Pokud definujeme mocninu ideálu $\mathfrak{a}^k := \underbrace{\mathfrak{a} \cdots \mathfrak{a}}_{k}$, můžeme každý prvoideál jednoznačně rozložit na součin mocnin prvoideálů.

Jednoznačnost rozkladu pospolu s existencí inverzních prvoideálů (a tedy všech ideálů) v $\mathcal{O}_K$ nám umožňuje pozorovat mnoho paralel s celými čísly. Mimo jiné můžeme dělitelnost přeformulovat pomocí inkluze, pro ideály $\mathcal{O}_K$ platí ekvivalence $\mathfrak{a} \mid \mathfrak{b} \Leftrightarrow \mathfrak{b} \subseteq \mathfrak{a}$. Tato vlastnost je dokonce jednou z ekvivalentních definicí tzv. *Dedekindových oborů*, další z nich je jednoznačný rozklad ideálů na prvoideály či invertibilita každého ideálu (což platí díky invertibilitě prvoideálů). Invertibilita ideálů nám též umožňuje ideály krátit, ve smyslu implikace $\mathfrak{a}\mathfrak{b} = \mathfrak{a}\mathfrak{c} \Rightarrow \mathfrak{b} = \mathfrak{c}$ pro nenulový ideál $\mathfrak{a}$.

Čínská zbytková věta říká, že norma ideálů je multiplikativní, přičemž každý ideál $\mathcal{O}_K$ se jednoznačně rozkládá na součin prvoideálů. Dá se ukázat [51, Věta 4.3.18.], že norma mocniny prvoideálu $\mathfrak{p}$ je rovna příslušné mocnině normy $\mathfrak{p}$, tedy norma ideálů maximálního pořádku je kompletně multiplikativní.

Navíc jednoznačnost rozkladu nám pomůže omezit počet generátorů libovolného ideálu okruhu $\mathcal{O}_K$ číslem 2:

Věta 3.5.10. *Každý ideál $\mathcal{O}_K$ je generovaný nejvýše dvěma prvky.*

Důkaz. Dejme tomu, že $\mathfrak{a}$ není hlavní ideál a uvažme nenulové $x \in \mathfrak{a}$. Pak $(x) \subset \mathfrak{a}$, neboli $\mathfrak{a} \mid (x)$ podle předchozí diskuze. Uvažme pak rozklady ideálů $\mathfrak{a} = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_k^{a_k}$, $(x) = \mathfrak{p}_1^{b_1} \cdots \mathfrak{p}_k^{b_k}$ splňující $a_i \leqslant b_i$ pro každé i. Čínská zbytková věta 3.4.12 nám pak umožňuje najít y splňující $y \in \mathfrak{p}_k^{a_k} \setminus \mathfrak{p}_k^{a_k+1}$ pro každé k, což znamená $\mathfrak{a} = \mathfrak{p}_1^{a_1} \cdots \mathfrak{p}_k^{a_k} = (x) + (y) = (x, y)$. □

Nyní nastává vhodná chvíle se zamyslet nad naší volbou zúžit se pouze na maximální pořádky. Ve všech pořádcích opravdu platí, že každý ideál obsahuje součin nějakých prvoideálů a prvoideály se chovají podobně jako prvočísla, ztrácíme ale nutnou existenci inverzního prvoideálu a s ní i vyjádření všech ideálů jako součin prvoideálů i možnost krátit. Než si zmíníme ucelenou větu o faktorizaci ideálů v pořadcích, ukažme si příklad selhání faktorizace.

Příklad 3.5.11. Uvažme pořádek $\mathbb{Z}[2i] \subset \mathbb{Q}[i]$ a jeho ideál $(2, 2i)$. Ukážeme, že nemůže být rozložitelný na prvoideály. Platí totiž $(2, 2i)^2 = (4, 4i) = (2)(2, 2i)$, tedy pokud by byl tento ideál rozložitelný na prvoideály, musela by platit rovnost ideálů $(2, 2i) = (2)$, prvek $4 + 2i$ ale leží pouze v prvním ideálu. Problém zde nastává, protože 2 je sudé číslo, ideál $(2, 2i)$ tedy náleží do ideálu $\mathfrak{c} = \{x \in \mathbb{Q}(i) | x\mathbb{Z}[i] \subseteq \mathbb{Z}[2i]\}$, kde $\mathcal{O} = \mathbb{Z}[2i]$, a není invertibilní

jako $\mathcal{O}$-modul. Tento ideál $\mathfrak{c}$ je (vzhledem k inkluzi) největší ideál $\mathcal{O}_K$ obsažen v $\mathcal{O}$ a nese název *conductor ideal* (vodící ideál), více informací o něm se nachází na [11].

Věta 3.5.12. *Buď $\mathcal{O}$ pořádek číselného tělesa K a označme $\mathfrak{c} = \{x \in K | x\mathcal{O}_K \subseteq \mathcal{O}\}$. Každý ideál $\mathcal{O}$ nesoudělný s $\mathfrak{c}$ je součinem invertibilních prvoideálů a je sám jako $\mathcal{O}$-modul invertibilní, speciálně je též jednoznačně rozložitelný na součin invertibilních prvoideálů. Navíc neinvertibilních prvoideálů je pouze konečně mnoho.*

Důkaz se nachází na [11, Sec. 3.]. Speciálně ideály nesoudělné s $\mathfrak{c}$ se chovají prakticky identicky jako ideály maximálního pořádku, jsou všechny generované nejvýše dvěma prvky, můžeme je krátit, většina hezkých vlasností, které jsme zde zmínili. Ideály s $\mathfrak{c}$ soudělné u většiny těchto vlastností takovým či onakým způsobem selžou.

3.6 Grupa tříd ideálů a jednoznačnost rozkladu

Pojďme si nyní ideály pořádku $\mathcal{O}$ rozšířit na jeho moduly. Každý nenulový konečně generovaný $\mathcal{O}$-modul je roven $\mathfrak{a} = a_1\mathcal{O} + a_2\mathcal{O} + \cdots + a_k\mathcal{O}$ pro nenulová $a_i \in K$, přičemž víme, že existuje (nenulový) celý násobek každého z nich ležící v $\mathcal{O}_K$ a tedy i v $\mathcal{O}$. Pokud d je nejmenším společným násobkem všech těchto skalárů, $d\mathfrak{a}$ je ideálem $\mathcal{O}$ a naopak násobek $\mathcal{O}$-ideálu prvkem tělesa K je jistě konečně generovaný $\mathcal{O}$-modul.

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 $(\alpha) = \alpha\mathcal{O}$ *hlavním lomeným ideálem* R.

Mezi zástupce lomených ideálů v $\mathcal{O}_K$ patří například $\mathfrak{p}^{-1}$, inverz libovolného prvoideálu v $\mathcal{O}_K$. Hlavní lomené ideály nedávají příliš překvapivé příklady, v okruhu celých čísel tělesa $\mathbb{Q}$ je typickým příkladem $\frac{(3)}{2} = \frac{3}{2}\mathbb{Z}$.

Součet i součin ideálů pořádku $\mathcal{O}$ přirozeně generalizuje i na lomené ideály, z nich součin nás bude zajímat více.

Uvažme nyní $\mathfrak{a}, \mathfrak{b}$ lomené ideály pořádku $\mathcal{O}$ a definujme relaci *ekvivalence* $\sim$ s tím, že $\mathfrak{a}, \mathfrak{b}$ jsou ekvivalentní, pokud existují $x, y \in \mathcal{O}$ taková, že $\mathfrak{a} \cdot (x) = \mathfrak{b} \cdot (y)$, tedy pokud „podíl“ dvou takových ideálů je hlavní lomený ideál $\mathcal{O}$ (K je podílovým tělesem $\mathcal{O}$). Relace $\sim$ pak rozkládá množinu lomených ideálů $\mathcal{O}$ na třídy ekvivalence $[\mathfrak{a}]$, kde násobení hlavním (lomeným) ideálem ponechá třídu .

Věta 3.6.3. *Buďte $\mathfrak{a}, \mathfrak{b} \subseteq \mathcal{O}$ lomené ideály. Pak platí:*

$$[\mathfrak{a}] \cdot [\mathfrak{b}] = [\mathfrak{a}\mathfrak{b}].$$

Důkaz. Buďte $a, a' \in [\mathfrak{a}], b, b' \in [\mathfrak{b}]$ lomené ideály. Existují pak nenulové prvky $\alpha, \beta \in K$ takové, že $a' = \alpha \cdot a$ a $b' = \beta \cdot b$. Součin a', b' je roven $\alpha\beta ab$ a vždy tedy leží ve třídě $[ab]$. □

Součin dvou ideálů z dvou tříd spadá vždy do té samé třídy, můžeme pak přirozeně definovat na třídách násobení. To je jistě komutativní i asociativní a třída $[\mathcal{O}]$ hlavních lomených ideálů $\mathcal{O}$ skrz něj působí jako identita.

Věta 3.6.4. *Třídy invertibilních ideálů $\mathcal{O}$ tvoří grupu.*

Důkaz. Uvažme třídu obsahující nenulový lomený ideál $\mathfrak{a}$ pořádku $\mathcal{O}$. Pokud $\mathfrak{a}$ je invertibilní lomený ideál (existuje $\mathfrak{b}$ s $\mathfrak{ab} = \mathcal{O}$), třída $[\mathfrak{a}]$ je invertibilní též, s inverzem $[\mathfrak{b}]$. Naopak pokud je třída $[\mathfrak{a}]$ invertibilní, ve smyslu $[\mathfrak{a}][\mathfrak{b}] = [(1)]$, součin $\mathfrak{ab}$ je hlavní lomený ideál $x\mathcal{O}$, tedy platí $\mathfrak{a}\frac{\mathfrak{b}}{x} = \mathcal{O}$. Násobení tříd je zřejmě asociatvní, invertibilní třídy, neboli třídy invertibilních ideálů, pak tvoří s násobením tříd grupu. □

Definice 3.6.5. Buď $\mathcal{O}$ pořádek číselného tělesa K. Definujeme pak *grupu tříd ideálů* $Cl(\mathcal{O})$ jako grupu všech invertibilních tříd $[\mathfrak{a}]$ rozklad podle relace $\sim$ definované výše spolu s operací násobení ideálů.

Existuje ještě jeden způsob jak definovat grupu tříd ideálů, pro některé čtenáře možná přirozenější. Označíme-li množiny G, H invertibilních lomených, případně invertibilních hlavních lomených ideálů, spolu s operací násobení ideálů se obě stavají grupami, přičemž H je podgrupou G. V případě maximálního pořádku jsou G, H prostě množiny lomených, resp. hlavních lomených ideálů, protože díky jednoznačnosti rozkladu je každý ideál invertibilní. Grupu tříd ideálů pak můžeme zapsat jako faktorgrupu G/H.

Věta 3.6.6. *Každá třída ideálů $Cl(\mathcal{O})$ má reprezentanta z ideálů $\mathcal{O}$.*

Důkaz. Buď $\mathfrak{a}/m$ nějaký lomený ideál. Pak $\mathfrak{a}/m \cdot (m) = \mathfrak{a}$ je ideál $\mathcal{O}$ a leží ve stejné třídě jako $\mathfrak{a}/m$. □

Mluvíme-li o pořádcích v číselném tělese, nalezneme u nich pouze konečně mnoho takových tříd ideálů, i když okruhy obecně mohou mít grupu tříd ideálů nekonečnou. Tento fakt není na první pohled zjevný a nebudeme se jím nějak zvlášť zabývat. Klasické důkazy v každé třídě naleznou ideál normy nižší než počet tříd, z čehož konečnost po uvedení pár dalších tvrzení plyne, zaujatý čtenář ocení [51, Kap. 5].

Věta 3.6.7. *Grupa tříd ideálů pořádku $\mathcal{O}$ je konečná.*

S konečností grupy tříd ideálu se pak můžeme bavit o počtu jejích prvků.

Definice 3.6.8. *Třídové číslo $h_{\mathcal{O}}$* pořádku $\mathcal{O}$ definujeme jako počet prvků grupy $Cl(\mathcal{O})$.

Důkaz konečnosti třídového čísla též navádí na jeho nalezení, není to však jednoduchý proces. Často je redukován na rozkládání ideálů generovaných prvočísly pod danou hranici, viz například [51, Kap. 5.]. Obecný její výpočet, i s pomocí počítače, je obtížný, jak může dosvědčit fakt, že nejsou ani známa všechna reálná kvadratická tělesa, jejichž maximální pořádek má třídové číslo 1. Brzy si totiž ukážeme, že tyto okruhy jsou právě ty mající jednoznačný rozklad na ireducibilní prvky.

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

Věta 3.6.9. *Buď $\mathcal{O}$ pořádek a $\mathfrak{a}$ jeho ideál. Pak ideál $\mathfrak{a}^{h_\mathcal{O}}$ je hlavním ideálem $\mathcal{O}$.*

Jen takto banální poznatek o grupě tříd ideálů přirozeně spojuje grupu tříd ideálů s jednoznačností rozkladu na ireducibilní prvky. Pokud je totiž grupa tříd ideálů pořádku triviální, každý jeho (lomený) ideál je hlavní.

Věta 3.6.10. *Buď $\mathcal{O}_K$ maximální pořádek číselného tělesa. Pak každý prvek $\mathcal{O}_K$ se, až na permutaci a násobení jednotkou, jednoznačně rozkládá na ireducibilní prvky $\mathcal{O}_K$ právě pokud platí $h_{\mathcal{O}_K} = 1$.*

Důkaz. Nejprve ať je třídové číslo $\mathcal{O}_K$ rovno jedné. Každý jeho ideál je pak hlavní. Pokud pak $p_1 \cdots p_k = n = q_1 \cdots q_\ell$ jsou dva rozklady čísla n na (ne nutně různé) ireducibilní prvky, ideály generované příslušnými výrazy jsou:

$$(p_1) \cdots (p_k) = (n) = (q_1) \cdots (q_\ell).$$

Pokud by ideál generovaný například p_1 nebyl prvoideálem, byl by vyjádřitelný jako součin dvou (hlavních) ideálů $(p_1) = (a)(b) = (ab)$, tedy platí $p_1 \mid ab \mid p_1$ a existují $x, y \in \mathcal{O}_K$ splňující $p_1 = abx = p_1 xy$, x, y jsou pak jednotkami. Prvky p_1 a ab jsou asociované a díky ireducibilitě p je jedno z a, b jednotkou též, ideál (p_1) prvoideálem. Rovnosti výše jsou proto vyjádření prvoideálů a musí se tak příslušné množiny ideálů rovnat. To znamená, že množiny generátorů musí být, včetně násobnosti a bez ohledu na násobení jednotkou, shodné. Naopak ať $\mathcal{O}_K$ připouští jednoznačnost rozkladu na ireducibilní prvky a buď $\mathfrak{p}$ jeho prvoideál. Můžeme pak nenulový prvek $n \in \mathfrak{p}$ rozložit $n = p_1 \cdots p_k$ na (ne nutně různé) ireducibilní p_i. Jeden z těchto ireducibilních prvků, ať to je p_1, leží v prvoideálu $\mathfrak{p}$, tedy platí $(p_1) \subseteq \mathfrak{p}$. Díky ireducibilitě p_1 je (p_1) je prvoideál a maximalita prvoideálů říká $(p_1) = \mathfrak{p}$. Každý prvoideál je hlavní a tedy díky jednoznačnému rozkladu každého ideálu na prvoideály je i každý jiný ideál. □

Předchozí trvzení se samozřejmě přirozeně zobecňuje na ideály pořádků nesoudělné s vodícím ideálem. Soudělné ideály mnoho podobných hezkých vlastností ztrácí, několik pár z nich je k nalezení v [11, Ch. 3.].

Poznámka. Lze ukázat, že okruh celých algebraických čísel tělesa $\mathbb{Q}(\sqrt{d})$ s $d < 0$ má třídové číslo 1, neboli připouští jednoznačnost rozkladu, právě pokud je euklidovým okruhem.

Další zajímavé vlastnosti platí pro okruhy s třídovým číslem $2, 3, 4$ a více, například třídové číslo $\leqslant 2$ znamená, že byť se některé prvky rozkládají do více různých množin prvočinitelů, jejich počet (včetně násobnosti) zůstane vždy konzistentní. Například dříve zmíněný okruh $\mathbb{Z}[\sqrt{-14}]$ s dvěma rozklady 81 na různé počty faktorů má příslušnou grupu tříd ideálů čtyřprvkovou.

Příklad 3.6.11. Každý pořádek, který je euklidovým okruhem, má třídové číslo 1. Naopak okruh celých algebraických čísel $\mathbb{Z}[\sqrt{-5}] \subseteq \mathbb{Q}(\sqrt{-5})$ má třídové číslo 2. Ne každý jeho ideál

je hlavní, například $(2, 1+\sqrt{-5})$ je ideál s normou 2. Pokud by byl generovaný prvkem $a+b\sqrt{-5} \in \mathbb{Z}[\sqrt{-5}]$, bylo by $2 = N((a+b\sqrt{-5})) = N(a+b\sqrt{-5}) = a^2+5b^2$, což nemá řešení modulo 5.

Dále si zkonstruujeme přirozený injektivní homomorfismus vedoucí z grupy tříd ideálů pořádku do grupy tříd ideálů okruhu celých algebraických čísel, který nám poví o vztahů příslušných třídových čísel.

Věta 3.6.12. *Buď $\mathcal{O}$ pořádek číselného tělesa K. Pak $h(\mathcal{O}_K) \mid h(\mathcal{O})$.*

Důkaz. Uvažme třídu $[\mathfrak{a}] \in Cl(\mathcal{O}_K)$, kde $\mathfrak{a} \subseteq \mathcal{O}_K$ je ideál nesoudělný s vodícím ideálem $\mathfrak{c}$. Ideál $\mathfrak{a} \cap \mathcal{O}$ je ideál nesoudělný s $\mathfrak{c}$ a jeho $\mathcal{O}_K$ násobek $\mathcal{O}_K(\mathfrak{a} \cap \mathcal{O})$ je zjevně celý $\mathfrak{a}$, $\mathfrak{a} \cap \mathcal{O}$ je tedy invertibilní $\mathcal{O}$-modul a zobrazení $[\mathfrak{a} \cap \mathcal{O}] \mapsto [\mathcal{O}_K(\mathfrak{a} \cap \mathcal{O})] = [\mathfrak{a}]$ udává surjektivní homomorfismus $Cl(\mathcal{O}_K) \longrightarrow Cl(\mathcal{O})$, speciálně se příslušná třídová čísla dělí. □

O co víc, čistě pro zajímavost uveďme, že dokážeme s pomocí vodícího ideálu $\mathfrak{c}$ přesně určit vztah svazující $h(\mathcal{O})$ a $h(\mathcal{O}_K)$. Důkaz následujícího tvrzení není jednoduchý, uvedeme ho proto bez důkazu, ten je k nalezení na [11, Thm. 5.2.].

Věta 3.6.13. *Buď $\mathcal{O}$ pořádek číselného s pomocí vodícího ideálu $\mathfrak{c}$ tělesa K s vodičem $\mathfrak{c}$. Pak platí:*

$$\frac{h(\mathcal{O})}{h(\mathcal{O}_K)} = \frac{[(\mathcal{O}_K/\mathfrak{c})^\times : (\mathcal{O}/\mathfrak{c})^\times]}{[\mathcal{O}_K^\times : \mathcal{O}^\times]}.$$

Třídová čísla propojují rozklad v okruhu s jeho ideály, není to však ani zdaleka jediné, kde toto číslo působí. V analytické teorii čísel má své místo ve tvrzení známem jako „class number formula“ dokázané Peterem Dirichletem, spojující kvadratické formy, L-funkce, diskriminant číselného tělesa i číslo π (tentokrát opravdu ono číslo splňující $\pi \approx 3$) v jedné elegantní formuli. Opusťme ale nyní svět algebraické teorie čísel a pojďme užitkovat nabyté znalosti na teorii eliptických křivek.

Kapitola 4

Okruhy Endomorfismů

Jak napovídá název této sekce, endomorfismy na eliptické křivce 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 s teorií, kterou jsme si představili v předchozí kapitole, pomohu osvětlit funkčnost dalšího kryptografického schématu založeného na isogeních.

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 oboustranně „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. □

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 1.3.9 platí:

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

□

Víme, že isogenie působí na $E[\ell]$ jako 2×2 matice pod modulem ℓ, vztah isogenie a matice ji udávající nyní několikanásobně prohloubíme. Začněme s jejich minimálním polynomem.

Definice 4.1.6. Buď $f : X \longrightarrow Y$ funkce a $M \subseteq X$ množina. Pak *zúžení f na M* definujeme jako funkci $f|_M : M \longrightarrow Y$ splňující $f|_M(x) = f(x)$ pro všechna $x \in M$. Pro jednoduchost v případě isogenií značme $\phi|_n := \phi|_{E[n]}$.

Isogenie $\phi|_n$ je endomorfismem na volném $\mathbb{Z}_n$-modulu $E[n]$, který má rank nejvýše 2, v případě n nesoudělného s p právě 2. Matice M_n udávající akci $\phi|_n$ je určena volbou báze $E[n]$ jako $\mathbb{Z}_n$-modulu, vždy je však zachován determinant i stopa, můžeme proto zúžení takové isogenie přiřadit determinant i stopu.

Věta 4.1.7. *Buď $E/\mathbb{F}_p$ eliptická křivka a ϕ endomorfismus na ní. Pokud $p \nmid n$ je přirozené a $M \in \mathrm{SL}_2(\mathbb{Z}_n)$ matice působící jako ϕ na $E[n]$, platí:*

$$\mathrm{Tr}\,\phi \equiv \mathrm{Tr}\,M \pmod{n}, \qquad \deg\phi \equiv \det M \pmod{n}.$$

Důkaz. Ať ϕ není nulová mapa a buď $x^2 - sx + t$ charakteristický polynom $\phi|_n$, $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ matice udávající její akci, matice N reprezentující $\widehat{\phi}|_n$. Protože složení $\phi|_n$ a jejího duálu působí na $E[n]$ jako matice tI s $p \nmid b$, matice M je invertibilní a navíc:

$$N = tM^{-1} = \frac{t}{\det M}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Díky $\phi|_n + \widehat{\phi}|_n = s$ platí dále:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \frac{t}{\det M}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = sI = \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix}.$$

Platí poté $b - \frac{bt}{\det M} = 0$, $c - \frac{ct}{\det M} = 0$, buď tedy platí $b = c = 0$ a následně $a = d$, nebo $b = c = 0$ a následně $a = d$. První případ dává $t = \det M$ a $s = a + d = \operatorname{Tr} M$, jsme pak hotovi. Druhý případ říká, že $\phi|_n$ působí jako skalární násobení na $E[n]$. Tento případ je těžší a pro dostatečně vysoké n je jej též možné dotáhnout do konce, viz [63, Thm. 7.17], plný důkaz tvrzení užívá Weilových párování. □

Pro n dělitelné řádem všech bodů $P \in E(\mathbb{F}_p)$ dostatečně velké musí v kongruencích výše nastat rovnost, tedy nutně stupeň a stopa endomorfismu se rovnají determinantu, resp. stopě, matice udávající tuto isogenii.

Připomeňme, že každá 2×2 matice M splňuje charakteristickou rovnici $M^2 - \operatorname{Tr} M M + \det M I = 0$, isogenie i matice ji udávající splňují „stejnou“ kvadratickou rovnici.

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 $\pi = \pm\sqrt{-p} \notin \mathbb{Z}$. Platí pak inkluze $\mathbb{Z}[\pi] \subseteq \operatorname{End}(E)$. V další sekci okruh endomorfismů umístíme do kvadratického tělesa, čímž pak zásadně omezíme jeho možné tvary.

Pozastavme se ještě nad kvadratickým vztahem udávajícím Frobeniův endomorfismus, ne nutně již nad supersingulární křivkou. Ten nám pomůže poodhalit tajemství struktury grafů isogenií prvočíselného stupně ℓ, konkrétně kolik hran vychází z j-invariantu reprezentujícího příslušnou třídu isomorfismu. K tomu se na rovnice udávající isogenie musíme podívat ne jako celé, ale pouze modulo ℓ. Tato sekce postupuje volně podle [59, Sec. 6].

Nejprve si propojíme zúžení Frobeniova morfismu na ℓ-torze s rovnicí udávající Frobeniova morfismu modulo ℓ.

Lemma 4.1.8. *Buď $\phi : E \longrightarrow E'$ isogenie mezi křivkami nad $\mathbb{F}_p$ prvočíselného stupně ℓ. Pak ϕ je definovaná nad $\mathbb{F}_p$, právě pokud platí $\pi(\ker \phi) = \ker \phi$.*

Důkaz. Protože π je endomorfismus na E, E', π je na bodech obou křivek prostý. Druhou podmínku zadání můžeme proto relaxovat na $\pi \ker \phi \in \ker \phi$. Pokud je ϕ definovaná nad $\mathbb{F}_p$, tak je invariantní pod kompozicí s π a každý prvek jádra je zobrazen sám na sebe. Naopak ať platí $\pi(\ker \phi) = \ker \phi$. Jádro ϕ má prvočíselnou velikost a proto je generované každým svým afinním bodem, existuje $P \in \ker \phi$, že $\ker \phi = \langle P \rangle$. Protože π je endomorfismem na E, druhá podmínka je ekvivalentní s $\pi P \in \ker \phi$

Důsledek 4.1.9. *Nechť $P \in E[\ell]$. Pokud existuje celé λ splňující $\pi(P) = \lambda P$, pak λ je kořenem rovnice charakteristické rovnice π, $x^2 - tx + p$, modulo ℓ.*

Důkaz. Pokud platí $\pi(P) = \lambda P$, tak:

$$\mathcal{O} = (\pi^2 - t\pi + p)P = (\lambda^2 - t\lambda + p)P,$$

tedy protože řád P je ℓ, je λ kořenem polynomu $x^2 - tx + p$ modulo ℓ. □

Důsledek 4.1.10. *Buďte $\ell \neq p$ prvočíslo a bod $P \in E[\ell]$. Pak P splňuje kvadratickou rovnost $\pi^2 P - t\pi P + p \equiv 0 \pmod{\ell}$, právě pokud platí $\pi^2|_\ell P - t\pi|_\ell P + p = 0$.*

Důkaz. Předchozí propozice nám říká, že $\pi|_\ell$ permutuje prvky $E[\ell]$. Jediný prvek $E[\ell]$

Problém zjišťování akce π na ℓ-torzi můžeme tedy převést na řešení charakteristického polynomu π modulo ℓ. Speciálně P je kořenem $x^2 - tx + p$ pod modulem ℓ právě pokud platí $P \equiv \pi|_\ell \pmod{\ell}$???????.

Dejme tomu, že $\lambda \in E[\ell]$ je kořenem $x^2 - tx + p \pmod{\ell}$ v $\mathrm{End}(E)$, charakteristickou rovnici π můžeme rozložit jako:

$$(x - \lambda)(x - \mu) \equiv 0 \pmod{\ell},$$

čísla λ, μ jsou pak nutně vlastní čísla matice udávající lineární zobrazení $\pi|_\ell$ na $E[\ell]$. Protože $\pi|_\ell$ je lineární zobrazení na $E[\ell]$, je jednoznačně určené jeho akcí na (dvou) generátorech ℓ-torze.

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 3.1.12 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 opravdu 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:

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

Dále:

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

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

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

□

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

Důkaz. Položme $1/\alpha = \widehat{\alpha}/\,\mathrm{N}\,\alpha \in \mathrm{End}^0(E)$. Ukážeme, že toto $1/\alpha$ je hledaným inverzem, platí totiž $\alpha \cdot 1/\alpha = \alpha\widehat{\alpha}/\,\mathrm{N}\,\alpha = 1$, je tedy levým inverzem. Analogicky $(\widehat{\alpha}/\deg\alpha)\alpha = 1$, tedy $1/\alpha$ 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. Tento objekt je proto tělesem, právě pokud je násobení komutativní. Tímto způsobem algebru endomorfismů klasifikovat nebudeme, zvolíme trochu mazanější přístup. Nejprve vidíme, že každý prvek této algebry je nad racionálními čísly nejvýše kvadratický.

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

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

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

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 lišit bude. Významným výsledkem artibutovaným Maxu Deuringovi [19] je klasifikace plných algeber endomorfismů nad libovolným konečným tělesem. Ukáže se, že buď jsou isomorfní tělesu racionálních čísel (pouze křivky nad racionálními čísly), imaginárnímu kvadratickému tělesu, či tzv. *kvaternionové algebře*, tedy rozšíření $\mathbb{Q}(\alpha, \beta)$ s $\alpha\beta = -\beta\alpha$, kde na pořadí zápisu násobení jistě záleží. Obecně všechny supersingulární křivky mají algebru endomorfismů kvaternionovou algebru a obyčejné kvadratické těleso, hezký, poměrně elementární důkaz je k nalezení na [63, Thm. 13.17]. V případě naší „zjednodušené“ algebry endomorfismů však pro supersingulární křivky nastává opačný případ.

Věta 4.2.9. *Buďte $E/\mathbb{F}_p$ supersingulární křivka a libovolné $\alpha \in \mathrm{End}^0(E)$. Pak $\alpha \in \mathbb{Q}(\pi)$.*

Důkaz. Nejprve vidíme, že jak racionální čísla, tak Frobeniův prvek komutují s libovolným endomorfismem, $\pi = \pm\sqrt{-p}$ navíc není racionální. Zde vyžadujeme křivky i endomorfismy definované nad $\mathbb{F}_p$, supersingulární křivka $E/\mathbb{F}_{p^2} : y^2 = x^3 + x$ totiž splňuje $\pi = -2p$. Platí tedy $\alpha\pi = \pi\alpha$, přičemž lineární transformace tento vztah nezmění, speciálně $\alpha \mapsto \alpha - \frac{\mathrm{Tr}\,\alpha}{2}$:

$$\left(\alpha - \frac{\mathrm{Tr}\,\alpha}{2}\right)\pi = \pi\left(\alpha - \frac{\mathrm{Tr}\,\alpha}{2}\right),$$

tedy roznásobením opět platí $\alpha\pi = \pi\alpha$. Naše úprava ale způsobí:

$$\mathrm{Tr}\,\alpha \mapsto \mathrm{Tr}\left(\mathrm{Tr}\,\alpha - \frac{\mathrm{Tr}\,\alpha}{2}\right) = \mathrm{Tr}\,\alpha - \mathrm{Tr}\left(\frac{\mathrm{Tr}\,\alpha}{2}\right) = \mathrm{Tr}\,\alpha - \left(\frac{\mathrm{Tr}\,\alpha}{2} + \widehat{\frac{\mathrm{Tr}\,\alpha}{2}}\right) = 0,$$

kde užíváme nějaké základní vlastnosti stopy. Obdobně můžeme zvolit $\tilde{\pi} = \pi - \frac{\operatorname{Tr}\pi}{2}$, který má stopu nulovou, sám však nulový není. Dále, zvolme $\tilde{\alpha} = \alpha - \frac{\operatorname{Tr}\alpha\tilde{\pi}}{2\tilde{\pi}}$, tento prvek komutuje s $\tilde{\pi}$. Jeho stopa je rovna:

$$\operatorname{Tr}\tilde{\alpha} = \operatorname{Tr}\alpha - \operatorname{Tr}\frac{\operatorname{Tr}\alpha\tilde{\pi}}{2\tilde{\pi}} = \operatorname{Tr}\alpha - \left(\frac{\operatorname{Tr}\alpha\tilde{\pi}}{2}\right)\operatorname{Tr}\frac{1}{\tilde{\pi}} = -\left(\frac{\operatorname{Tr}\alpha\tilde{\pi}}{2}\right)\operatorname{Tr}\frac{1}{\tilde{\pi}}.$$

Platí ale:

$$\operatorname{Tr}\frac{1}{\tilde{\pi}} = \frac{1}{\tilde{\pi}} + \frac{1}{\widehat{\tilde{\pi}}} = \frac{\tilde{\pi} + \widehat{\tilde{\pi}}}{\tilde{\pi}\widehat{\tilde{\pi}}} = 0,$$

tedy $\operatorname{Tr}\tilde{\alpha} = 0$. Konečně, součin $\tilde{\alpha}$ a $\tilde{\phi}$ má stopu nulovou též:

$$\operatorname{Tr}\tilde{\pi}\tilde{\alpha} = \operatorname{Tr}\left(\alpha\tilde{\pi} - \frac{\operatorname{Tr}\alpha\tilde{\pi}}{2}\right) = \operatorname{Tr}\alpha\tilde{\pi} - \operatorname{Tr}\frac{\operatorname{Tr}\alpha\tilde{\pi}}{2} = 0,$$

což jsme chtěli. Čísla $\tilde{\alpha}, \tilde{\pi}$ proto splňující $\operatorname{Tr}\alpha = \operatorname{Tr}\tilde{\pi} = \operatorname{Tr}\tilde{\alpha}\tilde{\pi} = 0$. Pak tedy:

$$\tilde{\alpha}\tilde{\pi} = -\widehat{\tilde{\alpha}}\widehat{\tilde{\pi}} = -\widehat{\tilde{\pi}}\widehat{\tilde{\alpha}} = -(-\tilde{\pi})(-\tilde{\alpha}) = -\tilde{\pi}\tilde{\alpha} = -\tilde{\alpha}\tilde{\pi},$$

neboli $2\tilde{\alpha}\tilde{\pi} = 0$. Algebra $\operatorname{End}^0(E)$ je oborem integrity, tedy musí být jeden z $\tilde{\alpha}, \tilde{\pi}$ nulový. Frobeniův endomorfismus není racionálním číslem, musí proto $\tilde{\alpha}$ být nulové, z čehož plyne $\alpha = \frac{\operatorname{Tr}\alpha\tilde{\pi}}{2\tilde{\pi}} \in \mathbb{Q}(\pi)$. □

Důsledek 4.2.10. *Ať $E/\mathbb{F}_p$ je supersingulární křivka. Pak* $\operatorname{End}^0(E) = \mathbb{Q}(\pi)$.

Důkaz. Předchozí věta naznačuje inkluzi $\operatorname{End}^0 \subseteq \mathbb{Q}(\pi)$. Víme ale, že π má stupeň 2 nad racionálními čísly a každý jeho racionální násobek v algebře endomorfismů leží, čímž svíráme algebru endomorfismů z obou stran: $\mathbb{Q}(\pi) \subseteq \operatorname{End}^0(E) \subseteq \mathbb{Q}(\pi)$, nutně musí nastat rovnost $\operatorname{End}^0(E) = \mathbb{Q}(\pi)$. □

Speciálně, protože racionální čísla i π komutují s libovolným prvkem algebry endomorfismů, tento obor je komutativní.

Tak a nyní si můžeme užívat mnoho vlastností algebry endomorfismů (a tedy i okruhu endomorfismů) jako kvadratického tělesa, které dokazovat přímo by bylo bolestivé. Začněme s okruhem endomorfismů.

Věta 4.2.11. *Buď $E/\mathbb{F}_p$ supersingulární křivka. Pak* $\operatorname{End}(E)$ *je pořádkem v* $\mathbb{Q}(\pi)$.

Důkaz. Tvrzení je přímým důsledkem věty 3.2.16. □

O co víc, díky $\pi \notin \mathbb{Q}$ a větě 3.2.10 platí inkluze:

$$\mathbb{Z}[\pi] \subseteq \operatorname{End}(E) \subseteq \mathcal{O}_{\mathbb{Q}(\pi)}.$$

To nás navádí se podívat na zbytky, které p dává po dělení čtyřmi. Pokud $p \equiv 1 \pmod 4$, maximální pořádek v $\mathbb{Q}(\pi)$ je $\mathbb{Z}[\pi]$ a tak musí platit $\mathrm{End}(E) = \mathbb{Z}[\pi]$. V případě $p \equiv -1 \pmod 4$ máme zase řetězec inkluzí:

$$\mathbb{Z}[\pi] \subseteq \mathrm{End}(E) \subseteq \mathbb{Z}\left[\frac{1+\pi}{2}\right].$$

Poznámka. Výraz $\frac{1+\pi}{2}$ jako takový nedává v okruhu endomorfismů smysl, protože je formálně roven $\frac{1}{2} \otimes (1+\pi)$. Isogenie $[2]$ na E je ale surjektivní, tedy můžeme tento výraz považovat jako prvek splňující $2\frac{1+\pi}{2} = 1+\pi$ a šťastně s ním pracovat jako s endomorfismem.

Při hledání okruhu endomorfismů nás už pouze volba prvočísla nezachrání.

Příklad 4.2.12. Podívejme se na supersingulární křivky $E_1/\mathbb{F}_{19} : y^2 = x^3 + x$ a $E_2/\mathbb{F}_{19} : y^2 = x^3 - x$. Obě křivky mají shodný Frobeniův prvek $\sqrt{-19}$ a tedy i algebru endomorfismů $\mathbb{Q}(\sqrt{-19})$. Křivka E_1 má okruh endomorfismů pouze pořádek $\mathbb{Z}[\pi]$, zato okruh endomorfismů křivky E_2 je maximální pořádek $\mathbb{Z}\left[\frac{1+\pi}{2}\right]$.

Krom komutativity samotného okruhu endomorfismů můžeme ukázat i komutativitu isogenií s endomorfismy.

Lemma 4.2.13. *Buďte E_1, E_2 dvě křivky nad $\mathbb{F}_p$ a $\phi : E_1 \longrightarrow E_2$ isogenie. Libovolný endomorfismus $\alpha \in \mathrm{End}(E_1) \cap \mathrm{End}(E_2)$ pak s ϕ komtuje, tedy $\phi\alpha = \alpha\phi$.*

Důkaz. Endomorfismus $\alpha \in \mathrm{End}(E_1) \cap \mathrm{End}(E_2)$ můžeme vyjádřit jako $\frac{a+b\pi}{2}$, kde a, b jsou celá čísla, protože $\mathrm{End}(E_i) \subseteq \mathbb{Z}\left[\frac{1+\pi}{2}\right]$. Případ $\alpha = 0$ je zřejmý, jinak jsou endomorfismy surjektivní, pro každý bod $P \in E$ tedy existuje Q splňující $2Q = P$. Pak pro každý $P \in E$ platí:

$$\phi\alpha P = \phi\frac{a+b\pi}{2}P = \phi(a+b\pi)Q = (a+b\pi)\phi Q = \frac{a+b\pi}{2}\phi 2Q = \frac{a+b\pi}{2}\phi P = \alpha\phi P.$$

□

Těleso $\mathbb{Q}(\pi)$, je ale vždy zachováno, protože všechny supersingulární křivky nad $\mathbb{F}_p$ mají stejný Frobeniův prvek. Pro zajímavost uveďme, že i algebra endomorfismů obyčejné křivky je pod akcí isogenie zachována.

Věta 4.2.14. *Buďte E, E' křivky nad $\mathbb{F}_p$. Pokud jsou tyto dvě křivky jsou nad $\mathbb{F}_p$ isogenní, platí $\mathrm{End}^0(E) \cong \mathrm{End}^0(E')$.*

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

V případě obyčejných křivek dmsamdsaofopasfopsaf, můžeme však jednoduchost okruhů endomorfismů využít při charakterizaci isogenií mezi křivkami.

Věta 4.2.15. *Buďte E, E' křivky nad $\mathbb{F}_p$, mezi kterými existuje isogenie ϕ stupně ℓ. Pak nastává jeden z následujících případů:*

- $\mathrm{End}(E) = \mathrm{End}(E')$*, pak ϕ nazveme horizontální. V opačném případě ji nazveme vertikální a bere následující formy:*
- $[\mathrm{End}(E) : \mathrm{End}(E')] = \ell$*,*
- $[\mathrm{End}(E') : \mathrm{End}(E)] = \ell$*.*

Důkaz. Existují $\vartheta, \upsilon \in \mathrm{End}^0(E)$ taková, že $\mathrm{End}(E) = \mathbb{Z}[\vartheta]$ a $\mathrm{End}(E') = \mathbb{Z}[\upsilon]$ jsou v $\mathrm{End}^0(E)$ pořádky. Endomorfismus $\phi\vartheta\widehat{\phi} \in \mathbb{Z}[\upsilon]$ má duál $\phi\widehat{\vartheta}\widehat{\phi}$ díky komutativitě endomorfismů a isogenií. Spočtěme si jeho normu a stopu:

$$\mathrm{N}\,\phi\vartheta\widehat{\phi} = \phi\vartheta\widehat{\phi}\phi\widehat{\vartheta}\widehat{\phi} = \phi\vartheta\ell\widehat{\vartheta}\widehat{\phi} = \phi\ell\vartheta\widehat{\vartheta}\ell\widehat{\phi} = \phi\ell(\mathrm{N}\,\vartheta)\widehat{\phi} = \ell(\mathrm{N}\,\vartheta)\phi\widehat{\phi} = \ell^2\,\mathrm{N}\,\vartheta = \mathrm{N}\,\ell\vartheta,$$
$$\mathrm{Tr}\,\phi\vartheta\widehat{\phi} = \phi\vartheta\widehat{\phi} + \phi\widehat{\vartheta} = \phi(\vartheta + \widehat{\vartheta})\widehat{\phi} = \phi(\mathrm{Tr}\,\vartheta)\widehat{\phi} = \phi\widehat{\phi}\,\mathrm{Tr}\,\vartheta = \ell\,\mathrm{Tr}\,\vartheta = \mathrm{Tr}\,\ell\vartheta.$$

Endomorfismus $\phi\vartheta\widehat{\phi}$ tedy splňuje stejnou kvadratickou rovnici jako $\ell\vartheta$ a je buď roven jemu, nebo jeho duálu. Tak či tak $\ell\vartheta = \phi\vartheta\widehat{\phi} \in \mathbb{Z}[\upsilon]$ a analogicky $\ell\upsilon \in \mathbb{Z}[\vartheta]$.

V případě vertikálních isogenií mezi supersingulárními křivkami víme, že možné stupně rozšíření pouze 1 a 2, tedy mezi takovými křivkami existují vertikální isogenie stupně nejvýše 2 a všechny ostatní nutně zachovají okruh endomorfismů.

Dále si ukážeme, že endomorfismy komutují s isogeniemi.

4.3 Isogenie generované ideály

Mějme v této sekci $E/\mathbb{F}_p$ supersingulární eliptickou křivku a $\mathrm{End}(E)$ pořádek v kvadratickém tělese $\mathbb{Q}(\pi)$ s vodičem dělícím 2. Libovolný invertibilní ideál $\mathfrak{a} \subseteq \mathrm{End}(E)$ se jednoznačně rozkládá na $(\pi)^r\mathfrak{b}$, kde $r \geqslant 0$ a $\pi \notin \mathfrak{b}$ je jednoznačně rozložitelný na prvoideály. Z každého takového ideálu zkonstruujeme isogenii vycházející z E, který má mnoho společného s jeho jádrem.

Definice 4.3.1. Buď E supersingulární křivka a invertibilní ideál $\mathfrak{a} \subset \mathrm{End}(E)$. Definujme pak $\mathfrak{a}$-*torsor* jako:

$$E[\mathfrak{a}] := \bigcap_{\alpha \in \mathfrak{a}} \ker \alpha.$$

Definice 4.3.2. Buď $\mathfrak{a} \subset \mathrm{End}(E)$ invertibilní ideál, který se rozkládá jako $(\pi)^r\mathfrak{b}$ s $\pi \notin \mathfrak{b}$. Pak isogenii $\phi_\mathfrak{a} : E \longrightarrow E/\mathfrak{a}$ definujeme jako separabilní isogenii $\psi : E \longrightarrow E/\mathfrak{b}$, kde:

$$\ker \psi = E[\mathfrak{b}],$$

složenou s r iteracemi Frobeniova endomorfismu.

Tato křivka $E/\mathfrak{a}$ (i isogenie) jsou definované nad $\mathbb{F}_p$ a křivka je až na isomorfismus jednoznačně určena, opodstatňující notaci $E/\mathfrak{a}$.

Pojďme tyto isogenie vycházející z ideálů zkoumat. Přirozeně bychom chtěli vědět, jaký stupeň takové isogenie mají.

Lemma 4.3.3. *Buď $\mathfrak{a} \subset \mathrm{End}(E)$ invertibilní ideál. Pak platí $\deg \phi_\mathfrak{a} = N(\mathfrak{a})$.*

Důkaz. Bez újmy na obecnosti uvažme $\pi \notin \mathfrak{a}$ a tedy ϕ_a separabilní.

Důsledek 4.3.4. *Buďte $\mathfrak{a}, \mathfrak{b} \subset \mathrm{End}(E)$ invertibilní ideály. Pak platí $\deg \phi_\mathfrak{a}\phi_\mathfrak{b} = \deg \phi_{\mathfrak{ab}}$.*

Důkaz. Věta je přímým důsledkem předchozího lemmatu a multiplikativity normy ideálů. □

Věta 4.3.5. *Buďte $\mathfrak{a}, \mathfrak{b} \in \mathrm{End}(E)$ invertibilní ideály. Pak platí $\ker \phi_\mathfrak{a}\phi_\mathfrak{b} = \ker \phi_{\mathfrak{ab}}$.*

Důkaz. Bez újmy na obecnosti ať $\mathfrak{a}, \mathfrak{b}$ neobsahují π, pak isogenie $\phi_\mathfrak{a}, \phi_\mathfrak{b}$ jsou separabilní. Uvažme P libovolný bod v jádře $\phi_\mathfrak{a}\phi_\mathfrak{b}$. Ekvivalentně tento bod splňuje $\phi_\mathfrak{b}(P) \in \ker \phi_\mathfrak{a} = E[\mathfrak{a}]$. Každý endomorfismus $\alpha \in \mathfrak{a}$ pak splňuje $\alpha\phi_\mathfrak{b}(P) = \mathcal{O}$. Nyní užijeme větu 4.2.13, dle které můžeme prohodit pořadí aplikace našich isogenií:

$$\phi_\mathfrak{b}\alpha(P) = \mathcal{O}$$

pro každý $\alpha \in \mathfrak{a}$, neboli $\alpha(P) \in \ker \phi_\mathfrak{b} = E[\mathfrak{b}]$. Pro každý endomorfismus $\beta \in \mathfrak{b}$ proto platí $\beta\alpha(P) = \mathcal{O}$. Speciálně pro každý endomorfimus $\gamma \in \mathfrak{ba} = \mathfrak{ab}$ je pak $\gamma(P) = 0$, protože pro libovolné dva endomorfismy, které nulují P, tuto vlastnost sdílí i jejich součet. Platí tedy $P \in E[\mathfrak{ab}]$ a tak $\ker \phi_\mathfrak{a}\phi_\mathfrak{b} \subseteq \ker \phi_{\mathfrak{ab}}$, tedy díky separabilitě obou isogenií:

$$\deg \phi_\mathfrak{a}\phi_\mathfrak{b} \leqslant \deg \phi_{\mathfrak{ab}},$$

přičemž věta 4.3.4 dokonce diktuje v nerovnosti výše rovnost. Musí tedy nastat i rovnost jader příslušných (separabilních) isogenií, tj. $\ker \phi_\mathfrak{a}\phi_\mathfrak{b} = \ker \phi_{\mathfrak{ab}}$. □

Skládání isogenií má díky (skoro) jednoznačnosti separabilní isogenie ten samý efekt jako násobení příslušných ideálů, hlavní ideály nám zde reprezentují endomorfismy. Dvě křivky isogenní s E jsou proto spolu isomorfní, právě pokud se příslušné isogenie liší až na endomorfismus. Ideály příslušící těmto isogeniím jsou pak až na násobení hlavním lomeným ideálem shodné, tj. leží ve stejné třídě grupy tříd ideálů. Tuto úvahu shrnuje následující tvrzení:

Věta 4.3.6. *Buďte $\phi : E \longrightarrow E/\mathfrak{a}$ a $\psi : E \longrightarrow E/\mathfrak{b}$ isogenie nad $\mathbb{F}_p$ a $\mathcal{O} = \mathrm{End}(E)$. Pak jsou cílové křivky spolu nad $\overline{\mathbb{F}}_p$ isomorfní, právě pokud $\mathfrak{a}$ a $\mathfrak{b}$ leží ve stejné třídě $Cl(\mathcal{O})$.*

Každému j-invariantu přísluší dvě třídy křivek isogenních nad $\mathbb{F}_p$, křivka a její kvadratické twisty. Počet takových j-invariantů je proto roven $2h(\mathrm{End}(E))$.

Kapitola 5

CSIDH

Okruh endomorfismů definovaných nad $\mathbb{F}_p$ pro supersingulární eliptickou křivku má strukturu pořádku v imaginárním kvadratickém tělese a jak jsme zmínili, podobnou strukturu tvoří plný okruh endomorfismů příslušící křivce obyčejné. Toto pozorování nás nabádá vzkřísit nápady Couveigna, Rostovstseva a Stolbunova [14], [56] a adaptovat je do supersingulárního prostředí. Subexponenciální útok Childs-Jao-Soukhareva na tyto protokoly závisí na faktu, že grupa tříd ideálů a tedy i okruh endomorfismů je komutativní, což byl jedním z důvodů, proč SIDH užívá křivky supersingulární. Tento útok je proto opět hrozbou ?????????

Závěr

zu ende

Literatura

[1] Azarderakhsh, Reza, Matthew Campagna, Craig Costello, Luca De Feo, Basil Hess, Amir Jalali, Brian Koziel, Brian LaMacchia, Patrick Longa, Michael Naehrig, Joost Renes, Vladimir Soukharev a David Urbanik: *SIKE: Supersingular Isogeny Key Encapsulation*. 2017.

[2] Bisson, Gaetan a Andrew V. Sutherland: *Computing the Endomorphism Ring of an Ordinary Elliptic Curve Over a Finite Field*. 2009. Dostupné z: https://arxiv.org/abs/0902.4670.

[3] Bottinelli, Paul, Victoria de Quehen, Christopher Leonardi, Anton Mosunov, Filip Pawlega a Milap Sheth. ISARA Corporation, Waterloo, Canada. 2019. Dostupné z: https://eprint.iacr.org/2019/1333.

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

[5] Cervantes-Vázquez, Daniel, Eduaro Ochoa-Jiménez a Francisco Rodríguez-Henríquez: *eSIDH: the revenge of the SIDH*. 2020.

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

[7] Deng, Yu-Hao, Xing Ding, Lin Gan, Peng Hu, Yi Hu, Ming-Cheng Chen, Xiao Jiang, Hao Li, Li Li, Yuxuan Li, Nai-Le Liu, Chao-Yang Lu, Yi-Han Luo, Jian-Wei Pan, Li-Chao Peng, Jian Qin, Hui Wang, Zhen Wang, Zhen Wang, Guangwen Yang, Lixing You, Han-Sen Zhong:*Quantum computational advantage using photons*. Science Magazine. 2020. Dostupné z: https://science.sciencemag.org/content/370/6523/1460.full

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

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

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

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

[12] Costello, Craig: *B-SIDH: supersingular isogeny Diffie-Hellman using twisted torsion*. Microsoft Research, USA, 2019. Dostupné z: https://eprint.iacr.org/2019/1145.

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

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

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

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

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

[18] 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.

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

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

[21] Eisenträger, Sean H., Kristin Lauter, Travis Morrison a Christpoher Petit: *Supersingular Isogeny Graphs and Endomorphism Rings: Reductions and Solutions*. Advances in Cryptology – EUROCRYPT 2018, Lecture Notes in Computer Science, pages 329–368. Springer International Publishing, 2018.

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

[23] 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.

[24] Galbraith, Steven D., Florian Hess a Nigel P. Smart: *Extending the GHS Weil descent attack*. EUROCRYPT 2002, Springer LNCS 2332 29-44, 2002.

[25] Galbraith, Steven D. a Anton Stolbunov: *Improved Algorithm for the Isogeny Problem for Ordinary Elliptic Curves*. Applicable Algebra in Engineering, Communication and Computing, Vol. 24, No. 2, 2013. Dostupné z: https://arxiv.org/abs/1105.6331.

[26] Galbraith, Steven D., Christopher Petit, Barak Shani a Yan Bo Ti: *On the security of supersingular isogeny cryptosystems*. International Conference on the Theory and Application of Cryptology and Information Security. Springer, 2016.

[27] Griffiths, Robert B.: *Hilbert Space Quantum Mechanics*. 2014.

[28] 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.

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

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

[31] Jao, David a David Urbanik: *Extra Secrets from Automorphisms and SIDH-based NIKE*, 2018.

[32] Johnson, Don, Alfred Menenzes a Scott Vanstone: *The Elliptic Curve Digital Signature Algorithm (ECDSA)*. Certicom a Department of Combinatorics & Optimization, University of Waterloo, Ontario, Canada. 2001.

[33] 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.

[34] Karamlou, Amir H, Willieam A. Simon, Amara Katabarwa, Travis L. Scholten, Borja Peropandre a Yudong Cao: *Analyzing the Performance of Variational Quantum Factoring on a Superconducting Quantum Processor*. Zapata Computing, Boston; Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge a IBM Quantum, IBM T. J. Watson Research Center, New York, 2020. Dostupné z: https://www.zapatacomputing.com/publications/analyzing-the-performance-of-variational-quantum-factoring-on-a-superconducting-quantum-processor/.

[35] Kohel, David R.: *Endomorphism rings of elliptic curves over finite fields*. University of California, Berkley, 1996.

[36] Koblitz, Neal: *Elliptic curve cryptosystems*. Mathematics of Computation. 48 (177): 203–209, 1987.

[37] Lagarias, Jeffrey C. a Andrew M. Odlyzko: *Effective Versions of the Chebotarev Density Theorem*. Algebraic Number Fields,L-Functions and Galois Properties (A. Fröhlich, ed.), pp. 409–464. New York, London: Academic Press, 1977.

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

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

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

[41] Menezes, Afred, Tatsuki Okamoto a Scott Vanstone: *Reducing Elliptic Curve Logarithms to Logarithms in a Finite Field*. IEEE Transactions on Information Theory 39, 1993.

[42] Miller, Victor: *Use of elliptic curves in cryptography*. Advances in Cryptology—CRYPTO '85, Lecture Notes in Computer Science, vol 218. Springer, pp 417–426, 1986.

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

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

[45] Olšák, Radek: *Method of Animation*. *i*KS soustředění, Strmilov, 2020.

[46] Perutka, Tomáš: *Vyjadřování prvočísel kvadratickými formami*. Středoškolská odborná činnost. Brno: Masarykova univerzita, 2017. Dostupné z: https://socv2.nidv.cz/archiv39/getWork/hash/ff6e75d5-f922-11e6-848a-005056bd6e49.

[47] Perutka, Tomáš: *Užití dekompoziční grupy k důkazu zákona kvadratické reciprocity*. Středoškolská odborná činnost. Brno: Masarykova univerzita, 2018. Dostupné z: https://socv2.nidv.cz/archiv40/getWork/hash/1984482c-1298-11e8-90e4-005056bd6e49.

[48] Pezlar, Zdeněk: *Zajímavá využití algebraické teorie čísel*. Středoškolská odborná činnost. Brno: Masarykova univerzita, 2020. Dostupné z: https://socv2.nidv.cz/archiv42/getWork/hash/921aa7aa-568d-11ea-9fea-005056bd6e49.

[49] Pizer, Arnold K.: *Ramanujan graphs and Hecke operators*. Bulletin of the American Math Society, 23, 1990.

[50] Proos, John a Christof Zalka: *Shor's discrete logarithm quantum algorithm for elliptic curves*. Department of Combinatorics & Optimization, University of Waterloo, Ontario, Canada, 2008. Dostupné z: https://arxiv.org/abs/quant-ph/0301141.

[51] 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/.

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

[53] 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.

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

[55] Shenguy, Zhang: *Promised and Distributed Quantum Search Computing and Combinatorics*. Proceedings of the Eleventh Annual International Conference on Computing and Combinatorics, Berlin, Heidelberg, 2005.

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

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

[58] Schoof, René: *Elliptic Curves Over Finite Fields and the Computation of Square Roots* mod *p*. Journal de Théorie des Nombres de 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.

[59] Schoof, René: *Counting points on elliptic curves over finite fields*. Journal de Théorie des Nombres de Bordeaux 7, 1995. Dostupné z: https://www.mat.uniroma2.it/~schoof/ctg.pdf.

[60] 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.

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

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

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

[64] Tani, Seiichiro: *Claw Finding Algorithms Using Quantum Walk*. Theoretical Computer Science, 410(50):5285-5297, 2009.

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

[66] Vélu, Jacques: *Isogénies entre courbes elliptiques*. Comptes Rendus de l'Académie des Sci-ences de Paris, 1971.

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

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