UNIVERZA V LJUBLJANI

FAKULTETA ZA MATEMATIKO IN FIZIKO

SCHOOFOV ALGORITEM

JERNEJ BARBIČ

Delo je pripravljeno pod mentorstvom
doc. dr. Aleksandra Jurišića.

LJUBLJANA 2000

Povzetek

V tem delu bomo definirali eliptične krivulje in grupo na eliptični krivulji. Podali bomo podroben opis Schoofovega algoritma za določanje moči grupe racionalnih točk na eliptični krivulji, definirani nad poljubnim končnim obsegom. Vse obravnavane pojme bomo umestili v širši kriptografski okvir in nakazali glavne smernice razvoja kriptografske znanosti na tem področju. Poleg tega bomo navedli predstavitev osnovnih algebraičnih lastnosti končnih obsegov s poudarkom na lastnostih, ki so pomembne za kriptografijo.

Ključne besede: eliptične krivulje, končni obsegi, aritmetična teorija števil.

Abstract

In this thesis elliptic curves and elliptic groups are introduced. We give a detailed description of the basic Schoof algorithm for determining the cardinality of the group of rational points on an elliptic curve over a finite field. We demonstrate the cryptographic importance of the presented material and comment on the current development of the cryptographical science in this field. We also give an overview of the basic algebraic properties of finite field with implications to cryptography.

Keywords: elliptic curves, finite fields, computational number theory.

Math. subj. class (2000): primary 14H52, 11T71, 11Y16.

Zahvaljujem se mojemu mentorju doc. dr. Aleksandru Jurišiću za vso pomoč pri pisanju tega dela.

Zahvaljujem se tudi mojim staršem in vsem mojim prijateljem, ki mi stojijo ob strani.

PROGRAM DIPLOMSKEGA DELA

Schoofov algoritem

Delo naj predstavi matematične osnove, potrebne za razumevanje in implementacijo Schoofovega algoritma za določanje moči grupe racionalnih točk na eliptični krivulji, definirani nad poljubnim končnim obsegom. Glavni cilji so podroben opis Schoofovega algoritma, analiza kompleksnosti in pomen tega algoritma za kriptosistem z eliptičnimi krivuljami. Izdela naj se tudi predstavitev osnovnih algebraičnih lastnosti končnih obsegov, ki se uporabljajo v kriptografiji.

Literatura:

R. Schoof, Elliptic Curves Over Finite Fields and the Computation of Square Roots mod p, *Mathematics of Computation* **44** (1985), 483-494.

A. Enge, *Elliptic curves and their applications to cryptography: an elementrary introduction*, doktorska disertacija, Univerza v Augsburgu, 1997.

I. Blake, G. Seroussi, N. Smart, *Elliptic curves in cryptography*, Cambridge University Press, 1999.

Kazalo

Uvod

To delo govori o eliptičnih krivuljah, končnih obsegih in njihovem sožitju z računalniki. Eliptične krivulje so v matematiki s teoretičnega vidika algebraične geometrije prisotne že dolgo. Z razvojem računalnikov pa so doživele svoj preporod. Na Internetu lahko najdemo veliko strani, posvečenih tej temi. Eliptične krivulje so namreč temelj eliptične javne kriptografije. Poleg kriptosistemov DES in RSA je to najbolj razširjeni način za varno izmenjavo zaupnih podatkov preko kanala, kjer lahko nekdo prisluškuje. Tipičen primer takega kanala je Internet. Zaradi njegovega hitrega širjenja so potrebe računalniške industrije po obvladovanju eliptičnih krivulj vse večje, kar na eliptično kriptografijo usmerja žaromete javnosti in celotnemu področju prinaša izdatne finančne injekcije.

To diplomsko delo je spopad z ogromnimi naravnimi števili. Glavni cilj dela je opis Schoofovega (izg. Škofovega) algoritma, ki je eden od temeljnih kamnov varne eliptične kriptografije. Naloga Schoofovega algoritma je iz dveh koeficientov, s katerima je določena eliptična krivulja, izračunati moč grupe na eliptični krivulji. Za kriptografijo namreč potrebujemo krivulje, katerih moč grupe je deljiva z dovolj velikim praštevilom. Svetovni rekord v določanju moči grupe na eliptični krivulji se v letu 2000 giblje v območju 1000-mestnih naravnih števil.

Schoofov algoritem je pravzaprav velika iznajdba, ki je potisnila meje računsko obvladljivih eliptičnih krivulj daleč naprej. Moderne izboljšave Schoofovega algoritma predstavljajo enega od vrhuncev sodobne znanosti in tehnologije. Schoofov algoritem se za resne potrebe izvaja na velikih superračunalnikih, ki imajo pogosto del čipov zgrajenih namensko za potrebe eliptične kriptografije. Lastnikom običajnih PC-jev pa ostane le možnost, da računsko prednost superračunalnikov izničimo z novo bistro idejo.

S strogo teoretičnega vidika sodi to diplomsko delo v algebro, predvsem v teorijo končnih obsegov. Dotakne pa se tudi algebraične geometrije. Vendar imajo v tem delu prednost ideje in algoritmi, ki vodijo do praktično uporabnih kriptografskih postopkov.

V slovenski literaturi ni nobenega drugega dela o Schoofovem algoritmu ali o eliptični kriptografiji. Želim, da bi moje delo bilo primerno za vsakega matematika, ki želi razumeti teoretično ozadje eliptične kriptografije in dobiti občutek, kateri problemi na tem področju predstavljajo najtrše orehe. O neposrednih protokolih eliptične kriptografije sicer ni kaj dosti besede. Dobri uvodni referenci za to sta [Stinson] in [Jurišić].

Posamezna poglavja imajo naslednjo vlogo. Poglavje o končnih obsegih bralca seznani s splošnimi algebraičnimi osnovami teorije končnih obsegov. Govori tudi o reševanju kvadratnih enačb v končnih obsegih. To znanje je za pisanje učinkovitih algoritmov eliptične kriptografije nujno. V drugem poglavju definiramo eliptične krivulje in nanje uvedemo strukturo Abelove grupe. Nato sledi poglavje, ki je neke vrste ogrevanje za spopad s Schoofovim algoritmom. V njem se bo bralec lahko naučil potrebne algebre za razumevanje Schoofovega algoritma. Zadnje, četrto poglavje je posvečeno natančnemu opisu Schoofovega algoritma. V njem je mogoče najti precej nasvetov za učinkovito izvedbo algoritma, hkrati pa opozorimo tudi na nekatere pasti in slabosti osnovnega Schoofovega algoritma. Na koncu na kratko predstavimo še druge sorodne algoritme in nakažemo, v katero smer se na tem področju razvija kriptografska znanost.

Poglavje 1

Končni obsegi

V tem poglavju bomo predstavili tiste pojme iz teorije končnih obsegov, ki so pomembni za kriptografijo. Najprej bomo navedli osnovne splošne algebraične lastnosti končnih obsegov. Zatem bomo definirali sled, raziskali mrežo podobsegov končnega obsega in konstruirali razpadne obsege končnih obsegov. Sledi razdelek, v katerem bomo predstavili dva standardna razreda baz končnega obsega nad svojim praobsegom: polinomske in normalne baze. Ta razdelek je verjetno najpomembnejši v celotnem poglavju, saj je osrednjega pomena za eliptično kriptografijo. Zatem se bomo posvetili še reševanju kvadratnih enačb v končnih obsegih.

1.1 Osnove

V tem razdelku bomo navedli osnovne algebraične lastnosti končnih obsegov. Dokaze tistih trditev, ki jih bomo zgolj navedli, je mogoče najti v [Vidav] ali [Fraleigh].

Po Wedderburnovem izreku so vsi končni obsegi komutativni. Moč vsakega končnega obsega je oblike p^n, kjer je p praštevilo in $n \in \mathbb{N}$. Vsi končni obsegi enake moči so med seboj izomorfni. Za vsako praštevilo p in vsak $n \in \mathbb{N}$ torej obstaja do izomorfizma natančno natanko en obseg moči p^n. V tem delu ga bomo označevali s $\mathbb{F}_{p^n}$, nekateri avtorji pa uporabljajo oznako $GF(p^n)$ (Galois Field). Na tistih mestih tega dela, kjer število n ne bo igralo pomembnejše vloge, bomo pogosto uporabljali standardno oznako $q = p^n$.

Karakteristika obsega $\mathbb{F}_{p^n}$ je p. V vsakem obsegu $\mathbb{F}_{p^n}$ lahko torej najdemo ustrezen praobseg $\mathbb{F}_p$. V njem so natanko tisti elementi obsega $\mathbb{F}_{p^n}$, ki zadoščajo enačbi $X^p - X = 0$. Seveda so obsegi $\mathbb{F}_p$ tudi predstavniki končnih obsegov, dobimo jih za $n = 1$. Kot smo že nakazali, bomo v tem delu praobsege označevali s $\mathbb{F}_p$, oznako $\mathbb{Z}_n$ pa bomo rezervirali za končno ciklično grupo moči n.

Multiplikativna grupa obsega $\mathbb{F}_q$ ima moč $q-1$. Zato po Lagrangevem izreku (natančneje, njegovi neposredni posledici) iz teorije grup sledi, da za vsak neničeln element $a \in \mathbb{F}_q$ velja $a^{q-1} = 1$. Potemtakem za vsak element a iz obsega $\mathbb{F}_q$ velja $a^q = a$. Torej je vsak element obsega $\mathbb{F}_q$ ničla polinomske enačbe $X^q - X = 0$. Ker je stopnja te enačbe q, je vsak element obsega $\mathbb{F}_q$ enostavna ničle te enačbe. Zato polinom $X^q - X \in \mathbb{F}_p[X]$ v obsegu $\mathbb{F}_q$ razpade na linearne faktorje. Velja torej

$$X^q - X = \prod_{a \in \mathbb{F}_q} (X - a).$$

Naj bo a neničeln element obsega $\mathbb{F}_q$. Potem iz $a^{q-1} = 1$ sledi $a^{q-2} = a^{-1}$. S tem smo dobili eksplicitno formulo za računanje inverza v obsegu $\mathbb{F}_q$. Pokazati je možno [Vidav, str. 284], da je multiplikativna grupa obsega $\mathbb{F}_q$ ciklična, torej izomorfna grupi $(\mathbb{Z}_{q-1}, +)$. Vseh različnih generatorjev multiplikativne grupe obsega $\mathbb{F}_q$ je zato $\varphi(q-1)$. Pri tem φ označuje številsko-teoretično Eulerjevo funkcijo. Pripomnimo, da bomo v zadnjih dveh poglavjih tega dela črko φ rezervirali za Frobeniusov endomorfizem grupe na eliptični krivulji, kar je v literaturi o eliptičnih krivuljah ustaljena praksa.

1.2 Sled

V tem razdelku bomo obravnavali sled. Ta funkcional ima v teoriji končnih obsegov pomebno vlogo.

Vsak obseg je vektorski prostor nad poljubnim svojim podobsegom. Tako je $\mathbb{F}_{p^n}$ vektorski prostor nad $\mathbb{F}_p$. Dimenzija vektorskega prostora $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$ je n, kar bomo pokazali v nadaljevanju. Opomnimo, da iz tega sledi, da je aditivna grupa obsega $\mathbb{F}_{p^n}$ izomorfna grupi

$$\mathbb{Z}_p \oplus \mathbb{Z}_p \oplus \ldots \oplus \mathbb{Z}_p \qquad (n \text{ faktorjev}).$$

Trditev 1.1 *Preslikava* $Tr : \mathbb{F}_{p^n} \longrightarrow \mathbb{F}_p$, *definirana s predpisom*

$$x \longmapsto x^{p^0} + x^{p^1} + x^{p^2} + \ldots + x^{p^{n-1}},$$

je funkcional na vektorskem prostoru $\mathbb{F}_{p^n}$ *nad* $\mathbb{F}_p$.

Dokaz: Linearnost preslikave Tr je enostavno preveriti. Pokažimo, da ta preslikava res slika v $\mathbb{F}_p$. Izberimo poljuben $x \in \mathbb{F}_{p^n}$. Potem je

$$\Big(\mathrm{Tr}(x)\Big)^p = \left(\sum_{i=0}^{n-1} x^{p^i}\right)^p = \sum_{i=0}^{n-1} x^{p^{i+1}} = \mathrm{Tr}(x). \tag{1.1}$$

Pri zadnjem enačaju smo upoštevali, da v obsegu $\mathbb{F}_{p^n}$ za vsak element x velja $x^q = x$. Ker enačbi $X^p = X$ v $\mathbb{F}_{p^n}$ zadoščajo natanko elementi iz praobsega $\mathbb{F}_p$, iz enačbe (1.1) sledi, da preslikava Tr res slika v $\mathbb{F}_p$. ∎

Definicija 1.2 *Funkcional* Tr *imenujemo* **sled**.

Pripomnimo še, da lahko s podobnim sklepom kot v izpeljavi enačbe (1.1) pokažemo, da za vsak $x \in \mathbb{F}_q$ in vsak $i \in \mathbb{N}$ velja

$$\mathrm{Tr}(x^{p^i}) = \mathrm{Tr}(x). \tag{1.2}$$

Enakosti (1.2) in (1.1) se bosta izkazali za zelo uporabni. Naslednja trditev pa bo zelo pomembna v razdelku 1.6 o rešljivosti kvadratnih enačb v končnih obsegih.

Trditev 1.3 *Sled je neničeln funkcional. Moč njegovega jedra je natanko p^{n-1}.*

Dokaz: Jedro sledi je enako množici ničel polinoma

$$\mathrm{Tr}(X) = \sum_{i=0}^{n-1} X^{p^i},$$

ki je stopnje $p^{n-1} < p^n$. Torej je moč jedra kvečjemu p^{n-1} in je zato sled neničeln funkcional. Vsak neničeln funkcional je surjektiven, zato je zaloga vrednosti sledi enaka $\mathbb{F}_p$ in je torej njena dimenzija enaka 1. Jedro sledi označimo s $\ker \mathrm{Tr}$, zalogo vrednosti pa z $\mathrm{im}\,\mathrm{Tr}$. Potem iz dimenzijske enačbe

$$\dim(\ker \mathrm{Tr}) + \dim(\mathrm{im}\,\mathrm{Tr}) = \dim(\mathbb{F}_{p^n})$$

sledi $\dim(\ker \mathrm{Tr}) = n - 1$. Torej je

$$\ker \mathrm{Tr} \cong \mathbb{Z}_p \oplus \mathbb{Z}_p \oplus \ldots \oplus \mathbb{Z}_p \qquad (n-1 \text{ faktorjev}).$$

Zato je moč jedra sledi enaka p^{n-1}. ∎

Posledica 1.4 *Polinom $\mathrm{Tr}(X) = \sum_{i=0}^{n-1} X^{p^i}$ v $\mathbb{F}_{p^n}$ razpade na linearne faktorje, ki so med seboj paroma različni.*

Dokaz: Stopnja polinoma $\mathrm{Tr}(X)$ je enaka p^{n-1}, po prejšnji trditvi pa ima natanko p^{n-1} različnih ničel. ∎

1.3 Mreža podobsegov končnega obsega

Cilj tega razdelka je popolna karakterizacija vseh podobsegov danega končnega obsega. Karakterizirali ne bomo zgolj izomorfnostnih razredov podobsegov, ampak dejansko vse različne podobsege. To znanje bomo v naslednjem razdelku potrebovali za opis razpadnih obsegov končnih obsegov, ki jih bomo kasneje srečali v sami definiciji eliptičnih krivulj. Opozorimo še, da bomo pri izpeljavi iskane karakterizacije bistveno uporabili osnovni izrek Galoisjeve teorije.

Če imata dva končna obsega različni karakteristiki, očitno nobenega od njiju ni mogoče vložiti v drugega kot podobseg. Dodaten potreben pogoj za obstoj vložitve nam podaja naslednja trditev.

Trditev 1.5 *Dimenzija vektorskega prostora $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$ je enaka n. Za poljuben podobseg K obsega $\mathbb{F}_{p^n}$ velja, da je njegova moč enaka p^m za nek $m \in \mathbb{N}$, za katerega je $m|n$.*

Dokaz: Pokažimo najprej drugi del trditve. Podobseg vsakega obsega ima vedno isto karakteristiko kot celoten obseg. Zato ima tudi K karakteristiko p in ima torej moč oblike p^m za nek $m \in \mathbb{N}$. Vsak obseg je vektorski prostor nad poljubnim svojim podobsegom. Tako je $\mathbb{F}_{p^n}$ vektorski prostor nad K. Ta vektorski prostor je končno dimenzionalen, saj je celoten $\mathbb{F}_{p^n}$ končno ogrodje za samega sebe. Zato ima neko končno dimenzijo d. Torej je vektorski prostor $\mathbb{F}_{p^n}$ nad K izomorfen direktni vsoti d primerkov vektorskega prostora K nad K :

$$\mathbb{F}_{p^n} \cong K \oplus K \oplus \ldots \oplus K \qquad (d \text{ faktorjev}).$$

Izomorfizem je bijekcija, torej ohranja tudi moč množic. Ker so v našem primeru vektorski prostori končne množice, mora veljati

$$p^n = |\mathbb{F}_{p^n}| = |K|^d = p^{md}. \tag{1.3}$$

Torej je $m|n$, s čimer smo pokazali drugi del trditve. V primeru $K = \mathbb{F}_p$ je seveda $m = 1$. Zato nam enačba (1.3) v tem primeru pove, da je $d = n$. S tem smo pokazali tudi prvi del trditve. ∎

Opomnimo, da zgornja trditev (1.5) sledi tudi iz glavnega izreka tega razdelka, to je izreka (1.20). Trditev (1.5) smo na tem mestu navedli predvsem zaradi njenega prvega dela, ki ga bomo v nadaljevanju bistveno potrebovali za izpeljavo izreka (1.20).

Trditev (1.5) torej pravi, da je vsak morebiten podobseg obsega $\mathbb{F}_{p^n}$ izomorfen obsegu moči $\mathbb{F}_{p^m}$, za nek $m|n$. Še vedno pa moramo odgovoriti na naslednji vprašanji:

1. Kakšno moč imajo končni obsegi, ki jih dejansko najdemo med podobsegi danega končnega obsega?

2. Če za nek končen obseg $\mathbb{F}_{p^n}$ obstaja podobseg moči p^k, koliko je vseh različnih podobsegov danega končnega obsega $\mathbb{F}_{p^n}$, ki imajo moč p^k?

Drugo vprašanje je smiselno, ker je moč obsega $\mathbb{F}_{p^n}$ končna. Na ti dve vprašanji nam bo dala odgovor Galoisjeva teorija. Najprej pa se spomnimo dveh definicij iz algebre. Pripomnimo, da konstantnih polinomov po definiciji ne štejemo med nerazcepne polinome.

Definicija 1.6 *Naj bosta K in k komutativna obsega. Naj bo K algebraična razširitev obsega k, kar označimo s $K\backslash k$. Pravimo, da je algebraična razširitev $K\backslash k$* **normalna**, *če je izpolnjen naslednji pogoj: vsak nerazcepen polinom iz kolobarja polinomov $k[X]$, ki ima v K ničlo, v kolobarju polinomov $K[X]$ razpade na linearne faktorje.*

Transcendentne razširitve po definiciji niso normalne. Tako razširitvi $\mathbb{R}\backslash\mathbb{Q}$ in $\mathbb{C}\backslash\mathbb{Q}$ nista normalni. Razširitev $\mathbb{C}\backslash\mathbb{R}$ pa je normalna. Iz definicije normalnosti namreč neposredno sledi, da je v splošnem vsaka algebraična razširitev $K\backslash k$, kjer je K algebraično zaprt obseg, normalna. O algebraično zaprtih obsegih bo govora v razdelku 1.4.

Vsaka razširitev $K\backslash k$, kjer sta K in k končna obsega, je algebraična, saj je vedno končne stopnje, to je, $\dim_k K < \infty$. Pokazati pa je možno še več, namreč da je razširitev $K\backslash k$ za poljubna končna obsega $k \leq K$ normalna [Vidav, str. 284].

Definirajmo še pojem separabilnosti, ločeno za nekonstantne polinome in za razširitve obsegov. Za konstantne polinome pa pojma separabilnosti ne definiramo. Pojem separabilnosti najprej definirajmo za nerazcepne polinome iz $k[X]$, kjer je k nek komutativen obseg. Spomnimo se, da je razpadni obseg danega nerazcepnega polinoma $p \in k[X]$ taka minimalna razširitev obsega k, da polinom p v kolobarju polinomov nad to razširitvijo razpade na linearne faktorje.

Definicija 1.7 *Naj bo k poljuben komutativen obseg. Pravimo, da je nerazcepen polinom $p \in k[X]$* **separabilen**, *če je v svojem razpadnem obsegu brez večkratnih ničel. Za poljuben polinom iz $k[X]$ pravimo, da je* **separabilen**, *če je vsak njegov nerazcepen faktor separabilen.*

Lahko se torej zgodi, da ima separabilen polinom r v svojem razpadnem obsegu tudi večkratne ničle. V tem primeru mora polinom r seveda biti razcepen in veljati mora, da vsak nerazcepen faktor polinoma r k tej stopnji večkratne ničle prispeva kvečjemu eno ničlo. Končno definirajmo še separabilnost razširitve $K \backslash k$.

Definicija 1.8 *Za algebraično razširitev $K \backslash k$ pravimo, da je* **separabilna**, *če za vsak element $a \in K$ velja, da je njegov minimalni polinom $p \in k[X]$ separabilen polinom v $k[X]$. Če razširitev ni algebraična, jo podobno kot pri pojmu normalnosti razglasimo za neseparabilno.*

Pripomnimo, da je minimalni polinom po definiciji vedno nerazcepen. V definiciji separabilnosti razširitev tako potrebujemo definicijo separabilnosti nerazcepnih polinomov.

Preprosto se je prepričati, da so za obsege k karakteristike 0 vsi polinomi iz $k[X]$ separabilni. Torej je vsaka algebraična razširitev obsegov s karakteristiko 0 separabilna. Tako je na primer razširitev $\mathbb{C} \backslash \mathbb{R}$ separabilna. Neseparabilnih algebraičnih razširitev v tem delu ne bomo srečali.

Pojem separabilnosti je neke vrste naravna posplošitev dejstva, da ima vsak nerazcepen polinom druge stopnje iz $\mathbb{R}[X]$ v $\mathbb{C}$ dve različni ničli. V običajnem procesu učenja matematike se vsakdo najprej sreča z reševanjem enačb z realnimi koeficienti, kjer se podzavestno navadi, da imajo v realnem nerešljive enačbe različne kompleksne korene. V splošnem pa ni tako, zato je pojem separabilnosti sprva težje doumljiv. S separabilnostjo sta povezani še naslednji dve trditvi. Dokaz prve od njiju je mogoče najti v [Vidav, str. 178].

Trditev 1.9 *Naj bo k poljuben komutativen obseg in naj bo r poljuben nekonstanten polinom iz $k[X]$. Potem je polinom r v svojem razpadnem obsegu brez večkratnih ničel natanko tedaj, ko sta si polinoma r in njegov odvod r' tuja.* ■

Trditev 1.10 *Naj bo k poljuben komutativen obseg. Nerazcepen polinom r iz $k[X]$ je separabilen natanko takrat, ko je njegov odvod r' neničeln.*

Dokaz: Naj bo r nerazcepen polinom iz $k[X]$. Konstantnih polinomov po dogovoru ne štejemo med nerazcepne, zato je $\deg r \geq 1$. Ker velja $\deg(r') < \deg(r)$ in je r nerazcepen, sta si r in r' tuja natanko tedaj, ko je $r' \neq 0$. Naša trditev zdaj sledi iz trditve (1.9). ■

Izrek 1.11 *Naj bo $q = p^n$, kjer je p praštevilo in $n \in \mathbb{N}$. Vsi nekonstantni polinomi iz $\mathbb{F}_q[X]$ so separabilni. Vsaka algebraična razširitev obsega $\mathbb{F}_q[X]$ je separabilna. Za poljubna končna obsega $k \leq K$ je razširitev $K \backslash k$ separabilna.*

Dokaz: Najprej ugotovimo, kateri nerazcepni polinomi iz $\mathbb{F}_q[X]$ niso separabilni. Po zadnji trditvi (1.10) je potrebno poiskati vse nerazcepne polinome iz $\mathbb{F}_q[X]$ z ničelnim odvodom. Naj bo r tak polinom. Ker je karakteristika obsega $\mathbb{F}_q$ enaka p, iz $r' = 0$ sledi, da so neničelni koeficienti polinoma r lahko kvečjemu pri potencah oblike X^{kp}. Zato velja

$$r(X) = \sum_{k=0}^{m} a_k X^{kp} = \sum_{k=0}^{m} (a_k^{p^{n-1}})^p X^{kp} = \left(\sum_{k=0}^{m} a_k^{p^{n-1}} X^k \right)^p.$$

Pri tem je $m \geq 1$, ker je $\deg r \geq 1$. S tem smo prišli v protislovje z nerazcepnostjo polinoma r. Zato so vsi nerazcepni polinomi iz $\mathbb{F}_q[X]$ separabilni. Potemtakem so vsi nekonstantni polinomi iz $\mathbb{F}_q[X]$ separabilni. Zato je vsaka algebraična razširitev obsega $\mathbb{F}_q[X]$ separabilna. Vsaka razširitev končnega obsega do končnega obsega je algebraična, zato torej tudi separabilna. ∎

Razširitev $\mathbb{F}_{p^n}$ obsega $\mathbb{F}_p$ je normalna in separabilna. Take razširitve imenujemo **Galoisjeve**. Naslednji korak na poti do karakterizacije podobsegov končnega obsega je določitev grupe avtomorfizmov obsega $\mathbb{F}_{p^n}$, ki fiksirajo $\mathbb{F}_p$. To grupo imenujemo **Galoisjeva grupa** razširitve $\mathbb{F}_{p^n} \setminus \mathbb{F}_p$. Označimo jo z G. Iz Galoisjeve teorije vemo, da je moč Galoisjeve grupe vsake Galoisjeve razširitve $K \setminus k$ enaka stopnji te razširitve, torej $\dim_k K$ [Fraleigh, str. 482]. Zato je

$$|G| = \dim \mathbb{F}_{p^n} = n \quad (\text{nad } \mathbb{F}_p).$$

Pokažimo še, da je G ciklična grupa. Dovolj je v G poiskati element reda n. Tak element je Frobeniusov avtomorfizem.

Definicija 1.12 Frobeniusov avtomorfizem *je avtomorfizem obsega $\mathbb{F}_{p^n}$, ki element x preslika v element x^p. Označimo ga s $\mathcal{F}$.*

Trditev 1.13 *Red Frobeniusovega avtomorfizma v Galoisjevi grupi G razširitve $\mathbb{F}_{p^n} \setminus \mathbb{F}_p$ je enak n. Grupa G je ciklična grupa reda n.*

Dokaz: Naj bo θ generator ciklične grupe $\mathbb{F}_{p^n}^*$. Za vsak $k \geq 1$ očitno velja $\mathcal{F}^k(x) = x^{p^k}$. Naj bo za nek $k \geq 1$ preslikava $\mathcal{F}^k$ identiteta na $\mathbb{F}_{p^n}$. Potem v posebnem primeru velja tudi

$$\theta^{p^k} = \mathcal{F}^k(\theta) = \theta.$$

Torej je $\theta^{p^k-1} = 1$. Ker je red elementa θ enak $p^n - 1$, je $k \geq n$. Torej je red Frobeniusovega endomorfizma vsaj n. Ker pa je $|G| = n$, je ta red enak n. Torej je G ciklična grupa reda n, $\mathcal{F}$ pa je eden od njenih $\varphi(n)$ generatorjev. ∎

Za določitev vseh podobsegov danega končnega obsega bomo zdaj uporabili osnovni izrek Galoisjeve teorije [Fraleigh, str. 485]. Ta pravi, da je mreža podgrup grupe G anti-izomorfna mreži vseh podobsegov obsega $\mathbb{F}_{p^n}$. V ta namen potrebujemo klasifikacijo vseh podgrup ciklične grupe moči m.

Lema 1.14 *Podgrupe grupe $(\mathbb{Z}_m, +)$ so natanko podgrupe oblike $\langle d \rangle = \{d, 2d, \ldots, (m/d)d\}$, kjer je d nek delitelj števila m. Pri tem seveda razumemo, da v $\mathbb{Z}_m$ velja $(m/d)d = m = 0$.*

Dokaz: Očitno je za vsak delitelj d števila m množica $\langle d\rangle$ podgrupa grupe $\mathbb{Z}_m$. Pokažimo, da so to natanko vse podgrupe. Naj bo H podgrupa grupe $\mathbb{Z}_m$ in naj bo a najmanjši neničelni element v H. Potem $a|m$, sicer bi lahko v H našli nek neničeln element, ki je manjši od a. Iz istega razloga je vsak element iz H deljiv z a. Sledi, da je $H = \langle a\rangle$. ■

Vse podgrupe ciklične grupe $\mathbb{Z}_m$ so torej ciklične in za vsak delitelj d dobimo natanko eno podgrupo, to je podgrupo, generirano z elementom d. Ta podgrupa ima moč m/d.

Definicija 1.15 *Naj* $(\mathrm{Lat}, \leq)$ *označuje mrežo podobsegov obsega* $\mathbb{F}_{p^n}$.

Poudarimo, da v zgornji definiciji ne gre za mrežo izomorfnostnih razredov podobsegov, ampak dejansko za mrežo vseh podobsegov. Oznaka Lat izhaja iz angleške besede lattice.

Definicija 1.16 *S* $(\mathrm{Sub}, \leq)$ *označimo mrežo podgrup grupe* G.

Oznaka Sub izvira iz angleške besede subgroup.

Definicija 1.17 *Naj* $(\mathcal{M}_n, |)$ *označuje mrežo, katere elementi so vsi naravni delitelji števila* n *(vključno z* n *in* 1*), relacija pa je običajna deljivost števil.*

Po lemi (1.14) je tako mreža $(\mathrm{Sub}, \leq)$ antiizomorfna mreži $(\mathcal{M}_n, |)$. Kanoničen antiizomorfizem mrež $(\mathcal{M}_n, |) \longrightarrow (\mathrm{Sub}, \leq)$ številu k, kjer je $k|n$, priredi podgrupo grupe G, generirano s $\mathcal{F}^k$. Torej številu n ustreza podgrupa $\{\mathrm{id}\}$, številu 1 pa ustreza podgrupa G.

Po osnovnem izreku Galoisjeve teorije je mreža $(\mathrm{Lat}, \leq)$ antiizomorfna mreži $(\mathrm{Sub}, \leq)$. Kanonični antiizomorfizem mrež $(\mathrm{Sub}, \leq) \longrightarrow (\mathrm{Lat}, \leq)$ podgrupi $H \leq G$ priredi obseg fiksnih točk grupe H, to je obseg

$$\mathbb{F}_{p^n}^H = \{x \in \mathbb{F}_{p^n} \mid \quad g(x) = x \text{ za vsak element } g \in H\}.$$

Seveda velja $\mathbb{F}_p \leq \mathbb{F}_{p^n}^H \leq \mathbb{F}_{p^n}$. Podgrupi $\langle \mathcal{F}^d\rangle$, kjer je $d|n$, torej priredimo podobseg $\mathbb{F}_{p^n}^{\mathcal{F}^d}$. Podgrupi $\{\mathrm{id}\}$ tako ustreza podobseg $\mathbb{F}_{p^n}^{\{\mathrm{id}\}} = \mathbb{F}_{p^n}$, podgrupi G pa ustreza podobseg $\mathbb{F}_{p^n}^G = \mathbb{F}_p$.

Pokažimo, da ima podobseg $\mathbb{F}_{p^n}^{\mathcal{F}^d}$ natanko p^d elementov. V ta namen najprej pokažimo naslednjo lemo, od tod pa potem želeno trditev.

Lema 1.18 *Množica rešitev enačbe* $X^{p^d} = X$ *v* $\mathbb{F}_{p^n}$ *ima natanko* p^d *elementov.*

Dokaz: Očitno ima ta enačba kvečjemu p^d rešitev. Pokažimo, da jih ima v $\mathbb{F}_{p^n}$ natanko p^d. Naj bo θ generator ciklične grupe $\mathbb{F}_{p^n}^*$. Zapišimo

$$p^n - 1 = p^{\frac{n}{d}d} - 1 = \left(p^d - 1\right)\left(1 + \ldots + (p^d)^{\frac{n}{d}-1}\right).$$

Torej $p^d - 1 \,|\, p^n - 1$. Označimo $\kappa = \theta^{(p^n-1)/(p^d-1)}$. Ker je red elementa θ enak $p^n - 1$, so $1, \kappa, \kappa^2, \ldots, \kappa^{p^d-2}$ paroma različni elementi, ki vsi zadoščajo enačbi $X^{p^d-1} = 1$. Če jim dodamo še element 0, dobimo v obsegu $\mathbb{F}_{p^n}$ vsaj p^d paroma različnih elementov, ki zadoščajo enačbi $X^{p^d} = X$. ■

Trditev 1.19 *Velja*

$$\left|\mathbb{F}_{p^n}^{\langle\mathcal{F}^d\rangle}\right| = p^d.$$

Dokaz: Preprosto je videti, da je v definiciji $\mathbb{F}_{p^n}^{\langle\mathcal{F}^d\rangle}$ dovolj opazovati, kaj fiksira generator Galoisjeve grupe, da je torej

$$\mathbb{F}_{p^n}^{\langle\mathcal{F}^d\rangle} = \{x \in \mathbb{F}_{p^n} \mid \quad \mathcal{F}^d(x) = x\}. \tag{1.4}$$

Potemtakem obseg $\mathbb{F}_{p^n}^{\langle\mathcal{F}^d\rangle}$ sovpada z množico rešitev enačbe $X^{p^d} = X$ v $\mathbb{F}_{p^n}$. Po lemi (1.18) pa ima ta natanko p^d elementov. ∎

Zdaj lahko izpolnimo cilj tega razdelka.

Izrek 1.20 *Podobsegi obsega $\mathbb{F}_{p^n}$ so izomorfni natanko obsegom $\mathbb{F}_{p^d}$, za katere je $d|n$. Za vsak tak d obstaja v $\mathbb{F}_{p^n}$ natanko en podobseg moči $\mathbb{F}_{p^d}$. Pri tem za vsak $x \in \mathbb{F}_{p^n}$ velja*

$$x \in \mathbb{F}_{p^d} \iff x^{p^d} = x.$$

Dokaz: Če komponiramo antiizomorfizma $(\mathcal{M}_n, |) \longrightarrow (\mathrm{Sub}, \leq)$ in $(\mathrm{Sub}, \leq) \longrightarrow (\mathrm{Lat}, \leq)$, dobimo izomorfizem mrež $(\mathcal{M}_n, |)$ in $(\mathrm{Lat}, \leq)$. Številu d, kjer je $d|n$, pri tej bijektivni korespondenci ustreza podobseg $\mathbb{F}_{p^n}^{\langle\mathcal{F}^d\rangle}$, ki ima moč p^d. Po drugi strani pa je vsak podobseg obsega $\mathbb{F}_{p^n}$ take oblike, torej je njegova moč enaka p^d za neko število d, ki deli n. Še več, dva različna podobsega zaradi bijektivnosti korespondence ne moreta deliti istega d. Zato za vsako število d, ki deli n, obstaja v $\mathbb{F}_{p^n}$ natanko en podobseg moči p^d. Zadnji del izreka sledi potem sledi iz enakosti (1.4). ∎

1.4 Razpadni obsegi

Eliptičnih krivulj ne moremo definirati, če ne poznamo pojma razpadnega obsega. Kljub temu je ta pojem v kriptografski literaturi, ki je orientirana predvsem proti praktični uporabi, pogosto omenjen na hitro in brez posebne razlage ali komentarjev. Ta razdelek poskuša odpraviti ravno to pomankljivost. Končni rezultat tega razdelka je konstrukcija razpadnih obsegov končnih obsegov. V ta namen najprej navedemo nekaj splošne algebraične teorije. Vseeno pa pripomnimo, da je možno eliptične krivulje in Schoofov algoritem zadovoljivo razumeti tudi brez posebnega znanja o razpadnih obsegih končnih obsegov. Tak pristop ubere večina knjig.

1.4.1 Splošno o razpadnih obsegih

Definicija 1.21 *Naj bo K poljuben komutativen obseg. Pravimo, da je obseg K* **algebraično zaprt***, če vsak nekonstanten polinom iz kolobarja polinomov $K[X]$ v kolobarju $K[X]$ razpade na linearne polinome.*

Z drugimi besedami, v algebraično zaprtih obsegih K je vsak polinom stopnje večje ali enake 2 razcepen element glavnega kolobarja $K[X]$. Povedano na še drugačen način, vsak nekonstanten polinom iz $K[X]$ ima ničlo v K. Definirajmo še pojem algebraičnega zaprtja obsega.

Definicija 1.22 *Vsako algebraično razširitev komutativnega obsega K, ki je algebraično zaprta, imenujemo* **algebraično zaprtje** *komutativnega obsega K. Drug izraz za algebraično zaprtje je tudi* **razpadni obseg**.

Pokazati je možno [Fraleigh, str. 418-422], da algebraično zaprtje obstaja za vsak komutativen obseg in da so za dani komutativni obseg K vsa algebraična zaprtja med seboj izomorfna. Zato za algebraično zaprtje komutativnega obsega K uvedemo enotno oznako $\overline{K}$. Izomorfizem med dvema algebraičnima zaprtjima istega komutativnega obsega lahko vedno izberemo tako, da je na K identiteta.

Algebraično zaprti obsegi imajo pomembno lastnost, da so v nekem smislu maksimalni. Natančneje to opišeta naslednji dve trditvi.

Trditev 1.23 *Algebraično zaprti obsegi nimajo netrivialnih algebraičnih razširitev.*

Dokaz: Naj bo namreč $L \leq M$, kjer je L algebraično zaprt obseg in M neka njegova razširitev in naj bo $a \in M$. Ker je minimalni polinom elementa a nerazcepen element kolobarja $L[X]$, obseg L pa je algebraično zaprt, mora biti ta minimalni polinom stopnje 1. To pa pomeni, da je $a \in L$. Sledi $L = M$. ■

Trditev 1.24 *Vsaka algebraična razširitev danega komutativnega obsega K ima isti razpadni obseg kot K. Vsaka algebraična razširitev danega komutativnega obsega K je (do izomorfizma natančno) vsebovana v algebraičnem zaprtju $\overline{K}$.*

Dokaz: Naj bo L poljubna algebraična razširitev komutativnega obsega K. Potem je $\overline{L}$ algebraična razširitev algebraične razširitve komutativnega obsega K, torej je razširitev $\overline{L}\backslash K$ algebraična. Sledi, da je $\overline{L}$ razpadni obseg komutativnega obsega K. To pomeni, da je $L \leq \overline{L} = \overline{K}$. ■

Iz trditev (1.23) in (1.24) sledi, da je algebraično zaprtje maksimum mreže vseh algebraičnih razširitev komutativnega obsega K.

1.4.2 Razpadni obsegi končnih obsegov

Kot smo že omenili, so razpadni obsegi končnih obsegov pomembni za kriptografijo, ker so eliptične krivulje, ki jih najpogosteje srečamo v praksi, definirane nad algebraičnimi zaprtji končnih obsegov. Pri računanju in računalniški implementaciji se sicer omejimo na t.i. podgrupo racionalnih točk, to je točk na krivulji, katerih koordinati sta že kar v samem končnem obsegu. Vseeno pa veliko teoretičnih razmislekov o eliptičnih krivuljah, ki imajo daljnosežne posledice tudi za podgrupo racionalnih točk in posledično za praktično uporabo, ne more potekati brez razpadnih obsegov končnih obsegov. V Schoofovem algoritmu imajo na primer veliko vlogo torzijske podgrupe praštevilskega reda. Točke teh podgrup imajo v splošnem koordinate v razpadnem obsegu končnega obsega.

Opozorimo, da bomo v tem razdelku poskušali nazorno prikazati konstrukcijo razpadnih obsegov končnih obsegov. Strogo formalno konstrukcijo pa bomo nakazali na koncu razdelka.

Trditev 1.25 *Končni obsegi niso algrebraično zaprti.*

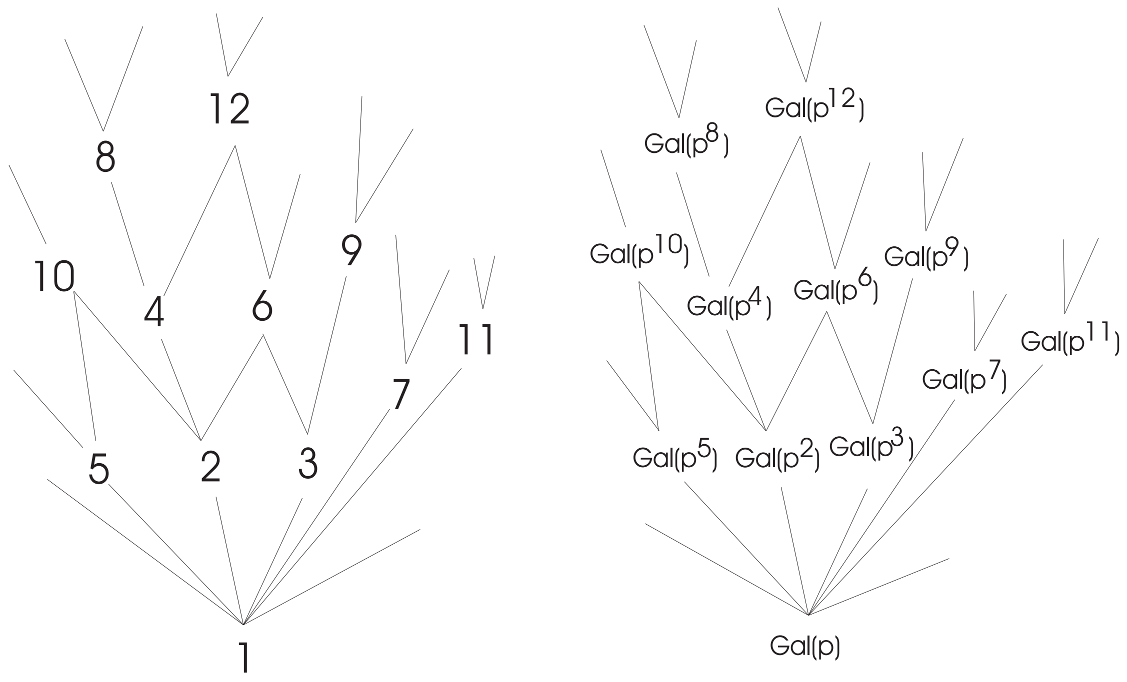
Dokaz: Polinom
$$1+\prod_{a_i\in\mathbb{F}_{p^n}}(X-a_i)$$
očitno nima ničle v $\mathbb{F}_{p^n}$. ■

Možno je pokazati še več: za vsak obseg $\mathbb{F}_{p^n}$ obstaja nerazcepen polinom iz $\mathbb{F}_{p^n}[X]$, ki ima poljubno veliko stopnjo [Lidl-Niederreiter, str. 51].

Za vsak $q=p^n$ obstaja algebraično zaprtje obsega $\mathbb{F}_q$, ki ga označimo s $\overline{\mathbb{F}_q}$. Ker je $\overline{\mathbb{F}_q}$ razširitev obsega $\mathbb{F}_q$, je karakteristika obsega $\overline{\mathbb{F}_q}$ enaka p. Obseg $\overline{\mathbb{F}_q}$ ni končen, je pa komutativen. Komutativnost bomo pokazali v nadaljevanju. Ker je obseg $\mathbb{F}_{p^n}$ algebraična razširitev obsega $\mathbb{F}_p$, ima po trditvi (1.24) praobseg $\mathbb{F}_p$ isti razpadni obseg kot vsak končen obseg $\mathbb{F}_{p^n}$. Torej velja $\overline{\mathbb{F}_p}=\overline{\mathbb{F}_{p^n}}$ in je za vsako karakteristiko p zato potrebno konstruirati le algebraično zaprtje praobsega $\mathbb{F}_p$.
V razdelku 1.3 o podobsegih končnega obsega smo videli, da je mreža podobsegov obsega $\mathbb{F}_{p^n}$ pravzaprav iste oblike kot mreža $\mathcal{M}_n$ (definicija (1.17)). Mrežo $\mathcal{M}_n$ lahko enostavno povečamo do večje mreže $(\mathbb{N},|)$. Premislimo, če lahko v primeru končnih obsegov storimo analogno spremembo. Obseg $\mathbb{F}_{p^n}$ želimo torej povečati do nekega večjega obsega F. Hočemo, da bo temu obsegu ustrezala mreža $(\mathbb{N},|)$ na enak način, kot obsegu $\mathbb{F}_{p^n}$ ustreza mreža $\mathcal{M}_n$. Mreža $(\mathbb{N},|)$ ima neskončno elementov. Vseeno si poskusimo ustvariti sliko, zato mrežo narišemo na list papirja v obliki Hassejevega diagrama. Čisto na dnu je element 1, nad njim pa potem še preostala naravna števila. Ko potujemo navzgor, postane drevo zaradi zapletenosti relacije deljivosti popolnoma nepregledno. Zdaj iz tega diagrama izbrišimo vsa števila, veje pa ohranimo. Na mesto, kjer je prej stalo naravno število n, zdaj napišimo $\mathbb{F}_{p^n}$. To storimo za vsa naravna števila n. Dobimo podobno drevo, kot prej pri podobsegih danega končnega obsega, le da to drevo zdaj vsebuje vse končne obsege karakteristike p in ne le nekaterih. Dobljeno drevo vsebuje kot poddrevesa drevesa podobsegov za vse končne obsege karakteristike p. Neskončno drevo, ki smo ga našli, je pravzaprav Hassejev diagram za mrežo podobsegov iskanega razpadnega obsega F praobsega $\mathbb{F}_p$. Ta Hassejev diagram je skupaj z mrežo $(\mathbb{N},|)$ ilustriran na sliki 1.1.

Obseg F si intuitivno predstavljamo kot unijo vseh končnih obsegov karakteristike p. Pri tej uniji razumemo, da so nekateri obsegi podobsegi drugih obsegov, kot smo ugotovili v razdelku 1.3 o mreži podobsegov končnega obsega. Torej elementi vsakega obsega $\mathbb{F}_{p^n}$ v tej uniji prekrijejo elemente vseh podobsegov obsega $\mathbb{F}_{p^n}$.

Navedimo zdaj nekaj osnovnih lastnosti razpadnih obsegov. Poljubna dva obsega $\mathbb{F}_{p^n}$ in $\mathbb{F}_{p^m}$ je možno vložiti v neko skupno razširitev. Najmanjša taka razširitev je obseg $\mathbb{F}_{p^{\mathrm{lcm}(n,m)}}$. V mreži $\mathrm{Lat}(\mathbb{F}_{p^{\mathrm{lcm}(n,m)}})$ velja
$$\mathbb{F}_{p^n}\cup\mathbb{F}_{p^m}=\mathbb{F}_{p^{\mathrm{lcm}(n,m)}},$$
kar je analog enakosti $n\cup m=\mathrm{lcm}(n,m)$ v mreži $\mathcal{M}_{\mathrm{lcm}(n,m)}$. Podobno je analog zveze $n\cap m=\gcd(n,m)$ zveza
$$\mathbb{F}_{p^n}\cap\mathbb{F}_{p^m}=\mathbb{F}_{p^{\gcd(n,m)}}.$$
Torej sta v F obsega $\mathbb{F}_{p^n}$ in $\mathbb{F}_{p^m}$ vsebovana v obsegu $\mathbb{F}_{p^{\mathrm{lcm}(n,m)}}$. Na ta način lahko elemente obsega $\mathbb{F}_{p^n}$ seštevamo in množimo z elementi obsega $\mathbb{F}_{p^m}$.



Slika 1.1 Mreža naravnih števil za relacijo deljivosti in njej izomorfna mreža podobsegov razpadnega obsega $\overline{\mathbb{F}_p}$. Končne obsege $\mathbb{F}_{p^n}$ smo zaradi preglednosti tokrat označili z $\mathrm{Gal}(p^n)$.

Mreža podobsegov obsega F je izomorfna mreži $(\mathbb{N}, |)$, kjer $|$ označuje relacijo deljivosti naravnih števil. Vsakemu naravnemu številu d ustreza natanko en podobseg obsega F. Ta podobseg je izomorfen obsegu $\mathbb{F}_{p^d}$. Vsak končen obseg iz F je vsebovan v nekem večjem končnem obsegu. V F torej lahko najdemo neskončne naraščajoče verige podobsegov, podobno kot lahko najdemo neskončne verige naravnih števil za relacijo deljivosti. Preprosto karakterizacijo podobsegov obsega F nam podaja naslednja trditev.

Trditev 1.26 *Za vsak element x iz konstruiranega obsega F velja*

$$x \in \mathbb{F}_{p^n} \iff x^{p^n} = x.$$

Dokaz: To je enostavna posledica izreka (1.20). ∎

Trditev 1.27 *Razširitev $F \backslash \mathbb{F}_p$ je algebraična.*

Dokaz: Za vsak element a obsega F obstaja dovolj velik končen obseg $\mathbb{F}_{p^n}$, da je $a \in \mathbb{F}_{p^n}$ $(\subset F)$. Ker je razširitev $\mathbb{F}_{p^n} \backslash \mathbb{F}_p$ algebraična, je a algebraičen element obsega F nad $\mathbb{F}_p$. Torej je razširitev $F \backslash \mathbb{F}_p$ algebraična. ∎

Trditev 1.28 *Konstruirani obseg F je algebraično zaprt.*

Dokaz: Naj bo $r = \sum_{i=0}^{s} a_i X^i \in F[X]$ nerazcepen polinom. Pokažimo, da je stopnja polinoma r kvečjemu ena. Vsak koeficient a_i polinoma r je vsebovan v nekem dovolj velikem končnem obsegu $\mathbb{F}_{p^{n_i}}$. Sledi, da so vsi koeficienti polinoma vsebovani v nekem dovolj velikem končnem obsegu, namreč v obsegu $\mathbb{F}_{p^v}$, kjer je $v = \mathrm{lcm}(n_0, n_1, \ldots, n_s)$. Na

r torej lahko gledamo kot na polinom iz $\mathbb{F}_{p^v}[X]$. Polinom r je nerazcepen tudi v $\mathbb{F}_{p^v}[X]$. V nasprotnem primeru bi bil razcepen tudi v $F[X]$, kar pa po predpostavki ni.

Vsak nerazcepen polinom $t \in \mathbb{F}_{p^v}[X]$ stopnje s razpade v obsegu $\mathbb{F}_{p^{vs}}$ na linearne faktorje. Če namreč tvorimo faktorski kolobar $\mathbb{F}_{p^v}[X]/(t)$, dobimo obseg moči p^{vs}, v katerem ima polinom t očitno vsaj eno ničlo. Ker je vsaka razširitev končnega obsega do nekega (večjega) končnega obsega normalna in separabilna, sledi, da ima polinom t v $\mathbb{F}_{p^{vs}}$ natanko s različnih ničel. Torej r razpade v $\mathbb{F}_{p^{vs}}[X]$ na linearne faktorje. Ker je $\mathbb{F}_{p^{vs}} \subset F$, polinom r v $F[X]$ razpade na linearne faktorje. ∎

Zgornji opis obsega F je sicer osnova za dobro predstavo, vendar strogo formalno ni povsem korekten. Orišimo še definicijo obsega F, ki je strogo formalno korektna. Ta poteka preko pojma direktne limite v kategoriji obsegov. Obseg F, ki ga bomo na ta način definirali, je isti kot zgoraj intuitivno opisani obseg F.

Obseg $\mathbb{F}_{p^n}$ je za vsak $k \in \mathbb{N}$ možno naravno vložiti v obseg $\mathbb{F}_{p^{kn}}$. Ustrezno inkluzijsko preslikavo označimo z $i_{n,kn}$. Obseg F definirajmo kot direktno limito sistema

$$(\mathbb{F}_p, \mathbb{F}_{p^2}, \mathbb{F}_{p^3}, \ldots),$$

kjer med pari, kjer je en obseg podobseg drugega, deluje ustrezna inkluzijska preslikava $i_{n,kn}$. Direktna limita je splošno znan algebraičen pojem, katerega natančen opis je naveden v mnogih knjigah o algebri. Njegova konstrukcija poteka v osnovi tako, da tvorimo disjunktno unijo

$$W = \coprod_{n \in \mathbb{N}} \mathbb{F}_{p^n}.$$

Nanjo uvedemo ekvivalenčno relacijo $\sim$, podano z naslednjim predpisom. Elementa $x \in \mathbb{F}_{p^n}$ in $y \in \mathbb{F}_{p^m}$ naj bosta ekvivalentna, če je eden od $\mathbb{F}_{p^n}, \mathbb{F}_{p^m}$ podobseg drugega in pri vložitvi postaneta x in y ista elementa večjega od $\mathbb{F}_{p^n}, \mathbb{F}_{p^m}$. Potem definiramo $F = W/\sim$. Na F uvedemo strukturo obsega po naslednjem pravilu: dva elementa iz F seštejemo tako, da poiščemo tak n, da imata oba elementa v $\mathbb{F}_{p^n}$ predstavnika v svojih ekvivalenčnih razredih. Ta predstavnika v $\mathbb{F}_{p^n}$ seštejemo. Ekvivalenčni razred vsote proglasimo za vsoto originalnih elementov. Enako pravilo velja za množenje. Pokazati se da, da je ta definicija nedvoumna in da res dobimo obseg.

1.5 Polinomske in normalne baze

Če želimo v praksi računati z elementi obsega $\mathbb{F}_{p^n}$, moramo te elemente v računalniku na nek način predstaviti. To je cilj tega razdelka. Znanje iz tega razdelka je izrednega pomena za eliptično kriptografijo. Nujno potrebno je tudi za praktično izvedbo Schoofovega algoritma. Polinomske in normalne baze in njihovo praktično realizacijo na računalnikih izčrpno opisuje ameriški standard IEEE P1363.

Vsak končen obseg $\mathbb{F}_{p^n}$ je vektorski prostor nad obsegom $\mathbb{F}_p$. Predstavitev obsega $\mathbb{F}_{p^n}$ običajno poiščemo tako, da izberemo neko bazo za vektorski prostor $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$. Vsaka baza tega vektorskega prostora ima seveda natanko n elementov. Potem poljuben element obsega $\mathbb{F}_{p^n}$ v računalniku predstavimo tako, da shranimo n koeficientov razvoja tega elementa po izbrani bazi. Vseh različnih urejenih baz vektorskega prostora $\mathbb{F}_{p^n}$ nad

$\mathbb{F}_p$ je toliko, kot je obrnljivih matrik v matrični algebri $\mathbb{F}_p^{n,n}$. To število [Menezes, str. 3] je enako

$$(p^n - 1)(p^n - p) \cdots (p^n - p^{n-1}).$$

Vsak element obsega tako v računalniku postane neka urejena n-terica koeficientov iz $\{0, 1, \ldots, p-1\}$. Torej moramo biti najprej sposobni v računalniku shranjevati števila $\{0, 1, \ldots, p-1\}$. Če je p majhen, s tem ni težav. Če pa je p veliko praštevilo, postane že to opravilo samo zase zahtevno. V kriptografski praksi najpogosteje delamo z naslednjimi obsegi:

1. Obsegi oblike $\mathbb{F}_{2^n}$, kjer je n veliko naravno število.
   Pri binarnih eliptičnih krivuljah, ki se uporabljajo v praksi za resne namene, se danes (leto 2000) eksponent n giblje med 100 in 1000. Ker računalniki postajajo vedno zmogljivejši, predvsem pa, ker postajajo algoritmi vedno učinkovitejši, to število n dokaj hitro raste.

2. Obsegi oblike $\mathbb{F}_p$, kjer je p veliko praštevilo.
   Število bitov, potrebnih za zapis števila p, se pri resnih kriptografskih uporabah giblje v podobnem območju, kot se giblje n v zgornjem primeru.

Ta razdelek se ukvarja s predstavitvijo obsega $\mathbb{F}_{p^n}$ s pomočjo elementov obsega $\mathbb{F}_p$. V primeru $n = 1$ seveda na ta način nič ne pridobimo. Torej se ta razdelek ne nanaša na zgornji primer 2. V tem primeru gre pravzaprav za izvedbo mnogomestne celoštevilske aritmetike, torej računanja z velikimi naravnimi števili. Nekaj o tem lahko najdemo v [Blake, poglavje 2].

Izbor baze je v veliki meri odvisen od tega, kaj bomo z elementi obsega $\mathbb{F}_{p^n}$ počeli. Pomembno je, katere operacije bomo izvajali in kako pogosto. Na podlagi izbora posamezne baze dobimo eksplicitna pravila, kako sešteti dve n-terici in kako ju zmnožiti. Ta pravila nam torej podajajo realizacijo seštevanja in množenja v $\mathbb{F}_{p^n}$. Koristno je, če ima baza, na podlagi katere predstavimo $\mathbb{F}_{p^n}$ kot n-terice števil iz $\mathbb{F}_p$, ugodne aritmetične lastnosti. To pomeni, da je običajne operacije v obsegu $\mathbb{F}_{p^n}$, kot so seštevanje elementov, množenje elementov in invertiranje elementov možno izvajati čim hitreje. Nekatere poddružine baz se odlikujejo s posebno preprostostjo in ugodnimi aritmetičnimi lastnostmi. Ena družina takih baz so polinomske baze.

1.5.1 Polinomske baze

Polinomske baze so za kriptografske namene tradicionalno najpogosteje uporabljane baze. Na voljo so v vsaki karakteristiki p in za vsak obseg $\mathbb{F}_{p^n}$, o čemer bomo več povedali v nadaljevanju. V moderni kriptografiji je najzanimivejši primer, ko je karakteristika enaka 2 in je n velik. V tem primeru so koeficienti a_i, ki predstavljajo nek element $a \in \mathbb{F}_{2^n}$, števila 0 in 1. Predstavitev v polinomski bazi obsega $\mathbb{F}_{2^n}$ torej poteka v naravnem jeziku računalnikov, v binarni kodi. Druge karakteristike se v praksi pojavljajo redkeje. Razlog za to je, da se že sami obsegi $\mathbb{F}_{p^n}$, kjer je $p \geq 3$ in $n > 1$, za kriptografske namene redkeje uporabljajo, torej tudi ni praktične potrebe po njihovih bazah. Vendar njihova teoretična obravnava ni bistveno težja od primera karakteristike 2, zato bomo v tem delu obravnavali polinomske baze za vse karakteristike p.

Definicija 1.29 *Podmnožica vektorskega prostora $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$ se imenuje* **polinomska baza**, *če je oblike*

$$\mathcal{P} = \{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\},$$

kjer je α neka ničla nekega nerazcepnega polinoma r stopnje n iz $\mathbb{F}_p[X]$.

Uporabo izraza baza v zgornji definiciji polinomske baze bo upravičila trditev (1.31). Vsak nerazcepen polinom iz kolobarja $\mathbb{F}_p[X]$ stopnje n nam torej za vsako svojo ničlo podaja neko bazo za vektorski prostor $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$. Za različne nerazcepne polinome dobimo v splošnem seveda različne baze za isti vektorski prostor $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$. V nadaljevanju bomo pokazali, kako za dan obseg $\mathbb{F}_{p^n}$ poiskati tak nerazcepen polinom stopnje n iz $\mathbb{F}_p[X]$, da bo dobljena polinomska baza imela čim lepše aritmetične lastnosti.

Lema 1.30 *Polinom r iz definicije polinomske baze je minimalni polinom elementa α.*

Dokaz: Minimalni polinom elementa α mora deliti nerazcepni polinom r, to pa je možno le, če je r kar enak minimalnemu polinomu. ∎

Trditev 1.31 *Množica $\mathcal{P}$ iz definicije (1.29) je baza za vektorski prostor $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$.*

Dokaz: Zaradi leme (1.30) element α ni ničla nobenega netrivialnega polinoma iz $\mathbb{F}_p[X]$, ki ima stopnjo manjšo ali enako n. Zato so elementi iz $\mathcal{P}$ paroma različni. Iz istega razloga velja še, da je $\mathcal{P}$ linearno neodvisna množica. Ker ima moč n, je baza. ∎

Omenili smo že, da za vsak $n \in \mathbb{N}$ obstaja nerazcepen polinom $r \in \mathbb{F}_p[X]$ stopnje n [Lidl-Niederreiter, str. 51]. Torej polinomske baze res obstajajo za vsako karakteristiko p in za vsak $n \in \mathbb{N}$. Poljuben element $a \in \mathbb{F}_{p^n}$ lahko razvijemo po bazi:

$$a = a_0 + a_1\alpha + a_2\alpha^2 + \ldots + a_{n-1}\alpha^{n-1}, \qquad a_i \in \mathbb{F}_p. \tag{1.5}$$

Dva elementa $a = \sum_{i=0}^{n-1} a_i\alpha^i$ in $b = \sum_{i=0}^{n-1} b_i\beta^i$ seštejemo tako, da v $\mathbb{F}_p$ seštejemo istoležne koeficiente. V primeru $p = 2$ je to v računalniku preprosto izvesti, saj gre za običajen logični XOR dveh n-bitnih spremenljivk.

Množenje v $\mathbb{F}_{p^n}$ je v računalniku nekoliko težje izvesti. Opišimo algebraično idejo, ki bo pojasnila, od kje pravzaprav izvira definicija polinomskih baz. S tem bomo dobili tudi dober opis množenja v obsegu $\mathbb{F}_{p^n}$. Ker je $\mathbb{F}_p$ obseg, je $\mathbb{F}_p[X]$ glavni kolobar. Ideal, generiran z nerazcepnim elementom, je v glavnem kolobarju vedno maksimalni. Za nerazcepni polinom r iz definicije polinomske baze je potemtakem faktorski kolobar $\mathbb{F}_p[X]/(r)$ obseg. Moč tega obsega je $p^{\deg(r)} = p^n$, torej je ta obseg izomorfen $\mathbb{F}_{p^n}$. Za dano polinomsko bazo obsega $\mathbb{F}_{p^n}$, določeno z nerazcepnim polinomom r in njegovo ničlo α, obstaja naravni izomorfizem obsegov $\mathbb{F}_{p^n}$ in $\mathbb{F}_p[X]/(r)$. Ta izomorfizem je natančno določen s predpisom $\alpha \mapsto X + (r)$. V smislu tega izomorfizma so elementi obsega $\mathbb{F}_{p^n}$ polinomi iz $\mathbb{F}_p[X]$, ki imajo stopnjo manjšo od n. To pravzaprav narekuje že enačba (1.5), v kateri moramo le črko α zamenjati z X. Zgoraj opisano seštevanje v tej luči postane kar običajno seštevanje polinomov. Množenje elementov obsega $\mathbb{F}_{p^n}$ pa predstavimo kot običajno množenje polinomov, s tem, da produkt okrajšamo modulo r.

Za hitro izvedbo množenja moramo poleg množenja polinomov s koeficienti iz $\mathbb{F}_p$ znati učinkovito izvesti tudi redukcijo danega polinoma po modulu polinoma r. Zato je ugodno, če je polinom r, ki podaja posamezno polinomsko bazo obsega $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$, čim preprostejši. Stopnja nerazcepnega polinoma r sicer mora biti n, tako da, vsaj kar se stopnje tiče, poenostavitve niso možne. Lahko pa poskusimo poiskati tak polinom r, ki ima čim manj členov, torej čim več ničelnih koeficientov. V kolobarju $\mathbb{F}_2[X]$ ne obstajajo nerazcepni polinomi, ki imajo natanko sodo mnogo neničelnih členov. Za vsak tak polinom je namreč element 1 ničla, zato je tak polinom razcepen.

Definicija 1.32 *Polinome oblike* $r(X) = X^n + X^k + 1 \in \mathbb{F}_2[X]$, *kjer je* $1 \leq k \leq n-1$, *imenujemo* **trinomi**. *Podobno polinome iz* $\mathbb{F}_2[X]$, *ki imajo natanko pet neničelnih členov, imenujemo* **pentomi**.

Pripomnimo, da stopnja polinoma ni neposredno povezana s pojmoma trinom in pentom. Trinomi in pentomi imajo lahko poljubno veliko stopnjo. V primeru karakteristike 2 so trinomi najpreprostejši kandidati za nerazcepne polinome r. Na žalost pa za nekatere n v kolobarju $\mathbb{F}_2[X]$ ni nerazcepnih trinomov stopnje n. Empirični rezultati kažejo, da je kar približno polovica vseh naravnih števil n takih, s teoretičnega vidika pa vprašanje eksistence nerazcepnih trinomov še ni razjasnjeno [Blake, str. 19]. Kadar za določen n ne obstaja nerazcepni trinom, si lahko pomagamo tako, da poiščemo nerazcepen pentom stopnje n. Empirični rezultati kažejo, da za vrednosti $n < 10000$ v primeru, ko ne obstaja nerazcepen trinom stopnje n, vedno obstaja nerazcepen pentom stopnje n [Blake, str. 19]. Pri današnjem stanju tehnike območje $n < 10000$ zadostuje za večino praktičnih potreb. V knjigi [Handbook, str. 158,159] je možno najti tabelo, ki navede tiste $n \leq 1478$, za katere v $\mathbb{F}_2[X]$ obstaja nerazcepen trinom oblike $X^n + X^k + 1$, kjer je $1 \leq k \leq n-1$. Za vsak tak n je navedeno še najmanjše tako število k, $1 \leq k \leq n-1$, da je trinom $X^n + X^k + 1$ nerazcepen. Pripomnimo, da je enostavno pokazati, da iz nerazcepnosti polinoma $X^n + X^k + 1$ sledi nerazcepnost polinoma $X^n + X^{n-k} + 1$.

1.5.2 Normalne baze

Polinomske baze imajo hudega tekmeca: normalne baze. Podobno kot polinomske baze so tudi normalne baze posebna družina baz za vektorski prostor $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$, kjer je n naravno število, p pa praštevilo. Opomnimo, da pogosto govorimo o bazah končnega obsega $\mathbb{F}_{p^n}$, pri tem pa imamo vedno v mislih bazo za vektorski prostor $\mathbb{F}_{p^n}$ nad $\mathbb{F}_p$.

Definicija 1.33 *Podmnožica* $\mathcal{N}$ *vektorskega prostora* $\mathbb{F}_{p^n}$ *nad* $\mathbb{F}_p$ *se imenuje* **normalna baza**, *če sta izpolnjeni zahtevi:*

1. $\mathcal{N}$ *je baza za vektorski prostor* $\mathbb{F}_{p^n}$ *nad* $\mathbb{F}_p$,
2. $\mathcal{N} = \{\theta^{p^0}, \theta^{p^1}, \theta^{p^2}, \ldots, \theta^{p^{n-1}}\}$ *za nek element* θ *obsega* $\mathbb{F}_{p^n}$.

V kriptografski praksi se uporablja skoraj brez izjeme primer $p = 2$. Definicijo za splošen p smo navedli le, ker je definicijo za primer $p = 2$ možno enostavno posplošiti na splošen p.

Definicija 1.34 *Za element* θ *iz zgornje točke 2. rečemo, da* **generira normalno bazo**. *Imenujemo ga* **normalen element**.

V obsegu $\mathbb{F}_{p^n}$ je lahko več normalnih elementov. Če je θ normalen element, generirata za vsak $i \in \{0, \ldots, n-1\}$ elementa θ in θ^{p^i} isto normalno bazo, saj v obsegu $\mathbb{F}_{p^n}$ velja enakost $\theta^{p^n} = \theta$. Množica $\{\theta^{p^0}, \theta^{p^1}, \theta^{p^2}, \ldots, \theta^{p^{n-1}}\}$ je lahko za nekatere $\theta \in \mathbb{F}_{p^n}$ linearno odvisna in torej ni vedno baza. Taka sta, na primer, elementa 0 in 1. Torej normalne baze generirajo kvečjemu le nekateri elementi obsega $\mathbb{F}_{p^n}$. Pri tem velja, da različni elementi θ lahko generirajo različne normalne baze. Pokazati je možno, da za vsak $n \in \mathbb{N}$ obstaja vsaj en element $\theta \in \mathbb{F}_{p^n}$, ki generira neko normalno bazo [Menezes, str. 80].

Vsak element a obsega $\mathbb{F}_{p^n}$ lahko zapišemo v normalni bazi:

$$a = \sum_{i=0}^{n-1} a_i \theta^{p^i}, \qquad a_i \in \mathbb{F}_p. \tag{1.6}$$

V primeru $p = 2$ torej lahko vsak element $a \in \mathbb{F}_{2^n}$ enačimo z nekim zaporedjem ničel in enic dolžine n, kar je pisano na kožo računalnikov. Od tu dalje se bomo ukvarjali le še s primerom $p = 2$.

Trditev 1.35 *Elementi, ki generirajo normalno bazo obsega $\mathbb{F}_{2^n}$, imajo sled enako ena.*

Dokaz: Ker je $\{\theta^{2^0}, \theta^{2^1}, \theta^{2^2}, \ldots, \theta^{2^{n-1}}\}$ baza, v posebnem vsota vseh baznih elementov ne more biti 0, saj bi to bila ničelna netrivialna linearna kombinacija baznih elementov. Vsota vseh baznih elementov pa je enaka ravno sledi elementa θ. Ker je karakteristika enaka 2, je vrednost sledi lahko le nič ali pa ena. Torej je $\mathrm{Tr}(\theta) = 1$. ■

Enakost $\sum_{i=0}^{n-1} \theta^{2^i} = 1$ pa nam pove še nekaj drugega, da namreč element 1 obsega $\mathbb{F}_{2^n}$ v vsaki normalni bazi predstavimo z n-terico $(1, 1, \ldots, 1)$. Element 0 seveda v vsaki bazi predstavimo z $(0, 0, \ldots, 0)$.

Predstavitev v normalni bazi nam omogoča enostaven izračun sledi poljubnega elementa. To je pomembno pri reševanju kvadratnih enačb v končnih obsegih. Z njimi se srečamo, na primer, pri algoritmu za kompresijo točke na eliptični krivulji. Ta algoritem za faktor dve zmanjša potrebno pasovno širino komunikacijskega kanala, po katerem poteka izmenjava sporočila, ki smo ga zašifrirali s pomočjo eliptičnih krivulj. Če je a kot v (1.6), je za $p = 2$

$$\mathrm{Tr}(a) = \sum_{i=0}^{n-1} a_i \,\mathrm{Tr}(\theta^{2^i}) = \sum_{i=0}^{n-1} a_i \,\mathrm{Tr}(\theta) = \sum_{i=0}^{n-1} a_i. \tag{1.7}$$

Podobno kot pri polinomskih bazah dva elementa obsega $\mathbb{F}_{2^n}$ seštejemo tako, da v $\mathbb{F}_2$ seštejemo istoležne koeficiente. Torej je ta operacija pravzaprav izračun logičnega XOR med dvema zaporedjema ničel in enic dolžine n.

Normalne baze imajo lepo lastnost, da je v njih kvadriranje izjemno enostavno. Kvadrat elementa a, ki je oblike (1.6) za $p = 2$, je namreč

$$a^2 = \sum_{i=0}^{n-1} a_i^2 \theta^{2^{i+1}} = \sum_{i=0}^{n-1} a_i \theta^{2^{i+1}} = a_{n-1}\theta + \sum_{i=1}^{n-1} a_{i-1} \theta^{2^i}.$$

Potemtakem so koeficienti elementa a^2 enaki $(a_{n-1}, a_0, a_1, \ldots, a_{n-2})$. Dobljeni so torej s cikličnim premikom koeficientov elementa a za eno mesto v desno.

Iz tega sledi, da je tudi korenjenje v normalnih bazah enostavno. En koren elementa a mora namreč biti element, ki ga dobimo s cikličnim premikom koeficientov elementa a za eno mesto v levo. V razdelku 1.6 o kvadratnih enačbah bomo pokazali, da ima v obsegu $\mathbb{F}_{2^n}$ vsak element natanko en kvadratni koren. Torej je koren, ki smo ga pravkar našli, tudi edini.

Na žalost pa je množenje v primeru normalnih baz razmeroma zapleteno. Rešitev za ta problem so *optimalne normalne baze*. Optimalne normalne baze so podmnožica normalnih baz, ki imajo nekatere dodatne lastnosti, zaradi katerih je množenje v njih enostavno. Z vidika poenostavitve množenja igrajo enako vlogo, kot jo imajo nerazcepni trinomi v primeru polinomskih baz. Optimalne normalne baze ne obstajajo za vsak obseg $\mathbb{F}_{2^n}$. Vrednosti n, za katere obstajajo, so vseeno precej pogoste. Optimalne normalne baze so izčrpno obdelane v [Menezes, 5. poglavje].

1.6 Kvadrati in kvadratne enačbe

V tem poglavju bomo obravnavali rešljivost kvadratnih enačb v končnih obsegih. Glavna referenca za ta razdelek bo za nas [Vidav].

Definicija 1.36 *Naj bo K poljuben obseg. Element $x \in K$ imenujemo* **kvadrat**, *če obstaja $y \in K$, tako da je $x = y^2$. Elemente, ki niso kvadrati, pa imenujemo* **nekvadrati**.

V splošnem obseg K vsebuje tako kvadrate kot nekvadrate. Če pa je K algebraično zaprt, so vsi elementi kvadrati. Po zgornji definiciji sta elementa 0 in 1 vedno kvadrata.

1.6.1 Obsegi karakteristike p, kjer je $p > 2$

Naj velja char $K \neq 2$. Oglejmo si kvadratno enačbo

$$aX^2 + bX + c = 0, \qquad a \neq 0,\ a, b, c \in K. \tag{1.8}$$

Trditev 1.37 *Kvadratna enačba (1.8) ima v obsegu K rešitev natanko tedaj, ko je diskriminanta $b^2 - 4ac$ kvadrat v K.*

Dokaz: Ker je char $K \neq 2$, je $2 \neq 0$ in v K obstaja element $1/2$. Levo stran enačbe (1.8) dopolnimo do popolnega kvadrata:

$$\left(X + \frac{b}{2a}\right)^2 + \frac{c}{a} - \frac{b^2}{4a^2} = 0$$

Vidimo, da ima ta enačba rešitev natanko takrat, ko je element $b^2/(4a^2) - c/a$ kvadrat. Element obsega je kvadrat natanko takrat, kadar je njegov produkt s poljubnim kvadratom tudi kvadrat. Zato ima enačba (1.8) rešitev natanko takrat, ko je diskriminanta kvadrat. ∎

Če je diskriminanta kvadrat, je število različnih rešitev enako 1 ali pa 2. Ker je karakteristika različna od 2, se lahko hitro prepričamo, da ima enačba (1.8) dve različni rešitvi natanko takrat, ko je diskriminanta neničeln kvadrat. Torej dobimo eno dvojno rešitev natanko v primeru, ko je diskriminanta enaka 0. Če eno rešitev enačbe poznamo, lahko drugo določimo preko Viètovih formul.

Naj bo zdaj $K = \mathbb{F}_{p^n}$, kjer je $p \neq 2$. Izberimo poljuben element $u \in \mathbb{F}_{p^n}$ in zapišimo enačbo

$$X^2 - u = 0. \tag{1.9}$$

Če je $u = 0$, je diskriminanta te enačbe enaka 0 in ima ta enačba dvojno rešitev $X = 0$. Privzemimo torej, da velja $u \neq 0$. Očitno ima enačba (1.9) rešitev natanko takrat, ko je u kvadrat. Naslednja trditev nam pove, koliko je elementov u obsega $\mathbb{F}_{p^n}$, za katere je enačba (1.9) rešljiva.

Trditev 1.38 *Natanko polovica neničelnih elementov obsega $\mathbb{F}_{p^n}$ je kvadratov, druga polovica pa so nekvadrati.*

Dokaz: Oglejmo si preslikavo

$$\gamma : \mathbb{F}_{p^n} \longrightarrow \mathbb{F}_{p^n}$$
$$\gamma : x \longmapsto x^2.$$

Ta preslikava ni homomorfizem obsegov, je pa homomorfizem multiplikativne grupe obsega $\mathbb{F}_{p^n}$. Lastnost, ki je za nas pomembna, je, da je njena slika enaka množici vseh kvadratov v $\mathbb{F}_{p^n}$. Oglejmo si, kako je z injektivnostjo preslikave γ. Naj bo $x^2 = y^2$, $x, y \in \mathbb{F}_{p^n}$. Sledi $(x - y)(x + y) = 0$. Ker so vsi obsegi celi kolobarji, sledi, da je bodisi $x = y$, bodisi $x = -y$. Neničelne elementi obsega $\mathbb{F}_{p^n}$ razdelimo v $(q-1)/2$ parov oblike $\{x, -x\}$. Vsi pari vsebujejo dva elementa in so paroma disjunktni. Preslikava γ preslika različne pare v različne elemente obsega $\mathbb{F}_{p^n}$. Sledi, da ima slika preslikave γ toliko elementov, kot je parov, to je $(q-1)/2$. Torej imamo v $\mathbb{F}_{p^n}$ res natanko $(q-1)/2$ neničelnih kvadratov in prav toliko nekvadratov. ∎

V algoritmih s področja teorije števil in kriptografije moramo pogosto računati potence elementov neke grupe velike moči. V ta namen običajno uporabimo algoritem **kvadriraj in množi** (square and multiply), To je splošen algoritem za hiter izračun n-te potence danega elementa g v poljubni grupi G (lahko nekomutativni). Algoritem poteka tako, da pretečemo vse bite binarne predstavitve števila n od najpomembnejšega bita po vrsti do najmanj pomembnega. Začetni vmesni rezultat je g. Pri vsakem bitu vmesni rezultat kvadriramo. Če je tekoči bit binarne predstavitve števila n enak 1, dobljeni rezultat še pomnožimo z g. Na koncu algoritma je vmesni rezultat enak g^n. Za izvedbo algoritma smo porabili $\lceil \log_2(n) \rceil$ kvadriranj in eno množenje manj, kot je bitov 1 v binarni predstavitvi števila n, torej kvečjemu $\lfloor \log_2(n) \rfloor$. Algoritem kvadriraj in množi je podrobno opisan v [Stinson, str. 127].

Naslednja trditev nam omogoča, da lahko s pomočjo algoritma kvadriraj in množi preverimo, če je dan element $x \in \mathbb{F}_{p^n}$ kvadrat ali ne. Za to potrebujemo $O(\log q)$ množenj v obsegu $\mathbb{F}_{p^n}$. Algoritem kvadriraj in množi v tem primeru poteka v multiplikativni grupi obsega $\mathbb{F}_{p^n}$.

Trditev 1.39 *Veljata naslednji karakteristični lastnosti:*

1. *Neničeln element x iz $\mathbb{F}_{p^n}$ je kvadrat natanko tedaj, ko je $x^{(p^n-1)/2} = 1$.*

2. *Element x iz $\mathbb{F}_{p^n}$ je nekvadrat natanko tedaj, ko je $x^{(p^n-1)/2} = -1$.*

Dokaz: Ker je q lih, lahko faktoriziramo

$$X^q - X = X(X^{q-1} - 1) = X(X^{(q-1)/2} - 1)(X^{(q-1)/2} + 1).$$

Ničle polinoma na levi strani te enakosti so natanko vsi elementi obsega $\mathbb{F}_{p^n}$ in vse ničle so enostavne. Sledi, da je vsak element obsega $\mathbb{F}_{p^n}$ ničla natanko enega od treh faktorjev na desni strani enakosti, in sicer enostavna. Faktor X je rezerviran za element 0. Torej za natanko polovico vseh neničelnih elementov x obsega $\mathbb{F}_{p^n}$ velja $x^{(q-1)/2} = 1$. Za drugo polovico pa velja $x^{(q-1)/2} = -1 \neq 1$. Pokažimo, da prva polovica sovpada z neničelnimi kvadrati, druga pa z nekvadrati. Naj bo neničeln element $x \in \mathbb{F}_{p^n}$ kvadrat. Torej je $x = y^2$ za nek neničeln $y \in \mathbb{F}_{p^n}$. Sledi $x^{(q-1)/2} = y^{q-1} = 1$. Torej so vsi neničelni kvadrati v prvi polovici elementov obsega $\mathbb{F}_q$. Ker pa je neničelnih kvadratov enako število kot nekvadratov, v drugi polovici ne bomo našli nobenega kvadrata. V njej so natanko vsi nekvadrati. ∎

Posledica 1.40 *Neničeln element x obsega $\mathbb{F}_p$ je kvadrat natanko tedaj, ko velja*

$$x^{(p-1)/2} \equiv 1 \pmod{p}.$$

Dokaz: V zgornjo trditev vstavimo $n = 1$. ∎

Na tem mestu uvedimo Jacobijev simbol, ki nam za dan element praobsega $\mathbb{F}_p$ pove, ali je element kvadrat ali ne. Jacobijev simbol je seveda samo oznaka, ki nam sama po sebi v ničemer ne olajša določanja kvadratov.

Definicija 1.41 *Naj bo p poljubno praštevilo in naj bo $x \in \{0, 1, \ldots, p-1\}$.* **Jacobijev simbol** *definiramo z naslednjim predpisom:*

$$\left(\frac{x}{p}\right) = \begin{cases} 1 & ;\ 0 \neq x \text{ je kvadrat v } \mathbb{F}_p \\ -1 & ;\ 0 \neq x \text{ ni kvadrat v } \mathbb{F}_p \\ 0 & ;\ x = 0. \end{cases}$$

Definicija 1.42 *Definicijo Jacobijevega simbola lahko razširimo na $x \in \mathbb{Z}$. Če je x_0 predstavnik kongruenčnega razreda števila x po modulu p, tako da je $0 \leq x_0 \leq p-1$, definirajmo $\left(\frac{x}{p}\right) = \left(\frac{x_0}{p}\right)$.*

Definicija 1.43 *Definicijo dalje razširimo na poljuben $x \in \mathbb{Z}$ in $n \in \mathbb{N}$, $n \geq 2$. Predpis naj bo takšen, da bo Jacobijev simbol v spodnjem faktorju številsko-teoretično multiplikativen. Če je $n = p_1^{i_1} p_2^{i_2} \cdots p_k^{i_k}$, naj bo torej*

$$\left(\frac{x}{n}\right) = \prod_{j=1}^{k} \left(\frac{x}{p_j}\right)^{i_j}.$$

Z Jacobijevim simbolom se bomo ponovno srečali v Schoofovem algoritmu.

Recimo, da smo za nek neničeln element $u \in \mathbb{F}_p$ ugotovili, da je kvadrat. Zdaj bi želeli določiti njegov kvadratni koren. Recimo, da smo našli nek koren $x \in \mathbb{F}_{p^n}$, torej $x^2 = u$. Iz Viètove formule sledi, da je potem drugi koren enak $-x$. Ker je $u \neq 0$, sta x in $-x$ med seboj različna. Drugih korenov seveda ni, saj je stopnja enačbe $X^2 - u = 0$ enaka 2.

V primeru, ko je $q \equiv 3 \pmod 4$, je koren preprosto določiti. V tem primeru je namreč $(q+1)/4$ naravno število in velja

$$(u^{(q+1)/4})^2 = u^{(q+1)/2} = u \cdot u^{(q-1)/2} = u.$$

Torej je

$$\sqrt{u} = u^{(q+1)/4}.$$

Ta formula nam podaja determinističen algoritem za izračun kvadratnega korena v primeru $q \equiv 3 \pmod 4$. Če spet uporabimo algoritem kvadriraj in množi, lahko koren deterministično določimo z $O(\log q)$ množenji v obsegu $\mathbb{F}_{p^n}$.

Ker je q liho število, ostane le še primer, ko je $q \equiv 1 \pmod 4$. Ta primer je računsko zahtevnejši in zanj zaenkrat ni znan učinkovit determinističen algoritem.

1.6.2 Obsegi s karakteristiko $p = 2$

V obsegih s karakteristiko 2 običajni pristop za reševanje kvadratne enačbe, to je dopolnitev do popolnega kvadrata, odpove. Obravnavali bomo le končne obsege. Torej naj bo $p = 2$ in $q = 2^n$.

Trditev 1.44 *Naj bo u poljuben element obsega $\mathbb{F}_q$. Enačba $X^2 = u$ je rešljiva za vsak $u \in \mathbb{F}_q$. Pri vsakem $u \in \mathbb{F}_q$ ima natanko eno, dvojno rešitev.*

Dokaz: Recimo, da je $x^2 = y^2$ za neka $x, y \in \mathbb{F}_q$. Potem bodisi $x = y$ bodisi $x = -y$. Ker pa smo v obsegu s karakteristiko 2, je $-y = y$, torej imamo pravzaprav samo eno možnost, da je namreč $x = y$. Iz tega izluščimo sledeče: ko x preteče vse elemente obsega $\mathbb{F}_q$, so elementi x^2 med seboj paroma različni. Torej jih je natanko toliko, kolikor je moč obsega. Sledi, da je vsak element obsega $\mathbb{F}_q$ kvadrat in da ima enačba $X^2 = u$ za vsak $u \in \mathbb{F}_q$ natanko eno rešitev. Zato mora ta rešitev biti dvojna. ∎

Iz te trditve neposredno sledi, da velja razcep

$$X^2 - u = (X - \sqrt{u})(X + \sqrt{u}) = (X + \sqrt{u})(X + \sqrt{u}).$$

Situacija glede obstoja korenov je torej v primeru karakteristike 2 bistveno drugačna kot v primeru, ko karakteristika ni 2. V končnih obsegih $\mathbb{F}_{2^n}$ nekvadratov sploh ni. Tudi koren elementa u lahko enostavno poiščemo. Velja namreč

$$(u^{2^{n-1}})^2 = u^{2^n} = u, \quad \text{torej } \sqrt{u} = u^{2^{n-1}}.$$

S pomočjo algoritma kvadriraj in množi lahko koren deterministično poiščemo z $O(\log q) = O(n)$ množenji elementov obsega $\mathbb{F}_q$.

Vse to pa še ne pomeni, da je vsaka kvadratna enačba v obsegu $\mathbb{F}_{2^n}$ rešljiva. Oglejmo si torej splošno enačbo oblike

$$aX^2 + bX + c = 0, \qquad a, b, c \in \mathbb{F}_q,\ a \neq 0.$$

To enačbo pomnožimo z a^{-1} in dobimo enačbo

$$X^2 + uX + v = 0, \qquad u, v \in \mathbb{F}_q.$$

Če je $u = 0$, trčimo ob primer, ki smo ga že obravnavali. Naj bo torej $u \neq 0$. Enačbo pomnožimo z u^{-2} in uvedimo novo neznanko $Y = u^{-1}X$. Enačba se poenostavi v obliko

$$Y^2 + Y = w, \qquad \text{kjer je } w = vu^{-2}. \tag{1.10}$$

Bralca naj opozorimo, da smo v obsegu s karakteristiko 2, zato lahko člene v enačbah prestavljamo z ene strani enačbe na drugo kar brez spremembe predznaka. Izberemo pač tisto obliko, ki se nam zdi preglednejša. Če je y ena rešitev zadnje enačbe (1.10), nam Viètova formula pove, da je druga rešitev enaka $y + 1$. Ker $1 \neq 0$, sta ti dve rešitvi vedno različni.

V obsegih s karakteristiko različno od 2 kriterij za rešljivost enačbe podaja diskriminanta. Če pa je karakteristika enaka 2, diskriminanta nima kakšnega posebnega pomena. Vseeno lahko najdemo preprost kriterij za rešljivost enačbe (1.10). Ta kriterij je povrhu še enostavno računsko uporaben. Njegov dokaz pa nam podaja učinkovito metodo za iskanje rešitve enačbe. Kriterij podaja naslednja trditev.

Trditev 1.45 *Enačba (1.10) ima rešitev v $\mathbb{F}_{2^n}$ natanko tedaj, ko je sled elementa w enaka* 0.

Najprej opomnimo, da v primeru obsegov s karakteristiko 2 sled slika v praobseg $\mathbb{F}_2$. Torej je vrednost sledi nekega elementa lahko le 0 ali 1.

Dokaz:
($\Rightarrow$) Naj ima enačba (1.10) rešitev y. Uporabimo sled na obeh straneh enačbe (1.10). Dobimo

$$\mathrm{Tr}(w) = \mathrm{Tr}(y^2 + y) = \mathrm{Tr}(y^2) + \mathrm{Tr}(y) = \mathrm{Tr}(y) + \mathrm{Tr}(y) = 0.$$

Pri tem smo uporabili aditivnost sledi in lastnost $\mathrm{Tr}(x^2) = \mathrm{Tr}(x)$.

($\Leftarrow$) Naj bo $\mathrm{Tr}(w) = 0$. Opisali bomo eksplicitno konstrukcijo ene rešitve enačbe (1.10). Ta konstrukcija je praktično uporabna, kadar imamo elemente obsega v računalniku predstavljene v normalni bazi. Naj bo θ poljuben generator poljubne normalne baze za $\mathbb{F}_{2^n}$. Rešitev y bomo predstavili v tej normalni bazi. Naj elementu y ustrezajo koeficienti $(y_0, y_1, \ldots, y_{n-1})$, kjer moramo elemente y_i še določiti. Elementu w pa naj ustreza predstavitev $(w_0, w_1, \ldots, w_{n-1})$. Poskusimo z nastavkom $y_0 = 0$. Enačba (1.10) dobi obliko

$$\begin{array}{cllllll} & (y_{n-1}, & 0, & y_1, & y_2, & \ldots, & y_{n-2}) \\ + & (0, & y_1, & y_2, & y_3, & \ldots, & y_{n-1}) \\ \hline = & (w_0, & w_1, & w_2, & w_3, & \ldots, & w_{n-1}) \end{array}$$

Iz drugega, tretjega, ..., zadnjega stolpca po vrsti dobimo določilne enačbe za $y_1, y_2, \ldots, y_{n-1}$:

$$\begin{aligned} y_1 &= w_1, \\ y_i &= w_i + y_{i-1}, \qquad i = 2, 3, \ldots, n-1. \end{aligned} \tag{1.11}$$

Edini problem, ki še ostane, je, da ne vemo, če je zadoščeno tudi enačbi v prvem stolpcu $y_{n-1} = w_0$. Tu uporabimo predpostavko, da je $\mathrm{Tr}(w) = 0$. Enačbe iz (1.11) seštejemo med sabo. Členi $y_i,\ 1 \leq i \leq n-2$ se pokrajšajo in dobimo enačbo

$$y_{n-1} = \sum_{i=1}^{n-1} w_i = \mathrm{Tr}(w) + w_0 = w_0,$$

kar je ravno enačba prvega stolpca. Pri drugem enačaju v zgornji enakosti smo uporabili formulo (1.7), ki smo jo izpeljali v podrazdelku 1.5.2 o normalnih bazah. ∎

Če po postopku iz zgornjega dokaza določimo normalno predstavitev ene rešitve y, imamo normalno predstavitev za drugo rešitev $y+1$ praktično že na dlani. Ker je element $1 \in \mathbb{F}_{2^n}$ predstavljen z $(1, 1, \ldots, 1)$, namreč drugo rešitev dobimo tako, da negiramo vse bite iz normalne predstavitve elementa y.

V razdelku 1.2 o sledi smo pokazali, da je sled neničeln funkcional, zato obstajajo elementi $w \in \mathbb{F}_q$ z neničelno sledjo. Pokazali smo, da je takih elementov v $\mathbb{F}_{2^n}$ natanko $2^n - 2^{n-1}$. Torej v vsakem obsegu oblike $\mathbb{F}_{2^n}$ obstajajo nerešljive kvadratne enačbe. Našo predstavitev kvadratnih enačb v končnih obsegih strnimo z naslednjo ugotovitvijo.

Trditev 1.46 *V vsakem končnem obsegu obstajajo kvadratne enačbe s koeficienti iz tega obsega, ki v tem obsegu nimajo rešitve.*

Dokaz: Če karakteristika ni 2, so to že kar enačbe oblike $X^2 - u = 0$, kjer je u nekvadrat. Če pa je karakteristika 2, poiščemo element w s sledjo enako 1 in uporabimo trditev (1.45). ∎

Tako na primer enačba $X^2 + X + \theta = 0$ nima rešitve v $\mathbb{F}_{2^n}$, če je θ generator kakšne normalne baze v $\mathbb{F}_{2^n}$.

Poglavje 2

Eliptične krivulje

Eliptične krivulje so posebna družina algebraičnih krivulj, ki so v matematiki prisotne že od začetka 19. stoletja. Z njimi so se ukvarjali mnogi znani matematiki, med njimi tudi akademik prof. dr. Ivan Vidav. Njegova knjiga [Vidav EC] je zelo primerna kot uvod v teorijo eliptičnih krivulj in eliptičnih funkcij. Schoofov algoritem je leta 1985 na široko odprl vrata za uporabo eliptičnih krivulj v kriptografiji, kar je precej spodbudilo raziskovanje na tem področju, za eliptične krivulje pa so začeli zanimati tudi v nematematičnih vodah. To poglavje je namenjeno definiciji eliptičnih krivulj in uvedbi grupne strukture na eliptične krivulje. V prvem razdelku definiramo eliptično krivuljo in navedemo njene osnovne geometrijske lastnosti. V drugem razdelku na eliptično krivuljo dodamo algebraično strukturo.

2.1 Weierstrassova enačba eliptične krivulje

V nadaljevanju bomo definirali eliptične krivulje in navedli njihove osnovne geometrijske lastnosti. Eliptične krivulje imajo dve obliki: projektivno in afino. V tem delu se bomo v naslednjih poglavjih ukvarjali le z afinimi eliptičnimi krivuljami, ki jim bomo rekli kar eliptične krivulje. Definirali pa bomo oba tipa eliptičnih krivulj. Zato bomo najprej navedli osnovna dejstva iz projektivne geometrije, ki so potrebna za razumevanje definicije projektivne eliptične krivulje. Zatem bomo s pomočjo Weierstrassove enačbe definirali projektivne in afine eliptične krivulje. V zadnjem delu razdelka bomo za različne tipe eliptičnih krivulj poiskali njihove poenostavljene kanonične predstavnike.

2.1.1 Kratek izlet v projektivno geometrijo

V tem podrazdelku bomo navedli osnovna dejstva projektivne geometrije, ki so potrebna za obravnavo projektivne eliptične krivulje. Več o projektivni ravnini in projektivnih eliptičnih krivuljah lahko bralec najde v [Vidav EC]. Oznaka K naj povsod v tem razdelku pomeni poljuben komutativen obseg, $\overline{K}$ pa naj pomeni njegovo algebraično zaprtje.

Definicija 2.1 *Točki* (x, y, z) *in* (x', y', z') *naj bosta ekvivalentni, če obstaja neničeln element* $\lambda \in K$*, tako da je* $x = \lambda x'$*,* $y = \lambda y'$ *in* $z = \lambda z'$*.* **Projektivna ravnina** $\mathbb{P}^2(K)$ *je faktorska množica množice* $K \times K \times K \setminus \{(0, 0, 0)\}$ *glede na to ekvivalenčno relacijo.*

Definicija 2.2 *Ekvivalenčni razred točke* (x, y, z) *označimo z* $[x, y, z]$.

Intuitivno si lahko projektivno ravnino predstavljamo kot množico, ki jo dobimo, če običajni ravnini $K \times K$ dodamo **točke v neskončnosti**. Običajno ravnino $K \times K$ lahko vložimo v projektivno ravnino $\mathbb{P}^2(K)$ in sicer tako, da točko (x, y) preslikamo v $[x, y, 1]$.

Definicija 2.3 *Točke projektivne ravnine, ki ustrezajo neki točki iz* $K \times K$, *imenujemo* **končne točke**. *To so torej točke, ki ne ležijo v neskončnosti. Analitično jih prepoznamo po neničelni tretji projektivni koordinati.*

Definicija 2.4 *Točke v neskončnosti so tiste točke projektivne ravnine, ki imajo zadnjo koordinato enako* 0, *torej točke oblike* $[x, y, 0]$ *za neka* $x, y \in K$.

Točke v neskončnosti v običajni ravnini $K \times K$ nimajo ustreznega predstavnika. Vseeno si lahko intuitivno predstavljamo njihovo lego. Točke v neskončnosti so namreč v bijektivni korespondenci s premicami v $K \times K$, ki gredo skozi izhodišče $(0, 0)$. Za posamezno tako premico si lahko predstavljamo, da ustrezna točka v neskončnosti projektivne ravnine leži v "neskončnosti" ravnine $K \times K$ in sicer "na koncu" dane premice. Vsaka premica skozi izhodišče ima seveda dva taka konca. Predstavljamo si, da ustrezna točka v neskončnosti leži na "obeh koncih ravnine hkrati"oziroma, da ta dva konca identificiramo. Analitičen odnos med premico in točko v neskončnosti pa je takle: če je smerni vektor premice skozi izhodišče enak (x, y), ima potem ustrezna točka v neskončnosti projektivne koordinate $[x, y, 0]$.

Definicija 2.5 **Projektivna premica** *v projektivni ravnini je rešitev homogene enačbe prve stopnje* $aX + bY + cZ = 0$, *kjer so* a, b, c *neki elementi iz* K, *ki niso hkrati vsi enaki* 0.

S pomočjo te definicije lahko enostavno vidimo, da je množica točk v neskončnosti projektivne ravnine $\mathbb{P}^2(K)$ enaka projektivni premici, ki jo dobimo za $a = b = 0, c = 1$.

Definicija 2.6 *To projektivno premico imenujemo* **premica v neskončnosti**.

Vse definicije pojmov, ki jih definiramo v zvezi s projektivno ravnino, morajo biti usklajene z ekvivalenčno relacijo na $K \times K \times K$, s pomočjo katere smo definirali projektivno ravnino. Te definicije morajo torej biti invariantne na množenje vseh koordinat z istim neničelnim elementom obsega K. V definiciji projektivne premice to na primer dosežemo s homogenostjo ustrezne linearne enačbe.

S stališča eliptičnih krivulj in grup na eliptičnih krivuljah, ki jih bomo definirali v nadaljevanju, je glavna prednost projektivne ravnine pred običajno, da naravno in enakovredno vsebuje tudi točke v neskončnosti, za katere tako ni potrebna ločena obravnava. Za splošno matematiko verjetno najpomembnejša lastnost projektivne ravnine pa je, da se v njej vsaki dve projektivni premici sekata. Ta lastnost je s stališča končnih geometrij pravzaprav eden od osnovnih aksiomov, ki definirajo projektivno ravnino. V tem delu bomo ostali pri intuitivnem opisu projektivne ravnine. Dve premici v projektivni ravnini se ali sekata v eni točki ali pa sovpadata. Vzporednih premic, kot jih poznamo iz običajne ravnine, tukaj ni. Dve vzporedni različni premici iz običajne ravnine se, potem ko običajni ravnini dodamo točke v neskončnosti, sečeta v točki v neskončnosti, ki se nahaja v smeri premice skozi izhodišče, ki je vzporedna danima vzporednima premicama.

2.1.2 Definicija projektivnih in afinih eliptičnih krivulj

Definicija 2.7 **Projektivna eliptična krivulja** *je množica tistih elementov projektivne ravnine* $\mathbb{P}^2(\overline{K})$*, ki zadoščajo enačbi*

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3, \tag{2.1}$$

kjer so a_1, a_2, a_3, a_4, a_6 *elementi obsega* K.

Za različne a_1, a_2, a_3, a_4, a_6 seveda v splošnem dobimo različne krivulje.

Definicija 2.8 *Enačba (2.1) se imenuje* **homogena Weierstrassova enačba**.

Polinom, ki nastopa na levi strani homogene Weierstrassove enačbe, je **homogen**, to je, vsi členi imajo isto stopnjo. Kot smo zgoraj že omenili, je to nujno, če želimo, da homogena Weierstrassova enačba (2.1) sploh določa neko podmnožico projektivne ravnine.

Morda ni odveč posebej opozoriti, da elementi projektivne eliptične krivulje ležijo v projektivni ravnini zaprtja obsega K. Homogeni polinom na levi strani homogene Weierstrassove enačbe, ki definira krivuljo, pa ima koeficiente v K.

Projektivna ravnina $\mathbb{P}(\overline{K})$ seveda vsebuje tudi projektivno ravnino $\mathbb{P}(K)$. V splošnem se zgodi, da večina točk na projektivni eliptični krivulji ne leži v $\mathbb{P}(K)$, nekaj (lahko tudi neskončno) pa jih vseeno leži tudi v $\mathbb{P}(K)$. Definirajmo še afini ekvivalent projektivne eliptične krivulje.

Definicija 2.9 **Afina eliptična krivulja** *ali na kratko kar* **eliptična krivulja** *naj bo množica tistih elementov produkta* $\overline{K} \times \overline{K}$*, ki zadoščajo enačbi*

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6, \tag{2.2}$$

kjer so a_1, a_2, a_3, a_4, a_6 *neki elementi obsega* K.

Definicija 2.10 *Enačba (2.2) se imenuje* **afina Weierstrassova enačba** *ali na kratko kar* **Weierstrassova enačba**. *Kadar jo primerjamo s kratko Weierstrassovo enačbo, ki jo bomo definirali v nadaljevanju, ji rečemo* **splošna Weierstrassova enačba**. *Afino eliptično krivuljo ponavadi označimo z* E *ali z* $E(\overline{K})$. *Pri tem je v literaturi v navadi, da v* E *poleg rešitev enačbe (2.2) dodatno damo še* **točko v neskončnosti**, *o kateri bo govora kasneje.*

Spet velja opozorilo, da koordinate točk na afini krivulji ležijo v $\overline{K}$ in ne nujno v K. Tako kot prej ima tudi sedaj definicijski polinom krivulje koeficiente v K. Zato je za vsak obseg L, kjer je $K \leq L \leq \overline{K}$, na mestu naslednja definicija.

Definicija 2.11 *Naj bo* L *vmesni obseg med* K *in* $\overline{K}$. *Potem naj oznaka* $E(L)$ *pomeni množico vseh tistih točk* (x, y) *krivulje* E*, za katere obe koordinati* x *in* y *ležita v obsegu* L *(in ne zgolj v* $\overline{K}$*). Po definiciji v množico* $E(L)$ *dodamo tudi točko v neskončnosti.*

Definicija 2.12 *Naj bo* L *tak vmesni obseg, da ima krivulja* $E(L)$ *končno moč. Potem to moč označimo z* $\#E(L)$.

Ta definicija upravičuje zgornjo oznako $E \equiv E(\overline{K})$. Krivulje E ne smemo zamešati s krivuljo $E(K)$, čeprav pogosto za obe krivulji uporabljamo neformalno oznako eliptična krivulja.

Definicija 2.13 *Množico $E(K)$ imenujemo* **množica racionalnih točk na eliptični krivulji**. *Koordinate točk iz te množice torej ležijo v K.*

Navedimo sedaj nekaj osnovnih dejstev o številu točk na krivuljah E in $E(K)$.

Trditev 2.14 *Krivulja E je ne glede na svoje parametre a_1, a_2, a_3, a_4, a_6 vedno neskončna množica.*

Dokaz: Če v afino Weierstrassovo enačbo vstavimo $X = x$, kjer je x nek poljuben element obsega $\overline{K}$, dobimo kvadratno enačbo po spremenljivki Y. Ker je obseg $\overline{K}$ algebraično zaprt, ima ta enačba vedno rešitev. To pomeni, da za vsak $x \in \overline{K}$ obstaja tak $y \in \overline{K}$, da je točka (x, y) na krivulji E. Ker končni obsegi niso algebraično zaprti, je $\overline{K}$ neskončen obseg. Torej je krivulja E neskončna množica. ■

Kako pa je s kardinalnostjo množice racionalnih točk $E(K)$? Pri nekaterih obsegih je tudi ta množica vedno neskončna.

Trditev 2.15 *Za poljubno eliptično krivuljo nad obsegom $\mathbb{R}$ je množica racionalnih točk $E(\mathbb{R})$ neskončna.*

Dokaz: Diskriminanta kvadratne enačbe po spremenljivki Y, ki jo dobimo, če v Weierstrassovo enačbo vstavimo $X = x$, kjer je $x \in \mathbb{R}$, je polinom tretje stopnje spremenljivke x. Za vsak polinom tretje stopnje iz kolobarja $\mathbb{R}[X]$ je množica točk iz $\mathbb{R}$, pri katerih je vrednost polinoma pozitivna, neskončna. Za vsak x iz te neskončne množice na krivulji $E(\mathbb{R})$ obstaja vsaj ena točka, katere X-koordinata je enaka x. ■

V posebnem primeru, ko je $K = \mathbb{F}_q$ končen obseg, pa je množica $E(K)$ končna.

Trditev 2.16 *Za moč množice racionalnih točk $E(K)$ na poljubni eliptični krivulji nad $\mathbb{F}_q$ velja ocena*

$$\#E(\mathbb{F}_q) \leq 2q + 1.$$

Dokaz: Za vsak $x \in \mathbb{F}_q$ ima enačba po Y, ki jo dobimo, če v afino Weierstrassovo enačbo vstavimo $X = x$, stopnjo 2 in tako kvečjemu dve rešitvi. Iz tega neposredno sledi iskana ocena. ■

To oceno precej izboljša Hassejev izrek.

Izrek 2.17 (Hasse, 1930) *Naj bo E eliptična krivulja nad končnim obsegom $\mathbb{F}_q$. Potem je*

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

■

Dokaz tega izreka je precej zapleten in zahteva veliko znanja algebraične geometrije. Eno različico dokaza je možno najti v [Silverman, str. 131]. V delu [Enge, str. 98-100] pa je naveden dokaz Hassejevega izreka, ki hkrati dokaže še karakteristično enačbo Frobeniusovega endomorfizma (3.14). S to za Schoofov algoritem zelo pomembno enačbo se bomo srečali v naslednjem poglavju.

Za velike končne obsege in dano eliptično krivuljo je kljub Hassejevemu izreku natančna določitev števila $\#E(\mathbb{F}_q)$ težka naloga. Glavni cilj tega diplomskega dela je opis učinkovitega algoritma za reševanje tega problema.

2.1.3 Primerjava projektivnih in afinih eliptičnih krivulj

Projektivna eliptična krivulja in afina eliptična krivulja sta dva različna obraza "istega" matematičnega objekta. Projektivne eliptične krivulje in afine eliptične krivulje so namreč v bijektivni korespondenci. Par korespondentnih krivulj dobimo, če v homogeni Weierstrassovi enačbi in afini Weierstrassovi enačbi vzamemo iste koeficiente a_1, a_2, a_3, a_4, a_6.

Oglejmo si projektivno eliptično krivuljo E_P in njej ustrezno afino krivuljo E_A. Točki (x, y) na E_A ustreza točka $[x, y, 1]$ na E_P. Obratno, točki $[x, y, z]$ na E_P ustreza točka $(x/z, y/z)$ na E_A, če le z ni enak 0. Kaj pa točke na E_P, kjer je $z = 0$? Iz prejšnjega razdelka vemo, da je $z = 0$ karakteristična lastnost točk projektivne ravnine, ki ležijo v neskončnosti. Iz homogene Weierstrassove enačbe takoj sledi, da je edina točka na E_P, ki leži v neskončnosti, točka $[0, 1, 0]$. Ta točka se v projektivni ravnini nahaja "na koncu ordinatne osi", to je v neskončnosti v smeri ordinatne osi.

Definicija 2.18 *Točko* $[0, 1, 0]$ *na homogeni eliptični krivulji označimo z* $\mathcal{O}$.

Točka $\mathcal{O}$ na afini eliptični krivulji E_A nima ustrezne ekvivalentne točke, si pa lahko predstavljamo, da leži v neskončnosti "na koncu krivulje E_A".

V primeru $K = \mathbb{R}$ si lahko lego točke $\mathcal{O}$ glede na afino eliptično krivuljo lepo nazorno predstavljamo. Afina Weierstrassova enačba je stopnje dve v Y in stopnje tri v X. Ko $X \to \infty$, tako na afini eliptični krivulji velja približno $Y = X^{\frac{3}{2}}$. Krivulja tako uide v neskončnost "hitreje"v smeri osi Y kot v smeri osi X, tangenta na krivuljo pa postaja z $X \to \infty$ vedno bolj navpična. Zato gre afina krivulja proti neskončnosti natanko proti projektivni točki $[0, 1, 0] = \mathcal{O}$.

Kot smo zgoraj že omenili, točko $\mathcal{O}$ vedno dodamo v množico E_A. Po tem dogovoru eliptična krivulja $E = E_A$ sestoji iz rešitev dane afine Weierstrassove enačbe, ki so urejeni pari elementov iz $\overline{K} \times \overline{K}$ in iz točke $\mathcal{O}$.

V nadaljevanju bomo uporabljali le afine krivulje. Projektivne eliptične krivulje se sicer uporabljajo pri nekaterih hitrih algoritmih za izvedbo grupne operacije, a se z njimi v tem delu ne bomo ukvarjali [Blake, str. 58]. Afine krivulje so pogosto preglednejše, ker delamo le z dvema spremenljivkama. Slaba stran afinih krivulj pa je, da moramo poleg končnih točk na krivulji vedno posebej obravnavati še točko v neskončnosti.

2.1.4 Diskriminanta, j-invarianta in izomorfizmi eliptičnih krivulj

Vsaki eliptični krivulji lahko priredimo dve pomembni količini: diskriminanto in j-invarianto. Na množico vseh eliptičnih krivulj nad danim komutativnim obsegom je možno naravno vpeljati relacijo izomorfnosti. To relacijo bomo definirali in pokazali, kako je povezana z diskriminanto in j-invarianto.

Navedimo najprej dva elementarna pojma iz algebraične geometrije, namreč pojma algebraične množice in varietete. S stališča algebraične geometrije so eliptične krivulje izčrpno obdelane v delu [Silverman]. Algebraična geometrija je zaenkrat tudi edino pravo orodje za globlje razumevanje teoretičnega ozadja eliptičnih krivulj.

Definicija 2.19 *Naj bo $p \in K[X,Y]$ nek nekonstanten polinom dveh spremenljivk in A množica rešitev enačbe $p(X,Y) = 0$ v $K \times K$. Potem rečemo, da je A* **algebraična množica***.*

Možno je pokazati [Enge, str. 15], da je za vsako izbiro parametrov a_1, a_2, a_3, a_4, a_6 polinom, ki določa eliptično krivuljo, nerazcepen element kolobarja $K[X,Y]$. Zato nobena eliptična krivulja ni unija manjših algebraičnih množic.

Definicija 2.20 *Algebraične množice, ki niso unije manjših algebraičnih množic, imenujejo* **varietete***.*

Vsaka eliptična krivulja je torej varieteta. Definirati moramo še pojem singularne točke na algebraični krivulji.

Definicija 2.21 *Naj bo $p \in K[X,Y]$ nek polinom dveh spremenljivk in naj bo (x,y) točka na algebraični krivulji $p(X,Y) = 0$. Potem rečemo, da je (x,y)* **singularna točka krivulje***, če velja*

$$\frac{\partial p}{\partial X}(x,y) = \frac{\partial p}{\partial Y}(x,y) = 0.$$

Če je $K = \mathbb{R}$, so singularne točke krivulje točke, v katerih tangenta na krivuljo ni dobro definirana. Tipične take točke so samopresečišča in osti krivulje.

Definicija 2.22 *Za eliptično krivuljo pravimo, da je* **singularna***, če ima vsaj eno singularno točko. Eliptične krivulje, ki niso singularne, imenujemo* **nesingularne***.*

V splošnem je eliptična krivulja lahko singularna ali pa nesingularna. Nas bodo zanimale le nesingularne eliptične krivulje. Odslej privzemimo, da so vse eliptične krivulje, s katerimi delamo, nesingularne. Preprost kriterij za singularnost nam podaja diskriminanta eliptične krivulje, ki jo bomo zdaj definirali.

Definicija 2.23 *Naj bo E eliptična krivulja nad poljubnim komutativnim obsegom K. Uvedimo najprej naslednje konstante, ki nastopajo v definiciji diskriminante. Te konstante bodo nastopale tudi v nekaterih formulah v nadaljevanju.*

$$\begin{aligned} b_2 &= a_1^2 + 4a_2, \quad b_4 = a_1 a_3 + 2a_4, \quad b_6 = a_3^2 + 4a_6, \\ b_8 &= a_1^2 a_6 + 4a_2 a_6 - a_1 a_3 a_4 + a_2 a_3^2 - a_4^2, \\ c_4 &= b_2^2 - 24 b_4, \quad c_6 = -b_2^3 + 36 b_2 b_4 - 216 b_6. \end{aligned}$$

Diskriminanto *označimo z Δ. Podana je z enačbo*

$$\Delta = -b_2^2 b_8 - 8b_4^3 - 27b_6^2 + 9b_2 b_4 b_6.$$

V zadnjem delu tega razdelka bomo navedli poenostavljene formule za diskriminanto za enostavnejše primere eliptičnih krivulj. Pojem diskriminante je važen predvsem zaradi naslednje ugotovitve, katere dokaz lahko najdemo v [Enge, str. 20].

Trditev 2.24 *Eliptična krivulja je singularna natanko takrat, ko je njena diskriminanta enaka* 0. ∎

Druga pomembna konstanta, ki jo priredimo eliptični krivulji, je *j-invarianta*. Definirana je le za nesingularne eliptične krivulje.

Definicija 2.25 *Naj bo E nesingularna eliptična krivulja nad poljubnim komutativnim obsegom K. Njena* **j-invarianta** *je podana z enačbo*

$$j = \frac{c_4^3}{\Delta}.$$

Nad algebraično zaprtimi obsegi se iz j-invariante vidi, kdaj sta dve krivulji izomorfni. Obsegi, s katerimi delamo v praksi, sicer niso algebraično zaprti, tako da nam taka karakterizacija ne pomaga preveč. V obsegih s karakteristiko 2 ali 3 je j-invarianta v tesni zvezi s pojmom supersingularnosti, ki ga bomo definirali v razdelku 3.1. V kriptografiji srečamo j-invarianto v Schoof-Atkin-Elkiesovem algoritmu, kjer igra ključno vlogo. O tem algoritmu bo tekla beseda v zadnjem razdelku tega dela.

Nekatere eliptične krivulje je možno z uvedbo primerne substitucije prevesti na druge. Taki dve eliptični krivulji imenujemo *izomorfni*. Skupnih imata veliko pomembnih lastnosti. Sledi natančna definicija izomorfizma.

Definicija 2.26 *Naj bosta E in E' eliptični krivulji, podani z Weierstrassovima enačbama*

$$E: \quad Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6,$$

$$E': \quad Y'^2 + a_1'X'Y' + a_3'Y' = X'^3 + a_2'X'^2 + a_4'X' + a_6',$$

kjer so $a_i, a_i' \in K$, $i = 1, 2, 3, 4, 6$. Pravimo, da sta eliptični krivulji E in E' **izomorfni**, *če obstaja neka transformacija oblike*

$$\begin{bmatrix} X \\ Y \end{bmatrix} \longmapsto \begin{bmatrix} u^2 & 0 \\ u^2 s & u^3 \end{bmatrix} \begin{bmatrix} X' \\ Y' \end{bmatrix} + \begin{bmatrix} r \\ t \end{bmatrix}, \qquad r, s, t, u \in K,\ u \neq 0,$$

ki pretvori enačbo za E v enačbo za E'.

Definicija 2.27 *Taki transformaciji pravimo* **dopustna transformacija spremenljivk**.

Vsaka dopustna transformacija spremenljivk je obrnljiva, saj je

$$\det \begin{bmatrix} u^2 & 0 \\ u^2 s & u^3 \end{bmatrix} = u^5 \neq 0.$$

Hitro se lahko prepričamo, da je inverz vsake dopustne transformacije spremenljivk spet dopustna transformacija spremenljivk. Zato naša definicija izomorfizma podaja ekvivalenčno relacijo na množici vseh eliptičnih krivulj nad danim obsegom in je torej smiselna.

Primer: Konjugiranje

$$(X,Y) \longmapsto \overline{(X,Y)} = (X', -Y' - a_1X' - a_3)$$

je dopustna transformacija spremenljivk, ki jo dobimo za $u=-1, r=0, s=-a_1, t=-a_3$. Konjugiranje je sebi inverzna preslikava. Kasneje, ko bomo na eliptični krivulji definirali grupno strukturo, bomo videli, da to transformacijo pravzaprav lahko razumemo kot homomorfizem grupe na eliptični krivulji, ki poljubno točko preslika v njej inverzno točko.

Iz eksplicitnih formul za diskriminanto in j-invarianto se lahko prepričamo, da imata izomorfni krivulji isto j-invarianto. Diskriminanti pa se razlikujeta za multiplikativni faktor. Zato za izomorfni krivulji E in E' velja

$$\Delta(E) = 0 \iff \Delta(E') = 0$$

in je torej singularnost eliptične krivulje pravzaprav lastnost celotnega ekvivalenčnega razreda eliptičnih krivulj za relacijo izomorfnosti.

2.1.5 Poenostavitev Weierstrassove enačbe

Poskusimo zdaj v vsakem izomorfnostnem razredu eliptičnih krivulj poiskati čim bolj preprostega predstavnika. Tudi v tem razdelku naj bo K še vedno poljuben komutativen obseg.

Naj bo najprej $\mathrm{char}(K) \neq 2$. Potem lahko izraz na levi strani Weierstrassove enačbe dopolnimo do popolnega kvadrata. Dobimo

$$\left(Y + \frac{a_1X + a_3}{2}\right)^2 - \left(\frac{a_1X + a_3}{2}\right)^2 = X^3 + a_2X^2 + a_4X + a_6.$$

Če uporabimo dopustno transformacijo spremenljivk $(X,Y) \mapsto (X, Y - (a_1X + a_3)/2)$, dobimo izomorfno krivuljo z enačbo oblike

$$Y^2 = X^3 + a_2'X^2 + a_4'X + a_6'. \tag{2.3}$$

Če velja dodatno še $\mathrm{char}(K) \neq 3$, lahko tudi kubični polinom po X na desni strani enačbe (2.3) dopolnimo do kuba in se tako znebimo člena pri X^2. Zapišimo

$$X^3 + a_2'X^2 + a_4'X + a_6' = \left(X + \frac{a_2'}{3}\right)^3 + a_4'X + a_6' - a_2'^2X - \frac{a_2'^3}{27}.$$

Dopustna transformacija spremenljivk $(X,Y) \mapsto (X - a_2'/3, Y)$ nam tako spremeni krivuljo (2.3) v izomorfno krivuljo z enačbo oblike

$$Y^2 = X^3 + aX + b, \quad a, b \in K. \tag{2.4}$$

Definicija 2.28 *Enačba* (2.4) *se imenuje* **kratka Weierstrassova enačba.**

To je ena od dveh najpogosteje uporabljanih enačb na področju kriptografije z eliptičnimi krivuljami. Za vsako eliptično krivuljo nad obsegom karakteristike različne od 2 in 3 obstaja tej krivulji izomorfna krivulja, katere enačba ima obliko kratke Weierstrassove enačbe za neka a in b iz K. Za krivuljo, podano s kratko Weierstrassovo enačbo, se diskriminanta glasi $\Delta = -16(4a^3 + 27b^2)$.

Nadaljujmo s poenostavljanjem krivulj v primeru $\mathrm{char}(K) = 3$. Prišli smo že do enačbe (2.3), v kateri zaradi karakteristike 3 desne strani ne moremo dopolniti do kuba. Vseeno bi radi z neko dopustno transformacijo eliminirali kakšen člen na desni strani enačbe (2.3). Če je $a'_2 = 0$, je to že opravljeno. V tem primeru se izkaže, da je krivulja supersingularna. Supersingularnosti bo sicer posvečen razdelek (3.1), v katerem bomo navedli natančno definicijo in osnovne lastnosti supersingularnih krivulj. V primeru $a'_2 \neq 0$ pa lahko uporabimo dopustno transformacijo spremenljivk $(X, Y) \mapsto (X + a'_4/a'_2, Y)$, ki nam enačbo (2.3) poenostavi v obliko

$$Y^2 = X^3 + a''_2X^2 + a''_6, \qquad a''_4, a''_6 \in K.$$

Enačbe tega tipa predstavljajo kanonično obliko nesuperingularnih eliptičnih krivulj nad obsegi karakteristike 3. Kljub temu, da smo uspeli enačbo krivulje poenostaviti ravno toliko, kot jo bomo v primeru drugih karakteristik, se enačbe s karakteristiko 3 vseeno redkeje uporabljajo v kriptografske namene. Razlog je najbrž, da je trojiški sistem, ki bi bil naraven za predstavitev takih krivulj, računalniku zaenkrat tuj. V nadaljevanju tega dela se s krivuljami nad obsegi karakteristike 3 ne bomo ukvarjali.

Končno, oglejmo si še zadnji primer, ko je $\mathrm{char}(K) = 2$.

Definicija 2.29 *Eliptične krivulje, definirane nad obsegi s karakteristiko dve, imenujemo* **binarne eliptične krivulje**.

Poskusimo poenostaviti enačbo splošne binarne eliptične krivulje. Začnimo s splošno Weierstrassovo enačbo. Najprej obravnavajmo primer, ko je $a_1 = 0$. Potem dopustna transformacija $(X, Y) \mapsto (X - a_2, Y)$ eliminira člen pri X^2 na desni strani enačbe. Dobimo izomorfno krivuljo z enačbo oblike

$$Y^2 + a_3Y = X^3 + a'_4X + a'_6.$$

Podobno kot pri karakteristiki 3 je v tem primeru krivulja supersingularna.

Oglejmo si še podprimer, ko je $a_1 \neq 0$. V tem primeru bo eliptična krivulja nesupersingularna. Najprej želimo koeficient a_1 pri členu XY v splošni Weierstrassovi enačbi postaviti na 1. V splošno Weierstrassovo enačbo uvedimo dopustno transformacijo spremenljivk

$$(X, Y) \mapsto (a_1^2X, a_1^3Y).$$

Dobljeno enačbo okrajšamo z a_1^6 in dobimo enačbo

$$Y^2 + XY + a'_3Y = X^3 + a'_2X^2 + a'_4X + a'_6.$$

Naslednji korak je eliminacija člena a'_3Y. V zadnjih dveh členih na levi lahko Y izpostavimo. To nas navede na misel, da uporabimo dopustno transformacijo $(X, Y) \mapsto (X - a'_3, Y)$. Dobimo enačbo oblike

$$Y^2 + XY = X^3 + a''_2X^2 + a''_4X + a''_6.$$

Zdaj bi želeli odstraniti še linearni člen $a_4''X$ na desni strani enačbe. Ker je $\mathrm{char}(K) = 2$, se oblika enačbe nič ne spremeni, če namesto Y v enačbi pišem $Y + a$, kjer je a poljuben. Če to ugotovitev uporabimo za $a = a_4''$, to je, če izvedemo dopustno transformacijo $(X, Y) \mapsto (X, Y - a_4'')$, se linearni člen pri X na desni strani uniči. Tako dobimo končni rezultat, to je, enačbo oblike

$$Y^2 + XY = X^3 + a_2X^2 + a_6, \qquad a_4, a_6 \in K, \tag{2.5}$$

Za krivulje s tako enačbo velja $\Delta = a_6$.

Definicija 2.30 *Enačbo (2.5) imenujemo* **binarna nesupersingularna enačba**.

Ta enačba je kanonična oblika za enačbe nesupersingularnih eliptičnih krivulj nad obsegi karakteristike 2. Kratka Weierstrassova enačba opisuje izomorfnostne razrede eliptičnih krivulj nad obsegi karakteristike različne od 2 in 3, binarna nesupersingularna enačba pa nesupersingularne izomorfnostne razrede eliptičnih krivulj nad obsegi karakteristike 2. Med binarnimi krivuljami so slednje najbolj zanimive za uporabo v kriptografiji. Kriptosistemi, ki temeljijo na supesingularnih eliptičnih krivuljah, od leta 1993, ko je bil odkrit t.i. *napad MOV* (Menezes, Okamoto, Vanstone), namreč niso več varni.

Izomorfni eliptični krivulji imata izomorfni grupi na eliptični krivulji. Pri opisu Schoofovega algoritma bomo tako srečali le nesingularne krivulje nad končnimi obsegi velike karakteristike, podane v kratki Weierstrassovi obliki, in binarne nesupersingularne krivulje, podane v kanonični binarni nesupersingularni obliki.

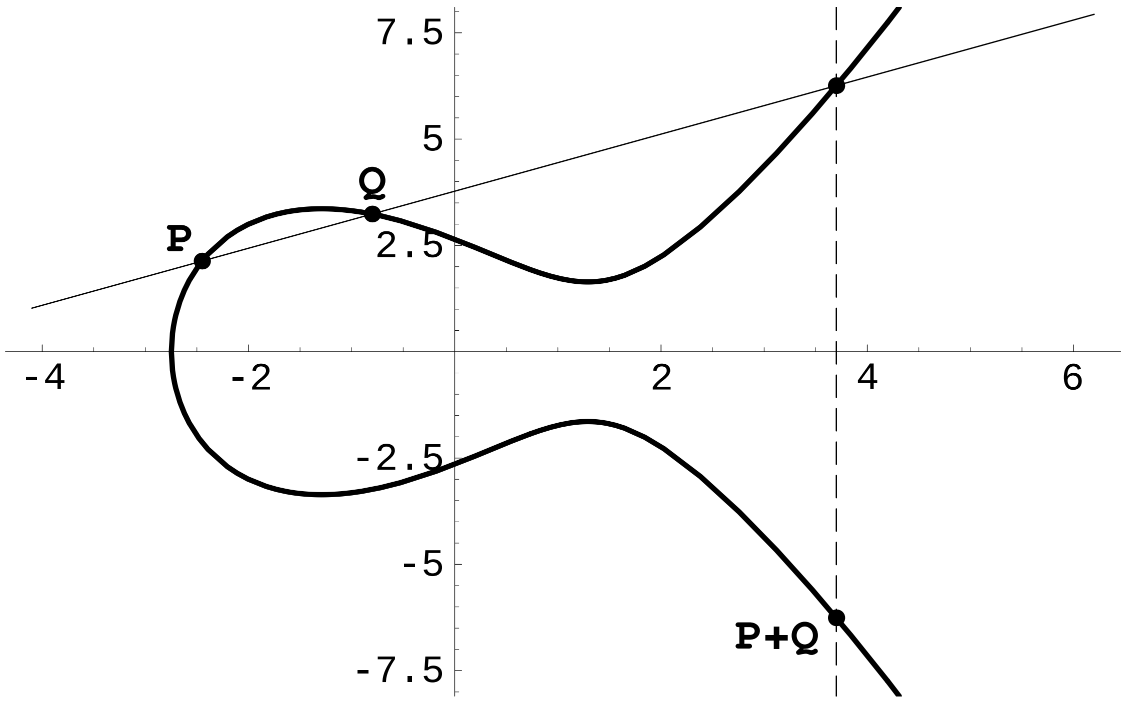
2.2 Grupa na eliptični krivulji

Eliptična krivulja je sama po sebi zanimiv matematični objekt, vendar za kriptografske namene postane zanimiva šele, ko nanjo uvedemo strukturo grupe. To bomo storili v tem razdelku.

2.2.1 Intuitivna predstava

Naj bo K poljuben komutativen obseg in naj bo E eliptična krivulja nad K, definirana s splošno afino Weierstrassovo enačbo iz prejšnjega razdelka, vključno s točko $\mathcal{O}$. Na množici E bomo definirali neko binarno operacijo, za katero bo množica E postala grupa. To grupo bomo imenovali **grupa na eliptični krivulji**. Opozorimo, da bomo v nadaljevanju oznako E uporabljali za eliptično krivuljo in tudi za grupo na njej, saj gre za isto množico. Pri tem bo vedno iz ostalega besedila jasno, na kaj mislimo. Grupa na eliptični krivulji je komutativna, zato bomo zanjo uporabljali aditivni zapis $+$. Nevtralni element grupe na eliptični krivulji bo točka $\mathcal{O}$.

Grupno operacijo na eliptični krivulji definiramo s pomočjo **principa sekante in tangente**. To je geometrijsko pravilo, ki za dve točki množice racionalnih točk $E(\mathbb{R})$ na eliptični krivulji nad $\mathbb{R}$ definira njuno vsoto. Podali ga bomo v nadaljevanju. S pomočjo elementarne analitične geometrije lahko princip sekante in tangente pretvorimo v algebraično obliko. Na ta način dobimo eksplicitne algebraične formule za grupno operacijo v grupi $E(\mathbb{R})$. Dobljene formule so pravzaprav smiselne za poljuben komutativen obseg,



Slika 2.1 Ilustracija principa sekante za grupo racionalnih točk $E(\mathbb{R})$ na eliptični krivulji nad $\mathbb{R}$ z enačbo $Y^2 = X^3 - 5X + 7$.

zato jih v enaki obliki uporabimo za definicijo grupnega zakona za eliptično krivuljo nad poljubnim komutativnim obsegom K. Te formule bomo navedli v naslednjem podrazdelku.

Orišimo napovedani geometrijski princip sekante in tangente za množico racionalnih točk $E(\mathbb{R})$ na dani eliptični krivulji E nad $\mathbb{R}$. Graf tipične množice $E(\mathbb{R})$ in ilustracijo principa sekante je možno videti na sliki 2.1. Naj bosta P in Q dve različni točki na $E(\mathbb{R})$, ki nista $\mathcal{O}$. Želimo definirati točko $P + Q$. V ta namen najprej v ravnini $\mathbb{R}^2$ skozi točki P in Q potegnemo premico. Zaradi nazornosti privzemimo, da premica v nobeni od točk P in Q ni tangentna na E. Če ta premica ni navpična, potem obstaja še tretje presečišče te premice z E. To je posledica dejstva, da ima enačba za krivuljo E stopnjo tri v neznanki X in da premica že seče krivuljo v dveh točkah. Točko, ki jo dobimo s prezrcaljenjem tega tretjega presečišča preko abscisne osi, proglasimo za $P + Q$.

Če je sekanta navpična, torej, če ima enačbo $X = a$, $a \in \mathbb{R}$, potem sta P in Q tudi edini presečišči sekante z E. V tem primeru v projektivnem smislu sekanta seče E v neskončni točki v smeri ordinatne osi. To pa je ravno točka $\mathcal{O}$. Kot prej moramo točko $\mathcal{O}$ še "prezrcaliti" preko abscisne osi. Intuitivno je lahko verjeti, da je zrcalna slika točke $\mathcal{O}$ kar točka $\mathcal{O}$. Točka $\mathcal{O}$ namreč predstavlja neskončno točko projektivne ravnine, ki leži v smeri ordinatne osi, torej v smeri ordinatne osi navzdol ali navzgor. V tem primeru je tako smiselno definirati $P + Q = \mathcal{O}$.

Definicija 2.31 *Zgoraj opisani postopek imenujemo* **princip sekante**. *Grafično ga prikazuje slika 2.1.*

Povejmo še, kako seštejemo točko P samo s seboj, to je, točko podvojimo. V tem primeru moramo v točki P potegniti tangento na E in poiskati presečišče te tangente z E. Kot prej to presečišče preslikamo čez abscisno os. Dobljena točka je točka $P + P$. Tudi v tem primeru je treba na enak način kot pri principu sekante posebej obravnavati primer, ko je tangenta navpična.

Definicija 2.32 *Temu pravilu pravimo* **princip tangente**.

Na domači strani kanadskega podjetja Certicom Corp. (`http://www.certicom.ca`) je na voljo interaktivna predstavitev grupne operacije na eliptični krivulji, kjer lahko z miško izberemo dve točki na krivulji, program pa nam grafično pokaže njuno vsoto in vse za to potrebne račune. Do te predstavitve na Internetu pridemo, če na Certicomovi domači strani po vrsti izberemo crypto research, elliptic curve cryptography, on-line tutorial, experiment: an elliptic curve model.

2.2.2 Eksplicitne algebraične formule za grupno vsoto

V nadaljevanju bomo podali eksplicitne formule za grupni zakon na eliptični krivulji, ki smo ga zgoraj ilustrirali s principom sekante in tangente. Najprej bomo obravnavali eliptično krivuljo s poljubnimi koeficienti a_1, a_2, a_3, a_4, a_6, potem pa bomo izpeljali še poenostavljene formule za posamezne kanonične oblike.

Grupno pravilo za eliptično krivuljo s poljubnimi koeficienti

Naj bo K poljuben komutativen obseg in E eliptična krivulja z enačbo

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6, \qquad a_i \in K.$$

Najprej definirajmo $P + \mathcal{O} = P$ za vsak $P \in E$. Torej je $\mathcal{O}$ enota grupe E. V prejšnjem razdelku smo že definirali konjugacijo poljubne točke $T = (x, y)$ s predpisom $\overline{T} = (x, -y - a_1x - a_3)$. V primeru $K = \mathbb{R}$ imamo za to lepo geometrijsko interpretacijo: konjugiranje pomeni zrcaljenje preko abscisne osi. Omenili smo tudi, da je konjugiranje involucija.

Naj bosta zdaj $\mathcal{O} \neq P = (x_1, y_1)$ in $\mathcal{O} \neq Q = (x_2, y_2)$ poljubni točki na E. Želimo definirati vsoto $P+Q$. Če je $Q = \overline{P}$, definiramo $Q+P = \mathcal{O}$. Vidimo torej, da je konjugacija točke ravno njen inverzni element v grupi. Če pa je $Q \neq \overline{P}$, označimo $P + Q = (x_3, y_3)$ in definirajmo x_3 in y_3. Ločiti moramo dva primera:

1. Če je $P = Q$, definiramo

   $$\lambda = \frac{3x_1^2 + 2a_2x_1 + a_4 - a_1y_1}{2y_1 + a_1x_1 + a_3} \qquad \text{in} \qquad \mu = \frac{-x_1^3 + a_4x_1 + 2a_6 - a_3y_1}{2y_1 + a_1x_1 + a_3}.$$

   Ker hkrati velja $P \neq -Q$, je pogoj $P = Q$ ekvivalenten pogoju $x_1 = x_2$. Ta primer ustreza metodi tangente, zato govorimo o **tangentnem grupnem pravilu**. Opomnimo, da imenovalca v izrazih za λ in μ ne moreta biti enaka 0. Če bi namreč bilo $y_1 = -y_1 - a_1x - a_3$, bi veljalo

   $$\overline{P} = \overline{(x_1, y_1)} = (x_1, -y_1 - a_1x - a_3) = (x_1, y_1) = P = Q,$$

   za kar smo zgoraj predpostavili, da ne velja.

2. Če je $P \neq Q$, definiramo

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} \qquad \text{in} \qquad \mu = \frac{y_1 x_2 - y_2 x_1}{x_2 - x_1}.$$

   Pogoj $P \neq Q$ je ekvivalenten pogoju $x_1 \neq x_2$. Ta primer ustreza metodi sekante, zato pravimo, da smo uporabili **sekantno grupno pravilo**.

V obeh zgornjih primerih potem definiramo

$$x_3 = \lambda^2 + a_1\lambda - a_2 - x_1 - x_2 \qquad \text{in} \qquad y_3 = -(\lambda + a_1)x_3 - \mu - a_3.$$

Z dolgotrajnim direktnim računom lahko pokažemo, da je tako definirana operacija asociativna. Pri tem računu je potrebno dosledno upoštevati vse možne kombinacije različnih predpisov v definiciji grupne operacije in za vsako od njih preveriti, da je spoštovan asociativnostni zakon. Definicije so simetrične glede na sumanda P in Q, zato je operacija komutativna. Iz definicij je očitno, da na E obstaja enota (točka $\mathcal{O}$) in da ima vsak element inverz. Zato E z zgoraj opisano operacijo postane Abelova grupa. Konjugiranje pa je grupni homomorfizem, ki vsakemu elementu priredi njegov inverz. Vsaka podgrupa grupe E je seveda invariantna na konjugiranje.

Ker so vsi koeficienti, ki nastopajo v formulah za vsoto, iz obsega K, velja, da je za vsak vmesni obseg $K \leq L \leq \overline{K}$ množica $E(L)$ podgrupa grupe $E = E(\overline{K})$. Zato je tudi podmnožica racionalnih točk $E(K)$ podgrupa grupe E. V splošnem je ta podgrupa seveda precej manjša od celotne grupe $E = E(\overline{K})$.

Grupno pravilo za kanoničen primer velike karakteristike $q = p > 3$

Za naši odlikovani kanonični obliki se grupno pravilo precej poenostavi. V kanoničnem primeru velike karakteristike $q = p > 3$ je $K = \mathbb{F}_p$, krivulja pa ima enačbo oblike

$$Y^2 = X^3 + aX + b, \qquad a, b \in \mathbb{F}_p, \qquad \Delta = -16(4a^3 + 27b^2) \neq 0.$$

Če prevedemo to na oznake splošnega primera, je $a_1 = a_2 = a_3 = 0$, $a_4 = a$, $a_6 = b$. Konjugirana točka točke $T = (x, y)$ je zdaj enaka $\overline{T} = (x, -y)$. Naj bosta $P = (x_1, y_1)$ in $Q = (x_2, y_2)$ poljubni netrivialni točki na E. Izrazi za λ, μ so takšni:

1. Če je $P = Q$:
$$\lambda = \frac{3x_1^2 + a}{2y_1} \qquad \text{in} \qquad \mu = \frac{-x_1^3 + ax_1 + 2b}{2y_1}.$$

2. Če je $P \neq Q$:
$$\lambda = \frac{y_2 - y_1}{x_2 - x_1} \qquad \text{in} \qquad \mu = \frac{y_1 x_2 - y_2 x_1}{x_2 - x_1}.$$

Formuli za vsoto pa sta

$$x_3 = \lambda^2 - x_1 - x_2 \qquad \text{in} \qquad y_3 = \lambda(x_1 - x_3) - y_1.$$

Grupno pravilo za kanonične nesupersingularne binarne krivulje

V tem primeru je $K = \mathbb{F}_{2^n}$, krivulja pa ima enačbo oblike

$$Y^2 + XY = X^3 + a_2X^2 + a_6, \qquad a_2, a_6 \in \mathbb{F}_{2^n}, \qquad \Delta = a_6 \neq 0.$$

V oznakah splošnega primera to pomeni $a_1 = 1,\ a_3 = a_4 = 0$. Konjugacije točke $T = (x, y)$ je tokrat enaka $\overline{T} = (x, x + y)$. Naj bosta $P = (x_1, y_1)$ in $Q = (x_2, y_2)$ poljubni netrivialni točki na E. Izraza za λ in μ sta v tem primeru naslednje oblike:

1. Če je $P = Q$:
$$\lambda = x_1 + \frac{y_1}{x_1} \qquad \text{in} \qquad \mu = x_1^2.$$

2. Če je $P \neq Q$:
$$\lambda = \frac{y_2 + y_1}{x_2 + x_1} \qquad \text{in} \qquad \mu = \frac{y_1x_2 + y_2x_1}{x_2 + x_1}.$$

Formuli za vsoto pa sta

$$x_3 = \lambda^2 + \lambda + a_2 + x_1 + x_2 \qquad \text{in} \qquad y_3 = (\lambda + 1)x_3 + \mu.$$

Poglavje 3

Algebraično ozadje Schoofovega algoritma

To poglavje je namenjeno predstavitvi tistih algebraičnih lastnosti grupe na eliptični krivulji, ki so pomembne za Schoofov algoritem. V prvem razdelku bomo definirali supersingularne eliptične krivulje in omenili njihove osnovne lastnosti. V naslednjem razdelku bomo govorili o torziji grupe na eliptični krivulji. Zatem sledi razdelek, v katerem bomo definirali Frobeniusovo preslikavo, ki je ključnega pomena za Schoofov algoritem. V zadnjem razdelku bomo uvedli delitvene polinome, ki so tesno povezani s torzijskimi podgrupami in množenjem točk z naravnim številom v grupi na eliptični krivulji. Delitveni polinomi ravno tako precej pomembni za Schoofov algoritem.

3.1 Supersingularne eliptične krivulje

Povrnimo dolg iz poglavja o eliptičnih krivuljah in povejmo nekaj o supersingularnosti eliptičnih krivulj. Supersingularne eliptične krivulje so poseben razred eliptičnih krivulj, za katere je razmeroma preprosto določiti grupno strukturo in kardinalnost množice racionalnih točk. Te lastnosti same po sebi kar kličejo po kriptografski uporabi. Za določanje moči množice racionalnih točk za splošne eliptične krivulje namreč potrebujemo Schoofov algoritem oziroma njegove izpeljanke. V praksi je pogosto potrebno generirati veliko število eliptičnih krivulj in za vsako od njih pognati Schoofov algoritem, kar je lahko precej zamudno. Pri supersingularnih krivuljah tega problema ni bilo. Vendar je leta 1993 prišlo do preobrata, ki je supersingularne krivulje, vsaj kar se neposredne uporabe tiče, v veliki meri potisnil v ozadje. Tega leta je bil namreč odkrit t.i. *napad MOV* na diskretni logaritem v supersingularnih eliptičnih krivuljah. Zaradi tega napada danes supersingularne krivulje veljajo za kriptografsko šibke in se jih pri kriptografski uporabi običajno izogibamo. V tem delu se zato ne bomo podrobneje posvečali supersingularnim krivuljam. Vsebina tega razdelka pa bo v nadaljevanju povsem zadostovala za razumevanje vseh obravnavanih pojmov, ki so povezani s supersingularnostjo. Dokaze vseh trditev in izrekov tega razdelka je možno najti v [Enge, razdelek 3.11].

Definicija 3.1 *Eliptična krivulja je* **supersingularna**, *če je* $\mathbb{Z}$*-algebra endomorfizmov grupe* E *nekomutativna.*

To je splošna definicija supersingularnosti eliptičnih krivulj, ki je v tem delu ne bomo neposredno potrebovali. Ker se bomo ukvarjali le z eliptičnimi krivuljami nad končnimi obsegi, bomo namesto te definicije povsod v nadaljevanju uporabljali karakteristični izrek (3.3). Preden ga navedemo, moramo vpeljati še naslednjo definicijo. Spomnimo se, da oznaka $\#E(\mathbb{F}_q)$ pomeni moč grupe racionalnih točk $E(\mathbb{F}_q)$ na krivulji E.

Definicija 3.2 *Naj bo E eliptična krivulja nad končnim obsegom $\mathbb{F}_q$. Z enačbo $\#E(\mathbb{F}_q) = q + 1 - t$ definiramo količino $t \in \mathbb{Z}$, ki jo imenujemo* **sled Frobeniusa**.

Hassejev izrek (2.17) lahko s pomočjo sledi Frobeniusa reformuliramo v enačbo $t^2 \leq 4q$. Iz razdelka 3.3 o Frobeniusovem endomorfizmu bo razvidno, zakaj smo si za količino t izbrali na prvi pogled tako nenavadno ime.

Izrek 3.3 *Eliptična krivulja E, definirana nad končnim obsegom $\mathbb{F}_{p^n}$, je supersingularna natanko tedaj, ko karakteristika p deli sled Frobeniusa t te krivulje.* ∎

Dokaz je možno najti v [Enge, razdelek 3.11]. S pomočjo tega izreka in Hassejevega izreka izpeljemo naslednjo trditev.

Trditev 3.4 *Eliptična krivulja, definirana nad končnim obsegom $\mathbb{F}_p$, kjer je $p \geq 5$ poljubno praštevilo, je supersingularna natanko takrat, ko je $t = 0$.*

Dokaz:
($\Leftarrow$) Trditev očitno sledi iz izreka (3.3).

($\Rightarrow$) Po Hassejevem izreku velja $t^2 \leq 4p$. Za supersingularne krivulje je po izreku (3.3) $p \mid t$. Če je $t \neq 0$, je tako $p^2 \leq t^2 \leq 4p$, kar je protislovje. ∎

V primerih $p = 2, 3$ pa supersingularnost opiše naslednji rezultat.

Izrek 3.5 *Eliptična krivulja, definirana nad končnim obsegom karakteristike* 2 *ali* 3, *je supersingularna natanko takrat, ko je njena j-invarianta enaka* 0. ∎

Dokaz je naveden v [Enge, razdelek 3.11].

Pripomnimo, da obstajajo tudi t.i. *anomalne eliptične krivulje* [Blake, str. 35], ki so, podobno kot supersingularne eliptične krivulje, nek poseben razred eliptičnih krivulj. Za anomalne eliptične krivulje nad končnimi obsegi je določitev moči grupe racionalnih točk zelo preprosta, praktična izvedba aritmetike pa precej enostavnejša kot za splošne eliptične krivulje. Zaradi slabih izkušenj s pomanjkljivo varnostjo supersingularnih eliptičnih krivulj se v kriptografiji anomalnim eliptičnim krivuljam že v naprej izogibamo. V tem trenutku sicer ni znano, da bi bil problem diskretnega logaritma za anomalne eliptične krivulje bistveno lažji od problema diskretnega logaritma za splošne eliptične krivulje. Celotna kriptografska skupnost pa (že nekaj časa) pričakuje skorajšnji pojav takega algoritma.

3.2 Torzija v eliptični grupi

Schoofov algoritem je prepleten z idejami, povezanimi s torzijo v grupi na eliptični krivulji. V tem razdelku bomo predstavili osnovne torzijske lastnosti grupe na eliptični krivulji. Zanimale nas bodo predvsem eliptične krivulje, definirane nad končnimi obsegi.

Definicija 3.6 *Naj bo E poljubna eliptična krivulja in m poljubno naravno število. Preslikavo*

$$[m]\,:\,E \longrightarrow E$$

$$[m]\,:\,P \longmapsto mP,$$

imenujemo **množenje z m** *in jo označimo z $[m]$. Oznaka mP pomeni vsoto m seštevancev $P+P+\ldots+P$ v grupi E.*

Enostavno je preveriti, da je za vsak m množenje z m endomorfizem grupe E. V algoritmih s področja kriptografske uporabe eliptičnih krivulj so množenja z m zelo pogosta operacija v grupi na eliptični krivulji. Z njimi je v zvezi naslednja definicija.

Definicija 3.7 *Naj bo $m \in \mathbb{N}$ in E eliptična krivulja. Potem pravimo, da je $P \in E$* **točka torzije m**, *če je $mP = \mathcal{O}$. Vse točke torzije m tvorijo podgrupo grupe E, ki jo označimo z $E[m]$. Imenujemo jo* **torzijska podgrupa reda m**.

Definicija 3.8 *Označimo $E[m]^* = E[m]\backslash\{\mathcal{O}\}$.*

Trditev 3.9 *Vsaka grupa na eliptični krivulji nad končnim obsegom moči $q = p^n$ je* **torzijska**, *kar pomeni, da za vsak $P \in E = E(\overline{\mathbb{F}_q})$ obstaja neko naravno število k, da je $kP = \mathcal{O}$.*

Dokaz: Naj bo $P = (x, y) \in E$, torej $x, y \in \overline{\mathbb{F}_q}$. V podrazdelku 1.4.2 o razpadnih obsegih končnih obsegov smo videli, da je vsak element razpadnega obsega končnega obsega $\mathbb{F}_q$ vsebovan v nekem dovolj velikem končnem obsegu karakteristike p. Zato obstajata taki naravni števili r in s, da je $x \in \mathbb{F}_{q^r}$ in $y \in \mathbb{F}_{q^s}$. Potem velja $x, y \in \mathbb{F}_{q^{rs}}$. Potemtakem točka P leži v podgrupi $E(\mathbb{F}_{q^{rs}})$. Ta podgrupa je končna, saj je $E(\mathbb{F}_{q^{rs}}) \subset \mathbb{F}_{q^{rs}} \times \mathbb{F}_{q^{rs}}$, ta kartezični produkt pa je končen. Z dokazom smo zdaj pri kraju. Če namreč označimo $k = \#E(\mathbb{F}_{q^{rs}})$, po Lagrangevem izreku iz teorije grup velja $kP = \mathcal{O}$. ∎

Opozorimo, da iz te trditve še ne sledi, da je vsaka grupa na eliptični krivulji E nad končnim obsegom končna. Tak zaključek bi lahko napravili, če bi bila grupa E končno generirana. Trditev (2.14) nam je povedala, da grupa na eliptični krivulji nikdar ni končna grupa. Zato lahko zaključimo le, da E nikdar ni končno generirana grupa. O strukturi torzijskih podgrup nad končnim obsegom znamo povedati naslednje:

Izrek 3.10 *Naj bo E grupa na eliptični krivulji nad končnim obsegom $\mathbb{F}_q$, kjer je $q = p^n$ za neko praštevilo p in neko naravno število n. Potem velja:*

1. *Če je $p \not| m$, potem je $E[m] \cong \mathbb{Z}_m \oplus \mathbb{Z}_m$.*
2. *Če je $m = p^k$, kjer je $k \in \mathbb{N}$ in je E supersingularna krivulja, potem je $E[m] = \{\mathcal{O}\}$.*
3. *Če je $m = p^k$, kjer je $k \in \mathbb{N}$ in E ni supersingularna krivulja, potem je $E[m] \cong \mathbb{Z}_m$.*

Dokaz: Ta izrek ni elementaren, njegov dokaz pa presega okvire tega dela. Celoten dokaz je možno najti v [Enge, str. 68-86]. ∎

Ta izrek nam torej poda algebraično strukturo vseh torzijskih podgrup, razen tistih, katerih red je deljiv s karakteristiko obsega p in hkrati ni potenca karakteristike p. Ta pomankljivost nas ne bo motila, ker se bomo v nadaljevanju ukvarjali le s torzijskimi podgrupami velikega praštevilskega reda $\ell \neq p$. Torej nam bo zadostoval že primer 1. zgornjega izreka, ki pravi, da je $E[\ell] \stackrel{\sim}{=} \mathbb{Z}_\ell \oplus \mathbb{Z}_\ell$. Naslednja trditev opiše strukturo te grupe.

Trditev 3.11 *Če je ℓ praštevilo, potem ima grupa $\mathbb{Z}_\ell \oplus \mathbb{Z}_\ell$ poleg trivialnih podgrup še natanko $\ell + 1$ podgrup. Vse te podgrupe imajo moč ℓ.*

Dokaz: Ker je $\left|\mathbb{Z}_\ell \oplus \mathbb{Z}_\ell\right| = \ell^2$ in je ℓ praštevilo, ima vsaka podgrupa moč 1, ℓ ali ℓ^2. Preštejmo torej, koliko je podgrup moči ℓ. Ker je ℓ praštevilo, je presek poljubnih dveh različnih podgrup moči ℓ trivialna podgrupa $\{\mathcal{O}\}$. Vsak netrivialen element iz $\mathbb{Z}_\ell \oplus \mathbb{Z}_\ell$ ima red ℓ, zato ima podgrupa, ki jo generira, moč ℓ. Torej je vsak netrivialen element grupe $\mathbb{Z}_\ell \oplus \mathbb{Z}_\ell$ vsebovan v neki podgrupi moči ℓ. Potemtakem podgrupe moči ℓ disjunktno pokrijejo $\mathbb{Z}_\ell \oplus \mathbb{Z}_\ell \backslash \{0^2\}$. Vsaka podgrupa vsebuje $\ell - 1$ netrivialnih elementov. Zato ima grupa $\mathbb{Z}_\ell \oplus \mathbb{Z}_\ell$ natanko $(\ell^2 - 1)/(\ell - 1) = \ell + 1$ različnih podgrup moči ℓ. ∎

Na grupo $E[\ell] \stackrel{\sim}{=} \mathbb{Z}_\ell \oplus \mathbb{Z}_\ell$ lahko gledamo kot na dvodimenzionalni vektorski prostor nad $\mathbb{F}_\ell$. S tega vidika so podgrupe moči ℓ natanko linearni podprostori prostora $E[\ell]$ dimenzije 1. Te podgrupe igrajo pomembno vlogo v Atkin-Elkiesovi izboljšavi Schoofovega algoritma, ki je predmet zadnjega razdelka tega diplomskega dela. Nastopajo namreč pri reševanju lastnega problema za ℓ-ti Frobeniusov operator, saj so naravni kandidati za lastne podprostore.

3.3 Frobeniusov endomorfizem eliptične grupe in njegove lastnosti

V tem razdelku se bomo ukvarjali le z grupami na eliptičnih krivuljah nad končnimi obsegi. Povsod v tem razdelku naj tako velja dogovor, da je $E = E(\overline{\mathbb{F}_q})$. V teoriji eliptičnih krivulj se srečamo z dvema pomembnima preslikavama, ki obe nosita ime po Frobeniusu. **Frobeniusov avtomorfizem** končnih obsegov $\mathbb{F}_q$ je preslikava, ki element x preslika v x^p, kjer je $p = \operatorname{char} \mathbb{F}_q$. To preslikavo smo definirali že v prvem poglavju v razdelku 1.3 o mreži podobsegov končnega obsega. V nadaljevanju tega dela pa se bomo ukvarjali le s *Frobeniusovim endomorfizmom grupe na eliptični krivulji E.*

Definicija 3.12 Frobeniusov endomorfizem grupe na eliptični krivulji $E(\overline{\mathbb{F}_q})$ *je preslikava*

$$\varphi : E \longrightarrow E$$

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

$$\varphi : \mathcal{O} \longmapsto \mathcal{O}.$$

Oznaka φ za Frobeniusov endomorfizem grupe na eliptični krivulji je v literaturi ustaljena praksa. Kot je razvidno iz definicije, smo morali za Frobeniusov endomorfizem navesti dva predpisa, enega za točko v neskončnosti $\mathcal{O}$ in enega za preostale točke na krivulji E. V definiciji Frobeniusovega endomorfizma smo pravzaprav že napovedali naslednjo trditev.

Trditev 3.13 *Preslikava φ je endomorfizem grupe E.*

Skica dokaza: Trditev preverimo z direktnim računom s pomočjo algebraičnih formul, ki definirajo grupno operacijo na eliptični krivulji. Pri tem moramo upoštevati vse različne primere, ki nastopajo v definiciji grupne operacije, zaradi česar je dokaz precej dolg, ni pa posebno zahteven. Za dokaz je ključnega pomena dejstvo, da so koeficienti eliptične krivulje E iz $\mathbb{F}_q$, kjer za vsak element x velja $x^q = x$. ■

Iz definicije Frobeniusovega endomorfizma neposredno sledi, da je φ injektiven, saj je njegovo jedro trivialno. Očitno se namreč edino element $\mathcal{O}$ preslika v $\mathcal{O}$.

Na tem mestu se spet srečamo s sledjo Frobeniusa t, ki smo jo definirali na začetku razdelka 3.1 o supersingularnih eliptičnih krivuljah. Definicija sledi Frobeniusa se pravzaprav sklicuje le na grupo racionalnih točk $E(\mathbb{F}_q)$ in ne na celotno krivuljo E. Zato je presenetljivo, da sled Frobeniusa nastopa tudi pri pojavih, ki nimajo neposredne zveze z $E(\mathbb{F}_q)$, ampak le z $E(\overline{\mathbb{F}_q})$. Primer take lastnosti podaja naslednji izrek, ki je bistvenega pomena za Schoofov algoritem. V Schoofovem algoritmu se bomo na ta izrek kar naprej sklicevali.

Izrek 3.14 *Za Frobeniusov endomorfizem eliptične krivulje E velja*

$$\varphi^2(P) - t\varphi(P) + qP = \mathcal{O} \qquad \text{za vsak } P \in E. \tag{3.1}$$

■

Za dokaz tega izreka potrebujemo t.i. *Weilovo parjenje* [Enge, str. 92] in veliko predpriprave iz algebraične geometrije, zato ga izpustimo. Dokaz je zelo soroden dokazu Hassejevega izreka. V [Enge, str. 98-100] je možno najti tako različico dokaza, ki hkrati dokaže naš izrek (3.14) in Hassejev izrek. Morda na tem mestu ni odveč komentar o pomenu simbolov, ki nastopajo v enačbi (3.1), ki jo podaja zgornji izrek (3.14), saj se v tem delu na tem mestu prvič srečamo z enačbami takega tipa. Oznaka $\varphi^2(P)$ pomeni seveda $\varphi(\varphi(P))$. Znaka $-$ in $+$ sta po vrsti simbola za odštevanje in seštevanje v Abelovi grupi E. Oznaka qP pomeni, kot smo že omenili na začetku razdelka, vsoto q sumandov $P + P + \ldots + P$. Pri tem gre seveda za seštevanje v grupi E. Analogen je pomen izraza $t\varphi(P)$.

Vsaka Abelova grupa je $\mathbb{Z}$-modul. Zato je tudi grupa na eliptični krivulji E modul nad $\mathbb{Z}$. Preslikava φ v tej luči postane endomorfizem $\mathbb{Z}$-modula E. Zato lahko na osnovno zvezo gledamo tudi kot enakost

$$\varphi^2 - [t]\varphi + [q] = 0$$

v $\mathbb{Z}$-algebri endomorfizmov $\mathbb{Z}$-modula E. Pri tem sta $[q]$ in $[t]$ endomorfizma množenja s q in s t v grupi E. Pogled z vidika algebre endomorfizmov nam bo koristil, ko se bomo specializirali na zožitve preslikave φ na torzijske podgrupe praštevilskega reda.

Podgrupa $E(\mathbb{F}_q)$ je invariantna za endomorfizem φ. Če je namreč $(x, y) \in E(\mathbb{F}_q)$, je potem $x, y \in \mathbb{F}_q$, torej je $(x^q, y^q) = (x, y) \in E(\mathbb{F}_q)$. Za φ so invariantne tudi vse torzijske podgrupe $E[m]$, kjer je $m \in \mathbb{N}$. Ker je φ injektivna, velja namreč za vsako naravno število $m \in \mathbb{N}$ in za vsako točko $P \in E$

$$mP = \mathcal{O} \iff m\varphi(P) = \mathcal{O}.$$

Definicija 3.15 *Naj bo zdaj ℓ poljubno praštevilo. Homomorfizem φ grupe E lahko po pravkar povedanem zožimo na grupo $E[\ell]$. To zožitev imenujemo* **ℓ-ti Frobeniusov operator** *in ga označimo s φ_ℓ.*

Izraz operator uporabljamo zato, ker je $E[\ell]$ vektorski prostor nad obsegom $\mathbb{F}_\ell$, preslikava φ_ℓ pa je očitno linearna na tem vektorskem prostoru. Ker je operator φ_ℓ injektiven in $E[l]$ vektorski prostor s končno elementi, je operator φ_ℓ tudi surjektiven. Torej je φ_ℓ avtomorfizem vektorskega prostora $E[\ell]$ nad $\mathbb{F}_\ell$.

Definicija 3.16 *V nadaljevanju bomo uporabljali oznaki $t_\ell = t \bmod \ell$ in $q_\ell = q \bmod \ell$.*

Izrek (3.14) lahko zdaj prepišemo v enačbo

$$\varphi_\ell^2(P) - t_\ell \varphi_\ell(P) + q_\ell P = \mathcal{O},$$

ki velja za vsak $P \in E[\ell]$. Ker zdaj delamo v torzijski podgrupi reda ℓ, smo koeficiente te enačbe reducirali modulo ℓ. V $\mathbb{F}_\ell$-algebri endomorfizmov vektorskega prostora $E[\ell]$ nad $\mathbb{F}_\ell$ torej zadošča operator φ_ℓ enakosti

$$\varphi_\ell^2 - t_\ell \varphi_\ell + q_\ell = 0. \tag{3.2}$$

Torej polinom $X^2 - t_\ell X + q_\ell$ uniči operator φ_ℓ. Ta polinom je pravzaprav karakteristični polinom za operator φ_ℓ na dvodimenzionalnem vektorskem prostoru $E[\ell] \cong \mathbb{F}_\ell \oplus \mathbb{F}_\ell$ nad $\mathbb{F}_\ell$ [Enge, str. 114].

3.4 Delitveni polinomi

V tem razdelku bomo uvedli *delitvene polinome*, ki so tesno povezani s torzijskimi podgrupami in operacijo množenja točke na eliptični krivulji z naravnim številom. To povezavo bodo natančno opisale trditve tega razdelka. Delitvene polinome v tem delu obravnavamo zato, ker imajo v Schoofovem algoritmu pomembno vlogo.

3.4.1 Definicija delitvenih polinomov

Eliptična krivulja E naj bo v tem razdelku definirana nad poljubnim komutativnim obsegom K. Podana naj bo s splošno Weierstrassovo enačbo s koeficienti $a_i \in K$, $i = 1, 2, 3, 4, 6$. Delitveni polinomi so posebna družina polinomov, ki je neposredno povezana s torzijo grupe na eliptičnimi krivulji in množenji z m, ki smo jih definirali v razdelku 3.2 o torziji grupe na eliptični krivulji. Izraz delitveni polinomi je pravzaprav le skupno ime za več različnih družin polinomov. Za vsako od teh družin uporabljamo ime delitveni polinomi, zato mora biti vedno iz konteksa jasno, katero družino konkretno imamo na nekem mestu v mislih. V literaturi najpogosteje srečamo družine delitvenih polinomov $\psi_m, \theta_m, \omega_m, \overline{f}_m$ in f_m, kjer zavzame indeks m vrednosti $0, 1, 2, \ldots$. Vse te družine bomo v nadaljevanju definirali. Pri tem v splošnem velja $\psi_m, \theta_m, \omega_m \in K[X, Y]$ in $\overline{f}_m, f_m \in K[X]$. Koeficienti vsakega delitvenega polinoma torej ležijo v obsegu K, nad katerim je definirana eliptična krivulja. Delitvene polinome vedno prirejamo posamezni eliptični krivulji. Različne eliptične krivulje imajo tako v splošnem različne delitvene polinome, čeprav uporabljamo zanje isto oznako.

Pri mnogih enačbah tega in naslednjih poglavij zaradi preglednosti delitveni polinomi nastopajo brez argumentov. V takšnem primeru naj se vedno razume, da je argument X, če gre za polinom ene spremenljivke, oziroma, da sta argumenta X in Y, če gre za polinom dveh spremenljivk.

Definirajmo najprej polinome ψ_m. Postavimo

$$\begin{aligned}
\psi_0 &= 0, \quad \psi_1 = 1, \quad \psi_2 = 2Y + a_1X + a_3, \\
\psi_3 &= 3X^4 + b_2X^3 + 3b_4X^2 + 3b_6X + b_8 \qquad \text{in} \\
\psi_4 &= (2X^6 + b_2X^5 + 5b_4X^4 + 10b_6X^3 + 10b_8X^2 + (b_2b_8 - b_4b_6)X + b_4b_8 - b_6^2)\psi_2.
\end{aligned}$$

Količine b_2, b_4, b_6, b_8 smo uvedli v definiciji (2.23) diskriminante v podrazdelku 2.1.4. Preostale polinome ψ_m definiramo s pomočjo rekurzivnih formul

$$\begin{aligned}
\psi_{2m+1} &= \psi_{m+2}\psi_m^3 - \psi_{m-1}\psi_{m+1}^3, & m \geq 2, \\
\psi_{2m} &= \frac{1}{\psi_2}(\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2)\psi_m, & m > 2.
\end{aligned}$$

V zvezi z zadnjo vrstico pripomnimo, da je z indukcijo mogoče pokazati, da je števec v izrazu za ψ_{2m} vedno deljiv s ψ_2^2. Zato so za $m \geq 1$ vsi polinomi ψ_{2m} deljivi s ψ_2.

Definicija 3.17 *S pomočjo enačbe eliptične krivulje v splošni Weierstrassovi obliki lahko izrazimo količino Y^2 s količinama X in Y in parametri krivulje. Z upoštevanjem te transformacije lahko vsak polinom dveh spremenljivk X, Y pretvorimo v obliko $s(X)Y + t(X)$, kjer sta s in t neka (morda ničelna) polinoma iz $K[X]$. Pravimo, da smo polinom r* **okrajšali za enačbo krivulje E***. Polinom $s(X)Y + t(X)$ se imenuje* **reducirani predstavnik polinoma** *r. Označimo ga s $r \bmod E$.*

Vrednosti delitvenih polinomov ψ_m bomo v nadaljevanju vedno računali le v točkah eliptične krivulje. Če v netrivialni točki P krivulje E ovrednotimo polinom $\psi_m \bmod E$, $m \geq 0$, je rezultat isti element obsega $\overline{K}$, kot če v P ovrednotimo kar originalni polinom ψ_m. Zato naj od zdaj dalje velja dogovor, da oznaka ψ_m pomeni reduciranega predstavnika predhodno definiranega polinoma ψ_m. Z originalnimi polinomi ψ_m pa se ne bomo več ukvarjali. Enak dogovor naj že v naprej velja tudi za ostale družine delitvenih polinomov, ki jih bomo šele definirali. V jeziku algebraične geometrije to pomeni, da namesto v kolobarju $K[X, Y]$ delamo v kolobarju $K[X, Y]/E$. Pri tem E označuje ideal v $K[X, Y]$, generiran s polinomom dveh spremenljivk, ki ga dobimo, če v splošni Weierstrassovi enačbi premečemo vse člene na eno stran. Ko smo govorili o varietetah, smo že omenili, da je ta polinom vedno nerazcepen. Ker je $K[X,Y]$ Gaussov kolobar, je kolobar $K[X,Y]/E$ cel komutativen kolobar z enico. Vendar v splošnem ni Gaussov [Enge, str. 14]. Kolobar $K[X,Y]/E$ bomo v nadaljevanju označevali s $K[X,Y] \bmod E$.

Naslednji izrek podaja ključno lastnost polinomov ψ_m.

Izrek 3.18 *Naj bo P od $\mathcal{O}$ različna točka eliptične krivulje E in naj bo $m \in \mathbb{N}$. Potem je $P \in E[m]$ natanko takrat, ko je $\psi_m(P) = 0$.* ∎

Dokaz tega izreka je možno najti v [Enge, str. 91]. Opomnimo, da citirani izrek uporablja terminologijo *deliteljev (angl. divisors)*, ki so predhodno uvedeni v citiranem delu.

Definirajmo zdaj družini θ_m in ω_m. Naj bo $P = (x,y) \in E \setminus \mathcal{O}$. Koordinate točke $2P = (w,z)$ dobimo s pomočjo tangentnega grupnega zakona. Iz algebraičnih formul očitno sledi, da sta w in z racionalni funkciji koordinat x in y. S tem premislekom lahko nadaljujemo in ugotovimo, da sta za vsak $m \in \mathbb{N}$ koordinati točke mP racionalni funkciji koordinat x in y. Za vsak m torej lahko konstruiramo par racionalnih funkcij (g_m, h_m), ki ležita v $K(X,Y)$ in imata lastnost, da za vsako točko $(x,y) \in E \setminus E[m]$ velja

$$m(x,y) = (g_m(x,y), h_m(x,y)).$$

V točkah iz $E[m]$ zgornja formula očitno ne bi mogla veljati. Funkciji g_m in h_m imata pole natanko v točkah iz $E[m]$ in sta v splošnem med seboj različni. Njuni eksplicitni obliki podaja naslednja trditev.

Trditev 3.19 *Naj bo E eliptična krivulja nad obsegom K in naj bo $m \in \mathbb{N}$. Potem obstajata polinoma $\theta_m, \omega_m \in K[X,Y]$, tako da za vsako točko $P = (x,y) \in E \setminus E[m]$ velja*

$$mP = (g_m, h_m) = \left(\frac{\theta_m(x,y)}{\psi_m^2(x,y)}, \frac{\omega_m(x,y)}{\psi_m^3(x,y)} \right).$$

Skica dokaza: Ta trditev je naloga 3.7. iz knjige [Silverman, str. 105]. Preverimo jo lahko z upoštevanjem definicije grupne operacije in rekurzivnih formul za deliteljske polinome.∎

Ker sta racionalni funkciji g_m, h_m in polinom ψ_m enolično določeni, sta tudi polinoma θ_m, ω_m iz zgornje trditve enolično določena. Oba polinoma se lepo izražata z osnovnimi polinomi ψ_m [Blake, str. 40]:

$$\theta_m = X\psi_m^2 - \psi_{m-1}\psi_{m+1}, \quad m \geq 1.$$

Če je $\mathrm{char}(K) \neq 2$, je polinome ω_m možno določiti iz rekurzivne zveze

$$2\psi_m\omega_m = \psi_{2m} - (a_1\theta_m + a_3\psi_m^2)\psi_m^2, \quad m \geq 1.$$

V primeru $\mathrm{char}(K) = 2$ pa je formula za ω_m nekoliko bolj zapletena. Za nesupersingularni primer jo bomo navedli v nadaljevanju.

S pomočjo indukcije zlahka pokažemo, da za lihe m polinom ψ_m pravzaprav sploh ni odvisen od Y, torej, da leži kar v $K[X]$. Za sode m pa velja, da je polinom ψ_m produkt polinoma ψ_2 z nekim polinomom iz $K[X]$. Zato je smiselna naslednja definicija.

Definicija 3.20

$$\overline{f}_m = \begin{cases} \psi_m & ; \quad \textit{če je } m \textit{ lih} \\ \psi_m/\psi_2 & ; \quad \textit{če je } m \textit{ sod.} \end{cases}$$

Polinomi $\overline{f}_m$ torej ležijo v $K[X]$. Ker so odvisni samo od ene spremenljivke, je računanje z njimi preprostejše. Z njihovo pomočjo lahko tudi poenostavimo izrek (3.18).

Trditev 3.21 *Naj bo $P = (x, y)$ točka iz $E(\overline{K})\backslash\{\mathcal{O}\}$, ki ni reda 2, torej $2P \neq \mathcal{O}$. Naj bo $m \geq 3$. Potem je $P \in E[m]$ natanko takrat, ko je $\overline{f}_m(x) = 0$.*

Dokaz: Za lihe m sledi trditev neposredno iz izreka (3.18). Za sode m pa velja $\psi_m = \psi_2\overline{f}_m$. Ker je $P \notin E[2]$, je $\psi_2(P) \neq 0$ in torej velja

$$P \in E[m] \iff \psi_m(P) = 0 \iff \overline{f}_m(x) = 0. \qquad \blacksquare$$

S pomočjo rekurzivnih formul za polinome ψ_m lahko zlahka izpeljemo rekurzivne formule za polinome $\overline{f}_m$. V nadaljevanju teh formul ne bomo potrebovali, zato jih ne bomo posebej navedli. S pomočjo indukcije lahko iz rekurzivnih formul za polinome ψ_n izpeljemo, da je stopnja polinoma $\overline{f}_m$ kvečjemu $(m^2 - 1)/2$, če je m lih, in kvečjemu $(m^2 - 4)/2$, če je m sod.

Analizirajmo podrobneje primer, ko je K končen obseg, $m = \ell$ pa neko praštevilo, različno od karakteristike p. Ta rezultat bo pomemben pri oceni časovne zahtevnosti Schoofovega algoritma.

Trditev 3.22 *Naj bo ℓ praštevilo, različno od $p = \mathrm{char}(\mathbb{F}_q)$. Potem je stopnja delitvenega polinoma $\overline{f}_\ell$ natanko $(\ell^2 - 1)/2$.*

Dokaz: Iz razdelka 3.2 o torziji grupe na eliptični krivulji vemo, da za praštevilo $\ell \neq p$ velja $E[\ell] \cong \mathbb{Z}_\ell \oplus \mathbb{Z}_\ell$. Ker je enačba eliptične krivulje stopnje dve v spremenljivki Y, za vsak $x \in \overline{\mathbb{F}_q}$ obstajata na E kvečjemu dve točki, ki imata prvo koordinato enako x. Torej med prvimi koordinatami točk iz $E[\ell]\backslash\{\mathcal{O}\}$ najdemo vsaj $(\ell^2 - 1)/2$ različnih elementov obsega $\overline{\mathbb{F}_q}$. Vsi ti elementi pa so po trditvi (3.21) ničle polinoma $\overline{f}_\ell$. Torej ima ta polinom vsaj $(\ell^2 - 1)/2$ različnih ničel. Potemtakem je njegova stopnja vsaj $(\ell^2 - 1)/2$. Ker smo v odstavku pred trditvijo utemeljili neenakost v drugo smer, je trditev dokazana. $\blacksquare$

V nadaljevanju obravnavo delitvenih polinomov zožimo na oba kanonična primera.

3.4.2 Kanoničen primer karakteristike $p > 3$

V tem primeru je enačba krivulje E v kratki Weierstrassovi obliki. Rekurzivne formule za ψ_m se precej poenostavijo. Velja namreč

$$\begin{aligned}
\psi_0 &= 0, \quad \psi_1 = 1, \quad \psi_2 = 2Y, \\
\psi_3 &= 3X^4 + 6aX^2 + 12bX - a^2, \\
\psi_4 &= 4Y(X^6 + 5aX^5 + 20bX^3 - 5a^2X^2 - 4abX - 8b^2 - a^3), \\
\psi_{2m+1} &= \psi_{m+2}\psi_m^3 - \psi_{m-1}\psi_{m+1}^3, && m \geq 2, \\
\psi_{2m} &= \frac{1}{2Y}(\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2)\psi_m, && m > 2.
\end{aligned}$$

Množenje z m je zdaj oblike

$$mP = \left(x - \frac{\psi_{m-1}\psi_{m+1}}{\psi_m^2}, \frac{\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2}{4y\psi_m^3}\right).$$

Tudi zveza med ψ_m in f_m se poenostavi, ker je zdaj $\psi_2 = 2Y$. Formulo za množenje z m zlahka prepišemo v obliko, v kateri nastopajo polinomi $\overline{f}_m$.

3.4.3 Nesupersingularni binarni primer

Privzamemo lahko, da je enačba krivulje v kanonični binarni nesupersingularni obliki. V tem primeru se polinomi ψ_m poenostavijo v

$$\begin{aligned}
\psi_0 &= 0,\ \psi_1 = 1,\ \psi_2 = X, \\
\psi_3 &= X^4 + X^3 + a_6, \\
\psi_4 &= X^6 + a_6X^2, \\
\psi_{2m+1} &= \psi_{m+2}\psi_m^3 + \psi_{m-1}\psi_{m+1}^3, & m \geq 2, \\
\psi_{2m} &= \frac{1}{X}(\psi_{m+2}\psi_{m-1}^2 + \psi_{m-2}\psi_{m+1}^2)\psi_m, & m > 2.
\end{aligned}$$

Vidimo, da so v tem primeru vsi polinomi ψ_m odvisni le od spremenljivke X. Na tem mestu uvedemo zadnjo od družin delitvenih polinomov, namreč družino f_m. Definirali jo bomo le za oba kanonična primera.

Definicija 3.23 *V karakteristiki dve v kanoničnem nesupersingularnem primeru preprosto definiramo $f_m = \psi_m$. Za karakteristiko $p \geq 3$ pa definiramo $f_m = \overline{f}_m$.*

Vidimo, da v vsakem primeru velja $f_m \in K[X]$. V karakteristiki dve v nesupersingularnem primeru velja $f_m = \psi_2\overline{f}_m = X\overline{f}_m$.

Nadaljujmo z analizo vloge delitvenih polinomov v karakteristiki dve. Množenje z $m \geq 2$ ima v tem primeru za točke $P = (x, y) \in E\backslash E[m]$ obliko

$$mP = \left(x + \frac{f_{m-1}f_{m+1}}{f_m^2}, x + y + \frac{(x^2 + x + y)f_{m-1}f_mf_{m+1} + f_{m-2}f_{m+1}^2}{xf_m^3}\right). \tag{3.3}$$

To formulo je mogoče najti v [Blake, str. 42]. Iz tega izraza lahko tudi razberemo formulo za ω_m, ki smo jo za ta primer predhodno obljubili. Formulo dobimo tako, da drugo koordinato zgornjega rezultata izenačimo s ω_m/f_m^3 in iz dobljene enačbe izrazimo ω_m. Te formule ne bomo navedli, saj polinome ω_m potrebujemo le za izračun točke mP, za kar že imamo zgornjo formulo (3.3).

Poglavje 4

Schoofov algoritem

V tem poglavju bomo podrobno opisali Schoofov algoritem. Večina standardnih referenc se podrobnemu opisu in rigorozni utemeljitvi vseh korakov Schoofovega algoritma v večji ali manjši meri izogne in navede le bistvene korake. Schoofov algoritem je namreč dokaj zapleten, njegova analiza pa hitro postane zelo tehnična in mukotrpna. Poleg tega je za podrobno analizo potrebno ločiti obravnavo karakteristike dve in ostalih karakteristik. Že sama priprava Schoofovega algoritma tako zahteva precej simboličnega računanja. Formule, ki jih dobimo s pomočjo teh izračunov, je potrebno vgraditi neposredno v izvorno kodo Schoofovega algoritma. Šele zatem lahko Schoofov algoritem sploh poženemo.

Sam sem se želel dokopati do zgoraj omenjenih formul in Schoofov algoritem razdelati do konca. Zato sem sam do konca izdelal in dokazal vse korake Schoofovega algoritma, ki jih nisem našel v omenjenih referencah. Končni rezultat tega je, da v tem delu navedem eksplicitne formule za vse polinome in druge enačbe, ki jih je potrebno vgraditi v Schoofov algoritem. Vse dolgotrajne izračune sem dodatno preveril s paketom za simbolno računanje (Mathematica).

Namen Schoofovega algoritma je izračun moči grupe $E(\mathbb{F}_q)$, kjer je $q = p^n$, p praštevilo, $n \in \mathbb{N}$ in E poljubna eliptična krivulja, definirana nad končnim obsegom $\mathbb{F}_q$. Spomnimo se, da oznaka $E(\mathbb{F}_q)$ pomeni podgrupo racionalnih točk grupe na eliptični krivulji E. Moč podgrupe $E(\mathbb{F}_q)$ pogosto označujemo z $\#E(\mathbb{F}_q)$. Določitev moči grupe $E(\mathbb{F}_q)$ je za eliptično kriptografijo zelo pomembna. Večina praktičnih izvedb Schoofovega algoritma se omeji zgolj na krivulje nad obsegi velike karakteristike, ki so podane v kratki Weierstrassovi obliki, in na binarne nesupersingularne krivulje, podane v kanonični obliki za ta primer.

Schoofov algoritem je leta 1985 v reviji Mathematics of Computations prvi objavil René Schoof (beri: Škof), ki je trenutno profesor matematike na univerzi v Rimu. Algoritem je dramatična izboljšava v primerjavi s prej znanimi metodami. Shanksova metoda velikega in majhnega koraka, ki je, skupaj s svojimi manjšimi izboljšavami, pred letom 1985 bila najučinkovitejša znana metoda, potrebuje v najhitrejši izvedbi za izračun moči grupe $E(\mathbb{F}_q)$ reda $O(q^{\frac{1}{4}+\epsilon})$ bitnih XOR operacij, kjer lahko $\epsilon > 0$ izberemo poljubno majhen. Več o tem v nadaljevanju.

Časovna zahtevnost Schoofovega algoritma je $O(\log^8 q)$ bitnih operacij. Schoofu je torej uspelo zmanjšati časovno zahtevnost iz eksponentne zahtevnosti na polinomsko zahtevnost v $\log q$. Ker število $\log_2 q$ določa velikost, ki je potrebna za predstavitev posameznega elementa obsega $\mathbb{F}_q$ v računalniku, je ta izboljšava bistvena. Pred iznajdbo

Schoofovega algoritma je bilo za tiste eliptične krivulje, katerih parameter $\log q$ je bil dovolj velik, da so bile kriptografsko varne, zelo težko ali nemogoče določiti moč grupe na eliptični krivulji. To je bila resna ovira za dejansko praktično uporabo eliptične kriptografije. Schoofov algoritem je to oviro odpravil in tako bistveno pripomogel k hitremu razmahu eliptične kriptografije. Opozorimo pa, da se za resne namene danes ne uporablja več osnovna verzija Schoofovega algoritma, kot jo bomo opisali v nadaljevanju. Ta je namreč pogosto še vedno prepočasna za moderne kriptografske potrebe. Zato namesto nje uporabljamo izboljšane Schoofove algoritme, ki dosežejo v praksi boljše rezultate. Kriptografska znanost je namreč od prve objave Schoofovega algoritma precej napredovala.

V prvem razdelku bomo razložili pomen določanja moči grupe na eliptični krivulji za kriptografijo. V drugem razdelku bomo spoznali Shanksovo metodo velikega in majhnega koraka. V tretjem razdelku bomo podali osnoven opis Schoofovega algoritma. Sledi razdelek, v katerem navedemo podrobno shemo Schoofovega algoritma. S pomočjo te sheme je možno v praksi tudi brez razumevanja podrobnosti izvesti Schoofov algoritem. V naslednjih dveh razdelkih bomo podrobno utemeljili vse korake Schoofovega algoritma. Sledita razdelka, ki sta posvečena oceni časovne in prostorske zahtevnosti Schoofovega algoritma. O izboljšavah Schoofovega algoritma bomo govorili v zadnjem razdelku tega poglavja.

4.1 Pomen določanja moči grupe racionalnih točk za kriptografijo

V tem razdelku bomo razložili, zakaj se je v kriptografiji sploh pojavila potreba po učinkovitem algoritmu za določanje moči grupe racionalnih točk na eliptični krivulji.

Eliptične krivulje nad končnimi obsegi so temelj *eliptičnega kriptosistema*. Beseda *kriptosistem* v kriptografiji označuje vsako shemo, model, oziroma protokol, ki omogoča varno izmenjavo zaupnih podatkov preko nekega javnega komunikacijskega sistema. Eliptični kriptosistem je član širše družine kriptosistemov, ki jih imenujemo *ElGamalovi kriptosistemi*. To so kriptosistemi, katerih kriptografska varnost temelji na težavnosti izračunavanja diskretnega logaritma v končnih cikličnih grupah. Za različne končne ciklične grupe dobimo različne predstavnike ElGamalovih kriptosistemov. Več o ElGamalovih kriptosistemih in problemu diskretnega logaritma lahko bralec najde v [Stinson, poglavje 5].

Opišimo zdaj, kako pridemo do eliptičnega kriptosistema. Zanj najprej potrebujemo eliptično krivuljo E, definirano nad nekim končnim obsegom $\mathbb{F}_q$. Koordinate točk na taki eliptični krivulji ležijo v razpadnem obsegu $\overline{\mathbb{F}_q}$. Elemente iz $\overline{\mathbb{F}_q}$ je v računalniku zelo težko predstaviti, zato se omejimo zgolj na grupo racionalnih točk $E(\mathbb{F}_q)$, to je točk krivulje E, ki imajo obe koordinati v $\mathbb{F}_q$. Ta grupa je sicer vedno končna, vendar v splošnem ni ciklična. Za vsak ElGamalov kriptosistem je cikličnost nujno potrebna. Za kriptografsko uporabo moramo zato v dani grupi $E(\mathbb{F}_q)$ poiskati neko veliko ciklično podgrupo H. Po izboru podgrupe H potekajo vsi algoritmi eliptičnega kriptosistema zgolj le še v H. S preostankom grupe $E(\mathbb{F}_q)$ (kaj šele s celotnim E) pa ne delamo več.

Pri vzpostavitvi eliptičnega kriptosistema običajno najprej na nek način določimo moč grupe $E(\mathbb{F}_q)$. Če ima ta moč neko določeno lastnost, ki jo bomo navedli v nadaljevanju, je teoretično zagotovljen obstoj vsaj ene uporabne ciklične podgrupe H grupe $E(\mathbb{F}_q)$.

Še več, v tem primeru lahko vsaj eno uporabno podgrupo H tudi dejansko najdemo, to je, najdemo njen generator v $E(\mathbb{F}_q)$ in določimo njeno moč. Vse to bomo opisali v nadaljevanju. Moč podgrupe H je s stališča kriptografske varnosti nujno potrebno poznati. Problem diskretnega logaritma v končni ciklični grupi znane moči je namreč občutljiv na *Pohlig-Hellmanov napad* [Blake, str. 80]. Ta napad je učinkovit le, če je moč ciklične grupe produkt majhnih med seboj paroma tujih faktorjev. Če ne poznamo moči grupe H, ne vemo, ali ta moč ima tako obliko ali ne. Če nasprotniku uspe določiti moč grupe H in če ta moč slučajno je take oblike, bo nasprotnik v grupi H lahko rešil problem diskretnega logaritma. Zato je za varnost eliptičnega kriptosistema nujno, da poznamo moč končne ciklične grupe H in lahko izločimo tiste grupe H, ki so občutljive na Pohlig-Hellmanov napad. Pripomnimo, da moči eliptične grupe $E(\mathbb{F}_q)$ po določitvi dobre podgrupe H in izračunu moči podgrupe H ne potrebujemo več.

Za izbiro eliptične krivulje E in določitev njene moči se v praksi uporabljajo trije pristopi [Blake, str. 101]:

1. Krivuljo E izberemo naključno in s pomočjo Schoofovega algoritma in njegovih izboljšav določimo $\#E(\mathbb{F}_q)$.

2. Krivuljo E izberemo s pomočjo teorije kompleksne multiplikacije. Za tako izbrane krivulje je določitev $\#E(\mathbb{F}_q)$ enostavna. Več o tem je možno najti v [Blake, poglavje 8].

3. Izberemo končen obseg $\mathbb{F}_q$, $q = p^n$, kjer je q majhen, in namesto grupe $E(\mathbb{F}_q)$ uporabimo (precej večjo) grupo $E(\mathbb{F}_{q^m})$, kjer je m neko naravno število. Za majhne q je $\#E(\mathbb{F}_q)$ enostavno določiti. Iz Weilovega izreka [Enge, str. 101] lahko potem enostavno določimo $E(\mathbb{F}_{q^m})$. Take krivulje imenujemo **Koblitzeve** [Blake, str. 101]. Če pogledamo na Koblitzevo krivuljo s perspektive obsega $\mathbb{F}_{q^m}$, je to pravzaprav eliptična krivulja E nad $\mathbb{F}_{q^m}$, ki ima vse koeficiente že v podobsegu $\mathbb{F}_q$. Ta pristop v praksi najpogosteje srečamo pri binarnih krivuljah. V tem primeru je $p = 2$, majhnost q pa je ekvivalentna majhnosti eksponenta n.

Metoda 1. ima očitno prednost pred ostalima dvema metodama, da je izbrana krivulja naključna in zato zelo splošna. Pri metodah 2. in 3. namreč krivulja pripada nekemu podrazredu eliptičnih krivulj. To pa je že neka informacija, zaradi katere je morda problem diskretnega logaritma enostavnejši. Vseeno pa v tem trenutku za nobenega od primerov 2. in 3. ni znano, da bi bi bil problem diskretnega logaritma bistveno lažji kot v splošnem.

Recimo, da smo ubrali enega od zgornjih pristopov 1., 2. ali 3. Naslednji algoritem potem poišče tako krivuljo E, da bo v grupi $E(\mathbb{F}_q)$ možno najti ciklično podgrupo H, katere moč je neko ogromno praštevilo. Krivulja $E(\mathbb{F}_q)$, ki jo bomo poiskali, bo namreč imela moč oblike $s \cdot r$, kjer je r veliko praštevilo in s neko majhno število. Eliptično krivuljo, ki je kriptografsko uporabna, torej poiščemo takole:

1. Z eno od zgornjih metod izberemo eliptično krivuljo E in določimo moč podgrupe $E(\mathbb{F}_q)$. V primeru Koblitzevih krivulj namesto z $E(\mathbb{F}_q)$ operiramo s podgrupo $E(\mathbb{F}_{q^n})$, česar v nadaljevanju ne bomo več posebej izpostavljali.

2. Preverimo odpornost grupe na eliptični krivulji proti vsem znanim učinkovitim napadom. Med drugim izločimo supersingularne krivulje. Če je krivulja občutljiva na katerikoli napad, se vrnemo v korak 1.

3. Poskusimo faktorizirati $\#E(\mathbb{F}_q)$. Če nam ne uspe, se vrnemo v korak 1. V nadaljevanju namreč faktorizacijo nujno potrebujemo.

4. Če je moč grupe $E(\mathbb{F}_q)$ oblike $s \cdot r$, kjer je s majhno število in r praštevilo, smo našli iskano krivuljo. V nasprotnem primeru se vrnemo v korak 1.

Opozorimo na naslednjo zanimivo situacijo. Krivulje, ki danes veljajo za kriptografsko varne, imajo sicer zelo veliko moč, tipično reda velikosti 2^{150} ali še več. Vseeno pa je ta moč dovolj majhna, da njena faktorizacija ni problem. Če bi ta moč v praksi namreč bila tako velika, da bi imeli resne težave s faktorizacijo, eliptični kriptosistem sploh ne bi bil praktično izvedljiv.

Po Cauchyjevem izreku iz teorije grup zdaj vemo, da v $E(\mathbb{F}_q)$ obstaja vsaj ena podgrupa moči r. Podgrup moči r je lahko več. Ena od takih podgrup bo naša iskana podgrupa $H \leq E(\mathbb{F}_q)$. Konstruirati podgrupo H pravzaprav pomeni poiskati neko točko P na krivulji $E(\mathbb{F}_q)$, ki ima red enak r. Opomnimo, da pri običajnih algoritmih za iskanje podgrupe H (torej točke P) ne moremo neposredno vplivati na to, katero od podgrup moči r bo naš algoritem izbral, vendar s tem varnost eliptičnega kriptosistema ni ogrožena. Podgrupo H lahko konstruiramo na več načinov. Opisal bom naslednjo konstrukcijo.

Iz algebre je dobro znano, da lahko vsako končno Abelovo grupo moči $s \cdot r$, kjer je r praštevilo, ki ne nastopa v razcepu naravnega števila s, zapišemo kot direktno vsoto neke Abelove grupe moči s in neke Abelove grupe moči r. Ker je r veliko praštevilo, s pa majhno število, je torej

$$E(\mathbb{F}_q) = M \oplus \mathbb{Z}_r,$$

kjer je M neka končna Abelova grupa moči s in kjer $\mathbb{Z}_r$ označuje ciklično grupo reda r. Zato je v grupi $E(\mathbb{F}_q)$ natanko s elementov, katerih red je manjši od r. Izberimo naključno točko $P \in E(\mathbb{F}_q)$. Verjetnost, da je $sP = \mathcal{O}$, je $s/(sr) = 1/r \ll 1$. Če ta možnost nastopi, izberemo drugo točko P in to po potrebi ponavljamo, dokler $sP \neq \mathcal{O}$. Privzemimo torej, da smo našli točko $P \in E(\mathbb{F}_q)$, za katero je $sP \neq \mathcal{O}$. Tedaj P generira neko ciklično podgrupo H, ki ima moč vsaj r. Red elementa rP je neko število iz $\{1, 2, \ldots, s\}$. Ker velja $r \mid \operatorname{red} P$, $\operatorname{red}(rP) \mid \operatorname{red} P$ in $\Big(r \operatorname{red}(rP)\Big)P = \mathcal{O}$, velja

$$\operatorname{red}(P) = r \cdot \operatorname{red}(rP) \qquad \text{in zato} \qquad \operatorname{red}\Big(\operatorname{red}(rP)P\Big) = r.$$

Torej element $\operatorname{red}(rP)P$ generira ciklično podgrupo praštevilske moči r. To je iskana podgrupa H. Pripomnimo, da je njen generator $\operatorname{red}(rP)P$ enostavno poiskati, saj je enostavno določiti $\operatorname{red}(rP) \leq s$.

4.2 Kaj zmoremo brez Schoofovega algoritma

V tem razdelku bomo opisali metode za določanje moči grupe racionalnih točk na eliptični krivulji, ki so bile na voljo pred iznajdbo Schoofovega algoritma. Vse te metode imajo eksponentno časovno zahtevnost glede na parameter $\log q$, kar bistveno omejuje njihov domet. Za potrebe moderne kriptografije so vse te metode prepočasne. Vseeno pa so lahko koristne za testiranje pravilnosti delovanja hitrih algoritmov na majhnih krivuljah.

Poleg tega se Shanksova metoda, ki jo bomo tudi opisali v tem razdelku, pogosto uporablja v v kombinaciji s Schoofovim algoritmom in njegovimi izpeljankami.

4.2.1 Eksplicitna formula za $\#E(\mathbb{F}_p)$

V primeru, ko je $q = p$, kjer je p veliko praštevilo in je krivulja E podana v kratki Weierstrassovi obliki s parametroma a in b, obstaja eksplicitna formula za $\#E(\mathbb{F}_p)$. Iz definicije Jacobijevega simbola sledi, da je za vsak $x \in \mathbb{F}_p$ na $E(\mathbb{F}_p)$ natanko

$$1 + \left(\frac{x^3 + ax + b}{p}\right)$$

različnih točk, katerih X-koordinata je enaka x. Ker moramo šteti še točko v neskončnosti, je potem

$$\#E(\mathbb{F}_p) = 1 + \sum_{x \in \mathbb{F}_p} \left(1 + \left(\frac{x^3 + ax + b}{p}\right)\right) = p + 1 + \sum_{x \in \mathbb{F}_p} \left(\frac{x^3 + ax + b}{p}\right).$$

Za izračun te vsote potrebujemo $O(p \log p)$ množenj v aritmetiki velikih celih števil, kar je eksponentna časovna zahtevnost glede na parameter $\log q = \log p$. Ta metoda je neuporabna za velike krivulje. Ker pa je precej preprosta, jo vseeno pogosto uporabljamo za majhne krivulje, kjer je $p < 10000$ [Blake, str. 102].

4.2.2 Shanksova metoda majhnih in velikih korakov

V tem podrazdelku bomo predstavili klasično metodo za določanje moči grupe racionalnih točk in pokazali, kako jo lahko izboljšamo s Shanksovo metodo velikih in majhnih korakov. Spomnimo se, da Hassejev izrek (2.17) pravi, da je $t^2 \leq 4q$, kjer je $\#E(\mathbb{F}_q) = q + 1 - t$. Ta ocena precej zoži območje, v katerem se lahko giblje moč grupe racionalnih točk.

Definicija 4.1 *Interval $(q + 1 - 2\sqrt{q}, q + 1 + 2\sqrt{q})$, na katerega ta ocena ocena omeji število $\#E(\mathbb{F}_q)$, imenujemo* **Hassejevo območje**. *Označimo ga s $\mathcal{H}$.*

Vidimo, da je v primerjavi s q velikost Hassejevega območja majhna.

Opišimo enega izmed načinov za določitev števila $\#E(\mathbb{F}_q)$. Izberemo naključno točko $P \in E(\mathbb{F}_q)$. Za izbor naključne točke obstaja učinkovit algoritem [Blake, str. 35]. Iz teorije grup vemo, da je $(\#E(\mathbb{F}_q))P = \mathcal{O}$. Zato obstaja vsaj en $m \in \mathcal{H}$, tako da je $mP = \mathcal{O}$. Če nam hkrati uspe pokazati, da za izbrani P drugih takih števil m v Hassejevem območju ni, mora veljati $m = \#E(\mathbb{F}_q)$. Če pa obstaja več takih števil, je red točke P enak najmanjši absolutni vrednosti razlike dveh takih števil. V tem primeru lahko zaključimo le, da število $\operatorname{red} P$ deli $\#E(\mathbb{F}_q)$ in poskusimo z drugim P. Iz informacij, ki jih na ta način dobimo, poskusimo enolično določiti $\#E(\mathbb{F}_q)$.

Definicija 4.2 *Zgoraj opisano metodo imenujemo* **klasična metoda**.

Pred odkritjem Schoofovega algoritma so vse pomembnejše metode za določanje $\#E(\mathbb{F}_q)$ temeljile na tej metodi. Uspešnost klasične metode je seveda močno odvisna od izbora točk P. Večji kot je red točke P, bolje je. Če želimo namreč, da bo najdeno število m edino v Hassejevem območju, mora biti red P večji od velikosti Hassejevega območja, to je $4\sqrt{q}$. V trenutku izbora naključne točke P seveda reda točke P v naprej ne poznamo in ga tudi nima smisla računati, ker bi za to porabili več časa, kot za operacije, ki jih s točko P nameravamo narediti. Kljub temu, da obstajajo zadostni pogoji, ki nam vsaj v nekaterih primerih teoretično zagotovijo obstoj ugodnih točk P, je v praksi take točke težko poiskati.

Naslednji, še hujši problem klasične metode pa je, da je Hassejevo območje, čeprav se zdi relativno ozko, še vedno zelo veliko in je izračunavanje mP za vse $m \in \mathcal{H}$ računsko zahtevno. Tak izračun poteka tako, da najprej določimo $(q+1)P$ s pomočjo algoritma kvadriraj in množi. Ta izračun je v primerjavi z nadaljevanjem hiter. Nadaljujemo tako, da po vrsti določamo pare

$$\Big\{\Big((q+1)+s\Big)P, \Big((q+1)-s\Big)P\Big\} \qquad \text{za } s = 0, 1, 2, \ldots, \lfloor 2\sqrt{q} \rfloor.$$

Pri tem naslednji par dobimo iz prejšnjega tako, da prvemu elementu prištejemo P, drugemu pa odštejemo P. Za izvedbo celotnega postopka za posamezen P potrebujemo $O(\sqrt{q})$ grupnih operacij, kar je veliko, saj gre za eksponentno časovno zahtevnost glede na parameter $\log q$.

Ta problem poskuša rešiti **Shanksova metoda**, ki popularno nosi ime **metoda velikih in majhnih korakov**. Shanksova metoda zniža število potrebnih grupnih operacij za obdelavo posamezne točke P z $O(\sqrt{q})$ na $O(\sqrt[4]{q})$. Vendar hkrati poveča prostorsko zahtevnost z $O(1)$ na $O(\sqrt[4]{q})$ elementov grupe na eliptični krivulji. Je torej neke vrste kompromis med časovno in prostorsko zahtevnostjo. V nadaljevanju bomo opisali Shanksovo metodo in ocenili njeno časovno in prostorsko zahtevnost.

Naj bo $P \in E(\mathbb{F}_q)$. Naša osnovna naloga, ki jo želimo s Shanksovim algoritmom razrešiti, je natanko določiti množico

$$\mathcal{R} = \Big\{ m \in \mathcal{H} \,\Big|\, mP = \mathcal{O} \Big\}. \tag{4.1}$$

Opozorimo, da se Shanksov pristop s problemom, kako izbrati točko P, ki ima velik red, ne ukvarja.

Najprej naredimo nekaj premislekov, ki so nujni za Shanksov algoritem. Naj bosta D in V neki poljubni naravni števili. Če preteče i množico $\{-V, \ldots, 0, \ldots, V\}$ in hkrati j preteče množico $\{-D, \ldots, 0, \ldots, D\}$, potem izraz $i \cdot (2D+1) + j$ preteče natanko cela števila

$$\mathcal{A}_{D,V} = \Big\{ -V \cdot (2D+1) - D, \ldots, 0, \ldots, V \cdot (2D+1) + D \Big\}.$$

To so pravzaprav cela števila, katerih zapis v $(2D+1)$-tem številskem sistemu ima dve števki (vodilna je morda 0), od katerih je vodilna po absolutni vrednosti manjša ali enaka V. Edina razlika je, da števka j ne teče od 0 do $2D$, ampak od $-D$ do D. Razlog za uvedbo množice $\mathcal{A}_{D,V}$ je, da želimo naravna števila, ki so vsebovana v Hassejevem območju, čim bolj učinkovito parametrizirati. To bomo storili tako, da jih bomo zapisali kot vsoto $q + 1 + r,\ r \in \mathbb{Z},\ r \in (-2\sqrt{q}, 2\sqrt{q})$ in potem r parametrizirali z i in j kot v definiciji množice $\mathcal{A}_{D,V}$. V ta namen običajno postavimo $D = \lfloor \sqrt[4]{q} \rfloor$ in poiščimo najmanjše število $V \in \mathbb{N}$, da bodo cela števila iz intervala $(-2\sqrt{q}, 2\sqrt{q})$ vsebovana v $\mathcal{A}_{D,V}$. Veljati mora $V(2D+1) + D \geq 2\sqrt{q}$, od koder dobimo, da je najmanjši V enak

$$V = \left\lceil \frac{2\sqrt{q}}{2D+1} - \frac{D}{2D+1} \right\rceil = \left\lceil \frac{2\sqrt{q}}{2\lfloor \sqrt[4]{q} \rfloor + 1} - \frac{\lfloor \sqrt[4]{q} \rfloor}{2\lfloor \sqrt[4]{q} \rfloor + 1} \right\rceil.$$

Reševanje osnovne naloge (4.1) se zdaj prevede na iskanje rešitev enačbe

$$\Big(q + 1 + i(2D+1) + j\Big)P = \mathcal{O}, \tag{4.2}$$

kjer sta i in j neznanki z lastnostma $-D \leq j \leq D$ in $-V \leq i \leq V$, števila q, D, V in točka P pa so parametri.

Če smo čisto natančni, potem opazimo, da za $i = V$ pri največjih j število $i(2D+1)+j$ lahko postane večje od $2\sqrt{q}$. Enako velja za $i = -V$ in najmanjše (negativne) j. To se zgodi, ker smo v definiciji M morali optimalno število

$$\frac{2\sqrt{q}}{2D+1} - \frac{D}{2D+1}$$

povečati s funkcijo $\lceil x \rceil$ na prvo večje naravno število, saj mora biti $V \in \mathbb{N}$. Vendar nas ta pojav ne bo motil, le v algoritmu moramo pri $i = \pm V$ prevelika oziroma premajhna števila j izločiti.

Dejanski izračuni potekajo takole. Najprej inicializiramo $S = \emptyset$, kjer je S množica, v katero bomo shranjevali dobljene rezultate. Izračunamo $Q = (2D+1)P$ in $R = (q+1)P$. Enačbo (4.2) prepišemo v obliko

$$R + iQ = -jP \tag{4.3}$$

Zdaj sta i in j vsak na svoji strani enačbe, torej sta ločena. Zato lahko posebej, neodvisno od i, izračunamo desne strani enačbe (4.3) za $j = -D, \ldots, D$ in rezultate shranimo v neko tabelo $\mathcal{T}$. Tem izračunom rečemo računanje majhnih korakov. Število potrebnih majhnih korakov je enako D (oznaka D pride iz dete). Tabela $\mathcal{T}$ sestoji torej iz $2D+1$ urejenih parov $(-jP, j) \in E(\mathbb{F}_q) \times \mathbb{Z}$. Zatem po vrsti izračunamo še leve strani enačbe (4.3) za $i = -V, \ldots, V$. Temu postopku pa rečemo računanje velikih korakov (oznaka V pride iz velikan). Pri vsakem i preverimo, če je leva stran enačbe (4.3) za ta i morda enaka prvi koordinati kakega urejenega para (P, j) iz $\mathcal{T}$. Če ni, preverimo naslednji i. Če pa je, smo tako našli števili i in j, ki rešita enačbo (4.3). Ustrezna rešitev m originalnega problema, ki ustreza i in j, je enaka $m = q + 1 + i(2D+1) + j$. Rešitev m shranimo v $\mathcal{S}$ in preverimo naslednji i. Na koncu postopka je množica $\mathcal{S}$ enaka iskani množici rešitev $\mathcal{R}$ osnovne naloge (4.1). Ker vemo, da $\mathcal{S}$ ni prazna, smo morali v zgornjem algoritmu vsaj za en i naleteti na trčenje z nekim j.

S tem smo opisali Shanksov algoritem. Opozorimo, da v kriptografiji obstaja več Shanksovih algoritmov, ki imajo različne namene. Obstaja na primer tudi Shanksov algoritem za diskretni logaritem. Vsi pa imajo skupno idejo, da ločeno izračunamo majhne korake, jih shranimo v tabelo, izračunamo še velike korake in potem iščemo trčenje velikih korakov z majhnimi koraki ali pa obratno. Če brez Shanksovega algoritma za nek problem potrebujemo v grobem n izračunov, potem s Shanksovim algoritmom potrebujemo v grobem $\sqrt{n}$ izračunov. Vendar moramo pri tem shraniti reda velikosti $\sqrt{n}$ podatkov. Pri izvedbi Shanksovega algoritma je pomembno, da je tabela $\mathcal{T}$ dobro organizirana, da lahko po njej čim hitreje iščemo. Dobra rešitev je na primer, da $\mathcal{T}$ predstavimo kot binarno drevo.

Vidimo tudi, da moramo zaradi iskanja po tabeli majhnih korakov v obseg $\mathbb{F}_q$ na nek način vpeljati relacijo urejenosti. Med končnimi obsegi imajo naravno urejenost le praobsegi $\mathbb{F}_p$, kjer je p praštevilo. Ostanke po modulu p namreč uredimo kar s standardno urejenostno relacijo celih števil. Obseg $\mathbb{F}_q$, kjer je $q = p^n,\ n \geq 2$, lahko s pomočjo dane polinomske baze uredimo tako, da polinome leksikografsko uredimo glede na koeficiente, ki so iz praobsega $\mathbb{F}_p$. Za različne polinomske baze dobimo različne ureditve obsega $\mathbb{F}_q$. Nobena med njimi ni posebej odlikovana.

Ocenimo časovno zahtevnost Shanksovega algoritma. Najprej pripomnimo, da vsaka posamezna osnovna operacija (seštevanje ali podvojevanje točke) v grupi E zahteva kvečjemu neko konstantno število operacij v $\mathbb{F}_q$. Torej je aritmetika v E asimptotsko enako zahtevna kot aritmetika v $\mathbb{F}_q$. Merska enota za našo analizo bo zato kar potrebno število operacij v $\mathbb{F}_q$. Za izračun majhnih korakov potrebujemo $O(\sqrt[4]{q})$ operacij v $\mathbb{F}_q$. Zraven moramo graditi še binarno drevo, kar za vsakega od $\sqrt[4]{q}$ korakov pomeni $\log D = (1/4)\log q$ primerjanj elementov iz $\mathbb{F}_q$. Torej skupno potrebujemo $O(\sqrt[4]{q}\log q)$ primerjanj elementov v $\mathbb{F}_q$. Pri tem moramo shraniti $O(\sqrt[4]{q})$ elementov obsega $\mathbb{F}_q$ in indeksov j. Pri izračunu velikih korakov je podobno. Spet potrebujemo $O(\sqrt[4]{q})$ operacij v $\mathbb{F}_q$ za izračun samih korakov. Poleg tega za iskanje po tabeli $\mathcal{T}$ potrebujemo še $O(\sqrt[4]{q}\log q)$ primerjanj elementov v $\mathbb{F}_q$. Za celoten algoritem torej potrebujemo $O(\sqrt[4]{q})$ operacij v $\mathbb{F}_q$ in $O(\sqrt[4]{q}\log q)$ primerjanj elementov iz $\mathbb{F}_q$. Pri tem med algoritmom potrebujemo prostor, da naenkrat shranimo $O(\sqrt[4]{q})$ elementov iz $\mathbb{F}_q$ in ravno toliko velikih naravnih števil j. Dobljena časovna zahtevnost je žal še vedno eksponentna v parametru $\log q$. Zato Shanksova metoda sama po sebi ni kos modernim zahtevam kriptografske industrije. V kombinaciji s Schoofovim algoritmom in drugimi izboljšavami pa je neprecenljive vrednosti. Pripomnimo, da se problemu s prostorom lahko delno izognemo z uporabo t.i. *Pollard ρ-metode*, ki je v tem delu ne bomo podrobneje obravnavali.

Na koncu navedimo še tehničen komentar. Opisani Shanksov algoritem v praksi še nekoliko izboljšamo. Ko računamo majhne korake, v tabelo $\mathcal{T}$ ne shranjujemo točk $-jP$, ampak le njihove X-koordinate. Parameter j teče samo od 0 do D, saj sta X-koordinati točk $-jP$ in jP enaki. Točko $-jP$ vedno izračunamo preko formule $-jP = -(j-1)P - P$ in ne direktno, kar bi bilo bistveno počasneje. Takoj, ko določimo $-jP$, torej na Y-koordinato točke $-(j-1)P$ pozabimo. Velike korake računamo tako, da iz $R + (i-1)Q$ in $R - (i-1)Q$ s prištevanjem oziroma odštevanjem točke Q določimo $R + iQ$ in $R - iQ$. Pri tem za vsako od teh dveh točk posebej preverimo, če se X-koordinata ujema s prvo koordinato kakšnega elementa iz $\mathcal{T}$. Če se velik korak $R + iQ$ ujema s parom (x, j'), moramo določiti še pravilen predznak za iskani j. To storimo tako, da direktno preverimo, katera od enakosti $R + iQ = -j'P,\ R - iQ = j'P$ drži. Če drži prva, je $j = j'$, sicer pa je $j = -j'$. Če pride do trčenja pri $R - iQ$, poteka postopek analogno. Na asimptotski čas izvajanja in velikost potrebnega skladišča podatkov pravkar opisane tehnične izboljšave ne vplivajo.

4.3 Osnoven opis Schoofovega algoritma

V tem razdelku bomo opisali osnovne korake, ki sestavljajo Schoofov algoritem. Podrobnosti bomo izdelali v naslednjih razdelkih.

Cilj Schoofovega algoritma je določitev moči grupe $E(\mathbb{F}_q)$, kjer je $q = p^n$, $n \in \mathbb{N}$, p praštevilo in E poljubna eliptična krivulja nad končnim obsegom $\mathbb{F}_q$. Namesto direktne določitve števila $\#E(\mathbb{F}_q)$ Schoofov algoritem določi sled Frobeniusa t. Ta ustreza zvezi $\#E(\mathbb{F}_q) = q + 1 - t$, zato lahko iz znanega števila t neposredno določimo $\#E(\mathbb{F}_q)$. Kot smo navedli v začetku podrazdelka 4.2.2 o Shanksovi metodi, Hassejev izrek pove, da je $\#E(\mathbb{F}_q) \in \mathcal{H} = (q + 1 - 2\sqrt{q}, q + 1 + 2\sqrt{q})$. To pomeni, da je $t \in (-2\sqrt{q}, 2\sqrt{q})$. Velikost tega intervala pa je enaka $4\sqrt{q}$. Naj bo $\mathcal{L}$ neka tako velika množica dovolj velikih praštevil, da velja

$$\prod_{\ell \in \mathcal{L}} \ell > 4\sqrt{q}. \tag{4.4}$$

Denimo, da smo uspeli določiti količine $t_\ell = t \bmod \ell$ za vsako praštevilo $\ell \in \mathcal{L}$. Potem lahko s pomočjo kitajskega izreka o ostankih [Stinson, str. 119-122] iz teh podatkov enolično določimo t. S tem smo podali osnoven opis Schoofovega algoritma, ki torej poteka takole:

1. Najprej določimo primerno množico praštevil $\mathcal{L}$.
2. Zatem za vsako praštevilo iz $\mathcal{L}$ določimo število t_ℓ.
3. S pomočjo kitajskega izreka o ostankih določimo t.

Prva točka je preprosta. V $\mathcal{L}$ želimo imeti taka praštevila ℓ, da bomo v drugi točki znali določiti t_ℓ in to čim hitreje. Kako izberemo ta praštevila, bomo povedali v razdelku 4.5.

Druga točka Schoofovega algoritma je računsko bistveno zahtevnejša od ostalih dveh in predstavlja jedro Schoofovega algoritma. Natančneje jo bomo predstavili v razdelku 4.6, tukaj pa omenimo samo bistvene korake. Za $\ell \in \mathcal{L}$ označimo $q_\ell = q \bmod \ell$. Spomnimo se, da smo v razdelku 3.3 o Frobeniusovem endomorfizmu pokazali, da za ℓ-ti Frobeniusov operator φ_ℓ za vsak $P \in E[\ell]^*$ velja

$$\varphi_\ell^2(P) - t_\ell \varphi_\ell(P) + q_\ell P = \mathcal{O}.$$

Iskano število t_ℓ je ena od vrednosti iz $\{0, 1, 2, \ldots, \ell - 1\}$. Poiščemo ga tako, da s spremenljivko τ pretečemo $0, 1, 2, \ldots, \ell - 1$ in vsakič testiramo, če je za vsak $P \in E[\ell]^*$ izpolnjena enačba

$$\varphi_\ell^2(P) + q_\ell P = \tau \varphi_\ell(P). \tag{4.5}$$

Ta test bo izpolnilo natanko eno število $\tau_0 \in \{0, 1, 2, \ldots, \ell - 1\}$. Potem je iskano število t_ℓ enako τ_0. Pripomnimo, da teh testov ne izvajamo tako, da generiramo netrivialne točke iz $E[\ell]$ in direktno preverimo, za kateri τ velja zgornja enakost (4.5). Točke iz $E[\ell]$ v splošnem namreč ne ležijo v $E(\mathbb{F}_q)$, ampak "globoko v E". Za določitev števila t_l bi sicer zadostovalo že, da poznamo eno samo netrivialno točko iz $E[\ell]$, ki bi jo vstavili v (4.5) in enolično določili število t_l. Vendar je v praksi težko poiskati že eno samo netrivialno točko iz $E[\ell]$ (kaj šele poiskati vse točke iz $E[\ell]$). Schoofov algoritem se problemu iskanja netrivialnih točk iz $E[\ell]^*$ elegantno izogne z uporabo največjega skupnega delitelja polinomov, kot bomo pokazali v naslednjih razdelkih.

Tretja točka algoritma je standarden postopek reševanja sistema linearnih kongruenc, ki ga omogoča kitajski izrek o ostankih. Gre za dobro znan in v kriptografiji in še kje pogosto uporabljan postopek, tako da ga v tem delu ne bomo eksplicitno predstavljali. Najti ga je mogoče v [Stinson, str. 119].

4.4 Shema Schoofovega algoritma

Zaradi boljšega pregleda najprej navedimo podrobno shemo Schoofovega algoritma. Vse korake bomo v naslednjih razdelkih utemeljili.

---

1. S pomočjo postopka iz razdelka 4.5 določimo množico $\mathcal{L}$.
   Označimo $\ell_{\max} = \max \mathcal{L}$. Pri tem oznaka $\max \mathcal{L}$ pomeni običajem maksimum končne množice $\mathcal{L} \subset \mathbb{R}$.
2. Izračunamo $q_\ell = q \bmod \ell$.
3. Izračunamo in shranimo delitvene polinome $\psi_0, \psi_1, \ldots, \psi_{\ell_{\max}}$. Opomnimo, da potrebujemo vse polinome z indeksi od 0 do $\ell_{\max}$ in ne le polinomov $\{\psi_\ell \mid \ell \in \mathcal{L}\}$.
4. V binarnem nesupersingularnem primeru je $t_2 = t \bmod 2 = 1$, kar v točki 6. uporabimo skupaj z rezultati točke 5.
5. Za vsa praštevila $\ell \in \mathcal{L}$ določimo število $t_\ell = t \bmod \ell$, tako da:
   5.1. Izračunamo $X^q, X^{q^2}, Y^q, Y^{q^2} \bmod E \bmod \psi_\ell$.
   5.2. S pomočjo pogoja iz trditve (4.11) ugotovimo, ali je ℓ tipa C ali ne.
   5.3. Če ℓ ni tipa C, poiščemo t_ℓ takole:
      5.3.1. Preverimo pogoj iz trditve (4.13). Če ni izpolnjen, torej če q_ℓ ni kvadraten ostanek v $\mathbb{F}_\ell$, je $t_\ell = 0$ (praštevilo ℓ pa je zagotovo tipa A).
      5.3.2. Sicer pa določimo $\omega = \sqrt{q_\ell}$ v $\mathbb{F}_\ell$ in preverimo še pogoja iz trditev (4.14) in (4.15), s čimer natanko določimo t_ℓ (in tudi tip praštevila ℓ).
   5.4. Če je ℓ tipa C, poiščemo t_ℓ takole:
      5.4.1. Najprej s pomočjo rekurzivnih formul določimo $\psi_0^q \bmod \psi_\ell,\ \ldots\ , \psi_{\ell-1}^q \bmod \psi_\ell$.
      5.4.2. Za $\tau = 1, \ldots, \lceil (\ell-1)/2 \rceil$ naredimo naslednje:
         5.4.2.1. Izvedemo X-test za τ. Če test ni uspešen, povečamo τ za ena in ponovimo korak 5.4.2.1.
         5.4.2.2. Če je test iz koraka 5.4.2.1. pri nekem τ uspešen, pa s pomočjo Y-testa določimo, katera od možnosti $t_\ell = \pm\tau$ je prava.
         S tem je določitev t_ℓ končana, preostalih τ ni potrebno preverjati.
6. Iz izračunanih števil $\{t_\ell \mid \ell \in \mathcal{L}\}$ s pomočjo kitajskega izreka o ostankih enolično določimo t in s tem $\#E(\mathbb{F}_q)$.

4.5 Določitev množice $\mathcal{L}$

V tem razdelku bomo opisali, kako konstruiramo množico praštevil $\mathcal{L}$, ki smo jo uvedli v razdelku 4.3.

Praštevilo 2 vključimo v množico $\mathcal{L}$ le v binarnem nesupersingularnem primeru, kjer brez računanja vemo, da je $t_2 = 1$. V primeru velike karakteristike $p > 3$ pa izračun t_2 povzroča nekaj težav, zato v tem primeru števila 2 ne vključimo v $\mathcal{L}$.

Kar pa se tiče lihih praštevil, je v množico $\mathcal{L}$ z izjemo praštevila p (primer $p > 3$) najugodneje dati kar liha praštevila $3, 5, 7, 11, \ldots$ po vrsti, dokler množica $\mathcal{L}$ ni dovolj velika. V Schoofovem algoritmu namreč zahtevnost izračuna števila t_ℓ hitro raste z rastočim ℓ, zato želimo delati s čim manjšimi ℓ. Hkrati z dodajanjem praštevil v $\mathcal{L}$ torej sproti preverjamo pogoj (4.4) in se ustavimo takoj, ko je ta pogoj izpolnjen. V ta namen mora algoritem znati računati z naravnimi števili reda velikosti $O(q)$. Ta velikost presega običajne vgrajene tipe v standardne prevajalnike, zato je potrebno to mnogomestno številsko aritmetiko izdelati posebej. Mnogomestno številsko aritmetiko potrebujemo tudi v zadnji fazi Schoofovega algoritma pri izreku o kitajskih ostankih. V primeru, ko obravnavamo eliptično krivuljo nad obsegom velike karakteristike $p = q > 3$, moramo mnogomestno številsko aritmetiko tako ali tako imeti zaradi obsega samega.

Navedli smo, da je pri konstrukciji množice $\mathcal{L}$ v primeru $p > 3$ karakteristika p izjema in je nikoli ne vključimo v $\mathcal{L}$. Problem je namreč, da torzijska podgrupa $E[p]$ ni izomorfna $\mathbb{Z}_p \oplus \mathbb{Z}_p$ in ne moremo uporabiti enakega postopka kot za preostala praštevila. Za najzanimivejše in praktično uporabne eliptične krivulje v primeru $p > 3$ je $q = p$ veliko praštevilo. Tedaj postane $\mathcal{L}$ dovolj velika že davno preden v zaporedju lihih praštevil pridemo do p, tako da se ta problem sploh ne pojavi.

4.6 Podroben opis izračuna števil t_ℓ

V tem razdelku bomo natančno opisali in utemeljili centralni in najzahtevnejši del Schoofovega algoritma, to je določitev števil $t_\ell = t \bmod \ell$ za vse $\ell \in \mathcal{L}$. Na nekaterih mestih bomo srečali precej dolge in zapletene enačbe. Vse enačbe, ki jih bomo navedli, so bile preverjene s programom za simbolno računanje (Mathematica).

Naša naloga je poiskati število $\tau \in \{0, 1, \ldots, \ell - 1\}$, tako da za vsak $P \in E[\ell]$ velja

$$\varphi_\ell^2(P) - \tau\varphi_\ell(P) + q_\ell P = \mathcal{O}. \tag{4.6}$$

Pri tem veljajo oznake iz razdelka 4.3 o osnovah Schoofovega algoritma. Kot smo že omenili, v binarnem nesupersingularnem primeru vedno velja $t_2 = 1$. To sledi neposredno iz izreka (3.3).

Spoprimimo se zdaj še z netrivialnim primerom $\ell > 2$. Od tu dalje naj bo ℓ torej praštevilo, za katerega velja $\ell \neq 2$ in $\ell \neq p$. Spomnimo se oznake $E[\ell]^* = E[\ell] \setminus \{\mathcal{O}\}$. V razdelku 3.4 o delitvenih polinomih smo navedli, da za binarne eliptične krivulje običajno polinome ψ_n označimo s f_n. V tistih delih tega razdelka, kjer obravnavamo oba kanonična primera hkrati, bomo zaradi preglednosti uporabljali le oznako ψ_n.

Trditev 4.3 *Število τ iz enačbe (4.6) je pri fiksnih E, q in ℓ enolično.*

Dokaz: Recimo, da bi obstajali različni števili τ_1, τ_2, obe iz $\mathbb{F}_l$, ki zadostita enačbi (4.6). Potem oba primerka te enačbe odštejemo in dobimo

$$(\tau_1 - \tau_2)\varphi_\ell(P) = \mathcal{O} \quad \text{za vsak } P \in E[\ell].$$

Ker je $E[\ell] \cong \mathbb{Z}_\ell \oplus \mathbb{Z}_\ell$ in je ℓ praštevilo, ima vsaka točka iz $E[\ell]^*$ v grupi E red ℓ. Za poljubno točko $P \in E[\ell]^*$ je $\varphi_\ell(P)$ neničelna točka grupe $E[\ell]$. Zato mora biti $\tau_1 - \tau_2 \equiv 0 \pmod{\ell}$. To pa ni mogoče, ker sta τ_1 in τ_2 različni števili iz $\mathbb{F}_\ell$. ■

Iskanje števila τ, ki zadosti enačbi (4.6), precej olajša naslednja pomembna trditev.

Trditev 4.4 *Naj bo $\tau \in \mathbb{F}_\ell$. Če obstaja tak $P' \in E[\ell]^*$, da zanj velja*

$$\varphi_\ell^2(P') + q_\ell P' = \tau \varphi_\ell(P'), \tag{4.7}$$

potem ta enakost velja za vsak $P \in E[\ell]^*$ in torej velja $t_\ell = \tau$.

Dokaz: Vemo, da tudi za naš P' velja originalna karakteristična enačba

$$\varphi_\ell^2(P') + q_\ell P' = t_\ell \varphi_\ell(P').$$

Če to odštejemo od enakosti (4.7), dobimo $(\tau - t_\ell)\varphi_\ell(P') = \mathcal{O}$. Ker je $P' \neq 0$, je $\varphi_\ell(P') \neq 0$. Torej je $t_\ell - \tau \equiv 0 \pmod{\ell}$. Sledi $t_\ell = \tau$. ■

Potemtakem za določitev pravega τ v Schoofovem algoritmu enačbe (4.6) ni potrebno izpolniti za vsak $P \in E[\ell]^*$. Dovolj jo je izpolniti le za neko točko iz $E[\ell]^*$, saj v tem primeru enačba potem velja za vse točke iz $E[\ell]^*$.

Pripomnimo še, da iz $\ell \neq p$ sledi $\ell \nmid q$. Torej je $q_\ell = q \bmod \ell \neq 0$. Zato za vsako točko $P = (x, y) \in E[\ell]^*$ velja $q_\ell P \neq \mathcal{O}$ in lahko uporabimo naslednjo formulo iz razdelka 3.4 o delitvenih polinomih

$$q_\ell P = \left(\frac{\theta_{q_\ell}(x,y)}{\psi_{q_\ell}^2(x,y)}, \frac{\omega_{q_\ell}(x,y)}{\psi_{q_\ell}^3(x,y)} \right). \tag{4.8}$$

V nadaljevanju bomo to formulo večkrat uporabili. Naslednja definicija ima ključno vlogo v opisu Schoofovega algoritma.

Definicija 4.5 *Za vsako praštevilo ℓ definiramo* **tip praštevila** *ℓ.*

$$\text{Tip praštevila } \ell \ = \begin{cases} A, & \text{če obstaja } P \in E[\ell]^*, \text{tako da je } \varphi_\ell^2(P) + q_\ell P = \mathcal{O}, \\ B, & \text{če obstaja } P \in E[\ell]^*, \text{tako da je } \varphi_\ell^2(P) - q_\ell P = \mathcal{O}, \\ C, & \text{sicer.} \end{cases}$$

Schoofov algoritem mora ločeno obravnavati vse te tri primere. Pripomnimo, da tip praštevila ni globalna lastnost praštevila, ampak je odvisen od števila q in krivulje E. Iz trditve (4.4) sledi, da je možnost A ekvivalentna pogoju $t_\ell = 0$.

Očitno je možnost C nezdružljiva z A in B. Iz naslednje leme sledi, da so primeri A, B, C pri danih q in E paroma disjunktni in izčrpajo vse možnosti.

Lema 4.6 *Praštevilo ℓ pri danih q in E ne more biti hkrati tipa A in B.*

Dokaz: Recimo, da je praštevilo ℓ pri nekih q in E hkrati tipa A in B. Potem je $t_\ell = 0$, hkrati pa za nek $P \in E[\ell]^*$ velja $\varphi_\ell^2(P) - q_\ell(P) = \mathcal{O}$. Zaradi $t_\ell = 0$ iz karakteristične enačbe dobimo, da za P velja $\varphi_\ell^2(P) + q_\ell P = \mathcal{O}$. Iz obeh enakosti, ki veljata za točko P, sledi, da je $2q_\ell P = \mathcal{O}$ in zato tudi $2q_\ell \equiv 0 \pmod{\ell}$. Ker je $\ell > 2$, mora biti $q_\ell \equiv 0 \pmod{\ell}$, kar pa je protislovje. ■

Schoofov algoritem za vsako praštevilo $\ell \in \mathcal{L}$ preveri, ali je ℓ tipa C ali ne. Za tip C poteka izračun števila t_ℓ drugače kot v primeru, ko je ℓ tipa A ali B. Vsaki od teh dveh možnosti (C oziroma A, B) bomo v nadaljevanju posvetili poseben podrazdelek. Zdaj pa najprej pokažimo, kako algoritem ugotovi, katera od možnosti nastopi. Pri tem lahko obravnavamo oba kanonična primera eliptičnih krivulj hkrati.

Lema 4.7 *Naj bosta $P = (x, y)$ in $Q = (x', y')$ poljubni točki na E. Potem velja*

$$(P = Q) \text{ ali } (P = -Q) \iff x = x'.$$

Dokaz:
($\Rightarrow$) Sledi neposredno iz definicije grupne operacije na E.
($\Leftarrow$) Naj bo $x = x'$ in $y \neq y'$. Če v enačbo krivulje vstavimo $X = x$, dobimo kvadratno enačbo po spremenljivki Y, ki ima dve različni rešitvi y in y'. Po Viètovih formulah je potem $y + y' = -(a_1 x + a_3)$. Zdaj neposredno iz definicije grupnega zakona sledi, da je $(x, y) = -(x, y')$, oziroma $P = -Q$. ■

Trditev 4.8 *Praštevilo $\ell > 2$, kjer je $\ell \neq 2, p$, je tipa A ali B natanko takrat, ko obstaja $P = (x, y) \in E[\ell]^*$, tako da je*

$$\psi_{q_\ell}^2 \cdot (x^{q^2} - x) + \psi_{q_\ell - 1}\psi_{q_\ell + 1} = 0. \tag{4.9}$$

Dokaz: Po definiciji φ in lemi (4.7) je ℓ tipa A ali B natanko takrat, ko obstaja $(x, y) \in E[\ell]^*$, tako da je

$$X\Big((x^{q^2}, y^{q^2})\Big) = X\Big(q_\ell(x, y)\Big).$$

Spomnimo se, da oznaka $X(P)$ pomeni X-koordinato točke P. Iz formule (4.8) sledi, da je zgornja enačba ekvivalentna enačbi

$$x^{q^2} = \frac{\theta_{q_\ell}(x, y)}{\psi_{q_\ell}^2(x, y)}.$$

Če odpravimo ulomek in upoštevamo formulo za θ_{q_ℓ} iz razdelka 3.4 o delitvenih polinomih, dobimo, da je

$$\psi_{q_\ell}^2 x^{q^2} = x\psi_{q_\ell}^2 - \psi_{q_\ell - 1}\psi_{q_\ell + 1},$$

kar je očitno ekvivalentno (4.9). ■

Pogoj iz trditve (4.8) je uporaben, ker ima naslednjo lastnost:

Trditev 4.9 *Če pogoj iz trditve (4.8) okrajšamo po modulu krivulje E, v njem ne nastopa spremenljivka Y.*

Dokaz: Spomnimo se, da so polinomi $\psi_n(X,Y)$, kjer je n liho število, pravzaprav odvisni le od spremenljivke X. Če pa je n sodo število, velja

$$\psi_n(X,Y) = \psi_2(X,Y)\cdot r(X) = (2Y + a_1X + a_3)\cdot r(X),$$

kjer je $r \in \mathbb{F}_q[X]$. Velja

$$\psi_2^2(X,Y) = 4Y^2 + 4a_1XY + 2a_1a_3X + 4a_3Y + a_1^2X^2 + a_3^2.$$

Za vsoto prvih dveh členov na desni strani te enačbe upoštevamo enačbo krivulje E in ugotovimo, da je

$$\psi_2^2 \equiv s(X) \pmod{E}, \quad \text{kjer je } s \text{ nek polinom iz } \mathbb{F}_q[X]. \tag{4.10}$$

S pomočjo tega rezultata lahko dokončamo dokaz. Če je q_ℓ liho število, je polinom ψ_{q_ℓ} odvisen le od spremenljivke X. Iz enačbe (4.10) sledi, da enako velja za produkt $\psi_{q_\ell-1}\psi_{q_\ell+1}$. Če pa je q_ℓ sodo število, sta $\psi_{q_\ell-1}$ in $\psi_{q_\ell+1}$ odvisni le od X. Spet zaradi (4.10) enako velja za $\psi_{q_\ell}^2$. Ne glede na parnost števila q_ℓ je torej pogoj iz trditve (4.8) odvisen le od X. ∎

Pogoj iz trditve (4.8) lahko prevedemo v obliko, ki je primerna za uporabo v algoritmu. Najprej se spomnimo definicije največjega skupnega delitelja (gcd) v glavnih kolobarjih.

Definicija 4.10 *Naj bo k komutativen obseg. Vsak ideal glavnega kolobarja $k[X]$ je glavni, torej oblike (a) za nek polinom $a \in k[X]$.* **Največji skupni delitelj** *polinomov f in g iz $k[X]$ je tisti polinom a glavnega kolobarja $k[X]$, za katerega je ideal (a) enak idealu, generiranem s polinomoma f in g. Ker je polinom a s tem določen samo do neničelne multiplikativne konstante natančno, je tudi največji skupni delitelj določen samo do neničelne multiplikativne konstante natančno. V tem delu bomo vedno izbrali tako multiplikativno konstanto, da je vodilni koeficient največjega skupnega delitelja enak 1. Največji skupni delitelj označimo z $\gcd_{k[X]}(f,g)$.*

Trditev 4.11 *Praštevilo ℓ je za dana q in E tipa A ali B natanko takrat, ko velja*

$$\gcd\Big(\psi_{q_\ell}^2(X^{q^2} - X) + \psi_{q_\ell-1}\psi_{q_\ell+1}, \psi_\ell\Big) \neq 1.$$

Dokaz: V razdelku 3.4 o delitvenih polinomih smo videli, da so X-koordinate točk iz $E[\ell]^*$ natanko ničle polinoma ψ_ℓ (in vse te ničle so enostavne). Ker delamo v algebraično zaprtem obsegu $\overline{\mathbb{F}_q}$, sledi trditev iz trditve (4.8). ∎

Ta trditev nam podaja računsko enostaven kriterij, s katerim lahko ugotovimo, ali je tip praštevila ℓ enak C ali ne. Pripomnimo, da moramo v praksi pred preverjanjem zgornjega pogoja in vseh naslednjih podobnih pogojev izračunati $X^{q^2} \bmod E \bmod \psi_\ell$. Več o tem bomo povedali v podrazdelku 4.6.2 o razrešitvi primera C.

Največji skupni delitelj (gcd) iz trditve (4.11) se pravzaprav nanaša na kolobar $\overline{\mathbb{F}_q}[X]$, kar je računsko neugodno. Ta problem razreši naslednja lema.

Lema 4.12 *Naj bo K razširitev obsega k in naj bosta $f, g \in k[X] \subset K[X]$ poljubna polinoma. Potem je*

$$\gcd_{k[X]}(f,g) = \gcd_{K[X]}(f,g).$$

Opozorimo najprej, da ta lema, upoštevajoč ekzaktno definicijo največjega skupnega delitelja (4.10) ni tako očitna, kot se zdi na prvi pogled. Za njen dokaz se opremo na Evklidov algoritem, ki v vsakem glavnem kolobarju izračuna največji skupni delitelj dveh polinomov glede na ta glavni kolobar.

Dokaz:

Evklidov algoritem v $K[X]$ nam vrne $\gcd_{K[X]}(f,g)$. Ker pa sta f in g vsebovana že v $k[X]$, so koeficienti vseh polinomov, ki nastopajo v $K[X]$-verziji Evklidovega algoritma, vsebovani že v k. Zato dobimo pri $K[X]$-verziji Evklidovega algoritma enake delitelje in ostanke kot pri $k[X]$-verziji Evklidovega algoritma in sta zato tudi končna rezultata enaka. ∎

Vsi polinomi ki se pojavijo v Schoofovem algoritmu, imajo koeficiente v $\mathbb{F}_q$ in ne zgolj v $\overline{\mathbb{F}_q}$. Iz leme (4.12) sledi, da lahko zato pogoj iz trditve (4.11) in podobne pogoje preverimo s pomočjo Evklidovega algoritma v $\mathbb{F}_q[X]$ in nam torej ni potrebno računati z elementi iz $\overline{\mathbb{F}_q}$.

Če preverimo pogoj iz trditve (4.11), torej izvemo, ali nastopi eden od primerov A, B ali nastopi primer C. Pokažimo najprej, kako algoritem obdela primera A, B.

4.6.1 Razrešitev primerov A in B

Recimo, da smo s pomočjo pogoja iz trditve (4.11) ugotovili, da je praštevilo ℓ tipa A ali B. Ne vemo pa še, katerega tipa izmed A, B je praštevilo ℓ. Čim kjerkoli v nadaljevanju ugotovimo, da je praštevilo ℓ tipa A, lahko zaključimo $t_\ell = 0$. Tega v nadaljevanju ne bomo venomer ponavljali. Naredimo najprej nekaj teoretičnih sklepov, ki nam bodo podali učinkovit algoritem za točno določitev tipa praštevila ℓ.

Premislimo, kaj mora veljati, če je je ℓ tipa B. Po definiciji lahko v tem primeru poiščemo točko $P \in E[\ell]^*$, za katero je $\varphi_\ell(P) = q_\ell P$. Potem je $\varphi_\ell^2(P) = q_\ell \varphi_\ell(P)$. Iz karakteristične enačbe Frobeniusovega endomorfizma sledi

$$t_\ell \varphi_\ell(P) = 2q_\ell P. \tag{4.11}$$

Ker je $t_\ell \neq 0$, ima t_ℓ v $\mathbb{F}_\ell$ inverz $1/t_\ell$. Zato lahko obe strani zadnje enačbe pomnožimo z $1/t_\ell$ in dobimo

$$\varphi_\ell(P) = \frac{2q_\ell}{t_\ell} P. \tag{4.12}$$

Če to enakost vstavimo nazaj v karakteristično enačbo Frobeniusovega endomorfizma, dobimo

$$\left(\frac{4q_\ell^2}{t_\ell^2} - q_\ell \right) P = \mathcal{O}. \tag{4.13}$$

Ker imajo vse netrivialne točke grupe $E[\ell]$ red ℓ in ker je $q_\ell \neq 0$, mora veljati

$$4q_\ell^2 \equiv t_\ell^2 \pmod{\ell}. \tag{4.14}$$

Tako smo pokazali

Trditev 4.13 *Če je ℓ tipa B, je število q_ℓ kvadraten ostanek v $\mathbb{F}_\ell$.* ∎

Ta pogoj preverimo z Jacobijevim simbolom $\left(\frac{q_\ell}{\ell}\right)$. Ker je $q_\ell \neq 0$, je ta Jacobijev simbol enak bodisi 1 bodisi -1. Če je enak -1, q_ℓ ni kvadratni ostanek, zato je ℓ tipa A. Če pa je Jacobijev simbol enak 1, še vedno ne vemo, katerega tipa je ℓ. Potrebni so nadaljni testi.

Pot pelje naprej takole. Ker je q_ℓ kvadratni ostanek, lahko ne glede na tip praštevila ℓ določimo kvadratni koren števila q_ℓ v $\mathbb{F}_\ell$. Korena sta dva. Oba sta neničelna in sta med seboj različna. Enega od njih poiščemo in označimo z ω. Praštevilo ℓ je razmeroma majhno (reda velikosti $O(\log q)$), zato lahko ω poiščemo kar z metodo grobe sile. Pripomnimo, da je naša izbira enega od korenov izmed obeh korenov čisto slučajna. Tistega že, ki ga pač naš algoritem za iskanje korenov izbere, označimo z ω, drugi koren pa je potem enak $-\omega = \ell - \omega$. Prav lahko pa bi bilo obratno. Tipa A in B nam zdaj dokončno loči naslednja trditev.

Trditev 4.14 *Naj bo*

$$\rho_\ell(X) = \gcd\left(\psi_\omega^2 \cdot (X^q - X) + \psi_{\omega-1}\psi_{\omega+1}, \psi_\ell\right)$$

in naj bo praštevilo ℓ pri danih E in q tipa A ali B. Potem je ℓ tipa B natanko tedaj, ko je $\rho_\ell \neq 1$.

Dokaz:
($\Rightarrow$) Naj bo ℓ tipa B. Potem obstaja točka $P \in E[\ell]^*$, ki zadošča enačbam (4.11), (4.12) in (4.13). Iz kongruence (4.14) sledi, da je bodisi $t_\ell = 2\omega$ bodisi $t_\ell = -2\omega$. Ker je $\omega \neq \ell - \omega$, sta primera disjunktna. Iz enakosti (4.12) sledi

$$\varphi_\ell(P) = \frac{2q_\ell}{\pm 2\omega} P = \frac{2\omega^2}{\pm 2\omega} P = \pm \omega P,$$

pri čemer nastopi natanko ena od možnosti $+, -$. Potemtakem za točko P velja

$$X\Big(\varphi_\ell(P)\Big) = X\Big(\omega P\Big).$$

Ta pogoj pa je, po enakem premisleku kot pri trditvi (4.11), ekvivalenten pogoju $\rho_\ell \neq 1$.
($\Leftarrow$) V tem primeru obstaja točka $P \in E[\ell]^*$, za katero velja $X(\varphi_\ell(P)) = X(\omega P)$. Torej ima P eno od lastnosti $\varphi_\ell P = \omega P$, $\varphi_\ell P = -\omega P$. Pripomnimo, da P ne more imeti obeh lastnosti hkrati, ker je sicer $2\omega P = \mathcal{O}$, število 2ω pa ni deljivo s praštevilom ℓ. Kakorkoli že, v obeh primerih velja $\varphi_\ell^2(P) = q_\ell P$. Recimo, da bi praštevilo ℓ bilo tipa A. Potem bi veljalo $t_\ell = 0$. Iz karakteristične enačbe Frobeniusovega endomorfizma bi potem sledilo $\varphi_\ell^2(P) = -q_\ell P$. Potemtakem bi moralo veljati $2q_\ell P = \mathcal{O}$. Ker število $2q_\ell$ ni deljivo s praštevilom ℓ, je to protislovje. ■

Izračun polinoma ρ_ℓ nam na primernem mestu v algoritmu torej pove, kateri od primerov A oziroma B nastopi. Ko to določimo, je potrebno ugotoviti le še, katera od možnosti $t_\ell = \pm 2\omega$ je prava v primeru B. Na to vprašanje nam odgovori naslednja trditev. Opozorimo, da v njej nastopajo polinomi iz družine $\{\omega_n\}_n$, hkrati pa tudi kvadratni koren števila q_ℓ v obsegu $\mathbb{F}_\ell$, ki smo ga dosedaj tudi označevali z ω. Da ne bi prišlo do zmede z oznakami, bomo samo v tej trditvi in v neposrednem komentarju za njo za kvadratni koren ω uporabljali oznako ν.

Trditev 4.15 *Naj bo praštevilo ℓ pri danih E in q tipa B. Naj bosta a in b polinoma iz $\mathbb{F}_q[X]$, ki ju dobimo, če enačbo*

$$Y^q\psi_\nu^3(X,Y) = \omega_\nu(X,Y) \tag{4.15}$$

reduciramo po modulu krivulje E, tako da dobimo enačbo oblike

$$a(X)Y + b(X) = 0, \qquad \text{kjer sta } a, b \in \mathbb{F}_q[X]. \tag{4.16}$$

Naj bodo $a_i \in \mathbb{F}_q,\ i = 1,2,3,4,6$ parametri krivulje E. Označimo

$$h(X) = b^2(X) - a_1Xa(X)b(X) - a_3a(X)b(X) -$$
$$-X^3a^2(X) - a_2X^2a^2(X) - a_4Xa^2(X) - a_6a^2(X).$$

Potem je $t_\ell = 2\nu$ natanko tedaj, ko je $\gcd(\rho_\ell, h) \neq 1$.

Pripomnimo najprej, da smo polinom ρ_ℓ definirali v formulaciji trditve (4.14). Polinoma a in b sta za dano enačbo (4.15) enolično določena, ker v kolobarju $\mathbb{F}_q[X,Y] \bmod E$ iz $t(X) + Yu(X) = 0$ sledi $t = u = 0$. Enačba (4.15) je enačba, ki jo dobimo, če zapišemo enakost Y-koordinat točk $\varphi_\ell\Big((X,Y)\Big)$ in $\nu(X,Y)$ in odpravimo ulomek. Ker je $\nu \neq 0$, za točke P iz $E[\ell]^*$ velja $\nu P \neq \mathcal{O}$ in $\psi_\nu(P) \neq \mathcal{O}$, zato pri tem ni težav z ničelnim imenovalcem. Izraz za h smo dobili tako, da smo iz enakosti $a(X) + Yb(X) = 0$ simbolično izrazili Y kot racionalno funkcijo spremenljivke X, to zvezo vstavili v enačbo krivulje E in odpravili ulomke.

Dokaz:
($\Rightarrow$) Naj bo $t_\ell = 2\nu$. Potem enakost (4.12) pove, da obstaja točka $P = (x,y) \in E[\ell]^*$, tako da je $\varphi_\ell(P) = \nu P$. Iz enakosti X-koordinat te enačbe potem sledi $\rho_\ell(x) = 0$. Enakost Y-koordinat pa nam da enačbo (4.15), v katero smo vstavili $X = x, Y = y$. Torej je $a(x)y + b(x) = 0$. Če je $a(x) = 0$, mora biti $b(x) = 0$ in zato $h(x) = 0$. Če pa je $a(x) \neq 0$, lahko izrazimo $y = -b(x)/a(x)$. To vstavimo v enačbo krivulje in dobimo $h(x) = 0$. V obeh primerih torej velja $h(x) = 0$, kar pomeni, da je $\gcd(\rho_\ell, h) \neq 1$.
($\Leftarrow$) V tem primeru obstaja $x \in \overline{\mathbb{F}_q}$, tako da je $\rho_\ell(x) = h(x) = 0$. Iz $\rho_\ell(x) = 0$ sledi, da obstaja $y \in \overline{\mathbb{F}_q}$, tako da je $(x,y) \in E[\ell]^*$ in velja $X\Big(\varphi_\ell((x,y))\Big) = X\Big(\nu(x,y)\Big)$. Recimo, da velja $a(x) = 0$. Zaradi $h(x) = 0$ mora potem veljati tudi $b(x) = 0$ in je za $X = x, Y = y$ zadoščeno enačbi (4.16). Zato par (x,y) zadosti tudi enačbi (4.15). Potem pa je $\varphi_\ell\Big((x,y)\Big) = \nu(x,y)$ in tako $t_\ell = 2\nu$.

Oglejmo si še primer $a(x) \neq 0$. Tokrat iz $h(x) = 0$ sledi, da je točka $(x, -b(x)/a(x))$ na krivulji E. Ker je tudi $(x,y) \in E$, po lemi (4.7) sledi, da je bodisi $(x, -b(x)/a(x)) = (x,y)$ bodisi $(x, -b(x)/a(x)) = -(x,y)$. V prvem primeru je potem za $X = x, Y = y$ očitno zadoščeno enačbi (4.16), od koder z enakim sklepom kot v primeru $a(x) = 0$ zaključimo, da je $t_\ell = 2\nu$. V drugem primeru pa dobimo, da točka $-(x,y)$ zadošča enačbi (4.16) in zato tudi enačbi (4.15). Ker enačbi $X\Big(\varphi_\ell((x,y))\Big) = X\Big(\nu(x,y)\Big)$ poleg točke (x,y) zadošča tudi točka $-(x,y)$, od tod sledi $\varphi_\ell\Big(-(x,y)\Big) = \nu\Big(-(x,y)\Big)$. Torej obstaja neka točka $P' \in E[\ell]^*$, tako da je $\varphi_\ell(P') = \nu P'$. Zato je $t_\ell = 2\nu$. ∎

Za testiranje pogoja iz te trditve moramo med izvajanjem algoritma določiti polinom $h \in \mathbb{F}_q[X]$, ki je odvisen od E, q, ℓ in ν. Pri tem upoštevamo spoznanja iz razdelka 3.4 o delitvenih polinomih, ki nam povedo, kakšno obliko imajo za našo krivuljo oziroma karakteristiko polinomi ω_n in ψ_n.

4.6.2 Primer C

Zdaj si oglejmo še, kako določimo število t_ℓ v primeru, ko smo s pomočjo trditve (4.11) ugotovili, da je praštevilo ℓ tipa C. V tem primeru vemo, da je $t_\ell \neq 0$. Iščemo torej $\tau \in \mathbb{Z},\ 0 < \tau < \ell$, za katerega za vsak $P \in E[\ell]^*$ velja

$$\varphi_\ell^2(P) + q_\ell P = \tau\varphi_\ell(P). \tag{4.17}$$

V nadaljevanju bomo za poljuben $P = (x, y) \in E[\ell]^*$ eksplicitno izračunali obe strani te enakosti. Rezultata sta seveda odvisna od x in y. Vsaka od strani enakosti tako postane urejen par nekih racionalnih funkcij dveh spremenljivk x in y. Vsi izračuni v tem podrazdelku veljajo za poljubno točko $P = (x, y) \in E[\ell]^*$. To bomo poudarili tako, da bomo za argumenta racionalnih funkcij in polinomov v tem razdelku namesto običajnih oznak X in Y uporabljali kar x in y.

Koordinati X oziroma Y leve in desne strani predelane enakosti (4.17) bomo izenačili in na podoben način kot v primerih A in B ugotovili, kateri τ je pravi. Opozorimo, da je večina premislekov in tudi izračunov, ki jih bomo navedli, namenjena izključno utemeljevanju korakov Schoofovega algoritma. Na izračune, ki jih je potrebno pri izvedbi Schoofovega algoritma dejansko izvesti, bomo posebej opozorili.

Poiščimo torej ekspliciten izraz za levo in desno stran enakosti (4.17). Ker imajo netrivialne točke grupe $E[\ell]$ red ℓ in ker je φ_ℓ injektivna, za vsak $P \in E[\ell]^*$ in vsak $\tau \in \{1, \ldots, \ell - 1\}$ velja $\tau\varphi_\ell(P) \neq \mathcal{O}$. Zato lahko uporabimo formulo za množenje točke $\varphi_\ell(P)$ z naravnim številom τ :

$$\tau\varphi_\ell(P) = \left(\frac{\theta_\tau\left(\varphi_\ell(P)\right)}{\psi_\tau^2\left(\varphi_\ell(P)\right)}, \frac{\omega_\tau\left(\varphi_\ell(P)\right)}{\psi_\tau^3\left(\varphi_\ell(P)\right)} \right).$$

Če upoštevamo še formulo (4.8), dobi enakost (4.17) obliko

$$(x^{q^2}, y^{q^2}) + \left(\frac{\theta_{q_\ell}(x,y)}{\psi_{q_\ell}^2(x,y)}, \frac{\omega_{q_\ell}(x,y)}{\psi_{q_\ell}^3(x,y)} \right) = \left(\frac{\theta_\tau(x^q, y^q)}{\psi_\tau^2(x^q, y^q)}, \frac{\omega_\tau(x^q, y^q)}{\psi_\tau^3(x^q, y^q)} \right). \tag{4.18}$$

V nadaljevanju bomo najprej sešteli sumanda na levi strani zgornje enakosti (4.18). Ker smo v primeru C, za vsak $P \in E[\ell]^*$ velja, da $\varphi_\ell^2(P)$ ni niti $q_\ell P$ niti $-q_\ell P$. Hkrati nobena od točk $\varphi_\ell^2(P)$ in $q_\ell P$ ni enaka $\mathcal{O}$. Torej sta sumanda na levi strani (4.18) za vsak $(x, y) \in E[\ell]^*$ netrivialna, med seboj različna in nista drug drugemu inverzna. Zato ju moramo sešteti z uporabo grupnega sekantnega pravila. Za sumanda na levi strani enakosti (4.18) vpeljimo oznaki

$$\begin{aligned} (x_1, y_1) &= (x^{q^2}, y^{q^2}) \qquad \text{in} \\ (x_2, y_2) &= \left(\frac{\theta_{q_\ell}}{\psi_{q_\ell}^2}, \frac{\omega_{q_\ell}}{\psi_{q_\ell}^3} \right). \end{aligned}$$

Zdaj torej seštejemo točki (x_1, y_1) in (x_2, y_2) po sekantnem grupnem pravilu. Njuno vsoto označimo z (x_3, y_3). Na ta način lahko za vsako točko $P = (x, y)$ iz $E[\ell]^*$ določimo obe strani enakosti (4.17) oziroma (4.18). Opomnimo, da naš račun morda ni pravilen za nekatere točke $P = (x, y)$ iz $E\backslash E[\ell]$, saj bi za te točke morda morali uporabiti tangentno grupno pravilo. Vendar nas to ne moti, ker pri iskanju števila t_ℓ operiramo le s točkami iz $E[\ell]$.

Formule, ki podajajo sekantno pravilo, so za obe obravnavani kanonični obliki (binarni nesupersingularni primer in primer velike karakteristike) različne. Vsako od kanoničnih oblik bomo zato v nadaljevanju obravnavali posebej.

Primer velike karakteristike $p = q > 3$

Uporabimo sekantno grupno pravilo in formule za ω_n in θ_n za ta primer. Vpeljimo pomožne količine

$$\begin{aligned}
\alpha(x,y) &= \psi_{q_\ell+2}\psi^2_{q_\ell-1} - \psi_{q_\ell-2}\psi^2_{q_\ell+1} - 4\psi^3_{q_\ell} y^{q^2+1}, \\
\beta(x,y) &= 4y\psi_{q_\ell}\Big((x - x^{q^2})\psi^2_{q_\ell} - \psi_{q_\ell-1}\psi_{q_\ell+1}\Big) \qquad \text{in} \\
\lambda(x,y) &= \frac{y_2 - y_1}{x_2 - x_1}.
\end{aligned}$$

Ker smo v primeru C, je $x_2 \neq x_1$ za vsak $(x, y) \in E[\ell]^*$ ali, kar je ekvivalentno, $\beta(x, y) \neq \mathcal{O}$ za vsak $(x, y) \in E[\ell]^*$. Zato je izraz λ na $E[\ell]^*$ dobro definiran. S kratkim računom lahko preverimo, da je $\lambda = \alpha/\beta$. Sekantni grupni zakon potem pravi, da je

$$\begin{aligned}
x_3 &= \lambda^2 - x_1 - x_2 = \frac{\alpha^2}{\beta^2} - x^{q^2} - \frac{\theta_{q_\ell}}{\psi^2_{q_\ell}} \qquad \text{in} \\
y_3 &= \lambda(x_1 - x_3) - y_1 = \frac{\alpha}{\beta}\Big(x^{q^2} - (\frac{\alpha^2}{\beta^2} - x^{q^2} - \frac{\theta_{q_\ell}}{\psi^2_{q_\ell}})\Big) - y^{q^2}.
\end{aligned}$$

Zdaj lahko poenostavimo enakost (4.18). Izpišimo najprej enakost X-koordinat te enačbe, ki se glasi $x_3 = \theta_\tau(x^q, y^q)/\psi^2_\tau(x^q, y^q)$. V enakost vstavimo dobljeno formulo za x_3, izrazimo delitvene polinome θ_n s ψ_n in x, odpravimo ulomke in dobimo enakost

$$\psi_\tau^{2q}\left[\beta^2\left(\psi_{q_\ell-1}\psi_{q_\ell+1} - (x + x^q + x^{q^2})\psi^2_{q_\ell}\right) + \alpha^2\psi^2_{q_\ell}\right] + \psi^q_{\tau-1}\psi^q_{\tau+1}\beta^2\psi^2_{q_\ell} = 0. \tag{4.19}$$

Polinom na levi strani te enakosti označimo z $g_X(x, y)$. Polinom g_X je odvisen tudi od y, saj spremenljivka y nastopa v pomožnih polinomih α in β. Pogoj (4.19) želimo zamenjati z ekvivalentnim pogojem, v katerem nastopa samo spremenljivka x. Ker v g_X spremenljivka y povsod nastopa na sodo potenco, jo lahko z zaporednim večkratnim upoštevanjem enačbe eliptične krivulje popolnoma eliminiramo. Enostavno je premisliti, da se pri zaporednem upoštevanju enačbe krivulje množica točk krivulje, ki zadostijo enačbi (4.19), ne spreminja. Na ta način izraz g_X preoblikujemo v polinom, ki je odvisen le od x. Ta polinom je z opisanim postopkom enolično določen. Označimo ga s $h_X(x)$.

Enak postopek ponovimo še za Y-koordinato enačbe (4.18). Delitveni polinom ω_n izrazimo s ψ_n, x in y in odpravimo ulomke. Dobimo enačbo

$$4y^q\psi_\tau^{3q}\Big[\alpha\Big(\psi_{q_\ell}^2[2\beta^2x^{q^2}+\beta^2x-\alpha^2]-\beta^2\psi_{q_\ell-1}\psi_{q_\ell+1}\Big)-\beta^3y^{q^2}\psi_{q_\ell}^2\Big]-$$
$$-\beta^3\psi_{q_\ell}^2(\psi_{\tau+2}\psi_{\tau-1}^2-\psi_{\tau-2}\psi_{\tau+1}^2)^q=0.$$

Polinom na levi strani zgornje enačbe označimo z $g_Y(x,y)$. Enostavno je videti, da tokrat po eliminaciji količin y^2 iz izraza za g_Y dobimo enačbo oblike $y\cdot h_Y(x)=0$, kjer je h_Y nek polinom, odvisen zgolj od spremenljivke x. V $E[\ell]^*$ ni nobene točke z Y-koordinato enako 0, saj imajo take točke red 2. Zato lahko enačbo $y\cdot h_Y(x)=0$ okrajšamo z y in dobimo ekvivalentno polinomsko enačbo ene spremenljivke $h_Y(x)=0$.

Ker se množica rešitev enačb $g_X(x,y)=0$ in $h_X(x,y)=0$ pri posameznih transformacijah ohranja, sledi, da je $\tau=t_\ell$ natanko takrat, kadar za vsak $P=(x,y)\in E[\ell]^*$ velja

$$h_X(x)=h_Y(x)=0.$$

Kako nadaljujemo od tu dalje, bomo pokazali v mini podrazdelku z naslovom Dokončna razrešitev primera C, kjer bomo združili oba obravnavana primera.

Nesupersingularni binarni primer

Ta primer se od prejšnjega razlikuje le v obliki formul za grupno sekantno vsoto in v obliki delitvenih polinomov. V binarnem primeru ponavadi delamo z družino delitvenih polinomov f_n. Enako kot prej se spomnimo formul za ω_n in θ_n za ta primer in izračunamo pomožne količine

$$\begin{aligned}
\alpha &= (y^{q^2}+y+x)xf_{q_\ell}^3+f_{q_\ell-2}f_{q_\ell+1}^2+(x^2+x+y)f_{q_\ell-1}f_{q_\ell}f_{q_\ell+1},\\
\beta &= xf_{q_\ell}^3(x+x^{q^2})+xf_{q_\ell-1}f_{q_\ell}f_{q_\ell+1},\\
\lambda &= \frac{y_2+y_1}{x_2+x_1}.
\end{aligned}$$

Pripomnimo, da v zgornjih formulah nastopa $f_{q_\ell-2}$, ki v primeru $q_\ell=1$ ni definiran. Če ga za potrebe Schoofovega algoritma definiramo kot $f_{-1}=1$, pa zgornje formule držijo tudi v tem primeru. S tem pravilno odpravimo tudi analogen problem, ki se v nadaljevanju pojavi pri $\tau=1$. Enako kot v primeru $q=p>3$ velja $\beta\neq\mathcal{O}$ in $\lambda=\alpha/\beta$. Sekantni grupni zakon potem pravi, da je

$$\begin{aligned}
x_3 &= \lambda^2+\lambda+a_2+x_1+x_2=\frac{\alpha^2}{\beta^2}+\frac{\alpha}{\beta}+a_2+x^{q^2}+x+\frac{f_{q_\ell-1}f_{q_\ell+1}}{f_{q_\ell}^2}\\
y_3 &= \frac{\alpha}{\beta}\Big(x_1+x_3\Big)+x_3+y_1=\\
&= \frac{\alpha}{\beta}x^{q^2}+\Big(\frac{\alpha}{\beta}+1\Big)\Big(\frac{\alpha^2}{\beta^2}+\frac{\alpha}{\beta}+a_2+x^{q^2}+x+\frac{f_{q_\ell-1}f_{q_\ell+1}}{f_{q_\ell}^2}\Big)+y^{q^2}.
\end{aligned}$$

Tako kot v primeru $q=p>3$ lahko zdaj poenostavimo enakost (4.18). Če dobljeno formulo za x_3 upoštevamo v enakosti (4.18), izrazimo delitvene polinome ψ_n,θ_n s f_n in odpravimo ulomke, se enakost X-koordinat enačbe (4.18) glasi

$$f_\tau^{2q}\Big(f_{q_\ell-1}f_{q_\ell+1}\beta^2+f_{q_\ell}^2[\alpha^2+\alpha\beta+\beta^2(a_2+x^{q^2}+x^q+x)]\Big)+\beta^2f_{q_\ell}^2f_{\tau-1}^qf_{\tau+1}^q=0 \quad (4.20)$$

Levo stran te enačbe tako kot v primeru $q = p > 3$ označimo z $g_X(x,y)$. Izraz $g_X(x,y)$ je odvisen tudi od y, ker spremenljivka y nastopa v polinomu α. Spremenljivke y v tem primeru ne moremo tako neposredno eliminirati kot v primeru $q = p > 3$. Z zaporednim upoštevanjem enačbe krivulje E lahko iz izraza $g_X(x,y)$ eliminiramo višje potence spremenljivke y in dobimo enačbo, ki je linearna v y. Koeficienta te enačbe sta polinomski funkciji spremenljivke x. Iz te enačbe simbolično izrazimo y kot racionalno funkcijo spremenljivke x in dobljeni izraz vstavimo v enačbo krivulje E. Dobljeni polinom ene spremenljivke iz $\mathbb{F}_q[X]$ označimo s $h_X(x)$. Ta postopek smo že srečali v dokončni razrešitvi primera B, natančneje v trditvi (4.15). Enako kot tam tudi tukaj velja, da je polinom h_X za dane q, E, ℓ, τ enolično določen. V komentarju k trditvi (4.15) smo tudi razložili, zakaj pri tem postopku v nobenem primeru nimamo težav z ničelnimi imenovalci.

Enak postopek ponovimo še za Y-koordinato enačbe (4.18). Po odpravi ulomkov dobimo enačbo

$$g_Y(x,y) = \Big(\alpha x^{q^2} + \beta(x^q + y^q + y^{q^2})\Big)\beta^2 x^q f_\tau^{3q} f_{q_\ell}^2 +$$
$$+(\alpha+\beta)x^q f_\tau^{3q}\Big((\alpha^2 + \alpha\beta + \beta^2(a_2 + x^{q^2} + x))f_{q_\ell}^2 + \beta^2 f_{q_\ell-1} f_{q_\ell+1}\Big) +$$
$$+\beta^3 f_{q_\ell}^2\Big((x^{2q} + x^q + y^q) f_{\tau-1}^q f_\tau^q f_{\tau+1}^q + f_{\tau-2}^q f_{\tau+1}^{2q}\Big) = 0.$$

To enačbo predelamo na enak način kot enačbo za X-koordinato in dobimo enolično določeno polinomsko enačbo ene spremenljivke $h_Y(x) = 0$.

Naslednjo trditev dokažemo na enak način, kot smo dokazali trditev (4.15), zato njen dokaz izpustimo.

Trditev 4.16 *Naj bo (x,y) poljubna točka iz $E[\ell]^*$ in $\tau \in \{1,\ldots,\ell-1\}$ poljubno število. Po zgoraj opisanem postopku pri danem τ iz polinomov dveh spremenljivk g_X in g_Y konstruirajmo polinoma ene spremenljivke h_X in h_Y. Potem velja $h_X(x) = h_Y(x) = 0$ natanko takrat, ko točka (x,y) zadosti enačbi (4.18).* ∎

Iz te trditve neposredno sledi, da je $\tau = t_\ell$ natanko takrat, kadar za vsak $P = (x,y) \in E[\ell]^*$ velja $h_X(x) = h_Y(x) = 0$.

Dokončna razrešitev primera C

Naša ločena razmišljanja obeh zgornjih kanoničnih primerov lahko zdaj spet združimo. Skupni zaključek obeh primerov je namreč, da je $\tau = t_\ell$ natanko takrat, kadar za vsak $(x,y) \in E[\ell]^*$ velja

$$h_X(x) = h_Y(x) = 0.$$

Pravi τ je torej natanko tisti, pri katerem ψ_ℓ deli polinoma h_X in h_Y. Schoofov algoritem za posamezno praštevilo ℓ v primeru C poteka tako, da zgornji pogoj po vrsti preverja za posamezne τ, dokler ne naleti na tisti τ, za katerega je $\tau = t_\ell$. Polinomov h_X in h_Y pri tem nikoli ne določimo direktno, ker bi v tem primeru morali v računalniku hraniti polinome, katerih stopnja je eksponentna v parametru $\log q$ in bi bilo zabave kmalu konec. Da se temu izognemo, mora algoritem znati izvajati elementarne operacije z elementi kolobarja $\mathbb{F}_q[X] \bmod \psi_\ell$, to je s polinomi ene spremenljivke s koeficienti v obsegu $\mathbb{F}_q$, reduciranimi modulo ψ_ℓ. Poleg tega pa mora znati računati tudi s polinomi dveh spremenljivk oblike $r(X)+Ys(X)$ po modulu enačbe krivulje E in hkrati po modulu ψ_ℓ. To zadnjo aritmetiko je možno enostavno izvesti s pomočjo aritmetike v kolobarju $\mathbb{F}_q[X] \bmod \psi_\ell$.

S tako aritmetiko se lahko omenjenim polinomom prevelike stopnje izognemo. Pri tem uporabimo preprosto dejstvo, da polinom ψ_ℓ deli h_X (oz. h_Y) natanko tedaj, ko deli polinom $h_X + r\psi_\ell$ (oz. $h_Y + r\psi_\ell$), kjer je r nek poljuben polinom ene spremenljivke. Algoritem lahko zato pri pretvorbi izrazov $g_X(x,y)$ v $h_X(x)$ in $g_Y(x,y)$ v $h_Y(x)$ vse vmesne rezultate sproti okrajša po modulu ψ_ℓ. Ta ugodnost pomeni velike prihranke na času izvajanja in potrebnem prostoru.

V izrazih za $g_X(x,y)$ in $g_Y(x,y)$ nastopata količini x^q in x^{q^2}. V algoritmu običajno najprej vsako od teh količin posebej reduciramo po modulu ψ_ℓ. Pri tem gre pravzaprav za potenciranje v kolobarju $\mathbb{F}_q[X] \bmod \psi_\ell$, ki ga opravimo s pomočjo algoritma kvadriraj in množi. V izrazih za $g_X(x,y)$ in $g_Y(x,y)$ se pojavljata tudi količini y^q in y^{q^2}. Vsako od teh dveh količin reduciramo z izmeničnimi redukcijami modulo ψ_ℓ in modulo E. Bralca opozorimo, da velja $y^{q^2} \neq (y^q)^2$, zato rezultata za y^{q^2} ne dobimo trivialno iz rezultata za y^q. Lahko pa izkoristimo enakost $y^{q^2} = (y^q)^q$ in si pri redukciji izraza y^{q^2} pomagamo z že reduciranim izrazom y^q. Isti komentar seveda velja tudi za par x^q, x^{q^2}.

Vse te račune je treba opraviti za vsak τ posebej. Delitvene polinome ψ_n, $n \in \{1, \ldots, \max \mathcal{L}\}$ ponavadi enkrat za vselej izračunamo v začetni fazi Schoofovega algoritma, saj jih potrebujemo za vsa praštevila ℓ. Pri tem oznaka $\max \mathcal{L}$ pomeni običajem maksimum končne množice $\mathcal{L} \subset \mathbb{R}$. Poleg običajnih delitvenih polinomov pri vsakem praštevilu ℓ potrebujemo tudi redukcije q-tih potenc $\psi_0^q, \ldots, \psi_{\ell-1}^q$ modulo ψ_ℓ. Za vsako praštevilo ℓ moramo zato najprej določimo te redukcije. Pri tem lahko izkoristimo, da so koeficienti formul za običajne polinome ψ_n iz obsega $\mathbb{F}_q$, kjer za vsak element x velja $x^q = x$. Zato za polinome ψ_n^q veljajo enake rekurzivne formule kot za polinome ψ_n.

Za vsak τ dobljene rezultate vstavimo v $g_X(x,y)$ oziroma $g_Y(x,y)$ in preverimo, če ψ_ℓ deli dobljena polinoma.

Definicija 4.17 *Zgornja dva testa imenujemo zaporedoma* $\boldsymbol{X}$**-test** *in* $\boldsymbol{Y}$**-test***.*

Pripomnimo, da je po prej povedanem X-test ekvivalenten preverjanju pogoja $\psi_\ell \,|\, h_X$, Y-test pa je ekvivalenten preverjanju pogoja $\psi_\ell \,|\, h_Y$. Opozorimo še, da začetnih enačb $g_X(x,y) = 0$ in $g_Y(x,y) = 0$ ne določi algoritem sam, ampak jih programer prepiše iz svojih zapiskov in vnese neposredno v izvorno kodo programa.

Algoritem običajno organiziramo na način, da za vsak posamezen τ najprej izvedemo X-test. Če je test negativen, trenutni τ ni pravi. Kot bo sledilo iz nadaljevanja, v tem primeru velja $t_\ell \neq \tau$ in $t_\ell \neq -\tau$. X-test nam tako za posamezen τ pove, ali velja $t_\ell \in \{\tau,\ \ell-\tau\}$ in na ta način pravzaprav ujame dve muhi na en mah. Zato je potrebno z X-testom testirati le števila $\tau = 1, 2, \ldots, \lceil (\ell-1)/2 \rceil$.

Recimo, da število τ opravi X-test. Z drugimi besedami to pomeni, da za vsak $P \in E[\ell]^*$ velja

$$X\Big(\varphi_\ell^2(P) + q_\ell P\Big) = X\Big(\tau\varphi_\ell(P)\Big).$$

Naj bo $P \in E[\ell]^*$ poljubna točka. Ker imata $\varphi_\ell^2(P) + q_\ell P$ in $\tau\varphi_\ell(P)$ isto X-koordinato, po lemi (4.7) nastopi natanko ena od možnosti

$$\varphi_\ell^2(P) + q_\ell P = \tau\varphi_\ell(P) \quad \text{bodisi} \quad \varphi_\ell^2(P) + q_\ell P = -\tau\varphi_\ell(P).$$

Po trditvi (4.4) pa je potem v prvem primeru $t_\ell = \tau$, v drugem primeru pa $t_\ell = -\tau = \ell - \tau$. Zato za vsak τ, ki je opravil X-test, nastopi natanko ena od naslednjih dveh možnosti:

1. $\varphi_\ell^2(P) + q_\ell P = \tau\varphi_\ell(P)$ za vsak $P \in E[\ell]^*$,
2. $\varphi_\ell^2(P) + q_\ell P = -\tau\varphi_\ell(P)$ za vsak $P \in E[\ell]^*$.

Prvi primer nastopi natanko takrat, ko je $t_\ell = \tau$. Ker vemo, da število τ zadosti X-testu, nastopi prvi primer torej natanko tedaj, ko število τ izpolni tudi Y-test. Ta zadnji pogoj preverimo. Če je izpolnjen, je $t_\ell = \tau$, sicer pa je $t_\ell = -\tau = \ell - \tau$. S tem smo dokončno enolično določili t_ℓ.

4.7 Aritmetika v kolobarjih polinomov nad končnim obsegom

V Schoofovem algoritmu so v ospredju različne aritmetične operacije s polinomi iz kolobarja $\mathbb{F}_q[X]$, kjer je $q = p^n$, p praštevilo in $n \in \mathbb{N}$. V tem razdelku bomo ocenili časovno zahtevnost posameznih takšnih operacij.

Najprej navedimo osnovna dejstva o predstavitvi končnega obsega v računalniku. O tem smo nekaj povedali že v razdelku 1.5 o polinomskih in normalnih bazah. Končni obseg $\mathbb{F}_q$ ima seveda p^n elementov. Zato lahko z $\lceil \log_2 p^n \rceil = \lceil n \log_2 p \rceil$ bitnimi mesti natančno določimo posamezne elemente obsega. Vsak element obsega torej zakodiramo v neko zaporedje ničel in enic dolžine $\lceil n \log_2 p \rceil$. Pri običajnih načinih predstavitve končnega obsega je časovna zahtevnost seštevanja $O(\log q)$ bitnih operacij, množenja $O(\log^2 q)$ bitnih operacij in deljenja $O(\log^3 q)$ bitnih operacij. Pri tem ena bitna operacija pomeni en izračun izraza oblike a XOR b, kjer sta $a, b \in \{0, 1\}$. Opomnimo, da na računalnikih ponavadi posamezne bite združimo v 32 ali 64-bitne besede in potem bitne operacije izvajamo na besedah in ne na posameznih bitih. Na asimptotske ocene časovne zahtevnosti ta tehnična podrobnost ne vpliva.

Ocenimo zdaj, kako hitro lahko izvajamo operacije v kolobarju polinomov $\mathbb{F}_q[X]$. Rezultat nam bo služil neposredno za oceno časovne zahtevnosti Schoofovega algoritma. Ker so množenja v $\mathbb{F}_q$ bistveno počasnejša od seštevanj, bomo šteli samo potrebna množenja in morebitna deljenja. Ključni rezultat sta naslednji dve lemi, ki ju je preprosto dokazati, zato dokaz izpustimo. Kot vedno naj oznaka deg pomeni stopnjo polinoma.

Lema 4.18 *Naj bosta $f, g \in \mathbb{F}_q[X]$ in naj bo $\deg f \geq \deg g$. Naj bo še $\alpha \in \mathbb{F}_q$. Potem za izračun polinoma $f + g$ ne potrebujemo množenj in deljenj, ampak samo $O(\deg f)$ seštevanj v obsegu $\mathbb{F}_q$. Za izračun polinoma αf potrebujemo $O(\deg f)$ množenj v obsegu $\mathbb{F}_q$ in nobenega deljenja. Za izračun polinoma fg pa potrebujemo $O\Big((\deg f)(\deg g)\Big)$ množenj v obsegu $\mathbb{F}_q$ in nobenega deljenja.* ■

Lema 4.19 *Naj bosta f in g kot v zgornji lemi. Potem polinomsko deljenje polinoma f s polinomom g, pri katerem izračunamo kvocient in ostanek, zahteva $O\Big((\deg g)(\deg f - \deg g)\Big)$ množenj elementov iz $\mathbb{F}_q$ in $O(\deg f - \deg g)$ deljenj elementov iz $\mathbb{F}_q$.* ■

Polinomsko deljenje polinoma stopnje $2n$ s polinomom stopnje n tako terja $O(n^2)$ osnovnih operacij v obsegu. V kolobarjih polinomov, okrajšanih po modulu nekega polinoma stopnje n, takšno deljenje običajno potrebujemo za izvedbo množenja.

Naslednja lema oceni kompleksnost izračuna največjega skupnega delitelja.

Lema 4.20 *Naj bosta f in g polinoma iz $\mathbb{F}_q[X]$ stopnje manjše ali enake n. Potem za izračun največjega skupnega delitelja $\gcd(f,g)$ potrebujemo $O(n^2)$ množenj elementov iz $\mathbb{F}_q$ in $O(n)$ deljenj elementov iz $\mathbb{F}_q$.*

Dokaz: Največji skupni delitelj izračunamo s pomočjo Evklidovega algoritma. Naj imata deljenec in delitelj na i-tem koraku Evklidovega stopnji po vrsti $d_i > e_i$. Za i-ti korak Evklidovega algoritma (polinomsko deljenja i-tega deljenca z i-tim deliteljem) tako porabimo eno deljenje v $\mathbb{F}_q$ in $O\Big(e_i(d_i - e_i)\Big) < O\Big(n(d_i - e_i)\Big)$ množenj v $\mathbb{F}_q$. Naj r označuje število korakov za celoten Evklidov algoritem. Ker je

$$\sum_{i=1}^{r}(d_i - e_i) = d_1 - e_r < d_1 < n,$$

potrebujemo za izračun največjega skupnega delitelja $O(n^2)$ množenj v obsegu $\mathbb{F}_q$. Na vsakem koraku Evklidovega algoritma se stopnja deljenca zmanjša vsaj za ena, zato je $r \leq n$. Za celoten algoritem je tako potrebno reda $O(n)$ deljenj. ■

Naslednji dve trditvi izmerita časovno zahtevnost dveh najpogostejših aritmetičnih operacij v Schoofovem algoritmu.

Trditev 4.21 *Naj bo ℓ praštevilo, f in g pa poljubna polinoma iz kolobarja $\mathbb{F}_q[X]$ s stopnjo manjšo od stopnje polinoma ψ_ℓ. Potem za izračun polinoma $fg \mod \psi_\ell$ potrebujemo $O(\ell^4)$ množenj in $O(\ell^2)$ deljenj v obsegu $\mathbb{F}_q$.*

Dokaz: Vemo, da je $\deg \psi_\ell = (\ell^2 - 1)/2$, torej je $\deg \psi = O(\ell^2)$. Po lemi (4.19) lahko to operacijo izvedemo z $O(\ell^4)$ množenji in $O(\ell^2)$ deljenji v obsegu $\mathbb{F}_q$. ■

Trditev 4.22 *Naj bosta $f, g \in \mathbb{F}_q[X,Y] \mod E \mod \psi_\ell$. Torej imata f in g obliko $f = f_1 + f_2Y$ in $g = g_1 + g_2Y$, kjer so $f_i, g_i \in \mathbb{F}_q[X] \mod \psi_\ell$. Potem za izračun polinoma $fg \mod \psi_\ell$ potrebujemo $O(\ell^4)$ množenj in $O(\ell^2)$ deljenj v obsegu $\mathbb{F}_q$.*

Dokaz: Zapišimo

$$fg = f_1g_1 + f_2g_2(X^3 + a_2X^2 + a_4X + a_6) + \Big(f_1g_2 + f_2g_1 - f_2g_2(a_1X + a_3)\Big)Y.$$

Torej moramo izvesti natanko 7 produktov dveh polinomov iz $\mathbb{F}_q[X] \mod \psi_\ell$. Po trditvi (4.21) torej skupno porabimo $O(\ell^4)$ množenj in $O(\ell^2)$ deljenj, vse seveda v obsegu $\mathbb{F}_q$. ■

Pripomnimo, da sta časovni zahtevnosti za izračun polinomov $f + g$ in αf, kjer je $\alpha \in \mathbb{F}_q$, v obeh zgornjih primerih dominirani s časovno zahtevnostjo izračuna polinoma fg. Natančneje, za izračun polinoma $f + g$ v nobenem primeru ne potrebujemo množenj in deljenj, za izračun polinoma αf pa potrebujemo $O(\ell^2)$ množenj in nobenega deljenja.

4.8 Ocena časovne in prostorske zahtevnosti Schoofovega algoritma

V tem razdelku bomo naredili natančno analizo časovne in prostorske zahtevnosti Schoofovega algoritma. Ocenimo najprej, kako velika je množica praštevil $\mathcal{L}$. Pokazati je možno [Enge, str. 136], da je velikost največjega praštevila v $\mathcal{L}$ pri opisani izbiri $\mathcal{L}$ reda $O(\log q)$, zato so vsa praštevila iz $\mathcal{L}$ tudi reda velikosti $O(\log q)$. Moč množice $\mathcal{L}$ je tako reda

$$O\left(\frac{\log q}{\log\log q}\right) < O(\log q).$$

Gradnja množice $\mathcal{L}$ zahteva reda $O(\log q)$ mnogomestnih množenj števil velikosti $O(q)$, kar skupaj znese $O(\log^3 q)$ bitnih operacij.

Naslednji korak predstavljata izračun in shramba delitvenih polinomov $\psi_0, \ldots, \psi_{\ell_{\max}}$. Ti polinomi morajo biti v algoritmu v nadaljevanju ves čas na voljo. Najprej premislimo, koliko prostora potrebujemo za njihovo shrambo. Stopnja delitvenega polinoma ψ_n je $O(n^2)$. Shraniti moramo $1+\ell_{\max} = O(\log q)$ polinomov stopnje reda $O(\ell_{\max}^2) = O(\log^2 q)$, torej skupno reda $O(\log^3 q)$ elementov obsega $\mathbb{F}_q$. To skupaj znese reda $O(\log^4 q)$ bitov prostora. Delitvene polinome izračunamo preko rekurzivnih formul, zato za izračun vsakega od njih porabimo konstanto število polinomskih množenj in seštevanj. Delitveni polinomi imajo stopnjo reda $O(\log^2 q)$, zato za izračun vsakega delitvenega polinoma posebej po lemi (4.18) porabimo $O(\log^4 q)$ operacij v obsegu, oziroma $O(\log^6 q)$ bitnih operacij. Vseh potrebnih delitvenih polinomov je $O(\log q)$, torej za vse skupaj porabimo $O(\log^7 q)$ bitnih operacij.

Vse operacije, ki jih bomo opisali v nadaljevanju, moramo ponoviti za vsak $\ell \in \mathcal{L}$. Najprej moramo določiti število q_ℓ, kar zahteva le eno mnogomestno deljenje, torej $O(\log^2 q)$ bitnih operacij. Sledi izračun polinomov $X^q, X^{q^2}, Y^q, Y^{q^2} \bmod E \bmod \psi_\ell$. Vsakega od teh polinomov lahko z algoritmom kvadriraj in množi določimo z $O(\log q)$ operacijami v kolobarju $\mathbb{F}_q[X] \bmod \psi_\ell \bmod E$. Po lemi (4.21) posamezna taka operacija zahteva $O(\log^6 q)$ bitnih operacij. Skupno torej za izračun vseh štirih polinomov za posamezen ℓ potrebujemo $O(\log^7 q)$ bitnih operacij. Za njihovo shrambo pa potrebujemo $O(\log^2 q)$ bitov prostora. Sledi preverjanje pogoja, če je ℓ tipa C ali ne. Pri tem najprej porabimo $O(\log^6 q)$ bitnih operacij za izračun polinoma iz trditve (4.11), potem pa še $O(\log^6 q)$ bitnih operacij za izračun največjega skupnega delitelja (gcd) tega polinoma in ψ_ℓ. Izračun največjega skupnega delitelja dveh polinomov stopnje $O(\log^2 q)$ namreč po lemi (4.20) terja $O(\log^4 q)$ elementarnih operacij v obsegu $\mathbb{F}_q$. Za odločitev, ali je ℓ tipa C ali ne, torej porabimo $O(\log^6 q)$ bitnih operacij.

Če ne nastopi primer C, moramo najprej izračunati Jacobijev simbol $\left(\frac{q_\ell}{\ell}\right)$. Za to potrebujemo $O(\log^2 \ell) = O(\log^2(\log q))$ bitnih operacij [Stinson, str. 134]. Za iskanje kvadratnega korena ω z metodo grobe sile porabimo $O(l \log^2 \ell) = O(\log q \log^2(\log q))$ bitnih operacij. Sledi preverjanje pogojev iz trditev (4.14) in (4.15), ki imata enako časovno zahtevnost kot preverjanje pogoja iz trditve (4.11), ki smo ga opravili zgoraj.

Če pa nastopi primer C, moramo najprej določiti polinome $\psi_0^q, \ldots, \psi_{\ell-1}^q \bmod \psi_\ell$. Za določitev vsakega od njih potrebujemo konstanto število množenj polinomov po modulu ψ_ℓ, torej $O(\log^4 q)$ elementarnih operacij v obsegu $\mathbb{F}_q$. Skupno za njihovo določitev torej potrebujemo $O(\log^7 q)$ bitnih operacij. Prostor, ki ga zasedejo, pa je $O(\log^4 q)$ bitov. V nadaljevanju primera C moramo v najslabšem primeru za vsak τ izvesti X-test in enkrat še Y-test. Za posamezen X-test moramo določiti polinom h_X in preveriti, če je deljiv s ψ_ℓ. Za to potrebujemo konstantno število operacij v $\mathbb{F}_q[x] \bmod \psi_\ell$, torej $O(\log^6 q)$ bitnih operacij. Časovna zahtevnost Y-testa je enaka. V najslabšem primeru, ko moramo preveriti vse τ, to pomeni $O(\log^7 q)$ bitnih operacij.

Za posamezen ℓ torej potrebujemo $O(\log^7 q)$ bitnih operacij in $O(\log^4 q)$ bitov veliko skladišče, ki ga lahko izpraznimo po vsakem ℓ. Za obravnavo vseh $\ell \in \mathcal{L}$ potemtakem potrebujemo $O(\log^8 q)$ bitnih operacij in $O(\log^4 q)$ bitov veliko skladišče. Takšna je kompleksnost glavnega dela Schoofovega algoritma.

Ocenimo še zahtevnost zadnjega dela Schoofovega algoritma, torej izvedbo standardnega postopka reševanja sistema kongruenc $t \equiv t_\ell \bmod \ell$, $\ell \in \mathcal{L}$ [Stinson, str. 119]. Za to potrebujemo $O(|\mathcal{L}|^2)$ množenj in $O(|\mathcal{L}|)$ seštevanj naravnih števil bitne dolžine $O(\log q)$. Za vse to skupaj potrebujemo $O(\log^4 q)$ bitnih operacij. Poleg tega potrebujemo še skupno reda $O(|\mathcal{L}|)$ osnovnih operacij v končnih obsegih $\mathbb{F}_\ell$, $\ell \in \mathcal{L}$, kar je v primerjavi z $O(\log^4 q)$ zanemarljivo. Del algoritma, ki se nanaša na kitajski izrek o ostankih, je tako računsko nezahteven v primerjavi z jedrom Schoofovega algoritma.

Tako smo dokazali naslednji izrek.

Izrek 4.23 *Časovna zahtevnost Schoofovega algoritma je $O(\log^8 q)$ bitnih operacij, pri čemer ena bitna operacija pomeni izračun logičnega* XOR*-a dveh elementov iz $\{0, 1\}$. Algoritem je tudi prostorsko razmeroma zahteven, saj potrebujemo med izvajanjem algoritma $O(\log^4 q)$ bitov veliko skladišče.* ∎

Z uporabo določenih hitrih metod za množenje je možno časovno kompleksnost Schoofovega algoritma znižati na $O(\log^{5+\varepsilon} q)$ bitnih operacij za vsak $\varepsilon > 0$. Vendar je v praksi q običajno premajhen, da bi te asimptotske izboljšave prišle do izraza [Blake, str. 111,112]. Največ računanja v standardnem Schoofovem algoritmu, kot smo ga predstavili mi, zahteva izračun polinomov $X^q, X^{q^2}, Y^q, Y^{q^2} \bmod E \bmod \psi_\ell$ in X-test v primeru C. Slednji je sicer hiter, vendar ga moramo velikokrat ponoviti. Glavni problem Schoofovega algoritma je, da imajo delitveni polinomi razmeroma veliko stopnjo $O(\log^2 q)$. To Schoofov algoritem precej upočasni. Glede na današnji nivo kriptografije je verzija Schoofovega algoritma, opisana v tem delu, pogosto nesprejemljivo prepočasna. V praksi se zato uporabljajo izboljšave Schoofovega algoritma. Nekaj jih bomo navedli v naslednjem razdelku.

4.9 Kam pelje pot naprej

V tem poglavju bomo na kratko orisali dve izboljšavi Schoofovega algoritma.

4.9.1 Kombinacija Schoofovega algoritma in Shanksove metode

Za velike ℓ utegnejo v Schoofovem algoritmu postati izračuni, potrebni za določitev števila t_ℓ, preobsežni. Pomagamo si lahko tako, da s pomočjo Schoofovega algoritma ne določimo vseh števil $\{t_\ell \mid \ell \in \mathcal{L}\}$, ampak pač le tiste, ki jih zmoremo. S tem iskanega števila t nismo enolično določili, smo ga pa določili modulo neko veliko število M.

Število t zatem poskusimo natančno določiti s pomočjo Shanksove metode. Ker poznamo $t \bmod M$, so lahko majhni koraki Shanksove metode veliki M namesto 1, kot so v originalni Shanksovi metodi. Veliki koraki Shanksove metode se s tem tudi ustrezno povečajo. Zaradi vsega tega tako za prečesanje celotnega Hassejevega območja potrebujemo manj računanja kot brez informacije $t \bmod M$.

4.9.2 Schoof-Atkin-Elkiesov algoritem

Ta algoritem odpravi glavno pomankljivost Schoofovega algoritma in pri iskanju števila t_ℓ namesto delitvenih polinomov za testiranje enakosti v $E[\ell]^*$ uporablja polinome, katerih stopnja je linearna v ℓ.

Karakteristična enakost za operator φ_ℓ na vektorskem prostoru $E[\ell]$ nad $\mathbb{F}_\ell$ se glasi [Blake, str. 114]

$$\Delta_{\varphi_\ell}(X) = X^2 - t_\ell X + q_\ell = 0.$$

Za Atkin-Elkiesov algoritem je bistvenega pomena, ali ničle tega polinoma ležijo v $\mathbb{F}_\ell$ ali ne, torej, ali ima operator φ_ℓ kakšno lastno vrednost ali ne. V razdelku 1.6 o kvadratnih enačbah smo videli, da je ta pogoj je ekvivivalenten pogoju, da je diskriminanta $t^2 - 4q$ kvadratni ostanek v $\mathbb{F}_\ell$.

Definicija 4.24 *Praštevila ℓ, za katere diskriminanta $t_\ell^2 - 4q_\ell$ je kvadratni ostanek v $\mathbb{F}_\ell$, imenujemo* **Elkiesova praštevila**. *Preostala praštevila pa imenujemo* **Atkinova praštevila**.

Ti dve lastnosti sta za dano praštevilo ℓ seveda odvisni od števila q in krivulje E. Ker števila t_ℓ ne poznamo, ne moremo direktno določiti, ali je dano praštevilo ℓ Elkiesovo ali Atkinovo. Vseeno pa nam to lahko uspe po ovinku, pri katerem se poslužimo t.i. *modularnih polinomov* [Blake, str. 114]. Ta pristop nam v primeru, da je ℓ Atkinovo, da tudi neko določeno informacijo o številu t_ℓ.

Če je ℓ Elkiesovo praštevilo, poskusimo določiti neko lastno vrednost λ za operator φ_ℓ. Če nam uspe najti neko lastno vrednost λ, je potem po Viètovih formulah druga lastna vrednost enaka $\mu = q/\lambda$. Težav z deljenjem z nič pri tem nimamo, ker je φ_ℓ injektiven in torej velja $\lambda \neq 0$. S pomočjo druge Viètove formule lahko potem določimo t_ℓ. Velja namreč

$$\lambda + \mu = t_\ell, \quad \text{torej} \qquad t \equiv \lambda + \frac{q}{\lambda} \pmod{\ell}.$$

Pri iskanju lastne vrednosti iščemo točko $P = (x, y) \in E[\ell]^*$ in število $\lambda \in \{1, 2, \ldots, \ell - 1\}$, tako da velja

$$(x^q, y^q) = \lambda(x, y). \tag{4.21}$$

Če bi tako točko iskali podobno kot v Schoofovem algoritmu, nam sicer ne bi bilo treba izračunati polinomov $x^{q^2}, y^{q^2} \bmod E \bmod \psi_\ell$, vendar na asimptotski časovni zahtevnosti algoritma ne bi bistveno pridobili. Tehnike, povezane z modularnimi polinomi, nam omogočijo, da lahko za iskanje točke iz (4.21) namesto polinoma ψ_ℓ uporabimo nek faktor polinoma ψ_ℓ, ki ima stopnjo $(\ell - 1)/2$. To algoritem precej pospeši. Časovna zahtevnost Elkiesovega dela algoritma je tako $O(\log^6 q)$ bitnih operacij. Časovna zahtevnost Atkinovega dela algoritma pa je eksponentna v parametru $\log q$, vendar je Atkinov del algoritma v praksi vseeno koristen. Z uporabe hitre aritmetike v obsegu $\mathbb{F}_q$ postane časovna zahtevnost Elkiesovega dela algoritma $O(\log^{4+\varepsilon} q)$ bitnih operacij za poljuben $\varepsilon > 0$. Ocene časovne zahtevnosti Schoof-Elkies-Atkinovega algoritma so povzete po [Blake, str. 117].

V zadnji fazi Schoof-Atkin-Elkiesovega algoritma podobno kot pri Schoofovem algoritmu informacije o številu t, ki smo jih za različna praštevila zbrali bodisi v Elkiesovem primeru bodisi v Atkinovem primeru, zberemo skupaj in s pomočjo kitajskega izreka o ostankih enolično določimo t. Ta del algoritma je hiter. Tudi tukaj je seveda zopet možna kombinacija s Shanksovo metodo.

Literatura

[Vidav] *I. Vidav: Algebra*, Mladinska knjiga, 1972

[Vidav EC] *I. Vidav: Eliptične krivulje in eliptične funkcije*, Društvo matematikov, fizikov in astronomov Slovenije, 1991

[Schoof] *R.Schoof: Elliptic Curves Over Finite Fields and the Computation of Square Roots mod p*, Mathematics of Computation, **44** (1985), strani 483-494

[Enge] *A. Enge: Elliptic curves and their applications to cryptography: an elementrary introduction*, doktorska disertacija, Univerza v Augsburgu, 1997
(disertacija je bila pred kratkim izdana tudi v knjižni obliki pri založbi Kluwer Academic Publishers)

[Blake] *I. Blake, G. Seroussi, N. Smart: Elliptic curves in cryptography*, Cambridge University Press, 1999

[Lidl-Niederreiter] *R. Lidl, H. Niederreiter: Encyclopedia of Mathematics and Its Applications: Finite Fields*, Cambridge University Press, 1987

[Menezes] *J. Menezes: Applications of finite fields*, Kluwer Academic Publishers, 1993

[Stinson] *D. R. Stinson : Cryptography: Theory and Practice*, Boca Raton : CRC, 1995

[Jurišić] *A. Jurišić, A. Menezes: Elliptic Curves and Cryptography*, Dr. Dobb's Journal, **264** (1997), strani 26-36

[Silverman] *J. Silverman: The Arithmetic of Elliptic Curves*, Springer-Verlag, 1986

[Handbook] *J. Menezes, P. Oorschot, S. Vanstone: Handbook of Applied Cryptography*, CRC Press, 1997

[Fraleigh] *J. Fraleigh: A first course in abstract algebra*, Addison-Wesley, 1994