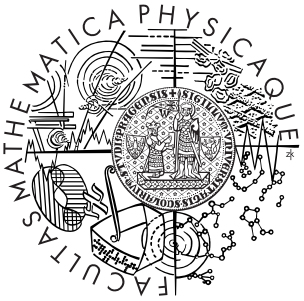
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Adéla Skoková

Podpůrné algoritmy číselného síta

Katedra algebry

Vedoucí diplomové práce:	prof. RNDr. Aleš Drápal, CSc., DSc.
Studijní program:	Matematika
Studijní obor:	Matematické metody informační bezpečnosti

Praha 2013

Děkuji vedoucímu mé diplomové práce Prof. Aleši Drápalovi za jeho vedení a cenné rady. Dále bych chtěla poděkovat svému konzultantovi Lukáši Perutkovi, který mi pomohl pochopit detaily algoritmu.

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ..... Podpis autora

Název práce: Podpůrné algoritmy číselného síta

Autor: Adéla Skoková

Katedra: Katedra algebry

Vedoucí diplomové práce: prof. RNDr. Aleš Drápal, CSc., DSc.

Abstrakt: V předložené diplomové práci studujeme hlavně první fázi algoritmu číselného síta, generování polynomů. Nejprve popisujeme celé číselné síto pro pochopení role polynomů a jejich vliv na celý algoritmus. Pak se věnujeme jejich vlastnostem a ohodnocování. Nakonec uvádíme algoritmy pro generování polynomů, se kterými přišel Thorsen Kleinjung. Jedná se o zatím nepřekonané algoritmy na získávání vhodných polynomů.

Klíčová slova: Číselné síto, GNFS, Číselné těleso, Kleinjungův algoritmus

Title: Supporting algorithms of number field sieve

Author: Adéla Skoková

Department: Department of Algebra

Supervisor: prof. RNDr. Aleš Drápal, CSc., DSc.

Abstract: In this work we study the first part of the algorithm of number field sieve, generating of polynomials. At first we describe all the algorithm of the sieve for understanding of the role of polynomials and their impact on the entire algorithm. Then we present their characteristics and evaluation. The last part is about the most effective know algorithms of generating polynomials, invented by Thorsen Klinjung.

Keywords: GNFS, Number sieve, Number field, Kleinjung algorithm

Obsah

Kapitola 1

Úvod

Rozkladu celých čísel na jejich dělitele byl vždy přikládán určitý význam. Čím větší jsou prvočísla, která dělí dané číslo, tím delší dobu rozkladu lze očekávat. Dnes máme různé rychlejší metody než zkoušení všech prvočísel postupně od nejmenších až do druhé odmocniny rozkládaného čísla, ale tyto metody nejsou stále dosti efektivní pro rozklad čísel složených pouze z prvočísel o velikosti několik stovek bitů. Taková čísla se dnes používají v asymetrické kryptografii. Celé zabezpečení systému je pak založeno pouze na nesnadnosti rozkladu zveřejněného čísla.

Dnes běžně používáme elektronický podpis nebo časové razítko, které mají časově omezenou platnost. Po vypršení této doby je asymetrické zabezpečení považováno za neplatné, protože pravděpodobnost, že došlo k prolomení veřejného klíče již překročila určitou mez.

Popišme stručně systém RSA, který je jednou z hlavních metod asymetrické kryptografie. Skládá se ze dvou klíčů: veřejného a soukromého. Ve veřejném klíči se uvádí celé číslo N a veřejný exponent e. Zakódování zprávy pak vypadá tak, že zprávu m umocníme veřejným exponentem e a počítáme modulo N. Získat zpět původní zprávu můžeme pomocí opačného exponentu k e, který se často značí d. Ten získáme pomocí Eulerovy funkce čísla N, tedy pomocí rozkladu čísla N na jeho prvočíselné dělitele, v případě RSA se jedná o dvě velká prvočísla p a q.

Zřejmě stačí „pouze“ rozložit číslo N na prvočíselné dělitele a dopočítat potřebný opačný exponent, abychom získali původní zprávu m.

Asymetrická kryptografie by neměla smysl, kdyby rozhodnutí, zda dané číslo je složené, bylo řádově stejně rychlé, jako nalezení jeho rozkladu na prvočíselné dělitele. Pro RSA dnes uvažujeme číslo N běžně o velikostech mezi 1024 bity až 4096 bity. Taková čísla zatím nejsme schopni rozložit i pomocí nejsilnějších algo-

ritmů dříve než za desítky let. Díky novým technologiím se ale tyto postupy stále urychlují. Co se zdálo nerozložitelné před deseti lety, není už dnes považováno za úplně bezpečné. Například Laboratoře RSA vyhlásily roku 1991 výzvu o rozložení některého ze seznamu prvočísel. Roku 2007 tuto soutěž ukončily se slovy, že dnešní znalosti kryptografie jsou mnohem dále [14]. Zatím největší rozložené číslo z této soutěže RSA-768 se podařilo skupině lidí okolo Thorstena Kleinjunga [9]. K této faktorizaci použili algoritmu, který se nazývá číselné síto.

Algoritmus, který česky nazýváme číselné síto, nebo také metoda síta nad číselným tělesem, je v současné době nejsilnějším nástrojem pro faktorizaci čísel běžně používaných jako veřejné klíče pro RSA. V této práci se budeme věnovat algoritmům, které číselné síto využívá hlavně v první fázi. Nejprve předestřeme matematický aparát potřebný k pochopení dále popsaných algoritmů. Pak stručně popíšeme celý algoritmus a následně rozebereme algoritmy okolo takzvané první fáze.

Kapitola 2

Potřebné pojmy obecné komutativní algebry

V celém textu budeme vycházet ze znalosti základní obecné algebry. V této kapitole stručně shrneme několik pojmů a poznatků z obecné komutativní algebry, o které se opírá algoritmus číselného síta. Nebudeme uvádět všechny důkazy. Pouze takové, které mají vypovídající hodnotu k tématu práce. Neuvedené důkazy lze nalézt například ve skriptech o komutativních okruzích [11].

Okruhem budeme vždy myslet komutativní okruh s jednotkou. Tělesa budeme rovněž uvažovat pouze komutativní a obor (integrity) definujeme jako komutativní okruh bez dělitelů nuly ($ab = 0 \rightarrow a = 0 \vee b = 0$). Každé komutativní těleso je oborem integrity. V tomto textu nebudeme uvažovat triviální obory, kdy jednotka a nula splývají.

2.1 Základní pojmy

Rozšířením těles $T \subseteq U$ rozumíme dvojici těles, kde těleso U obsahuje těleso T. Analogicky definujeme i rozšíření okruhů.

Mějme $T \subseteq U$, $T \subseteq W$ dvojice rozšíření těles. Pak homomorfizmus z U do W, který je identický na T, nazýváme T-**homomorfizmus**. Množinu všech T-homomorfizmů z U do W budeme dále značit $\hom_T(U, W)$.

Nechť $T \subseteq U$ je rozšíření těles. Prvek $\alpha \in U$ nazýváme **algebraický nad** T, je-li kořenem nějakého polynomu z $T[x]$.

Nechť $R \subseteq S$ je rozšíření okruhů a prvek $\alpha \in S$. Všechny polynomy $f \in R[x]$ s kořenem α tvoří v $R[x]$ ideál. Tento ideál je hlavní, je-li R těleso. Generátor tohoto ideálu v nazýváme **minimální polynom** prvku α. Je-li R těleso, za generátor považujeme monický polynom. Budeme ho dále značit $f_{\alpha,R}$. Označení budeme zjednodušovat na f_α, pokud bude zřejmé, o jaké R se jedná.

Rozšíření těles $T \subseteq U$ je **algebraické**, pokud je každý prvek z U algebraický nad T. Množina $\{\alpha \in U | \alpha$ je algebraický nad $T\}$ se nazývá **algebraický uzávěr** tělesa T v U. Těleso je **algebraicky uzavřené**, pokud nemá vlastní algebraické rozšíření. **Algebraické číslo** je každé komplexní číslo algebraické nad $\mathbb{Q}$.

Poznámka 2.1. *Nechť $T \subseteq U$ je rozšíření těles a ať V je algebraický uzávěr T v U.*

- *Algebraické rozšíření tělesa T, které je algebraicky uzavřené v U, je právě těleso V.*
- *Algebraický uzávěr T v U je podtěleso U.*
- *Algebraický uzávěr T v U je jednoznačný až na T-isomorfizmus.*
- *Pro algebraické rozšíření T' tělesa T a těleso K nad T algebraicky uzavřené existuje vnoření T' do K. Současně pokud máme $T \subseteq T' \subseteq K$, tak každý homomorfismus z T' do K lze rozšířit na automorfismus K.*
- *Běžně značíme $\overline{T}$ algebraický uzávěr tělesa T.*

Například platí, že algebraický uzávěr tělesa $\mathbb{R}$ je $\mathbb{C}$. Těleso $\mathbb{C}$ je algebraicky uzavřené. Tento fakt je nazývaný Fundamentální věta algebry. Algebraický uzávěr tělesa $\mathbb{Q}$ není těleso $\mathbb{C}$. Nejedná se totiž o algebraické rozšíření.

Je-li $R \subseteq S$ rozšíření okruhů a $M \subseteq S$, pak $R[M]$ označuje nejmenší podokruh S, který obsahuje $R \cup M$. Je-li $T \subseteq U$ rozšíření tělesa a $M \subseteq U$, tak $T(M)$ značí nejmenší podtěleso U obsahující $T \cup M$. Mějme pro $\alpha \in U$. Připoměňmě strukturu $T[\alpha] = \{g(\alpha) | g \in T[x]\}$. Je dobře známo, že α je algebraický prvek nad T, právě když $T[\alpha] = T(\alpha)$. V takovém případě

$$T[\alpha] \cong T[x]/(f_\alpha).$$

Nechť $T \subseteq U$ je rozšíření těles. Algebraické rozšíření $T \subseteq T(\alpha)$ nazýváme **jednoduché**, pokud existuje $\alpha \in U$ takové, že stupeň jednoduchého rozšíření $[T(\alpha) : T] = \deg f_{\alpha,T}$.

Uvažme multiplikativní zobrazení $\mu_r : M \to M$, definované $\mu_r(a) = ra$. Pak $M(+)$ je modul, pokud je na něm definováno skalární násobení takové, že $r \longmapsto \mu_r$ je homomorfismem okruhů $R \to \mathrm{End}(M(+))$. Jádro tohoto homomorfismu označíme $\mathrm{Ann}_R(M)$ a nazveme **anihilátor** R-modulu M. Modul M se nazývá **beztorzní**, pokud každé μ_r je pro $r \neq 0$ injektivní. Tento pojem má tedy smysl uvažovat pouze pro obory integrity. Zřejmě $\mathbb{Z}$-moduly jsou totéž jako abelovské grupy.

Připomeňme, že okruh je konečně generovaný, jestliže je generován nějakou konečnou množinou prvků.

Definice 2.2. *Podmnožina X generující R-modulu M se nazývá **volná báze**, pokud pro každý modul M' a každý výběr prvků $a_x \in M'$, kde $x \in X$, existuje homomorfismus $\varphi : M \to M'$ takový, že $\varphi(x) = a_x$ pro každé $x \in X$.*

***Volný modul** je takový R-modul, pro který lze nalézt alespoň jednu volnou bázi.*

Tvrzení 2.3. *Podmoduly volného modulu konečné hodnosti nad obory hlavních ideálů jsou volné.*

Tvrzení 2.4. *Konečně generovaný R-modul nad oborem hlavních ideálů je beztorzní právě tehdy, když je volný.*

Připomeňme například, že $\mathbb{Z}$-modul je konečně generován, pokud ho můžeme zapsat jako lineární kombinaci konečné báze nad $\mathbb{Z}$.

Definice 2.5. *R-Modul M nazýváme **věrný**, pokud* $\mathrm{Ann}_R(M) = 0$.

Definice 2.6. *Mějme $T \subseteq U \subseteq W$ algebraická rozšíření těles, kde W je algebraicky uzavřené. Mohutnost množiny* $\hom_T(U, W)$ *nazveme **stupeň separability** U nad T a označíme $[U : T]_S$.*

Připomeňme, že **rozkladové nadtěleso** polynomu f z $T[x]$ se rozumí nejmenší nadtěleso tělesa T, ve kterém lze polynom f rozložit na součin polynomů stupně jedna. Rozkladové nadtěleso je vždy určeno jednoznačně až na T-izomorfismus.

Definice 2.7. *Polynom $f \in T[x]$ nazýváme **separabilní polynom**, nemá-li ve svém rozkladovém nadtělese vícenásobné kořeny.*

*Rozšíření $T \subseteq U$ je **separabilním rozšířením**, má-li každý prvek $\alpha \in U$ separabilní minimální polynom $f_{\alpha,T}$.*

Tvrzení 2.8. *Nechť $T \subseteq U$ je rozšíření těles konečného stupně. Pak je ekvivalentní*

- $[U : T]_S = [U : T]$,
- $T \subseteq U$ *je separabilní,*
- $U = T[\alpha_1, \ldots, \alpha_k]$ *pro* $\alpha_1, \ldots, \alpha_k \in U$ *separabilní.*

Nechť $T \subseteq U$ je rozšíření těles. Je-li těleso T charakteristiky 0, je každý ireducibilní polynom $f(x) \in T[x]$ separabilní. Tedy lze ve svém rozkladovém nadtělese rozložit právě na $\deg(f)$ různých lineárních členů. Pro těleso T charakteristiky p je ireducibilní polynom $f(x) \in T[x]$ separabilní, právě když není tvaru $f(x) = g(x^p)$ pro nějaký polynom $g(x) \in T[x]$. Prvky $\alpha, \beta \in U$ nazýváme **konjugované** nad T pokud existuje polynom $f(x) \in T[x]$ ireducibilní nad T takový, že α a β jsou jeho kořeny.

Nechť $T \subseteq U$ je rozšíření těles. Nechť $\alpha \in U$ je kořenem $f(x) \in T[x]$. Rozklad polynomu na lineární členy v $\overline{U}$ lze zapsat ve tvaru

$$f(x) = \prod_{\sigma \in \hom_T(U, \overline{T})} (x - \sigma(\alpha)).$$

Hlavní tvrzení 2.9. *Separabilní rozšíření konečného stupně je jednoduché.*

2.2 Celistvost a číselná tělesa

Toto téma podáme v úplnosti včetně důkazů, protože z něj vychází podstatná část teorie pro algoritmus. Uvedené důkazy mají vypovídající hodnotu pro téma práce a pro pochopení základní struktury, se kterou algoritmus dále pracuje.

Definice 2.10. *Mějme $R \subseteq S$ rozšíření okruhů. Prvek $r \in S$ nazýváme **celistvým** nad R, pokud je kořenem nějakého monického polynomu $f(x) \in R[x]$.*

*Okruh S nazýváme **celistvý** nad R, pokud je každý prvek z S celistvý nad R.*

Pojem celistvého prvku nad tělesem je analogický pojmu algebraického prvku nad tělesem. Všimněme si, že jsou-li $R = T$ a $S = U$ tělesa, je prvek $\alpha \in U$ celistvý nad T právě když je algebraický nad T.

Definice 2.11. *Nechť $T \subseteq U$ je rozšíření těles a S okruh prvků z tělesa U a $R = S \cap T$. Báze $\alpha_1, \ldots, \alpha_n \in S$ tělesa U nad T se nazývá **celistvá**, jestliže každý prvek $s \in S$ lze vyjádřit ve tvaru $s = \sum r_i \alpha_i$, kde $r_i \in R$.*

Poznámka 2.12. *Rozšiřme přirozeně pojem adjungované matice i pro práci v okruzích. Uvažujme čtvercovou matici* $\mathbf{C}$ *tvaru* $n \times n$*. Adjungovaná matice k matici* $\mathbf{C}$ *je tvaru* $\{(-1)^{i+j} \det \bar{\mathbf{C}}_{ji}\}$*, kde* $\mathbf{C}_{ji}$ *je matice* $\mathbf{C}$*, ze které jsme vypustili* j*-tý řádek a* i*-tý sloupec.*

Je-li matice $\mathbf{D}$ *adjungovaná k matici* $\mathbf{C}$*, pak* $\mathbf{CD} = \mathbf{DC} = (\det \mathbf{C})I$*, kde* I *je jednotková matice tvaru* $n \times n$*.*

Tvrzení 2.13. *Mějme* $R \subseteq S$ *rozšíření okruhů a* I *ideál okruhu* R*. Dále mějme* S*-modul* M*, konečně generovaný* n *prvky* $g_1, \ldots, g_n$*.*

Nechť platí pro prvek $s \in S$*, že* $sM \subseteq IM$*. Pak existují prvky* $a_i \in I^{n-i}$*, kde* $i \in \{0, \ldots, n-1\}$ *takové, že*

$$s^n + a_{n-1}s^{n-1} + \ldots + a_1 s + a_0 \in \mathrm{Ann}_R(M).$$

Důkaz. Mějme translaci indukovanou prvkem $s \in S$.Podle předpokladů víme, že pro generátory $g_1, \ldots, g_n$ existují hodnoty b_{ij} pro $i, j = 1, \ldots, n$, které určují translaci daným prvkem. Sestavme tyto prvky do matice $\mathbf{B} = \{b_{ij}\}_{i,j=1}^n$. Tato matice vyjadřuje translaci chápanou jako endomorfismus R-modulu. Pak zřejmě platí

$$sg_i = \sum_{j=1}^{n} b_{ij} g_j \in IM.$$

Položme $\mathbf{C} = s\mathbf{I} - \mathbf{B}$, kde $\mathbf{I}$ je jednotková matice o velikosti $n \times n$. Uvažme vektor generátorů $\mathbf{g} = (g_1, \ldots, g_n)$.

$$\mathbf{C}\,\mathbf{g}^T = (sg_1, \ldots, sg_n)^T - (\sum_{j=1}^{n} b_{1j}g_j, \ldots, \sum_{j=1}^{n} b_{nj}g_j)^T = (0, \ldots, 0)^T.$$

Matice $\mathbf{C}$ je zřejmě regulární čtvercová matice. Buď $\mathbf{D}$ matice adjungovaná k matici $\mathbf{C}$. Pak $\mathbf{D}\,\mathbf{C} = (\det \mathbf{C})\,\mathbf{I}$. Tím získáme vztah

$$0 = \mathbf{D}\,\mathbf{C}\,\mathbf{g}^T = (\det \mathbf{C})\,\mathbf{I}\,\mathbf{g}^T = ((\det \mathbf{C})\,g_1, \ldots, (\det \mathbf{C})\,g_n)^T.$$

Každý koeficient vektoru na pravé straně je roven nule. Dále pro libovolný prvek $m \in M$ existují prvky $r_1, \ldots, r_n \in R$ tak, že

$$m = \sum_{j=1}^{n} r_j g_j,$$

$$(\det \mathbf{C})\, m = (\det \mathbf{C}) \sum_{j=1}^{n} r_j g_j = \sum_{j=1}^{n} r_j ((\det \mathbf{C})\, g_j) = 0.$$

To znamená, že $\det \mathbf{C} \in \mathrm{Ann}_R(M)$ a že $\det(x\mathbf{I} - \mathbf{B}) = f(x)$ je monický polynom stupně n s koeficienty z ideálu okruhu R. Tedy $f(s) \in \mathrm{Ann}_R(M)$. □

Následuje tvrzení, které vypovídá o podstatné vlastnosti celistvých prvků. Z tohoto tvrzení vycházíme při důkazech dalších důležitých vlastností celistvých prvků.

Tvrzení 2.14. *Mějme $R \subseteq S$ rozšíření okruhů a prvek $s \in S$. Pak je ekvivalentní:*

1. *s je celistvé nad R;*
2. *$R[s]$ je konečně generované jako R-modul;*
3. *existuje podokruh S' okruhu S takový, že $R[s] \leq S'$ a S' je konečně generované jako R-modul;*
4. *existuje věrný $R[s]$-modul, který je konečně generovaný jako R-modul.*

Důkaz. 1. ⇒ 2. Mějme $s \in R$ a monický polynom $f \in R[x]$ stupně d takový, že $f(s) = 0$. Potom $R[s] \cong R[x]/(f)$ je generován prvky $1, s, \ldots, s^{n-1}$.

2. ⇒ 3. ⇒ 4. Ze sebe přímo plynou. Stačí brát za uvažovaný podmodul $R[s]$ a podokruh S, který obsahuje $R[s]$ je věrný $R[s]$-modul.

4. ⇒ 1. Ať M je $R[s]$-modul, který je konečně generovaný jako R-modul. Pak existuje monický polynom $f \in R[x]$ takový, že $f(s)M = 0$, podle předchozího tvrzení 2.13. Pokud je M navíc věrný, pak podle definice platí $\mathrm{Ann}(M) = 0$ a tedy $f(s)M = 0 \Rightarrow f(s) = 0$, což znamená, že s je celistvý prvek nad R. □

Důsledek 2.15. *Mějme $R \subseteq S$ rozšíření okruhů a prvky $u_1, \ldots, u_n \in S$ celistvé nad R. Pak $R[u_1, \ldots, u_n]$ je celistvé nad R a navíc je jako R-modul konečně generovaný.*

Důkaz. Mějme prvek $s \in R[u_1, \ldots, u_n]$. Pak zřejmě $R[s] \subseteq R[u_1, \ldots, u_n]$ a je podle tvrzení 2.14 konečně generovaný jako R-modul. Podle předchozího tvrzení stačí ukázat, že $R[u_1, \ldots, u_n]$ je jako R-modul konečně generovaný a tedy je věrným $R[s]$-modulem. Postupujme indukcí podle n.

Pro $n = 1$ podle tvrzení 2.14 případu *1. ⇔ 2.* je $R[u_1]$ konečně generovaný R-modul.

Nechť platí indukční předpoklad, že $R' = R[u_1, \ldots, u_{n-1}]$ je konečně generovaný R-modul s generátory $a_1, \ldots, a_k$. Pak $R'[u_n]$ je konečně generovaný jako R'-modul prvky $b_1, \ldots, b_l$, protože prvek u_n je celistvý nad R a tedy i nad R'. Z toho plyne, že $R'[u_n]$ je konečně generovaný jako R-modul prvky $a_1 b_1, a_1 b_2, \ldots, a_k b_l$. □

Důsledek 2.16. *Mějme $R \subseteq S$ rozšíření okruhů. Pak množina všech prvků z okruhu S, které jsou celistvé nad R, tvoří podokruh S.*

Definice 2.17. *Podokruh okruhu S složený ze všech celistvých prvků nad R nazýváme* ***celistvý uzávěr*** *R v S.*

Obor R je ***celistvě uzavřený****, pokud je roven svému celistvému uzávěru ve svém podílovém nadtělese.*

Příklad 2.18. *Okruh $\mathbb{Z}$ je celistvě uzavřený, protože ve svém podílovém nadtělese $\mathbb{Q}$ již nemá žádné další prvky, které by byly celistvé nad $\mathbb{Z}$.*

Nyní se budeme zabývat koeficienty minimálních polynomů.

Tvrzení 2.19. *Mějme rozšíření $R \subseteq T \subseteq U$, kde T a U jsou tělesa a R je okruh. Pro libovolné $\alpha \in U$ celistvé nad R jsou koeficienty jeho minimálního polynomu $f_{\alpha,T}$ celistvé nad R.*

Důkaz. Mějme $U \subseteq W$ rozšíření těles, kde W je rozkladové nadtěleso polynomu $f_{\alpha,T}$. Minimální polynom $f_{\alpha,T}$ je ireducibilní nad T.

Mějme monický polynom $g \in R[x]$, který má kořen α. Pak zřejmě $f_{\alpha,T}|g$. Máme-li libovolný prvek $\beta \in W$, který je kořenem $f_{\alpha,T}$, pak je také kořenem g a je celistvý nad R. Z toho plyne, že všechny kořeny polynomu $f_{\alpha,T}$ v W jsou celistvé nad R. Koeficienty polynomu $f_{\alpha,T}$ leží v okruhu generovaném jeho kořeny a tedy jsou celistvé nad R. □

Tvrzení 2.20. *Nechť $T \subseteq U$ je rozšíření těles. Dále mějme R celistvě uzavřený obor integrity, který má podílové těleso T. Pokud je $\alpha \in U$ celistvé nad R, potom minimální polynom $f_{\alpha,T}$ leží v $R[x]$.*

Důkaz. Podle předchozího tvrzení má polynom $f_{\alpha,T}$ s koeficienty celistvá nad R. Obor R je celistvě uzavřený a tedy $f_{\alpha,T} \in R[x]$. □

Tvrzení 2.21. *Nechť $T \subseteq U$ je rozšíření těles. Buď T podílové těleso okruhu R a $\alpha \in U$ algebraické nad T. Pak existuje $r \in R$ takové, že $r\alpha$ je celistvé nad T.*

Důkaz. Buď $f_{\alpha,T}(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0$. Protože T je podílové těleso okruhu R, lze uvažovat $a_i = \frac{b_i}{c_i}$, pro $b_i, c_i \in R$, kde $i = 0, \ldots, n-1$. Položme $r = \prod_{i=0}^{n-1} c_i$. Prvek $r\alpha \in S$ je celistvý nad R, protože

$$\begin{aligned} r^d f(\alpha) &= r^d\alpha^d + r^d a_{d-1}\alpha^{d-1} + \ldots + r^d a_1\alpha + r^d a_0 \\ &= (r\alpha)^d + r a_{d-1}(r\alpha)^{d-1} + \ldots + r^{d-1} a_1(r\alpha) + r^d a_0. \end{aligned}$$

□

Důsledek 2.22. *Pro libovolnou n-tici prvků $\alpha_1, \ldots, \alpha_n \in U$ algebraických nad T existuje nenulový prvek $r \in T$ tak, že $r\alpha_1, \ldots, r\alpha_n$ jsou celistvé nad T.*

Důkaz. Podle tvrzení 2.21 najdeme pro každé $\alpha_i \in U$ jeho $r_i \in R$. Společné r pro všechna α_i je tvaru $r = r_1 \ldots r_n$. □

Příklad 2.23. *Mějme okruh $R = \mathbb{Z}$, těleso $T = \mathbb{Q}$ jeho algebraické rozšíření $U = \mathbb{Q}[i]$. Celistvý uzávěr $\mathbb{Z}$ v $\mathbb{Q}[i]$ je tvaru $S = \mathbb{Z}[i]$. Podle tvrzení 2.21 pro libovolný prvek $\alpha \in \mathbb{Q}[i]$ existuje jeho celočíselný násobek, který patří do $\mathbb{Z}[i]$.*

Tento příklad je zřejmý. Tvrzení 2.21 však platí pro obecné těleso U, tedy nejen pro případy racionálních a celých čísel.

Definice 2.24. ***Algebraickým celým číslem*** *značíme prvek z $\mathbb{C}$, který je celistvý nad $\mathbb{Z}$.*

Tvrzení 2.25. *Množina algebraických celých čísel tvoří podokruh komplexních čísel. Označme ho $\boldsymbol{A}$.*

Důkaz. Jedná se o speciální případ důsledku 2.16, kdy $S = \mathbb{C}$ a $R = \mathbb{Z}$. □

Příklad 2.26. *Aplikujme tvrzení 2.20 na příkladu algebraických celých čísel. Nechť $R = \mathbb{Z}$, $T = \mathbb{Q}$ a $U = \mathbb{C}$. Pro celistvý prvek $\alpha \in \mathbb{C}$ platí, že jeho minimální polynom f_α má koeficienty v $\mathbb{Z}$. Má-li $\alpha \in \mathbb{C}$ minimální polynom tvaru $f_\alpha \in \mathbb{Z}[x]$, pak se podle definice jedná o algebraické celé číslo. Tedy prvek $\alpha \in \mathbb{C}$ je algebraickým celým číslem, právě když je celistvý nad $\mathbb{Z}$.*

Definice 2.27. ***Číselné těleso*** *K (někdy zvané algebraické číselné těleso) je každé nadtěleso konečného stupně tělesa racionálních čísel $\mathbb{Q}$.*

Stupněm číselného tělesa *$[K : \mathbb{Q}]$, rozumíme stupeň rozšíření K nad $\mathbb{Q}$.*

Bývá zvykem předpokládat navíc, že číselné těleso je podtělesem $\mathbb{C}$. Touto konvencí se budeme dále řídit.

Zřejmě každé číselné těleso je nekonečné a je charakteristiky 0. Prvky číselného tělesa jsou algebraická čísla, protože se jedná o konečné, a tedy algebraické rozšíření tělesa $\mathbb{Q}$. Každé číselné těleso je podtělesem tělesa všech algebraických komplexních čísel.

Naopak ale neplatí, že by každé algebraické rozšíření racionálních čísel bylo číselným tělesem. Například těleso všech algebraických čísel je algebraickým rozšířením racionálních čísel, ale není číselným tělesem, protože není konečného stupně.

Číselné těleso K je separabilní rozšíření $\mathbb{Q}$, protože je podle definice konečné. Stupeň číselného tělesa je roven stupni separability podle tvrzení 2.8. Navíc je každé číselné těleso K je jednoduché rozšíření $\mathbb{Q}$ podle tvrzení 2.9.

Všechny prvky číselného tělesa jsou algebraická čísla. Není však pravda, že by všechny tyto prvky byly algebraickými celými čísly.

Určeme množinu $O_K = \{\alpha \in K; \alpha \text{ je celistvé nad } \mathbb{Q}\}$. Zřejmě je O_K celistvý uzávěr $\mathbb{Z}$ v číselném tělese K. Jedná se o podmnožinu okruhu všech algebraických celých čísel, $O_K \subseteq \mathbf{A}$. Podle důsledku 2.16, kdy $R = \mathbb{Z}$ a $S = K$, je O_K okruh.

Definice 2.28. *Okruh O_K budeme nazývat* ***okruh celých algebraických čísel*** *číselného tělesa K.*

Okruh O_K je uzavřený na operace sčítání a násobení. Tedy součet a součin algebraických celých čísel z O_K je opět algebraické celé číslo z O_K.

Podle tvrzení 2.20 mají všechny prvky z O_K minimální polynom s celočíselnými koeficienty. Dále víme, že podle tvrzení 2.21 lze pro libovolný prvek K najít takový celočíselný násobek, který je prvkem O_K. Z toho okamžitě plyne, že $\mathbb{Q}O_K = K$. Také platí $O_K \cap \mathbb{Q} = \mathbb{Z}$, protože pro $K_1 \subseteq K_2$ rozšíření číselných těles máme $O_{K_1} = O_{K_2} \cap K_1$.

Podstatný je tvar okruhu algebraických celých čísel O_K. Vzhledem k tomu, že pro $K = \mathbb{Q}$ získáváme $O_K = \mathbb{Z}$, bylo by intuitivní předpokládat, že pro $K = \mathbb{Q}[\alpha]$ je $O_K = \mathbb{Z}[\alpha]$, ale tak tomu vždy není. Například pro $K = \mathbb{Q}[i\sqrt{3}]$ platí $\mathbb{Z}[i\sqrt{3}] \subsetneq O_K$, protože prvek $\frac{1+i\sqrt{3}}{2} \notin \mathbb{Z}[\sqrt{3}]$ patří do O_K a je kořenem monického polynomu

$$g(x) = x^2 - x + 1 = \frac{1}{4}\left((2x-1)^2 + 3\right).$$

Je zřejmé, že platí $\mathbb{Z}[\alpha] \subseteq O_K$. Navíc pro f monický je okruh $\mathbb{Z}[\alpha]$ celistvý nad $\mathbb{Z}$. Při výpočtech se budeme však převážně pohybovat právě v $\mathbb{Z}[\alpha]$.

Příklad 2.29. *V případě, kdy* $K = \mathbb{Q}[\sqrt{c}]$*, pro* $\sqrt{c} \notin \mathbb{Q}$*, je*

$$O_K = \begin{cases} \mathbb{Z}[\sqrt{c}] \text{ pro } c \equiv 2, 3 (mod\ 4) \\ \mathbb{Z}[\frac{1}{2}, \sqrt{c}] \text{ pro } c \equiv 1 (mod\ 4) \end{cases}$$

2.3 Norma a číselné těleso

Pro číselné síto je norma podstatný pojem. Často bývá norma uváděna společně se stopu. Vhledem k tomu, že číselné síto využívá pouze normu, stopu vynecháme. Toto téma uvedeme bez důkazů až na část o okruhu algebraických celých číslech číselného tělesa K.

Nechť $T \subseteq U$ je rozšíření těles. Mějme prvek $\alpha \in U$ a zobrazení μ_α definované $\mu_\alpha(x) = \alpha x$ pro všechna $x \in U$. Determinant lineárního multiplikativního zobrazení μ_α vektorového prostoru U (nad T) není závislý na zvolené bázi U. Budeme ho proto značit pouze $\det(\mu_\alpha)$.

Definice 2.30. ***Normou*** *prvku* $\alpha \in T$ *rozumíme hodnotu* $\mathrm{N}_{U|T}(\alpha) = \det(\mu_\alpha)$.

Zřejmě platí, že $\det(\mu_\alpha) \in U$, protože všechny hodnoty v matici zobrazení μ_α jsou prvky z U pro libovolnou volbu báze. Vzhledem k vlastnostem determinantu matice je norma multiplikativní zobrazení. Pro $\alpha, \beta \in U$ platí $\mathrm{N}_{U|T}(\alpha\beta) = \mathrm{N}_{U|T}(\alpha)\,\mathrm{N}_{U|T}(\beta)$.

Poznámka 2.31. *Nechť* $T \subseteq U$ *je rozšíření těles stupně* n*. Pak pro prvky* $\alpha \in T$ *platí* $\mathrm{N}_{U|T}(\alpha) = \alpha^n$*. Matice zobrazení* μ_α *je v takovém případě tvaru* $\alpha\mathbf{I}$ *pro všechny prvky z* $\alpha \in T$*, při libovolné volbě báze.*

Norma také bývá definována jako součin konjugovaných prvků. Připomeňme ekvivalenci těchto definic.

Hlavní tvrzení 2.32. *Nechť* $T \subseteq U$ *je rozšíření těles konečného stupně* n*, které je separabilní. Dále mějme prvek* $\alpha \in U$ *a* $\mathbf{M}_\alpha$ *matici zobrazení* μ_α *vzhledem k nějaké bázi. Pak platí*

$$\det(x\mathbf{I} - \mathbf{M}_\alpha) = \prod_{\sigma \in Hom_T(U, \overline{T})} (x - \sigma(\alpha)),$$

$$\mathrm{N}_{U|T}(\alpha) = \prod_{\sigma \in Hom_T(U, \overline{U})} \sigma(x).$$

Důsledek 2.33. *Mějme konečné separabilní rozšíření těles $T \subseteq U$ stupně n a prvek α z celistvého uzávěru T v U. Pak jeho norma $\mathrm{N}_{U|T}(\alpha)$ leží v T.*

Podívejme se nyní na normu prvků z číselného tělesa K. Podle tvrzení 2.8 máme při stupni rozšíření $[K : \mathbb{Q}] = n$ právě n různých vnoření číselného tělesa K do $\mathbb{C}$, které zachovávají identitu na $\mathbb{Q}$.

Zaměřme se na prvky z O_K. Norma každého prvku $\alpha \in O_K$ je podle důsledku 2.33 celočíselná.

Tvrzení 2.34. *Mějme $\alpha \in O_K$. Jeho norma je rovna ± 1, právě když je α jednotkou v O_K.*

Důkaz. Dokažme nejprve implikaci „$\Leftarrow$“. Nechť máme jednotku $\alpha \in O_K$. Zřejmě i $\frac{1}{\alpha} \in O_K$. Vzhledem k multiplikativitě normy platí $1 = \mathrm{N}_{K|\mathbb{Q}}(1) = \mathrm{N}_{K|\mathbb{Q}}(\alpha)\,\mathrm{N}_{K|\mathbb{Q}}(\frac{1}{\alpha})$. Z tvrzení 2.33 plyne, že $\mathrm{N}_{K|\mathbb{Q}}(\alpha), \mathrm{N}_{K|\mathbb{Q}}(\frac{1}{\alpha}) \in \mathbb{Z}$. Jediní celočíselní dělitelé 1 v oboru celých čísel jsou pouze ± 1.

Pro „$\Rightarrow$“ mějme prvek $\alpha \in O_K$ s normou rovnou ± 1. Pak je absolutní člen minimálního polynomu $f_\alpha(x)$ roven ± 1. Potom prvek $\frac{1}{\alpha} \in K$ je kořenem polynomu $x^d f(\frac{1}{x})$, což je opět monický polynom s koeficienty v $\mathbb{Z}$. Tedy $\frac{1}{\alpha} \in O_K$ a α je jednotkou v O_K. □

Hlavní tvrzení 2.35. *Abelovská grupa O_K je volná a její hodnost je rovna $n = [K : \mathbb{Q}]$.*

Důkaz. Chceme dokázat, že O_K je isomorfní $\mathbb{Z}^n$. Zřejmě $\mathbb{Z}^n$ lze vnořit do O_K. Okruh O_K je stupně n, protože pro $K = \mathbb{Q}[\alpha]$ nastává $\mathbb{Z}[\alpha] \subseteq O_K$. Ukažme, že se jedná o isomorfismus.

Každý prvek $\gamma \in O_K$ můžeme zapsat jako lineární kombinaci prvků $\alpha_1, \ldots, \alpha_n$ báze K jako vektorového prostoru nad $\mathbb{Q}$.

$$\gamma = \sum_{i=1}^{n} c_i \alpha_i,\ c_i \in \mathbb{Q}.$$

Tedy O_K lze vnořit do $\mathbb{Q}^n$. Stačí ukázat, že pro libovolný prvek $\gamma \in O_K$ jsou jmenovatelé všech jeho koeficientů c_i omezeni stejnou konstantou B. V takovém případě existuje vnoření O_K do $\mathbb{Z}^n$.

Pro spor předpokládejme, že v O_K existuje posloupnost prvků, jejichž koeficienty c_{ij} jsou nezkratitelné a mají jmenovatele jdoucí do nekonečna:

$$\gamma_j = \sum_{i=1}^{n} c_{ij} \alpha_i,\ c_{ij} \in \mathbb{Q}.$$

Víme, že podle důsledku 2.33 je norma $\mathrm{N}_{K|\mathbb{Q}}(\gamma_j) \in \mathbb{Z}$. Norma prvku je definována jako $\det(\mu_{\gamma_j})$ kdy při volbě báze, kterou uvažujeme, odpovídají prvky matice právě hodnotám c_{ij}. Zvolme $c'_{ij} = c_{ij} - \lfloor c_{ij} \rfloor$. Potom prvky tvaru $\sum_{i=1}^{n} c'_{ij}\alpha_i$ leží opět v O_K, ale jejich norma již bude příliš malá na to, aby byla celým číslem. Tím získáváme spor. Mez pro tyto jmenovatele označme B. Pak máme

$$O_K \subseteq \frac{1}{B} \oplus_{i=1}^{n} \alpha_i \mathbb{Z}.$$

Pravá strana inkluze je volná Abelovská grupa stupně n a okruh algebraických celých čísel O_K je volný.

□

Dosud jsme pracovali pouze s normou prvku, rozšiřme tento pojem na normu ideálu. Mějme okruh R a jeho ideál I. Definujme normu ideálu jako řád faktorokruhu nad ideálem.

Definice 2.36. ***Normou*** *ideálu I okruhu R rozumíme $\mathcal{N}(I) = |R/I|$.*

Příklad 2.37. *Mějme $a \in \mathbb{Z}$ nenulové. Normu ideálu (a) okruhu O_K spočteme podle definice $\mathcal{N}((a)) = |O_K/(a)|$.*

Podle tvrzení 2.35 víme, že O_K je abelovská grupa hodnosti n. Mějme celistvou bázi $B_1 = \{\beta_1, \ldots, \beta_n\}$ této abelovské grupy. Zřejmě $O_K = \oplus_{i=1}^{n} \beta_i \mathbb{Z}$. Ideál (a) má zřejmě stejnou hodnost jako O_K, protože má volnou bázi $B_2 = \{a\beta_1, \ldots, a\beta_n\}$ a tedy platí $(a) = \oplus a\beta_i\mathbb{Z}$.

$$\mathcal{N}((a)) = |O_K/(a)| = \oplus_{i=1}^{n} \beta_i\mathbb{Z}/ \oplus a\beta_i\mathbb{Z} \cong (\mathbb{Z}/a\mathbb{Z})^n$$

Tím jsme získali, že $\mathcal{N}((a)) = |a|^n$ pro $a \in \mathbb{Z}$.

Zjistit počet prvků není vždy triviální úkol, proto je vhodné uvážit rychlejší cestu výpočtu této normy při speciálních výchozích podmínkách.

Nás bude dále zajímat norma hlavních ideálů okruhu algebraických celých čísel O_K. Hlavně pak práce s hlavními ideály tohoto okruhu.

Nyní dokažme základní vlastnost okruhu algebraických celých čísel O_K a to, že je isomorfní $\mathbb{Z}^n$.

Hlavní tvrzení 2.38. *Mějme hlavní ideál I okruhu O_K generovaný nenulovým prvkem $a \in K$. Pak $\mathcal{N}(I) = \left|\mathrm{N}_{K|\mathbb{Q}}(a)\right|$.*

Důkaz. Mějme báze B_1 a B_2 definované stejně jako v příkladu 2.37. Podle tvrzení 2.3 a toho, že O_K má strukturu modulu můžeme říci, že i každý jeho ideál J lze uvažovat jako volný modul. Ideál J vnímaný jako Abelova grupa má bázi $B_3 = \{d_1\beta_1, \ldots, d_n\beta_n\}$, kde $d_1, \ldots, d_n \in \mathbb{Z}$. Každý nenulový hlavní ideál okruhu O_K má hodnost stejnou jako je stupeň K, což je zřejmé podle jeho celistvé báze, viz příklad 2.37. Každý ideál obsahuje nějaký hlavní ideál a tedy má hodnost rovnou stupni. Tedy n je rovno hodnosti číselného tělesa K.

Zřejmě B_1, B_2 a B_3 jsou báze K nad $\mathbb{Q}$. Všechny prvky těchto bází jsou zřejmě nenulové a mají stejný počet prvků, protože volné $\mathbb{Z}$-moduly mají hodnost rovnou velikosti volné báze. Navíc je-li volný $\mathbb{Z}$-modul obsažen v nějakém vektorovém prostoru, tak jeho hodnost je nejvýše dimenze.

Mějme matici identického zobrazení přecházející od báze B_1 k B_3, označme ji $[id]_{B_3B_1}$. Taková matice je diagonální s prvky $d_1, \ldots, d_n \in \mathbb{N}$ na hlavní diagonále. Označme $\mathbf{I}$ jednotkovou matici $n \times n$. Matice $[id]_{B_2B_1} = a\mathbf{I}$ je rovna matici zobrazení μ_a vzhledem k libovolné bázi. Ukažme, že matice přechodu mezi různými bázemi ideálů z O_K mají determinant rovný ± 1.

Uvažme matice $[id]_{B_3B_2}$ a $[id]_{B_2B_3}$. Zřejmě jsou tyto matice navzájem inverzní. Obě matice jsou typu $\mathbb{Z}^{n\times n}$. Dále víme, že determinanty obou těchto matic jsou celočíselné. Nutně tedy musí být rovny invertibilním prvkům v $\mathbb{Z}^* = \{\pm 1\}$. To také znamená, že oba determinanty jsou rovny buď $+1$, nebo jsou oba rovny -1.

Při přechodu k vyjádření báze ideálu nad jinou celistvou bází se determinant, jak víme, v absolutní hodnotě nemění. V diagonální matici je tento determinant roven počtu prvku. Při vyjadřování báze B_2 nad B_1 dostaneme matici zobrazení μ_a, což je podle definice norma prvku a a tedy $\mathcal{N}(I) = \left|\mathrm{N}_{K|\mathbb{Q}}(a)\right|$. □

2.4 Dedekindovy obory

Toto téma je podstatné pro vlastnosti základních struktur algoritmu, ale obecné důkazy uvedených tvrzení v této sekci nemají výpovědní hodnotu pro téma práce, proto je neuvedeme.

Definice 2.39. *Dedekindův obor D je obor integrity, pro který platí, že:*

- *je celistvě uzavřený,*
- *je noetherovský,*
- *každý jeho nenulový prvoideál je maximální.*

Tvrzení 2.40. *Nechť $T \subseteq U$ je rozšíření těles konečného stupně. Buď T podílové těleso Dedekindova oboru D. Pak celistvý uzávěr oboru D v U je opět Dedekindův obor.*

Tvrzení 2.41. *Buď D Dedekindův obor a T jeho podílové těleso. Pro libovolný nenulový vlastní ideál I oboru D existují nenulové prvoideály $P_1, \ldots, P_k$ v oboru D tak, že*

$$\prod_{i=1}^{k} P_i \subseteq I.$$

Poznámka 2.42. *Mějme ideál I oboru R, který má podílové těleso T. Položme*

$$I^{-1} = \{t \in T;\ tI \subseteq R\}.$$

Zřejmě $II^{-1} \subseteq R$ a rovnost nemusí vždy nastat. Pokud existuje takový ideál J, že $IJ = R$ pak již $J = I^{-1}$.

*Mějme ideály I a J v Dedekindově oboru. Říkejme, že J **dělí** I pokud existuje ideál P takový, že $I = JP$ tedy $J \supset I$.*

Tvrzení 2.43. *Buď D Dedekindův obor a T jeho podílové těleso. Pro nenulový ideál I a nenulový prvoideál P oboru D platí*

$$I \subsetneq IP^{-1}.$$

Pro nevlastní ideál $I = D$ zřejmě platí, že $D \subsetneq P^{-1}$. Z maximality prvoideálů v D plyne, že pro libovolný nenulový prvoideál vždy platí

$$P \subsetneq P^{-1}P = D \subsetneq P^{-1}.$$

Z těchto tvrzení lze přímo dokázat podstatná vlastnost Dedekindových oborů. Nejenže platí inkluze $\prod_{i=1}^{k} P_i \subseteq I$, ale dokonce pro vhodné prvoideály P_i Dedekindova oboru D nastává rovnost.

Hlavní tvrzení 2.44. *Každý vlastní nenulový ideál Dedekindova oboru lze jednoznačně, až na pořadí, vyjádřit jako součin prvoideálů.*

Rozklad ideálu na součin prvoideálů je velice užitečný. Obecně neplatí, že by byl možný v libovolném oboru. Někdy bývají Dedekindovy obory přímo definovány jako obory, kde je takový rozklad možný.

Příklad 2.45. *Okruh algebraických celých čísel O_K číselného tělesa K stupně n je Dedekindův obor.*

Již víme, že O_K je celistvě uzavřený ve svém podílovém nadtělese, kterým je K. Podle tvrzení 2.35 víme, že O_K má celistvou bázi, čili je to konečně generovaný $\mathbb{Z}$-modul. Tedy je jako $\mathbb{Z}$-modul noetherovský. Dokažme tedy poslední vlastnost, že každý nenulový prvoideál je maximální v O_K.

Stačí ukázat, že pro každý nenulový prvoideál P okruhu O_K je O_K/P těleso. Mějme $p \in P$ nenulový prvek s minimálním polynomem $f_{p,\mathbb{Z}}(x) = \sum_{i=0}^{n} c_i x^i$. Pak je absolutní člen $c_0 \in P$, tedy ho můžeme vyjádřit pomocí $f_{p,\mathbb{Z}}(p) = 0$ následovně

$$c_0 = -(c_n p^n + c_{n-1} p^{n-1} + \ldots + a_1 p) \in P.$$

Vzhledem k tomu, že $f_{p,\mathbb{Z}}$ je ireducibilní polynom, je absolutní člen c_0 nenulový. Zřejmě pak $c_0 \in P \cap \mathbb{Z}$. Buď $(b_1, \ldots, b_n)$ báze P jako konečně generovaného $\mathbb{Z}$-modulu (tvrzení 2.35). Pak $(b_1 \,(mod\ P), \ldots, b_n \,(mod\ P))$ generuje $O_K/(P \cap \mathbb{Z})$ jako vektorový prostor nad $\mathbb{Z}/P$. To znamená, že O_K/P je konečný obor integrity a tedy těleso.

Pro libovolný nenulový ideál I okruhu O_K platí $I \cap \mathbb{Z} \neq 0$. Je-li I navíc prvoideál, pak obsahuje právě jedno prvočíslo. Kdyby jich obsahoval více, pak by již nebyl vlastní. Naopak musí obsahovat alespoň jedno prvočíslo, protože se jedná o prvoideál.

Tvrzení 2.46. *Pro prvoideál P okruhu O_K platí, že existuje prvočíslo p takové, že $P \cap \mathbb{Z} = p\mathbb{Z}$.*

Důkaz. Podle předchozích úvah vyplývá, že $P \cap \mathbb{Z} \supseteq p\mathbb{Z}$. Navíc $P \cap \mathbb{Z}$ je zřejmě prvoideál v $\mathbb{Z}$ a tedy nastává rovnost. □

Jak zmíníme dále, platí ještě obecnější tvrzení a to i pro ideály vynásobené inverzním prvkem k libovolnému prvku z Dedekindova oboru.

Definice 2.47. *Mějme R obor integrity a T jeho podílové těleso.* ***Lomený ideál*** *J nazýváme konečně generovaný R-podmodul tělesa T.*

Podle definice je lomený ideál ideálem oboru R právě v případě, kdy je jeho podmnožinou.

Tvrzení 2.48. *J je lomený ideál tělesa T, právě když existuje $a \in R \backslash \{0\}$ tak, že aJ je ideál v R.*

Tvrzení 2.49. *Pro lomený ideál J Dedekindova oboru D je i J^{-1} lomený ideál.*

Hlavní tvrzení 2.50. *Mějme D Dedekindův obor a T jeho podílové těleso. Potom každý lomený ideál J Dedekindova oboru D lze jednoznačně až na pořadí vyjádřit ve tvaru*

$$J = \prod_{i=1}^{k} P_i^{n_i},$$

kde P_i jsou po dvou různé nenulové prvoideály D a $n_1, \ldots, n_k \in \mathbb{Z}\backslash\{0\}$.

Tvrzení 2.51. *Každý ideál Dedekindova oboru je lomený.*

Hlavní tvrzení 2.52. *Lomené ideály Dedekindova oboru tvoří volnou grupu. Bází této grupy jsou všechny nenulové prvoideály.*

Vraťme se zpět k normám ideálů. Následující tvrzení budeme využívat v algoritmu pro rozklad normy hlavních ideálů, ale platí obecně pro libovolný ideál okruhu O_K. Rozklad normy nám totiž dává informaci o tom, které prvoideály patři do rozkladu daného ideálu Dedekindova oboru O_K.

Tvrzení 2.53. *Mějme těleso T a v něm Dedekindův obor R. Dále nenulový prvoideál P oboru R. Potom P^i/P^{i+1} je vektorový prostor nad R/P dimenze 1, pro každé $i \in \mathbb{N}$.*

Důkaz. Zřejmě R/P je těleso, protože prvoideál P je v Dedekindově oboru maximální.

Vzhledem k jednoznačnosti rozkladů na prvoideály v Dedekindově oboru nutně platí $P^{i+1} \subsetneq P^i$. Tedy existuje takový ideál I, že $P^{i+1} \subsetneq I \subseteq P^i$, tvrzení 2.43. Pak platí, že $P \subsetneq IP^{-i} \subseteq R$. Prvoideál P je maximální a tedy musí nastat $IP^{-i} = R$, tím musí platit $I = P^i$. □

Tvrzení 2.54. *Mějme ideál I okruhu O_K. Dále mějme po dvou různé prvoideály $P_1, \ldots, P_n$ takové, že $I = \prod_{i=1}^{n} P_i^{r_i}$, kde $r_1, \ldots, r_n \in \mathbb{N}$. Pak platí*

$$\mathcal{N}(I) = \prod_{i=1}^{n} \mathcal{N}(P_i)^{r_i}.$$

Důkaz. Podle definice $\mathcal{N}(I) = [O_K : I] = |O_K/I|$. Tedy máme podle rozkladu ideálu I:

$$\mathcal{N}(I) = \left| \prod_{i=1}^{n} O_K/P_i^{r_i} \right| = \prod_{i=1}^{n} |O_K/P_i^{r_i}| = \prod_{i=1}^{n} \prod_{j=1}^{r_i} [P_i^{j-1} : P_i^j].$$

Poslední rovnost vychází z Lagrangeovy věty, která říká, že počet prvků grupy je roven řádu grupy krát její index. Dále pro index podgrupy platí podle druhé věty o izomorfizmu $\left|(O_K/P_i^j)/(P_i^{j-1}/P_i^j)\right| = \left|O_K/P_i^j\right|$ pro $j = 2, 3, \ldots$.

Zřejmě lze P_i^{j-1}/P_i^j brát jako vektorový prostor dimenze 1 nad tělesem O_K/P_i vzhledem k předchozímu tvrzení 2.53. Tedy $\left|P_i^{i-1}/P_i\right| = |O_K/P_i| = \mathcal{N}(P_i)$. Tím již rovnou plyne

$$\mathcal{N}(I) = \prod_{i=1}^{n} \left|P_i^{i-1}/P_i\right|^{r_i} = \prod_{i=1}^{n} \mathcal{N}(P_i)^{r_i}.$$

□

Z této vlastnosti norem okamžitě získáme i možnost rozkladu na menší počet činitelů, než přímo na všechny prvoideály.

Tvrzení 2.55. *Mějme ideály I, J okruhu O_K. Pak $\mathcal{N}(IJ) = \mathcal{N}(I)\mathcal{N}(J)$.*

Důkaz. Mějme po dvou různé prvoideály $P_1, \ldots, P_n$ takové, že $I = \prod_{i=1}^{n} P_i^{r_i}$ a $J = \prod_{i=1}^{n} P_i^{s_i}$, kde $r_1, \ldots, r_n, s_1, \ldots, s_n, \in \mathbb{Z}$. Pak

$$\mathcal{N}(IJ) = \mathcal{N}(\prod_{i=1}^{n} P_i^{r_i+s_i}) = \prod_{i=1}^{n} \mathcal{N}(P_i)^{r_i+s_i} = \prod_{i=1}^{n} \mathcal{N}(P_i)^{r_i} \prod_{i=1}^{n} \mathcal{N}(P_i)^{s_i} = \mathcal{N}(I)\mathcal{N}(J).$$

□

Tato tvrzení jsou velice užitečná při rozkládání ideálů na prvoideály, čímž se budeme zabývat v další sekci. Nejprve rozložíme normu ideálu okruhu O_K na prvočinitele. Podle prvočinitelů dohledáme prvoideály s příslušnou normou patřící do rozkladu.

2.5 Rozklad na prvoideály v rozšířeních

Uvažujme stále číselné těleso $K = \mathbb{Q}[\alpha]$, kde $\alpha \in \mathbb{C}$ je algebraický prvek nad $\mathbb{Z}$ s minimálním polynomem f_α stupně n rovnému stupni číselného tělesa. Pokud v této sekci mluvíme o ideálech, jedná se o ideály okruhu O_K.

Definice 2.56. *Mějme prvočíslo p takové, že $P \cap \mathbb{Z} = p$, kde P je prvoideál O_K. Potom říkáme, že se jedná o prvoideál **nad prvočíslem** p.*

Vzhledem k předchozím úvahám těleso O_K/P je charakteristiky p, právě když P je prvoideál nad prvočíslem p.

Tvrzení 2.57. *Pro každé prvočíslo p existuje prvoideál okruhu O_K takový, že norma $\mathcal{N}(I)$ je rovna mocnině p.*

Důkaz. Podle příkladu 2.37 víme, že existuje ideál s normou rovnou mocnině p. Využijeme-li tvrzení 2.54 získáme:

$$p^n = \mathcal{N}((p)) = \prod_{i=1}^{n} \mathcal{N}(P_i)^{r_i}.$$

Normy všech ideálů okruhu O_K jsou celočíselné, tvrzení 2.33. Mocninu prvočísla p^n tedy můžeme rozložit pouze na násobky p a 1 což nám dává existenci prvoideálu P splňujícího $\mathcal{N}(P) = p^k$, kde $k \in \{1, \ldots, n\}$. □

Definice 2.58. *Pro prvoideál P mějme $\mathcal{N}(P) = p^k$. Hodnotu k nazýváme **stupeň inerce**, nebo také **stupeň nehybnosti**.*

Uvažujme nyní rozklad hlavního ideálu nad prvočíslem p:

$$(p) = \prod_{i=1}^{n} P_i^{r_i}$$

Zřejmě pokud máme prvoideál P nad prvočíslem p, pak je již v tomto rozkladu, protože platí $(p) \subseteq P$ a protože pracujeme v Dedekindově oboru.

Definice 2.59. *Exponenty r_i z rozkladu ideálu (p) nazýváme **ramifikační index** P_i nad p, nebo také **index větvení**.*

Hlavní tvrzení 2.60 (Fundamentální rovnost)**.** *Mějme prvoideály $P_1, \ldots, P_r$ nad prvočíslem p. Nechť má každý z těchto prvoideálů stupeň inerce k_i a ramifikační index e_i. Pak platí*

$$\sum_{i=1}^{r} e_i k_i = n,$$

kde číslem n značíme stupeň číselného tělesa.

Důkaz. Vzhledem k tvrzením 2.54 a 2.57 víme, že platí:

$$\mathcal{N}((p)) = \prod_{i=1}^{r} \mathcal{N}(P_i)^{e_i} = \prod_{i=1}^{r} p^{k_i e_i} = p^{\sum_{i=1}^{r} k_i e_i}.$$

V příkladu 2.37 jsme ukázali, že $\mathcal{N}(p) = p^n$. Což okamžitě dává $\sum_{i=1}^{r} e_i k_i = n$. □

Pro číselné síto je podstatné, rozkládat hlavní ideály okruhu O_K na prvoideály. Pro tyto rozklady je třeba rozlišovat prvočísla podle toho, jestli dělí $|O_K/\mathbb{Z}[\alpha]|$. Prvočísla, která dělí $|O_K/\mathbb{Z}[\alpha]|$, bývají označována jako **speciální**. Rozložit prvoideál nad speciálním prvočíslem je mnohem složitější. Dále popíšeme prvoideály nad nespeciálními prvočísly.

Uvažujme normu ideálu rovnou určité hodnotě $\mathcal{N}(I) = c$. Ukažme, že máme pouze konečně mnoho prvoideálů dané normy.

Tvrzení 2.61. *Buď $c \in \mathbb{R}^+$. Pak existuje pouze konečně mnoho ideálů I Dedekindova oboru O_K takových, že $\mathcal{N}(I) \leq c$.*

Důkaz. Stačí dokázat, že existuje konečně mnoho prvoideálů P takových, že mají omezenou normu $\mathcal{N}(P) \leq c$.

Pro P nenulový prvoideál platí $\mathcal{N}(P) > 1$. Pokud prvoideál P je součástí primární dekompozice ideálu I, potom $\mathcal{N}(I) \geq \mathcal{N}(P)$.

Podle tvrzení 2.46 existuje prvočíslo p, že $P \cap \mathbb{Z} = p\mathbb{Z}$. Norma takového prvoideálu je pak rovna mocnině p, jak ukazujeme v důkazu tvrzení 2.57. Podle Fundamentální rovnosti 2.60 existuje pouze konečně mnoho prvoideálů, které mají normu rovnou p. Tedy existuje pouze konečně mnoho prvoideálů, které mají normu menší než dané číslo c.

□

Poznámka 2.62. *Dokonce platí silnější tvrzení. Pro libovolné prvočíslo p existuje pouze konečně mnoho prvoideálů P oboru O_K. Uvažujme primární dekompozici ideálu (p) ve smyslu pO_K, pouze prvoideály z dekompozice tohoto ideálu totiž splňují, že $P \cap \mathbb{Z} = p\mathbb{Z}$. Tedy pouze tyto prvoideály mají normu rovnou mocnině p.*

Tvrzení 2.63. *Mějme prvoideál P nad nespeciálním prvočíslem p. Potom platí, že $O_K = \mathbb{Z}[\alpha] + P$.*

Důkaz. Víme, že P je maximální prvoideál okruhu O_K a položme $q = |O_K/\mathbb{Z}[\alpha]|$. Zřejmě pro $q = 1$ tvrzení okamžitě platí. Uvažujme tedy $q > 1$. Mějme libovolný $\beta \in O_K$. Pro něj platí $q\beta \in \mathbb{Z}[\alpha]$. Podle předpokladu jsou p a q nesoudělná a tedy pro ně existují koeficienty $r, s \in \mathbb{Z}$ takové, že $pr + qs = 1$. Vzhledem k tomu, že $p \in P$ a $qs\beta \in \mathbb{Z}[\alpha]$ dokážeme součtem prvku z prvoideálu a $\mathbb{Z}[\alpha]$ získat libovolný prvek z O_K,

$$pr\beta + qs\beta = \beta.$$

□

Mějme p nespeciální prvočíslo. Buď $f_\alpha = \prod_{i=1}^{s} {t_i}^{e_i} \pmod p$ rozklad na ireducibilní navzájem nesoudělné polynomy. Zřejmě polynomy t_i nejsou dány jednoznačně, stačí nám pouze jednoznačnost modulo p.

Tvrzení 2.64. *Ideály tvaru $P_i = (p, t_i(\alpha))$ jsou buď prvoideály stupně nehybnosti* $\deg(t_i)$*, nebo $P_i = O_K$.*

Důkaz. Uvažujme O_K/P_i, to je buď rovno 1 nebo $\mathbb{Z}_p[x]/t_i\mathbb{Z}_p[x]$, což je těleso, protože t_i je ireducibilní modulo p. Navíc $\mathbb{Z}[x]/(p, t_i)$ je těleso. Tedy ideál (p, t_i) je maximální v $\mathbb{Z}[x]$. Uvažujme přirozený homomorfizmus $\varphi : \mathbb{Z}[x] \to O_K/P_i$ definovaný $\varphi(x) = \alpha + P_i$. Prvek $\alpha + P_i$ je generátorem O_K/P_i, podle tvrzení 2.63. Tedy $\mathbb{Z}[x]/\operatorname{Ker}\varphi \approx O_K/P_i$ a homomorfizmus φ je epimorfizmem. Zřejmě maximální ideál $(p, t_i) \subseteq \operatorname{Ker}\varphi$ a tedy nastává rovnost nebo je $\operatorname{Ker}\varphi = \mathbb{Z}[x]$. To již znamená, že O_K/P_i je rovno buď $\mathbb{Z}_p/t_i\mathbb{Z}_p$ nebo 1. □

Tvrzení 2.65. *Mějme prvoideály $P_i = (p, t_i(\alpha))$ okruhu O_K. Potom*

$$(p) \supseteq \prod_{i=1}^{r} P_i^{e_i}.$$

Důkaz. Nejprve uvažme, že prvoideály P_i jsou po dvou různé, nebo rovny celému O_K. Pro různé $i \neq j$ lze najít takové koeficienty $u, v \in \mathbb{Z}$ splňující $ut_i(\alpha) + vt_j(\alpha) = 1$. Takže $P_i + P_j = O_K$ pro libovolné dva indexy $i \neq j$.

Platí $\prod_{i=1}^{r} P_i^{e_i} \subseteq (p, t_1(\alpha), \ldots, t_r(\alpha))$. Navíc $\prod_{i=1}^{s} t_i(x)^{e_i} - f_\alpha \in p\mathbb{Z}[x]$. Po dosazení α získáme $\prod_{i=1}^{s} t_i(\alpha)^{e_i} \in p\mathbb{Z}[\alpha]$ a tedy patří také do hlavního ideálu (p) okruhu O_K. □

Hlavní tvrzení 2.66. *Mějme nespeciální prvočíslo p a rozklad polynomu na ireducibilní polynomy $f_\alpha = \prod_{i=1}^{s} {t_i}^{e_i} \pmod p$. Potom pro rozklad hlavního ideálu (p) okruhu O_K na prvoideály platí:*

$$(p) = \prod_{i=1}^{s} P_i^{e_i},$$

kde prvoideály $P_i = (p, t_i(\alpha))$ mají stupně nehybnosti $\deg(t_i)$.

Důkaz. Inkluzi $(p) \supseteq \prod_{i=1}^{r} P_i^{e_i}$ jsme již ukázali v tvrzení 2.65. Dokažme, že se jedná o rovnost.

Z rozkladu polynomu f_α máme $\deg(f_\alpha) = \prod_{i=1}^{s} \deg(t_i)^{e_i}$ čímž dostáváme fundamentální rovnost $d = \sum_{i=1}^{r} \deg(t_i)e_i$ (tvrzení 2.60).

Podle předchozího tvrzení 2.64 víme, že každý z prvoideálů P_i má stupeň nehybnosti buď $\deg(t_i)$, nebo platí $P_i = O_K$. Tedy vzhledem k fundamentální rovnosti musí pro všechny prvoideály P_i nastat pouze první případ. To nám přímo dává $(p) = \prod_{i=1}^{s} P_i^{e_i}$. □

Rozkládání nad speciálními prvočísly se nebudeme přímo zabývat. Jedná se o komplexní téma, kterému se věnuje například [1].

Kapitola 3

Číselné síto

V této kapitole popíšeme stručně a informativně celý algoritmus číselného síta. Proto zde nebude kladen nárok na podrobnou přesnost.

Základní princip číselného síta je Fermatova faktorizace přirozeného čísla N. Jejím cílem je najít dvě různá celá čísla x a y, jejichž čtverce se liší o celočíselné násobky čísla N.

$$x^2 \equiv y^2 \,(\mathrm{mod}\ N)$$

Předpokládejme, že $x > y$. Pokud rozdíl $x - y$ není roven jedné, využijeme rozkladu

$$(x + y)(x - y) = kN,\ k \in \mathbb{Z}.$$

Tedy nastane buď $\gcd(x - y; N) > 1$ a nebo $\gcd(x + y; N) > 1$. Tím máme velkou šanci, že při zjištění největšího společného dělitele získáme netriviálního dělitele čísla N. V opačném případě je největším společným dělitelem přímo číslo N a je třeba hledat jiné hodnoty x a y.

Fermatova faktorizace v úspěšném případě najde netriviálního dělitele čísla N. O úplný rozklad se jedná tehdy, pokud N je násobkem dvou prvočísel. Pokud tomu tak není, lze opět použít stejnou cestu na rozklad neprvočíselných dělitelů čísla N.

Existuje mnoho algoritmů a postupů k nalezení vhodných x a y. Míra jejich efektivity je závislá na velikosti čísla N. Čím větší prvočísla násobíme, tím je náročnější je zpět rozložit nebo najít dobrá x a y. V této kapitole popíšeme, jak získává hodnoty x a y algoritmus číselného síta.

3.1 Stručný přehled celého algoritmu

Mějme číslo N, které není prvočíslo. Cílem celého algoritmu je sestavit dva čtverce x^2 a y^2 tak, aby se daly použít k Fermatově faktorizaci čísla N. Číselné síto při hledání těchto dvojic pracuje v okruhu algebraických celých čísel číselného tělesa. Nezůstává tedy pouze v oboru celých čísel, do kterého patří jak N, tak i jeho hledaní dělitelé.

Nejprve v první části sestavíme číselná tělesa tvaru $K_1 = \mathbb{Q}(\alpha_1)$ a $K_2 = \mathbb{Q}(\alpha_2)$ pomocí dvou vhodně zvolených ireducibilních polynomů f_1 a f_2 s kořeny α_1, respektive $\alpha_2 \in \mathbb{C}$. Okruhy algebraických celých čísel O_{K_1} a O_{K_2} jsou Dedekindovy obory, jak jsme ukázali v příkladu 2.45. Připravili jsme je takto pro práci s jejich hlavními ideály.

V druhé části prosséváme v každém číselném tělese zvlášť. Vzhledem k tomu, že v obou tělesech vykonáváme odděleně stejné postupy, vynechme pro přehlednost indexy. Teoreticky se jedná o práci s hlavními ideály okruhu algebraických celých čísel číselného tělesa. Hledáme dostatečné množství ideálů, jejichž normy mají pouze malé dělitele.

Definice 3.1. *Celé číslo je B-hladké, pokud jsou všichni jeho prvočíselní dělitelé menší než B.*

Označme $F(x, y) \in \mathbb{Z}[x, y]$ zhomogenizovaný polynom původního ireducibilního polynomu f, který měl kořen α. Zvolíme konstantu B pro B-hladkost a interval, ze kterého budeme volit dvojice (a, b). Tyto dvojice reprezentují hlavní ideály $(a - b\alpha)$ okruhu O_K.

Teoreticky hledáme hlavní ideály $(a - b\alpha)$, které mají B-hladkou normu ideálu. Prakticky hledáme dvojice (a, b), které mají B-hladkou hodnotu $F(a, b)$. Tím pracujeme s normou hlavního ideálu $(a-b\alpha)$ okruhu O_K. V následující kapitole dokážeme, že norma hlavního ideálu $(a - b\alpha)$ je rovna právě $F(a, b)$.

Postup, kterým vybíráme a určujeme vhodná (a, b), nazýváme prosséváním. Všechny vybrané dvojice (a, b), které získáme prosetím (tedy mají B-hladkou hodnotu $F(a, b)$), zaznamenáme pro další zpracování. Množinu všech vybraných dvojic označme ζ. Její prvky nazývejme relace. Relace (a, b) určují hlavní ideály $(a - b\alpha)$ okruhu O_K. Ty budeme v dalších fázích rozkládat na hladké prvoideály a hledat jejich spárování. Tedy budeme hledat podmnožinu ζ' množiny ζ tak, že součin všech hlavních ideálů určených prvky ze ζ' nám dá čtverec v $\mathbb{Z}[\alpha]$.

V další části zpracujeme všechny relace (a, b) z množiny ζ. V příkladu 2.45 jsme ukázali, že O_K je Dedekindův obor a každý jeho ideál se jednoznačně rozkládá na

prvoideály, tvrzení 2.44. Označme B-hladkým ideálem každý ideál okruhu O_K, jehož norma je B-hladká. Tvrzení 2.61 nám dává, že takových ideálů je pouze konečný počet.

Vytvoříme velkou řídkou matici $\mathbf{M}$. Každý řádek je určen jednou relací ze ζ a reprezentuje hlavní ideál okruhu O_K. Sloupce reprezentují B-hladké prvoideály okruhu O_K. Vzhledem k tomu, že O_K je Dedekindův obor, tvrzení 2.54 a sekci 2.5 máme představu, jak připravit všechny prvoideály potřebné pro rozklad ideálů určených relacemi ze ζ. Strukturu těchto prvoideálů určuje věta 2.66 a tvar hlavních ideálů $(a - b\alpha)$, které budeme rozkládat. Vzhledem k volbě polynomu f_α získáváme polynomy t_i lineární. Navíc v praxi zanedbáváme speciální prvočísla, protože se jim algoritmus de facto vyhýbá.

Hodnoty v buňkách matice $\mathbf{M}$ reprezentují mocniny prvoideálů v rozkladu daného ideálu určeného relací podle řádku matice. Zajímá nás pouze, zdali je mocnina lichá nebo sudá. Hodnoty v matici jsou tedy uvedeny modulo 2.

Následně je z množiny ζ vybrána podmnožina ζ' tak, že po vynásobení všech ideálů $(a - b\alpha)$ určených relacemi ze ζ', získáme ideál, jež má všechny prvoideály ze svého rozkladu v sudé mocnině. Jedná se tedy o řešení rovnice $\mathbf{M}\mathbf{x}^T = \mathbf{0}$ v $\mathbb{Z}_2$, kde vektor $\mathbf{x}$ odpovídá na otázku, které z dvojic ze ζ vybraných proséváním použijeme dále pro sestavení čtverců v O_K.

Tím jsme získali součin ideálů $\prod_{(a,b)\in\zeta'} (a - b\alpha) = I^2$, kde I je ideál O_K. Součin hlavních ideálů však nemusí být hlavní ideál. Navíc i když je hlavní, tak ještě z daných dat nevyplývá, že je generován čtvercem. Existuje postup využívající kvadratické charaktery, díky kterému lze získat, že součin ideálů určených relacemi ze ζ' je generován čtvercem. Následně se získá odmocnina z tohoto součinu, která je generována prvkem z $\mathbb{Z}[\alpha]$. Její nalezení vyžaduje samostatný postup, který je závěrečnou fází algoritmu.

3.2 Fáze algoritmu

Projděme algoritmus podrobněji postupně po fázích. Cílem této práce je popsat pomocné algoritmy k hledání vhodných polynomů pro první fázi algoritmu. Uvedeme proto informativně idealizovaný algoritmus ve stručnosti a s odkazy na reálné zpracování.

3.2.1 První fáze (volba polynomů)

V první fázi hledáme dva různé ireducibilní monické polynomy f_1 a f_2 ze $\mathbb{Z}[x]$, které mají společný kořen m modulo N. Tyto dva polynomy nemají obecně žádný společný kořen.

$$f_1(m) \equiv f_2(m) \,(\mathrm{mod}\ N)$$

Nad těmito polynomy sestavíme teoreticky pro $i = 1, 2$ číselná tělesa $K_i = \mathbb{Q}(\alpha_i)$, kde $\alpha_i \in \mathbb{C}$ je kořen polynomu $f_i(x)$. Hlavně však budeme pracovat v okruzích O_{Ki} a přesněji hlavně v $\mathbb{Z}[\alpha_i]$.

Jak už víme, budeme často používat homogenní polynomy $y^d f_i(x, y) = F_i(x, y)$, kde $d = \deg(f)$ a $i = 1, 2$. Je proto důležité najít polynomy f_1 a f_2 tak, aby dávaly na předpokládaném intervalu dostatečně mnoho relací při zvolené metodě prosévání, tedy B-hladkých hodnot po dosazení do zhomogenizovaného polynomu. To je podstatné pro získání mnoha hodnot, ze kterých budeme vybírat relace pro sestavení čtverce v číselném tělese. Existuje více způsobů, jak hledat zmíněné polynomy. Podle volby těchto metod se pak různě volí hodnota B a interval, ve kterém hledáme relace (a, b). Dále je podstatné, aby se s takovými polynomy dalo efektivně a rychle počítat, protože do nich bývá dosazováno velké množství hodnot. Metodám hledání těchto polynomů se budeme více zabývat v následujících kapitolách.

Definujme homomorfizmy φ_1, φ_2 pro převod nalezených hodnot v číselném tělese zpět do celých čísel. Přesněji se jedná o hodnoty z podokruhu $\mathbb{Z}[\alpha]$ daného číselného tělesa. Pro $i = 1, 2$ a $r \in \mathbb{Z}$ mějme

$$\varphi_i : \mathbb{Z}[\alpha_i] \to \mathbb{Z}_N,$$

$$\varphi_i(\alpha_i) = m \,(\mathrm{mod}\ N)\,,$$

$$\varphi_i(r) = r \,(\mathrm{mod}\ N)\,.$$

V sekci 3.2.5 ukážeme, že stačí uvažovat homomorfizmus pouze z $\mathbb{Z}[\alpha_i] \subseteq O_K$.

Nyní máme vztah $\varphi_1(\alpha_1) \equiv m \equiv \varphi_2(\alpha_2) \,(\mathrm{mod}\ N)$. Taková volba je podstatná pro kongruenci nalezených čtverců z $\mathbb{Z}[\alpha_i]$, označme je β_i^2. Polynomy f_1 a f_2 volíme se všemi výše uvedenými podmínkami proto, abychom získali kongruenci čtverců. Pokud bychom získali stejné čtverce $x^2 = y^2$, musíme začít znovu. Potřebujeme pouze kongruenci čtverců mod N, nikoliv rovnost. Tím již získáme pro Fermatovu faktorizaci kongruentní čtverce v $\mathbb{Z}_N$ ve tvaru

$$\beta_1^2 = \prod_{(a,b)\in\zeta'} (a - b\alpha_1),$$

$$\beta_2^2 = \prod_{(a,b)\in\zeta'} (a - b\alpha_2),$$

$$\varphi\left(\beta_1^2\right) = x^2 \equiv y^2 = \varphi\left(\beta_2^2\right) \pmod{N}.$$

3.2.2 Druhá fáze (prosévání)

Do druhé fáze algoritmu vstupujeme, když máme připravena číselná tělesa K_1 a K_2, jejich obory algebraických celých čísel O_{K_1}, O_{K_2}, oba polynomy $f_1(x)$ a $f_2(x)$ a tím také jejich zhomogenizované verze $F_1(x, y)$ a $F_2(x, y)$. Tato fáze probíhá v obou tělesech zvlášť stejným způsobem, proto dále vynecháme indexy. Zvolíme mez $B \in \mathbb{N}$ pro B-hladkost a určíme interval $I \subset \mathbb{Z}^2$, ze kterého budeme uvažovat hodnoty (a, b).

Nyní začneme hledat dvojice $(a, b) \in I$ takové, aby hlavní ideál $(a - b\alpha)$ měl B-hladkou normu. Takové dvojice nazýváme relace a množinu všech relací označme ζ. Z relací budeme dále schopni najít ideál, který má v rozkladu na prvoideály pouze sudé exponenty. Omezením normy jsme omezili i množinu prvoideálů, které dělí prvoideály určené prvky ζ, množinu těchto prvoideálů pracovně nazýváme B-hladké prvoideály. Dvojice $(a, b) \in I$ vybíráme tedy podle toho, zda je hodnota $F(a, b)$ B-hladká. To se dá očekávat okolo kořenů $F(x, y)$, kdy bude $F(a, b)$ poměrně malé. Podmínka B-hladkosti bývá někdy oslabována tak, že jeden nebo více dělitelů ji nemusí splňovat. Tím získáme více relací k vzájemnému zkombinování a omezenou množinu B-hladkých prvoideálů rozšíříme pouze o konečně mnoho dalších prvoideálů. Všechny nalezené dvojice (a, b) zaznamenáváme pro další výpočty spolu s prvočíselnými děliteli $F(a, b)$ a jejich stupni.

Tato fáze je časově nejnáročnější z celého algoritmu, protože je třeba nasbírat velké množství relací a tedy se užívá velký prosévací interval I. Nejběžnější jsou dnes dva postupy na hledání vhodných dvojic. Nazývají se mřížové a klasické prosévání. Jejich popis lze nalézt v [1].

3.2.3 Třetí fáze (konstrukce matice)

Další fáze nebývá vždy brána jako samostatná fáze. Pro implementaci je vhodné ji oddělit. Jedná se o zpracování relací a vytvoření matice $\mathbf{M}$ v tělese $\mathbb{Z}_2$. Řádky této matice jsou určeny relacemi (a, b) z ζ. Tyto relace reprezentují ideály $(a - b\alpha)$. Zřejmě $a - b\alpha \in O_K$, protože O_K je podle tvrzení 2.35 okruh a $a, b, \alpha, -1 \in O_K$. Tedy $(a - b\alpha)$ je hlavní ideál okruhu O_K. Sloupce jsou určeny B-hladkými prvoideály z

Dedekidova oboru O_K. Jedná se o prvoideály nad všemi prvočísly, která jsou menší než B. Strukturu takových prvoideálů uvádíme v 2.5.

Matice $\mathbf{M}$ poskytuje informaci o tom, kdy rozklad daného ideálu v O_K obsahuje které prvoideály s lichým exponentem. Budeme hledat takovou podmnožinu ζ, aby součin hlavních ideálů určených relacemi již neobsahoval žádný prvoideál v liché mocnině.

Nyní matici zmenšíme o sloupce, kde není uvedena žádná hodnota 1. Navíc je výhodné matici následně zmenšit o řádky a sloupce, kde se v celém sloupci vyskytuje jediná hodnota 1. Jedná se totiž o dvojici (a, b), kterou není s čím zkombinovat. Obsahuje ideál v liché mocnině, který by zůstal stále v liché mocnině po libovolní kombinaci. Navíc žádný z ideálů nebudeme kombinovat sám se sebou, tím bychom nic nezískali. Tato úprava může výrazně zrychlit další výpočty vzhledem k velikosti matice.

3.2.4 Čtvrtá fáze (lineární)

Tuto fázi nazýváme lineární, protože se jedná o výpočet lineárních rovnic o více neznámých. Hledáme způsob, jak spárovat liché exponenty prvoideálů tak, aby po vynásobení všech vybraných ideálů $(a_i - b_i\alpha)$, pro $i = 1 \ldots, n$ nastalo:

$$\prod_{i=1}^{n} (a_i - b_i\alpha) = \prod_{j=1}^{k} {P_j}^{r_j},$$

kde P_j jsou prvoideály O_K a $r_j \in 2\mathbb{Z}$ pro všechna $j = 1, \ldots, k$. Stále pracujeme v obou tělesech odděleně.

Jedná se o řešení soustavy lineárních rovnic $\mathbf{M}\mathbf{x}^T = 0$, kde matici $\mathbf{M}$ jsme sestavili v předchozí fázi. Matice $\mathbf{M}$ je běžně velká řídká matice řádově o milionu řádků. Výpočet pomocí Gaussovy metody by změnil vlastnost řídkosti. Při implementaci by se jednalo o problém. Práce s takovou maticí, která by nebyla řídká, by byla výpočetně velice náročná. Proto se při implementaci vyplácí použít Wiedemannovu blokovou metodu [15] nebo Lanczošovu blokovou metodu [16].

3.2.5 Pátá fáze (odmocninová)

V poslední fázi začneme tím, že vezmeme výsledky z lineární fáze. Stále pracujeme v obou tělesech odděleně. Podle vektoru řešení $\mathbf{x}$ určíme, které relace použít k získání čtverce $\mathcal{N}(\prod_{\zeta'}(a - b\alpha))$. To však neznamená, že součin prvků $\prod_{\zeta'}(a - b\alpha)$

je čtvercem v číselném tělese. Tímto problémem za zabývá [1] a ukazuje, že této vlastnosti můžeme docílit pomocí postupu s kvadratickými charaktery. Pak již platí, že ideál $\prod_{\zeta'}(a - b\alpha) = I^2$ je hlavní ideál $I = (\beta)$.

Nyní máme dva čtverce ve dvou různých číselných tělesech pro které platí:

$$\varphi_1(\beta_1^2) \equiv \varphi_2(\beta_2^2) \,(\mathrm{mod}\ N)\,.$$

Následně zjistíme odmocninu čtverců v číselných tělesech β_1 a β_2. Získat odmocninu z $\prod_{\zeta'}(a - b\alpha_i)$ však není triviální a navíc požadujeme, aby tato odmocnina ležela v $\mathbb{Z}[\alpha_i]$. K tomu byla dříve používána Newtonova iterační metoda a Couveignesovo vylepšení. Dnes je za nejlepší odmocninovou metodu považována Montgomeryho metoda.

Podrobnější popis a vysvětlení odmocninové fáze a metod odmocňování v číselném tělese je možno najít například v [1, 5].

Kapitola 4

Podklady k generování polynomů

V této kapitole se již zaměříme pouze na první fázi algoritmu číselného síta. Konkrétně na to, co očekáváme od generovaných polynomů.

Mějme číselné těleso $K = \mathbb{Q}[\alpha]$ stupně n. Pak minimální polynom prvku $\alpha \in \mathbb{C}$ ze $\mathbb{Z}[x]$ má také stupeň n. Algoritmus číselného síta postupuje tak, že nejprve vygeneruje ireducibilní polynom s celočíselnými koeficienty, jehož kořen pak slouží k sestavení číselného tělesa. Našim cílem bude popsat smysl těchto ireducibilních polynomů pro algoritmus. Také uvedeme jejich porovnání pro výběr toho nejvhodnějšího polynomu.

4.1 Číselné těleso a algoritmus

Předpokládejme nadále že, pracujeme v Dedekindově oboru O_K, kde $K = \mathbb{Q}[\alpha]$. V číselném sítě se počítá rozklad hlavních ideálů prvků $a - b\alpha$ na prvoideály, sekce 2.5. Do prosévací fáze, kterou jsme stručně uvedli v předchozí kapitole, vstupují dvojice (a, b) reprezentující čísla tvaru $a - b\alpha$. Prvky (a, b) se vybírají z oblasti, které se tradičně říká prosévací interval. Tomu se budeme věnovat v sekci 4.2.

K prvku α uvažujme jeho minimální polynom $f_{\alpha,\mathbb{Z}}(x)$, který je zřejmě stupně d. V praxi je d malé číslo (maximálně 8). Od této kapitoly dále se bude vždy jednat o minimálním polynomu nad celými čísly, budeme ho značit pouze f_α. Definujme zhomogenizovaný minimální polynom tvaru $y^d f_\alpha(\frac{x}{y}) = F_\alpha(x, y)$, který budeme v algoritmu hojně používat. Ukažme nejprve výpočet normy prvku $a - b\alpha$, kde $a, b \in \mathbb{Z}$. Zřejmě $a - b\alpha \in O_K$. Uvažujme dále pouze normu prvků $\mathrm{N}_{K|\mathbb{Q}}$, kterou budeme značit pouze N.

Nejprve dokážeme pomocné tvrzení o determinantu matice zobrazení $\mu_{a-b\alpha}$.

Tvrzení 4.1. *Mějme číselné těleso $a - b\alpha$, kde $a, b \in \mathbb{Z}$ a α je celistvý prvek nad $\mathbb{Z}$. Dále mějme $F_\alpha(x, y)$ zhomogenizovaný minimální polynom prvku α. Potom* $\det(\mu_{a-b\alpha}) = F(a, b)$

Důkaz. Vzhledem k tomu, že pracujeme v číselném tělese tvaru $\mathbb{Q}$ je velikost matice zobrazení $\mu_{a-b\alpha}$ dána stupněm minimálního polynomu prvku α. Označme ho d. Dále mějme $\mathbf{M} = (c_{ij})_{i,j=0}^{d-1}$ matici lineárního zobrazení $\mu_{a-b\alpha}$ vzhledem k bázi $\left\{1, \alpha, \ldots, \alpha^{d-1}\right\}$ číselného tělesa nad $\mathbb{Q}$. Zřejmě všechny hodnoty v matici musí být celočíselné.

Ukažme nyní indukcí podle stupně polynomu, že determinant této matice je roven právě $F_\alpha(a, b) = b^d f_\alpha(\frac{a}{b})$. Nejnižší uvažovaný stupeň polynomu volme $d = 2$. Matice lineárního zobrazení $\mu_{a-b\alpha}$ je v takovém případě tvaru:

$$\mathbf{M} = \begin{pmatrix} a & ba_0 \\ -b & a + ba_1 \end{pmatrix}$$

$$\det \mathbf{M} = a^2 + aba_1 + b^2 a_0 = F(a, b).$$

Nechť tvrzení platí pro stupeň d. Dokažme, že platí i pro $d + 1$. Zřejmě

$$\mathbf{M} = \begin{pmatrix} a & 0 & 0 & \ldots & 0 & ba_0 \\ -b & a & 0 & \ldots & 0 & ba_1 \\ 0 & -b & a & \ldots & 0 & ba_2 \\ \vdots & \ddots & \ddots & \ldots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -b & -ab + ba_d \end{pmatrix}$$

$$\det \mathbf{M} = a \begin{vmatrix} a & 0 & \ldots & 0 & ba_1 \\ -b & a & \ldots & 0 & ba_2 \\ 0 & -b & \ddots & 0 & ba_3 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & -b & -ab + ba_d \end{vmatrix} + (-1)^{d-1} ba_0 \begin{vmatrix} -b & a & 0 & \ldots & 0 \\ 0 & -b & a & \ldots & 0 \\ 0 & 0 & -b & \ddots & 0 \\ \vdots & \ldots & \ddots & \ddots & a \\ 0 & 0 & 0 & \ldots & -b \end{vmatrix}$$

První determinant dokážeme spočítat z indukčního předpokladu. Jeho výsledek je $\sum_{i=1}^{d} a_i a^{i-1} b^{d-i}$, což je homogenní polynom $F_\alpha(x, y)$ stupně $d - 1$, do kterého bylo dosazeno (a, b). Druhý determinant je pouze násobek všech prvků na diagonále,

kterých je $d-1$, protože se jedná o horní trojúhelníkovou matici. Výsledek determinantu je

$$\det \mathbf{M} = a \sum_{i=1}^{d} a_i a^{i-1} b^{d-i} + b^d a_0 = F(a,b).$$

□

Tím okamžitě dostáváme i normu prvku $a - b\alpha$ a také hlavního ideálu tvaru $(a - b\alpha)$ okruhu O_K, která je podstatná pro zjišťování rozkladu na prvoideály v Dedekindově oboru O_K.

Tvrzení 4.2. *Mějme ideál $(a - b\alpha)$ okruhu algebraických celých čísel číselného tělesa $K = \mathbb{Q}[\alpha]$, kde $a, b \in \mathbb{Z}$ a α je celistvý prvek nad $\mathbb{Z}$. Při zavedeném značení platí $\mathcal{N}((a - b\alpha)) = F_\alpha(a,b)$.*

Důkaz. Norma hlavního ideálu generovaného prvkem $a - b\alpha \in O_K$ je podle 2.38 rovna normě prvku $a - b\alpha$. $\mathcal{N}((a - b\alpha)) = \mathrm{N}(a - b\alpha) = \det(\mu_{a-b\alpha})$ podle předchozího tvrzení máme okamžitě $\mathcal{N}((a - b\alpha)) = F_\alpha(a,b)$.

□

Rozkladu na prvoideály se nebudeme v této práci nebudeme více zabývat. Jedná se však o podstatnou část dalších fází algoritmu číselného síta, které jsme shrnuli v předchozí kapitole. Teoretický základ je popsán například v [1].

4.2 Prosévací interval

Nyní již víme, že role polynomů není pouze o sestavení číselného tělesa, ale také o hledání normy hlavních ideálů O_K ve tvaru $a - b\alpha$, kde $a, b \in \mathbb{Z}$. Při generování polynomů je třeba uvažovat i to, pro jaké ideály budeme zjišťovat jejich normy. Podívejme se tedy na definiční obor homogenních polynomů. Budeme ho dále nazývat **prosévací interval**. V praxi je běžné volit obdélníkový nebo dnes častější zkosený interval.

Uvažujme nejprve obdélníkový prosévací interval pro dvojice (a, b). Definujme ho pro $A, B \in \mathbb{N}$ následovně

$$I = I_a \times I_b = [-A; A] \times [0; B].$$

Vzhledem k tomu, že se jedná o definiční obor homogenního polynomu $F(x,y) = \sum_i a_i x^i y^{d-i} \in \mathbb{Z}[x,y]$, stačí uvažovat pouze jednu z hodnot x, y zápornou. Kdybychom uvažovali $I_b = [-B; B]$, rozšířili bychom dvojnásobně počet relací, ale nezískali bychom tím mnoho navíc, protože hodnoty $F(a,b)$ by se lišily pouze o násobek inverzním prvkem. Uvažujeme tedy oblast o velikosti $2AB$. Nastavení parametrů A a B je určováno heuristicky. Homogenní polynom $F(x,y)$ bývá sestaven tak, že vedoucí koeficient je poměrně malý a další koeficienty pak postupně narůstají. Tedy největší koeficienty bývají a_1 a a_0, které násobí členy xy^{d-1} a y^d. To nám dává, že stačí volit hodnotu B menší než A. Jinak by hodnoty $F(a,b)$ zbytečně příliš narostly a tím bychom spotřebovali více času na získání méně relací. Pro velké hodnoty y tedy neočekáváme výrazné zlepšení počtu relací.

Zkosený prosévací interval je upravený obdélníkový interval. Pomocí parametru $s = \frac{A}{B}$, který nazýváme **zkosení**, změníme hodnoty A a B na nové hodnoty $A' = \sqrt{sAB}$ a $B' = \sqrt{\frac{AB}{s}}$.

$$I = I_a \times I_b = [-A'; A'] \times [0; B'] = \left[-\sqrt{sAB}; \sqrt{sAB}\right] \times \left[0; \sqrt{\frac{AB}{s}}\right].$$

Tedy se opět jedná o obdélníkový interval o velikosti $2AB$ se zkosenými parametry A' a B'. Možnou variací je volba jiné hodnoty s podle velikostí koeficientů polynomu.

Zkosení se ukazuje výhodnější vzhledem k empirickému pozorování nalezených relací na zkoseném a nezkoseném intervalu.

Prosévací interval nemusíme nutně volit pouze v geometrickém tvaru. Je vhodné k němu připojit i další hodnoty (a,b), pro které platí, že $\frac{a}{b}$ je blízko nějakého reálného kořene polynomu $f(x)$. Hodnoty $F(a,b)$ jsou pak poměrně malé a tedy i dobře splňují B-hladkost.

Další variace a jejich zásah do číselného síta je možno ji nalézt v [7].

4.3 Vlastnosti polynomu

Vybrané polynomy velice ovlivní efektivitu prosévání. To, zda byl vybrán dobrý polynom, se plně ukáže až při samotném průběhu číselného síta. Takový způsob testování polynomů je ovšem velice neefektivní. Proto se používají různá heuristická kritéria, která vycházejí z vlastnosti, jež lze od polynomů očekávat, má-li být

výpočet v prosévacím intervalu dostatečně rychlý. Taková kritéria hodnotí polynomy podle takzvané vlastnosti koeficientů a vlastnosti kořene.

Běžný postup bývá vygenerovat mnoho polynomů s malými koeficienty. Upravit je a pak z nich vybrat nejvhodnějšího kandidáta.

4.3.1 Hodnocení koeficientů

Pro hodnocení koeficientů polynomu můžeme uvažovat normu ve smyslu největší z absolutních hodnot jeho koeficientů. Ukazuje se, že je vhodné uvažovat i to, jaká je mocnina proměnné, která tento koeficient násobí. Je proto lepší počítat se sup-normou polynomu podle Kleinjunga namísto maximální normy pro všechny koeficienty polynomu.

Definice 4.3. *Mějme polynom s reálnými koeficienty* $f(x) = \sum_{i=0}^{d} a_i x^i$ *stupně* d *a* $s \in \mathbb{R}^+$. *Pak definujeme:*

$$\sup(f, s) = \max_{0 \le i \le d} |a_i s^{i - \frac{d}{2}}|$$

$$\sup(f) = \min_{s>0} \sup(f, s)$$

Poznámka 4.4. *Ukažme, že se jedná o normu. V tomto případě mluvíme o normě vektorového prostoru koeficientů polynomu* f *nad tělesem* $\mathbb{R}$ *i přesto, že se jedná pouze o celočíselné koeficienty. Uvážíme tedy její základní tři vlastnosti: pozitivní definitnost, pozitivní homogennost a subaditivitu.*

Zřejmě vždy platí $\sup(f, s) \ge 0$ *a rovnost nastává pouze pokud všechny koeficienty polynomu* f *jsou rovny nule. Pokud vynásobíme polynom* f *nějakým racionálním číslem* $r \in \mathbb{Q}$, *platí*

$$\sup(rf, s) = \max_{0 \le i \le d} |r a_i s^{i - \frac{d}{2}}| = r \max_{0 \le i \le d} |a_i s^{i - \frac{d}{2}}| = r \sup(f, s).$$

Navíc pokud budeme počítat sup-normu pro součet dvou polynomů f *a* g. *Označme* $\max(\deg f, \deg g) = d$ *a* $f + g = \sum_{i=0}^{d} (a_i + b_i) x^i$. *Pak platí vzhledem trojúhelníkové nerovnosti*

$$\sup(f + g, s) = \max_{0 \le i \le d} |(a_i + b_i) s^{i - \frac{d}{2}}| \le \max_{0 \le i \le d} \left(|a_i s^{i - \frac{d}{2}}| + |b_i s^{i - \frac{d}{2}}| \right)$$

$$\le \max_{0 \le i \le d} |a_i s^{i - \frac{d}{2}}| + \max_{0 \le i \le d} |b_i s^{i - \frac{d}{2}}| = \sup(f, s) + \sup(g, s).$$

Zřejmě výše zavedená sup-norma je velice podobná ℓ^∞ normě. Empiricky se ukazuje jako dobrý nástroj na měření normy polynomu.

Vhodnou formou jak měřit vlastnosti koeficientů polynomu je také logaritmus z L^2-normy, se kterou přichází Bai [2]. Tato norma uvažuje eliptickou prosévací oblast. Jedná se o normu zhomogenizovaného polynomu.

$$\log L^2(F) = \frac{1}{2}\log\left(s^{-d}\int_0^{2\Pi}\int_0^1 F^2(xs, y)\,\mathrm{d}x\,\mathrm{d}y\right)$$

Bai tvrdí, že empiricky se ukazuje, že tyto dvě normy spolu vždy nekorelují. Pokud se však najde optimální zkosení s není rozdíl příliš velký. Nejlepší výsledky dávají takové polynomy, které mají obě uvedené normy malé.

4.3.2 Vlastnost kořene

Dalším měřítkem pro výběr vhodného polynomu je takzvaná vlastnost kořene. Jedná se o měření počtu kořenů polynomu modulo malá prvočísla. Vhodným ukazatelem je $\alpha(f)$ funkce definovaná v [5]. Jedná se o přínos polynomu vůči různým prvočíslům menším než zvolená mez B.

Poznámka 4.5. *Připomeňme několik elementárních pojmů z teorie pravděpodobnosti.*

- *Střední hodnota je parametr rozdělení náhodné veličiny X, který je definován jako vážený průměr daného rozdělení. Značí se $E(X)$.*
- *Pro celá čísla je p-valuace nejvyšší mocnina prvočísla p, které dělí dané celé číslo.*

Uvažujme náhodnou dvojici (a, b) z prosévacího intervalu. Potom pravděpodobnost, že $F(a, b)$ bude B-hladké, je stejná jako pro libovolný podobný polynom stejného stupně. Přesto je však možno ohodnotit polynom obecně trochu jiným způsobem.

Uvažujme nyní pouze prvočísla p, která jsou navíc menší než zvolená mez B pro B-hladkost. Mějme střední hodnotu p-valuací hodnot homogenního polynomu $F(x, y)$ na prosévacím intervalu I. Označme ji $\mathrm{E}\left(v_p(F(a, b))\right)$ pro všechna $(a, b) \in I$. Pak pro B-hladké $F(a, b)$ platí

$$\log(F(a, b)) \approx \sum_{p \le B} \mathrm{E}\left(v_p(F(a, b))\right)\log(p).$$

Nechť q_p je počet kořenů homogenního polynomu $f(x)$ modulo p na prosévací oblasti I. Pro speciální prvočísla p (definována v sekci 2.5) platí následující pro F a analogicky i pro f na jejich definičním oboru:

$$\mathrm{E}(v_p(F)) \approx \frac{\sum_{(a,b)\in I} v_p(F(a,b))}{c}, \text{ pro vhodné } c \leq |\{F(a,b); (a,b) \in I\}|.$$

Pro nespeciální prvočísla p pak máme rozlišené případy pro polynom f a jeho zhomogenizovanou verzi F

$$\mathrm{E}(v_p(f)) = \frac{q_p}{p-1},$$
$$\mathrm{E}(v_p(F)) = \frac{pq_p}{p^2-1}.$$

Podle toho je definována funkce pro ohodnocení polynomů podle rozložení jejich funkčních hodnot. Uveďme ji pro homogenní verzi polynomu F, ale stejně platí i pro f.

$$\alpha(F) = \sum_{p \leq B} (1 - (p-1)\mathrm{E}(v_p(F))) \frac{\log p}{p-1}.$$

Čím menší hodnotu má tato funkce, tím je polynom lepší pro hledání relací. Funkční hodnoty homogenního polynomu $F(x,y)$ se chovají podobně, jako náhodná čísla o velikosti $F(x,y)e^{\alpha(F)}$. Jedná se o porovnání náhodného čísla a funkčních hodnot polynomu. Hodnoty $\alpha(F)$ jsou negativní, pokud má homogenní polynom více, než jeden kořen.

Kapitola 5

Generování polynomů

První fází algoritmu je vygenerování dvou ireducibilních nesoudělných polynomů ze $\mathbb{Z}[x]$, které splňují

$$f_1(m) \equiv f_2(m) \,(\mathrm{mod}\ N)\,, \text{ kde } m \in \mathbb{Z}.$$

Nechť $\alpha_i \in \mathbb{C}$ je kořen polynom f_i pro $i = 1, 2$. Tyto kořeny použijeme k sestavení číselných těles $K_1 = \mathbb{Q}[\alpha_1]$ a $K_2 = \mathbb{Q}[\alpha_2]$. Okruhy algebraických celých čísel těchto těles hrají klíčovou roli v celém algoritmu, pro přehlednost je označme O_{K_1} a O_{K_2}.

Vždy předpokládáme, že koeficienty každého polynomu jsou nesoudělné, jinak bychom celý polynom zkrátili společným dělitelem všech koeficientů. Původně bylo požadováno i to, aby polynomy byly monické, ale tato podmínka není nutná, jak ukážeme v sekci 5.2 o nemonických polynomech. Navíc nám nemoničnost umožní zmenšit polynomy a tím i zefektivnit algoritmus.

Zatím nebyl nalezen algoritmus, který by dával vhodné polynomy vyšších stupňů. Experimentálně nebylo zjištěno zlepšení při volbě vyšších exponentů při užití známých metod generování polynomů.

Ukázalo se, že volit polynomy náhodně není optimálním řešením. Prosévací fáze pak nemusí dávat dobré relace, nebo je hledá s výrazně vyšší složitostí, jak časovou tak i paměťovou. Existují různé metody, pomocí kterých získáme polynomy s mnohem lepšími vlastnostmi a tedy i větší efektivitou celého dalšího prosévání. K nalezení takových polynomů se používají různé algoritmy. Cílem této kapitoly je popsat nejlepší známé algoritmy popsané Kleinjungem. Začneme u základního algoritmu na generování polynomů a popíšeme i další postupy, ze kterých Kleinjung vycházel.

5.1 m-adický rozvoj

Nejjednodušší metodou k nalezení polynomu, splňujícího požadavky, je metoda zvaná m-adický rozvoj. Jedná se o rozklad faktorizovaného čísla N na násobky mocnin předem zvoleného čísla m, které budeme dále uvažovat jako kořen pro oba polynomy modulo N. Nejprve se zaměřme na první polynom f_1. Mějme rozvoj

$$N = \sum_{i=0}^{d} a_i m^i,$$

pro $a_i \in \{0, 1, \ldots, m-1\}$. První polynom pak může být tvaru $f_1(x) = \sum_{i=0}^{d} a_i x^i$.

Druhý polynom f_2 zvolíme jako minimální polynom m v $\mathbb{Z}[x]$.

$$f_2(x) = x - m$$

Pak již jistě platí základní podmínka $f_1(m) \equiv f_2(m) \pmod{N}$, kde $m \in \mathbb{N}$ a přitom jsou oba polynomy nesoudělné.

První polynom f_1 můžeme ještě trochu upravit. Není nutné používat pouze kladné koeficienty. Můžeme je tedy zmenšit, pokud budeme uvažovat i záporné koeficienty. Je-li koeficient větší než $\frac{m}{2}$ zmenšíme celý koeficient o m. Pak musíme také zvětšit následující koeficient o 1. To znamená, že pokud platí $a_i > \frac{m}{2}$, pak

- zmenšíme a_i na $a_i - m$,
- zvětšíme a_{i+1} na $a_{i+1} + 1$.

Tento posun neprovádíme s vedoucím koeficientem, abychom nezvětšili stupeň polynomu. V případě monického polynomu neměníme ani následující koeficient a_{d-1}, který by nám jinak porušil moničnost. Máme tedy

$$f_1(x) = \sum_{i=0}^{d} a_i x^i, \text{ kde } a_i \in \left\{ -\frac{m}{2} - 1, \ldots, \frac{m}{2} \right\}.$$

Uvažujme prozatím pouze monické polynomy. Nemonickým polynomům se budeme věnovat později.

Číselná tělesa jsou v tomto případě tvaru $K_1 = \mathbb{Q}[\alpha_1]$ pro polynom $f_1(x)$ s kořenem $\alpha_1 \in \mathbb{C}$ a $K_2 = \mathbb{Q}$, protože polynom f_2 je lineární s kořenem m v $\mathbb{Z}$.

Veškeré výpočty pro hledané relace (a, b) budou v K_2 probíhat jako rozklady na prvočísla a jednotku -1, protože veškeré normy prvků z $O_{K_2} = \mathbb{Z}$ jsou zřejmě rovny hodnotě generujícího prvku ideálu. Tedy $\mathcal{N}_2(a - bm) = a - bm$.

Vstupem m-adického generování polynomů jsou přirozená čísla N a m. Bylo by zbytečné uvažovat m záporné. Volbou čísla m volíme i stupeň polynomu f_1. Chceme-li nejprve určit stupeň d polynomu f_1. Je třeba volit $m \approx \sqrt[d]{N}$ pro monický polynom a $m \approx \sqrt[d+1]{N}$ pro nemonický polynom.

Polynomy nalezené pouze m-adickým rozkladem sice postačují, ale nemají tak dobré vlastnosti. Číselné síto s nimi pracuje pomaleji, než s polynomy z komplexnějších metod. Ty však vycházejí z tohoto základního způsobu generování polynomů.

5.2 Užití nemonických polynomů

Ukažme, že pro algoritmus číselného síta je možno využít i nemonické polynomy. Nechť $f(x) \in \mathbb{Z}[x]$ je nemonický ireducibilní polynom stupně d. Potom již jeho kořen $\alpha \in \mathbb{C}$ nemusí být celistvý nad $\mathbb{Z}$. Těleso $\mathbb{Q}[\alpha]$ je nadtěleso konečného stupně tělesa $\mathbb{Q}$. Tedy se stále jedná o číselné těleso. Označme ho opět K. Již však okamžitě neplatí, že nutně $\alpha \in O_K$. Tedy pro normy hlavních ideálů z okruhu algebraických celých čísel O_K není splněna podmínka pro tvrzení 4.2. Pro prvek α však můžeme najít celočíselný násobek, který je již celistvý.

Ireducibilní polynom $f(x) = \sum_{i=0}^{d} a_i x^i$ vynásobíme prvek a_d^{d-1} a převedeme proměnou na $\frac{x}{a_d}$.

$$\tilde{f}(x) = a_d^{d-1} f\left(\frac{x}{a_d}\right) = x^d + \sum_{i=0}^{d-1} a_d^i a_i x^i$$

Tím jsme získali monický ireducibilní polynom $\tilde{f}(x) \in \mathbb{Z}[x]$. Jeho kořenem je $\beta = a_d\alpha$, kde víme, že $a_d \in \mathbb{Z}$. Kořeny polynomu $\tilde{f}$ jsou celistvé nad $\mathbb{Z}$ a tedy $\beta \in K$ je algebraické celé číslo. Číselné těleso $K = \mathbb{Q}[\alpha]$ lze také zapsat ve tvaru $\mathbb{Q}[\beta]$, protože ${a_d}^{-1} \in \mathbb{Q}$.

Zřejmě pro $a, b \in \mathbb{Z}$ je $(a - b\alpha)$ lomený ideál okruhu O_K. Podle příkladu 2.45 víme, že O_K je Dedekindův obor. Tvrzení 2.50 říká, že i pro lomené ideály existuje jednoznačný rozklad na prvoideály v Dedekindově oboru. Lomený ideál $(a - b\alpha)$ lze tedy jednoznačně rozložit na prvoideály z O_K. Nevyplácí se úplně přejít na výpočet

norem těchto ideálů pomocí polynomu $\tilde{f}(x)$. Koeficienty polynomu $\tilde{f}(x)$ se mohou výrazně změnit oproti koeficientům $f(x)$, což se v praxi stává běžně. Tím se výrazně mění jak kořenové tak velikostní vlastnosti. Vyplácí se počítat normu lomených ideálů podle původního polynomu. Chceme stále používat polynom $F(x, y)$. To je možné s následující úpravou.

$$F(x,y) = y^d f\left(\frac{x}{y}\right) = a_d^{1-d} y^d \tilde{f}\left(\frac{a_d x}{y}\right) = a_d^{1-d} \tilde{F}(a_d x, y).$$

Lomený ideál $(a - b\alpha)$ lze zapsat jako součin hlavního lomeného ideálu generovaného celým číslem a ideálu Dedekindova oboru.

$$(a - b\alpha) = (a_d)^{-1}(a_d a - b\beta)$$

Díky těmto dvěma úvahám a tvrzením 2.38, 2.54 a 4.2 lze již snadno zjistit normu hlavního lomeného ideálu Dedekindova oboru O_K.

$$\begin{aligned}\mathcal{N}\left((a - b\alpha)\right) &= \mathcal{N}\left(a_d^{-1}(a_d a - b\beta)\right) = \mathcal{N}\left((a_d)^{-1}(a_d a - b\beta)\right) \\ &= \mathcal{N}\left((a_d)^{-1}\right)\mathcal{N}\left((a_d a - b\beta)\right) = a_d^{-d}\tilde{F}(a_d a, b) \\ &= a_d^{-1} F(a, b)\end{aligned}$$

Použití nemonických polynomů ovlivní výrazně prosévací fázi, ale také sestavení matice. Při prosévání stačí uvažovat pouze hodnoty $F(a, b)$ pro určeí vhodných relací, které zařadíme do množiny ζ. Tyto relace však nereprezentují hlavní ideály $(a - b\beta)$, ale hlavní lomené ideály $(a - b\alpha)$. Při sestavování matice je pak třeba uvážit i prvoideály, které patří do jednoznačného rozkladu lomeného ideálu $(a_d)^{-1}$, který víme, že existuje v O_K podle tvrzení 2.50. Je podstatné nezanedbat tyto lomené ideály pro správné spárování prvoideálů z rozkladů všech relací ze ζ. Pokud bychom je zanedbali, mohlo by se stát, že pro vybrané dvojice (a, b) získáme čtverec $\prod(a_d a - b\beta)$, ale $\prod(a_d)^{-1}$ čtvercem nebude. Chceme, aby $\prod(a - b\alpha)$ byl čtverec pro vhodný výběr relací (a, b) ze ζ. Řešením je například uvažovat pouze takové výběry dvojic (a, b), které mají sudý počet prvků. Každý lomený ideál $(a - b\alpha)$ obsahuje $(a_d)^{-1}$ a při sudém počtu takových lomených ideálů zřejmě nastane i spárování prvoideálů s lichým exponentem z rozkladu všech $(a_d)^{-1}$. Více se této problematice věnuje [1].

Při generování polynomů bývá běžný postup zvolit nejprve stupeň polynomu a vedoucí koeficient. Podle nich jsou dále voleny další koeficienty polynomu. Prvním

krokem algoritmu bývá navýšení a_d na další vhodnou hodnotu, ke které bývá sestaven polynom a algoritmus se vrací na začátek. Některé hodnoty koeficientu a_d nemusí dávat dobré polynomy. Dále není cílem zvolit a_d až příliš velké.

Ukažme nyní metodu takzvaného (m, p)-adického rozvoje čísla N, kde $m, p \in \mathbb{N}$. Požadujme navíc, aby p bylo prvočíslo, tento požadavek však není nutný. Postupně zjistíme všechny koeficinty a_j pro rozklad $N = \sum_{j=0}^{d} a_j m^j p^{d-j}$. Tyto koeficienty pak využijeme následovně $f(x) = \sum_{j=0}^{d} a_j x^j$. Zřejmě platí, že $p^d f\left(\frac{m}{p}\right) = N$. Tento polynom bude reprezentovat polynom f_1. Druhý polynom bude pak lineární tvaru $f_2(x) = px - m$. Společný kořen obou nemonických polynomů modulo N je tedy zřejmě $\frac{m}{p}$. Tím modifikujeme jeden ze základních požadavků na generované polynomy. Pokud požadavek na kořen obou polynomů oslabíme z celých čísel pouze na čísla racionální, potřebujeme, aby platilo, že existuje p^{-1} modulo N. Pokud by však neexistovalo, tak jsme získali netriviálního dělitele N, což přesně chceme. Společný kořen obou polynomů modulo N tedy máme mp^{-1}. Podle toho se také opraví homomorfizmy φ_i pro $i = 1, 2$ na $\varphi_i(\alpha_i) = mp^{-1} \pmod{N}$.

Chceme, aby polynom f měl pokud možno co nejmenší koeficienty. Předtím, než začneme takový polynom generovat podle zvoleného vedoucího koeficientu, je třeba otestovat, zda má kombinace m, p, a_d řešení. Chceme získat rozklad $N = \sum_{j=0}^{d} a_j m^j p^{d-j}$ a podle něj ireducibilní polynom $f(x) = \sum_{j=0}^{d} a_j x^j$ s kořenem $\frac{m}{p}$. Pokud však p dělí a_d nejsou koeficienty polynomu f nesoudělné.

Mějme $p \in \mathbb{Z}$ prvočíslo takové, že $p < a_d$. Předpokládejme, že p nedělí a_d. Vedoucí koeficient polynomu $p^d f\left(\frac{x}{p}\right)$ zřejmě není dělitelný p, ale všechny ostatní koeficienty jsou.

$$p^d f\left(\frac{x}{p}\right) \equiv a_d x^d \pmod{p}.$$

Dosadíme-li m za x získáme: $N \equiv a_d m^d \pmod{p}$. Pokud tato kongruence nemá řešení, ireducibilní polynom f s kořenem m modulo p neexistuje.

Ukažme nyní, jak vygenerovat vhodný nemonický polynom pomocí (m, p)-adického rozkladu. Pro pevně zvolená celé číslo m a prvočíslo p hledáme rozklad $N = \sum_{i=0}^{d} a_i m^i p^{d-i}$. Definujeme nejprve pomocné rekurzivní parametry:

- $r_d = N$,
- $r_i = \frac{r_{i+1} - a_{i+1} m^{i+1}}{p}$ pro $i = d - 1 \ldots, 0$.

Při této definici platí, že:

$$N = \sum_{j=i+1}^{d} (a_j m^j p^{d-j}) + r_i p^{d-i},$$

pro $i = d-1, \ldots, 0$. Jednotlivé koeficienty a_i získáme tak, aby byla splněna kongruence $r_i \equiv a_i m^i \,(\mathrm{mod}\ p)$. Definujeme je tedy následovně

$$a_i = \frac{r_i}{m^i} + q_i,$$

kde $0 \leq q_i < p$ bude takové, aby byla splněna kongruence a $a_i \in \mathbb{Z}$.

Vyjádřeme nejprve r_i z rozvoje čísla N. Pro všechna $i = 0, \ldots, d$ máme

$$r_i p^{d-i} = N - \sum_{j=i+1}^{d} a_j m^j p^{d-j},$$

$$r_i = \sum_{j=0}^{i} a_j m^j p^{i-j}.$$

Polynom f_1 získáme jako vyjádření parametru $\frac{r_d}{p}$. Pak $f_1(x) = \sum_{j=0}^{d} a_j x^j$ s kořenem $x = \frac{m}{p}$.

Při (m,p)-adickém generování polynomů volíme vedoucí koeficient z určitého intervalu, tedy dokážeme určit jeho maximální velikost. Máme-li pevně zvolený vedoucí koeficient a_d, stupeň polynomu $\deg f(x) = d$ a kořen $\frac{m}{p}$, dovedeme určit mez pro velikosti ostatních koeficientů:

Tvrzení 5.1. *Mějme $N, d \in \mathbb{N}$ a prvočísla $a_d > p \in \mathbb{N}$. Dále mějme $\widetilde{m} = \sqrt[d]{\frac{N}{a_d}}$ reálné a m celé tak, že $m \geq \widetilde{m}$ a $N \equiv a_d m^d \,(mod\ p)$ má řešení. Potom existuje polynom $f(x) = \sum_{i=0}^{d} a_i x^i$ takový, že platí $p^d f\left(\frac{m}{p}\right) = N$ a jeho koeficienty jsou omezeny:*

- $|a_{d-1}| < p + d a_d \frac{m - \widetilde{m}}{p}$,
- $|a_i| < p + m$ *pro* $i = 0, \ldots, d-2$.

Důkaz. Existenci takového polynomu jsme právě právě ukázali při popisu generování pomocnými parametry r_i pro $i = d, \ldots, 0$.

Máme pevně zvolené a_d a m. Ostatní koeficienty polynomu máme definovány, podle způsobu konstrukce $a_i = \frac{r_i}{m^i} + q_i$, kde $0 \leq q_i < p$.

Podívejme se na omezení druhého koeficientu a_{d-1}. Zřejmě platí

$$|r_{d-1}| = \frac{1}{p}\left|N - a_d m^d\right| = \frac{1}{p}\left|a_d \widetilde{m}^d - a_d m^d\right| = \frac{a_d}{p}\left|\widetilde{m}^d - m^d\right| < \frac{a_d}{p}(m - \widetilde{m}) d m^{d-1}$$

Použijeme-li trojúhelníkovou nerovnost získáme

$$|a_{d-1}| = \left|\frac{r_{d-1}}{m^{d-1}} + q_{d-1}\right| \leq \left|\frac{r_{d-1}}{m^{d-1}}\right| + |q_{d-1}|\,.$$

Složíme-li obě nerovnosti dohromady spolu s $q_{d-1} < p$, získáme:

$$|a_{d-1}|\, m^{d-1} - |q_{d-1}|\, m^{d-1} = |r_{d-1}| < \frac{a_d}{p}(m - \widetilde{m}) d m^{d-1}$$

$$|a_{d-1}| < \frac{a_d}{p}(m - \widetilde{m}) d + p.$$

Omezení koeficientů a_i pro $i = 0, \ldots, d-2$ se dokáže obdobně

$$|r_{i-1}| = \frac{1}{p}\left|r_i - a_i m^i\right| = \frac{1}{p}\left|r_i - (\frac{r_i}{m^i} + q_i) m^i\right| = \frac{q_i m^i}{p} < m^i,$$

$$|a_i| = \left|\frac{r_i}{m^i} + q_i\right| < m + q_i < m + p.$$

□

Tedy má smysl uvažovat i nemonické polynomy.

5.3 Montgomery - Murphyho algoritmus

Popišme algoritmus Montgomery - Murphyho, ze kterého dále vychází nejpoužívanější algoritmy. Výsledné polynomy jsou tvaru $f_1(x) = \sum_{i=0}^{d} a_i x^i$ a $f_2(x) = x - m$, kde $m, a_0, \ldots, a_d \in \mathbb{Z}$.

Uvažujme mez $k \in \mathbb{N}$ pro vedoucí koeficient. Ten nejprve inicializujeme $a_d = 0$. Pak ho postupně navyšujeme až do chvíle, kdy překročíme mez k. Pak ukončíme generování. Z nalezených polynomů vybereme kandidáty podle kritérií popsaných v předchozí kapitole.

Vždy, když zvolíme nové a_d, sestavíme podle něj $m = \left\lfloor \sqrt[d]{\frac{N}{a_d}} \right\rfloor$ a následující dva koeficienty polynomu a_{d-1} a a_{d-2} pomocí m-adického rozvoje čísla N. Pokud nejsou tyto dva koeficienty dostatečně malé, zvolené a_d nevede k vhodnému polynomu. Vrátíme se zpět k hledání vedoucího koeficientu. Pokud ale vyhovují, pokračujeme jako v m-adické metodě a sestavíme celý polynom $f_1(x) = \sum_{i=0}^{d} a_i x^i$.

Tímto však generování polynomu nekončí. Dalším krokem je optimalizovat nalezené polynomy. Murphy přichází s dvěma metodami:

- **Translace**, nebo-li posunutí kořene polynomu. $f_{i,new}(x) = f_i(x + t)$, kde $i = 1, 2$ a $t \in \mathbb{N}$.
- **Rotace**, nebo-li přičtení násobku druhého polynomu k prvnímu. Nový polynom tedy získáme tímto způsobem $f_{1,new}(x) = f_1(x) + c(x).f_2(x)$, kde polynom $c(x) \in \mathbb{Z}[x]$ je menšího stupně než polynom $f_1(x)$.

Translace má vliv pouze na velikost polynomu. Mění jeho koeficienty, ale neměníme kořenové vlastnosti při lineární změně kořene o t nepříliš velké.

Při rotaci zřejmě zůstane stejný kořen, ale změníme jak velikosti koeficientů, tak i kořenové vlastnosti.

Užitím těchto dvou metod upravíme velikosti polynomů tak, aby měly lepší normu. Detailnějším popisem tohoto algoritmu se již zabývá [1], případně přímo Murphyho práce [5].

5.4 Kleinjungův algoritmus

Nyní přecházíme k popisu Kleinjungova algoritmu. Jedná se (m, p)-adickou metodu, rozšířenou o omezení všech koeficientů polynomu.

Budeme opět postupovat přes postupnou volbu vedoucích koeficientů a_d, podle kterých sestavíme polynomy. Pevně zvolené a_d tedy považujme za konstantu. Jako první pak dopočteme předběžné kořeny $\widetilde{m} = \sqrt[d]{\frac{N}{a_d}}$, ponechané v reálném tvaru. Konečný celočíselný kořen m určíme až později. Nebude řádově jinde než $\widetilde{m}$. Budeme generovat takové polynomy, jejichž poslední dva koeficienty a_0 a a_1 budou nejvýše řádově o velikosti $\widetilde{m}$.

Chceme generovat pouze takové polynomy, které nebudou mít libovolně velké koeficienty. Uvažujme sup-normu polynomu, z definice 4.3. Pokud za s z definice dosadíme zkosení prosévacího intervalu I, získáváme sup-normou informace o relativní velikosti koeficientů při používání hodnot z I. Zvolme mez M pro sup-normu

jako $\sqrt[d+1]{N}$ (nebo $\sqrt[d]{N}$ v případě monického polynomu). Větší mez je zbytečně velká, protože daný polynom by pak měl obor hodnot příliš rozsáhlý. Menší mez by naopak nemusela vést k dostatečným polynomům, pokrývající dělitele blízké odmocnině z N. Jeden z dělitelů čísla N je menší nebo roven jeho odmocnině. Tedy neexistuje-li takový dělitel, pak je N prvočíslo.

Tvrzení 5.2. *Mějme faktorizované číslo $N \in \mathbb{N}$, mez $M \in \mathbb{Z}$, zkosení s zvoleného prosévacího intervalu a polynom $f(x) = \sum_{i=0}^{d} a_i x^i \in \mathbb{Z}[x]$, pro který platí $\sup(f, s) = M$. Dále mějme odhad velkosti kořene $\widetilde{m} = \sqrt[d]{\frac{N}{a_d}}$. Nechť nastává maximum sup-normy v j-té souřadnici. Potom platí*

$$s \leq \left(\frac{M}{|a_j|}\right)^{\frac{2}{2j-d}},$$

$$s \geq \left(\frac{\widetilde{m}}{M}\right)^{\frac{2}{d-2}}.$$

Důkaz. Při daném polynomu f a zvolené mezi pro sup-normu $M \geq \sup(f, s)$ získáváme i omezení pro zkosení:

$$M \geq \sup(f, s) = \max_{0 \leq i \leq d} \left| a_i s^{i - \frac{d}{2}} \right|.$$

Nechť maximum nastává právě pro $i = j \in \{0, \ldots, d\}$, pak máme:

$$M \geq \left| a_j s^{j - \frac{d}{2}} \right|.$$

$$\frac{M}{|a_j|} \geq s^{j - \frac{d}{2}}$$

Tím okamžitě získáme omezení shora pro zkosení:

$$\left(\frac{M}{|a_j|}\right)^{\frac{2}{2j-d}} \geq s.$$

Pro omezení zdola použijeme vlastnost, že k nejhorším případům vygenerovaných polynomů patří, pokud nastane $|a_1| \approx \widetilde{m}$. Víme, že sup-norma polynomu má být menší než mez M a tedy i v případě, kdy maximum nastává právě pro hodnotu $i = 1$.

$$M \geq \left| a_1 s^{1 - \frac{d}{2}} \right|$$

V případě pro $|a_1|$ odpovídající přibližně velikostí $\widetilde{m}$ máme

$$\frac{M}{\widetilde{m}} \geq s^{1-\frac{d}{2}},$$

$$s^{d-2} \geq \left(\frac{\widetilde{m}}{M}\right)^2,$$

$$s \geq \left(\frac{\widetilde{m}}{M}\right)^{\frac{2}{d-2}}.$$

□

Koeficienty polynomu volíme řádově menší, než $\sqrt[d+1]{N}$. Není tedy třeba volit vyšší zkosení s, než zvolená mez, protože bychom tím nic nezískali.

Tvrzení 5.3. *Mějme mez $M \in \mathbb{Z}$, zkosení s zvoleného prosévacího intervalu a polynom $f(x) = \sum_{i=0}^{d} a_i x^i \in \mathbb{Z}[x]$, pro který platí $\sup(f, s) = M$. Potom pro jeho koeficienty platí omezení:*

- $|a_i| \leq M \left(\left(\frac{M}{\widetilde{m}}\right)^{\frac{2}{d-2}} \right)^{\frac{d}{2}-i} =: a_{i,max}$ *pro* $0 \leq i < d$.

Navíc platí:

- $|a_d| \leq \left(\frac{M^{2d-2}}{N}\right)^{\frac{1}{d-3}} =: a_{d,max}$.

Důkaz. Podle předchozího tvrzení 5.2 o omezení pro zkosení s již přímo dostaneme mez pro vedoucí koeficienty polynomu. Zřejmě platí, že omezení zdola je menší než omezení shora.

$$\left(\frac{\widetilde{m}}{M}\right)^{\frac{2}{d-2}} \leq \left(\frac{M}{|a_i|}\right)^{\frac{2}{2i-d}}$$

$$|a_i| \leq M \left(\frac{M}{\widetilde{m}}\right)^{\frac{2i-d}{d-2}} = a_{i,max} \text{ pro } 0 \leq i < d.$$

Rozlišme nyní případ, kdy maximum nastává pro $i = d$. Vzhledem k tomu, že $\widetilde{m} = \sqrt[d]{\frac{N}{a_d}}$, získáme omezení

$$|a_d| \leq M^{\frac{2d-2}{d-2}} \left(\frac{|a_d|}{N}\right)^{\frac{2d-d}{d(d-2)}}$$

$$|a_d|^{d-2-1} \leq \frac{M^{2d-2}}{N}$$
$$|a_d| \leq \left(\frac{M^{2d-2}}{N}\right)^{\frac{1}{d-3}} = a_{d,max}$$

□

Přejděme k heuristické části Kleinjungova algoritmu. Zavedeme pomocné hodnoty, díky kterým vygenerujeme hledaný polynom $f(x) = \sum a_i x^i$. Mějme opět podmínku, že existuje řešení pro $N \equiv a_d x^d \pmod{p}$ stejně jako v sekci 5.2 o použití nemonických polynomů. Zde pouze neuvažujme vlastnost, že p musí být prvočíslo. Zvolme $p \leq a_{d-1,max}$ jako součin různých malých prvočísel $p = \prod_{i=1}^{l} p_i$, pro která platí $p_i \equiv 1 \pmod{d}$. Vzhledem k tomu, že hledáme dělitele čísla N můžeme předpokládat $\gcd(N, p) = 1$, v opačném případě bychom při výpočtu $p^1 \pmod{N}$ získali netriviálního dělitele N.

Tvrzení 5.4. *Nechť $N, d, a_d, p \in \mathbb{N}$ tak, že $p = \prod_{i=1}^{l} p_i$, kde $p_1, \ldots, p_l$ jsou navzájem různá malá prvočísla, pro která platí $p_i \equiv 1 \pmod{d}$. Potom $N \equiv a_d x^d \pmod{p}$ buď nemá řešení, nebo jich má d^l pro $0 \leq x < p$.*

Důkaz. Nejprve ukažme proč $N \equiv a_d x^d \pmod{p_i}$ má právě d nebo 0 řešení. Pokud platí $\gcd(p, a_d N) = 1$, pak rovnice má řešení, v opačném případě nemá řešení. Uvažujme tedy pouze případ, kdy kongruence má řešení. Nutně platí $\gcd(p_i, a_d N) = 1$ pro všechna $i = 1, \ldots, l$. V takovém případě existuje $r \in \mathbb{Z}$, že $\gcd(p_i, r) = 1$ a platí

$$r a_d N \equiv 1 \pmod{p_i}.$$

Navíc zřejmě platí $\gcd(p_i, a_d) = 1$. Dosaďme vyjádření $N \equiv a_d x^d \pmod{p_i}$ a získáme

$$r a_d^2 x^d \equiv 1 \pmod{p_i}.$$

Pro hledané řešení kongruence platí $\gcd(x^d, p_i) = 1$. Můžeme tedy nastavit řešení x pro kongruenci $N \equiv a_d x^d \pmod{p_i}$ ve tvaru

$$x = c \prod_{j \neq i} p_j, \ c \in \mathbb{Z}, \ \gcd(p_i, c) = 1.$$

Počet řešení je určen počtem možných c, aby stále platilo $x < p$, tedy $c < p_i$ o kterém víme, že $p_i \equiv 1 \pmod{d}$. Zřejmě p_i jsou prvočísla a tedy $p_i > d$. Počet možných c je pak roven právě d.

Řešení pro každé p_i není řešením pro p_j, kde $i \neq j$. Tedy kongruence $N \equiv a_d x^d \pmod{p}$ má d^l řešení. □

Dále předpokládejme pouze případ, kdy řešení kongruence $N \equiv a_d x^d \pmod{p}$ existují. Potom tato řešení můžeme zapsat jako součet řešení hodnot x_{i,μ_i} pro jednotlivá p_i, kde $i \in \{1, \ldots, l\}$ a $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_l) \in \{1, \ldots, d\}^l$,

$$x_{\boldsymbol{\mu}} = \sum_{i=1}^{l} x_{i,\mu_i}.$$

Zřejmě platí $\frac{p}{p_i} | x_{i,\mu_i}$ a má cenu uvažovat pouze $0 \leq x_{i,\mu_i} < p$.

Posuňme řešení $x_{\boldsymbol{\mu}}$ k $\widetilde{m}$. Vezměme $m_0 \in \mathbb{N}$ blízko $\widetilde{m}$, takové, že p dělí m_0. Zvolme například nejmenší možné $m_0 > \widetilde{m}$. Pro řešení kongruence $N \equiv a_d x^d \pmod{p}$ blízká $\widetilde{m}$ pak platí

$$m_{\boldsymbol{\mu}} = m_0 + x_{\boldsymbol{\mu}} = \sum_{i=1}^{l} m_{i,\mu_i}.$$

Členy m_{i,μ_i} pro $1 = 2, \ldots, l$ volíme tak, že položíme

- $m_{1,\mu_1} = m_0 + x_{1,\mu_1}$ a
- $m_{i,\mu_i} = x_{i,\mu_i}$ pro $1 < i \leq l$.

Nyní se budeme věnovat druhému největšímu koeficientu a_{d-1}. Najdeme pomocný koeficient $a_{d-1,\boldsymbol{\mu}}$ pro rozklad čísla N. Mějme dané p a $m_0 + x_{\boldsymbol{\mu}}$. Definujme pomocné hodnoty $e_{i,j} \in \{0, \ldots, d-1\}$, kde $i = 1, \ldots, l$ a $j = 1, \ldots, d$, takto:

- $e_{1,j} \equiv a_{d-1,(j,1,\ldots,1)} \pmod{p}$,
- $e_{i,1} = 0$ pro $i > 1$,
- $e_{i,j} \equiv a_{d-1,(1,\ldots,1,j,1,\ldots,1)} - a_{d-1,(1,\ldots,1)} \pmod{p}$ pro $i > 1, j > 1$, kde v koeficientu $a_{d-1,(1,\ldots,1,j,1,\ldots,1)}$ se j nachází na i-tém místě vektoru.

Empiricky se při implementacích ukazuje, že je vhodné volit $e_{i,j}$ právě takto. Protože tím získáme:

$$a_{d-1,\boldsymbol{\mu}} = \sum_{i=1}^{l} e_{i,\mu_i}$$

splňující

$$a_{d-1,\boldsymbol{\mu}} m_\mu^{d-1} \equiv \frac{N - a_d m_\mu^d}{p} \pmod{p}.$$

Tedy uvedené $a_{d-1,\boldsymbol{\mu}}$ lze použít pro (m,p)-adický rozklad N podle kořene $m_{\boldsymbol{\mu}}$ a parametru p. Navíc z uvedené kongruence je možné určit pomocná $e_{i,j} \pmod p$ pro pevně zvolené $a_{d-1,\boldsymbol{\mu}}$, jak jsme je definovali výše. Zřejmě tato $e_{i,j}$ nejsou definována jednoznačně. To však není pro implementaci potřeba.

Zvolme dva vektory $\boldsymbol{\mu}$ a $\boldsymbol{\mu}'$ z $\{1,\dots,d\}^l$, které se liší pouze v jediné souřadnici. Nechť se jedná o i-tou souřadnici. Pak platí:

$$a_{d-1,\boldsymbol{\mu}} - a_{d-1,\boldsymbol{\mu}'} = \sum_{i=1}^{l} e_{i,\mu_i} - \sum_{i=1}^{l} e_{i,\mu'_i} = e_{i,\mu_i} - e_{i,\mu'_i}.$$

Dále definujme $\tilde{\boldsymbol{\mu}} \in \{1,\dots,d\}^l$, které naopak má s $\boldsymbol{\mu}$ stejnou pouze i-tou souřadnici a ve všech ostatních souřadnicích se liší. Tedy pro dané i máme $\boldsymbol{\mu}_i = \tilde{\boldsymbol{\mu}}_i \neq \boldsymbol{\mu}'_i$ a naopak $\boldsymbol{\mu}_j = \boldsymbol{\mu}'_j \neq \tilde{\boldsymbol{\mu}}_j$ pro ostatní koeficienty $j \in \{1,\dots,d\}\backslash\{i\}$.

Předpokládejme, že platí kongruence:

$$a_{d-1,\boldsymbol{\mu}} - a_{d-1,\boldsymbol{\mu}'} \equiv a_{d-1,\tilde{\boldsymbol{\mu}}} - a_{d-1,\tilde{\boldsymbol{\mu}}'} \pmod{p_k} \text{ pro všechna } k \in \{1,\dots,l\},$$

Potom druhý koeficient $a_{d-1,\boldsymbol{\mu}}$ takto definovaný splňuje uvedenou kongruenci.

Přistupme ke třetímu koeficientu $a_{d-2,\boldsymbol{\mu}}$ a určeme jeho velikost vzhledem k $m_{\boldsymbol{\mu}}$.

$$\begin{aligned}
\frac{a_{d-2,\boldsymbol{\mu}}}{m_{\boldsymbol{\mu}}} &\approx \frac{a_{d-2,\boldsymbol{\mu}}}{m_0} \approx \frac{N - a_d m_{\boldsymbol{\mu}}^d - a_{d-1,\boldsymbol{\mu}} m_{\boldsymbol{\mu}}^{d-1} p}{p^2 m_0^{d-1}} \\
&\approx \frac{N - a_d m_0^d - a_d d\,(m_{\boldsymbol{\mu}} - m_0)\, m_0^{d-1} - a_{d-1,\boldsymbol{\mu}} m_0^{d-1} p}{p^2 m_0^{d-1}} \\
&= \frac{N - a_d m_0^d}{p^2 m_0^{d-1}} - \frac{a_d d\,(m_{\boldsymbol{\mu}} - m_0) + a_{d-1,\boldsymbol{\mu}} p}{p^2} \\
&= \frac{N - a_d m_0^d}{p^2 m_0^{d-1}} - \frac{a_d d x_{\boldsymbol{\mu}}}{p^2} - \frac{a_{d-1,\boldsymbol{\mu}}}{p}
\end{aligned}$$

Pokud $\frac{a_{d-2,\boldsymbol{\mu}}}{m_{\boldsymbol{\mu}}}$ je velice blízko k celému číslu, pak můžeme získat hodnotu tohoto koeficientu $a_{d-2,\boldsymbol{\mu}}$ dostatečně malou. Postupuje se tak, že se přičítá $(px - m_{\boldsymbol{\mu}})\, x^{d-2}$ k polynomu f.

Vzhledem k aproximaci definujeme pomocné hodnoty pro $i = 1,\dots,l$ a $j = 1,\dots,d$:

- $f_0 = \frac{N - a_d m_0^d}{p^2 m_0^{d-1}}$,

- $f_{i,j} = -\frac{a_d dx_{i,j}}{p^2} - \frac{e_{i,j}}{p}$.

Tím můžeme $\frac{a_{d-2,\boldsymbol{\mu}}}{m_{\boldsymbol{\mu}}}$ přibližně zapsat jako sumu:

$$\frac{a_{d-2,\boldsymbol{\mu}}}{m_{\boldsymbol{\mu}}} \approx f_0 + \sum_{i=1}^{l} f_{i,\mu_i}.$$

5.4.1 Kleinjungův algoritmus - postup

Tímto máme všechny potřebné parametry pro vygenerování polynomu. Ukažme celý Kleinjungův algoritmus. Vstupní hodnoty jsou:

- faktorizované číslo N,
- stupeň prvního polynomu $\deg f_1(x) = d \geq 4$,
- mez M pro sup-normu,
- mez l pro počet prvočísel dělících první koeficient,
- maximální velikost těchto prvočísel p_{max}.

Nejprve nastavíme vedoucí koeficient $a_d = 0$ a získáme podmnožinu vhodných prvočísel

$$P = \{p \equiv 1 \,(\mathrm{mod}\ d)\ \text{ prvočíslo}; p < p_{max}; \gcd(p, N) = 1\}.$$

Pak postupně zvyšujeme vedoucí koeficient a_d, dokud nepřesáhne mez $a_{d,max}$ v tu chvíli algoritmus skončí.

$$a_{d,max} = \left(\frac{M^{2d-2}}{N}\right)^{\frac{1}{d-3}}$$

Máme pevně zvolený vedoucí koeficient a_d, podle kterého vygenerujeme celý polynom. Nejprve spočteme pomocný kořen $\widetilde{m} = \sqrt[d]{\frac{N}{a_d}}$ a meze pro další dva koeficienty

$$a_{d-1,max} = \frac{M^2}{\widetilde{m}},$$

$$a_{d-2,max} = \left(\frac{M^{2d-6}}{\widetilde{m}^{d-4}}\right)^{\frac{1}{d-2}}.$$

Z množiny prvočísel P vybereme její podmnožinu ke zvolenému a_d takto

$$\widetilde{P}(a_d) = \left\{ p \in P; \frac{a_d}{N} \text{ je d-tá nenulová mocnina } (\mathrm{mod}\ r) \right\}.$$

To znamená, že prvky $p \in \widetilde{P}(a_d)$ splňují, že $a_d x^d \equiv N \,(\mathrm{mod}\ p)$ má právě d řešení. Z množiny $\widetilde{P}$ pak vybereme podmnožiny $\widetilde{P'}$ s alespoň l prvky, pro které bude jejich součin menší než mez druhého nejvyššího koeficientu.

$$r = \prod_{p \in \widetilde{P'}} p \leq a_{d-1,max}$$

Pak spočteme $x_{i,j}$, m_0 a $e_{i,j}$, dále f_0 a $f_{i,j}$ jak jsme popsali v této sekci. Nastavíme $\epsilon = \frac{a_{d-2,max}}{m_0}$ a nalezneme vektory $\boldsymbol{\mu}$ splňující, že $\left| f_0 + \sum_{i=1}^{l} f_{i,\mu_i} \right|$ leží v ϵ okolí nějakého celého čísla. Tím získáme hledané polynomy.

V tomto bodě se postupuje tak, že spočteme dva seznamy

$$f_0 + \sum_{i=1}^{\left[\frac{l}{2}\right]} f_{i,\mu_i} (\mathrm{mod}\ \mathbb{Z}) \text{ a } - \sum_{i=\left[\frac{l}{2}+1\right]}^{l} f_{i,\mu_i} (\mathrm{mod}\ \mathbb{Z}).$$

Ty potom seřadíme a postupně hledáme prvky z druhého seznamu, které se nachází v ϵ-okolí prvků z prvního seznamu.

Pak se opět vrátíme ke zvyšování a_d a generování nových polynomů s vyšším vedoucím koeficientem.

Z nalezených polynomů vybereme nejvhodnější polynom pomocí ohodnocení kořenových a velikostních vlastností, které jsme popsali v předchozí kapitole. Pro několik nejlepších polynomů se spustí testovací prosévání, které určí konečného kandidáta. Toho označíme f_1 a můžeme o něm říci, že má malé první dva koeficienty, které mají největší vliv na zrychlení algoritmu.

5.5 Kleinjungův druhý algoritmus

Kleinjung ukazuje v [4], že lze zmenšit paměť potřebnou k prosévání. Je třeba hodnotu zkosení s nastavit podobnou jako hodnotu zkosení prosévací oblasti. Kleinjung navrhuje změnit prosévací oblast z obdélníkové na zkosenou, kterou jsme uvedli v sekci o prosévacím intervalu. Tato změna vede ke zmenšení sup-normy polynomu

f vzhledem k parametru s, který opět ztotožníme se zkosením. Změnou oproti předchozímu algoritmu je navíc zmenšení druhého koeficientu a_{d-1}.

Nejprve algoritmus hledá zápis čísla N do tří členů pomocí (m,p)-adického rozvoje. Obě hodnoty m a p volíme stejně, jako v předchozím Kleinjungově algoritmu.

$$N = a_d m^d + a_{d-1} m^{d-1} p + p^2 R$$

Navíc volba m a p bude nyní záviset i na tom, že chceme získat podíl $\frac{R}{m^{d-2}}$ dostatečně malý. Zřejmě se jedná o hodnotu velice blízkou třetímu koeficientu a_{d-2}.

Nejprve upravme faktorizované číslo N na $\tilde{N}$ následujícím způsobem. Mějme m a p celá čísla, kde m je blízká $\sqrt[d]{\tilde{N}}$ a p je součinem l malých prvočísel. Definujme

$$R = a_d \sum_{i=2}^{d} \binom{d}{i} m^{d-i} \left(\frac{a_{d-1}}{d a_d} p\right)^{i-2}.$$

Pak pro d-tou mocninu členu $\left(m + \frac{a_{d-1}}{d a_d} p\right)$ platí

$$a_d \left(m + \frac{a_{d-1}}{d a_d} p\right)^d = a_d m^d + a_{d-1} m^{d-1} p + p^2 R.$$

Nechť máme rozvoj čísla N podle m a p tvaru $N = \sum_{i=0}^{d} a_i m^i p^{d-i}$. Jeho první dva členy můžeme vyjádřit i pomocí předchozího

$$N = a_d \left(m + \frac{a_{d-1}}{d a_d} p\right)^d - p^2 R + \sum_{i=0}^{d-2} a_i m^i p^{d-i}.$$

Položme $\tilde{N} = d^d a_d^{d-1} N$. Tím máme

$$\tilde{N} = d^d a_d^{d-1} N = (d a_d m + a_{d-1} p)^d - d^d a_d^{d-1} p^2 R + d^d a_d^{d-1} \sum_{i=0}^{d-2} a_i m^i p^{d-i}.$$

Hodnoty d a a_d je možno považovat za konstanty pro naše výpočty, protože jsou pevně zvoleny hned v prvním kroku a během výpočtů se nemění. Naopak ostatní hodnoty jsou variabilní. Dále uvažujme pomocný kořen $\tilde{m} = d a_d m + a_{d-1} p$ a zbytek násobený p^2 položíme $\tilde{R}$ čímž získáme:

$$\tilde{N} = \tilde{m}^d - d^d a_d^{d-1} p^2 R + d^d a_d^{d-1} \sum_{i=0}^{d-2} a_i m^i p^{d-i}$$

$$\tilde{N} = \tilde{m}^d + p^2\tilde{R}.$$

Zřejmě pro malé hodnoty p platí $\tilde{m} \approx \sqrt[d]{\tilde{N}}$.

Tvrzení 5.5. *Mějme rozklad $\tilde{N}$, $\tilde{m}$ a zbytek $\tilde{R}$, jak jsme je popsali v předchozím odstavci. Pak platí odhad*

$$|a_{d-2}| \approx \frac{\left|\tilde{R}\right|}{d^2 a_d \tilde{m}^{d-2}}.$$

Důkaz. Vzhledem k nastavení v předchozím odstavci můžeme $\tilde{R}$ vyjádřit ve tvaru

$$\begin{aligned}
\tilde{R} = \frac{\tilde{N} - \tilde{m}^d}{p^2} &= \sum_{i=2}^{d} m^{d-i} p^{i-2} \left(d^d a_d^{d-1} a_{d-i} - \binom{d}{i} (a_d d)^{d-i} a_{d-1}^i \right) \\
&= \sum_{i=2}^{d} \left(\frac{\tilde{m} - a_{d-1}p}{da_d} \right)^{d-i} p^{i-2} \left(d^d a_d^{d-1} a_{d-i} - \binom{d}{i} (a_d d)^{d-i} a_{d-1}^i \right) \\
&= \sum_{i=2}^{d} \left(\tilde{m} - a_{d-1}p \right)^{d-i} p^{i-2} \left(d^i a_d^{i-1} a_{d-i} - \binom{d}{i} a_{d-1}^i \right)
\end{aligned}$$

Pak můžeme vydělit $\tilde{R}$ prvkem $d^2 a_d \tilde{m}^{d-2}$, čímž získáme

$$\begin{aligned}
\frac{\left|\tilde{R}\right|}{d^2 a_d \tilde{m}^{d-2}} &= \sum_{i=2}^{d} \frac{\left(\tilde{m} - a_{d-1}p \right)^{d-i} p^{i-2}}{\tilde{m}^{d-2}} \left(d^{i-2} a_d^{i-2} a_{d-i} - \binom{d}{i} a_{d-1}^i d^{-2} a_d^{-1} \right) \\
&= \sum_{i=2}^{d} \left(1 - \frac{a_{d-1}p}{\tilde{m}} \right)^{d-i} \left(\frac{p}{\tilde{m}} \right)^{i-2} \left((da_d)^{i-2} a_{d-i} - \binom{d}{i} a_{d-1}^i d^{-2} a_d^{-1} \right) \\
&\approx \sum_{i=2}^{d} \left(\frac{p}{\tilde{m}} \right)^{i-2} (da_d)^{i-2} a_{d-i} \\
&\approx a_{d-2} + \frac{p}{m} a_{d-3} + \left(\frac{p}{m} \right)^2 a_{d-4}.
\end{aligned}$$

Nový kořen odhadujeme hodně velký $\tilde{m} \approx \sqrt[d]{\tilde{N}} > m$ pro malé hodnoty p. Navíc koeficienty a_{d-3} a a_{d-4} odpovídají velikostí m a tedy $\frac{|\tilde{R}|}{d^2 a_d \tilde{m}^{d-2}}$ odpovídá velikostí a_{d-2}. □

5.5.1 Kleinjungův druhý algoritmus - postup

Kleinjungův druhý algoritmus postupuje následujícím způsobem. Nejprve najde vhodné $p, \tilde{m}$, podle kterých dopočte a_{d-1} a m. Pak dopočte rozvoj N podle p a m. Vstupní hodnoty jsou:

- faktorizované číslo N,
- stupeň prvního polynomu $degf_1(x) = d \geq 4$,
- vedoucí koeficient a_d polynomu f_1,
- mez P_1 pro volbu pomocných hodnot.

Ze vstupu máme okamžitě i rozšířené faktorizované číslo $\tilde{N} = d^d a_d^{d-1}$. Zvolme $\tilde{m}_0 = \left\lfloor \sqrt[d]{\tilde{N}} \right\rfloor \in \mathbb{N}$, jako celou dolní část. Pro hodnoty $q \in [P_1, 2P_1]$, spočteme všechna r, aby platilo:

$$\tilde{n} \equiv (\tilde{m_0} + r)^d \,(\mathrm{mod}\ q)\,.$$

Zjištěné r pak povýšíme na výpočet mod q^2. Nechť máme $r \equiv r' \,(\mathrm{mod}\ q)$ a platí

$$\tilde{n} \equiv (\tilde{m_0} + r')^d \,(\mathrm{mod}\ q)\,.$$

Když získáme taková r' pro všechna vhodná q, podíváme se, která různá q mají stejné hodnoty r'. Pokud nenajdeme takové r', které splňuje kongruence pro dvě různé hodnoty q, označme je q_1 a q_2, algoritmus končí a je třeba ho spustit s vyšší hodnotou a_d. Najdeme-li takové řešení, pak platí:

$$\tilde{n} \equiv (\tilde{m_0} + r')^d \left(\mathrm{mod}\ q_1^2 q_2^2\right)$$

Zvolíme $p = q_1 q_2$ a $\tilde{m} = \tilde{m_0} + r'$. Velikost p tedy závisí na nalezených hodnotách q_1 a q_2, které chceme vybrat co největší, ale samotné p je vždy menší než $16P_1^4$.

Nyní spočteme druhý koeficient a_{d-1} pomocí vztahu $\tilde{m} = da_d m + a_{d-1}p$. Pokud kongruence $a_{d-1} \equiv \frac{\tilde{m}}{p} \,(\mathrm{mod}\ da_d)$ nemá řešení, spouštíme algoritmus s vyšší hodnotou a_d. Pokud má řešení získáme koeficient a_{d-1} menší než da_d. Zřejmě chceme zvolit hodnoty p_1 a p_2 v předchozím kroku co největší právě proto, aby druhý koeficient a_{d-1} byl co nejmenší.

Hodnotu m, která se nachází blízko $\sqrt[d]{n}$. Dopočteme z uvedeného vzorce pro $\tilde{m}$.

$$m = \frac{\tilde{m} - a_{d-1}p}{da_d}$$

Nyní již dopočteme koeficienty a_i pro $i \in \{0, \ldots, d-2\}$ pomocí rozvoje N. Zřejmě jejich hodnoty budou maximálně o velikosti násobku m.

$$N = \sum_{i=0}^{d} a_i m^i p^{d-i}$$

Kleinjung dále uvádí další 'triky'. Namísto volby q_1 a q_2 ze stejného intervalu, můžeme volit $p = cq$ pro $c \in [P_1, P_2]$ a $q \in [P_2, P_3]$, kde $P_1 < P_2 < P_3 \in \mathbb{N}$. Navíc je výhodné přidat podmínku, že q je prvočíslo. Čímž zvýšíme pravděpodobnost, že $\gcd(da_d, p) = 1$. Pak se postupuje stejně jako v algoritmu. Tímto, nebo jiným omezením na volbu čísel, která dají p, zrychlíme algoritmus, protože projdeme méně násobků 2, které nepřidají nejvhodnější relace.

Kleinjung ve své práci [4] uvádí testování na několika případech, kdy pro čísla RSA 512 576 a 640 dával tento druhý algoritmus lepší polynomy, než předchozí algoritmy. Výsledky měření jsou empirické. Tento algoritmus byl použit při faktorizaci čísla RSA 768 [9] z listu RSA challenge [14].

Kapitola 6

Závěr

Generování polynomů je výchozí fáze algoritmu číselného síta. Algoritmy, které jsme popsali v předchozí kapitole jsou v dnešní době považovány za nejefektivnější. Jejich pomocí získáváme polynomy, které vedou k rychlejšímu chodu celého algoritmu. Tato zjištění jsou stále pouze na heuristické úrovni. Vzhledem k velikosti čísel, na které má smysl používat algoritmus číselného síta pro získání faktorizace, trvá průběh celého algoritmu několik let pro jedinou hodnotu čísla N. Tato skutečnost vede k tomu, že se stále jedná o řádně neprozkoumaný problém.

Pokud by se podařilo urychlit chod algoritmu. Například pro čísla používaná v RSA 1024 na méně než deset let. Bylo by nutné změnit celkový pohled na dnešní asymetrickou kryptografii. Celé číselné síto a obecně faktorizace velkých čísel tedy stále čeká na nové postupy a zlepšení stávajících metod.

Literatura

[1] L. Perutka, *Hledání optimálních strategií číselného síta*, Diplomová práce, Karlova Universita, Matematicko-fyzikální fakulta, 2009

[2] S. Bai, *Polynomial Selection for the Number Field Sieve*, Doctor Thesis, Australian National University, 2011

[3] T. Kleinjung, *On Polynomial Selection For The General Number Field Sieve*, Mamathematics of Computtion, Vol.75, No.256, 2006, 2037C2047.

[4] T. Kleinjung, *In CADO workshop on integer factorization*, INRIA Nancy, 2008. http://cado.gforge.inria.fr/workshop/slides/kleinjung.pdf

[5] B. Murphy, *Polynomial Selection for the Number Field Sieve Integer Factorisation Algorithm*, Ph.D. thesis, The Australian National University, 1999.

[6] T. Prest, P. Zimmermann, *Non-linear polynomial selection for the number field sieve*

[7] N. Coxon, *On Nonlinear polynomial selection for the number field sieve*, Cornell University, 2011.

[8] R. S. Williams, *Cubic Polynomials in the Number Field Sieve*, Master thesis, Texas Tech University, 2010.

[9] T. Kleinjung, K. Aoki, J. Franke, etc.*Factorization of a 768-bit RSA modulus version 1.4*, 2010

[10] P. L. Montgomery *Small Geometric Progressions Modulo N*, Microsoft Research, 1993 (revised 1995, 2005, 2008).

[11] A. Drápal *text k přednášce Komutativní Okruhy*

[12] P. Jedlička *studijní text - Kapitola 3 a 4*

[13] D. Stanovský, L. Barto *Počítačová algebra* matfyzpress, Praha 2011.

[14] http://www.rsa.com/rsalabs/node.asp?id=2091

[15] B. Schmidt, H. Aribowo, Hoang-Vu Dang *Iterative Sparse Matrix-Vector Multiplication for Integer Factorization on GPUs* Springer, 2011.

[16] P. L. Montgomery *A block Lanczos algorithm for finding dependencies over GF(2)* Springer, 1995.

[17] *http://projecteuclid.org/DPubS?service=UI\&version=1.0\&verb=Display\&handle=euclid.em/1047915103*