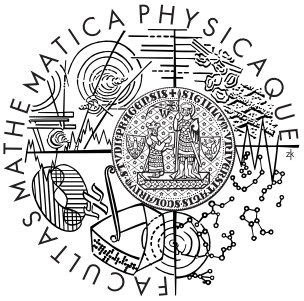
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

**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 2015

Děkuji vedoucímu mé diplomové práce Prof. Aleši Drápalovi za jeho vedení a cenné rady. Rovněž děkuji kolegům, kteří programovali ostatní části algoritmu a pomáhali mi s pochopením detailních částí algoritmu a odstraněním chyb, jmenovitě Lukáši Perutkovi, Anežce Pejlové, Janu Jeronýmu Zvánovcovi a Robertu Bashirovi.

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. Druhý Kleinjungův algoritmus byl naprogramován. 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. The second Kleinjung algoritm was also programmed.

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 [17]. Zatím největší rozložené číslo z této soutěže RSA-768 se podařilo skupině lidí okolo Thorstena Kleinjunga [11]. 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 ty, které mají vypovídající hodnotu k tématu práce. Neuvedené důkazy lze nalézt například ve skriptech o komutativních okruzích [12].

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

## 2.1 Základní pojmy

$R$-modul je každá Abelova grupa $M(+)$, na které je definované skalární násobení $M \times R \rightarrow M$ prvky $r$ tak, že zobrazení $\mu_r : M \rightarrow M$, kde $\mu_r(a) = ra$, určuje homomorfismus okruhů $R \rightarrow \mathrm{End}(M(+))$, ve smyslu $r \mapsto \mu_r$. Jádro tohoto homomorfismu označíme $\mathrm{Ann}_R(M)$ a nazveme **anihilátor** $R$-modulu $M$. Modul $M$ se nazývá **beztorzní**, pokud každé $\mu_r$ je pro $r \neq 0$ injektivní. Tento pojem má tedy smysl uvažovat pouze pro obory integrity. $R$-modul $M$ nazýváme **věrný**, pokud $\mathrm{Ann}_R(M) = 0$. $R$-modulům lze také říkat moduly nad okruhem $R$.

Podmnožina $X$ generující $R$-modulu $M$ se nazývá **volná báze**, pokud pro každý

$R$-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. Řekneme, že volný modul má konečnou hodnost, pokud v něm existuje konečná volná báze.

**Tvrzení 2.1.1.** *Mějme volný $R$-modul $M$ konečné hodnosti $n$. Pro každý podmodul $M'$ modulu $M$ existuje volná báze $\{\beta_1, \ldots, \beta_n\}$, pro kterou existují $\{a_1, \ldots, a_n\}$ tak, že podmodul $M'$ lze vyjádřit ve tvaru $M' = a_1\beta_1 R \bigoplus \ldots \bigoplus a_n\beta_n R$.*

**Poznámka 2.1.2.** *Z minulého tvrzení plyne, že podmoduly volného modulu konečné hodnosti $n$ nad obory hlavních ideálů jsou volné a mají hodnost menší nebo rovnou $n$.*

**Tvrzení 2.1.3.** *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 jej můžeme zapsat jako lineární kombinaci báze nad $\mathbb{Z}$. Je-li beztorzní Abelova grupa konečně generovaná, je možné v ní nalézt volnou bázi (často pouze bázi). Potom každý prvek takové grupy lze vyjádřit jako celočíselná kombinace prvků této báze.

Přistupme nyní od modulů k tělesům. Rozšířením těles $T \subseteq U$ rozumíme dvojici těles, kde těleso $U$ obsahuje těleso $T$. Analogicky definujeme i rozšíření okruhů. Těleso $U$ lze uvažovat jako vektorový prostor nad tělesem $T$. Dimenze takového vektorového prostoru nazýváme stupeň rozšíření těles a značíme $[U : T]$. Říkáme, že rozšíření je konečného stupně, pokud $[U : T] < \infty$. Stejně tak lze uvažovat pojem báze rozšíření ve smyslu báze vektorového prostoru $U$ nad tělesem $T$.

Mějme $T \subseteq U$, $T \subseteq W$ dvojice rozšíření těles. Pak homomorfizmus z $U$ do $W$, který je identický na $T$, nazýváme $T$-**isomorfismus**. V případě $U = W$ se jedná o $T$-automorfismus. Množinu všech $T$-isomorfismů z $U$ do $W$ budeme dále značit $\hom_T(U, W)$.

Nechť $R \subseteq S$ je rozšíření okruhů a prvek $\alpha \in S$. Všechny polynomy $f \in R[x]$ s kořenem $\alpha$ tvoří v $R[x]$ ideál. Je-li $R$ těleso, je tento ideál vždy hlavní a jako generátor můžeme uvažovat monický polynom. Takový monický polynom budeme značit $f_{\alpha,R}$ a nazývat **minimální polynom** prvku $\alpha$ v tělese $R$. Označení budeme zjednodušovat na $f_\alpha$, pokud bude zřejmé, o jaké $R$ se jedná.

Nechť $T \subseteq U$ je rozšíření těles. Prvek $\alpha \in U$ nazýváme **algebraický nad** $T$, je-li kořenem nějakého nenulového polynomu z $T[x]$. Všechny prvky $\alpha \in U$ algebraické nad $T$ tvoří podtěleso $U$, které se nazývá **algebraický uzávěr** tělesa $T$ v $U$.

Rozšíření těles $T \subseteq U$ je **algebraické**, pokud je každý prvek z $U$ algebraický nad $T$. Těleso $T$ je **algebraicky uzavřené v** $U$, pokud neexistuje $t \in U \setminus T$ algebraické nad $T$. Těleso je **algebraicky uzavřené**, pokud nemá vlastní algebraické rozšíření.

**Algebraické číslo** je každé komplexní číslo algebraické nad $\mathbb{Q}$. Množina všech algebraických čísel tvoří podtěleso $\mathbb{C}$, protože se jedná o algebraický uzávěr $\mathbb{Q}$ v $\mathbb{C}$.

**Tvrzení 2.1.4.** *Pro každé těleso $T$ existuje algebraicky uzavřené těleso $U$ takové, že $T \subseteq U$ je algebraické rozšíření těles. Dále je-li $T \subseteq T'$ algebraické rozšíření těles, existuje $T$-homomorfismus $T' \to U$. V případě, kdy je $T'$ algebraicky uzavřené je tento $T$-homomorfismus isomorfismem.*

Toto tvrzení tedy říká, že uvedené těleso $U$ je určeno jednoznačně až na $T$-isomorfismus. Je zvykem ho označovat $\bar{T}$ a nazývat **algebraickým uzávěrem** tělesa $T$. Dále budeme uvažovat vždy pouze jednu pevně zvolenou možnost $\bar{T}$.

**Příklad 2.1.5.**

- *Algebraický uzávěr tělesa $\mathbb{R}$ je těleso $\mathbb{C}$.*
- *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. Množina všech algebraických čísel je pouze podtěleso $\mathbb{C}$.*

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 těles a $M \subseteq U$, tak $T(M)$ značí nejmenší podtěleso $U$ obsahující $T \cup M$. Mějme prvek $\alpha \in U$. Připomeňme strukturu $T[\alpha] = \{g(\alpha) | g \in T[x]\}$. Je dobře známo, že $\alpha$ 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. Toto rozšíření nazýváme **jednoduché**, pokud $U = T(\alpha)$ pro nějaké $\alpha \in U$. Stupeň jednoduchého rozšíření $[T(\alpha) : T] = \deg f_{\alpha,T}$. Uvažujme bázi pro rozšíření ve smyslu báze vektorových prostorů. V takovém případě za bázi $T(\alpha)$ nad $T$ můžeme zvolit množinu $\left\{1, \alpha, \ldots, \alpha^{\deg f_{\alpha,T}-1}\right\}$.

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 $[U : T]_S$.

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

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 2.1.6.** *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 $\alpha$ a $\beta$ jsou jeho kořeny. Navíc pro každé takové dva kořeny existuje $\sigma \in \hom_T(U, \overline{U})$ takové, že $\sigma(\alpha) = \beta$.

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

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

**Hlavní tvrzení 2.1.7.** *Separabilní rozšíření konečného stupně je jednoduché.*

## 2.2 Celistvost a číselná tělesa

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. Jsou-li $R = T$ a $S = U$ tělesa, je prvek $\alpha \in U$ celistvý nad $T$, právě když je algebraický nad $T$.

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

**Tvrzení 2.2.1.** *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 $R[s] \subseteq S'$ a $S'$ je konečně generované jako $R$-modul;*
4. *existuje věrný $R[s]$-modul, který je konečně generovaný jako $R$-modul.*

**Důsledek 2.2.2.** *Mějme $R \subseteq S$ rozšíření okruhů a prvky $u_1, \ldots, u_n \in S$ celistvé nad $R$. Pak okruh $R[u_1, \ldots, u_n]$ je celistvý nad $R$ a navíc je jako $R$-modul konečně generovaný.*

**Důsledek 2.2.3.** *Mějme $R \subseteq S$ rozšíření okruhů. Pak množina všech prvků z okruhu $S$, které jsou celistvé nad $R$, tvoří podokruh $S$.*

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.

Nyní se budeme zabývat koeficienty minimálních polynomů.

**Tvrzení 2.2.4.** *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 $R$.*

**Tvrzení 2.2.5.** *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 $R[x]$.*

**Příklad 2.2.6.** *Okruh $\mathbb{Z}$ je celistvě uzavřený, protože ve svém podílovém nadtělese $\mathbb{Q}$ již nemá žádné další prvky, které by byly celistvé nad $\mathbb{Z}$.*

**Tvrzení 2.2.7.** *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 $T$.*

**Důsledek 2.2.8.** *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 $T$.*

**Příklad 2.2.9.** *Mějme okruh $R = \mathbb{Z}$, těleso $T = \mathbb{Q}$ a jeho algebraické rozšíření $U = \mathbb{Q}(i)$. Celistvý uzávěr $\mathbb{Z}$ v $\mathbb{Q}(i)$ je tvaru $S = \mathbb{Z}[i]$. Podle tvrzení 2.2.7 pro libovolný prvek $\alpha \in \mathbb{Q}(i)$ existuje jeho celočíselný násobek, který patří do $\mathbb{Z}[i]$.*

Tvrzení 2.2.7 zřejmě platí pro obecné těleso $U$. Nejen pro případy racionálních a celých čísel. Nás budou později ale zajímat právě tyto případy.

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

**Tvrzení 2.2.10.** *Množina algebraických celých čísel tvoří podokruh komplexních čísel.*

*Důkaz.* Jedná se o speciální případ důsledku 2.2.3, kdy $S = \mathbb{C}$ a $R = \mathbb{Z}$. □

**Příklad 2.2.11.** *Aplikujme tvrzení 2.2.5 na příkladu algebraických celých čísel. Nechť $R = \mathbb{Z}$, $T = \mathbb{Q}$ a $U = \mathbb{C}$. Pro prvek $\alpha \in \mathbb{C}$ celistvý nad $\mathbb{Z}$ platí, že jeho minimální polynom $f_\alpha$ 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.*

**Číselné těleso** $K$ (někdy zvané algebraické číselné těleso) je každé nadtěleso konečného stupně tělesa $\mathbb{Q}$. **Stupněm číselného tělesa**, rozumíme stupeň rozšíření $[K : \mathbb{Q}]$. Bývá zvykem předpokládat navíc, že číselné těleso je podtělesem $\mathbb{C}$. Touto konvencí se budeme dále řídit.

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 $\mathbb{Q}$.

Všechny prvky číselného tělesa jsou algebraická čísla. Není však pravda, že by všechny tyto prvky byly algebraickými celými čísly. Stejně tak neplatí, že by každé algebraické rozšíření racionálních čísel bylo číselným tělesem. Například těleso všech algebraických čísel je algebraickým rozšířením racionálních čísel. Není však číselným tělesem, protože není konečného stupně.

Číselné těleso $K$ je separabilní rozšíření $\mathbb{Q}$, protože je charakteristiky nula. Stupeň číselného tělesa je roven stupni separability podle tvrzení 2.1.6. Navíc každé číselné těleso $K$ je jednoduché rozšíření $\mathbb{Q}$ podle tvrzení 2.1.7.

Označme $O_K$ celistvý uzávěr $\mathbb{Z}$ v číselném tělese $K$. Jedná se o podmnožinu okruhu všech algebraických celých čísel. Podle důsledku 2.2.3, kdy $R = \mathbb{Z}$ a $S = K$, je $O_K$ okruh a je tedy uzavřený na operace sčítání a násobení. Okruh $O_K$ budeme nazývat **okruh celých algebraických čísel** číselného tělesa $K$.

Podle tvrzení 2.2.5 mají všechny prvky z $O_K$ minimální polynom s celočíselnými koeficienty. Dále víme, že podle tvrzení 2.2.7 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 monický $f_{\alpha,\mathbb{Z}}$ je okruh $\mathbb{Z}[\alpha]$ celistvý nad $\mathbb{Z}$. Při výpočtech se budeme však převážně pohybovat právě v $\mathbb{Z}[\alpha]$. V tomto okruhu lze v praxi počítat snáze.

**Příklad 2.2.12.** *V případě, kdy $K = \mathbb{Q}(\sqrt{c})$, pro $\sqrt{c} \notin \mathbb{Q}$, je*

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

Každá báze $K$ nad $\mathbb{Q}$ , která je obsažena v $O_K$ a je současně (volnou) bází $O_K$ nad $\mathbb{Z}$ se nazývá celistvá. Množina $\{\alpha_1, \ldots, \alpha_n\} \subseteq O_K$ je tedy volnou bází,jestliže každé $\beta \in O_K$ lze jednoznačně vyjádřit jako $\sum k_i\alpha_i$, kde $k_i \in \mathbb{Z}$.

## 2.3 Norma a číselné těleso

Pro číselné síto je norma podstatný pojem. Často bývá norma uváděna společně se stopou. Vhledem k tomu, že potřebujeme pro náš popis číselného síta 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 $\mu_\alpha$ definované $\mu_\alpha(x) = \alpha x$ pro všechna $x \in U$. Determinant lineárního multiplikativního zobrazení $\mu_\alpha$ vektorového prostoru $U$ nad $T$ není závislý na zvolené bázi. Budeme ho proto značit pouze $\det(\mu_\alpha)$. **Normou** prvku $\alpha \in U$ rozumíme hodnotu $\mathrm{N}_{U|T}(\alpha) = \det(\mu_\alpha)$.

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

**Poznámka 2.3.1.** *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 $\mu_\alpha$ 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 2.3.2.** *Nechť $T \subseteq U$ je rozšíření těles konečného stupně $n$, které je separabilní. Dále mějme prvek $\alpha \in U$. Pak platí*

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

Norma $\mathrm{N}_{U|T}(\alpha)$ je v separabilním případě rovna poslednímu koeficientu minimálního polynomu $f_{\alpha,T}$ až na případné znaménko.

**Důsledek 2.3.3.** *Mějme $\alpha \in O_K$. Pak jeho norma $\mathrm{N}_{K|\mathbb{Q}}(\alpha)$ je celočíselná.*

Podívejme se nyní na normu prvků z číselného tělesa $K$. Podle tvrzení 2.1.6 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 2.3.3 celočíselná. Vzhledem k multiplikativitě normy je zajímavé uvažovat takové prvky z $O_K$, pro jejichž normu existuje inverzní prvek.

**Tvrzení 2.3.4.** *Mějme $\alpha \in O_K$. Pak platí $\mathrm{N}_{K|\mathbb{Q}}(\alpha) = \pm 1$, právě když je $\alpha$ jednotkou v $O_K$.*

*Důkaz.* Dokažme nejprve implikaci „⇐“. 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 2.3.3 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 $\pm 1$.

Pro „⇒“ mějme prvek $\alpha \in O_K$ s normou rovnou $\pm 1$. Pak je absolutní člen minimálního polynomu $f_\alpha(x)$ roven $\pm 1$. 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 $\alpha$ je jednotkou v $O_K$. □

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

**Hlavní tvrzení 2.3.5.** *Abelovská grupa $O_K$ je volná a její hodnost je rovna $n = [K : \mathbb{Q}]$.*

*Důkaz.* Mějme bázi $\{\beta_1, \ldots, \beta_n\}$ prostoru $K$ nad $\mathbb{Q}$. Podle tvrzení 2.2.7 existuje $d > 0$ takové, že $d\beta_1, \ldots, d\beta_n \in O_K$ a tedy hodnost $O_K$ je větší nebo rovna $n$. Tím získáváme vztah:

$$\bigoplus_{i=1}^{n} d\beta_i\mathbb{Z} \subseteq O_K \subseteq \bigoplus_{i=1}^{n} \beta_i\mathbb{Z} \subset \bigoplus_{i=1}^{n} \beta_i\mathbb{Q} = K.$$

Podgrupa $O_K$ volné Abelovské grupy $\bigoplus_{i=1}^{n} \beta_i\mathbb{Z}$ je volná (tvrzení 2.1.1) a hodnosti, která nepřesahuje hodnost $\bigoplus_{i=1}^{n} \beta_i\mathbb{Z}$. Hodnost $O_K$ je tedy menší nebo rovna $n$. Čímž plyne, že je rovna právě $n$. □

Mějme $R$ celistvě uzavřený obor integrity s podílovým tělesem $T$. Dále ať $U$ je separabilní rozšíření $T$ konečného stupně $n$ a $S$ celistvý uzávěr $R$ v $U$. Mějme $\{\alpha_1, \ldots, \alpha_n\} \in S$ bázi $U$ nad $T$. Tato báze se nazývá **celistvá**, jestliže každý prvek $s \in S$ lze jednoznačně vyjádřit ve tvaru $s = \sum r_i\alpha_i$, kde $r_i \in R$. Tedy celistvá báze $S$ je vlastně volná báze $S$, pokud $S$ chápeme jako $R$-modul.

Dosud jsme pracovali pouze s normou prvku, rozšiřme tento pojem na normu ideálu. Mějme okruh $O_K$ a jeho ideál $I$. **Normou** ideálu $I$ okruhu $O_K$ rozumíme index $|O_K/I| = \mathcal{N}(I)$. Z tvrzení 2.3.5 plyne konečnost tohoto výrazu.

**Příklad 2.3.6.** *Mějme $a \in \mathbb{Z}$ nenulové. Normu ideálu $(a)$ okruhu $O_K$ spočteme podle definice $\mathcal{N}((a)) = |O_K/(a)|$.*

*Podle tvrzení 2.3.5 víme, že $O_K$ je volná abelovská grupa hodnosti $n$. Mějme bázi $B_1 = \{\beta_1, \ldots, \beta_n\}$ této abelovské grupy, $O_K = \bigoplus_{i=1}^{n} \beta_i\mathbb{Z}$. Ideál $(a)$ má stejnou hodnost jako $O_K$, protože má volnou bázi $B_2 = \{a\beta_1, \ldots, a\beta_n\}$. Tedy $(a) = \bigoplus a\beta_i\mathbb{Z}$.*

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

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

Zjistit počet prvků okruhu faktorizovaného podle ideálu 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 $O_K$. Ta nám usnadní další práci s hlavními ideály tohoto okruhu.

**Hlavní tvrzení 2.3.7.** *Mějme hlavní ideál $I$ okruhu $O_K$ generovaný nenulovým prvkem $a \in O_K$. Pak $\mathcal{N}(I) = \left|\mathrm{N}_{K|\mathbb{Q}}(a)\right|$.*

*Důkaz.* Je-li $B_1 = \{\beta_1, \ldots, \beta_n\}$ volná báze Abelovské grupy $O_K$, tak podle tvrzení 2.3.5 je $n = [K : \mathbb{Q}]$. Také $B_2 = \{a\beta_1, \ldots, a\beta_n\}$ je volnou bází Abelovské grupy $I = aO_K$. Podle tvrzení 2.1.1 lze $B_1$ zvolit tak, aby pro vhodná $d_1 \geq d_2 \geq \ldots \geq d_n \geq 1$ bylo $B_3 = \{d_1\beta_1, \ldots, d_n\beta_n\}$ volnou bází $I$. Máme $d_n \geq 1$, neboť ze tvaru $B_2$ plyne, že $I$ má hodnost $n$.

Uvažujme matice přechodu mezi těmito bázemi. Protože $B_2$ i $B_3$ jsou volné báze $I$, jsou matice přechodu $[id]_{B_2}^{B_3}$ a $[id]_{B_3}^{B_2}$ celočíselné. Tyto matice jsou vzájemně inverzní, proto jsou jejich determinanty rovny $\pm 1$.

Matice $[id]_{B_1}^{B_3}$ je diagonální matice, s prvky $d_1, \ldots, d_n$ na hlavní diagonále. Matice $[id]_{B_1}^{B_2}$ je matice zobrazení $\mu_a$, jejíž determinant je roven $|O_K/I|$.

Z rovnosti $[id]_{B_2}^{B_3} = [id]_{B_2}^{B_1}[id]_{B_1}^{B_3}$ plyne, že $\left|\det[id]_{B_1}^{B_3}\right| = \left|\det[id]_{B_1}^{B_2}\right|$. Odtud již okamžitě dostáváme dokazované tvrzení, protože $\left|\det[id]_{B_1}^{B_2}\right| = |O_K/I| = \left|\mathrm{N}_{K|\mathbb{Q}}(a)\right|$. □

## 2.4 Dedekindovy obory

**Dedekindův** obor $D$ je obor integrity, pro který platí, že:

- je celistvě uzavřený,
- je noetherovský,

- každý jeho nenulový prvoideál je maximální.

Uvažujme dále pouze nenulové ideály.

**Příklad 2.4.1.** *Okruh celých čísel $\mathbb{Z}$ je Dedekindův obor. Všechny jeho ideály jsou tvaru $(n)$, kde $n \in \mathbb{Z}$. Pro prvoideály se pak jedná o prvočísla.*

**Tvrzení 2.4.2.** *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 2.4.3.** *Buď $D$ Dedekindův obor a $T$ jeho podílové těleso. Pro libovolný vlastní nenulový ideál $I$ oboru $D$ existují nenulové prvoideály $P_1, \ldots, P_k$ v oboru $D$ tak, že*

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

**Hlavní tvrzení 2.4.4.** *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.

Následující tvrzení je klíčové pro celou práci. Jedná se o okruh algebraických celých čísel.

**Tvrzení 2.4.5.** *Okruh algebraických celých čísel $O_K$ číselného tělesa $K$ je Dedekindův obor.*

*Důkaz.* Podle tvrzení 2.4.2 stačí, že $\mathbb{Z}$ je Dedekindův obor v $\mathbb{Q}$. Potom celistvý uzávěr $\mathbb{Z}$ v číselném tělese $\mathbb{Q} \subseteq K$ je Dedekindův obor.

□

**Poznámka 2.4.6.** *Pro libovolný nenulový ideál $I$ okruhu $O_K$ platí $I \cap \mathbb{Z} \neq 0$, protože ideál $I$ je hodnosti rovné $n = [K : \mathbb{Q}]$ a má volnou bázi.*

*Je-li $I$ navíc prvoideál, pak $I \cap \mathbb{Z}$ je prvoideál a platí, že 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 2.4.7.** *Pro prvoideál $P$ okruhu $O_K$ platí, že existuje prvočíslo $p$ takové, že $P \cap \mathbb{Z} = p\mathbb{Z}$.*

*Důkaz.* Průnik prvoideálu s podokruhem je vždy prvoideál. Prvoideály okruhu $\mathbb{Z}$ jsou právě hlavní ideály $p\mathbb{Z}$, kde $p$ je prvočíslo. □

Nyní přejděme od ideálů ke konečně generovaným modulům. Mějme $R$ obor integrity a $T$ jeho podílové těleso. **Lomený ideál** $J$ nazýváme konečně generovaný $R$-podmodul tělesa $T$. Tedy lomený ideál je ideálem oboru $R$ právě v případě, kdy je jeho podmnožinou. Z definice Dedekindových oborů je každý ideál lomený.

**Tvrzení 2.4.8.** *$J$ je lomený ideál tělesa $T$, právě když existuje $a \in R\backslash\{0\}$ tak, že $aJ$ je konečně generovaný ideál v $R$.*

Nechť $T$ je podílové těleso Dedekindova oboru $R$. Mějme ideál $I$ oboru $R$. Položme

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

**Tvrzení 2.4.9.** *Pro lomený ideál $J$ Dedekindova oboru $D$ je také $J^{-1}$ lomený ideál.*

**Hlavní tvrzení 2.4.10.** *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 $n_1, \ldots, n_k \in \mathbb{Z}\backslash\{0\}$.*

**Důsledek 2.4.11.** *Lomené ideály Dedekindova oboru tvoří volnou grupu vůči operaci násobení ideálů. Bází této grupy jsou všechny nenulové prvoideály.*

Uvažujme, že hledáme rozklad lomeného ideálu okruhu $O_K$ na prvoideály. Norma každého prvku $\alpha \in O_K$ je podle důsledku 2.3.3 celočíselná. Tuto vlastnost dále využijeme.

Rozklad ideálů v Dedekindově oboru je zajímavý i při počítání jeho normy. Zaměříme se na rozklad normy hlavních ideálů. Rozklad normy nám totiž dává informaci o tom, které prvoideály patří do rozkladu daného ideálu.

**Tvrzení 2.4.12.** *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.* Každý prvoideál $P$ je v Dedekindově oboru maximální, proto je $R/P$ těleso. Vzhledem k jednoznačnosti rozkladů na prvoideály v Dedekindově oboru nutně platí $P^{i+1} \subsetneq P^i$ pro každé $i \in \mathbb{N}$. Tedy $P^i/P^{i+1}$ je modulem nad $R$, pro který platí, že $P\,(P^i/P^{i+1}) = 0$. Proto je $P^i/P^{i+1}$ vektorovým prostorem nad $R/P$. Každý podprostor prostoru $P^i/P^{i+1}$ je tvaru $I/P^{i+1}$, kde $P^{i+1} \subseteq I$ je lomený ideál v $R$. Protože mezi $P^i$ a $P^{i+1}$ neleží žádný další ideál, nemá $P^i/P^{i+1}$ vlastní podprostor. Je tedy dimenze 1.

□

**Tvrzení 2.4.13.** *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.* Z definice $\mathcal{N}(I) = |O_K/I|$. Podle rozkladu ideálu $I$, Čínské věty o zbytcích a Lagrangeovy věty máme postupně rovnosti:

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

Je možné $P_i^{j-1}/P_i^j$ uvažovat jako vektorový prostor dimenze 1 nad tělesem $O_K/P_i$ vzhledem k předchozímu tvrzení 2.4.12. Tedy $\left|P_i^{j-1}/P_i^j\right| = |O_K/P_i| = \mathcal{N}(P_i)$. Tím již rovnou plyne

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

□

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

**Tvrzení 2.4.14.** *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 $r_1, \ldots, r_n, s_1, \ldots, s_n \in \mathbb{N}$. 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 $f_\alpha$ stupně $n$ rovnému stupni číselného tělesa. Pokud v této sekci mluvíme o ideálech, jedná se o ideály okruhu $O_K$.

Mějme prvočíslo $p$ takové, že $P \cap \mathbb{Z} = p\mathbb{Z}$, kde $P$ je prvoideál $O_K$. Potom říkáme, že se jedná o prvoideál **nad prvočíslem** $p$. Vzhledem k předchozím úvahám je $O_K/P$ konečné těleso charakteristiky $p$, právě když $P$ je prvoideál nad prvočíslem $p$.

**Tvrzení 2.5.1.** *Pro každé prvočíslo $p$ existuje prvoideál $P$ okruhu $O_K$ takový, že norma $\mathcal{N}(P)$ je rovna mocnině $p$.*

*Důkaz.* Podle příkladu 2.3.6 víme, že existuje ideál s normou rovnou mocnině $p$. Využijeme-li tvrzení 2.4.13 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 2.3.3. 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 $1 \leq k \leq n$. □

Pro prvoideál $P$ mějme jeho normu $\mathcal{N}(P) = p^k$. Hodnotu $k$ nazýváme **stupeň inerce**, nebo také **stupeň nehybnosti**. Její význam odpovídá $|O_K/P : \mathbb{Z} + P/P|$. V tomto případě platí $\mathbb{Z} + P/P \cong \mathbb{Z}/p\mathbb{Z}$. Uvažujme nyní rozklad hlavního ideálu nad prvočíslem $p$

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

Pokud máme prvoideál $P$ nad prvočíslem $p$, pak je již v tomto rozkladu, protože platí $(p) \subseteq P$ a protože pracujeme v Dedekindově oboru. Exponent $r_i$ z rozkladu ideálu $(p)$ nazýváme **ramifikační index** $P_i$ nad $p$, nebo také **index větvení**.

**Hlavní tvrzení 2.5.2** (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 2.4.13 a 2.5.1 víme, že pro výše zavedené značení 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 2.3.6 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 umět rozkládat hlavní ideály okruhu $O_K$ na prvoideály. Získat okruh $O_K$ by bylo v praxi velice složité. Proto budeme pracovat v $\mathbb{Z}[\alpha]$. Z úvah na konci sekce 2.2 víme, že pro každé celistvé $\alpha \in O_K$ je velikost faktorokruhu $|O_K/\mathbb{Z}[\alpha]|$ konečná. Pro výpočty v $\mathbb{Z}[\alpha]$ bude potřeba rozlišit 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. Rozkladům nad speciálními prvočísly se v této práci věnovat nebudeme.

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

**Tvrzení 2.5.3.** *Buď $c \in \mathbb{R}^+$. Pak existuje pouze konečně mnoho ideálů $I$ Dedekindova oboru $O_K$ takových, že $1 < \mathcal{N}(I) \leq c$.*

*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 2.4.7 existuje prvočíslo $p$, že $P \cap \mathbb{Z} = p\mathbb{Z}$. Norma takového prvoideálu je pak rovna mocnině $p$, jak ukazujeme v důkazu tvrzení 2.5.1. Podle Fundamentální rovnosti 2.5.2 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 2.5.4.** *Pro libovolné prvočíslo $p$ existuje pouze konečně mnoho prvoideálů $P$ oboru $O_K$ takových, že $P \cap \mathbb{Z} = p\mathbb{Z}$. Uvažujme primární dekompozici ideálu $(p)$. 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 2.5.5.** *Ať $\alpha \in \mathbb{C}$ je algebraické celé číslo, $K = \mathbb{Q}[\alpha]$ a prvočíslo $p$ nedělí $O_K/\mathbb{Z}[\alpha]$ (tedy je nespeciální). Potom $O_K = \mathbb{Z}[\alpha] + (p)$. Speciálně $O_K = \mathbb{Z}[\alpha] + I$ pro každý ideál $I \subseteq O_K$ takový, že $p \in I$.*

*Důkaz.* Položme $q = |O_K/\mathbb{Z}[\alpha]|$. Pro $q = 1$ tvrzení okamžitě platí. Uvažujme tedy $q > 1$. Mějme libovolný prvek $\beta \in O_K$. Pro něj platí $q\beta \in \mathbb{Z}[\alpha]$. Podle předpokladu jsou čísla $p$ a $q$ nesoudělná a tedy pro ně existují koeficienty $r, s \in \mathbb{Z}$ takové, že $pr + qs = 1$. Proto $\beta = sq\beta + pr\beta \in \mathbb{Z}[\alpha] + (p)$.

□

Mějme $p$ nespeciální prvočíslo. Buď $f_\alpha = \prod_{i=1}^{k} {s_i}^{e_i} \pmod p$ rozklad na navzájem nesoudělné ireducibilní polynomy. Ať jsou polynomy $t_i(x) \in \mathbb{Z}[x]$ zdvižené polynomy $s_i(x) \in \mathbb{Z}_p[x]$, ve smyslu Henselova zdvihnutí. Pro zjednodušení dále budeme používat značení $f_\alpha = \prod_{i=1}^{s} {t_i}^{e_i} \pmod p$. Polynomy $t_i$ nejsou obecně dány jednoznačně. Stačí nám pouze jednoznačnost modulo $p$.

**Tvrzení 2.5.6.** *Mějme $\alpha \in \mathbb{C}$ celistvé nad $\mathbb{Z}$, číselné těleso $K = \mathbb{Q}[\alpha]$, minimální polynom $f_\alpha$ a jeho rozklad $f_\alpha = \prod_{i=1}^{s} {t_i}^{e_i} \pmod p$, kde $p$ je nespeciální prvočíslo a kde $t_1(x), \ldots, t_s(x) \in \mathbb{Z}[x]$.*

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

*Důkaz.* Definujme homomorfismus okruhů $\varphi : \mathbb{Z}[x] \to O_K/P_i$ tak, že $\varphi(x) = \alpha + P_i$. Protože $p \in P_i$, je $p\mathbb{Z}[x] \subseteq \operatorname{Ker}(\varphi)$. Proto lze definovat homomorfismus okruhů $\psi : \mathbb{Z}_p[x] \to O_K/P_i$ tak, že $\psi(x) = \alpha$. Tedy $\operatorname{Im}(\psi) \supseteq (\mathbb{Z}[\alpha] + P_i)/P_i = O_K/P_i$, podle tvrzení 2.5.5. Máme $t_i \in \operatorname{Ker}\psi$. Ideál $(t_i)$ je maximální. $\operatorname{Ker}(\psi)$ je buď rovno tomuto ideálu, nebo celému $\mathbb{Z}_p[x]$. Proto je okruh $O_K/P_i \cong \mathbb{Z}_p[x]/\operatorname{Ker}(\psi)$ buď triviální (tedy $P_i = O_K$), nebo je to těleso isomorfní $\mathbb{Z}_p[x]/(t_i)$. Je-li $O_K/P_i$ těleso, musí $P_i$ být maximální ideál a tedy prvoideál.

□

**Hlavní tvrzení 2.5.7.** *Mějme $\alpha \in \mathbb{C}$ celistvé nad $\mathbb{Z}$, číselné těleso $K = \mathbb{Q}[\alpha]$ a okruh $O_K$. Dále mějme nespeciální prvočíslo $p$ a rozklad polynomu na ireducibilní*

*polynomy* $f_\alpha = \prod_{i=1}^{k} {t_i}^{e_i} \pmod p$, *kde* $t_1, \ldots, t_k \in \mathbb{Z}[x]$. *Potom pro rozklad hlavního ideálu* $(p)$ *okruhu* $O_K$ *na prvoideály platí:*

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

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

*Důkaz.* Mějme $\pi : \mathbb{Z}[x] \to \mathbb{Z}_p[x]$ projekci modulo prvočíslo $p$, tedy $\pi(\sum a_i x^i) = \sum (a_i \pmod p) x^i$ a $\pi' : O_K \to O_K/(p)$ projekci modulo prvoideál $(p)$, tedy $\pi'(\alpha) = \alpha + (p)$.

Definujme homomorfismus okruhů $\varphi : \mathbb{Z}[x] \to O_K$ tak, že $\varphi(x) = \alpha$. Protože $\alpha \in \mathbb{C}$ je algebraické celé číslo, je $\operatorname{Ker}\varphi = (f_\alpha) = f_\alpha\mathbb{Z}$. Jistě $\varphi(p\mathbb{Z}[x]) = (p) = pO_K$. Protože $\operatorname{Ker}(\pi'\varphi) = \varphi^{-1}((p))$, je $\operatorname{Ker}(\pi'\varphi) = \operatorname{Ker}\varphi + p\mathbb{Z}[x] = (f_\alpha) + p\mathbb{Z}[x]$. Protože $\operatorname{Im}(\varphi) = \mathbb{Z}[\alpha]$, je $\pi'\varphi$ surjektivní podle tvrzení 2.5.5. Jelikož $p\mathbb{Z}[x] = \operatorname{Ker}\pi \subseteq \operatorname{Ker}(\pi'\varphi)$, existuje podle Věty o homomorfismu homomorfismus $\psi : \mathbb{Z}_p[x] \to O_K/(p)$ takový, že $\psi\pi = \pi'\varphi$.

Je $\psi(\sum a_i x^i) = \sum a_i \alpha^i$ pro každé $a = \sum a_i x^i \in \mathbb{Z}_p[x]$. Homomorfismus $\psi$ je surjektivní, protože $\pi'\varphi$ je surjektivní. Proto je každý maximální ideál v $O_K/(p)$ obrazem nějakého maximálního ideálu v $\mathbb{Z}_p[x]/\operatorname{Ker}\psi \cong O_K/(p)$, tedy nějakého maximálního ideálu $\mathbb{Z}_p[x]$, který obsahuje $\operatorname{Ker}\psi$.

Protože $f_\alpha \in \operatorname{Ker}(\pi'\varphi)$, je $\pi(f_\alpha) \in \operatorname{Ker}\psi$. Máme $\pi(f_\alpha) = \prod_{i=1}^{k} {s_i}^{e_i}$, kde je $\pi(t_i) = s_i$.

$\operatorname{Ker}\psi$ je hlavní ideál $\mathbb{Z}_p[x]$. Proto existuje monický polynom $g \in \mathbb{Z}_p[x]$ takový, že $\operatorname{Ker}(\psi) = (g)$. Přitom $g$ dělí $\prod_{i=1}^{k} {s_i}^{e_i}$. Po případné změně indexů můžeme tedy předpokládat, že $g = \prod_{i=1}^{k'} {s_i}^{d_i}$, kde $1 \le k' \le k$ a $1 \le d_i \le e_i$.

Struktura maximálních ideálů okruhu $\mathbb{Z}_p[x]/(g)$ kopíruje strukturu maximálních ideálů $\mathbb{Z}_p[x]$, které obsahují $(g)$. Jsou to tedy právě ideály $(s_j)/(g)$, $1 \le j \le k'$. Zobrazení $\psi$ je surjektivní, proto je podle První věty o isomorfismu $\mathbb{Z}_p[x]/(g) \cong O_K/(p)$. Homomorfismus zobrazuje maximální ideály na maximální ideály, přičemž $\psi((s_i)) = P_i$, $1 \le i \le k'$. Proto $P_i/(p)$, $1 \le i \le k'$, jsou všechny maximální ideály okruhu $O_K/(p)$, a tedy $P_i$, $1 \le i \le k'$, jsou všechny maximální ideály $O_K$, které obsahují $(p)$.

Dále platí, že pro $x_i \ge 0$ je $\prod (s_i)^{x_i} \subseteq (g)$ právě když $x_i \ge d_i$, $1 \le i \le k'$. Proto je $\prod P_i^{x_i} \subseteq (p)$, právě když $x_i \ge d_i$. Protože $O_K$ je Dedekindův, platí $\prod P_i^{d_i} = (p)$, $1 \le i \le k'$. Jelikož $\psi((s_i)) = P_i$, $1 \le i \le k'$, tak je $\mathbb{Z}_p[x]/(s_i) \cong O_K/P_i$. Protože $\mathbb{Z}_p[x]/(s_i)$ je těleso řádu $p^{\deg(t_i)}$, je stupeň nehybnosti ideálu $P_i$

roven $\deg(t_i)$. Fundamentální rovnost 2.5.2 říká, že $n = [K : \mathbb{Q}] = \sum_{i=1}^{k'} d_i \deg(t_i)$. Současně máme $n = \deg(f_\alpha) = \sum_{i=1}^{k} e_i \deg t_i$. Z $d_i \leq e_i$ plyne, že $k' = k$ a $d_i = e_i$, pro všechna $1 \leq i \leq k$.

□

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.

Mějme v celé kapitole číslo $N \in \mathbb{N}$, které není prvočíslo. Uvažujme okruh $0, 1, \ldots, N-1$ s operacemi modulo $N$, který budeme značit $\mathbb{Z}_N$. Vzhledem k tomu, že $N$ není prvočíslo, existují v okruhu $\mathbb{Z}_N$ netriviální dělitelé nuly. Právě takové netriviální dělitele nuly hledáme.

Základní princip číselného síta je Fermatova faktorizace $N$. Jejím cílem je najít dvě různá celá čísla $x$ a $y$, jejichž čtverce se liší o celočíselné násobky čísla $N$. Předpokládejme, že platí $x > y \in \mathbb{N}$ a

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

Pokud rozdíl $x - y$ není roven jedné, využijeme rozkladu

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

Nastane buď $\gcd(x-y; N) > 1$ a nebo $\gcd(x+y; N) > 1$. Máme velkou šanci, že při zjištění největšího společného dělitele těchto dvojic 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 je násobkem víc prvočísel, získáme pouze rozklad na dva celočíselné dělitele $N$.

Existuje mnoho algoritmů a postupů k nalezení vhodných dvojic $x$ a $y$. Míra jejich efektivity je závislá na velikosti čísla $N$. Čím větší prvočísla násobíme, tím

je náročnější najít jejich rozklad zpět, nebo najít dobrá čísla $x$ a $y$. V této kapitole popíšeme, jak získává hodnoty $x$ a $y$ algoritmus číselného síta.

## 3.1 Stručný přehled celého algoritmu

Cílem celého algoritmu je najít vhodnou dvojici $(x, y)$ tak, aby se dala použít k Fermatově faktorizaci čísla $N$. Číselné síto při hledání dvojic $(x, y)$ nezůstává pouze v oboru celých čísel $\mathbb{Z}$, do kterého patří jak $N$, tak i jeho hledaní dělitelé. V algoritmu vytvoříme dva čtverce ve dvou algebraických strukturách, které jsou si rovny při zobrazení do $\mathbb{Z}_N$. Bude třeba vhodně nastavit tyto algebraické struktury a zobrazení mezi nimi.

Ukazuje se, že vhodnými strukturami jsou číselná tělesa. Mějme číselná tělesa tvaru $K_1 = \mathbb{Q}(\alpha_1)$ a $K_2 = \mathbb{Q}(\alpha_2)$. Hodnoty $\alpha_1, \alpha_2 \in \mathbb{C}$ jsou určeny jako kořeny dvou vhodně zvolených ireducibilních polynomů $f_1$ respektive $f_2$. Pojednáváme je jako algebraická celá čísla, tedy uvažujeme polynomy $f_i$ monické, ale v dalších kapitolách ukážeme, proč v praxi není tato vlastnost nutná. Výpočty pak probíhají způsobem, který bere v úvahu, že popsaný model se vztahuje pro vhodný celočíselný násobek kořene příslušného nemonického polynomu. Tato úvaha nám dává zobrazení $\mathbb{Z}[x] \to \mathbb{Z}[\alpha_i]$, pro $i = 1, 2$, kde $\psi_i : x \to \alpha_i$.

Nyní chceme zobrazit $\mathbb{Z}[\alpha_i] \to \mathbb{Z}_N$. Přidáme požadavek, aby oba polynomy $f_1$ a $f_2$ měly společný kořen modulo $N$. Označme tento kořen $m$. Důvodem je, jak ukážeme dále, že pak při návratu do $\mathbb{Z}$ dobře získáme hodnoty kongruentní modulo $N$. Tím můžeme definovat zobrazení $\varphi_i : \mathbb{Z}[\alpha_i] \to \mathbb{Z}_N$ určené $\varphi_i(\alpha_i) = m$, pro $i = 1, 2$. V první části algoritmu tedy najdeme polynomy $f_1$ a $f_2$, které určí číselná tělesa.

V druhé části pracujeme 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 3.1.1.** *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 $\alpha$. Zvolíme hodnotu $B$ pro $B$-hladkost a oblast, ze které budeme volit dvojice $(a, b)$. Tyto dvojice reprezentují hlavní ideály $(a - b\alpha)$

okruhu $O_K$. Nás budou zajímat takové ideály $(a-b\alpha)$, které mají $B$-hladkou normu ideálu. V algoritmu hledáme dvojice $(a, b)$, které mají $B$-hladkou hodnotu $F(a, b)$. V následující kapitole dokážeme, že norma hlavního ideálu $(a - b\alpha)$ v $O_K$ je rovna právě $F(a, b)$.

Postup, kterým vybíráme a určujeme vhodné dvojice $(a, b)$, nazýváme *prosévání*. 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 $\zeta$. Její prvky nazývejme *relace*. Relace $(a, b)$ přímo určují hlavní ideály $(a-b\alpha)$ okruhu $O_K$. Podle tvrzení 2.4.5 je $O_K$ Dedekindův obor. Každý jeho ideál se jednoznačně rozkládá na prvoideály, tvrzení 2.4.4. Ideály $(a - b\alpha)$ budeme v dalších fázích rozkládat v $O_K$ na prvoideály.

V další části budeme hledat podmnožinu $\zeta'$ množiny $\zeta$. Vybrané prvky ze $\zeta'$ určí, které ideály $(a - b\alpha)$ můžeme spolu vynásobit tak, abychom získali čtverec ve smyslu, kdy každý prvoideál z jeho rozkladu v $O_K$ bude v sudé mocnině. Označme $B$-hladkým ideálem každý ideál okruhu $O_K$, jehož norma je $B$-hladká. Tvrzení 2.5.3 nám dává, že takových ideálů je pouze konečný počet. Vytvoříme velkou řídkou matici $\mathbf{M}$. Každý řádek je určen jednou relací ze $\zeta$ a reprezentuje tak hlavní ideál okruhu $O_K$. Sloupce reprezentují $B$-hladké prvoideály okruhu $O_K$. Vzhledem k tvrzení 2.4.13, sekci 2.5 a tomu, že $O_K$ je Dedekindův obor, máme představu, jak připravit všechny prvoideály potřebné pro rozklad ideálů určených relacemi ze $\zeta$. Věta 2.5.7 určuje strukturu těchto prvoideálů a tvar hlavních ideálů $(a - b\alpha)$, které budeme rozkládat. Vzhledem k volbě polynomu $f_\alpha$ získáváme lineární polynomy $t_i$, definované v 2.5.7. Navíc v praxi zanedbáváme speciální prvočísla, protože se jim algoritmus de facto vyhýbá.

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

Podmnožina $\zeta'$ je z množiny $\zeta$ vybrána následovně. Po vynásobení všech ideálů $(a - b\alpha)$ určených relacemi ze $\zeta'$, získáme ideál, jež má všechny prvoideály ze svého rozkladu v sudé mocnině. Jedná se 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 $\zeta$, vybraných prosévání, 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 $O_K$. Je součin hlavních ideálů v $O_K$ hlavní ideál? I když je hlavní, tak ještě z daných dat nutně nevyplývá, že má generátor čtverec. Existuje postup využívající kvadratické charaktery, díky kterému lze získat, že součin ideálů určených relacemi ze $\zeta'$ je gene-

rován čtvercem. Následně získáme 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.

$$f_1(m) \equiv f_2(m) \equiv 0 \,(\mathrm{mod}\ N)$$

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 $f_i(x)$. Důležité pro další výpočty jsou okruhy $O_{K_i}$ a hlavně okruhy $\mathbb{Z}[\alpha_i]$.

Jak už víme, budeme často používat homogenní polynomy $y^{d_i} f_i(x/y) = F_i(x, y)$, kde $d_i = \deg(f_i)$ budeme dále zjednodušeně značit pouze $d$, pro $i = 1, 2$. Je důležité najít polynomy $f_1$ a $f_2$ tak, aby při zvolené metodě prosévání dávaly na předpokládané oblasti dostatečně mnoho relací, tedy $B$-hladkých hodnot po dosazení do zhomogenizovaného polynomu. To je podstatné pro získání mnoha hodnot, ze kterých budeme vybírat relace pro sestavení čtverce ve smyslu ideálů. Existuje více způsobů, jak hledat zmíněné polynomy. Podle volby těchto metod je volena hodnota $B$ a oblast, ve které hledáme relace $(a, b)$. Dále je podstatné, aby se s polynomy $F_i(x, y)$ 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 $\varphi_1, \varphi_2$ pro převod nalezených hodnot v $\mathbb{Z}[\alpha_1]$, resp. $\mathbb{Z}[\alpha_2]$, do celých čísel modulo $N$. Pro $i = 1, 2$ a $r \in \mathbb{Z}$ mějme

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

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

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 $\mathbb{Z}[\alpha_i]$. Předpokládejme na chvíli, že jsme získali čtverce ve smyslu ideálů, které mají hlavní generátor opět čtverec. Označme je pro zjednodušení $\beta_1^2$ a $\beta_2^2$. 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. Chceme získat pro Fermatovu faktorizaci kongruentní čtverce v $\mathbb{Z}_N$ ve tvaru

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

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

$$\varphi_1\left(\beta_1^2\right) = x^2 \equiv y^2 = \varphi_2\left(\beta_2^2\right)\,(\mathrm{mod}\ N)\,.$$

Lze uvažovat i varianty pro více polynomů [10], ale těmi se tato práce nebude více zabývat.

### 3.2.2 Druhá fáze (prosévání)

Do druhé fáze algoritmu vstupujeme, když máme připraveny oba ireducibilní polynomy $f_1(x)$ a $f_2(x)$ v $\mathbb{Z}[x]$ a také jejich zhomogenizované verze $F_1(x, y)$ a $F_2(x, y)$. Tím máme teoreticky určena číselná tělesa $K_1$ a $K_2$, jejich obory algebraických celých čísel $O_{K_1}$, $O_{K_2}$, a podokruhu $\mathbb{Z}[\alpha_1]$ a $\mathbb{Z}[\alpha_2]$. Druhá fáze algoritmu probíhá v obou tělesech zvlášť stejným způsobem, proto dále vynecháme indexy.

Zvolíme mez $B \in \mathbb{N}$ pro $B$-hladkost a určíme oblast $I \subset \mathbb{Z}^2$, ze které 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. Množinu všech vybraných relací označme $\zeta$. Z ideálů určených relacemi budeme dále schopni zkombinovat ideál, který má v rozkladu na prvoideály pouze sudé exponenty. Omezením normy jsme omezili i množinu prvoideálů, které dělí ideály určené prvky $\zeta$. Množinu těchto prvoideálů pracovně nazýváme $B$-hladké prvoideály. Dvojice $(a, b) \in I$ vybíráme podle toho, zda je hodnota $F(a, b)$ $B$-hladká. To se dá očekávat pro hodnoty 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. 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 mocninami.

Tato fáze je časově nejnáročnější z celého algoritmu. Je třeba nasbírat velké množství relací a tedy se užívá velká prosévací oblast $I$. 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 někdy 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 $\zeta$. Relace reprezentují ideály $(a - b\alpha)$. Zřejmě $a - b\alpha \in O_K$, protože $O_K$ je podle tvrzení 2.3.5 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 Dedekindova oboru $O_K$. Jedná se o prvoideály nad všemi prvočísly, která jsou menší než $B$. Strukturu takových prvoideálů uvádíme v 2.5. Případně je tato množina rozšířena o konečně mnoho dalších prvoideálů, pokud jsme oslabili podmínku $B$-hladkosti pro určení relací.

Matice $\mathbf{M}$ poskytuje informaci o tom, kdy rozklad daného ideálu v $O_K$ obsahuje které prvoideály s lichým exponentem. Budeme hledat takovou podmnožinu $\zeta$, aby součin hlavních ideálů určených relacemi již neobsahoval ve svém rozkladu na prvoideály 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 o dvojici $(a, b)$, kterou není s čím zkombinovat. Obsahuje ideál v liché mocnině, který by zůstal stále v liché mocnině po libovolné 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. Jedná se o výpočet lineárních rovnic o více neznámých. Stále pracujeme v obou tělesech odděleně. Hledáme způsob, jak spárovat liché ex-

ponenty prvoideálů tak, aby po vynásobení všech vybraných ideálů 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 $j = 1, \ldots, k$.

Výběr relací získáme řešením soustavy lineárních rovnic $\mathbf{M}\mathbf{x}^T = 0$. Matici $\mathbf{M}$ jsme sestavili v předchozí fázi. Matice $\mathbf{M}$ je běžně velká řídká matice řádově o milionu řádků. Práce s takovou maticí, která by nebyla řídká, by byla výpočetně velice náročná. Pro velkou řídkou matici se při implementaci vyplácí použít Wiedemannovu blokovou metodu [20] nebo Lanczošovu blokovou metodu [21].

### 3.2.5 Pátá fáze (odmocninová)

V poslední fázi začneme tím, že vezmeme výsledky z lineární fáze. Zpočátku pracujeme v obou tělesech odděleně. Podle vektoru řešení $\mathbf{x}$ určíme, které relace použít k získání čtverce ve smyslu ideálů $\prod_{\zeta'}(a - b\alpha)$. Součin generátorů hlavních ideálů ale nedává generátor součinu hlavních ideálů. Je tedy třeba jiného postupu. Tímto problémem za zabývá [1] a ukazuje, že požadované vlastnosti můžeme docílit pomocí postupu s kvadratickými charaktery, které také popisuje [13]. Pak již získáme požadované 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 $\beta_1$ a $\beta_2$. 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, ref. [18]. Dnes je za nejlepší odmocninovou metodu považována Montgomeryho metoda, ref. [19].

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 popíšeme důležitost ireducibilních polynomů pro algoritmus. Budeme se věnovat oblast, ze které dosazujeme prvky do zvoleného polynomu. Také uvedeme metody porovnání získaných polynomů pro výběr toho nejvhodnějšího polynomu pro další fáze.

Algoritmus číselného síta postupuje tak, že nejprve vygeneruje ireducibilní polynomy s celočíselnými koeficienty, které pak slouží k práci s normami polynomů číselného tělesa. Volba polynomů tedy ovlivní další chod celého algoritmu. Od této kapitoly dále budeme vždy uvažovat minimální polynomy nad celými čísly. Značení zjednodušíme z $f_{\alpha,\mathbb{Z}}(x)$ pouze na $f_\alpha$.

## 4.1 Číselné těleso a algoritmus

Předpokládejme nadále že, pracujeme v Dedekindově oboru $O_K$, kde $K = \mathbb{Q}(\alpha)$. Prvek $\alpha \in \mathbb{C}$ celistvý nad $\mathbb{Z}$ je využíván pouze v teoretických úvahách. Nikdy ho nijak nereprezuntujeme a pro veškeré výpočty při prosévání používáme výhradně vygenerované polynomy. Do jejich zhomogenozovaných verzí pak dosazujeme kandidáty na relace, dvojice $(a, b)$, při prosévací fázi. Tyto prvky jsou vybírány z oblasti, které se tradičně říká prosévací oblast. Tomu se budeme věnovat v sekci 4.2.

Uvažujme v této sekci pouze monické polynomy $f$. Mějme k $\alpha \in \mathbb{C}$ celistvému nad $\mathbb{Z}$ jeho minimální polynom $f_\alpha = \sum_{i=0}^{d} a_i x^i$, který je stupně $d$. V praxi je $d$ malé číslo, v současnosti maximálně 8, často bývá rovno 6. Definujme zhomogenizovaný polynom tvaru $F_\alpha(x, y) = y^d f_\alpha(\frac{x}{y})$, který budeme v algoritmu hojně používat. Ukažme nejprve výpočet normy prvku $a - b\alpha \in O_K$, kde $a, b \in \mathbb{Z}$. Uvažujme dále

pouze normu prvků $\mathrm{N}_{K|\mathbb{Q}}$, kterou budeme zjednodušeně značit N.

Nejprve dokážeme pomocné tvrzení o determinantu matice multiplikativního zobrazení $\mu_{a-b\alpha}(x) = (a - b\alpha)x$.

**Tvrzení 4.1.1.** *Mějme $n \in \mathbb{N}$ a matici $\mathbf{M}$ s komplexními prvky o velikosti $n+1 \times n+1$ tvaru*

$$\mathbf{M} = \begin{pmatrix} a & 0 & 0 & \dots & 0 & ba_0 \\ -b & a & 0 & \dots & 0 & ba_1 \\ 0 & -b & a & \dots & 0 & ba_2 \\ \vdots & \ddots & \ddots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -b & a + ba_n. \end{pmatrix}$$

*Pak platí* $\det \mathbf{M} = a^{n+1} + \sum_{i=0}^{n} a_i a^i b^{n-i+1}$.

*Důkaz.* Postupujme pomocí matematické indukce. Pro $n = 1$ je matice tvaru $2 \times 2$:

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

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

Nechť tvrzení platí pro $n-1$. Dokažme, že platí i pro $n$. Determinant matice $n+1 \times n+1$ vypočteme následovně

$$\det \mathbf{M} = a \begin{vmatrix} a & 0 & \dots & 0 & ba_1 \\ -b & a & \dots & 0 & ba_2 \\ 0 & -b & \ddots & 0 & ba_3 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -b & a + ba_n \end{vmatrix} + (-1)^n ba_0 \begin{vmatrix} -b & a & 0 & \dots & 0 \\ 0 & -b & a & \dots & 0 \\ 0 & 0 & -b & \ddots & 0 \\ \vdots & \dots & \ddots & \ddots & a \\ 0 & 0 & 0 & \dots & -b \end{vmatrix}$$

První determinant dokážeme spočítat z indukčního předpokladu. Jeho výsledek je $a\left(a^n + \sum_{i=1}^{n} a_i a^{i-1} b^{n-i+1}\right)$. Druhý determinant je pouze násobek všech prvků na diagonále, kterých je $n$, protože se jedná o horní trojúhelníkovou matici. Výsledek determinantu je tvaru

$$\det \mathbf{M} = a\left(a^n + \sum_{i=1}^{n} a_i a^{i-1} b^{n-i+1}\right) + b^{n+1} a_0 = a^{n+1} + \sum_{i=0}^{n} a_i a^i b^{n-i+1}.$$

□

Matice zobrazení $\mu_{a-b\alpha}$ vzhledem k bázi $\left\{1, \alpha, \ldots, \alpha^{d-1}\right\}$ číselného tělesa $\mathbb{Q}(\alpha)$ nad $\mathbb{Q}$ je právě matice $\mathbf{M}$. Její determinant je tímto zajímavý, jako norma prvku $a - b\alpha$ a, jak nyní dokážeme, také jako norma hlavního ideálu $(a - b\alpha)$ okruhu $O_K$, která je podstatná pro zjišťování rozkladu na prvoideály v Dedekindově oboru $O_K$. Následující věta bude pro algoritmu stačit pouze pro monické polynomy, jak ukážeme později.

**Tvrzení 4.1.2.** *Mějme prvek $\alpha \in \mathbb{C}$ celistvý nad $\mathbb{Z}$ a $F_\alpha(x, y) = \sum_{i=0}^{d} a_i x^i y^{d-i}$ zhomogenizovaný minimální polynom prvku $\alpha$. Dále pro $a, b \in \mathbb{Z}$ uvažujme ideál $(a - b\alpha)$ okruhu $O_K$ číselného tělesa $K = \mathbb{Q}(\alpha)$. Potom platí*

$$\mathcal{N}\left((a - b\alpha)\right) = \left|F_\alpha(a, b)\right|.$$

*Důkaz.* Máme-li prvek $\alpha \in \mathbb{C}$ celistvý nad $\mathbb{Z}$. Pak je jeho zhomogenizovaný polynom po dosazení dvojice $(a, b)$ roven $F_\alpha(a, b) = \det \mathbf{M}$, kde matice $\mathbf{M}$ je maticí zobrazení $\mu_{a-b\alpha}$.

Norma hlavního ideálu generovaného prvkem $a - b\alpha \in O_K$ je podle 2.3.7 rovna normě prvku $a - b\alpha$. Tím přímo získáváme

$$\mathcal{N}\left((a - b\alpha)\right) = \left|\mathrm{N}(a - b\alpha)\right| = \left|\det(\mu_{a-b\alpha})\right| = \left|F_\alpha(a, b)\right|.$$

□

Rozkladem na prvoideály se nebudeme v této práci 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] nebo [13].

## 4.2 Prosévací oblast

Časově nejnáročnější část algoritmu číselného síta je hledání normy hlavních ideálů $O_K$. Jedná se o ideály tvaru $(a - b\alpha)$, kde $a, b \in \mathbb{Z}$ a $\alpha \in \mathbb{C}$ kořen polynomu $\mathbb{Z}[x]$, který, jak ukážeme později v sekci 5.2, nemusí být monický. Při hodnocení získaných polynomů je třeba uvažovat i to, pro jaké ideály budeme zjišťovat jejich normy, tedy hodnoty $F(a, b)$, jak jsme ukázali v předchozí sekci. Podívejme se na oblast, ze které budeme dosazovat do homogenních polynomů. Budeme ji dále nazývat **prosévací oblast**. Velikost této oblasti určuje, jak dlouho bude prosévání trvat. Pro větší oblast se zvýší jak počet zvolených relací, tak i čas prosévání. Proto je důležité přemýšlet o optimální velikosti prosévací oblasti.

Budeme-li dále hovořit o dvojicích $(a, b)$ z prosévací oblasti, budeme vždy uvažovat pouze nesoudělné $a$ a $b$. Dvojice, která by měla společného dělitele většího než jedna, by nepřinesla nic nového oproti dvojici vydělené tímto společným dělitelem. Zkoušením soudělných dvojic jen výrazně zvýšíme čas prosévání. Dále stačí, aby nanejvýš jedna z hodnot $a$ a $b$ byla záporná, vzhledem k tomu, že nás zajímá $B$-hladkost hodnot z homogenního polynomu. Označme $I_a$ (resp. $I_b$) interval, ze kterého volíme $a$ (resp. $b$).

Uvažujme nejprve *obdélníkovou prosévací oblast* pro nesoudělné dvojice $(a, b)$. Definujme ji pro $U \in \mathbb{N}$ následovně

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

Máme oblast o velikosti $2U^2$, ze které volíme hodnoty dosazované do polynomu $F(x, y) = \sum_i a_i x^i y^{d-i} \in \mathbb{Z}[x, y]$. Nastavení parametru $U$ je určováno heuristicky. V praxi bývá hodnota $U$ určena velikostí faktorizovaného čísla. Pro 100 ciferné číslo uvažuje měření v [1] hodnoty v řádu milionů. Pro RSA-768 [11] je prosévací oblast o velikosti v řádu miliard.

Vzhledem k tomu, že koeficienty u $x^d$ a $x^{d-1}y$ jsou malé a u $xy^{d-1}$ a $y^d$ velké, lze očekávat, že pro generování srovnatelně velkých hodnot $F(a, b)$ lze hodnoty $a$ vybírat z většího intervalu než hodnoty $b$. Této změně velikosti $I_a$ a $I_b$ říkáme *zkosení*. Míra změny (tedy **zkosení**) $s$ závisí na vlastnostech polynomu $F$. Meze pro intervaly změníme na nové hodnoty $A = U\sqrt{s}$ a $B = \frac{U}{\sqrt{s}}$. Prosévací oblast je pak určena

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

Stále se jedná o oblast velikosti $2U^2$. Zkosení $s$ volíme podle vygenerovaného polynomu. Bai [2] doporučuje, aby bylo voleno $s \approx \sqrt[3]{(\frac{a_{d-3}}{a_d})}$ pro polynomy stupně $d = 5, 6$. V takovém případě prokosíme podle koeficientu, kde již $b$ začíná mít větší exponent. Naopak další koeficienty, kde má $b$ největší exponent, jsou příliš velké a tím bychom získali příliš malý rozsah hodnot pro $b$.

Analogicky lze uvažovat *kruhovou prosévací oblast*, přesněji řečeno se jedná o polokruh o zvoleném poloměru, kdy opět uvažujeme $b$ pouze kladná. Pro poloměr $U \in \mathbb{N}$ a $\theta \in [0, \pi]$ máme

$$a = U \cos\theta,$$

$$b = U \sin\theta,$$

kde $\theta \in [0, \pi]$. Po prokosení získáme *eliptickou prosévací oblast*, kde $U \in \mathbb{N}$ a $\theta \in [0, \pi]$, tvaru

$$a = U\sqrt{s}\cos\theta,$$
$$b = \frac{U}{\sqrt{s}}\sin\theta.$$

Prosévací oblast 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 nalézt v [9].

## 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. Vycházejí z vlastností, které lze od polynomů očekávat, má-li být výpočet v prosévací oblasti dostatečně rychlý a přinést hodně dobrých relací. Taková kritéria hodnotí polynomy podle takzvaných vlastnosti koeficientů a kořenových vlastností.

Běžný postup bývá vygenerovat mnoho polynomů s relativně malými koeficienty. Upravit je a pak z nich vybrat nejvhodnějšího kandidáta. Případně pár kandidátů, na kterých je spuštěno testovací prosévání. Metodám předpovědi počtu relací podle testovacího prosévání se věnuje [16]. Tato metoda provede krátké prosévání na vybraných polynomech a podle výsledku odhaduje počet relací.

### 4.3.1 Hodnocení koeficientů

Od vygenerovaných polynomů požadujeme, abychom získali velký počet nesoudělných dvojic $(a, b)$, které povedou k hladkým hodnotám po dosazení do polynomů. Takové vlastnosti lze ohodnotit podle koeficientů a tím poznat vhodnost polynomu dříve, než pomocí zkušebního prosévání. Popišme jak takové hodnocení vypadá.

Murphy ve své práci [5] využívá Dickmanovu funkci $\rho$, definovanou v [14], jmenovitě

$$u\rho'(u) + \rho(u-1) = 0 \text{ pro } u > 1$$
$$\rho(u) = 1 \text{ pro } 0 \leq u \leq 1.$$

Dickmanova funkce udává odhad frekvence hladkých hodnot podle dané meze. Dickman dokázal, že pro počet $B$-hladkých čísel menších než mez $M$, označme $\psi(M, B)$, platí $\psi(M, M^{1/a}) \approx M\rho(a)$. Tedy lze uvažovat $a = \frac{\log M}{\log B}$.

Pravděpodobnost, že jsou $a$ a $b$ nesoudělná je přibližně rovna $\frac{6}{\pi^2}$, jak je dokázáno například v [15]. Ohodnocení polynomů $F_1$ a $F_2$ podle meze $B$-hladkosti je určeno na nějaké oblasti $I$ následovně

$$\frac{6}{\pi^2} \iint_I \rho\left(\frac{\log|F_1(x,y)|}{\log B}\right) \rho\left(\frac{\log|F_2(x,y)|}{\log B}\right) dxdy.$$

Uvedenou aproximaci lze více zjednodušit. Pro $f_2$ lineární polynom je druhý člen konstantní. Dále zanedbejme $\frac{6}{\pi^2}$ pro vzájemné porovnávání získaných polynomů. Tedy máme

$$\iint_I \rho\left(\frac{\log|F(x,y)|}{\log B}\right) dxdy.$$

Výpočet Dickmanovy funkce $\rho$ je mírně časově náročný, obzvláště pokud bychom takto chtěli ověřovat každý nalezený polynom. Naším požadavkem na hodnoty $|F_1(a,b)|$ a $|F_2(a,b)|$ je, aby byly v průměru malé pro volené dvojice $(a,b)$ z prosévací oblasti. Heuristicky tedy požadujeme malé koeficienty pro tyto dva polynomy. Naopak příliš malé koeficienty by nedávaly dostatek relací. Nejlepší se ukazují být polynomy s menším vedoucím koeficient, který řádově nemívá více jak 20 cifer. Ostatní koeficienty pak postupně řádově rostou.

Pro hodnocení koeficientů polynomu můžeme uvažovat normu ve smyslu největší z absolutních hodnot jeho koeficientů. Je vhodné uvažovat i to, jaká je mocnina proměnné, která tento koeficient násobí. Jednou z možností je počítat se sup-normou polynomu podle Kleinjunga [3].

Mějme množinu polynomů stupně nejvýše $d$ nad tělesem reálných čísel. Tuto množinu lze uvažovat jako vektorový prostor uspořádaných $d$-tic koeficientů s operacemi sčítání po koeficientech a násobení všech koeficientů prvky $r \in \mathbb{R}$.

Mějme dáno $s \in \mathbb{R}^+$ a $d \in \mathbb{N}$. Pro polynom $f(x) = \sum a_i x^i \in \mathbb{Z}[x]$ stupně nejvýše $d$. Položme

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

Dále index $d$ explicitně neuvedeme. Předpokládejme, že je zřejmý z kontextu. Definujme

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

Hodnoty $\sup(f, s)$ a $\sup(f)$ existují pro každý polynom zvoleného stupně $d \in \mathbb{N}$. Ukažme, že $\sup(f, s)$ i $\sup(f)$ jsou normami na vektorovém prostoru všech polynomů stupně nejvýše $d$.

**Poznámka 4.3.1.** *Připomeňme definici normy. Mějme vektorový prostor $V$ nad nějakým tělesem $T \subseteq \mathbb{C}$. Funkce $N$ je norma, pokud splňuje následující podmínky pro všechna $u, v \in V$ a všechna $r \in T$:*

- $N(v) = 0$ *právě tehdy, když* $v = 0$;
- $N(rv) = |r|N(v)$;
- $N(u + v) \le N(u) + N(v)$.

**Tvrzení 4.3.2.** *Mějme $s \in \mathbb{R}^+$ a polynom $f$ stupně nejvýše $d \in \mathbb{N}$. Sup-norma podle zkosení* $\sup(f, s)$ *je norma na vektorovém prostoru polynomů stupně nejvýše $d$.*

*Důkaz.* Uvážíme základní tři vlastnosti normy: pozitivní definitnost, pozitivní homogennost a subaditivitu.

Vždy platí, že $\sup(f, s) \ge 0$ a rovnost nastává pouze pokud všechny koeficienty polynomu $f$ jsou rovny nule. Pokud vynásobíme polynom $f$ libovolným číslem $r \in \mathbb{R}$, platí

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

Vzhledem k definici uvažujme pevný stupeň $d$. Pro součet dvou polynomů $f$ a $g$ stupně $d$ platí následující. Označme $f + g = \sum_{i=0}^{d} (a_i + b_i)x^i$. Pokud nejsou koeficienty $a_i + b_i$ definovány, pro $0 \le i \le d$, uvažujme je rovny 0. Pak platí vzhledem trojúhelníkové nerovnosti

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

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

□

**Důsledek 4.3.3.** *Mějme polynom $f$ stupně nejvýše $d \in \mathbb{N}$. Sup-norma* $\sup(f)$ *je norma na vektorovém prostoru polynomů stupně nejvýše $d$.*

Výše zavedená sup-norma je velice podobná $\ell^\infty$ normě. Empiricky se ukazuje jako dobrý nástroj na měření normy polynomu pro číselné síto.

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

### 4.3.2 Kořenová vlastnost

Dalším měřítkem pro výběr vhodného polynomu je takzvaná kořenová vlastnost. Jedná se o měření počtu kořenů polynomu modulo malá prvočísla. Zajímá nás přínos polynomu vůči různým prvočíslům menším než zvolená mez $B$. Takové polynomy dávají potom více hladkých výstupů a tedy relací. Nalézání takových hodnot pak vychází lépe, než volby náhodných čísel. Tato vlastnost se ukazuje být měřitelná. Popíšeme postup podle Murphyho [5].

**Poznámka 4.3.4.** *Pro celá čísla je p-valuace nejvyšší mocnina prvočísla $p$, které dělí dané celé číslo. Budeme ji dále značit $v_p$.*

Požadujeme, aby hodnota $F(a,b)$ byla s dobrou pravděpodobností $B$-hladká. Navíc je vhodné, aby polynom $F(x,y)$ měl mnoho nesoudělných kořenů modulo malá prvočísla. Uvažujme pouze prvočísla $p$ menší než zvolená mez $B$.

Střední hodnotou pro $p$-valuace $F(a,b)$, označme ji $E(v_p(F(a,b)))$, je v diskrétním případě vážený průměr. Uvažujme náhodnou nesoudělnou dvojici $(a,b)$ z předem zvolené prosévací oblasti $I$. Pro $B$-hladké $F(a,b)$ platí

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

Nechť počet kořenů homogenního polynomu $F(x, y)$ modulo $p$ na prosévací oblasti $I$ je roven $q_p$. Pro speciální prvočísla $p$ (definována v sekci 2.5) platí následující. Pro homogenní polynom $F$ na jeho 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 $p \leq B$ uvažujme případy, kdy máme kořeny $F$ modulo $p^k$. Počet takových kořenů je $p^k + p^{k-1}$, jak ukazuje Bai v článku [6] pomocí afinního a projektivního prostoru. Pravděpodobnost, že náhodná nesoudělná dvojice $(a, b)$ z prosévací oblasti je tímto kořenem, je $\frac{1}{p^k+p^{k-1}}$. Tím získáváme, že

$$\mathrm{E}(v_p(F)) = \sum_{k=1}^{\infty} \frac{q_p}{p^{k-1}(p+1)} = \frac{pq_p}{p^2-1}.$$

Podle těchto pozorování je určena Murphyho $\alpha(F)$ funkce, kterou definujeme podle [5]. Funkce $\alpha(F)$ slouží k ohodnocení polynomů podle rozložení jejich funkčních hodnot.

$$\begin{aligned}\alpha(F) &= \sum_{p\leq B} (1 - (p-1)\mathrm{E}(v_p(F))) \frac{\log p}{p-1} \\ &= \sum_{p\leq B} \left(1 - \frac{pq_p}{p+1}\right) \frac{\log p}{p-1}.\end{aligned}$$

Čím menší hodnotu má tato funkce, tím je polynom lepší pro hledání relací. Hodnota $\alpha(F)$ bývá často záporná, protože požadujeme, aby $F(x, y)$ mělo víc než jeden kořen. Myšlenkou $\alpha(F)$ funkce je porovnání náhodného čísla a funkčních hodnot polynomu $F$. Funkční hodnoty dobrého homogenního polynomu $F(x, y)$ se chovají podobně, jako náhodná čísla o velikosti $F(x, y)e^{\alpha(F)}$.

### 4.3.3 Murphyho $E$ funkce

Murphy v [5] také přichází s dalším zajímavým způsobem, jak ohodnotit nalezené polynomy. Doporučuje pro malý počet nejlepších kandidátů na polynomy $f_1$ a $f_2$ použít následující ohodnocení, které budeme dále nazývat Murphyho $E$ funkce.

Mějme dvojici polynomů $f_1$ a $f_2$ a zkosení $s$. Uvažujme jejich homogenní verze $F_1$ a $F_2$ na eliptické prosévací oblasti tvaru

$$x = \sqrt{s}\cos\theta,$$

$$y = \frac{1}{\sqrt{s}} \sin\theta,$$

kde $\theta \in [0, \pi]$. Tuto oblast uniformě rozdělme na $K$ částí. Podoblasti určeme $\theta_i$, kde $1 \leq i \leq K$. Mějme $j = 1, 2$ a uvažujme meze pro $B$-hladkost $B_{F_j}$. Položme

$$u_{F_j}(\theta_i) = \frac{\log |F_j(\sqrt{s}\cos\theta_i, \frac{1}{\sqrt{s}}\sin\theta) + \alpha(F_j)|}{\log B_{F_j}}.$$

Hodnotu $\alpha(F_2)$, kde $\deg F_2 = 2$, lze aproximovat na konstantu $0,56996$, jak dokazují R. Barbulescu, A. Lachand v [8], Proposition 3.2.

Pomocí Dickmanovy funkce $\rho$ je definována Murphyho $E$ funkce

$$E(F_1, F_2) = \sum_{i=1}^{K} \rho(u_{F_1}(\theta_i))\rho(u_{F_2}(\theta_i)).$$

Tato funkce dává měřitelné ohodnocení dvojice polynomů. Obecně není nutné uvažovat pouze eliptickou oblast pro dosazené hodnoty $x_i$ a $y_i$. Heuristicky, funkce $E$ za pomoci funkce $\rho$ aproximuje počet relací nalezených polynomy $f_1$ a $f_2$ na dané prosévací oblasti. Důvod je takový, že $E(F_1, F_2)$ lze uvažovat jako aproximaci integrálu uvedeného na začátku sekce 4.3.1 Hodnocení koeficientů.

Murphyho $E$ funkce je v současnosti považována za spolehlivé ohodnocení polynomů bez testovacího prosévání [7].

# Kapitola 5

# Generování polynomů

Volit polynomy pro číselné síto 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. Cílem této kapitoly je popsat algoritmy, považované v současnosti za nejefektivnější. Poslední uvedený algoritmus byl také naprogramován a jeho implementace je popsána i v následující kapitole.

Mějme přirozené číslo $N$, které není prvočíslo. Předpokládejme dále, že se jedná o číslo složené z velkých prvočísel. V této kapitole se budeme věnovat vygenerování dvou ireducibilních nesoudělných polynomů ze $\mathbb{Z}[x]$, které splňují

$$f_1(m) \equiv f_2(m) \equiv 0 \,(\mathrm{mod}\ N)\,,\ \text{kde } m \in \mathbb{Z}.$$

Vždy předpokládáme, že koeficienty každého polynomu jsou nesoudělné. V opačném případě bychom celý polynom zkrátili společným dělitelem 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.

V současné době jsou nejvíce používány metody, kdy generujeme polynom $f_2$ lineární a polynom $f_1$ stupně nejvýše 8. Dosud nebyl nalezen algoritmus, který by dával vhodné polynomy vyšších stupňů než 8. Experimentálně nebylo zjištěno zlepšení při volbě vyšších exponentů užitím známých metod generování polynomů.

Polynom $f_2$ nemusí být nutně lineární. Williams uvažuje v [22] užití dvou kvadratických, nebo dvou kubických polynomů. Prest a Zimmermann zkoumají užití

dvou polynomů stejného stupně v [23]. Takovým metodám se ale nebudeme v této práci více zabývat.

Začneme u základního algoritmu na generování polynomů, potom popíšeme další složitější postupy.

## 5.1 $m$-adický rozvoj

Nejjednodušší metodou k nalezení polynomů, splňujících požadavky, je metoda zvaná $m$-adický rozvoj. Jedná se o rozklad faktorizovaného čísla $N$ na násobky mocnin předem zvoleného čísla $m \in \mathbb{N}$, které budeme dále uvažovat jako kořen pro oba polynomy modulo $N$. Začněme s prvním polynomem $f_1$. Mějme rozvoj

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

pro $0 \leq a_i \leq m-1$. 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 $\mathbb{Z}[x]$. Tímto je tvaru $f_2(x) = x - m$. Pak již jistě platí základní podmínka společného kořene modulo $N$

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

kde $m \in \mathbb{N}$. Přitom jsou oba polynomy nesoudělné.

První polynom $f_1$ lze dále upravit. Není nutné používat pouze kladné koeficienty. Můžeme je tedy zmenšit, pokud budeme uvažovat i záporné koeficienty. Polynomy s velkým vedoucím koeficientem $a_d$ nedávají dobré výsledky. Uvažujme $a_d < \frac{m}{2}$. Ostatní koeficienty $a_i$, jsou-li větší než $\frac{m}{2}$, upravíme následovně:

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

Tento posun neprovádíme s vedoucím koeficientem i kdyby nastalo $a_d > \frac{m}{2}$, abychom nezvětšili stupeň polynomu. Je však běžné volit $m$ takové, aby bylo $a_d$ relativně malé. V případě monického polynomu neměníme ani následující koeficient $a_{d-1}$, aby neporušil moničnost. 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 monický polynom $f_1$ s kořenem $\alpha_1 \in \mathbb{C}$ a $K_2 = \mathbb{Q}$ pro polynom $f_2(x) = x - m$, kde $m \in \mathbb{N}$. Veškeré výpočty pro hledané relace $(a, b)$ budou v $K_2$ probíhat jako rozklady na prvočísla a jednotku $-1$. Důvod je takový, že normy prvků z $O_{K_2} = \mathbb{Z}$ jsou rovny hodnotě generujícího prvku ideálu. Tedy $\mathcal{N}_2(a - bm) = |a - bm|$.

Vstupem $m$-adického generování polynomů jsou přirozená čísla $N$ a $m$. Volbou čísla $m$ volíme i stupeň polynomu $\deg(f_1) = d$. Chceme-li nejprve určit stupeň $d$, 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, které vycházejí z tohoto základního způsobu generování polynomů.

## 5.2 Užití nemonických polynomů

Ukažme, jak pro algoritmus číselného síta využít i nemonické polynomy. Nechť $f(x) \in \mathbb{Z}[x]$ je nemonický polynom stupně $d$ ireducibilní nad $\mathbb{Z}[x]$. Potom již jeho kořen $\alpha \in \mathbb{C}$ není 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 $\alpha \in O_K$. Pro normy hlavních ideálů z okruhu algebraických celých čísel $O_K$ není splněna podmínka pro tvrzení 4.1.2. Ukažme, jak tento problém vyřešit.

Ireducibilní polynom $f(x) = \sum_{i=0}^{d} a_i x^i$ převedeme na monický polynom, protože takto získáme zpět jeho vlastnosti pro výpočet normy. Jak ukážeme v této sekci, budeme ale počítat s původním nemonickým polynomem, protože má lepší vlastnosti. Polynom $f(x)$ vynásobíme prvkem $a_d^{d-1}$. Jeho proměnou $x$ převedeme na $\frac{x}{a_d}$. Tím získáme upravený polynom

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

Máme 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 $\mathbb{Z}$. Tedy $\beta \in K$ je algebraické celé číslo. Číselné těleso $K = \mathbb{Q}(\alpha)$ lze také zapsat ve tvaru $\mathbb{Q}(\beta)$, protože ${a_d}^{-1} \in \mathbb{Q}$.

Pro $a, b \in \mathbb{Z}$ je $(a - b\alpha)$ lomený ideál okruhu $O_K$. Podle tvrzení 2.4.5 víme, že $O_K$ je Dedekindův obor. Tvrzení 2.4.10 říká, že v Dedekindově oboru i pro lomené ideály existuje jednoznačný rozklad na prvoideály v celočíselných mocninách. Lomený ideál $(a-b\alpha)$ lze tedy stále jednoznačně rozložit. 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. Normu lomených ideálů budeme počítat pomocí původního polynomu. 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 $O_K$.

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

Díky těmto dvěma úvahám a tvrzením 2.3.7, 2.4.13 a 4.1.2 lze již přímo zjistit normu hlavního lomeného ideálu Dedekindova oboru $O_K$ následovně.

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

Použití nemonických polynomů ovlivní výrazně prosévací fázi, ale také sestavení matice. Při prosévání stačí uvažovat pouze hodnoty $F(a, b)$ pro určení vhodných relací, které zařadíme do množiny $\zeta$. 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 $(a_d)^{-1}$ v $O_K$. Je podstatné nezanedbat tyto lomené ideály pro správné spárování prvoideálů z rozkladů všech ideálů generovaných relacemi ze $\zeta$. 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 $\zeta$. 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 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 pak vedoucí koeficient. Podle nich jsou dále voleny další koeficienty polynomu. Některé hodnoty koeficientu $a_d$ nemusí dávat dobré polynomy. Je vhodné nevolit vedoucí koeficient $a_d$ až příliš velký. Prvním krokem algoritmu bývá nastavení (většinou navýšení) $a_d$ na vhodnou hodnotu, ke které bývá sestaven zbytek polynomu. Takto je vygenerováno více polynomů, které jsou pak porovnány metodami popsanými v sekcích 4.3.2 a 4.3.3. Z nich je vybrán vhodný polynom pro další fáze.

## 5.3 $(m, p)$-adický rozvoj

Možnou variantou $m$-adického rozvoje je metoda $(m, p)$-adického rozvoje čísla $N$, kde $m, p \in \mathbb{N}$. Uvažujme v této sekci $p$ jako prvočíslo. Tento požadavek není nutný. V novějších metodách se jedná o součin pár prvočísel. Metody postavené na $(m, p)$-adickém rozvoji hledají hodnoty $m$ a $p$ více sofistikovaně. Volbou těchto hodnot také určíme, zda bude polynom $f_1$ monický. Druhý polynom budeme uvažovat lineární a nemonický tvaru $f_2(x) = px - m$.

Při získávání polynomu $f_1$ postupujeme analogicky jako v $m$-adické metodě. Zjistíme všechny koeficienty $a_j$ pro rozklad

$$N = \sum_{j=0}^{d} a_j m^j p^{d-j}.$$

Tyto koeficienty využijeme jako koeficienty polynomu $f_1(x) = \sum_{j=0}^{d} a_j x^j$. Platí, že

$$p^d f_1\left(\frac{m}{p}\right) = N.$$

Společný kořen obou nemonických polynomů modulo $N$ je tvaru $mp^{-1} \pmod N$ a budeme ho také značit $\frac{m}{p}$. Tím modifikujeme jeden ze základních požadavků na generované polynomy. Požadavek na kořen obou polynomů oslabíme z celých čísel pouze na čísla racionální. Potřebujeme, aby existoval prvek $p^{-1}$ modulo $N$. Pokud by takový prvek neexistoval, získali jsme netriviálního dělitele $N$, což přesně chceme. Předpokládejme dále, že $p^{-1}$ modulo $N$ existuje. V praxi hledáme takové $p$, aby byl jeho inverzní prvek modulo $N$ rychle nalezitelný. Podle toho také upravíme homomorfizmy $\varphi_i : \mathbb{Z}[\alpha_i] \to \mathbb{Z}_N$, kde $i = 1, 2$, na $\varphi_i(\alpha_i) = mp^{-1} \pmod N$.

Mějme zvolená čísla $m$ a $p$. Dalším krokem bývá zvolit vedoucí koeficient $a_d$ a začít generovat polynom $f_1$. Předtím, než začneme takový polynom generovat,

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 $f_1(x) = \sum_{j=0}^{d} a_j x^j$ s kořenem $mp^{-1}$ modulo $N$. Pokud $p$ a $a_d$ mají společného dělitele většího než 1, nejsou koeficienty polynomu $f_1$ nesoudělné. Takovému případu se chceme vyhnout.

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

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

Dosadíme-li $m$ za $x$ získáme: $N \equiv a_d m^d \,(\mathrm{mod}\ p)$. Pokud tato kongruence nemá řešení, ireducibilní polynom $f_1$ s kořenem $m$ modulo $p$ neexistuje.

Pro $p$ neprvočíselné, je třeba, aby uvedené kongruence měly řešení pro všechny jeho prvočíselné dělitele.

Ukažme, jak vygenerovat vhodný nemonický polynom pomocí $(m, p)$-adického rozkladu. Pro zvolené $m \in \mathbb{N}$ 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 pro $i = d-1, \ldots, 0$, že

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

Jednotlivé koeficienty $a_i$ získáme tak, aby byla splněna kongruence pro celočíselnost pomocných hodnot $r_i \equiv a_i m^i \,(\mathrm{mod}\ p)$. Definujeme je pro vhodná $0 \leq q_i < p$ následovně

$$a_i = \frac{r_i}{m^i} + q_i \in \mathbb{Z}.$$

Vyjádřeme $r_i$ 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 určité oblasti, 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 5.3.1.** *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 následovně:*

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

*Důkaz.* Existenci takového polynomu jsme právě 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 $a_{d-1}$. 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 $i = 0, \ldots, d-2$ dokážeme obdobně

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

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

□

## 5.4 Montgomery - Murphyho algoritmus

Popišme algoritmus Montgomery - Murphyho, ze kterého dále vychází dnes nejpoužívanější algoritmy. Tento postup dává ireducibilní polynomy se společným kořenem $m$ modulo $N$. Výsledné polynomy jsou tvaru $f_1(x) = \sum_{i=0}^{d} a_i x^i$ a $f_2(x) = x - m$, kde $m, a_0, \ldots, a_d \in \mathbb{Z}$. Výsledný polynom $f_1$ nebývá monický.

Uvažujme mez $k \in \mathbb{N}$ pro vedoucí koeficient. Ten nejprve inicializujeme $a_d = 0$. Postupně hodnotu $a_d$ navyšujeme až do chvíle, kdy překročíme mez $k$. Tím ukončíme generování polynomů. Ke každé hodnotě $a_d$ hledáme ireducibilní polynom s malými koeficienty. Pokud nevycházejí malé, pro dané $a_d$, negenerujeme celý polynom. Z nalezených polynomů vybereme vhodné kandidáty na polynom $f_1$ podle kritérií popsaných v předchozí kapitole.

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

Tímto Montgomery - Murphyho algoritmus nekončí. Dalším krokem je optimalizace nalezených polynomů. Murphy přichází se dvěma metodami:

- **Translace**, nebo-li posunutí kořene polynomu. $f_{i,new}(x) = f_i(x + t)$, kde $i = 1, 2$ a $t \in \mathbb{N}$. Tato metoda má vliv pouze na velikost polynomu. Mění jeho koeficienty, ale neměníme kořenové vlastnosti při lineární změně kořene o $t$ nepříliš velké.

- **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 $f_1(x)$. V praxi bývá stupně nejvýše 2. Kořen $f_1(x)$ zůstane stejný, ale změníme jak velikosti koeficientů, tak i kořenové vlastnosti.

Užitím těchto dvou metod upravíme velikosti polynomů tak, aby měly lepší normu. Proto také tyto metody využívají i další algoritmy pro generování polynomů. Detailnějším popisem Montgomery - Murphyho algoritmu se již zabývá [1], případně přímo Murphyho práce [5].

## 5.5 Kleinjungův první algoritmus

Kleinjung popsal v [3] algoritmus na generování polynomů, který budeme dále nazývat Kleinjungův první algoritmus. Jedná se $(m, p)$-adickou metodu na výpočet polynomu $f_1$. Polynom $f_2$ bude lineární, proto o něm nebudeme více hovořit. Hodnoty $m$ a $p$ spočteme tak, že tím získáme omezení všech koeficientů polynomu a že platí $N \equiv a_d m^d \pmod{p}$, potřebné pro $(m, p)$-adickou metodu.

Podle velikosti čísla $N$ zvolíme stupeň polynomu $d$. Postupně volíme vedoucí koeficient $a_d$, podle kterého získáme $m$, $p$ a další koeficienty polynomu. Pevně zvolené $a_d$ a $d$ považujme v dalších úvahách jako konstanty. První určíme předběžný kořen $\widetilde{m} = \sqrt[d]{\frac{N}{a_d}}$, ponechaný v reálném tvaru. Konečný celočíselný kořen $m$ určíme až později. Nebude řádově jinde než $\widetilde{m}$. Stejně tak poslední dva koeficienty $a_0$ a $a_1$ budou nejvýše řádově o velikosti $\widetilde{m}$.

Popišme úvahy pro omezením zkosení a koeficientů polynomu. Uvažujme sup-normu polynomu, uvedenou v sekci 4.3.1. Pokud za $s$ z definice dosadíme zkosení prosévací oblasti $I$, získáváme sup-normou informace o relativní velikosti koeficientů při používání hodnot z oblasti $I$. 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.

Jak zvolit správně hodnotu zkosení $s$, pomáhají určit následují dvě tvrzení. Uvažujme dále, že polynom $f_1$ má nejmenší vedoucí koeficient $a_d$, další koeficienty pak již větší a největší poslední dva koeficienty, které řádově nepřesahují $\widetilde{m}$.

**Tvrzení 5.5.1.** *Mějme číslo $N$ a mez $M \in \mathbb{N}$, zkosení $s \in \mathbb{R}^+$ zvolené prosévací oblasti a polynom $f_1(x) = \sum_{i=0}^{d} a_i x^i \in \mathbb{Z}[x]$, pro který platí $\sup(f_1, s) = M$. Nechť Dále mějme odhad velkosti kořene $\widetilde{m} = \sqrt[d]{\frac{N}{a_d}}$. Potom platí*

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

*Důkaz.* Mějme polynom $f_1$ s koeficienty podle požadavků v této sekci a mez pro sup-normu

$$M \geq \sup(f_1, s) = \max_{0 \leq i \leq d} \left| a_i s^{i-\frac{d}{2}} \right|.$$

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 to 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 $|a_1|$ odpovídá přibližně velikostí $\widetilde{m}$ máme

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

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

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

□

**Tvrzení 5.5.2.** *Mějme číslo $N$ a mez $M \in \mathbb{N}$, zkosení $s \in \mathbb{R}^+$ zvolené prosévací oblasti a polynom $f_1(x) = \sum_{i=0}^{d} a_i x^i \in \mathbb{Z}[x]$, pro který platí $\sup(f_1, s) = M$. Dále mějme odhad velkosti kořene $\widetilde{m} = \sqrt[d]{\frac{N}{a_d}}$. Nechť nastává maximum sup-normy v $j$-té souřadnici. Potom platí*

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

*Důkaz.* Mějme polynom $f_1$ s koeficienty podle požadavků v této sekci a mez pro sup-normu

$$M \geq \sup(f_1, s) = \max_{0 \leq i \leq d} \left| a_i s^{i-\frac{d}{2}} \right|.$$

Nechť maximum nastává právě pro $0 \leq i = j \leq d$, 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 tvaru

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

□

Tím jsme určili omezení shora i zdola pro zkosení $s$ pro polynomy generované tak, aby jejich koeficienty splňovaly výše uvedené požadavky. Díky těmto omezením dokážeme určit omezení pro další koeficienty polynomu. Pro každý index $i = d, \ldots, 0$ máme omezení pro $a_i$ shora.

**Tvrzení 5.5.3.** *Mějme mez $M \in \mathbb{Z}$, zkosení $s \in \mathbb{R}^+$ zvolené prosévací oblasti a polynom $f(x) = \sum_{i=0}^{d} a_i x^i \in \mathbb{Z}[x]$, pro který platí $\sup(f, s) = M$. Předpokládejme, že sup-norma nabývá svého maxima vždy právě pro koeficient, pro který hledáme omezení. Pak platí*

- $|a_i| \leq M \left( \left( \frac{M}{\widetilde{m}} \right)^{\frac{2}{d-2}} \right)^{\frac{d}{2}-i} =: a_{i,max}$ *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 tvrzení 5.5.1 a 5.5.2 o omezení pro zkosení $s$ dostaneme mez pro vedoucí koeficienty polynomu. Nechť maximum pro sup-normu nastává v $j$-té souřadnici.

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

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

Uvažujme, že pro každý koeficient $a_i$ existuje možnost zvýšit ho na $|a_i'| > |a_i|$, kdy již nastává maximum sup-normy. Mějme tedy zvlášť pro každý koeficient takový případ. Tím již získáme omezení nejen pro index $j$.

Rozlišme nyní případ $i = d$. 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 úvahám v Kleinjungově prvním algoritmu, které vedou k získání $m$ a $p$. Jedná se převážně o heuristické úvahy. Zavedeme pomocné hodnoty. Máme stále podmínku, že existuje řešení pro $N \equiv a_d x^d \pmod{p}$, jak jsme popsali v sekci 5.2 o použití nemonických polynomů. Již 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 získali netriviálního dělitele čísla $N$, které chceme rozkládat.

**Tvrzení 5.5.4.** *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 kongruence $N \equiv a_d x^d \pmod{p}$ 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. Pak 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 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 řešení kongruence platí $\gcd(x^d, p_i) = 1$. Nastavme 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 $p_i > d$. Kongruence $p_i \equiv 1 \pmod{d}$ způsobuje, že počet možných $c$ je pak roven právě $d$ v $\mathbb{Z}_{p_i}[x]$.

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 $x_{i,\mu_i}$ pro jednotlivá $p_i$, kde $1 \leq i \leq l$ a $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_l) \in \{1, \ldots, d\}^l$,

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

Platí, že $\frac{p}{p_i} | x_{i,\mu_i}$, a má tedy 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 $m_{i,\mu_i}$ 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 $0 \leq e_{i,j} \leq d-1$, 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. Tato $e_{i,j}$ nejsou definována jednoznačně. To však není pro implementaci potřeba.

Zvolme dva vektory $\boldsymbol{\mu}$ a $\boldsymbol{\mu}'$ z $\{1,\ldots,d\}^l$, 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 $\tilde{\boldsymbol{\mu}} \in \{1,\ldots,d\}^l$, 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 $j \neq i$, kde $1 \leq j \leq d$.

Předpokládejme, že platí kongruence:

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

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 $i = 1,\ldots,l$ a $j = 1,\ldots,d$:

- $f_0 = \frac{N - a_d m_0^d}{p^2 m_0^{d-1}}$,
- $f_{i,j} = -\frac{a_d d x_{i,j}}{p^2} - \frac{e_{i,j}}{p}$.

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

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

### 5.5.1 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 $P$. 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 $\epsilon$ okolí nějakého celého čísla. Tím získáme hledané polynomy.

V tomto bodě se postupuje tak, že spočteme dva seznamy

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

Ty potom seřadíme a postupně hledáme prvky z druhého seznamu, které se nachází v $\epsilon$-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.

## 5.6 Kleinjungův druhý algoritmus

Kleinjung přišel o dva roky později s dalším algoritmem na generování polynomů [4]. V základním principu se opět jedná o $(m, p)$-adický rozvoj, kde se algoritmus soustředí na nalezení čísel $m$ a $p$ pro předem zvolený stupeň polynomu $d$ a vedoucí koeficient $a_d$, které budeme opět uvažovat jako konstanty pro další úvahy.

Přínos Kelinjungova druhého algoritmu oproti prvnímu je v následující úvaze. Nejprve hledáme zápis čísla $N$ do prvních dvou členů pomocí $p$ a odhadnutého $m$.

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

Spočtení přesné hodnoty $m$ závisí také na tom, že chceme získat podíl $\frac{R}{m^{d-2}}$ dostatečně malý. Tento podíl je velice blízkou třetímu koeficientu $a_{d-2}$. Právě v tom je zlepšení druhého Kleinjungova algoritmu oproti prvnímu, kde není hodnota třetího koeficientu tolik omezena.

Pro usnadnění výpočtů upravme faktorizované číslo $N$ na $\tilde{N}$ podle následujících úvah. Předpokládejme, že máme vhodná $m$ a $p$ celá čísla, kde $m$ je blízké $a_d\sqrt[d]{N}$ a $p$ je součinem $l$ malých prvočísel, kde $1 \leq l \leq 4$. Definujme

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

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

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

Mějme 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

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

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

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

Dále uvažujme pomocný kořen $\tilde{m} = da_d m + a_{d-1} p$ a zbytek násobený $p^2$ položíme $\tilde{R}$ čímž získáme:

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

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

Pro malé hodnoty $p$ je $\tilde{m} \approx \sqrt[d]{\tilde{N}}$. Právě tyto úvahy umožňují odhadnout velikost třetího koeficientu $a_{d-2}$ podle zbytku $\tilde{R}$.

**Tvrzení 5.6.1.** *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 $\tilde{N}$, $\tilde{m}$ a $\tilde{R}$ v předchozím odstavci můžeme $\tilde{R}$ vyjádřit ve tvaru

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

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

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

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 $m$. Tedy $\frac{\left|\tilde{R}\right|}{d^2 a_d \tilde{m}^{d-2}}$ odpovídá velikostí $a_{d-2}$. □

Rozeberme nyní možnosti volby $p = \prod_1^l p_i$. Mějme $l = 2$. Obě prvočísla volíme z různých intervalů $p_1 \in P = [P_1, P_2]$ a $p_2 \in Q = [P_2, P_3]$ pro $P_1 < P_2 < P_3$. Zde jsou možné variace pro intervaly $P$ a $Q$. Například se může jednat o stejný interval, kdy pro $P$ volíme hodnoty rovné 1 modulo 4 a pro $Q$ volíme hodnoty rovné 3 modulo 4. Nejedená se pouze o součin dvou libovolných prvočísel $p_1$ a $p_2$ ze zvolených intervalů. Je podstatné, aby bylo splněno

$$\tilde{N} \equiv (\tilde{m}_0 + r)^d \left(\mathrm{mod}\ {p_1}^2 {p_2}^2\right).$$

Důvod pro tuto kongruenci vysvětlíme při výpočtu výsledné hodnoty $m$. Nyní hledáme řešení $r_i$ kongruence modulo ${p_i}^2$ pro $i = 1, 2$. Zaznamenáme nalezené dvojice $(p_1, r_1)$ a $(p_2, r_2)$ zvlášť pro $p_1 \in P$ a pro $p_2 \in Q$. Vybereme taková $p_1$ a $p_2$, pro která nalezneme kolizi řešení $r_1$ s $r_2$. Čímž získáme $p = p_1 p_2$.

Bai přichází ve své práci [2] s variací podle Zimmermanna pro $l = 4$. Uvažuje hodnotu $p = p_1 p_2 q_1 q_2$ jako součin prvočísel, kde $p_1, p_2 \in P = [\sqrt{B}, 2\sqrt{B}]$ a $q_1, q_2 \in P = [2\sqrt{B}, 3\sqrt{B}]$. Pro různé dvojice $p_1, p_2$ (resp. $q_1, q_2$) řešíme $r$ v kongruenci

$$\tilde{N} \equiv (\tilde{m}_0 + r)^d \left(\mathrm{mod}\ {p_1}^2 {p_2}^2\right).$$

Pak hledáme kolize řešení $r$ pro $p_1, p_2$ a $q_1, q_2$. Tím získáme

$$\tilde{N} \equiv (\tilde{m}_0 + r)^d \left(\mathrm{mod}\ {p_1}^2 {p_2}^2 {q_1}^2 {q_2}^2\right).$$

Velikost $p = p_1 p_2 q_1 q_2$ odpovídá velikosti $B^2$ a třetí koeficient $a_{d-2}$ je o velikosti $\frac{\tilde{m}_0}{B^2}$.

Další možnost úpravy podle Kleinjunga je varianta pro $l = 3$, kdy $p = p_1 p_2 q$. Jedná se o přepoužívání již nalezených kořenů kongruence $\tilde{N} \equiv (\tilde{m}_0 + r_p)^d \,(\mathrm{mod}\ p^2)$. Určíme dvě množiny prvočísel $P$ a $Q$, které nemají společné prvky. Pro všechna $p \in P$ nalezneme dvojice řešení $(p, r_p)$ a najdeme v nich kolize na $r_p$. Dále pro $q \in Q$ opět řešíme stejnou kongruenci modulo $q^2$ a zaznamenáme nalezené dvojice $(q, r_q)$. Pak pro každou dvojici $(q, r_q)$ a všechny dvojice $(p, r_p)$ spočteme $i_p \in [0, q^2)$ z kongruence

$$r_q + i_q q^2 \equiv r_p \left(\mathrm{mod}\ p^2\right),$$

takové, že platí

$$\tilde{N} \equiv (\tilde{m}_0 + r_q + i_p q^2)^d \left(\mathrm{mod}\ p^2\right).$$

Hodnoty $(p, i_p)$ zaznamenáme. Pokud nalezneme kolizi na $i_p$ pro různá $p_1$ a $p_2$, potom platí

$$\tilde{N} \equiv (\tilde{m}_0 + r_q + i_q q^2)^d \left(\mathrm{mod}\ {p_1}^2 {p_2}^2 q^2\right).$$

Položme $m_2 = p_1 p_2 q$ a $r = r_q + i_p q^2$, čímž získáme $\tilde{N} \equiv (\tilde{m}_0 + r)^d \,(\mathrm{mod}\ {m_2}^2)$. Pro uvedenou kongruenci je třeba uvažovat podmínku, aby platila nesoudělnost mezi $m_2$ a $a_d$, potažmo i $d$. Běžně volíme hodnotu $a_d$ dělitelnou malými prvočísly tak, že je soudělná s $d \leq 8$. Stačí tedy uvažovat pouze taková prvočísla z $P$ a z $Q$, která nedělí $a_d$.

Zde máme možnost uvažovat omezení intervalů $P$ a $Q$ tak, abychom získali kolize a přitom urychlili hledání $p$ tak, aby bylo nesoudělné s $a_d$.

Dopočet $m$ je následující. Položme $\tilde{m} = \tilde{m}_0 + r$. Vzhledem k požadavku

$$\tilde{N} \equiv (\tilde{m}_0 + r)^d \left(\mathrm{mod}\ p^2\right),$$

a tomu, jak jsem získali $p$, máme

$$N \equiv a_d \left(\frac{\tilde{m}_0 + r}{da_d}\right)^d \left(\mathrm{mod}\ p^2\right).$$

Hodnota $m$ musí splňovat $N \equiv a_d m^d \pmod{p}$. Potřebujeme tedy $m \equiv \frac{\tilde{m}_0 + r}{da_d} \pmod{p}$. Keinjung [4] volí přímo rovnost, Bai [2] uvádí $m = \frac{\tilde{m}_0 + r - a_{d-1}p}{da_d}$, kde $a_{d-1}$ volí dělitelné $da_d$, konkrétně tvaru $a_{d-1} \equiv \frac{\tilde{m}_0 + r}{p} \pmod{da_d}$. Takto získané $m$ již využijeme s $p$ na zisk polynomu $f_1$ pomocí $(m, p)$-adického rozvoje čísla $N$.

### 5.6.1 Kleinjungův druhý algoritmus - postup

Kleinjungův druhý algoritmus postupuje následujícím způsobem. Nejprve najde vhodné $p, \tilde{m}$, podle kterých dopočte $m$, potřebné pro $(m, p)$-adický rozvoj čísla $N$. Vstupní hodnoty jsou:

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

Ze vstupu máme okamžitě $\tilde{N} = d^d a_d^{d-1}$ a $\tilde{m}_0 = \left\lfloor \sqrt[d]{\tilde{N}} \right\rfloor \in \mathbb{N}$, zvolené jako celou dolní část. Pro hodnoty $P_1 \leq p_i \leq 2P_1$, spočteme $r_i$ z kongruence

$$\tilde{n} \equiv (\tilde{m}_0 + r_i)^d \left(\mathrm{mod}\ p_i^2\right).$$

Zaznamenáme dvojice $(p_i, r_i)$ a hledáme kolize pro různá $i$. Pokud kolizi nenalezneme, navýšíme hodnotu $a_d$ a začneme znovu. Najdeme-li kolizi, označme ji $r'$, pak platí:

$$\tilde{n} \equiv (\tilde{m}_0 + r')^d \left(\mathrm{mod}\ p_{i_1}^2 p_{i_2}^2\right).$$

Zvolíme $p = p_{i_1} p_{i_2}$ a $\tilde{m} = \tilde{m}_0 + r'$. Velikost $p$ tedy závisí na nalezených hodnotách $p_{i_1}$a $p_{i_2}$. Při nálezu více kolizí můžeme generovat více polynomů pro jeden vedoucí koeficient.

Hodnotu $m$ spočteme podle $\tilde{m} = da_d m + a_{d-1}p$, tedy pomocí $a_{d-1} \pmod{da_d}$. Máme tedy

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

Nyní již dopočteme koeficienty $a_i$ pro $0 \leq i \leq d-2$ pomocí $(m,p)$-adického rozvoje $N$. Jejich hodnoty budou maximálně o velikosti násobku $m$.

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

Kleinjung ve své práci [4] uvádí testování na několika případech, kdy pro čísla RSA 512, 576 a 640 dával druhý Kleinjungův algoritmus lepší polynomy, než předchozí algoritmy. Výsledky měření jsou empirické. Tento algoritmus byl použit při faktorizaci čísla RSA 768 [11] ze seznamu RSA challenge [17].

# Kapitola 6

# Program

Cílem této kapitoly je popsat implementaci Kleinjungova druhého algoritmu na generování polynomů a její výsledky. Jedná se o část přidanou do programu číselného síta popsaného v práci Lukáše Perutky [1], kód je napsaný v jazyce C++. Nejprve popíšeme zvolenou variantu programu s poznámkami pro další vývoj a potom výsledky získané ze současného stavu algoritmu.

## 6.1 Zvolené parametry

Pro generování polynomů metodou druhého Keinjungova postupu byla naprogramována varianta algoritmu popsaného v sekci 5.6.1. Zvolené modifikace popíšeme v této sekci.

Pro Kleinjungův druhý algoritmus byl použit stejný rámec jako pro Kleinjungův první algoritmus obsahující $(m, p)$-adický rozvoj. Metody zlepšení, popsané v sekci 5.4, a ohodnocení polynomů, popsané v sekci 4.3, jsou rovněž stejné pro oba algoritmy.

Obecně se ukazuje vhodné volit stupeň polynomu $d$ podle velikosti faktorizovaného čísla. Pro $N$ o velikosti méně než 105 decimálních hodnot volíme $d = 4$. Pro větší pak $d = 5$ až po 175 decimálních hodnot. Hodnotu $d$ navyšujeme o 1 vždy po 70-ti decimálních hodnotách faktorizovaného čísla $N$ až po $d = 8$. Zde je velký prostor pro úpravy. Hodnoty jsou voleny empiricky na základě ohodnocení získaných polynomů. V případě výsledků, popsaných v následující sekci, vychází nejlépe právě popsaný způsob nastavení.

Vedoucí koeficient je volen stejně v obou algoritmech jako násobky $k60$, kde $k \in \mathbb{N}$. Tedy je dělitelný 2, 3, 5 a $k$. Možnou variantou by bylo volit dělitelnost navíc i číslem 7 a tedy $a_d = k210$, pro $k \in \mathbb{N}$. Při testování na 100 ciferných hodnotách faktorizovaného čísla $N$ bylo dosaženo podobných výsledků pro obě varianty. Uvedeny výsledek v následující sekci je pro první variantu $a_d = k60$, $k \in \mathbb{N}$, protože jeho ohodnocení vyšlo s malým rozdílem lépe.

K sestavení druhého kořene byla zvolena následující varianta. Hodnota $p$ je složena ze 3 prvočísel. Dvě prvočísla $p_1$, $p_2$ jsou volena ze stejné množiny $P$ a třetí $q$ z jiné množiny $Q$. Množiny $P$ a $Q$ nastavíme ve fázi příprav před tím, než začneme postupně volit vedoucí koeficient $a_d$. Zvolili jsme variantu, kdy do obou množin rozdělujeme prvočísla z intervalu $[10, 1000]$. Spodní mez volíme tak, abychom přímo vyloučili prvočísla, která mohou dělit stupeň polynomu $d$, a omezili prvočísla, která dělí vedoucí koeficient $a_d$. Horní mez je nastavena heuristicky, čím je větší, tím více dvojic je nalezeno, ale tím déle generování trvá. Pro současné omezení byla nalezena hodnota $p$ pro většinu zvolených koeficientů $a_d$. Do množiny $P$ řadíme prvočísla $1 \pmod 4$ a do množiny $Q$ všechna ostatní. Kontrola, zda $p_1$, $p_2$ nebo $q$ nedělí $a_d$ probíhá při zaznamenávání dvojic řešení $(p_j, r_{p_j})$, resp. $(q, r_q)$. Hodnota $p$ pak v praxi vychází v průměru $6 - 8$ ciferná.

Pro všechna $p_j \in P$ dopočteme řešení $r_{p_j}$ z $\tilde{N} \equiv (\tilde{m}_0 + r_{p_j})^d \left(\mathrm{mod}\ p_j^2\right)$, pokud existuje. Nalezené dvojice $(p_j, r_{p_j})$ zaznamenáme. Pak pro každé $q \in Q$ opět nalezneme $r_q$ z $\tilde{N} \equiv (\tilde{m}_0 + r_q)^d \pmod{q^2}$ a zaznamenáme $(q, r_q)$. Pak tento seznam postupně projdeme a pro každé $q$ projdeme seznam $(p_j, r_{p_j})$, kde vyřešíme $i_{p_j}$ z $\tilde{N} \equiv (\tilde{m}_0 + r_q + i_{p_j} q^2)^d \pmod{(p_j q)^2}$ a zaznamenat dvojice $(p_j, i_{p_j})$. Nyní najdeme kolize mezi zaznamenanými dvojicemi. Tím získáme $p_1$, $p_2$ a $q$ a řešení $r' = r_q + i_p q^2$, které potřebujeme pro výpočet $m$ a $p$.

## 6.2 Výsledky

Uvedený program pro generování polynomů byl testován pro 110 a 120 ciferná čísla. Popišme výsledky pro číslo zvolené z wikipedie, článku RSA numbers [24], tvaru

$$N = 35794234179725868774991807832568455403003778024228226193532908190484670252364677411513516111204504060317568667.$$

Druhý Kleinjungův algoritmus byl spuštěn pro $a_d \in [60; 507600]$, kde $a_d = k60$, $k \in \mathbb{N}$. Podle ohodnocení Murphyho $\alpha$ funkcí byly vybrány výsledné polynomy vygenerované pomocí $m = 993610614932819879291$ a $p = 1563367$. Ohodnocení pomocí Murphyho $E$ nebudeme uvádět. První polynom je tvaru

$$\begin{aligned} f_1(x) =& 36960x^5 \\ &+1599104x^4 \\ &-9191978630014622379 26x^3 \\ &-893467474274113484 13x^2 \\ &-478893325288364132 82x \\ &-304364548712496087654817236864872 64. \end{aligned}$$

Druhý polynom je tvaru

$$f_2(x) = 4817891x - 728322830495799313686.$$

Jedná se o polynomy před zkosením. Tyto polynomy byly vybrány podle hodnoty $\alpha = -1,9647$. Algoritmus našel 23 polynomů s hodnotou $\alpha$ lepší než $-1,6$.

Na první polynom jsou následně aplikovány translace a rotace, popsané v 5.4. Zvolený upravený polynom vychází tvaru

$$\begin{aligned} f_1(x) =& 174660x^5 \\ &+134383x^4 \\ &-141624959209894293 6x^3 \\ &-262916520657205217 389x^2 \\ &+107999584970348579818728334753105567758903089985879227723677699 4 \\ &\quad 430192488473179027976672613147146729824023981605172651663665470x \\ &-166319360854336812920841635519782574348710758578254010694463657 1 \\ &\quad 422496432248695703084075824246902616628782978084101880071003091 00. \end{aligned}$$

Druhý polynom zůstává stejný. Ohodnocení polynomu je nyní $\alpha = -2,7164$. Obecně po úpravách bylo získáno 10 polynomů s $\alpha$ lepším než $-2,5$. Úpravy tímto změní vlastnost, kdy jsou poslední koeficienty řádově o velikosti $m$. Ohodnocení polynomů vychází však lepší než pro polynomy, které byly pouze vygenerovány.

# Kapitola 7

# Závěr

V první části práce byl popsán algoritmus hledání polynomů a algoritmus hledání odmocniny v číselném sítu. Pro implementaci byl zvolen algoritmus hledání polynomů, kterému je věnován zbytek práce. V dnešní době jsou považovány za nejefektivnější dvě varianty generování polynomů podle Kleinjunga. Samotné porovnání obou Kleinjungových algoritmů je v současnosti pouze teoretické. Druhý Kleinjungův algoritmus obsahuje jednodušší výpočty, než první Kleinjungův algoritmus. Obsahuje výpočty pro které lze snáze nahlédnout jejich opodstatnění. Není však obecně dokázáno ani porovnáno, že by jeden z těchto dvou algoritmů generoval lepší polynomy.

Pro implementaci byl zvolen Druhý Kleinjungův algoritmus. V současném stavu produkuje polynomy s dobrým ohodnocením. Dále by bylo zajímavé zaměřit se na přizpůsobení metod pro zlepšení vygenerovaných polynomů Kleinjungově druhému algoritmu. Při současném stavu ztrácíme vlastnost, kdy jsou poslední dva koeficienty o velikosti prvního kořene $m$. Ostatní koeficienty si řádově odpovídají.

Pokud by se podařilo výrazně urychlit chod celého algoritmu číselného síta například pomocí lepších polynomů, bylo by nutné změnit celkový pohled na dnešní asymetrickou kryptografii. Celé číselné síto a obecně faktorizace velkých čísel 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 Comptuttion, 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] S. Bai, *Root optimization of polynomials in the number field sieve*, 2012

[7] S. Bai, C. Bouvier, A. Kruppa, P. Zimmermann *Better polynomials for GNFS*, 2014

[8] R. Barbulescu, A. Lachand *Some mathematical remarks on the polynomial selection in NFS*, 2014

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

[10] Elkenbracht-Huizing M.*Marije Elkenbracht-Huizing* Experimental Mathematics, 1996

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

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

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

[14] K. Dickman *On the Frequency of Numbers Containing Prime Factors of a Certain Relative Magnitude* Ark. Mat., Astronomi och Fysik 22A, 1-14, 1930

[15] J.E Nymann *On the probability that k positive integers are relatively prime*, 1970

[16] W. Ekkelkamp, *Predicting the Sieving Effort for the Number Field Sieve*, 2008

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

[18] A. K. Lenstra, H. W. Lenstra Jr. *The Development of the Number Field Sieve*, Springer, 1993

[19] E. Thomé *Square Root Algorithms for the Number Field Sieve*, 2012

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

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

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

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

[24] *http://en.wikipedia.org/wiki/RSA_numbers#RSA-110*