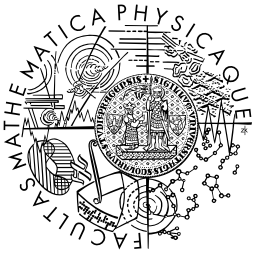

MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

DIPLOMOVÁ PRÁCE

Samuel Staško

Prosívání ve faktorizačních algoritmech

Katedra algebry

Vedoucí diplomové práce: doc. Mgr. Pavel Příhoda, Ph.D.

Studijní program: Matematika

Studijní obor: Matematika pro informační technologie

Praha 2023

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

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 má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ....

Podpis autora

Děkuji doc. Mgr. Pavlu Příhodovi, Ph.D. za cenné rady a pečlivou a ochotnou pomoc v celém průběhu tvorby práce.

Název práce: Prosívání ve faktorizačních algoritmech

Autor: Samuel Staško

Katedra: Katedra algebry

Vedoucí diplomové práce: doc. Mgr. Pavel Příhoda, Ph.D., Katedra algebry

Abstrakt: Kvadratické a číselné síto jsou dvě tradiční faktorizační metody. Uvádíme zde princip fungování obou těchto algoritmů, přičemž se zaměřujeme především na výpočet asymptotické složitosti. Největší důraz klademe na rozbor prosívací fáze. Hlavním cílem práce je však popis různých modifikací, odhad jejich časové složitosti a porovnání praktické využitelnosti se základními verzemi. Kromě několika známých variant prezentujeme vlastní návrhy jak kvadratického, tak číselného síta a podrobně analyzujeme jejich výhody či nevýhody.

Klíčová slova: faktorizace, prosívání, složitost, kvadratické síto, číselné síto

Title: Sieving in factoring algorithms

Author: Samuel Staško

Department: Department of Algebra

Supervisor: doc. Mgr. Pavel Příhoda, Ph.D., Department of Algebra

Abstract: The quadratic sieve and the number field sieve are two traditional factoring methods. We present here a principle of operation of both these algorithms, focusing mainly on the calculation of asymptotic complexity. The greatest emphasis is placed on the analysis of the sieving phase. However, the main goal of this work is to describe various modifications, estimate their time complexity and compare their practical usability with the basic versions. Apart from several well-known variants, we present our own proposals of both quadratic and number field sieve and analyze their advantages and disadvantages in detail.

Keywords: factorization, sieving, complexity, quadratic sieve, number field sieve

Obsah

Úvod **2**

1 Kvadratické síto **3**
1.1 Obecný princip . . . 3
1.2 Složitost . . . 5
1.3 Modifikace . . . 10
1.3.1 Varianta s velkým prvočíslem . . . 10
1.3.2 MPQS . . . 13
1.3.3 Vlastní verze MPQS . . . 14

2 Číselné síto **27**
2.1 Obecný princip . . . 27
2.2 Složitost . . . 31
2.3 Modifikace . . . 35
2.3.1 Kvadratické charaktery . . . 37
2.3.2 Verze s více polynomy . . . 39
2.3.3 Verze s více počítači . . . 44
2.3.4 Verze s předvýpočtem hladkých relací . . . 47
2.4 Randomizované číselné síto (RNFS) . . . 50
2.4.1 Obecný princip . . . 51
2.4.2 Složitost . . . 57
2.4.3 Randomizované číselné síto s více polynomy . . . 60

Závěr **64**

Seznam použité literatury **65**

Úvod

Rozklad čísel na součin prvočísel je bezpochyby jednou z nejstudovanějších oblastí matematiky. Motivací k rozvoji teorie v tomto směru dává především kryptografie, zejména kryptosystém RSA, který stojí na předpokladu, že nedokážeme efektivně faktorizovat. Tuto premisu se navzdory dlouhotrvající snaze pořád nepodařilo spolehlivě vyvrátit. My se budeme věnovat dvěma algoritmům, které k tomu, společně s metodou eliptických křivek (ECM), mají nejblíže.

V první kapitole si představíme kvadratické síto. To je poměrně jednoduchý algoritmus vhodný pro faktorizaci menších čísel (do 100 cifer). První polovina kapitoly je věnována vysvětlení jeho principu a výpočtu složitosti, přičemž vycházíme zejména z Crandall a Pomerance (2001). Dále se zabývá různými modifikacemi základní verze, které se snaží urychlit výpočet v praxi, přestože asymptotická složitost zůstává stejná. Velký prostor pro taková vylepšení nabízí především verze s více polynomy (MPQS). Tady si kromě standardně používané varianty ukážeme i vlastní návrh. Ten představujeme v druhé polovině kapitoly. Stejně jako ostatní modifikace sice nesnižuje asymptotickou časovou složitost, heuristicky ovšem pozorujeme vyšší frekvenci nalézání hladkých relací, díky čemu lze jisté zrychlení očekávat. Experimenty na nižších hodnotách navíc využíváme k návrhu (resp. potvrzení našich hypotéz) volby optimálních parametrů i pro větší vstupy a odhadu zlepšení oproti základní verzi kvadratického síta.

Druhá kapitola se zabývá číselným sítem, které je asymptoticky nejrychlejším známým faktorizačním algoritmem. Jeho podstata vychází z kvadratického síta, které v jistém smyslu zobecňuje. Na rozdíl od něj už nepracuje pouze v $\mathbb{Z}$, vyžaduje si proto také výpočty v číselných tělesech. První sekce kapitoly je tedy kromě principu z velké části věnována teorii v této oblasti. Následně si představíme podrobný výpočet asymptotické složitosti tohoto algoritmu včetně odůvodnění volby jednotlivých parametrů. Poslední, rozsahově největší část obsahuje popis různých modifikací základní verze. U žádné z nich samozřejmě nechybí analýza časové složitosti. Navíc zde prezentujeme vlastní návrh verze s více počítači (a několika jejich variant). Samostatnou sekci vyčleňujeme pro moderní modifikaci zvanou randomizované číselné síto, původně prezentovanou v Lee a Venkatesan (2018). Jeho hlavní výhodou je nahrazení veškerých heuristických předpokladů rigorózní analýzou při zachování časové složitosti (v průměrném případě). V závěrečné sekci 2.4.3 navrhneme, jak jej lze upravit tak, abychom byli schopni analogických výsledků dosáhnout také pro jednu z nejznámějších modifikací číselného síta, verzi s více polynomy.

1. Kvadratické síto

1.1 Obecný princip

Nechť N je liché přirozené číslo dělitelné alespoň dvěma různými prvočísly. Chceme najít $x, y \in \mathbb{Z}$ splňující

- $x^2 \equiv y^2 \pmod{N}$
- $x \not\equiv \pm y \pmod{N}$

Pak je totiž $\mathrm{NSD}(N, x \pm y)$ vlastním dělitelem N.

Zvolme *hladkou mez* $B \in \mathbb{N}$ a položme $B_{\mathbb{Z}} = \{-1\} \cup \{p \text{ prvočíslo} : p \leq B\}$. Tuto množinu budeme nazývat *faktorizační báze*. Řekneme, že (x, y), kde $x, y \in \mathbb{Z}$, je *$B_{\mathbb{Z}}$-hladká relace*, jestliže

- $x^2 \equiv y \pmod{N}$
- y je *$B_{\mathbb{Z}}$-hladké*, neboli je součinem prvků z $B_{\mathbb{Z}}$

V této kapitole budeme pořád pracovat nad $\mathbb{Z}$, pro přehlednost proto budeme psát pouze „B-hladké“, nebo „hladké“.

Naším cílem je nalézt hladké relace (x_i, y_i), $i = 1, \ldots, k$, takové, že $\prod_{i=1}^{k} y_i = y^2$ pro nějaké y. Je-li $x = \prod_{i=1}^{k} x_i$, potom $x^2 \equiv y^2 \pmod{N}$. Následně ověříme, zda platí $x \not\equiv \pm y \pmod{N}$. Pokud ne, musíme hledat dál. Statisticky bychom ale brzy měli dospět k vyhovující dvojici x, y.

Tvrzení 1.1. *Pro náhodně vybranou dvojici nesoudělných celých čísel x, y splňující $x^2 \equiv y^2 \pmod{N}$ platí $x \not\equiv \pm y \pmod{N}$ s pravděpodobností alespoň $1/2$.*

Důkaz. Z Čínské zbytkové věty plyne, že existuje 2^k odmocnin z 1 modulo N, kde k je počet prvočíselných dělitelů N. Dvě z nich jsou triviální, a sice ± 1. V naší situaci je $k > 1$, tudíž existují alespoň 4. Takže těch netriviálních je alespoň polovina. Jestliže $x^2 \equiv y^2 \pmod{N}$, tak xy^{-1} je odmocnina z 1. Protože x a y byly zvoleny náhodně, měli bychom nejméně v polovině případů dostat netriviální odmocninu. Ekvivalentně $x \not\equiv \pm y \pmod{N}$. □

Jednotlivé hladké relace budeme hledat přirozeným způsobem — zvolíme x, dopočítáme $y = x^2 \pmod{N}$ a ověříme, zda je y B-hladké. Samozřejmě, největší šanci na úspěch budeme mít, pokud y (v absolutní hodnotě) bude co nejmenší. Ideální by tedy bylo začít volbou $x = \lceil \sqrt{N} \rceil$ a postupně ho zvětšovat. Všimněme si, že až po $x = \lfloor \sqrt{2N} \rfloor$ lze y určit předpisem $x^2 - N$. Připustíme-li navíc záporné hodnoty pravých stran relací (což nám umožní zahrnutí -1 do faktorizační báze), bude to platit rovněž pro $x = \{1, \ldots, \lceil \sqrt{N} \rceil\}$. Náš *prosívací interval* pro výběr hodnot x bude tedy

$$I = \{\lceil \sqrt{N} \rceil - M, \ldots, \lceil \sqrt{N} \rceil + M\}.$$

Volbě parametru M se budeme věnovat později, nyní si ukážeme tzv. prosívací fázi kvadratického síta.

Zásadní podmínkou, kterou musí relace splňovat, aby tato fáze fungovala, je následující: jsou-li (x, y) a (x', y') relace, pak pro každé $q \in \mathbb{N}$

$$q \mid (x - x') \Rightarrow q \mid (y - y')\,.$$

Snadno ověříme, že toto v našem případě platí. Díky tomu lze simultánně prosívat více hodnot najednou. Postup vypadá takto:

1. Pro každé $x \in I$ spočteme $y(x) = x^2 - N$. Můžeme předpokládat $\sqrt{N} \notin \mathbb{Q}$ (jinak bychom mohli faktorizovat), takže $y(x) \neq 0$.
2. Pro každé prvočíslo $p \in B_{\mathbb{Z}}$ najdeme $C_p := \{c \in \mathbb{Z}_p : y(c) \equiv 0 \pmod{p}\}$.
3. Pro $c \in C_p$ máme $J_{c,p} := \{c + kp : k \in \{\lceil \frac{\lceil\sqrt{N}\rceil - M - c}{p} \rceil, \ldots, \lfloor \frac{\lceil\sqrt{N}\rceil + M - c}{p} \rfloor\}\}$.
4. Pro každé $c \in C_p$ a $x \in J_{c,p}$ vydělíme hodnotu $y(x)$ nejvyšší mocninou p, kterou je dělitelná, čili $p^{v_p(y(x))}$.

Hladké relace jsou teď přesně ty, kde $y(x) = \pm 1$.

Otázkou zůstává, kolik hladkých relací (x_i, y_i) potřebujeme najít, abychom z nich dokázali nakombinovat dvojici x, y splňující $x^2 \equiv y^2 \pmod{N}$. Nechť m je B-hladké. Potom

$$m = (-1)^{e_0} \prod_{i=1}^{\pi_B} p_i^{e_i},$$

kde p_i je i-té prvočíslo. (Připomeňme, že π_B značí počet prvočísel menších nebo rovných B, přičemž podle prvočíselné věty je $\pi_B \approx B/\ln B$.) Definujme vektor mocnin $v(m) = (e_0, e_1, \ldots, e_{\pi_B})$. Pokud má být $\prod y_i$ čtverec, musí mít $\sum v(y_i)$ všechny souřadnice sudé. Stačí proto, když budeme jednotlivé vektory mocnin počítat modulo 2. Nyní tedy $\sum v(y_i)$ musí být nulový vektor. Jinými slovy, hledáme řešení homogenní soustavy lineárních rovnic nad $\mathbb{F}_2$.

Tím se dostáváme k závěrečné, lineární fázi kvadratického síta. Do řádků matice soustavy zapisujeme vektory mocnin, sloupce budou indexovány prvky faktorizační báze. Těch je $\pi_B + 1$, abychom tedy měli jistotu, že nalezneme netriviální řešení, potřebujeme alespoň $\pi_B + 2$ řádků (vektorů mocnin).

Poznámka. Ve skutečnosti budeme prosívat jenom asi polovinou (heuristický odhad) prvočísel z faktorizační báze, konkrétně těmi, jejichž Legendrův symbol $\left(\frac{N}{p}\right) = 1$. Pro taková prvočísla p bude C_p dvouprvková množina, s jedinou výjimkou, a sice $C_2 = \{1\}$.

Z toho plyne, že asi $\pi_B/2$ souřadnic (příslušných prvočíslům $p \in B_{\mathbb{Z}}$, pro něž $x^2 \equiv N \pmod{p}$ nemá řešení) všech vektorů je nulových. Takže abychom měli jistotu, že takovou množinu najdeme, nepotřebujeme najít až $\pi_B + 2$ různých vektorů (hladkých relací), stačí nám jich přibližně $\pi_B/2$. My budeme ale dál pracovat s nejhorším odhadem, tedy že jich potřebujeme $\pi_B + 2$. Jak uvidíme později, asymptotickou složitost to nijak neovlivní.

Dokončeme ještě diskuzi o řešení soustavy. Jeden způsob, kterým ho lze najít, je Gaussova eliminační metoda. Jelikož se ale jedná o řídkou matici, můžeme použít také Wiedemannovu metodu (viz Wiedemann (1986)), která je asymptoticky rychlejší. My se této fázi ovšem dál věnovat nebudeme, takže to ponecháme bez podrobnějšího rozboru.

1.2 Složitost

Doposud jsme si vystačili s předpokladem, že máme nějakou předem zvolenou hladkou mez B. Je však zřejmé, že výpočetní složitost algoritmu bude na její hodnotě záviset. Základním úkolem je najít $\pi_B + 2$ B-hladkých relací. Čím větší B, tím snáz se nám budou hladké relace hledat. Na druhou stranu jich ale budeme potřebovat víc.

Rádi bychom tedy našli jakousi „zlatou střední cestu“, čili optimální hodnotu B, při níž na hladké relace narazíme dostatečně často a zároveň jich nebudeme potřebovat příliš moc. To nás přivádí ke stěžejní otázce: jaká je pravděpodobnost, že $x^2 - N$ bude B-hladké? Definujme pro $x, y \in \mathbb{R}^+$

$$\psi(x, y) := |\{n \in \mathbb{Z} : |n| \leq x \ \& \ n \text{ je } y\text{-hladké}\}|.$$

Následující tvrzení nám nastíní odpověď.

Věta 1.2. *Nechť $x \geq 1$ a B je hladká mez. Jestliže $3 \ln B \leq \ln x < B$, pak*

$$\psi(x, B) = 2xu^{-u+o(u)},$$

kde $u = \frac{\ln x}{\ln B}$.

Poznámka. Záležet nám bude především na následující (ekvivalentní) interpretaci: pravděpodobnost, že náhodně vybrané celé číslo z intervalu $[-x, x]$ bude B-hladké, je přibližně u^{-u}.

Důkaz. Rigorózní důkaz je hodně zdlouhavý, představíme si tady proto pouze jeho ideu, aniž bychom zabíhali do některých technických detailů. Úplnou verzi lze najít v Canfield a kol. (1983).

Zásadním pozorováním je, že hodnota $\psi(x, B)$ se rovná počtu uspořádaných nezáporných $(\pi_B + 1)$-tic $(e_0, \ldots, e_{\pi_B})$, kde $e_0 \in \{0, 1\}$, splňujících

$$\left| (-1)^{e_0} \prod_{i=1}^{\pi_B} p_i^{e_i} \right| \leq x,$$

neboli dvojnásobku počtu uspořádaných nezáporných π_B-tic $(e_1, \ldots, e_{\pi_B})$ splňujících

$$\prod_{i=1}^{\pi_B} p_i^{e_i} \leq x.$$

Po aplikování logaritmu na obě strany nerovnosti dostaneme

$$\sum_{i=1}^{\pi_B} e_i \ln p_i \leq \ln x.$$

V tomto okamžiku přichází první hrubý odhad. Pro většinu prvočísel p_i bude jejich přirozený logaritmus řádově stejný jako $\ln B$. My je proto tímto způsobem aproximujeme všechny. Následně získáme

$$\sum_{i=1}^{\pi_B} e_i \ln B \leq \ln x,$$

$$\sum_{i=1}^{\pi_B} e_i \leq u.$$

Kolik takových π_B-tic existuje? Můžeme si to představit tak, že rozdělujeme $\lfloor u \rfloor$ jedniček, každou do jedné z π_B pozic, nebo nikam (každé tedy můžeme přiřadit číslo od 1 do $\pi_B + 1$). Jedná se tedy o $\lfloor u \rfloor$-členné kombinace s opakováním z $\pi_B + 1$ prvků. Jejich počet je $\binom{\lfloor u \rfloor + \pi_B}{\pi_B}$.

Nyní jsme schopni spočítat kýženou pravděpodobnost.

$$\begin{aligned}\ln \frac{\psi(x,B)}{2x} &= \ln \psi(x,B) - \ln 2 - \ln x \\ &\approx \ln\left(2\binom{\lfloor u \rfloor + \pi_B}{\pi_B}\right) - \ln 2 - u \ln B \\ &= \ln \binom{\lfloor u \rfloor + \pi_B}{\pi_B} - u \ln B.\end{aligned}$$

Použitím vzorce pro kombinační čísla a Stirlingovy aproximace, která říká $\ln(n!) \approx n \ln n - n$ (viz např. Conrad (2016)), dostaneme

$$\begin{aligned}\ln \frac{\psi(x,B)}{2x} &\approx \ln \frac{(\lfloor u \rfloor + \pi_B)!}{\lfloor u \rfloor ! \pi_B!} - u \ln B \\ &\approx (\lfloor u \rfloor + \pi_B) \ln(\lfloor u \rfloor + \pi_B) - (\lfloor u \rfloor + \pi_B) - (\lfloor u \rfloor \ln u - \lfloor u \rfloor) - \\ &\quad - (\pi_B \ln \pi_B - \pi_B) - u \ln B.\end{aligned}$$

Dále odhadneme $\lfloor u \rfloor \approx u$ a $\ln(\lfloor u \rfloor + \pi_B) \approx \ln \pi_B$ podle $u = \frac{\ln x}{\ln B} < \frac{B}{\ln B} \approx \pi_B$, díky čemu se nám většina členů pokrátí a zbude pouze

$$\ln \frac{\psi(x,B)}{2x} \approx u \ln \pi_B - u \ln u - u \ln B \approx -u \ln u - u \ln \ln B.$$

Tady lze opět z předpokladů odvodit, že $\ln \ln B$ je zanedbatelné v porovnání s $\ln u$. Nyní jsme se tedy konečně dopracovali k

$$\frac{\psi(x,B)}{2x} \approx u^{-u},$$

což jsme chtěli dokázat. □

Nás tedy zajímá $\psi(N,B)$. Máme tudíž

$$u = \frac{\ln N}{\ln B}.$$

Je asi jasné, že $u \geq 3$ (B musí být určitě řádově nižší než N, aby byl algoritmus reálně proveditelný). Naopak, druhá podmínka, která po snadné úpravě říká $\ln N < B$, již na první pohled zřejmá není. Na konci výpočtu se k ní vrátíme a zkontrolujeme, zda je splněna.

Zatím (přirozeně) předpokládáme, že $|x^2 - N| \leq N$. Tento odhad lze ovšem vylepšit. Uvědomme si, že pokud $\sqrt{N} - N^\epsilon \leq x \leq \sqrt{N} + N^\epsilon$ pro malé $\epsilon > 0$, pak

$$\left|x^2 - N\right| \leq \left|\left(\sqrt{N} + N^\epsilon\right)^2 - N\right| = 2\sqrt{N}N^\epsilon + N^{2\epsilon} = O\left(N^{\frac{1}{2}+\epsilon}\right).$$

Tady se nabízí otázka, jestli je to legitimní omezení hodnot prosívacího intervalu. Opět se na věc lze dívat intuitivně (očekáváme řádově mnohem nižší počet pokusů generování hladkých relací než N), nebo si počkat na výsledek a zpětně to ověřit. Za těchto podmínek každopádně pracujeme s $\psi\left(N^{1/2}, B\right)$. Dostaneme tedy

$$u = \frac{\ln N}{2 \ln B}.$$

Poznámka. Lze namítat, že Věta 1.2 počítá s výběrem z celého intervalu, který je vycentrovaný kolem nuly, nikoliv z nějaké jeho speciální podmnožiny, kterou vygeneruje polynom $x^2 - N$. To je sice pravda, my ale budeme heuristicky předpokládat, že na tom nezáleží (neboli že tyto hodnoty jsou v podstatě náhodné). Navíc, kontrolní testy provedené na dostupných hodnotách naznačují, že tato hypotéza je celkem přesná.

Nyní tedy víme, že očekávaný počet pokusů pro nalezení jedné hladké relace je u^u. Již dříve jsme spočítali, že jich potřebujeme $\pi_B + 2$. To nám naznačuje, jak volit parametr M. Zatím ovšem jenom v závislosti na B i N. Na konci výpočtu, až nalezneme optimální B, se k tomu vrátíme a vyjádříme ho vzhledem k oběma těmto proměnným zvlášť.

Než budeme pokračovat, udělejme si přehled některých funkcí a tvrzení, které použijeme.

Definice. Ať $n \in \mathbb{N}$. Pak

$$\vartheta(n) = \sum_{\substack{p \leq n \\ p \text{ prvočíslo}}} \ln p$$

nazýváme *první Čebyševovou funkcí.*

O této funkci je známo, že se asymptoticky chová jako $O(n)$. Přesvědčit se o tom lze třeba v článku Rosser a Schoenfeld (1962).

Věta 1.3 (Druhá Mertensova)**.**

$$\lim_{n\to\infty} \left(\sum_{\substack{p \leq n \\ p \text{ prvočíslo}}} \frac{1}{p} - \ln \ln n \right) = M,$$

kde $M \approx 0{,}26$ *je Meissel–Mertensova konstanta.*

Důkaz může čtenář opět najít v Rosser a Schoenfeld (1962).

Definice. Pro $s \in \mathbb{N}$ definujeme *prvočíselnou zeta funkci* $P(s)$ předpisem

$$P(s) = \sum_{p \text{ prvočíslo}} \frac{1}{p^s}.$$

Nás bude zajímat především $P(2)$, konkrétně informace, že konverguje k hodnotě přibližně 0,45.

Jak dlouho nám tedy zabere prosívání intervalu $I = [-M,M]$? Nejdřív nalezneme kořeny $x^2 - N$ modulo p pro každé $p \in B_{\mathbb{Z}}$, $\left(\frac{N}{p}\right) = 1$. To lze provést pomocí Tonelliho-Shanksova algoritmu za cenu

$$\sum_{p \leq B} \ln p = \vartheta(B) \approx B$$

aritmetických operací (za ně považujeme jakékoli operace proveditelné v čase $O\left(\ln^2 N\right)$).

Poznámka. Je pravdou, že Tonelliho-Shanksův algoritmus má v nejhorším případě složitost až $O\left(\ln^4 p\right)$. To se stane, když v rozkladu $p-1 = 2^s t$, kde t liché, vyjde s velké. Pokud ale počítáme kořeny pro více prvočísel p, což je přesně naše situace, dává větší smysl uvažovat průměrný případ, jehož složitost je pouze $O\left(\ln^3 p\right)$. Podrobněji viz Crandall a Pomerance (2001), Algoritmus 2.3.8.

Pro všechna $p \in B_{\mathbb{Z}}$, $\left(\frac{N}{p}\right) = 1$ jsme takto sestavili množiny C_p a $J_{c,p}$. Pokračujeme postupným dělením. V nejhorším případě je $|C_p| = 2$ a $|J_{c,p}| = 2M + 1$. Provedeme tedy nanejvýš

$$\begin{aligned} &2\left(\sum_{\substack{p\le B\\ \left(\frac{N}{p}\right)=1}} \frac{2M+1}{p} + \sum_{\substack{p^2\le B\\ \left(\frac{N}{p}\right)=1}} \frac{2M+1}{p^2} + \sum_{\substack{p^3\le B\\ \left(\frac{N}{p}\right)=1}} \frac{2M+1}{p^3} + \cdots\right) \\ &\approx 4M\left(\sum_{p\le B}\frac{1}{p} + \sum_{p^2\le B}\frac{1}{p^2} + \sum_{p^3\le B}\frac{1}{p^3} + \cdots\right) \end{aligned}$$

dělení.

Podle druhé Mertensovy věty je $\sum_{p\le B} 1/p \approx \ln\ln B$. Zbylé členy v závorce odhadneme seshora a ukážeme, že jsou asymptoticky zanedbatelné v porovnání s tím prvním:

$$\sum_{k=2}^{\infty}\sum_{p^k\le B}\frac{1}{p^k} \le \sum_{k=2}^{\infty}\sum_{p\le B}\frac{1}{p^k} = \sum_{p\le B}\sum_{k=2}^{\infty}\frac{1}{p^k} = \sum_{p\le B}\frac{1/p^2}{1-1/p} = \sum_{p\le B}\frac{1}{p^2-p} \approx \sum_{p\le B}\frac{1}{p^2}.$$

Dopracovali jsme se k výrazu velice podobnému již zmíněné prvočíselné zeta funkci $P(2)$. Můžeme jej proto seshora odhadnout hodnotou 0,45. Takže v kombinaci s hledáním kořenů dohromady provedeme asi

$$\begin{aligned} T(B) &\approx B + 4M(\ln\ln B + 0{,}45) \\ &\approx B + 4M\ln\ln B \\ &\approx B + 4\pi_B\, u^u \ln\ln B \\ &\approx 4\frac{B}{\ln B}\, u^u \ln\ln B \end{aligned}$$

operací, kde $u = \frac{\ln N}{2\ln B}$.

Nyní se pokusíme najít minimum této funkce. Aby se nám líp počítalo, budeme ho hledat pro její logaritmus

$$S(B) = \ln T(B) = \ln 4 + \ln B + u\ln u + \ln\ln\ln B - \ln\ln B \approx \ln B + u\ln u.$$

Dosaďme za u a derivujme:

$$\begin{aligned} \frac{dS}{dB} &= (\ln B)' + \left(\frac{\ln N}{2\ln B}\right)' \ln\frac{\ln N}{2\ln B} + \frac{\ln N}{2\ln B}\left(\ln\frac{\ln N}{2\ln B}\right)' \\ &= \frac{1}{B} - \frac{\ln N}{2}\cdot\frac{1}{\ln^2 B}\cdot\frac{1}{B}\cdot\ln\frac{\ln N}{2\ln B} + \frac{\ln N}{2\ln B}\cdot\frac{2\ln B}{\ln N}\cdot\left(-\frac{\ln N}{2}\cdot\frac{1}{\ln^2 B}\cdot\frac{1}{B}\right) \\ &= \frac{1}{B} - \frac{\ln N}{2B\ln^2 B}\left(\ln\frac{\ln N}{2\ln B} + 1\right) \\ &= \frac{1}{B} - \frac{\ln N}{2B\ln^2 B}\left(\ln\ln N - \ln\ln B - \ln 2 + 1\right). \end{aligned}$$

Tento výraz zjednodušíme aproximací $\ln 2 \approx 1$ a položíme roven nule:

$$\begin{aligned} 0 &= \frac{2\ln^2 B - \ln N\,(\ln\ln N - \ln\ln B)}{2B\ln^2 B} \\ 0 &= \ln^2 B - \frac{\ln N\,(\ln\ln N - \ln\ln B)}{2} \\ \ln B &= \frac{\sqrt{2}}{2}\sqrt{\ln N\,(\ln\ln N - \ln\ln B)}. \end{aligned}$$

Z toho snadno odvodíme dolní a horní mez pro $\ln B$:

$$\frac{\sqrt{2}}{2}\sqrt{\ln N} < \ln B < \frac{\sqrt{2}}{2}\sqrt{\ln N \ln\ln N}.$$

Samozřejmě nevíme, kde přesně mezi těmito hodnotami $\ln B$ leží. Kdyby bylo uprostřed, tak by platilo

$$\ln B = \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{\ln N} + \sqrt{\ln N \ln\ln N}}{2} \approx \frac{\sqrt{\ln N \ln\ln N}}{2\sqrt{2}}.$$

Zkusme si to ale vyjádřit obecněji jako

$$\ln B = c\sqrt{\ln N \ln\ln N}$$

pro nějakou konstantu c. Její hodnotu odhadneme dosazením do rovnosti, z níž jsme odvodili meze pro $\ln B$:

$$\begin{aligned} 0 &= c^2 \ln N \ln\ln N - \frac{\ln N\left(\ln\ln N - \ln c - \frac{1}{2}\ln\ln N - \frac{1}{2}\ln\ln\ln N\right)}{2}, \\ 0 &= \left(c^2 - \frac{1}{4}\right)\ln N \ln\ln N + \frac{1}{4}\ln N \ln\ln\ln N + \frac{1}{2}\ln N \ln c. \end{aligned}$$

Chceme-li pravou stranu co nejvíc přiblížit nule, musíme vynulovat asymptoticky největší člen, kterým je $\ln N \ln\ln N$. Potřebujeme tedy

$$\begin{aligned} 0 &= c^2 - \frac{1}{4}, \\ c &= \pm\frac{1}{2}. \end{aligned}$$

Záporné c můžeme zřejmě vyloučit. Naopak, $c = \frac{1}{2}$ splňuje odvozené meze

$$\frac{\sqrt{2}}{2}\sqrt{\ln N} < c\sqrt{\ln N \ln\ln N} < \frac{\sqrt{2}}{2}\sqrt{\ln N \ln\ln N}$$

(pro $N > \exp(\exp(2)) \approx 1618$).

Tím jsme se dopracovali ke kýženému výsledku. Jednoduchou úpravou dojdeme k tomu, že optimální hodnota hladké meze v kvadratickém sítu je přibližně

$$B = \exp\left(\frac{\sqrt{\ln N \ln\ln N}}{2}\right).$$

V průběhu výpočtu jsme udělali dva předpoklady, které teď ověříme a ukážeme, že výsledek je korektní.

- $\ln N < B$: víme, že $\ln B \approx \frac{\sqrt{\ln N \ln\ln N}}{2} > \ln\ln N$ pro každé $N > 3$, z čeho již plyne tvrzení.
- $\sqrt{N} - N^{\epsilon} \leq x \leq \sqrt{N} + N^{\epsilon}$ pro malé $\epsilon > 0$: toto si vyžaduje vyjádřit M pomocí N, což jsme doposud také neudělali. Teď jsme toho už schopni:

$$\begin{aligned}
\ln M &\approx \ln B - \ln\ln B + u \ln u \\
&\approx \ln B + u \ln u \\
&\approx \frac{\sqrt{\ln N \ln\ln N}}{2} + \frac{\sqrt{\ln N}}{\sqrt{\ln\ln N}} \ln \frac{\sqrt{\ln N}}{\sqrt{\ln\ln N}} \\
&\approx \frac{\sqrt{\ln N \ln\ln N}}{2} + \frac{\sqrt{\ln N}}{2\sqrt{\ln\ln N}} (\ln\ln N - \ln\ln\ln N) \\
&\approx \frac{\sqrt{\ln N \ln\ln N}}{2} + \frac{\sqrt{\ln N}}{2\sqrt{\ln\ln N}} \ln\ln N \\
&\approx \sqrt{\ln N \ln\ln N} \\
M &\approx \exp\left(\sqrt{\ln N \ln\ln N}\right).
\end{aligned}$$

  Vidíme, že M je subexponenciální v N, takže náš odhad je v pořádku. (Z výpočtu mimochodem plyne též $M \approx B^2$.)

Nyní dosadíme za u a spočítáme složitost prosívací fáze:

$$\begin{aligned}
T(B) &= \exp(S(B)) \\
&= \exp(\ln B + u \ln u) \\
&= \exp(\ln M) \\
&= \exp\left(\sqrt{\ln N \ln\ln N}\right) \\
&= B^2.
\end{aligned}$$

Složitost lineární fáze podrobně studovat nebudeme, lze nicméně ukázat, že s využitím rychlých metod pro řešení soustav s řídkou maticí (jednou z možností je již zmiňovaná Wiedemannova metoda) vychází řádově taky B^2. Celková složitost kvadratického síta je tudíž

$$L(N) = \exp\left(\sqrt{\ln N \ln\ln N}\right).$$

1.3 Modifikace

V následující sekci si ukážeme několik variant algoritmu, které jsou sice asymptoticky stejně časově náročné, v realitě přesto typicky běží o něco rychleji. Tato diskrepance je způsobená tím, že během výpočtu složitosti docela hodně členů zanedbáváme nebo odhadujeme jednoduššími, čím ztrácíme drobné rozdíly mezi podobnými verzemi. To si ostatně ilustrujeme na každé z představených variant.

1.3.1 Varianta s velkým prvočíslem

V praxi se ukazuje, že frekvence nalézání hladkých relaci s hodnotami parametrů, které se používají v kvadratickém sítu, je poměrně malá. Z výpočtů ovšem

víme, že výrazné zvětšení hladké meze B není řešením, protože jsme ji stanovili jako optimální vzhledem k složitosti celého algoritmu. Pořád bychom ale chtěli najít způsob, jak počet relací zvýšit. K tomu nám poslouží parciální relace.

Definice. Nechť $x, y \in \mathbb{Z}$. Řekneme, že (x, y) je *parciální relace*, jestliže

- $x^2 \equiv y \pmod{N}$,
- $y = py'$, kde p je prvočíslo větší než B a y' je B-hladké.

Prvočíslo p se nazývá *velké prvočíslo parciální relace* (x, y).

Jak takové relace hledat? Zůstaneme-li u této obecné definice, tak celkem těžko. Museli bychom totiž všechny hodnoty na konci prosívání různé od ± 1 prohnat nějakým prvočíselným testem. Ačkoli zdaleka nejsou tak velké jako N, pořád by to byla značná komplikace. Navíc, jak si záhy ukážeme, příliš velká prvočísla nejsou vhodná ani pro následné zpracování v lineární fázi.

Řešení je naštěstí jednoduché — stačí omezit p seshora. Všimněme si, že pokud nám po prosívání někde zbude číslo různé od ± 1, které je zároveň v absolutní hodnotě menší než B^2, tak je to prvočíslo (větší než B). Přidáme-li tedy podmínku $p < B^2$ (v praxi se používá spíše $p < cB$ pro nějakou konstantu $c < B$, aby velkých prvočísel nebylo příliš moc, na tom ale v tuto chvíli nezáleží), poznáme parciální relace okamžitě.

Označme $\mathcal{P}$ množinu nalezených parciálních relací a seskupme ji podle velkého prvočísla jednotlivých relací. Relace $(x, y) \in \mathcal{P}$ se nazývá *singleton*, pokud má velké prvočíslo, které nemá žádná jiná relace z $\mathcal{P}$. Tyto relace můžeme zahodit, neboť kombinací s žádnou jinou evidentně nikdy nedostaneme čtverec (ve vektoru mocnin bude toto velké prvočíslo mít vždy exponent 1).

Předpokládejme, že máme $k \geq 2$ relací se stejným velkým prvočíslem p, čili $x_i^2 - N \equiv y_i p \pmod{N}$, $i = 1, \ldots, k$. Potom pro $i = 2, \ldots, k$ platí

$$(x_1 x_i)^2 \equiv y_1 y_i p^2 \pmod{N}.$$

Vidíme tedy, že můžeme k vektorů mocnin s velkým prvočíslem nahradit $k - 1$ vektory bez velkého prvočísla. Jiné kombinace nemá smysl uvažovat, protože pro $(x_j x_l)^2$, $j, l \neq 1$, bude vektor mocnin totožný s vektorem $(x_1 x_j)^2 (x_1 x_l)^2$. Tím se dostáváme do situace, kterou už známe, ovšem s množstvím nových řádků do matice lineární fáze.

Ještě bychom si měli rozmyslet, jestli nám po vyřazení singletonů zůstane dostatečný počet parciálních relací. Narozeninový paradox naštěstí naznačuje, že ano. Máme $\pi_{B^2} - \pi_B = \frac{B^2}{2 \ln B} - \frac{B}{\ln B} \approx \frac{B^2}{2 \ln B} = \pi_{B^2}$ velkých prvočísel. První shodu proto očekáváme asi po $\sqrt{\pi_{B^2}} = \frac{B}{\sqrt{2 \ln B}}$ pokusech, přičemž s jejich přibývajícím počtem pravděpodobnost dál roste. Připomeňme, že dohromady provedeme asi $M = B^2$ pokusů, což znamená, že množina P by i po zahození singletonů rozhodně měla být uspokojivě velká.

Zatím to tedy vypadá, že tento přístup je efektivnější než základní verze kvadratického síta. V praxi se opravdu často bez parciálních relací neobejdeme. Proč navzdory tomu složitost asymptoticky nevychází lépe?

Zajímá nás pravděpodobnost nalezení parciální relace. Zkusme replikovat postup z důkazu Věty 1.2. Označme

$$\psi'(x, B) := \psi(x, B) + \left| \{ n \in \mathbb{Z} : |n| \leq x \ \& \ n = pn', \ n' \ B\text{-hladké}, \ B < p < B^2 \} \right|.$$

Hodnotu prvního sčítance už známe, zbývá tudíž spočítat pouze druhý. Ten je roven počtu uspořádaných π_B-tic $(e_1, \ldots, e_{\pi_B})$ splňujících

$$p \prod_{i=1}^{\pi_B} p_i^{e_i} \leq x,$$

kde $B < p < B^2$, což po zlogaritmování dává

$$\ln p + \sum_{i=1}^{\pi_B} e_i \ln p_i \leq \ln x.$$

V dalším kroku jsme logaritmy prvočísel odhadli pomocí $\ln B$. To už ovšem nelze udělat pro velké prvočíslo p. U něj musíme aproximovat pomocí $\ln B^2$. Pak tedy dostaneme

$$\begin{aligned} 2 \ln B + \sum_{i=1}^{\pi_B} e_i \ln B &\leq \ln x, \\ \sum_{i=1}^{\pi_B} e_i &\leq u - 2. \end{aligned}$$

Takových π_B-tic je podle stejné úvahy jako ve vzorovém důkazu $\binom{\lfloor u-2 \rfloor + \pi_B}{\pi_B}$. Hledaná pravděpodobnost je tudíž

$$\begin{aligned} \ln \frac{\psi'(x,B)}{2x} &= \ln \left(2 \left(\binom{\lfloor u \rfloor + \pi_B}{\pi_B} + \binom{\lfloor u-2 \rfloor + \pi_B}{\pi_B} \right) \right) - \ln 2 - \ln x \\ &\approx \ln \left(2 \binom{\lfloor u \rfloor + \pi_B}{\pi_B} \right) - u \ln B \\ &= \ln 2 + \ln \binom{\lfloor u \rfloor + \pi_B}{\pi_B} - u \ln B \\ &= \ln \frac{\psi(x,B)}{2x} + \ln 2. \end{aligned}$$

Vidíme, že jedinou změnou je člen $\ln 2$, což asymptoticky žádnou roli nehraje. Nakonec totiž dospějeme k tvaru

$$\ln \frac{\psi(x,B)}{2x} \approx -u \ln u - u \ln \ln B + \ln 2,$$

kde je $\ln 2$ zanedbatelné.

Na tomto místě tedy rozdíl ztratíme a dojdeme k témuž výsledku jako v základní verzi (s hladkými relacemi), byť skutečná pravděpodobnost zřejmě musí být větší. Bohužel se ukazuje, že nemáme dostatečně jemné nástroje, abychom ji byli schopni rozlišit.

Co se týče lineární fáze, její složitost zůstává $O(B^2)$, neboť počet sloupců matice je pořád $O(B)$. Celková asymptotická složitost se proto oproti základní verzi nijak nezmění.

1.3.2 MPQS

Druhým způsobem, jak zvýšit pravděpodobnost nalezení hladkých relací, je snížit absolutní hodnotu čísel vstupujících do prosívací fáze. To má navíc za důsledek další významnou výhodu — menší absolutní hodnota obvykle také znamená méně prvočísel, kterými je dělitelná, což přináší časovou úsporu.

Docílit toho v základním nastavení je ovšem problematické. Délku prosívacího intervalu totiž nelze libovolně zkracovat, jelikož parametr M, který ji určuje, je volen tak, abychom dostali potřebný počet hladkých relací. Pokud bychom délku intervalu zvětšili, zvětší se též $|x^2 - N|$.

Řešením je místo $x^2 - N$ použít více polynomů. V průběhu výpočtu je budeme postupně měnit a to nám umožní udržet absolutní hodnotu v nižší mezi. Existují různé metody, jak tyto polynomy volit. My si teď ukážeme tu nejčastěji používanou, Montgomeryho (viz Pomerance (1984)).

Mějme kvadratický polynom $f(x) = ax^2 + 2bx + c$ s celočíselnými koeficienty splňující $b^2 - ac = N$. Potom

$$af(x) = a^2x^2 + 2abx + ac = (ax + b)^2 - N,$$

takže

$$(ax + b)^2 \equiv af(x) \pmod{N}.$$

V klasickém kvadratickém sítu požadujeme na pravé straně B-hladkou hodnotu, abychom ji mohli použít jako další řádek matice lineární fáze. Analogicky tedy chceme, aby a i $f(x)$ byly B-hladké. Protože parametr a zvolíme předem, stačí, když to bude libovolný čtverec krát B-hladká hodnota.

Řekněme, že budeme prosívat interval délky $2M$ (stejně jako doposud). Chtěli bychom co nejnižší $|f(x)|$, proto ho vycentrujeme kolem vrcholu paraboly, což je bod $\frac{-b}{a}$. Takže máme

$$I = \left[\frac{-b}{a} - M, \frac{-b}{a} + M\right]$$

a hledáme a, b, c takové, že

$$\left|f\left(\frac{-b}{a}\right)\right| \approx \left|f\left(\frac{-b}{a} - M\right)\right| = \left|f\left(\frac{-b}{a} + M\right)\right|.$$

Protože $b^2 - ac = N$, platí

$$\left|af\left(\frac{-b}{a}\right)\right| = N,$$

$$\left|af\left(\frac{-b}{a} - M\right)\right| = \left|af\left(\frac{-b}{a} + M\right)\right| = a^2M^2 - N.$$

Z toho plyne, že potřebujeme $a \approx \frac{\sqrt{2N}}{M}$. Následně vezmeme b jako řešení kongruence $b^2 \equiv N \pmod{a}$ a $c = (b^2 - N)/a$. V této situaci je tudíž

$$|f(x)| \leq \frac{M}{\sqrt{2}}\sqrt{N}.$$

V základní verzi kvadratického síta, kde se používá pouze polynom $x^2 - N$, jsou absolutní hodnoty omezeny přibližně $2M\sqrt{N}$. V porovnání s tím je tedy nyní

máme asi $2\sqrt{2}$-krát menší. Ani to nám však, co se časové složitosti týče, nepomůže. Jak jsme viděli, ta záleží na mocnině N, kterou dokážeme omezit prosívané hodnoty. MPQS sice významně snižuje konstantu před ní, mocnina samotná ale zůstává nezměněná, konkrétně $1/2$. Ve výpočtu proto pořád pracujeme s parametrem $u = \frac{\ln N}{2 \ln B}$. Dospějeme tudíž k témuž výsledku.

Na druhou stranu nesmíme zapomenout na přidanou cenu za inicializaci jednotlivých polynomů. Jak to bude fungovat? Zvolíme prvočíslo $p \approx (2N)^{1/4} / M^{1/2}$ takové, že $\left(\frac{N}{p}\right) = 1$, $p \equiv 3 \pmod 4$ a položíme $a = p^2$. Vidíme, že a je čtverec krát B-hladká hodnota velikosti přibližně $\sqrt{2N}/M$, jak jsme požadovali. Za těchto podmínek určíme řešení kongruence $b^2 \equiv N \pmod a$ snadno předpisem

$$b = N^{\left(p^2-p+2\right)/4}.$$

Nakonec už pouze jednoduše spočteme $c = \left(b^2 - N\right)/a$.

To, kolik polynomů budeme potřebovat, záleží na volbě M. Čím větší bude, tím víc ušetříme na jejich inicializaci. Naopak, menší M znamená nižší hodnoty vstupující do prosívací fáze. Můžeme si každopádně pomoct paralelizací. Prosívat musíme pro každé f zvlášť, tuto práci tedy lze provádět simultánně na více počítačích. Tady je už potom nutné zohlednit praktické výpočetní možnosti a implementační problémy, které se teoreticky obtížně kvantifikují (třeba vzhledem k složitosti inicializace jednotlivých polynomů). V praxi se nicméně ukazuje, že MPQS navzdory stejné asymptotické složitosti často oproti základnímu nastavení přináší významnou časovou úsporu.

1.3.3 Vlastní verze MPQS

Výše uvedená verze reprezentuje samozřejmě jenom jeden z mnoha způsobů, jak volit polynomy pro MPQS. Nyní si představíme vlastní návrh, který pořád snižuje absolutní hodnoty vstupující do prosívací fáze, ovšem bez nutnosti jakýchkoli výpočtů souvisejících s inicializací jednotlivých polynomů. Tím se zbavíme největší nevýhody Montgomeryho verze. Na oplátku se objeví některé jiné potíže, kterými jsme se dosud nemuseli zabývat.

Budeme pracovat s množinou polynomů

$$\{f_i(x) = x^2 - i^2 N : i \in \{1, \ldots, k\}\}.$$

Pro každý z nich potřebujeme nalézt kořeny modulo p. To však jde ze znalosti kořenů f_1 modulo p velmi snadno, neboť

$$x^2 \equiv N \pmod p \Rightarrow (ix)^2 \equiv i^2 N \pmod p.$$

Dále vidíme, že základní verze kvadratického síta je speciálním případem, kdy $k = 1$. V ní, jak už víme, prosíváme interval

$$\left[\lceil\sqrt{N}\rceil - M, \lceil\sqrt{N}\rceil + M\right] \cap \mathbb{Z}$$

s tím, že absolutní hodnoty jsou omezeny přibližně $2M\sqrt{N}$. Všimněme si, že stejné meze dosáhneme, pokud bychom kterýkoli polynom f_i prosívali na intervalu

$$\left[\lceil i\sqrt{N}\rceil - \frac{M}{i}, \lceil i\sqrt{N}\rceil + \frac{M}{i}\right] \cap \mathbb{Z}.$$

Samozřejmě, délka tohoto intervalu je pro $i > 1$ menší. Nic nám ale nebrání vzít si takových polynomů víc a celkovou délku navýšit podle potřeby. Volbou $k > 1$ tedy dostáváme nové příležitosti najít hladké relace se stejnou absolutní hodnotou. Parametr M je ale nastaven tak, abychom jich dostali přesně tolik, kolik potřebujeme. Hledat další relace by proto bylo zbytečné.

Nabízí se tedy snížit M, čím by se zkrátil původní prosívací interval ($i = 1$). Tento hendikep vykompenzujeme přidáním dalších polynomů tak, aby celková délka intervalů zůstala přibližně stejná. Nižší M ale znamená nižší absolutní hodnoty, které budeme prosívat a to v praxi urychluje výpočet (přestože to nutně neznamená zlepšení asymptotické složitosti, viz sekce 1.3.2).

Nejvýraznějším problémem navrhované verze je opakování nalezených hladkých relací. Není vyloučeno, že to nastane i u Montgomeryho verze, zde jsou ale polynomy vzájemně „provázanější“. Ukažme si to na konkrétním příkladě. Předpokládejme, že $t = x^2 - N$ je hladké a vezměme $i \leq k$. Podívejme se na $t' = i^2 (x^2 - N) = (ix)^2 - i^2 N$. Tato hodnota je zřejmě hladká právě tehdy, když je hladké i. Ovšem i pokud je t' hladké, dává nám modulo 2 stejný vektor mocnin jako t.

Buď $M' = \frac{M}{c}$. Označme I_i prosívací interval příslušný polynomu f_i. Pro jednoduchost nyní předpokládejme, že je místo $\lceil i\sqrt{N} \rceil$ vycentrovaný kolem $i\sqrt{N}$, neboli

$$I_i = \left[i\sqrt{N} - \frac{M'}{i}, i\sqrt{N} + \frac{M'}{i} \right] \cap \mathbb{Z}.$$

Mějme tedy $x \in I_1$ splňující $x^2 - N$ je hladké. Zajímá nás, pro které i platí $ix \in I_i$. Dostáváme

$$ix \in I_i \Leftrightarrow \left| ix - i\sqrt{N} \right| = i \left| x - \sqrt{N} \right| \leq \frac{M'}{i} \Leftrightarrow \left| x - \sqrt{N} \right| \leq \frac{M'}{i^2}.$$

Vidíme, že čím blíž je x k $\sqrt{N}$, tím častěji narazíme na tutéž hladkou relaci. Dobrou zprávou je, že maximální vzdálenost klesá kvadraticky vzhledem k i, takže shod bude rychle ubývat. Na druhou stranu, v blízkosti středu intervalu jsou absolutní hodnoty nejnižší, tudíž výskyt hladkých relací nejvyšší. Tuto situaci lze samozřejmě zobecnit na $x \in I_j$, $j \in \{1, \ldots, k\}$, $x^2 - j^2 N$ hladké. Potom

$$ix \in I_{ij} \Leftrightarrow \left| x - j\sqrt{N} \right| < \frac{1}{i^2} \cdot \frac{M'}{j}.$$

Zásadní otázkou nyní je, zda se relace nebudou opakovat až natolik, že bychom jich nakonec získali ještě méně, než v klasickém kvadratickém sítu. Nastavme si tedy parametry tak, aby měly prosívací intervaly přibližně stejnou celkovou délku a zkoumejme počet získaných hladkých relací.

Zvolme si nějakou konstantu $c > 1$ a řekněme, že chceme, aby byly maximální absolutní hodnoty c-krát menší než původně, čili $\frac{2M\sqrt{N}}{c}$. Z toho plyne, že počáteční prosívací interval (pro $i = 1$) bude tvaru

$$\left[\lceil \sqrt{N} \rceil - \frac{M}{c}, \lceil \sqrt{N} \rceil + \frac{M}{c} \right] \cap \mathbb{Z}.$$

Nyní chceme vzít k co největší takové, že celková délka intervalů nepřesáhne $2M$ (jinak by případný vyšší počet nalezených hladkých relací v této verzi mohl být

způsoben vyšším počtem zkoumaných hodnot). Počítejme:

$$\sum_{i=1}^{k} \frac{2M}{ic} \leq 2M,$$
$$\sum_{i=1}^{k} \frac{1}{i} \leq c,$$
$$k \leq \exp(c-1),$$

kde v posledním kroku používáme odhad $\sum_{i=1}^{k} 1/i \leq 1 + \ln k$.

Musíme si ještě uvědomit jednu důležitou věc. Pokud vezmeme moc velké c, pak budou pro $i \approx k$ příslušné intervaly příliš krátké (v extrémním případě jednobodové). Například existuje-li přirozené číslo $n_0 < k$ takové, že I_{n_0} je jednobodový, potom pro všechna $n_0 \leq i \leq k$ pořád pracujeme pouze s výrazem $\left(\lceil i\sqrt{N} \rceil\right)^2 - i^2 N$. Na první pohled by se mohlo zdát, že je to pro nás výhodné, neboť tyto hodnoty jsou poměrně malé a tudíž s velkou pravděpodobností hladké. Má to ale jeden háček.

Do celkové délky intervalů vždy započítáváme výraz $\frac{2M}{ic}$ (přesněji by to mělo být $\frac{2M+1}{ic}$, jednička v čitateli ovšem hraje zanedbatelnou roli). Pro jednobodové intervaly je jeho hodnota menší než 1, přičemž často je hodně blízko nule, jelikož klesá s rostoucím i. Ve skutečnosti tedy součet délek intervalů může být výrazně větší než ten, který očekáváme dle výpočtů. To má za následek dvě věci:

1) Nalezneme-li více hladkých relací oproti základní verzi algoritmu, může k tomu přispívat právě větší počet zkoumaných hodnot.

2) Zvyšuje se složitost prosívací fáze: jednak víme (viz str. 8), že dohromady provádíme přibližně $2 \cdot$ (délka intervalu) $\cdot \ln \ln B$ dělení – jestliže zvýšíme celkovou délku intervalů, zvýšíme také počet provedených dělení. Navíc, s každým dalším intervalem musíme nanovo počítat množiny $J_{c,p}$ (def. viz str. 4), díky čemu také drobně naroste časová náročnost.

Protože naším hlavním zájmem je ukázat, že navrhovaná modifikace je oproti klasickému kvadratickému sítu výhodnější, tento problém vyřešíme tím, že intervaly délky 1 raději vůbec nebudeme zpracovávat (nebudeme v nich hledat hladké relace). Případně si můžeme stanovit jinou minimální délku intervalu (označme ji d) a všechny kratší zanedbat.

Naším cílem tedy bude odhadnout optimální $c > 1$ vzhledem k počtu hladkých relací na intervalech

$$\left[\lceil i\sqrt{N} \rceil - \frac{M}{ic}, \lceil i\sqrt{N} \rceil + \frac{M}{ic}\right] \cap \mathbb{Z}, \quad i \in \{1, \ldots, k\}$$

délky větší než 1. Výsledek porovnáme s případem $c = 1$, který reprezentuje standardní verzi kvadratického síta.

Čím větší c zvolíme, tím nižší absolutní hodnoty budeme prosívat. Zároveň se ale budou jednotlivé intervaly zkracovat, což od jistého momentu způsobí, že jejich délka klesne pod d a přestanou nám přibývat hladké relace. Dá se tedy očekávat, že s rostoucím c se bude zpočátku jejich celkový počet zvyšovat, až do chvíle, kdy začnou být intervaly I_i pro i blízko k příliš krátké. Dál už bude počet relací

nejspíš klesat. Tuto hypotézu si ověříme prostřednictvím testů na dostupných hodnotách, viz tabulky níže.

Otázkou zůstává, jak odhadnout optimální c, aniž bychom pro každou hodnotu museli relace počítat předem.

Jako nejjednodušší způsob se nabízí stanovit si zmíněnou minimální délku intervalu d a následně vzít největší c, pro které budou všechny intervaly dlouhé alespoň d. Hledáme tedy maximální c splňující

$$|I_{\exp(c-1)}| \geq d,$$

neboli

$$c \exp(c-1) \leq \frac{2M}{d-1}.$$

Už jsme si rozmysleli, že potřebujeme zvolit $d > 1$. Zatím tedy vezmeme $d = 3$ (intervaly mají vždy lichou délku). Na základě výsledků testů pak zanalyzujeme, zda není lepší vzít nějakou větší hodnotu.

Druhou možností je odhadnout počet relací pravděpodobnostně. Označme

$$\begin{aligned} u_c &= v_c^{-v_c}, \text{ kde} \\ v_c &= \frac{\ln x_c}{\ln B}, \text{ kde} \\ x_c &= \frac{2M\sqrt{N}}{c} \end{aligned}$$

Jinými slovy, u_c označuje pravděpodobnost nalezení hladké hodnoty v intervalu

$$\left[\lceil i\sqrt{N}\rceil - \frac{M}{ic}, \lceil i\sqrt{N}\rceil + \frac{M}{ic}\right] \cap \mathbb{Z}$$

pro libovolné $i \in \mathbb{N}$.

V klasické verzi s intervalem $\left[\lceil\sqrt{N}\rceil - M, \lceil\sqrt{N}\rceil + M\right] \cap \mathbb{Z}$ a jedním polynomem tedy očekáváme $2Mu_1$ hladkých relací. Ve vlastní verzi pracujeme s intervaly

$$I_i = \left[\lceil i\sqrt{N}\rceil - \frac{M}{ic}, \lceil i\sqrt{N}\rceil + \frac{M}{ic}\right] \cap \mathbb{Z}$$

pro nějaké předem zvolené c. Kdybychom nebrali v potaz opakování relací, celkem bychom jich měli

$$\sum_{i=1}^{\exp(c-1)} \frac{2M}{ic} u_c = \frac{2Mu_c}{c} \sum_{i=1}^{\exp(c-1)} \frac{1}{i} \approx \frac{2Mu_c}{c} \cdot (c-1) \approx 2Mu_c.$$

Opakování je bohužel příliš časté na to, abychom ho mohli zanedbat. Pokusme se ho tedy kvantifikovat.

Definice. Nechť $i, j \in \mathbb{N}$, $i \leq k$, $j \mid i$.

- R_i buď množina všech hladkých relací $(x, x^2 - i^2N)$ z intervalu I_i,
- U_i buď množina „unikátních“ hladkých relací z intervalu I_i, neboli takových $h_i = (x, x^2 - i^2N)$, že neexistuje $l \mid i$, $l \neq i$, pro které by $\left(\frac{l}{i}x, \frac{l^2}{i^2}x^2 - l^2N\right)$ byla hladká relace z I_l,

- $O_{i,j}$ buď množina unikátních hladkých relací z I_j „opakujících“ se v I_i, čímž se myslí následující. Nechť je $h_j = (x, x^2 - j^2N)$ hladká relace z I_j. Řekneme, že $h_j \in O_{i,j}$ právě tehdy, když

  I. $\left(\frac{i}{j}x, \frac{i^2}{j^2}x^2 - i^2N\right) = \left(\frac{i}{j}x, \frac{i^2}{j^2}\left(x^2 - j^2N\right)\right)$ je hladká relace z I_i,
  II. $h_j \in U_j$.

Nás budou zajímat především velikost množin U_i. Z uvedených definic plyne, že pro každé přirozené číslo $i \leq k$ platí

$$|U_i| = |R_i| - \sum_{\substack{j|i \\ j \neq i}} |O_{i,j}|,$$

přičemž, jak už víme,

$$|R_i| = \frac{2Mu_c}{ic}.$$

Zbývá tedy určit velikosti jednotlivých $O_{i,j}$. Podmínku I. z jejich definice lze přepsat na dvojici ekvivalentních:

- $\frac{i}{j}x \in I_i$,
- $\frac{i}{j}$ je hladké.

Druhou z nich ovšem můžeme opomenout. Už totiž víme, že $\frac{i}{j}x \in I_i$ se dá ekvivalentně vyjádřit podmínkou

$$\frac{i}{j}x \in I_i = I_{\frac{i}{j}j} \Leftrightarrow \left|x - j\sqrt{N}\right| < \frac{1}{\left(\frac{i}{j}\right)^2} \cdot \frac{M}{jc} = \frac{M}{\frac{i^2}{j}c}.$$

Pro pořádek připomínáme, že toto jsme odvodili za předpokladu, že středy jednotlivých prosívacích intervalů jsou $i\sqrt{N}$ místo $\lceil i\sqrt{N} \rceil$, což je ale zanedbatelný rozdíl.

„Opakující“ se relace se tedy nachází v prostřední (i^2/j^2)-tině intervalu I_j. Jenomže pokud $\frac{i}{j}$ není hladké, pak je zřejmě větší než B, tudíž $\frac{i^2}{j^2} > B^2$. Pro $c \geq 2$ (což je minimální hodnota, se kterou pracujeme v naší verzi) je ovšem maximální velikost prosívacího intervalu $M + 1 \approx B^2$. Prostřední (i^2/j^2)-tina I_j je proto prakticky vždy jednobodový interval ($\lceil j\sqrt{N} \rceil$). Tuto situaci tedy můžeme klidně zanedbat a podmínku I. z definice $O_{i,j}$ nahradit podmínkou $\frac{i}{j}x \in I_i$. Tím si výrazně zjednodušíme výpočet.

Z toho tedy plyne, že hladkých relací z I_j „opakujících“ se v I_i bude přibližně

$$\frac{2Mu_{\frac{i^2}{j^2}c}}{\frac{i^2}{j}c}.$$

Mohlo by se zdát, že tento výraz je roven $|O_{i,j}|$. Zatím jsme ovšem nezohlednili druhou podmínku, a sice že tyto relace zároveň leží v U_j. Pokud bychom ji ignorovali a od celkového počtu relací odečítali tento podíl pro každého dělitele $j \mid i$, nejspíš bychom některé relace odečítali vícekrát. Snadno se totiž může stát, že se relace vedoucí k témuž vektoru mocnin vyskytne v intervalech I_l, I_j a I_i, kde $l \mid j \mid i$, $l \neq j \neq i$. Abychom zjistili, kolik jich bude, zobecníme si stávající značení.

Definice. Nechť $i \neq j \in \mathbb{N}$, $j \mid i$, $t \in \mathbb{N}$. Buď $h_j = (x, x^2 - j^2N)$ hladká relace z intervalu I_j. Řekneme, že $h_j \in O_{i,j,t}$ právě tehdy, když

- $\frac{i}{j}x \in \left[\lceil i\sqrt{N}\rceil - \frac{M}{tic}, \lceil i\sqrt{N}\rceil + \frac{M}{tic}\right]$,
- $h_j \in U_j$.

Jinými slovy, $O_{i,j,t}$ je množina unikátních hladkých relací $h_j = (x, x^2 - j^2N)$ z I_j takových, že $\left(\frac{i}{j}x, \frac{i^2}{j^2}x^2 - i^2N\right)$ je hladká relace z I_i, která se nachází ve vzdálenosti nejvýše $\frac{M}{tic}$ od jeho středu.

Jak tedy bude vypadat obecný vzorec pro $|O_{i,j,t}|$? Představme si nejdřív, že chceme určit $|O_{i,j,1}|$. To je ekvivalentní výpočtu $|O_{i,j}|$. Už tedy víme, že pokud nebereme v potaz druhou podmínku z definice, vyjde nám

$$\frac{2Mu_{\frac{i^2}{j^2}c}}{\frac{i^2}{j}c}.$$

Od tohoto výrazu potřebujeme pro každé přirozené $l \mid j$, $l \neq j$ odečíst počet relací $h_l = (x, x^2 - l^2N)$, z intervalu I_l takových, že $h_j = (\frac{j}{l}x, \frac{j^2}{l^2}x^2 - j^2N)$ je hladká relace z I_j nacházející se v jeho prostřední $\frac{i^2}{j^2}$-tině, tj. v intervalu

$$\left[\lceil j\sqrt{N}\rceil - \frac{M}{\frac{i^2}{j^2}jc}, \lceil j\sqrt{N}\rceil + \frac{M}{\frac{i^2}{j^2}jc}\right] \cap \mathbb{Z}.$$

Navíc požadujeme $h_l \in U_l$, abychom některé relace neodečítali opakovaně. Vidíme tedy, že pro všechna l (splňující uvedené podmínky) odečítáme přesně ty relace, které náleží do $O_{j,l,\frac{i^2}{j^2}}$. Tudíž

$$|O_{i,j,1}| = \frac{2Mu_{\frac{i^2}{j^2}c}}{\frac{i^2}{j}c} - \sum_{\substack{l \mid j \\ l \neq j}} \left|O_{j,l,\frac{i^2}{j^2}}\right|.$$

Teď už není těžké si rozmyslet, že

$$|O_{i,j,t}| = \frac{2Mu_{\frac{i^2}{j^2}t^2c}}{\frac{i^2}{j}t^2c} - \sum_{\substack{l \mid j \\ l \neq j}} \left|O_{j,l,\frac{i^2}{j^2}t^2}\right|.$$

Vraťme se nyní k výpočtu unikátních relací. V novém značení počítáme

$$|U_i| = |R_i| - \sum_{\substack{j \mid i \\ j \neq i}} |O_{i,j,1}|.$$

Dosazením za $|R_i|$ a rekurzivním výpočtem $|O_{i,j,1}|$ jsme takto schopni určit očekávaný počet nalezených hladkých relací pro každé i a c.

Bohužel, složitost vzorce prakticky znemožňuje nalezení optimálního c (vzhledem k počtu hladkých relací), případně porovnání s klasickou verzí kvadratického síta ($c = 1$, kde očekáváme $2Mu_1$ hladkých relací) jinak než dosazením konkrétních hodnot N a c. Počet intervalů navíc roste exponenciálně v c, takže rekurzivní

výpočet relací pro všechny dělitele všech čísel až do $\exp(c-1)$ by brzy začal být příliš náročný.

Nabízí se proto různá zjednodušení, které nám pořád můžou napovědět, jak volit parametr c. My si vybereme dvě - jedno poslouží jako „optimistický“ a jedno jako „pesimistický“ odhad počtu hladkých relací. Na základě testů na dostupných hodnotách N následně ověříme, zda tyto odhady dávají dobrou představu o reálném počtu relací pro různá c.

- pesimistický odhad (E_{pes}): budeme počítat pouze relace z intervalů I_1 a I_p, kde p je prvočíslo. Výhodou je, že vždycky odečítáme opakující se hladké relace jedině vzhledem k intervalu I_1, nehrozí proto žádná rekurze. Odhadovaný počet relací tedy vyjde jako

$$\begin{aligned}\tilde{R} &= \frac{2Mu_c}{c} + \sum_{\substack{p \text{ prvočíslo} \\ p<\exp(c-1)}} \left(\frac{2Mu_c}{pc} - \frac{2Mu_{p^2c}}{p^2c}\right) \\ &= \frac{2M}{c}\left(u_c + \sum_{\substack{p \text{ prvočíslo} \\ p<\exp(c-1)}} \frac{pu_c - u_{p^2c}}{p^2}\right).\end{aligned}$$

  Motivace tohoto odhadu je následující: čím víc dělitelů nějaké číslo j má, tím nižší šance na nalezení unikátní hladké relace v intervalu I_j máme. Je totiž hodně intervalů s menším indexem, v nichž se již stejné relace mohly objevit. Intuitivně tedy lze předpokládat, že do počtu unikátních relací přispějí nejvíc právě intervaly s prvočíselným indexem.

- optimistický odhad (E_{opt}): pro všechny intervaly budeme uvažovat opakování relací pouze vzhledem k intervalu I_1. Hodně hladkých relací tedy započítáme vícekrát. Na druhou stranu, výpočet bude opět poměrně jednoduchý. Podobně jako v předchozím odhadu dostaneme

$$\begin{aligned}\tilde{R} &= \frac{2Mu_c}{c} + \sum_{i=2}^{\exp(c-1)} \left(\frac{2Mu_c}{ic} - \frac{2Mu_{i^2c}}{i^2c}\right) \\ &= \frac{2M}{c}\left(u_c + \sum_{i=2}^{\exp(c-1)} \frac{iu_c - u_{i^2c}}{i^2}\right).\end{aligned}$$

Vybrali jsme 6 různých hodnot, na kterých jsme otestovali naše hypotézy. Zkoumali jsme čísla tvaru $10^e + 1$, kde e je celé číslo z intervalu $[15, 20]$. Všechny jsou dělitelné alespoň dvěma různými prvočísly, takže splňují podmínky použití kvadratického síta. Parametry M a B byly voleny dle výpočtů v sekci 1.2, tj.

$$\begin{aligned}M &= \exp\left(\sqrt{\ln N \ln\ln N}\right), \\ B &= \exp\left(\frac{\sqrt{\ln N \ln\ln N}}{2}\right).\end{aligned}$$

V polovině případů ($e = 16, 17, 18$) bylo B kvůli příliš malému počtu nalezených hladkých relací vzato dvakrát větší. „Malý počet“ v tomto případě myslíme zejména vzhledem k tomu, aby bylo možné sledovat rozdíly mezi jednotlivými

hodnotami c. Najdeme-li totiž u všech příliš málo hladkých relací, obtížněji nahlédneme vztahy mezi jejich počtem a hodnotou c.

Nutně tedy parametr B nenastavujeme tak, aby výsledný počet hladkých relací korespondoval s π_B (třeba u $e = 20$ v nejlepším případě dostaneme 50 relací, přičemž pro dané N je $\pi_B = 135$, u $e = 16$ máme relací naopak zbytečně moc). To samozřejmě neznamená, že B volíme úplně libovolně, primárním cílem tohoto experimentu je však odhadnout závislost počtu unikátních hladkých relací na c.

Ve všech případech uvažujeme běžně používanou variantu s velkým prvočíslem. Kromě B-hladkých hodnot tedy bereme v potaz i ty, které jsou navíc dělitelné jedním prvočíslem v intervalu (B, B^2). Princip této varianty byl podrobně popsán v sekci 1.3.1.

Uveďme nyní jednotlivé výsledky. Pro přehlednost zvýrazňujeme zelenou barvou maxima ve vybraných důležitých sloupcích.

$N = 10^{15} + 1$					
c	$\sum \lvert U_i \rvert$	$\sum \lvert R_i \rvert$	$\lvert I_{\exp(c-1)} \rvert$	E_{opt}	E_{pes}
1	48	48	127243	21	21
2	32	49	23405	15	15
3	36	82	5741	14	13
4	38	102	1583	14	12
5	44	140	467	15	11
6	54	171	143	18	11
7	55	181	45	21	11
8	57	191	15	24	11
9	60	202	5	28	11
10	63	216	1	31	11
11	61	203	1	29	10
12	59	192	1	27	10
13	55	175	1	26	9
14	52	164	1	25	9
15	50	160	1	23	9

$N = 10^{16} + 1$					
c	$\sum \lvert U_i \rvert$	$\sum \lvert R_i \rvert$	$I_{\exp(c-1)}$	E_{opt}	E_{pes}
1	137	137	202845	125	125
2	106	152	37311	88	88
3	119	268	9151	89	81
4	128	377	2525	101	76
5	132	523	743	116	73
6	144	788	227	134	71
7	148	983	71	153	70
8	155	1400	23	172	70
9	165	2241	7	191	69
10	165	3883	3	208	68
11	161	4142	1	200	65
12	149	3892	1	188	61
13	139	3681	1	177	58
14	129	3500	1	167	55
15	123	3330	1	159	53

$N = 10^{17} + 1$					
c	$\sum \lvert U_i \rvert$	$\sum \lvert R_i \rvert$	$I_{\exp(c-1)}$	E_{opt}	E_{pes}
1	125	125	319465	161	161
2	100	154	58763	113	113
3	122	287	14411	116	105
4	124	401	3977	131	98
5	131	473	1171	151	94
6	140	542	359	174	91
7	152	631	113	198	90
8	170	683	37	222	89
9	179	753	11	244	88
10	183	791	3	266	87
11	180	784	1	271	84
12	173	749	1	254	79
13	164	708	1	240	75
14	154	664	1	227	72
15	153	634	1	215	68

$N = 10^{18} + 1$					
c	$\sum \lvert U_i \rvert$	$\sum \lvert R_i \rvert$	$I_{\exp(c-1)}$	E_{opt}	E_{pes}
1	132	132	497557	206	206
2	106	155	91521	145	145
3	119	286	22445	149	135
4	122	383	6193	169	125
5	138	559	1823	195	120
6	148	871	559	224	117
7	153	1162	177	254	115
8	165	1635	57	284	114
9	176	2611	19	312	112
10	189	4673	7	338	110
11	199	8814	3	363	108
12	187	8446	1	341	102
13	171	7962	1	321	97
14	165	7555	1	304	92
15	161	7184	1	288	88

$N = 10^{19} + 1$					
c	$\sum \lvert U_i \rvert$	$\sum \lvert R_i \rvert$	$I_{\exp(c-1)}$	E_{opt}	E_{pes}
1	29	29	766997	63	63
2	25	43	141081	45	45
3	31	85	34601	44	40
4	36	115	9547	48	37
5	38	111	2809	55	36
6	43	124	861	63	35
7	45	134	271	73	35
8	47	146	87	82	34
9	46	145	29	92	34
10	42	147	9	102	34
11	46	164	3	111	34
12	47	141	1	111	32
13	45	147	1	105	31
14	43	150	1	99	29
15	42	144	1	95	28

$N = 10^{20} + 1$					
c	$\sum \lvert U_i \rvert$	$\sum \lvert R_i \rvert$	$\lvert I_{\exp(c-1)} \rvert$	E_{opt}	E_{pes}
1	29	29	1171127	82	82
2	23	38	215417	58	58
3	24	83	52831	57	52
4	28	152	14577	63	49
5	31	261	4289	72	47
6	32	535	1315	84	46
7	38	954	415	96	45
8	41	1374	133	108	45
9	41	2272	43	120	44
10	47	4156	15	132	44
11	50	7911	5	143	43
12	50	13577	1	150	42
13	49	12859	1	142	40
14	47	12233	1	135	39
15	47	11674	1	128	37

Poznámka (Analýza výsledků). Z nasbíraných dat lze vyvodit několik zajímavých závěrů:

(i) Námi navrhovaná metoda konzistentně přináší (pro vhodně zvolené c) lepší výsledky než standardní kvadratické síto. V některých případech (třeba $e = 20$) je zlepšení dokonce až na úrovni 70%. Tento rozdíl je například pro $e = 17$ zásadní pro úspěšnost celého algoritmu. Potřebujeme totiž nasbírat $\pi_B + 1$ hladkých relací, což je v dané situaci 140. Vidíme, že pro $c = 1$ jich nemáme dostatek. To ještě neznamená, že metoda přesto nemůže fungovat, jistotu však máme pouze u $c \in [6, 15]$.

(ii) Maximum unikátních hladkých relací se často nabývá pro největší (nebo skoro největší) c takové, že délka nejkratšího prosívacího intervalu je větší než jedna, případně nejmenší c takové, že délka nejkratšího intervalu je 1. To potvrzuje náš předpoklad, že optimální c lze dobře odhadnout pomocí délky $I_{\exp(c-1)}$. Říkali jsme, že hledáme maximální c splňující

$$\left| I_{\exp(c-1)} \right| \geq d.$$

Jako pravděpodobně nejvhodnější se ukazuje nastavit c tak, aby nejkratší interval byl délkou co nejblíže jedné, ale přímo se jí nerovnal, nebo tak, aby bylo nejmenší takové, že $\left| I_{\exp(c-1)} \right| = 1$. Můžeme proto ponechat $d = 3$. Z výsledků též vidíme, že ve všech případech bychom při takto určeném c dostali více relací než v klasické verzi. Navíc to většinou ($e = 16, 17, 18, 20$) vychází dokonce jako nejlepší volba.

(iii) Podobně kvalitní nápovědu k volbě c dává i „optimistický“ odhad E_{opt}. Všimněme si, že pokud bychom určovali optimální c podle jeho maxima, došli bychom prakticky ke stejným hodnotám jako u odhadu založeném na délce nejkratšího intervalu (rozdíl by byl v absolutní hodnotě nejvýše jedna). Oproti němu má ale jednu výhodu – dokáže nám naznačit, jaké zlepšení v porovnání se základní verzi můžeme přibližně očekávat. Ilustrujme to na příkladě $e = 20$.

Oba zmíněné odhady by nás vedly k optimální volbě c (11 nebo 12). E_{opt} pro $c = 12$ předpovídá 150 relací. Skutečně jich získáme pouze 50, přímé porovnání proto rozhodně nepomůže. Dejme tedy tyto hodnoty do poměru s příslušnými hodnotami pro základní verzi:

$$\frac{(E_{opt})_{c=12}}{(E_{opt})_{c=1}} = \frac{150}{82} \approx 1{,}83$$
$$\frac{(\sum |U_i|)_{c=12}}{(\sum |U_i|)_{c=1}} = \frac{50}{29} \approx 1{,}72.$$

Vidíme, že tyto poměry jsou u sebe mnohem blíž. U zbylých zkoumaných N mohou být rozdíly o něco větší, dává nám to nicméně alespoň hrubý odhad poměrného zlepšení vůči klasickém kvadratickému sítu (pro dané c).

(iv) „Pesimistický“ odhad E_{pes} se bohužel ukazuje jako prakticky obtížně použitelný. Nekopíruje totiž trend počtu unikátních relací, chová se jako nerostoucí funkce. To je pravděpodobně způsobeno klesající hustotou prvočísel, zanedbáváme tudíž pořád víc a víc intervalů. Něco by se dalo vyčíst z rychlosti poklesu – je patrné, že v okolí optimálního c zpravidla E_{pes} klesá nejpomaleji. Eventuálně bychom mohli zkusit tuto funkci aproximovat nějakou křivkou (nabízí se třeba polynom stupně 3) a hledat bod s maximální derivací. Takový odhad by nám ovšem nejspíš neposkytoval dostatečnou rozlišovací schopnost. Z experimentálních dat je totiž zřejmé, že tato derivace by byla dost podobná (hodně blízká nule) v okolí poměrně širokého rozsahu hodnot c.

(v) Dle očekávání s rostoucím c výrazně roste rozdíl mezi $\sum |R_i|$ a $\sum |U_i|$, až do momentu, kdy začneme prosívat intervaly délky 1. Maximum všech nalezených relací se tedy nabývá pro stejné nebo velmi podobné c jako u unikátních relací. Toto je spíše zajímavostí nežli užitečným faktem, který by nám v praxi jakkoli pomohl.

Na základě těchto poznatků by postup pro libovolné N (klidně řádově větší) mohl vypadat následovně:

1) Odhadneme optimální $\tilde{c}$ pomocí délky nejkratšího prosívacího intervalu, tj. tak, aby pro $d = 3$ platilo

$$\tilde{c} \exp(\tilde{c} - 1) \approx \frac{2M}{d-1}.$$

2) Spočteme $(E_{opt})_{\tilde{c}}$. Můžeme případně tento odhad dopočítat také pro c' v okolí $\tilde{c}$, čili například $c' \in [\tilde{c} - 2, \tilde{c} + 2]$. Pokud bychom viděli, že E_{opt} nabývá na tomto intervalu maxima v jednom z krajních bodů, nabízelo by se počítat odhad pro další hodnoty c' v příslušném směru, až dokud nezačne očekávaný počet relací klesat. Následně vezmeme $c'' = \operatorname{argmax}(E_{opt})_{c'}$. Výsledné odhadované optimální c můžeme určit nějakou kombinací $\tilde{c}$ a c'', třeba jejich průměrem. Nakonec určíme E_{opt} pro klasickou verzi ($c = 1$) a spočteme poměr odhadů nalezených relací

$$q = \frac{(E_{opt})_c}{(E_{opt})_1},$$

abychom se ujistili, že použitím vlastní verze MPQS lze očekávat zlepšení (a jak velké).

3) Jestliže $q > 1$, zvolíme vlastní verzi kvadratického síta s odhadovaným optimálním c. Pokud bychom dostali $q \leq 1$ (což se jeví jako krajně nepravděpodobné), zůstaneme u klasické verze.

Zkusme se podívat, jak by tento postup fungoval na uvedených datech.

- $e = 16$, $e = 18$: oba odhady by se shodovaly s optimální hodnotou c,
- $e = 20$: odhady vedou k různé volbě c, které je ale pokaždé nejlepší možné,
- $e = 15$, $e = 17$, $e = 19$: k optimálnímu c vede vždy pouze jeden z odhadů, ten druhý se však liší jenom o 1 a znamenal by zanedbatelně nižší počet nalezených hladkých relací, který by nicméně pořád významně překonal základní verzi.

Výsledky jsou tedy velmi příznivé. Přesto bohužel nemůžeme s jistotou říci, že uvedená metoda bude fungovat obecně. Vidíme však, že ve všech zkoumaných případech jsme dokázali pomocí výpočetně jednoduchých odhadů najít c, které povedlo ke zvýšení počtu nalezených hladkých relací. Často jsme dokonce zvolili nejlepší možnou hodnotu.

Problémem může být chybějící metoda pro efektivní detekování opakujících se relací. Nejvíc by pomohla u těch N, jejichž odmocnina je hodně blízko celého čísla, jako třeba $10^e + 1$ pro e sudé. Poměr počtu všech a unikátních relací je pak největší (což souvisí s tím, že jednotlivé prosívací intervaly mají největší průnik), jak je ostatně patrné z tabulek.

Na závěr ještě zmiňme, že pro hodnoty, které nás zajímají v praxi, tj. řádově $N = 10^{100}$, vychází podle naší metody jako optimální $c = 32$. Ověřit, zda bychom při těchto parametrech opravdu dostali víc relací (a o kolik), je bohužel daleko mimo naše výpočetní možnosti.

2. Číselné síto

2.1 Obecný princip

Náš úkol zůstává nezměněný. Na vstupu máme liché přirozené číslo N dělitelné alespoň dvěma různými prvočísly a snažíme se ho faktorizovat tak, že najdeme kongruentní čtverce modulo N. Změní se ale postup, kterým je budeme hledat. Zatímco v kvadratickém sítu jsme na jedné straně čtverec již měli a pomocí hladkých relací nad celými čísly se totéž snažili dosáhnout na druhé straně, nyní budeme pracovat nad složitějšími okruhy a čtverce hledat na obou stranách současně.

Začneme volbou monického polynomu $f \in \mathbb{Z}[x]$ a $0 \neq m \in \mathbb{Z}$ tak, aby platilo $f(m) \equiv 0 \pmod{N}$. Můžeme si například zvolit $d = \deg f$, položit $m = \lfloor N^{1/d} \rfloor$ a určit koeficienty f podle zápisu N v bázi m. Potřebujeme ověřit, že je f opravdu monický.

Jestliže $d \geq 2$, $N > 2^{d^2}$ (jak se ukáže později, pro doporučované hodnoty d budou tyto podmínky bezpečně splněny), pak pro každé $i \in \{0, \ldots, d\}$ máme

$$\binom{d}{i} \leq 2^d - 2 \leq N^{\frac{1}{d}} - 2 \leq m - 1,$$

kde první nerovnost lze jednoduše dokázat indukcí pro $d \geq 2$ pomocí faktu, že kombinační číslo nabývá největší hodnoty, když $i = \lfloor d/2 \rfloor$. Dále platí

$$(m+1)^d = \sum_{i=0}^{d} \binom{d}{i} m^i.$$

Jelikož $\binom{d}{i} < m$, jedná se zároveň o zápis $(m+1)^d$ v bázi m, přičemž koeficient u m^d je roven 1. Protože $m^d \leq N < (m+1)^d$, má N u m^d (ve svém zápisu v bázi m) také 1. Tudíž je f monický.

Dále můžeme předpokládat, že f je ireducibilní. V opačném případě jsme totiž z rozkladu $f(x) = g(x)\,h(x)$ schopni dostat netriviální faktorizaci N. Detaily lze najít v Brillhart a kol. (1981).

Nechť $\alpha \in \mathbb{C}$ je kořen f. Nemusíme jej přesně počítat, stačí nám pracovat s ním coby symbolem určujícím číselné těleso $K = \mathbb{Q}[\alpha] \simeq \mathbb{Q}[x]/(f)$. Zajímat nás bude především jeho podokruh

$$\mathbb{Z}[\alpha] = \{\sum_{i=0}^{d-1} z_i \alpha^i : z_0, \ldots, z_{d-1} \in \mathbb{Z}\}.$$

Definujme homomorfismus $\varphi_0 : \mathbb{Z}[x] \to \mathbb{Z}_N$ předpisem

$$\varphi_0(h) := h(m) \mod N.$$

Jedná se o dosazovací homomorfismus s jádrem obsahujícím ideál (f). Indukuje proto homomorfismus $\overline{\varphi_0} : \mathbb{Z}[x]/(f) \to \mathbb{Z}_N$ daný vztahem

$$\overline{\varphi_0}(h + (f)) := h(m) \mod N.$$

Analogicky, dosazovací homomorfismus $\varphi_1 : \mathbb{Z}[x] \to \mathbb{Z}[\alpha]$, $h \mapsto h(\alpha)$ má jádro (f) a obraz $\mathbb{Z}[\alpha]$. Dle první věty o isomorfismu je tedy $\mathbb{Z}[\alpha] \simeq \mathbb{Z}[x]/(f)$. Konkrétně dostáváme isomorfismus $\overline{\varphi_1} : \mathbb{Z}[\alpha] \to \mathbb{Z}[x]/(f)$ určen předpisem

$$\overline{\varphi_1}\left(\sum_{i=0}^{d-1} z_i \alpha^i\right) := \sum_{i=0}^{d-1} z_i x^i + (f).$$

Složením $\overline{\varphi_1}$ a $\overline{\varphi_0}$ získáme klíčový homomorfismus $\varphi : \mathbb{Z}[\alpha] \to \mathbb{Z}_N$ daný vztahem

$$\varphi\left(\sum_{i=0}^{d-1} z_i \alpha^i\right) := \sum_{i=0}^{d-1} z_i m^i \bmod N.$$

Algoritmus se bude snažit hledat dvojice prvků $(a_1, b_1), \ldots, (a_k, b_k) \in \mathbb{Z} \times \mathbb{Z}$ takové, že

- $\forall i \in \{1, \ldots, k\} : \mathrm{NSD}(a_i, b_i) = 1$,
- $\prod_{i=1}^{k} (a_i - b_i\alpha) = \gamma^2$ pro nějaké $\gamma \in \mathbb{Z}[\alpha]$,
- $\prod_{i=1}^{k} (a_i - b_i m) = z^2$ pro nějaké $z \in \mathbb{Z}$.

Aplikováním homomorfismu φ na rovnost v druhé podmínce dostaneme

$$\begin{aligned}
\varphi\left(\prod_{i=1}^{k} (a_i - b_i\alpha)\right) &= \varphi\left(\gamma^2\right), \\
\prod_{i=1}^{k} \varphi(a_i - b_i\alpha) &= \varphi(\gamma)^2, \\
\prod_{i=1}^{k} (a_i - b_i m) &\equiv \varphi(\gamma)^2 \pmod N,
\end{aligned}$$

což po zohlednění třetí podmínky dává kýženou kongruenci

$$z^2 \equiv \varphi(\gamma)^2 \pmod N,$$

díky níž můžeme zkusit faktorizovat N prostřednictvím $\mathrm{NSD}(N, z \pm \varphi(\gamma))$.

V dalším průběhu budeme potřebovat následující pojmy.

Definice. Buď $f \in \mathbb{Z}[x]$ ireducibilní polynom stupně d. Jsou-li $\alpha_1, \ldots, \alpha_d$ jeho kořeny v $\mathbb{C}$, $\alpha = \alpha_1$, pak *normou* prvku $\beta = \sum_{i=0}^{d} c_i \alpha^i \in \mathbb{Z}[\alpha]$ rozumíme

$$N(\beta) = \prod_{j=1}^{d} \sum_{i=0}^{d} c_i \alpha_j^i.$$

Definice. Nechť $B \in \mathbb{N}$ je hladká mez. Řekneme, že $\beta \in \mathbb{Z}[\alpha]$ je $B_{\mathbb{Z}[\alpha]}$-*hladké*, jestliže jeho norma $N(\beta) \in \mathbb{Z}$ je $B_{\mathbb{Z}}$-hladká.

Poznámka. Nechť $f(x) = l_c(f)(x - \alpha_1)\cdots(x - \alpha_d)$, kde $l_c(f)$ je vedoucí koeficient f. Potom

$$\begin{aligned}
N(a - b\alpha) &= (a - b\alpha_1)\cdots(a - b\alpha_d) \\
&= b^d \left(\frac{a}{b} - \alpha_1\right)\cdots\left(\frac{a}{b} - \alpha_d\right) \\
&= l_c(f)^{-1} b^d f\left(\frac{a}{b}\right).
\end{aligned}$$

V našem případě je f monický a tudíž $N(a - b\alpha) \in \mathbb{Z}$.

Vraťme se k hledání dvojic (a_i, b_i). Postupovat budeme podobně jako v kvadratickém sítu. Tehdy jsme si zafixovali parametr M a následně pro všechna $x \in \{\lceil\sqrt{N}\rceil - M, \ldots, \lceil\sqrt{N}\rceil + M\}$ zkoumali, zda je $x^2 - N$ B-hladké pro zvolenou mez B. Nyní ovšem máme dvě proměnné, potřebujeme tudíž dvourozměrný prosívací interval. Určíme si proto $M \in \mathbb{N}$ a budeme procházet dvojice (a, b), kde $a \in \{-M, \ldots, M\}$, $b \in \{1, \ldots, M\}$, splňující

- $\mathrm{NSD}(a, b) = 1$,
- $a - bm$ je $B_{\mathbb{Z}}$-hladké,
- $a - b\alpha$ je $B_{\mathbb{Z}[\alpha]}$-hladké.

První podmínka nepotřebuje žádný komentář. S tou druhou jsme se již setkali u kvadratického síta, takže víme, jak postupovat. Prosíváme prvočísly p z faktorizační báze, vytvoříme vektory mocnin modulo 2 a zapíšeme je do řádků matice lineární fáze. Pokud jich nasbíráme alespoň $\pi_B + 2$, dokážeme nalézt prvky, jejichž součin bude čtverec. Teď ovšem navíc potřebujeme zohlednit třetí podmínku.

Místo $a - b\alpha$ budeme vlastně prosívat jejich normy, čímž problém převedeme do celých čísel, kde to už umíme. Takže pro každou dvojici (a, b) nyní potřebujeme dvojnásobně dlouhý vektor mocnin — $\pi_B + 1$ souřadnic příslušných $a - bm$ plus $\pi_B + 1$ souřadnic příslušných $N(a - b\alpha)$. V lineární fázi tedy hledáme řešení soustavy s maticí, která má $2\pi_B + 2$ sloupců. Potřebujeme tudíž nasbírat alespoň $2\pi_B + 3$ hladkých relací.

Důležitým důsledkem pro fungování číselného síta je multiplikativita normy. Snadno lze nahlédnout, že pro libovolné $\beta, \beta' \in \mathbb{Z}[\alpha]$ platí $N(\beta\beta') = N(\beta)\, N(\beta')$. Jestliže je tedy $\beta = \gamma^2$ pro nějaké $\gamma \in \mathbb{Z}[\alpha]$, pak $N(\beta) = N(\gamma)^2$ je čtverec v $\mathbb{Z}$. Z toho potom plyne, že pokud má být $\prod (a_i - b_i\alpha)$ čtverec v $\mathbb{Z}[\alpha]$, což je to, co nás primárně zajímá, musí nutně zároveň $\prod N(a_i - b_i\alpha)$ být čtvercem v $\mathbb{Z}$.

Obrácená implikace bohužel rozhodně neplatí. Vezměme si třeba okruh Gaussovských celých čísel $\mathbb{Z}[i]$. Je-li $p \equiv 1 \pmod 4$ prvočíslo, pak jej lze napsat ve tvaru $a^2 + b^2$ a následně v $\mathbb{Z}[i]$ rozložit na $(a + bi)(a - bi)$. To je součin dvou různých (neasociovaných) prvočinitelů, proto p není čtverec v $\mathbb{Z}[i]$, přestože $N(p) = p^2$.

Problémem tady je, že norma tyto dva prvočinitele „ztotožňuje“. Rádi bychom je tedy nějak rozlišili. V tomto konkrétním případě se nabízí jednoduché řešení. Místo normy bychom rozkládali přímo prvky $a - bi$, souřadnice ve vektorech mocnin by pak korespondovaly s jednotlivými prvočiniteli normy $\leq B$. Podobně je to možné udělat kdykoli je $\mathbb{Z}[\alpha]$ gaussovský obor. K tomu je potřeba znát popis prvočinitelů a fundamentálních jednotek (analogie -1 v $\mathbb{Z}$) tohoto okruhu. My si ale ukážeme postup, který nebude závislý na speciálních vlastnostech $\mathbb{Z}[\alpha]$.

Pro každé prvočíslo p označme

$$R(p) := \{r \in \{0, \ldots, p-1\} : f(r) \equiv 0 \pmod p\}.$$

Víme, že

$$N(a - b\alpha) = b^d f\left(\frac{a}{b}\right).$$

Z toho plyne, že pokud $\mathrm{NSD}(a, b) = 1$, $b \not\equiv 0 \pmod p$, potom

$$N(a - b\alpha) \equiv 0 \pmod p \Leftrightarrow a \equiv br \pmod p \text{ pro nějaké } r \in R(p).$$

Prosívání tedy bude probíhat následovně: pro všechna $1 \leq b \leq M$ vytvoříme pole hodnot $N(a - b\alpha)$, $-M \leq a \leq M$. Pak pro každé prvočíslo $p \leq B$, které nedělí b, a každé $r \in R(p)$ nalezneme $a \equiv br \pmod{p}$ a hodnoty na těchto pozicích v poli vydělíme nejvyšší možnou mocninou p. Na konci procesu jsou $B_{\mathbb{Z}[\alpha]}$-hladké právě ty $a - b\alpha$, které na příslušných pozicích mají ± 1.

Poznámka. V nejhorším případě může mít pro nějaké p polynom f až d kořenů mod p (neboli $|R(p)| = d$). Průměrně nicméně očekáváme $|R(p)| = 1$. Intuitivně předpokládáme, že funkční hodnoty f jsou rovnoměrně distribuované. Pak je pravděpodobnost, že p dělí $f(x)$, rovna $1/p$. Rigorózní náhled nabízí Pesiri (2007), Theorem 3.1.5.

V lineární fázi, stejně jako doposud, vyřešíme homogenní soustavu nad $\mathbb{F}_2$ s maticí, jejíž řádky obsahují vektory mocnin hladkých hodnot $N(a - b\alpha)$. Tím ovšem získáme pouze množinu dvojic $(a, b) \in \mathbb{Z} \times \mathbb{Z}$, označme ji S, pro níž je $N\left(\prod_{(a,b)\in S}(a - b\alpha)\right)$ čtverec. To je sice nutná podmínka toho, aby byl čtvercem (v $\mathbb{Z}[\alpha]$) samotný součin $\prod_{(a,b)\in S}(a - b\alpha)$, nikoli však postačující. Naštěstí se ukazuje, že tuto potíž lze téměř zcela obejít tím, že si společně s prvočísly p, které dělí $N(a - b\alpha)$, budeme uchovávat také příslušné $r \in R(p)$.

Jsou-li a, b nesoudělná celá čísla, p prvočíslo a $r \in R(p)$, definujeme

$$e_{p,r}(a - b\alpha) = \begin{cases} v_p(N(a - b\alpha)) & \text{pokud } a \equiv br \pmod{p}, \\ 0 & \text{jinak.} \end{cases}$$

Jejich význam si předvedeme na situaci $\mathbb{Z}[\alpha] = \mathcal{O}_K$, kde $\mathcal{O}_K$ značí okruh celistvých prvků $K = \mathbb{Q}[\alpha]$.

Připomeňme si nejdřív některé jeho základní vlastnosti. Je známo, že se jedná o Dedekindův obor. Z toho mimo jiné plyne, že

- každý nenulový prvoideál $\mathcal{O}_K$ je maximální,
- každý nenulový ideál $\mathcal{O}_K$ lze jednoznačně rozložit na součin prvoideálů.

Dále, je-li $P \subset \mathcal{O}_K$ prvoideál, pak obsahuje právě jedno prvočíslo $p \in \mathbb{Z}$ a $\mathcal{O}_K/P$ je těleso. Stupněm P rozumíme $k \in \mathbb{Z}$ takové, že $|\mathcal{O}_K/P| = p^k$. Pokud je prvoideál P stupně 1, zřejmě $\mathcal{O}_K/P \simeq \mathbb{Z}_p$ a máme homomorfismus $\varphi : \mathcal{O}_K \to \mathbb{Z}_p$ s jádrem P určen pomocí $\varphi(\alpha) = r \bmod p$. Naopak, pro prvočíslo p a $r \in R(p)$ existuje jediný homomorfismus $\mathcal{O}_K \to \mathbb{Z}_p$, který posílá α na $r \bmod p$. Z první věty o isomorfismu a vlastností $\mathcal{O}_K$ plyne, že jeho jádrem je prvoideál P stupně 1.

Máme tedy bijekci mezi $\{(p, r) : p \text{ prvočíslo}, r \in R(p)\}$ a prvoideály $P \subset \mathcal{O}_K$ stupně 1 (které jsou generovány p a $\alpha - r$). Číslo $e_{p,r}(a - b\alpha)$ tedy budeme interpretovat jako mocninu příslušného prvoideálu $P = (p, \alpha - r) \subset \mathcal{O}_K$ v rozkladu hlavního ideálu $(a - b\alpha)$. Pozice ve vektorech mocnin příslušné normám budou tudíž nově indexovány dvojicemi p, r (dle předchozí poznámky víme, že délka vektorů přesto zůstane přibližně stejná). Podrobný rozbor nejen této situace, ale také obecného případu, kdy nemáme zajištěno $\mathbb{Z}[\alpha] = \mathcal{O}_K$, lze najít v Buhler a kol. (1993).

Dosavadní popis algoritmu rozhodně není úplný, na některých místech činíme předpoklady, které splněné být nemusí (třeba $\mathbb{Z}[\alpha] = \mathcal{O}_K$). Cílem je vynechat technicky nejnáročnější případy, bez nichž se obejdeme při analýze časové složitosti a návrhu různých modifikací, které jsou hlavním objektem našeho zájmu. Všechny detaily se nachází v odkazované literatuře.

2.2 Složitost

Na začátek si zaveďme značení

$$L_n[\alpha,\beta] := \exp\left((\beta + o(1))(\ln n)^{\alpha}(\ln\ln n)^{1-\alpha}\right), \quad 0 \le \alpha \le 1.$$

Připomeňme, že $f(n) \in o(g(n))$, pokud $\lim_{n\to\infty}\frac{f(n)}{g(n)} = 0$. Tedy $o(1)$ vyjadřuje funkci, která pro $n \to \infty$ jde k nule.

Kromě toho budeme opět používat značení již známé z kapitoly o kvadratickém sítu, konkrétně $\psi(x,y) := |\{n \in \mathbb{Z} : |n| \le x \ \&\ n \text{ je } y\text{-hladké}\}|$.

Víme, že potřebujeme najít $2\pi_B+3$ hladkých relací $(a,b) \in \mathbb{Z}\times\mathbb{Z}$ takových, aby $a - bm$ i $N(a - b\alpha)$ byly B-hladké, přičemž $a \in \{-M,\dots,M\}$, $b \in \{1,\dots,M\}$, $\mathrm{NSD}(a,b) = 1$. Budeme tedy zkoumat, zda je hladký součin

$$G(a,b) = (a-bm)N(a-b\alpha) = (a-bm)b^d f\left(\frac{a}{b}\right).$$

Označme $f(x) = \sum_{i=0}^{d} c_i x^i$, kde $c_d = 1$. Potom

$$b^d f\left(\frac{a}{b}\right) = a^d + c_{d-1}a^{d-1}b + \cdots + c_0 b^d.$$

Z toho můžeme udělat horní odhad $|G(a,b)|$. Vzpomeňme si, že polynom f vznikl zápisem N v bázi $m = \lfloor N^{1/d}\rfloor$. Takže $c_i \le N^{1/d}$ pro každé $i \in \{0,\dots,d-1\}$. Stejné omezení platí samozřejmě taky pro m. Dále víme, že $|a|, |b| \le M$. Celkem tedy dostáváme

$$\begin{aligned}|G(a,b)| &\le \left(M + MN^{1/d}\right)(d+1)M^d N^{1/d}\\ &\le 2MN^{1/d}(d+1)M^d N^{1/d}\\ &\le 2(d+1)M^{d+1}N^{2/d}.\end{aligned}$$

Položme tedy $X = 2(d+1)M^{d+1}N^{2/d}$. Našim cílem je zvolit hladkou mez B tak, aby $\frac{\psi(X,B)}{2X}$, čili pravděpodobnost nalezení hladké relace (a,b), byla co největší. Následující tvrzení převzaté z Buhler a kol. (1993), Theorem 10.1, nám ukáže, jak takové B vypadá.

Věta 2.1. *Nechť g je funkce definovaná pro všechna $y \ge 2$ splňující $g(y) \ge 1$ a $g(y) = y^{1+o(1)}$ pro $y \to \infty$. Potom pro $x \to \infty$*

$$\frac{2xg(y)}{\psi(x,y)} \ge L_x\left[\frac{1}{2},\sqrt{2}\right]$$

kdykoli $y \ge 2$, přičemž rovnost nastává právě tehdy, když $y = L_x\left[\frac{1}{2},\frac{\sqrt{2}}{2}\right]$.

Důkaz. Postupně odhadneme velikost výrazu $\frac{2xg(y)}{\psi(x,y)}$ v závislosti na třech možných hodnotách y. Několikrát při tom použijeme Větu 1.2, připomeňme proto její znění. Jestliže $u = \frac{\ln x}{\ln y}$, pak $\psi(x,y) \approx 2xu^{-u}$. Jinými slovy, pravděpodobnost nalezení y-hladké hodnoty v intervalu $[-x,x]$ je přibližně u^{-u}.

- $y \ge L_x\left[\frac{1}{2},2\right]$: jelikož $\frac{2x}{\psi(x,y)} \ge 1$, tak $\frac{2xg(y)}{\psi(x,y)} \ge g(y) \ge L_x\left[\frac{1}{2},2\right]$.

- $y \leq L_x\left[\frac{1}{2}, \frac{1}{4}\right]$: začneme tím, že $\frac{2xg(y)}{\psi(x,y)} \geq \frac{2x}{\psi(x,y)} \geq \frac{2x}{\psi\left(x, L_x\left[\frac{1}{2}, \frac{1}{4}\right]\right)}$. Tento výraz bychom rádi vyjádřili pomocí Věty 1.2. Máme

$$u = \frac{\ln x}{\ln\left(L_x\left[\frac{1}{2}, \frac{1}{4}\right]\right)} = 4\left(\ln x\right)^{1/2}\left(\ln\ln x\right)^{-1/2}$$

a zajímá nás hodnota u^u (pro přehlednost vynecháváme členy $o\left(1\right)$, které asymptoticky stejně nic nezmění). Dostáváme

$$\begin{aligned} u^u &= \exp\left(u \ln u\right) \\ &= \exp\left(4\left(\ln x\right)^{1/2}\left(\ln\ln x\right)^{-1/2}\left(\ln 4 + \frac{1}{2}\ln\ln x - \frac{1}{2}\ln\ln\ln x\right)\right) \\ &= \exp\left(2\left(\ln x\right)^{1/2}\left(\ln\ln x\right)^{1/2} + h\left(x\right)\right), \end{aligned}$$

kde

$$\begin{aligned} h\left(x\right) &= 4\left(\ln x\right)^{1/2}\left(\ln\ln x\right)^{-1/2}\left(\ln 4 - \frac{1}{2}\ln\ln\ln x\right) \\ &= \left(\ln x\right)^{1/2}\left(\ln\ln x\right)^{1/2}\frac{4\left(\ln 4 - \frac{1}{2}\ln\ln\ln x\right)}{\ln\ln x} \\ &= \left(\ln x\right)^{1/2}\left(\ln\ln x\right)^{1/2} o\left(1\right), \end{aligned}$$

takže

$$u^u = \exp\left(\left(2 + o\left(1\right)\right)\left(\ln x\right)^{1/2}\left(\ln\ln x\right)^{1/2}\right) = L_x\left[\frac{1}{2}, 2\right].$$

Opět jsme dospěli ke stejnému výsledku, a sice

$$\frac{2xg\left(y\right)}{\psi\left(x, y\right)} \geq L_x\left[\frac{1}{2}, 2\right].$$

- $L_x\left[\frac{1}{2}, \frac{1}{4}\right] \leq y \leq L_x\left[\frac{1}{2}, 2\right]$: označme $y = L_x\left[\frac{1}{2}, \delta\right]$. Podobným výpočtem jako výše získáme $\frac{2x}{\psi(x,y)} = u^u = L_x\left[\frac{1}{2}, \frac{1}{2\delta}\right]$ pro $u = \frac{1}{\delta}\left(\ln x\right)^{1/2}\left(\ln\ln x\right)^{-1/2}$, čili

$$\frac{2xg\left(y\right)}{\psi\left(x, y\right)} = L_x\left[\frac{1}{2}, \delta + \frac{1}{2\delta}\right].$$

Protože $\left(\delta + \frac{1}{2\delta}\right)' = 1 - \frac{1}{2\delta^2} = 0 \Leftrightarrow \delta = \pm\frac{\sqrt{2}}{2}$ a z předpokladu $\frac{1}{4} \leq \delta \leq 2$, tak $\delta + \frac{1}{2\delta}$ nabývá minima v bodě $\frac{\sqrt{2}}{2}$ s hodnotou $\sqrt{2}$. Tudíž

$$\frac{2xg\left(y\right)}{\psi\left(x, y\right)} \geq L_x\left[\frac{1}{2}, \sqrt{2}\right].$$

Tím je tvrzení dokázáno. □

Tuto větu aplikujeme na $x = X = 2\left(d+1\right)M^{d+1}N^{2/d}$, $y = B$ a funkci g určenou předpisem $g\left(B\right) = 2\pi\left(B\right) + 3$. Z toho plyne, že optimální hodnotou B pro nalezení potřebného počtu hladkých relací je $L_X\left[\frac{1}{2}, \frac{\sqrt{2}}{2}\right]$.

Poznámka. Máme-li $a, b \in \{1, \ldots, M\}$, pak každé prvočíslo $p \leq B$ dělí přibližně M^2/p^2 dvojic (a, b). Proto můžeme počet dvojic, které nebudou mít žádného netriviálního společného dělitele, odhadnout zhruba jako

$$M^2 \left(1 - \sum_{p \leq B} \frac{1}{p^2}\right) \approx M^2 \left(1 - 0{,}45\right) = 0{,}55M^2.$$

Aproximaci $\sum 1/p^2$ jsme už diskutovali v první kapitole, viz str. 7-8. V naší situaci máme $a \in \{-M, \ldots, M\}$, očekáváme tedy dvakrát víc nesoudělných párů.

Dále nám tvrzení poskytuje odpověď na otázku, jak volit parametr M. Ten by měl splňovat rovnost $M^2 = L_X\left[\frac{1}{2}, \sqrt{2}\right]$, neboť dle Poznámky procházíme $O\left(M^2\right)$ dvojic (a, b).

V prosívací fázi tedy potřebujeme vygenerovat $L_X\left[\frac{1}{2}, \sqrt{2}\right] = B^2$ relací. Samotné prosívání funguje obdobně jako u kvadratického síta, počet operací, které vyžaduje, je proto přibližně $\ln \ln B$ krát délka intervalu. Ta je nyní $2M + 1$, přičemž proces opakujeme M krát (pro $b = 1, 2, \ldots, M$). Složitost tudíž vychází jako $O\left(M^2 \ln \ln B\right) = L_X\left[\frac{1}{2}, \sqrt{2}\right]$. Zbývá tento výraz vyjádřit vzhledem k N.

Rozeberme si nejdříve jednodušší případ, kdy zafixujeme stupeň d (přičemž $N \to \infty$). Vyjdeme z právě zmíněného vztahu $M = L_X\left[\frac{1}{2}, \frac{\sqrt{2}}{2}\right]$. Dosazením do vzorce pro X a zlogaritmováním obou stran dostaneme

$$\ln X \approx \ln 2 + \ln\left(d + 1\right) + \left(d + 1\right) \frac{\sqrt{2}}{2} \left(\ln X\right)^{1/2} \left(\ln \ln X\right)^{1/2} + \frac{2}{d} \ln N.$$

První dva členy jsou konstantní, můžeme je tedy rovnou zanedbat. Ze zbylých dvou asymptoticky převažuje ten poslední, neboť $\ln X = O\left(\frac{2}{d} \ln N\right)$. Celý výraz se tím pádem zjednoduší na

$$\ln X \approx \frac{2}{d} \ln N.$$

Časová složitost pro pevné d tedy vychází jako

$$\begin{aligned}
L_X\left[\frac{1}{2}, \sqrt{2}\right] &= \exp\left(\left(\sqrt{2} + o\left(1\right)\right) \left(\ln X\right)^{1/2} \left(\ln \ln X\right)^{1/2}\right) \\
&= \exp\left(\left(\sqrt{2} + o\left(1\right)\right) \left(\frac{2}{d} \ln N\right)^{1/2} \left(\ln \frac{2}{d} + \ln \ln N\right)^{1/2}\right) \\
&= \exp\left(\left(\frac{2}{\sqrt{d}} + o\left(1\right)\right) \left(1 + \frac{\ln \frac{2}{d}}{\ln \ln N}\right)^{1/2} \left(\ln N\right)^{1/2} \left(\ln \ln N\right)^{1/2}\right) \\
&= L_N\left[\frac{1}{2}, \frac{2}{\sqrt{d}}\right].
\end{aligned}$$

Složitost kvadratického síta je v novém značení $L_N\left[\frac{1}{2}, 1\right]$. Aby tedy byl tento algoritmus z časového hlediska lepší, potřebujeme volit $d \geq 5$.

Podívejme se teď na případ, kdy nemáme zafixované d (tedy $d \to \infty$, $N \to \infty$). Vyjdeme opět ze vztahu $M = L_X\left[\frac{1}{2}, \frac{\sqrt{2}}{2}\right]$. Tentokrát po aplikovaní logaritmu dostaneme o něco komplikovanější aproximaci

$$\ln X \approx \frac{d\sqrt{2}}{2} \left(\ln X\right)^{1/2} \left(\ln \ln X\right)^{1/2} + \frac{2}{d} \ln N.$$

Naším cílem je zvolit d tak, aby X bylo co nejmenší. Jinými slovy, snažíme se minimalizovat X vzhledem k d. Zderivujeme proto obě strany podle d, čím dostaneme

$$\frac{X'}{X} \approx \frac{\sqrt{2}}{2} (\ln X)^{1/2} (\ln \ln X)^{1/2} + \frac{dX' (1 + \ln \ln X)}{2\sqrt{2} X (\ln X)^{1/2} (\ln \ln X)^{1/2}} - \frac{2}{d^2} \ln N$$

a následně položíme $X' = 0$:

$$\begin{aligned} 0 &\approx \frac{\sqrt{2}}{2} (\ln X)^{1/2} (\ln \ln X)^{1/2} - \frac{2}{d^2} \ln N, \\ d &\approx 2^{3/4} (\ln N)^{1/2} (\ln X)^{-1/4} (\ln \ln X)^{-1/4}. \end{aligned}$$

Dosazením tohoto výsledku do prvotní aproximace získáme

$$\begin{aligned} \ln X &\approx 2^{5/4} (\ln N)^{1/2} (\ln X)^{1/4} (\ln \ln X)^{1/4}, \\ (\ln X)^{3/4} &\approx 2^{5/4} (\ln N)^{1/2} (\ln \ln X)^{1/4}, \\ \frac{3}{4} \ln \ln X &\approx \frac{5}{4} \ln 2 + \frac{1}{2} \ln \ln N + \frac{1}{4} \ln \ln \ln X, \\ \ln \ln X &\approx \frac{2}{3} \ln \ln N. \end{aligned}$$

To dosadíme zpátky do druhého řádku:

$$\begin{aligned} (\ln X)^{3/4} &\approx 2^{3/2} \cdot 3^{-1/4} (\ln N)^{1/2} (\ln \ln N)^{1/4}, \\ \ln X &\approx 4 \cdot 3^{-1/3} (\ln N)^{2/3} (\ln \ln N)^{1/3}. \end{aligned}$$

Časová složitost tohoto algoritmu je tudíž

$$\begin{aligned} L_X \left[\frac{1}{2}, \sqrt{2}\right] &= \exp\left(\left(\sqrt{2} + o(1)\right) (\ln X)^{1/2} (\ln \ln X)^{1/2}\right) \\ &= \exp\left(\left(\sqrt{2} + o(1)\right) \left(\frac{4}{3^{1/3}} (\ln N)^{2/3} (\ln \ln N)^{1/3}\right)^{1/2} \left(\frac{2}{3} \ln \ln N\right)^{1/2}\right) \\ &= \exp\left(\left(\frac{4}{3^{2/3}} + o(1)\right) (\ln N)^{1/3} (\ln \ln N)^{2/3}\right) \\ &= L_N \left[\frac{1}{3}, \frac{4}{3^{2/3}}\right] \approx L_N \left[\frac{1}{3}, 1.923\right]. \end{aligned}$$

Analogickým výpočtem lze zjistit, že této hodnoty nabývá pro

$$d \approx \left(\frac{3 \ln N}{\ln \ln N}\right)^{1/3}.$$

Poznámka. Připomeňme, že toto je optimální stupeň za předpokladu $N \to \infty$. V praxi ovšem nepočítáme s nekonečnými N. Mohli bychom se tedy pokusit optimalizovat d pouze pro náhodně volená N z nějakého intervalu konečné délky. Příklad takového výpočtu se nachází v práci Pejlová (2016), sekce 4.1.2, kde autorka pro $N \in [10^{10}, 10^{400}]$ navrhuje brát

$$d \approx \left(\frac{2.56 \ln N}{\ln \ln N}\right)^{1/3}.$$

Nesmíme zapomenout ani na hledání množin $R(p)$, neboli kořenů $f \bmod p$. To však nebude žádný problém. I kdybychom je hledali hrubou silou, dokážeme to pro každé p udělat pomocí $O(pd)$ násobení a sčítání v $\mathbb{Z}_p$. Takže dohromady provedeme maximálně $O(pd\pi_B) = O(B^2) = L_X\left[\frac{1}{2}, \sqrt{2}\right]$ operací. Složitost prosívací fáze se tedy nijak nezmění.

To samé platí také o ostatních fázích číselného síta. Tu lineární, čili řešení soustavy lineárních rovnic s řídkou maticí, dokážeme zvládnout ve stejném čase $O(\pi_B^2) = O(B^2) = L_X\left[\frac{1}{2}, \sqrt{2}\right]$.

Nakonec potřebujeme odmocnit samotné řešení soustavy $\prod_{i=1}^{k}(a_i - b_i\alpha) = \gamma^2$ v $\mathbb{Z}[\alpha]$. Postup tady nebudeme podrobně rozebírat, zmíníme jenom, že i to lze provést v čase $O(B^2)$, nebo dokonce menším, použijeme-li techniky rychlého násobení (viz Buhler a kol. (1993)). Každopádně to opět rozhodně nebude pomalejší než prosívací fáze.

Celková složitost číselného síta je proto opravdu $L_N\left[\frac{1}{3}, \frac{4}{3^{2/3}}\right]$.

2.3 Modifikace

Začneme tím, že si zobecníme parametry používané v základní verzi. V prosívací fázi zkoumáme hladkost prvků $a - bm$ a $a - b\alpha$. Zatím to děláme vzhledem ke stejné hladké mezi B. Nic nám však nebrání zvolit si ji samostatně pro každé kritérium. Potom tedy budeme hledat dvojice (a, b) splňující

- $\mathrm{NSD}(a, b) = 1$,
- $a - bm$ je B_1-hladké,
- $N(a - b\alpha)$ je B_2-hladké,

kde B_1 a B_2 definujeme jako

$$B_1 = L_N\left[\frac{1}{3}, \beta\right],$$
$$B_2 = L_N\left[\frac{1}{3}, \gamma\right].$$

Parametry β a γ nám tedy umožňují ovlivňovat hodnoty těchto mezí. Všimněme si, že pokud $\beta = \gamma = \frac{2}{3^{2/3}}$, pak dostáváme klasické číselné síto. Víme totiž, že $B^2 = L_X\left[\frac{1}{2}, \sqrt{2}\right] = L_N\left[\frac{1}{3}, \frac{4}{3^{2/3}}\right]$.

Výpočet složitosti nám napovídá, jak volit optimální stupeň d. To ovšem neznamená, že v jiných verzích nemůže být ideální hodnota odlišná. Zobecníme ji proto do tvaru

$$d \approx \delta\left(\frac{\ln N}{\ln\ln N}\right)^{1/3}.$$

Opět snadno nahlédneme přechod k základní verzi, stačí vzít $\delta = 3^{1/3}$.

Konečně položme

$$M = L_N\left[\frac{1}{3}, \epsilon\right].$$

Doposud jsme pracovali s předpokladem $M = L_X\left[\frac{1}{2}, \frac{\sqrt{2}}{2}\right] = L_N\left[\frac{1}{3}, \frac{2}{3^{2/3}}\right]$, používali jsme tedy vlastně $\epsilon = \frac{2}{3^{2/3}}$. Jak záhy uvidíme, někdy je také výhodnější vzít trochu jinou hodnotu.

Analogicky jako v kvadratickém sítu, i v tom číselném můžeme použít variantu s velkým prvočíslem $B_1 < p < B_1^2$, případně velkým prvoideálem P, který by splňoval podmínku $B_2 < N(P) < B_2^2$. Z toho vyplývající změny algoritmu jsou velmi podobné těm, které jsme popsali v předchozí kapitole, nebudeme je proto specifikovat znova. Ukážeme si místo toho vylepšení, se kterým jsme se zatím nesetkali.

Než se k tomu dostaneme, udělejme si malou odbočku k faktorizaci metodou eliptických křivek (ECM). Ta je velmi vhodná pro hledání malých prvočíselných dělitelů, což se nám bude hodit při posuzování hladkosti. Zatím jsme si vystačili s prosíváním. Brzy si však ukážeme modifikace, kde tento postup nelze obecně aplikovat. Místo něj použijeme právě ECM. Zajímá nás tedy, jak to v našich konkrétních případech ovlivní časovou složitost. Budeme vycházet především z následujícího tvrzení.

Věta 2.2 (Lenstra Jr (1987), Theorem 1.1)**.** *Nechť $b, n \in \mathbb{N}$ a n není mocninou prvočísla. Jestliže má n vlastního dělitele menšího než b, pak existuje $c > 0$ takové, že ECM najde nějakého vlastního dělitele n v průměrném čase nejvýše*

$$L_b\left[\frac{2}{3}, c\right] \ln^2 n.$$

My tento výsledek aplikujeme na $b = B = L_N\left[\frac{1}{3}, \cdot\right]$, za n dosadíme normu prvku $a - b\alpha$, která bude v absolutní hodnotě omezená $L_N\left[\frac{2}{3}, \cdot\right]$.

Věta 2.3. *Nechť $x, y > 0$. Je-li $n = L_N\left[\frac{2}{3}, x\right]$, $b = L_N\left[\frac{1}{3}, y\right]$, ECM dokáže faktorizovat libovolné b-hladké číslo nepřesahující n v průměrném čase nejvýše*

$$L_b\left[\frac{2}{3}, c\right] \ln^3 n = b^{o(1)}$$

pro nějaké $c > 0$.

Poznámka. Připomeňme význam ω-notace. Podle definice

$$f(n) \in \omega(g(n)) \Leftrightarrow \lim_{n\to\infty} \frac{f(n)}{g(n)} = \infty.$$

Důkaz. Nejprve ukážeme, že $b = \ln^{\omega(1)} n$.

$$b = L_N\left[\frac{1}{3}, y\right] = \exp\left(y (\ln N)^{1/3} (\ln\ln N)^{2/3}\right),$$
$$\ln^{\omega(1)} n = \ln^{\omega(1)} L_N\left[\frac{2}{3}, x\right] = \exp\left(\omega(1)\left(\ln x + \frac{2}{3}\ln\ln N + \frac{1}{3}\ln\ln\ln N\right)\right).$$

Chceme tedy

$$y (\ln N)^{1/3} (\ln\ln N)^{2/3} = \omega(1)\left(\ln x + \frac{2}{3}\ln\ln N + \frac{1}{3}\ln\ln\ln N\right),$$

neboli

$$\lim_{N\to\infty} \frac{y (\ln N)^{1/3} (\ln\ln N)^{2/3}}{\ln x + \frac{2}{3}\ln\ln N + \frac{1}{3}\ln\ln\ln N} = \infty.$$

Snadno lze odvodit, že limita výrazu na levé straně je rovna

$$\lim_{N\to\infty} \frac{3y}{2} \cdot \left(\frac{\ln N}{\ln\ln N}\right)^{1/3} \cdot \left(1 + \frac{\frac{3}{2}\ln x + \frac{1}{2}\ln\ln\ln N}{\ln\ln N}\right)^{-1},$$

což je zřejmě nekonečno.

Víme tedy, že $b = \ln^{\omega(1)} n$. Faktorizace pomocí ECM spočívá v opakovaném hledání netriviálních dělitelů nějakého b-hladkého $m \leq n$. Proces z Věty 2.2 tedy opakujeme nejvýše $\log_2 n$ krát. Takže asymptoticky dokážeme faktorizovat v průměrném čase nejvýše $L_b\left[\frac{2}{3}, c\right] \ln^3 n$. Zbývá ukázat, že je to rovno $b^{o(1)}$. Počítejme:

$$\begin{aligned}
L_b\left[\frac{2}{3}, c\right] \ln^3 n &= \exp\left(c\,(\ln b)^{2/3}\,(\ln\ln b)^{1/3}\right) b^{3/\omega(1)} \\
&= \exp\left(c\,(\ln b)^{2/3}\,(\ln\ln b)^{1/3}\right) \exp\left(o\,(1)\ln b\right) \\
&= \exp\left(\ln b\left(o\,(1) + c\left(\frac{\ln\ln b}{\ln b}\right)^{1/3}\right)\right) \\
&= b^{o(1)+c\left(\frac{\ln\ln b}{\ln b}\right)^{1/3}} \\
&= b^{o(1)}.
\end{aligned}$$

□

Přesuňme se nyní k samotným modifikacím.

2.3.1 Kvadratické charaktery

V rozboru číselného síta jsme se vlastně dopracovali k tomu, jak najít dvojice $(a_1, b_1), \ldots, (a_k, b_k) \in \mathbb{Z} \times \mathbb{Z}$ takové, že

- $\prod_{i=1}^{k} (a_i - b_i m) = z^2$ pro nějaké $z \in \mathbb{Z}$,
- $\prod_{i=1}^{k} (a_i - b_i \alpha)\,\mathcal{O}_K = I^2$ pro nějaký ideál $I \subset \mathcal{O}_K$.

Kvadratické charaktery nabízí způsob, jak zvýšit pravděpodobnost, že bude platit to, co nás opravdu zajímá, čili $\prod_{i=1}^{k} (a_i - b_i \alpha) = \gamma^2$ pro nějaké $\gamma \in \mathbb{Z}[\alpha]$. Pojďme si je tedy definovat.

Definice. Nechť $0 \neq P \subset \mathcal{O}_K$ je prvoideál liché normy. *Kvadratickým charakterem* P rozumíme zobrazení $\left(\frac{\cdot}{P}\right) \to \{-1, 0, 1\}$ určené předpisem

- $\left(\frac{\beta}{P}\right) = 0$, pokud $\beta \in P$,
- $\left(\frac{\beta}{P}\right) = 1$, pokud $\beta \notin P$ a $\beta + P$ je čtvercem v $\mathcal{O}_K/P$,
- $\left(\frac{\beta}{P}\right) = -1$, pokud $\beta + P$ není čtvercem v $\mathcal{O}_K/P$.

Idea je jednoduchá. Čím více prvoideálů P takových, že $\left(\frac{\beta}{P}\right) = 1$ pro nějaké $\beta \in \mathcal{O}_K$ najdeme, tím pravděpodobnější je, že je čtvercem samotné β (v $\mathcal{O}_K$). Zvolíme si tedy m prvoideálů $P_1, \ldots, P_m$ s lichou prvočíselnou normou větší než

B. Do vektorů mocnin (respektive matice lineární fáze) přidáme sloupce indexovány $\left(\frac{\cdot}{P_i}\right)$, přičemž pro dvojici (a,b) na pozici příslušnou P_i zapíšeme 1, pokud $\left(\frac{a-b\alpha}{P_i}\right) = -1$, v opačném případě zapíšeme 0.

Otázkou zůstává, jak takto definovaný kvadratický charakter v praxi spočítat. Na konci sekce 2.1 jsme nastínili, že prvoideály $P < \mathcal{O}_K$ stupně 1 jsou v bijekci s množinou $\{(p,r) : p \text{ prvočíslo}, r \in R(p)\}$. Reálně tedy budeme počítat charaktery tvaru $\left(\frac{a-br}{p}\right)$. Následující tvrzení ukazuje, jak přesně musí vypadat.

Věta 2.4. *Nechť $f \in \mathbb{Z}[x]$ je monický ireducibilní polynom s kořenem $\alpha \in \mathbb{C}$. Dále mějme prvočíslo p a celé číslo r splňující $f(r) \equiv 0 \pmod{p}$, $f'(r) \not\equiv 0 \pmod{p}$. Je-li S množina nesoudělných dvojic $(a,b) \in \mathbb{Z} \times \mathbb{Z}$ takových, že p nedělí $a - br$ pro žádné $(a,b) \in S$ a $f'(\alpha)^2 \prod_{(a,b)\in S}(a - b\alpha)$ je čtverec v $\mathbb{Z}[\alpha]$, potom*

$$\prod_{(a,b)\in S} \left(\frac{a-br}{p}\right) = 1.$$

Poznámka. Na první pohled nemusí být jasné, zdali vůbec $f'(\alpha)\beta \in \mathbb{Z}[\alpha]$ pro nějaké $\beta \in \mathcal{O}_K$. Důkaz, že to platí pro libovolné $\beta \in \mathcal{O}_K$, poskytuje například Crandall a Pomerance (2001), Lemma 6.2.3.

Důkaz. Buď $\gamma \in \mathbb{Z}[\alpha]$ prvek, pro nějž $f'(\alpha)^2 \prod_{(a,b)\in S}(a - b\alpha) = \gamma^2$. Uvažujme homomorfismus $\varphi : \mathbb{Z}[\alpha] \to \mathbb{Z}_p$, který posílá α na r mod p. Podle předpokladu pak dostaneme

$$\varphi\left(\gamma^2\right) = f'(r)^2 \prod_{(a,b)\in S} (a - br) \not\equiv 0 \pmod{p},$$
$$\varphi(\gamma)^2 \left(f'(r)^{-1}\right)^2 \equiv \prod_{(a,b)\in S} (a - br) \not\equiv 0 \pmod{p},$$

tedy

$$\prod_{(a,b)\in S} \left(\frac{a-br}{p}\right) = \left(\frac{\prod_{(a,b)\in S}(a-br)}{p}\right) = 1.$$

□

Poznámka. Tvrzení poskytuje návod, jak upravit algoritmus v obecném případě, kdy nemáme zaručeno, že $\mathbb{Z}[\alpha] = \mathcal{O}_K$. My se zabýváme především situací, v níž je tato rovnost splněna, proto bychom mohli vypustit předpoklad $f'(r) \not\equiv 0 \pmod{p}$ a navíc by nám stačilo, že je čtvercem v $\mathbb{Z}[\alpha]$ prvek $\prod_{(a,b)\in S}(a - b\alpha)$.

Rádi bychom věděli, že platí i opačná implikace, tedy

$$\prod_{(a,b)\in S} \left(\frac{a-br}{p}\right) = 1 \Rightarrow \prod_{(a,b)\in S} (a - b\alpha) = \gamma^2 \text{ pro nějaké } \gamma \in \mathbb{Z}[\alpha].$$

To nedokážeme určit s jistotou, nicméně jak jsme si už řekli, bude-li levá strana splněna pro dostatečné množství dvojic (p,r), s velkou pravděpodobností je implikace pravdivá. V praxi se ukazuje, že „dostatečné množství“ znamená přibližně $3 \ln N$. Již dříve jsme diskutovali, že průměrně očekáváme $|R(p)| = 1$, jinými slovy, v průměru pro každé testované prvočíslo najdeme jednu vhodnou dvojici (p,r). Postačí nám tedy projít prvních přibližně $3 \ln N$ prvočísel větších než B.

Jak to ovlivní složitost? Nyní tedy potřebujeme o asi $O(\ln N)$ víc hladkých relací, což dohromady dává přibližně $O(\pi_B)+O(\ln N)$. To je asymptoticky rovno $O(B) = L_N\left[\frac{1}{3}, \frac{2}{3^{2/3}}\right]$, neboť $\ln N = L_N[0,1]$. Vycházíme opět z Věty 2.1. Oproti základní verzi číselného síta se změní pouze funkce $g(y)$, kde $y = B$. Jak jsme ale ukázali, pořád bude pro $y \to \infty$ rovna $y^{1+o(1)}$. To znamená, že složitost prosívací fáze zůstane beze změny. I ta lineární nadále poběží v čase nanejvýš B^2.

Teď ale navíc musíme pro každou relaci vypočítat asi $\ln N$ kvadratických charakterů, dohromady tedy $B \ln N = L_N\left[\frac{1}{3}, \frac{2}{3^{2/3}}\right]$. Samotný výpočet charakteru je pro nás jednotková operace (asymptoticky trvá stejně jako například NSD). Kromě toho počítáme množiny $R(p)$, což, jak je nám už známo, dokážeme v základní verzi provést v čase $O(B^2)$. Výpočet $R(p)$ pro jedno p trvá $O(p)$. Nyní těchto množin počítáme o $3 \ln N$ víc. Časová složitost se tedy navýší o $O(p \ln N)$, což určitě není víc než $O(B^2)$, neboť $\ln N = L_N[0,1]$ a $p \leq B^2 = L_N\left[\frac{1}{3}, \frac{4}{3^{2/3}}\right]$, protože $O(\pi_B) + O(\ln N) = L_N\left[\frac{1}{3}, \frac{2}{3^{2/3}}\right] < L_N\left[\frac{1}{3}, \frac{4}{3^{2/3}}\right] = O(\pi_{B^2})$.

Nic z toho nám tedy nezmění celkovou složitost, která znovu vychází jako

$$L_N\left[\frac{1}{3}, \frac{4}{3^{2/3}}\right] = O\left(B^2\right).$$

Kvadratické charaktery jsou díky tomu spíše běžnou součástí číselného síta. Složitost se jejich přidáním asymptoticky nezhorší, na druhou stranu nám významně pomáhají zaručit, že algoritmus proběhne podle očekávání.

2.3.2 Verze s více polynomy

Znovu se dostáváme k modifikaci, kterou jsme již viděli v kapitole o kvadratickém sítu. Nyní bude ovšem princip úplně odlišný. Zatímco u MPQS bylo hlavním cílem snížit absolutní hodnotu prvků vstupujících do prosívací fáze, teď budeme zejména chtít některé výpočty využít opakovaně (absolutní hodnota koeficientů polynomů se dokonce zvětší). Tato metoda byla poprvé popsána v článku Coppersmith (1993).

Na vstupu tedy máme přirozené číslo N, které chceme faktorizovat. Zafixujeme si stupeň polynomů d a celé číslo m podle vzoru základní verze, tj.

$$d \approx \delta \left(\frac{\ln N}{\ln \ln N}\right)^{1/3},$$
$$m \approx N^{1/d}.$$

Důležité je tady slovo „zafixujeme“, jelikož stejné hodnoty budeme používat pro všechny polynomy.

Hledání hladkých relací si nyní rozdělíme do dvou částí. Pomineme pro tuto chvíli podmínku hladkosti v $\mathbb{Z}[\alpha]$. Zvolíme si tedy parametry $M = L_N\left[\frac{1}{3}, \epsilon\right]$ a $B_1 = L_N\left[\frac{1}{3}, \beta\right]$ a nalezneme dvojice (a, b) v rozsahu $|a| < M$, $0 < b < M$, splňující

- $\mathrm{NSD}(a, b) = 1$,
- $a - bm$ je B_1-hladké.

Tento proces budeme nazývat „celočíselná fáze“.

Dále položme $B_2 = L_N\left[\frac{1}{3}, \gamma\right]$ a definujme nové parametry

$$\begin{aligned} B_1' &= \pi_{B_1} + 1, \\ B_2' &= \pi_{B_2} + O\left(\ln N\right). \end{aligned}$$

Všimněme si, že B_1' určuje počet sloupců matice lineární fáze příslušných -1 a prvočíslům $p < B_1$ (celočíselná fáze), zatímco B_2' určuje počet sloupců příslušných dvojicím (p, r), $p < B_2$ prvočíslo, r kořen f modulo p pro jeden polynom f (budeme jich používat víc) včetně kvadratických charakterů (polynomiální fáze).

V obou případech zároveň platí, že B_i a B_i' mají tentýž zápis v L-notaci. Předveďme si důkaz pro $i = 1$ (pro přehlednost vynecháváme členy $o\left(1\right)$):

$$\begin{aligned} B_1' &= \frac{L_N\left[\frac{1}{3}, \beta\right]}{\ln L_N\left[\frac{1}{3}, \beta\right]} \\ &= \frac{\exp\left(\beta\left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3}\right)}{\exp\left(\ln\left(\beta\left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3}\right)\right)} \\ &= \exp\left(\beta\left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3} - \ln\left(\beta\left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3}\right)\right) \\ &= \exp\left(\left(\beta + o\left(1\right)\right)\left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3}\right) \\ &= L_N\left[\frac{1}{3}, \beta\right], \end{aligned}$$

neboť

$$\begin{aligned} \ln\left(\beta\left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3}\right) &= \left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3} \cdot \frac{\ln\left(\beta\left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3}\right)}{\left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3}} \\ &= o\left(1\right)\left(\ln N\right)^{1/3}\left(\ln\ln N\right)^{2/3}. \end{aligned}$$

Obdobně lze ukázat také $B_2' = L\left[\frac{1}{3}, \gamma\right]$.

Tím se dostáváme k „polynomiální fázi“. V ní sestavíme množinu polynomů stupně d s kořenem m modulo N velikosti

$$\frac{B_1'}{B_2'} = L_N\left[\frac{1}{3}, \beta - \gamma\right].$$

To dokážeme udělat docela snadno. Opět si pomůžeme zápisem čísla N v bázi m. Položíme $f_0\left(x\right) = \sum_{i=0}^{d} c_i x^i$, kde $N = \sum_{i=0}^{d} c_i m^i$, $0 \leq c_i < m$. Následně vezmeme

$$f_i\left(x\right) = f_0\left(x\right) + i\left(x - m\right), \quad 0 \leq i \leq \frac{B_1'}{B_2'}.$$

Zde nastává již zmíněný nárůst absolutní hodnoty koeficientů. U těchto polynomů může dosahovat až mB_1'/B_2'. To budeme muset zohlednit během výpočtu složitosti.

Pro každou dvojici (a, b) z celočíselné fáze a každý polynom f_i vytvoříme trojici (a, b, f_i) a budeme hledat ty, pro které platí

$$N_{f_i}\left(a - b\alpha_{f_i}\right) = b^d f_i\left(\frac{a}{b}\right) \text{ je } B_2\text{-hladké},$$

kde $\alpha_{f_i} \in \mathbb{C}$ je kořen f_i. Cílem je nastavit parametry $\beta, \gamma, \delta, \epsilon$ tak, aby tuto podmínku pro každý polynom splňovalo přibližně $2B_2'$ trojic. Naše matice lineární fáze má totiž B_1' sloupců příslušných celočíselné fázi plus B_2' sloupců příslušných polynomiální fázi pro každé f_i. To dohromady dává $2B_1'$ sloupců. Chceme tedy také $2B_1' = (B_1'/B_2') \cdot 2B_2'$ řádků. Každý z nich bude patřit k právě jedné z těchto trojic (a, b, f_i).

V lineární fázi klasicky nalezneme netriviální řešení soustavy nad $\mathbb{F}_2$. Označme S množinu trojic příslušných řádkům (určených tímto řešením), jejichž lineární kombinace dává nulu modulo 2. Můžeme ji rozdělit podle jednotlivých polynomů na podmnožiny $S_{f_i} = \{(a, b, f) \in S : f = f_i\}$. V každé z nich platí

$$\prod_{S_{f_i}} (a - b\alpha_{f_i}) = \omega_{f_i}^2$$

pro nějaké $\omega_{f_i} \in \mathbb{Z}[\alpha_{f_i}]$.

Buď $x_{f_i} = \varphi_{f_i}(\omega_{f_i})$, kde $\varphi_{f_i} : \mathbb{Z}[\alpha_{f_i}] \to \mathbb{Z}_N$ je homomorfismus, který posílá α_{f_i} na m. Dále ať $x = \prod_{f_i} x_{f_i}$. Potom dva kongruentní čtverce modulo N dostaneme prostřednictvím

$$x^2 \equiv \prod_{f_i} x_{f_i}^2 \equiv \prod_{f_i} \prod_{S_{f_i}} (a - mb) \equiv \prod_{S} (a - mb) \equiv y^2 \pmod{N}$$

pro vhodné $y \in \mathbb{Z}$. (Detailní popis samotného odmocňování opět vynecháme.)

Podívejme se tedy na složitost takto nastaveného algoritmu. Nyní jsme prosívání rozdělili do dvou částí, nelze proto aplikovat Větu 2.1. Naštěstí si v tomto případě vystačíme pouze s Větou 1.2.

V celočíselné fázi nejdřív hledáme dvojice $(a, b) \in \mathbb{Z} \times \mathbb{Z}$ splňující $|a| < M$, $0 < b < M$, $\mathrm{NSD}(a, b) = 1$. Jejich počet odhadujeme jako M^2 (konstanty stejně asymptoticky roli nehrají). Zajímá nás, kolik hladkých relací na konci této fáze získáme, takže tuto hodnotu vynásobíme pravděpodobností, že $a - mb$ je hladké. Horní mez absolutní hodnoty tohoto čísla je přibližně $MN^{1/d}$, čili podle Věty 2.1 je tato pravděpodobnost rovna u^{-u} pro $u = \ln\left(MN^{1/d}\right) / \ln B_1$.

Nyní nás čeká několik technických výpočtů. Víme, že $M = L_N\left[\frac{1}{3}, \epsilon\right]$. Zápis druhého členu v L-notaci je

$$N^{1/d} = \exp\left(\frac{1}{\delta} (\ln N)^{2/3} (\ln \ln N)^{1/3}\right) = L_N\left[\frac{2}{3}, \frac{1}{\delta}\right],$$

tudíž $MN^{1/d} = L_N\left[\frac{2}{3}, \frac{1}{\delta}\right]$. Teď můžeme zjednodušit u:

$$u = \frac{\ln\left(MN^{1/d}\right)}{\ln B_1} = \frac{\frac{1}{\delta} (\ln N)^{2/3} (\ln \ln N)^{1/3}}{\beta (\ln N)^{1/3} (\ln \ln N)^{2/3}} = \frac{(\ln N)^{1/3}}{\delta\beta (\ln \ln N)^{1/3}}.$$

Potom

$$\begin{aligned} u^{-u} &= \exp\left(-\frac{(\ln N)^{1/3}}{\delta\beta (\ln \ln N)^{1/3}} \left(\frac{1}{3} \ln \ln N - \ln(\delta\beta) - \frac{1}{3} \ln \ln \ln N\right)\right) \\ &= \exp\left(\left(\frac{-1}{3\delta\beta} + o(1)\right) (\ln N)^{1/3} (\ln \ln N)^{2/3}\right) \\ &= L_N\left[\frac{1}{3}, \frac{-1}{3\delta\beta}\right]. \end{aligned}$$

Počet hladkých relací, které obdržíme z celočíselné fáze, je tedy

$$M^2 u^{-u} = \frac{L_N\left[\frac{1}{3}, 2\epsilon\right]}{L_N\left[\frac{1}{3}, \frac{1}{3\delta\beta}\right]} = L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right].$$

Pro každý polynom f_i máme

$$|N_{f_i}(a - b\alpha_i)| \leq (d+1) M^d N^{1/d} \frac{B_1'}{B_2'} = L_N\left[\frac{2}{3}, \epsilon\delta + \frac{1}{\delta}\right],$$

neboť asymptoticky nejvýznamnější členy v tomto součinu jsou $N^{1/d} = L_N\left[\frac{2}{3}, \frac{1}{\delta}\right]$ a $M^d = L_N\left[\frac{1}{3}, \epsilon d\right] = L_N\left[\frac{2}{3}, \epsilon\delta\right]$. Tím pádem je pravděpodobnost, že $N_{f_i}(a - b\alpha_i)$ bude B_2-hladká, rovna u^{-u} pro

$$u = \frac{\ln L_N\left[\frac{2}{3}, \epsilon\delta + \frac{1}{\delta}\right]}{\ln B_2} = \frac{\epsilon\delta + \frac{1}{\delta}}{\gamma}\left(\frac{\ln N}{\ln\ln N}\right)^{1/3}.$$

Potom

$$\begin{aligned} u^{-u} &= \exp\left(-\frac{\epsilon\delta + \frac{1}{\delta}}{\gamma}\left(\frac{\ln N}{\ln\ln N}\right)^{1/3}\left(\ln\frac{\epsilon\delta + \frac{1}{\delta}}{\gamma} + \frac{1}{3}\ln\ln N - \frac{1}{3}\ln\ln\ln N\right)\right) \\ &= \exp\left(\left(-\frac{\epsilon\delta + \frac{1}{\delta}}{3\gamma} + o(1)\right)(\ln N)^{1/3}(\ln\ln N)^{2/3}\right) \\ &= L_N\left[\frac{1}{3}, -\frac{\epsilon\delta + \frac{1}{\delta}}{3\gamma}\right], \end{aligned}$$

takže pro každé f_i podmínku hladkosti splňuje

$$\frac{L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right]}{L_N\left[\frac{1}{3}, \frac{\epsilon\delta + \frac{1}{\delta}}{3\gamma}\right]} = L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta}{3\gamma} - \frac{1}{3\gamma\delta}\right]$$

relací. My potřebujeme, aby jich bylo alespoň $2B_2' = L_N\left[\frac{1}{3}, \gamma\right]$. Dostáváme tedy podmínku

$$2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta}{3\gamma} - \frac{1}{3\gamma\delta} \geq \gamma.$$

Nyní si rozmyslíme, jakou složitost mají jednotlivé části algoritmu:

- celočíselná fáze: to nám je už známo z kapitoly o kvadratickém sítu, vyjde tedy jako $O(M^2 \ln\ln B_1) = L_N\left[\frac{1}{3}, 2\epsilon\right]$,
- polynomiální fáze: tady musíme na rozdíl od předchozí fáze dělit jednotlivé relace samostatně (neboť (a, b) již nejsou rovnoměrně rozděleny). K tomu můžeme použít metodu eliptických křivek. Podle Věty 2.3 každou hodnotu zpracujeme v čase $B^{o(1)} = L_N\left[\frac{1}{3}, \beta o(1)\right] = L_N\left[\frac{1}{3}, o(1)\right]$. Pro každý polynom jich máme $L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right]$, složitost polynomiální fáze proto bude $\frac{B_1'}{B_2'} L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right] L_N\left[\frac{1}{3}, o(1)\right] = L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta} + \beta - \gamma\right]$ (člen $o(1)$ zde stejně jako ve většině případů pro přehlednost nepíšeme),

- lineární fáze: $O\left((B_1')^2\right) = L_N\left[\frac{1}{3}, 2\beta\right]$,
- odmocňování: pro každé f_i zabere nanejvýš $O\left(B_2^2\right)$ času, takže dohromady to můžeme odhadnout pomocí $B_1 B_2 = L_N\left[\frac{1}{3}, \beta + \gamma\right]$.

Celková složitost je tedy součtem těchto čtyř částkových složitostí. My se ji snažíme minimalizovat. Protože v součtu asymptoticky záleží pouze na největším členu, řešíme vlastně následující problém:

$$\begin{aligned} \text{minimalizuj} \quad & \max\left(2\epsilon, 2\epsilon - \frac{1}{3\delta\beta} + \beta - \gamma, 2\beta, \beta + \gamma\right) \\ \text{vzhledem k} \quad & 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta}{3\gamma} - \frac{1}{3\gamma\delta} \geq \gamma \\ & \beta, \gamma, \delta, \epsilon \geq 0. \end{aligned}$$

Optimálními hodnotami jsou

$$\begin{aligned} \beta = \epsilon &= \left(\frac{46 + 13\sqrt{13}}{108}\right)^{1/3} \approx 0.95094, \\ \gamma &= \frac{1}{3}\left(2\left(4 + \sqrt{13}\right)\right)^{1/3} \approx 0.82591, \\ \delta &= \frac{\left(2\left(16 - \sqrt{13}\right)\right)^{1/3}}{3^{2/3}} \approx 1.40175, \end{aligned}$$

takže jako časově nejnáročnější pak vychází celočíselná a lineární fáze, které stanovují složitost celého algoritmu na

$$L_N\left[\frac{1}{3}, 2\beta\right] \approx L_N\left[\frac{1}{3}, 1.90188\right].$$

To je v porovnání se základním číselným sítem poměrně malá úspora. Navíc se ukazuje, že při hodnotách N, se kterými se reálně pracuje, nakonec převáží přidané úsilí ve formě faktorizace metodou eliptických křivek místo prosívání během polynomiální fáze a výpočty ve větším množství číselných těles. Proto se tato varianta prakticky často nepoužívá.

Mohlo by nás napadnout, zdali by nějaká jiná volba počtu polynomů nevedla k lepšímu výsledku. Na první pohled nejspíš není zřejmé, proč by právě $L_N\left[\frac{1}{3}, \beta - \gamma\right]$ mělo být nejlepší možností. Pojďme si tedy tuto variantu zobecnit.

Označme $P = L_N\left[\frac{1}{3}, \rho\right]$ počet polynomů. Každému z nich bude v matici lineární fáze patřit B_2' sloupců. Tato matice proto bude mít

$$B_1' + PB_2' = L_N\left[\frac{1}{3}, \beta\right] + L_N\left[\frac{1}{3}, \rho + \gamma\right] = L_N\left[\frac{1}{3}, \max\left(\beta, \rho + \gamma\right)\right]$$

sloupců. Z toho plyne, že podmínku hladkosti by pro každý polynom mělo splňovat alespoň $L_N\left[\frac{1}{3}, \max\left(\beta - \rho, \gamma\right)\right]$ relací. Podívejme se, jak se změní složitosti jednotlivých částí algoritmu:

- celočíselná fáze: pořád $L_N\left[\frac{1}{3}, 2\epsilon\right]$,
- polynomiální fáze: $PL_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right] L_N\left[\frac{1}{3}, o\left(1\right)\right] = L_N\left[\frac{1}{3}, \rho + 2\epsilon - \frac{1}{3\delta\beta}\right]$,

- lineární fáze: $L_N\left[\frac{1}{3}, 2\max\left(\beta, \rho+\gamma\right)\right]$,
- odmocňování: pro každý polynom maximálně $O\left(B_2^2\right)$, což dohromady dává $O\left(PB_2^2\right) = L_N\left[\frac{1}{3}, \rho+2\gamma\right]$.

Řešíme tedy optimalizační problém

$$\begin{aligned}
\text{minimalizuj} \quad & \max\left(2\epsilon, \rho+2\epsilon-\frac{1}{3\delta\beta}, 2\max\left(\beta,\rho+\gamma\right), \rho+2\gamma\right) \\
\text{vzhledem k} \quad & 2\epsilon-\frac{1}{3\delta\beta}-\frac{\epsilon\delta}{3\gamma}-\frac{1}{3\gamma\delta} \geq \beta-\rho \\
& 2\epsilon-\frac{1}{3\delta\beta}-\frac{\epsilon\delta}{3\gamma}-\frac{1}{3\gamma\delta} \geq \gamma \\
& \beta,\gamma,\delta,\epsilon,\rho \geq 0.
\end{aligned}$$

Podle očekávání dostáváme stejné optimální hodnoty, přičemž

$$\rho = \beta-\gamma = \frac{\left(4\sqrt{13}-14\right)^{1/3}}{6} \approx 0.12503.$$

2.3.3 Verze s více počítači

Nyní navrhneme vlastní modifikaci, kterou se budeme snažit minimalizovat složitost v situaci, kdy máme k dispozici více počítačů. Vyjdeme z právě představené verze s více polynomy. Pokusíme se ji urychlit tím, že paralelizujeme výpočty v polynomiální fázi. Každému počítači můžeme přiřadit vlastní polynom, pro nějž bude hledat hladké relace, které se složí na čtverec v příslušném okruhu $\mathbb{Z}\left[\alpha_{f_i}\right]$. Výsledky z jednotlivých větví pak zkombinujeme v další lineární fázi tak, aby vyšel čtverec i v $\mathbb{Z}$.

Novým parametrem je tedy počet počítačů, označme jej

$$k = L_N\left[\frac{1}{3}, \kappa\right].$$

Nejdřív potřebujeme zajistit celočíselnou hladkost. Jinými slovy, prosíváme hodnoty $a-mb$. Postup se zatím neliší od verze s více polynomy, můžeme si proto pomoct výpočty z předchozí sekce. To znamená, že v celočíselné fázi dostaneme

$$L_N\left[\frac{1}{3}, 2\epsilon-\frac{1}{3\delta\beta}\right]$$

hladkých relací.

V dalším kroku pro každý polynom f_i, $i=1,\ldots,k$, dělíme $N_{f_i}\left(a-b\alpha_{f_i}\right)$. Jelikož nyní máme k polynomů místo B_1'/B_2', absolutní hodnota norem bude omezena výrazem

$$\left(d+1\right)M^d N^{1/d} k,$$

který ovšem v L-notaci dá totožný výsledek, tj.

$$L_N\left[\frac{2}{3}, \epsilon\delta+\frac{1}{\delta}\right].$$

Nezmění se tudíž ani pravděpodobnost hladkosti $N_{f_i}(a - b\alpha_{f_i})$, pro všechny f_i je očekávaný počet hladkých relací

$$L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta}{3\gamma} - \frac{1}{3\gamma\delta}\right].$$

Každý počítač nyní potřebuje z těchto relací v lineární fázi sestavit čtverec v okruhu $\mathbb{Z}[\alpha_{f_i}]$. Tyto relace se zapisují do matice s $L_N\left[\frac{1}{3}, \gamma\right]$ sloupci (π_{B_2} sloupců pro prvočinitele a $O(\ln N)$ pro charaktery). Z toho plyne podmínka

$$2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta}{3\gamma} - \frac{1}{3\gamma\delta} \geq \gamma.$$

Tímto procesem získáme k relací, z nichž v další lineární fázi potřebujeme vytvořit čtverec v $\mathbb{Z}$. Teď pracujeme s maticí s $\pi_{B_1} = L_N\left[\frac{1}{3}, \beta\right]$ sloupci. Další podmínkou je proto $\kappa \geq \beta$. Stačí tedy vzít $\kappa = \beta$ (neboli $k = B_1$).

Pojďme si rozmyslet, zda tímto postupem dokážeme vylepšit složitost. Opět si můžeme pomoct analýzou z předchozí sekce. Jednotlivé fáze vychází následovně:

- celočíselné prosívání: $O(M^2 \ln\ln B_1) = L_N\left[\frac{1}{3}, 2\epsilon\right]$,
- polynomiální fáze: pro každý polynom $L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right]$ pomocí metody eliptických křivek,
- první lineární fáze: pro každý polynom $O(B_2^2) = L_N\left[\frac{1}{3}, 2\gamma\right]$,
- druhá lineární fáze: $O(B_1^2) = L_N\left[\frac{1}{3}, 2\beta\right]$,
- odmocňování: pro každý polynom nanejvýš $O(B_2^2) = L_N\left[\frac{1}{3}, 2\gamma\right]$.

Je tedy potřeba řešit optimalizační problém tvaru

$$\begin{aligned} \text{minimalizuj} \quad & \max\left(2\epsilon, 2\epsilon - \frac{1}{3\delta\beta}, 2\gamma, 2\beta\right) \\ \text{vzhledem k} \quad & 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta}{3\gamma} - \frac{1}{3\gamma\delta} \geq \gamma \\ & \beta, \gamma, \delta, \epsilon \geq 0. \end{aligned}$$

Optimálními hodnotami jsou

$$\begin{aligned} \kappa = \beta = \epsilon &= \left(\frac{46 + 13\sqrt{13}}{108}\right)^{1/3} \approx 0.95094, \\ \gamma &= \frac{1}{3}\left(2\left(4 + \sqrt{13}\right)\right)^{1/3} \approx 0.82591, \\ \delta &= \frac{\left(2\left(16 - \sqrt{13}\right)\right)^{1/3}}{3^{2/3}} \approx 1.40175, \end{aligned}$$

což se shoduje s výsledky ve verzi s více polynomy. Mohlo by se tedy zdát, že obě modifikace jsou ekvivalentní. To však není pravda. Problémem je, že přidáním počítačů šetříme ve fázích, které nejsou asymptoticky dominantní.

Všimněme si, že polynomiální fáze nyní trvá $L_N\left[\frac{1}{3}, 2\epsilon - 3\delta\beta\right]$, zatímco ve verzi s více polynomy trvala $L_N\left[\frac{1}{3}, 2\epsilon - 3\delta\beta + \beta - \gamma\right]$, kde $\beta - \gamma > 0$. Drobnou úsporu získáme i v odmocňování, neboť $2\gamma < \beta + \gamma$.

O něco rychleji poběží dokonce i lineární fáze, přestože ji provádíme dvakrát. Ve verzi s více polynomy máme její složitost odhadnutou jako $O\left((B_1')^2\right)$. Podíváme-li se podrobněji na příslušný výpočet, zjistíme, že jsme k tomuto výsledku dospěli pomocí aproximace $B_1' + B_2' = O\left(B_1'\right)$, kde

$$\begin{aligned} B_1' &= \pi_{B_1} + 1, \\ B_2' &= \pi_{B_2} + O\left(\ln N\right). \end{aligned}$$

Složitost tedy v tomto případě vychází přesněji $O\left((B_1' + B_2')^2\right)$, což samozřejmě asymptoticky nic nemění. Ve verzi s více počítači ovšem lineární fáze dohromady zabere pouze $O\left((B_1')^2 + (B_2')^2\right)$. To bohužel nepostačuje k zisku nějaké asymptotické úspory. Tato fáze proto společně s celočíselným prosíváním (které se jako jediné nijak nezrychlí) stanovuje celkovou složitost algoritmu opět na

$$L_N\left[\frac{1}{3}, 2\epsilon\right] = L_N\left[\frac{1}{3}, 2\beta\right] \approx L_N\left[\frac{1}{3}, 1.902\right].$$

Můžeme si pro zajímavost rychle spočítat, kolik touto verzí ušetříme v porovnání s tím, kdybychom ji celou počítali s jediným počítačem. To by mohlo lépe ilustrovat přínos paralelizace.

Změnila by se tedy složitost polynomiálního fáze na $L_N\left[\frac{1}{3}, \beta + 2\epsilon - \frac{1}{3\delta\beta}\right]$, první lineární fáze společně s odmocňováním by nově trvaly $L_N\left[\frac{1}{3}, \beta + 2\gamma\right]$. To znamená, že bychom řešili optimalizační problém

$$\begin{aligned} \text{minimalizuj} \quad & \max\left(2\epsilon, \beta + 2\epsilon - \frac{1}{3\delta\beta}, \beta + 2\gamma, 2\beta\right) \\ \text{vzhledem k} \quad & 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta}{3\gamma} - \frac{1}{3\gamma\delta} \geq \gamma \\ & \beta, \gamma, \delta, \epsilon \geq 0, \end{aligned}$$

což by nás dovedlo k volbě

$$\begin{aligned} \beta &= \left(\frac{\sqrt{33} - 1}{36}\right)^{1/3} \approx 0.5089, \\ \gamma &= \left(\frac{\sqrt{33} + 17}{36}\right)^{1/3} \approx 0.85808, \\ \delta &= \left(\frac{3\left(\sqrt{33} + 17\right)}{32}\right)^{1/3} \approx 1.28711, \\ \epsilon &= \left(\frac{\frac{1}{3}\left(11\sqrt{33} + 69\right)}{32}\right)^{1/3} \approx 1.11252. \end{aligned}$$

Všechny fáze s výjimkou druhé lineární by pak určovaly složitost celého algoritmu na

$$L_N\left[\frac{1}{3}, 2\epsilon\right] = L_N\left[\frac{1}{3}, 2.225\right].$$

Zmiňme na závěr ještě jednu variantu – předpokládejme, že máme předem pevně daný počet počítačů. To je model, který by mohl nejlépe reflektovat situaci v praxi. Pořád platí, že potřebujeme přibližně B_1 polynomů. Každý počítač jich tedy musí „zpracovat“

$$\frac{B_1}{k} = L_N\left[\frac{1}{3}, \beta - \kappa\right].$$

Složitost bude ovlivněna následovně:

- polynomiální fáze: každý počítač provede $L_N\left[\frac{1}{3}, \beta - \kappa\right]$ výpočtů v čase $L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right]$, takže složitost této fáze bude $L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta} + \beta - \kappa\right]$,
- první lineární fáze & odmocňování: obdobnou logikou $L_N\left[\frac{1}{3}, 2\gamma + \beta - \kappa\right]$.

Budeme tedy hledat řešení problému tvaru

$$\begin{aligned} \text{minimalizuj} \quad & \max\left(2\epsilon, 2\epsilon - \frac{1}{3\delta\beta} + \beta - \kappa, 2\gamma + \beta - \kappa, 2\beta\right) \\ \text{vzhledem k} \quad & 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta}{3\gamma} - \frac{1}{3\gamma\delta} \geq \gamma \\ & \beta, \gamma, \delta, \epsilon, \kappa \geq 0. \end{aligned}$$

Můžeme případně přidat podmínku $\kappa \leq \beta$, víc počítačů nám už totiž stejně nepomůže.

Výsledky v obecném tvaru (v závislosti na κ) zde kvůli náročnosti jejich popisu uvádět nebudeme, dopočítat je pro konkrétní hodnoty (jak jsme to dělali doposud) samozřejmě není problém. Z předchozích variant můžeme také nahlédnout, že složitost v tomto případě vyjde jako $L_N\left[\frac{1}{3}, x\right]$, kde $1.902 \leq x \leq 2.225$, přičemž toto x klesá s rostoucím k.

2.3.4 Verze s předvýpočtem hladkých relací

Druhá modifikace prezentovaná v Coppersmith (1993) je ve skutečnosti velmi podobná základní verzi číselného síta. Rozdíl spočívá pouze v tom, že hledání relací (a, b), pro něž je $a - bm$ hladké, zahrneme do předvýpočtu. Výsledky (vektory mocnin) si zapíšeme do tabulky. Následně, poté co dostaneme na vstup číslo N, si (nám už dobře známým postupem) zvolíme polynom f a budeme testovat hladkost $N(a - b\alpha)$ všech (a, b) z této tabulky. Zbylé fáze probíhají beze změny.

Vidíme, že v tomto případě potřebujeme zvolit m předem a tudíž jej používat pro všechny výpočty. Algoritmus proto bude fungovat pouze pro N v intervalu $\left[m^d, m^{d+1}\right)$, což ale není nijak zásadní omezení, protože tento rozsah je docela velký.

Jak se tedy změní složitost, nezapočítáme-li prosívaní hodnot $a - bm$? Podle analýzy z verze s více polynomy dostaneme z předvýpočtu $L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right]$ hladkých relací. Tento výraz tedy udává i složitost „polynomiální fáze“.

Na konci nám zbude (opět viz analogický výpočet z verze s více polynomy)

$$\frac{L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right]}{L_N\left[\frac{1}{3}, \frac{\epsilon\delta + \frac{1}{\delta}}{3\beta}\right]} = L_N\left[\frac{1}{3}, 2\epsilon - \frac{2}{3\delta\beta} - \frac{\epsilon\delta}{3\beta}\right]$$

hladkých relací (používáme pouze jednu hladkou mez $B = B_1 = B_2 = L_N\left[\frac{1}{3}, \beta\right]$). Matice lineární fáze má $2\pi_B + 2 = L_N\left[\frac{1}{3}, \beta\right]$ sloupců, potřebujeme jich tedy alespoň tolik. Dostáváme podmínku

$$2\epsilon - \frac{2}{3\delta\beta} - \frac{\epsilon\delta}{3\beta} \geq \beta.$$

Lineární fázi a odmocňování zvládneme v čase $O\left(B^2\right) = L_N\left[\frac{1}{3}, 2\beta\right]$. Celková složitost je tudíž

$$L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right] + L_N\left[\frac{1}{3}, 2\beta\right].$$

To nás přivádí k optimalizačnímu problém tvaru

$$\begin{aligned} \text{minimalizuj} \quad & \max\left(2\epsilon - \frac{1}{3\delta\beta}, 2\beta\right) \\ \text{vzhledem k} \quad & 2\epsilon - \frac{2}{3\delta\beta} - \frac{\epsilon\delta}{3\beta} \geq \beta \\ & \beta, \delta, \epsilon \geq 0, \end{aligned}$$

jehož řešením je tentokrát

$$\begin{aligned} \beta &= \left(\frac{5 + 2\sqrt{6}}{18}\right)^{1/3} \approx 0.8193, \\ \delta &= \left(3\left(\sqrt{6} - 2\right)\right)^{1/3} \approx 1.1048, \\ \epsilon &= \left(\frac{5 + 2\sqrt{6}}{4\sqrt{6}}\right)^{1/3} \approx 1.0034. \end{aligned}$$

Protože $2\epsilon - \frac{1}{3\delta\beta} = 2\beta$, všechny fáze jsou stejně časově náročné. Pro každé faktorizované N tedy dostáváme složitost

$$L_N\left[\frac{1}{3}, 2\beta\right] \approx L_N\left[\frac{1}{3}, 1.6386\right]$$

za cenu předvýpočtu složitosti

$$L_N\left[\frac{1}{3}, 2\epsilon\right] \approx L_N\left[\frac{1}{3}, 2.0068\right].$$

To je už značná úspora oproti základní verzi. Má to však jeden háček. Potřebujeme totiž dlouhodobě držet v paměti tabulku velikosti řádově

$$L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right] \approx L_N\left[\frac{1}{3}, 1.6386\right],$$

což je z praktického (finančního) hlediska nereálné.

Ukážeme si ještě jedno schéma založené na tomto principu. Inspirováno je návrhem faktorizační metody prezentované v Schnorr (1982). Můžeme jej vnímat jako speciální případ výše uvedené verze.

Nejprve zvolíme m přibližně stejné velikosti jako čísla N, která později očekáváme na vstupu. Algoritmus pak bude fungovat pouze pro N splňující podmínku

$|N-m| < L_N\left[\frac{2}{3},\epsilon\right]$, což je oproti předchozímu $m^d \leq N < m^d \cdot L_N\left[\frac{2}{3},\frac{1}{\delta}\right]$ výrazně menší rozsah.

Dále položíme

$$\begin{aligned} f(\alpha) &= \alpha + (N-m), \\ B_1 = B_2 &= L_N\left[\frac{1}{3},\beta\right], \\ M &= L_N\left[\frac{2}{3},\epsilon\right]. \end{aligned}$$

Předvýpočet bude zahrnovat hledání dvojic $(a,-1)$, kde $|a| < M$ takových, že $a - bm = a + m$ je B_1-hladké. Pravděpodobnost, že tato situace nastane, je u^{-u}, kde

$$u = \frac{\ln N}{\ln B_1} = \frac{\ln N}{\beta (\ln N)^{1/3} (\ln\ln N)^{1/3}} = \frac{1}{\beta}(\ln N)^{2/3}(\ln\ln N)^{-2/3}.$$

Tedy $u^{-u} = L_N\left[\frac{2}{3},-\frac{2}{3\beta}\right]$ a očekávaný počet hladkých relací je tudíž

$$Mu^{-u} = L_N\left[\frac{2}{3},\epsilon - \frac{2}{3\beta}\right].$$

My jich budeme chtít alespoň $L_N\left[\frac{1}{3},\beta+\frac{\epsilon}{3\beta}\right]$, což bude splněno kdykoli $\epsilon > \frac{2}{3\beta}$ (v praxi bychom samozřejmě potřebovali $\epsilon - \frac{2}{3\beta} > \epsilon'$ pro nějaké malé ϵ').

Nyní pro nějaké konkrétní přirozené číslo N z rozsahu $|N-m| < L_N\left[\frac{2}{3},\epsilon\right]$, které chceme faktorizovat, nalezneme trojice $(a,-1,f)$ splňující

$$N_f(a-b\alpha) = b^d f\left(\frac{a}{b}\right) = -f(-a) = a - (N-m)$$

je B_2-hladké. Takovou trojici objevíme s pravděpodobností

$$u^{-u} = L_N\left[\frac{1}{3},-\frac{\epsilon}{3\beta}\right],$$

neboť $|a-(N-m)| < L_N\left[\frac{2}{3},\epsilon\right]$, takže

$$u = \frac{\ln L_N\left[\frac{2}{3},\epsilon\right]}{\ln L_N\left[\frac{1}{3},\beta\right]} = \frac{\epsilon}{\beta}(\ln N)^{1/3}(\ln\ln N)^{-1/3}.$$

Zbude nám tedy

$$\frac{L_N\left[\frac{1}{3},\beta+\frac{\epsilon}{3\beta}\right]}{L_N\left[\frac{1}{3},\frac{\epsilon}{3\beta}\right]} = L_N\left[\frac{1}{3},\beta\right] = B_1$$

relací. Z nich vytvoříme matici lineární fáze a výpočet dokončíme klasickým způsobem.

Jak v tomto případě vyjde složitost? Polynomiální fáze zabere $L_N\left[\frac{1}{3},\beta+\frac{\epsilon}{3\beta}\right]$ a řešení soustavy $L_N\left[\frac{1}{3},2\beta\right]$ času. Odmocňování je nyní opravdu zanedbatelné,

protože počítáme v $\mathbb{Z}$. Dostáváme se tedy k optimalizačnímu problému

$$\begin{aligned} \text{minimalizuj} \quad & \max\left(\beta + \frac{\epsilon}{3\beta}, 2\beta\right) \\ \text{vzhledem k} \quad & \epsilon > \frac{2}{3\beta} \\ & \beta, \epsilon \geq 0. \end{aligned}$$

Řešení leží na hranici $\epsilon = \frac{2}{3\beta}$, což můžeme, jak bylo zmíněno, ošetřit přidáním malého ϵ'. Tímto řešením je nicméně

$$\begin{aligned} \beta &= \frac{2^{1/3}}{3^{2/3}} \approx 0.6057, \\ \epsilon &= \frac{2^{2/3}}{3^{1/3}} \approx 1.1006. \end{aligned}$$

To dává pro jedno N složitost pouze

$$L_N\left[\frac{1}{3}, 2\beta\right] \approx L_N\left[\frac{1}{3}, 1.2114\right],$$

ovšem za nepřiměřeně vysokou cenu předvýpočtu $L_N\left[\frac{2}{3}, \epsilon\right]$ a též stejné paměťové náročnosti. Toto schéma má proto k praktické použitelnosti bohužel ještě dál, než ostatní představené modifikace. Zůstává tedy spíše jen zajímavým teoretickým konceptem.

2.4 Randomizované číselné síto (RNFS)

Analýza složitosti číselného síta a veškerých jeho doposud prezentovaných modifikacích z velké části stojí na heuristických odhadech. Základním pilířem všech verzí je předpoklad existence dostatečného množství hladkých relací, abychom měli jistotu, že v lineární fázi nalezneme netriviální řešení soustavy. Přestože umíme přibližně spočítat pravděpodobnost, zda bude obecně nějaké celé číslo v závislosti na jeho absolutní hodnotě hladké, nedokážeme to s určitostí říct i pro specifické hodnoty, s kterými v číselném sítě pracujeme. Problémem je třeba uplatnění této obecné pravděpodobnosti v polynomiální fázi, kde zkoumáme hladkost norem $N_f(a - b\alpha)$ pouze pro (a, b) takové, že $a - bm$ je hladké. Díky tomu nemáme garantováno, že hladkých relací opravdu najdeme tolik, kolik potřebujeme, případně s jakou pravděpodobností.

Ani úspěch této části algoritmu ale pořád nestačí k tomu, abychom dospěli ke kýžené kongruenci čtverců. Lineární fáze sice nyní nalezne netriviální řešení, může se nicméně stát, že nedostaneme čtverec v příslušném okruhu $\mathbb{Z}[\alpha]$ a nepodaří se nám odmocnit. Tomu se snažíme zabránit přidáním kvadratických charakterů (viz sekce 2.3.1). S jejich zvyšujícím se počtem roste také pravděpodobnost úspěšného odmocnění, nikdy ovšem nedosáhne 1. Navíc jsme i tady nuceni spoléhat na některé heuristické předpoklady.

Snaha vyřešit tyto problémy vedla k návrhu randomizovaného číselného síta prezentovaného v článku Lee a Venkatesan (2018). Ukážeme si, jak jejich algoritmus modifikuje klasické číselné síto tak, aby šlo stejnou časovou složitost dokázat

i pomocí rigorózní (nikoli heuristické) analýzy. Navíc nám RNFS poslouží jako inspirace k podobné úpravě Coppersmithovy verze s více polynomy, kterou jsme se zabývali v sekci 2.3.2.

Jak nám napovídá název, RNFS do volby některých parametrů zakomponuje náhodu. V základní verzi číselného síta máme přesný postup, jak volit polynom f v závislosti na d a m. Jeho koeficienty jsou určeny zápisem m v bázi N. RNFS tyto koeficienty bere pravděpodobnostně, přičemž zachová základní vlasnost, tj. že m je kořen f modulo N. Podobně randomizujeme volbu kvadratických charakterů v lineární fázi.

To nám umožňuje analyzovat složitost v *průměrném případě* na rozdíl od nejhoršího, jako tomu bylo doposud. V některých oblastech jsou pak výpočty jednodušší, díky čemu se dokážeme zbavit veškeré heuristiky. Výsledkem je určení *střední hodnoty* času potřebného k nalezení kongruentních čtverců modulo N a pravděpodobnosti selhání algoritmu. Průměrný čas sice vyjde stejně jako v základní verzi číselného síta, čili

$$L_N\left[\frac{1}{3}, \frac{4}{3^{2/3}}\right] \approx L_N\left[\frac{1}{3}, 1.923\right],$$

nyní už ovšem s rigorózní analýzou.

Důležité tady je, že střední hodnota se nepočítá přes vstupy do algoritmu, nýbrž přes jednotlivé parametry voleny v jeho průběhu. Jinými slovy, průměrujeme časovou složitost pro každé N zvlášť. Tento výsledek lze tudíž očekávat nezávisle na jeho hodnotě.

RNFS se částečně zabývá ještě jedním problémem, který jsme zatím nezmínili. Na výstupu dostáváme kongruenci $x^2 \equiv y^2 \pmod{N}$. Víme, že $\mathrm{NSD}\,(N, x \pm y)$ dává vlastního dělitele N právě tehdy, když $x \not\equiv \pm y \pmod{N}$. To dle Tvrzení 1.1 pro náhodně vybrané x, y nastane s alespoň poloviční pravděpodobností. Platí to ovšem i pro dvojici vygenerovanou číselným sítem? Nevíme. Pouze heuristicky předpokládáme, že ano. RNFS dokáže tuto naši hypotézu ve specifickém případě $N = pq$, $p \equiv q \equiv 3 \pmod{4}$ potvrdit. Této oblasti se dál věnovat nebudeme, podrobnou diskuzi a důkaz lze najít v Lee a Venkatesan (2018).

2.4.1 Obecný princip

Nadále budeme používat značení, které známe z klasického číselného síta. Zvolíme si hladké meze

$$B_1 = L_N\left[\frac{1}{3}, \beta\right],$$
$$B_2 = L_N\left[\frac{1}{3}, \gamma\right].$$

Kromě toho zafixujeme

$$d \approx \delta\left(\frac{\ln N}{\ln\ln N}\right)^{1/3},$$
$$M = L_N\left[\frac{1}{3}, \epsilon\right],$$

které nám stanoví stupeň polynomu f, resp. rozsah bodů, v nichž ho budeme vyhodnocovat. Jediným novým parametrem bude

$$k = L_N\left[\frac{2}{3}, \kappa\right],$$

které omezí absolutní hodnoty jeho koeficientů.

Pro fungování algoritmu je nezbytné, aby byly splněny následující podmínky:

(A) $\delta^{-1} < \kappa$,

(B) $\delta^{-1} < \frac{\epsilon\delta+\kappa}{2}$,

(C) $2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta+\kappa}{3\gamma} \geq \max(\beta, \gamma)$.

Jejich význam si vysvětlíme později. Prozatím je tedy ponecháme bez komentáře a přejdeme k samotnému principu RNFS.

Algoritmus si rozdělíme na dvě části. Ta první se bude zabývat vygenerováním dostatečného počtu hladkých relací. Můžeme ji tedy ztotožnit s celočíselnou plus polynomiální fází dle rozdělení, které jsme používali doposud. Popíšeme si nejzásadnější rozdíl oproti klasickému číselnému sítu a vysvětlíme, jak ho lze využít k odstranění heuristiky ve výpočtu složitosti.

Začneme hned popisem tohoto stěžejního rozdílu, kterým je způsob volby polynomu f. Doposud jsme brali $f = \sum a_i x^i$, přičemž a_i byly určeny zápisem N v bázi

$$m \approx L_N\left[\frac{2}{3}, \frac{1}{\delta}\right],$$

tj. $N = \sum a_i m^i$. Řekli jsme, že RNFS deterministickou volbu koeficientů f mění na pravděpodobnostní. Označme $f_{n,m}$ zmíněný polynom vzniklý zápisem m v bázi N. Nově definujme

$$R(x) = \sum_{i=0}^{d-1} c_i (x - m) x^i,$$

kde c_i jsou náhodně volená čísla z intervalu $\left[-L_N\left[\frac{2}{3}, \kappa - \delta^{-1}\right], L_N\left[\frac{2}{3}, \kappa - \delta^{-1}\right]\right]$. Podmínka (A) nám zajistí, že bude obsahovat i nenulová celá čísla.

Položme

$$f(x) = f_{n,m}(x) + R(x).$$

Tím jsme do volby f přidali požadovaný náhodný prvek. Všimněme si, že pořád platí jeho základní vlastnost, čili $f(m) \equiv 0 \pmod N$. Můžeme tady vidět jistou analogii s verzí s více polynomy – volba jednotlivých f_i by se dala popsat stejným způsobem, s tím, že $c_0 = i$ a $c_j = 0$ pro všechna $1 \leq j \leq d-1$.

Už víme, že absolutní hodnota koeficientů $f_{n,m}$ je nejvýše $L_N\left[\frac{2}{3}, \delta^{-1}\right]$. Snadno lze nahlédnout, že u polynomu $R(x)$ to bude maximálně $L_N\left[\frac{2}{3}, \kappa\right]$. Díky podmínce (A) jsou tedy koeficienty f v absolutní hodnotě opravdu omezeny $L_N\left[\frac{2}{3}, \kappa\right]$.

Na rozdíl od základní verze číselného síta budeme v průběhu algoritmu postupně volit různé dvojice (m, f) s výše uvedenými vlastnostmi a hledat příslušné hladké relace. Zavedeme si proto nové značení.

Definice. Mějme m, f jako výše. Řekneme, že $(a, b) \in X_{m,f}$ právě tehdy, když

- $|a| \leq M$, $1 \leq b \leq M$,

- $\mathrm{NSD}(a, b) = 1$,
- $a - mb$ je B_1-hladké,
- $l_c(f) N_f(a - b\alpha)$ je B_2-hladké, kde $l_c(f)$ je vedoucí koeficient f.

Poznámka. V okruzích $\mathbb{Z}[x]/(f)$, kde $f(x) = l_c(f)(x - \alpha_1) \cdots (x - \alpha_d)$, obecně platí

$$N_f(a - b\alpha) = (a - b\alpha_1) \cdots (a - b\alpha_d) = l_c(f)^{-1} b^d f\left(\frac{a}{b}\right).$$

Zatím jsme vždy pracovali s monickými polynomy, $l_c(f)$ tedy bylo rovno 1. To po přičtení polynomu $R(x)$ už nemusí platit. Protože algoritmus reálně pořád zkoumá hladkost hodnot $b^d f\left(\frac{a}{b}\right)$, v lineární fázi bude nutná drobná úprava. Nalezení kongruence čtverců nyní bude vyžadovat netriviální řešení soustavy se sudým počtem vektorů mocnin (jinými slovy, kongruence musí vzniknout součinem sudého počtu hladkých relací), abychom měli v součinu člen $l_c(f)$ na sudý exponent. Naštěstí to pro nás nepředstavuje žádný zásadní problém, v nejhorším případě najdeme dvě různá řešení s lichým počtem vektorů a sečteme je. Pak požadovanou kongruenci dostaneme ze vztahů

$$\prod_{i=1}^{k} l_c(f)(a_i - b_i m) = z^2 \text{ pro nějaké } z \in \mathbb{Z},$$
$$\prod_{i=1}^{k} l_c(f)(a_i - b_i \alpha) = \gamma^2 \text{ pro nějaké } \gamma \in \mathbb{Z}[\alpha],$$

neboť k je sudé.

Poznámka. Pro analýzu RNFS je klíčové, aby B_2-hladkost $l_c(f) N_f(a - b\alpha)$ byla „dostatečně“ náhodnou událostí. To samozřejmě záleží na polynomu f. Tady vstupuje do hry podmínka (A), která zaručuje, že jeho náhodnost se odvíjí především od polynomu $R(x)$, neboť má výrazně větší koeficienty. Ty jsou (z definice) uniformně náhodné, na rozdíl od $f_{n,m}(x)$.

Poznámka. Nyní si můžeme vysvětlit také význam podmínky (C). Z odvozených mezí pro koeficienty plyne, že

$$\left|b^d f\left(\frac{a}{b}\right)\right| \leq (d + 1) M^d k = L_N\left[\frac{2}{3}, \epsilon\delta + \kappa\right].$$

Dále si pomůžeme výpočty, které jsme prováděli v sekci 2.3.2 (verze s více polynomy). Na začátku máme $L_N\left[\frac{1}{3}, 2\epsilon\right]$ kandidátů na hladké relace (počet nesoudělných dvojic (a, b) takových, že $|a| \leq M$, $1 \leq b \leq M$). Podmínku celočíselné hladkosti ($a - bm$ je B_1-hladké) splňuje $L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta}\right]$ párů. Po provedení polynomiální fáze jich zbylo pouze $L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta + \delta^{-1}}{3\gamma}\right]$. Teď jsou ovšem absolutní hodnoty norem omezeny $L_N\left[\frac{2}{3}, \epsilon\delta + \kappa\right]$ místo $L_N\left[\frac{2}{3}, \epsilon\delta + \delta^{-1}\right]$, což způsobí, že nám zbude

$$L_N\left[\frac{1}{3}, 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta + \kappa}{3\gamma}\right]$$

hladkých relací. Matice lineární soustavy má $L_N\left[\frac{1}{3}, \beta\right] + L_N\left[\frac{1}{3}, \gamma\right]$ sloupců, potřebujeme tedy

$$L_N\left[\frac{1}{3}, \max(\beta, \gamma)\right]$$

vektorů mocnin, abychom měli jistotu nalezení netriviálního řešení. Z toho se odvodí podmínka (C).

Z definice $X_{m,f}$ vidíme, že tato množina obsahuje hladké relace vzhledem k dvojici parametrů m, f. Snahou bude maximalizovat šanci, že alespoň pro jednu z nich jich najdeme dostatečný počet (potřebný k nalezení netriviálního řešení v lineární fázi). Matice má $L_N\left[\frac{1}{3}, \beta\right] + L_N\left[\frac{1}{3}, \gamma\right]$ sloupců, potřebujeme tedy

$$L_N\left[\frac{1}{3}, \max\left(\beta, \gamma\right)\right]$$

hladkých relací.

Toho dosáhneme tzv. *stochastickým prohlubováním*. Základní podstatou této metody je, že pro každou dvojici (m, f) budeme náležení do $X_{m,f}$ zkoumat pro náhodné množství párů (a, b). Hloubka prohledávání (myšleno minimální $M' \leq M$ takové, že pro každé zkoumané (a, b) platí $|a| \leq M', 1 \leq b \leq M'$) bude v průměru zřejmě menší než u základní verze, aby se nezvýšila složitost. Můžeme se tedy spoléhat, že nám to pro některé m, f přesto bude stačit?

Předně si musíme ukázat, že aspoň v průměrném případě platí to, co jsme v číselném sítě heuristicky předpokládali pro každý, a sice že množina $X_{m,f}$ obsahuje dostatečný počet hladkých relací. Nyní tedy stojíme před slabším tvrzením, které již umíme dokázat. To je klíčová výhoda RNFS. Díky přechodu k průměrnému případu jsme schopni rigorózně dokázat tvrzení, na jejichž platnost jsme doposud museli spoléhat heuristicky.

Věta 2.5. *Jestliže platí podmínky (A), (B), (C), pak*

$$\mathbb{E}_{m,f}\left(|X_{m,f}|\right) \geq L_N\left[\frac{1}{3}, 2\epsilon - \frac{1 + o\left(1\right)}{3\delta\beta} - \frac{\left(\epsilon\delta + \kappa\right)\left(1 + o\left(1\right)\right)}{3\gamma}\right].$$

Poznámka. Zkusme si intuitivně rozmyslet, proč by něco takového mělo platit. Víme, že $L_N\left[\frac{1}{3}, 2\epsilon\right]$ je maximální možná velikost $X_{m,f}$. Výraz $\frac{1}{3\delta\beta}$ reprezentuje pravděpodobnost hladkosti $a - bm$, zatímco $\frac{\epsilon\delta+\kappa}{3\gamma}$ značí pravděpodobnost hladkosti $l_c\left(f\right) N_f\left(a - b\alpha\right)$.

Tato úvaha určuje osnovu důkazu. Zřejmě tedy opět budeme potřebovat výsledky jako ve Větě 1.2. Z toho se následně odvodí, že

$$\mathbb{P}_{a,b,m}\left[a - bm \text{ je } B_1\text{-hladké}\right] = L_N\left[\frac{1}{3}, -\frac{1 + o\left(1\right)}{3\delta\beta}\right].$$

Nesložitější je pak ukázat, že pravděpodobnost hladkosti normy je

$$L_N\left[\frac{1}{3}, -\frac{\epsilon\delta + \kappa}{3\gamma}\right]$$

i tehdy, když nebereme v potaz všechny dvojice (a, b), nýbrž pouze ty, pro něž je $a - bm$ hladké. K důkazu této části je nezbytná platnost podmínky (B).

Musíme si ale uvědomit, že část uvedených výsledků stojí na předpokladu ireducibility polynomu f. Způsob, kterým ho nyní generujeme, nám splnění této vlastnosti nezaručuje. Pro reducibilní polynomy žádnou podobnou teorii nemáme,

nezbývá tedy než předpokládat, že algoritmus v těchto případech selže. Naštěstí lze ukázat, že jich není moc. Přesněji

$$\mathbb{P}\left[f \text{ reducibilní}\right] \leq L_N\left[\frac{2}{3}, -\frac{\kappa - \delta^{-1} + o\left(1\right)}{3}\right].$$

Jelikož budeme samplovat nejvýše $L_N\left[\frac{1}{3}, \cdot\right]$ polynomů, tato pravděpodobnost se zahrne do $o\left(1\right)$. Takže střední hodnotu nijak neovlivní, když ji budeme počítat přes všechny (ne jen ireducibilní) polynomy.

Celý důkaz je hodně zdlouhavý a technický, proto ho tady nebudeme provádět. Najít ho lze v Lee a Venkatesan (2018) v kapitolách 5 a 8.

Pokračujme nyní v diskuzi o stochastickém prohlubování. Představme si, že prohledávání v normální hloubce ($|a| \leq M$, $1 \leq b \leq M$) s významnou pravděpodobností selže, tj. že nenalezneme $L_N\left[\frac{1}{3}, \max\left(\beta, \gamma\right)\right]$ hladkých relací, jak očekáváme. To znamená, že $|X_{m,f}|$ bude často menší, než je jejich průměrná velikost. Ve zbývajících případech tedy musí být $|X_{m,f}|$ velká. Tady by nám proto k nalezení kýženého počtu hladkých relací stačilo prohledávat do menší hloubky. Pojďme si tuto úvahu exaktně kvantifikovat.

Věta 2.6. *Buď W náhodná veličina se střední hodnotou μ a $K \geq 1$ takové, že $0 \leq W \leq K\mu$. Potom existuje $i \in \{0, \ldots, \lceil \log_2 K \rceil\}$ splňující*

$$\mathbb{P}\left[W \geq \frac{2^i \mu}{1 + \lceil \log_2 K \rceil}\right] \geq \frac{1}{2^{i+1}}.$$

Důkaz. Sporem. Ať pro každé $i \in \{0, \ldots, \lceil \log_2 K \rceil\}$ platí

$$\mathbb{P}\left[W \geq \frac{2^i \mu}{1 + \lceil \log_2 K \rceil}\right] < \frac{1}{2^{i+1}},$$

neboli

$$\mathbb{P}\left[W < \frac{2^i \mu}{1 + \lceil \log_2 K \rceil}\right] \geq 1 - \frac{1}{2^{i+1}}.$$

Označme

$$w\left(i\right) := \begin{cases} \frac{2^i \mu}{1 + \lceil \log_2 K \rceil} & i \in \{0, \ldots, \lceil \log_2 K \rceil\}, \\ 0 & i = -1. \end{cases}$$

a zkusme nadefinovat W tak, abychom maximalizovali její střední hodnotu.

Dosazením $i = \lceil \log_2 K \rceil$ do nerovnosti v předpokladu zjišťujeme, že v alespoň $\left(1 - 2^{-\lceil \log_2 K \rceil - 1}\right)$-tině případů nabývá W hodnoty menší než $w\left(\lceil \log_2 K \rceil\right)$. Střední hodnota bude největší, pokud ve zbylé $2^{-\lceil \log_2 K \rceil - 1}$-tině případů bude W nabývat maximální hodnoty $K\mu$.

Zároveň z předpokladu pro $i = \lceil \log_2 K \rceil - 1$ plyne, že W bude v alespoň $\left(1 - 2^{-\lceil \log_2 K \rceil}\right)$-tině případů nabývat hodnoty menší než $w\left(\lceil \log_2 K \rceil - 1\right)$. Takže chceme-li maximalizovat střední hodnotu, musí největší možný počet, tj.

$$\left(1 - 2^{-\lceil \log_2 K \rceil - 1}\right) - \left(1 - 2^{-\lceil \log_2 K \rceil}\right) = 2^{-\lceil \log_2 K \rceil - 1}\text{-tina}$$

hodnot W spadat do intervalu $[w(\lceil \log_2 K \rceil - 1), w(\lceil \log_2 K \rceil))$. Podobně dosazením všech $i \in \{0, \ldots, \lceil \log_2 K \rceil\}$ dostaneme, že W nabývá v 2^{-i-1}-tině případů hodnot z intervalu $[w(i-1), w(i))$. Střední hodnota W proto splňuje

$$\mu < \frac{K\mu}{2^{1+\lceil \log_2 K \rceil}} + \sum_{i=0}^{\lceil \log_2 K \rceil} \frac{1}{2^{i+1}} \cdot \frac{2^i \mu}{1 + \lceil \log_2 K \rceil} \leq \frac{\mu}{2} + \frac{\mu}{2} = \mu,$$

což je spor. □

Označme $\tau = 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta+\kappa}{3\gamma} + o(1)$. Potom v naší situaci máme

$$\begin{aligned} W &= |X_{m,f}|, \\ \mu &\geq L_N\left[\frac{1}{3}, \tau\right], \\ K &\leq L_N\left[\frac{1}{3}, 2\epsilon - \tau\right]. \end{aligned}$$

To znamená, že existuje $i^* \in \{0, \ldots, \lceil \log_2 K \rceil\}$ takové, že

$$\mathbb{P}\left[|X_{m,f}| \geq 2^{i^*} L_N\left[\frac{1}{3}, \tau\right]\right] \geq \frac{1}{2^{i^*+1}},$$

neboť $\lceil \log_2 K \rceil = O\left((\ln N)^{1/3} (\ln \ln N)^{2/3}\right)$ se absorbuje do členu $o(1)$. Naším cílem je toto i^* najít.

Budeme tedy postupně volit $i \in \{0, \ldots, \lceil \log_2 K \rceil\}$. Pro každé z nich vygenerujeme 2^{i+1} dvojic (m, f), přičemž pro každou tuto dvojici vygenerujeme $2^{-i} L_N\left[\frac{1}{3}, \max(\beta, \gamma) + 2\epsilon - \tau\right]$ dvojic (a, b), o nichž rozhodneme, zda patří do $X_{m,f}$ či nikoli.

Proč zrovna takovéto hodnoty? Nechť $i = i^*$. Pak je očekávaný počet párů (m, f) splňujících $|X_{m,f}| \geq 2^i L_N\left[\frac{1}{3}, \tau\right]$ po vyzkoušení 2^{i+1} možností alespoň 1. Selhání (nenalezení žádného vhodného páru) nastane pokaždé s nějakou konstantní pravděpodobností. Pokud se nám naopak vhodnou dvojici (m, f) najít podaří, potom

$$\mathbb{P}_{a,b}[(a, b) \in X_{m,f}] \geq 2^i L_N\left[\frac{1}{3}, \tau - 2\epsilon\right],$$

takže po otestování $2^{-i} L_N\left[\frac{1}{3}, \max(\beta, \gamma) + 2\epsilon - \tau\right]$ dvojic (a, b) očekáváme minimálně $L_N\left[\frac{1}{3}, \max(\beta, \gamma)\right]$ hladkých hodnot, což je přesně mez, které se snažíme dosáhnout. Navíc, pravděpodobnost selhání je opět nějaká konstanta.

Stochastické prohlubování jako celek proto selže také s konstantní pravděpodobností. To pro nás bude důležité pro minimalizaci pravděpodobnosti selhání RNFS, jak uvidíme později při analýze jeho složitosti.

Poznámka. Protože dvojice (a, b) pro všechna (m, f) volíme náhodně, zřejmě nelze k testování celočíselné hladkosti použít proces prosívání. S každou dvojicí (a, b) tedy musíme pracovat samostatně. Na podobný proces jsme už zvyklí z polynomiální fáze (testování hladkosti norem), kde se snažíme dané hodnoty faktorizovat pomocí metody eliptických křivek (ECM). Stejný postup aplikujeme i zde. Podle Věty 2.3 nám tudíž otestování hladkosti jedné hodnoty $a - bm$ zabere

$$B_1^{o(1)} = \exp(o(1) \ln B_1) = \exp\left(o(1)\, \beta (\ln N)^{1/3} (\ln \ln N)^{2/3}\right) = L_N\left[\frac{1}{3}, o(1)\right]$$

času. Tentýž výsledek samozřejmě platí i pro B_2-hladkost norem, dohromady tedy otestování jedné dvojice (a, b) vychází na $L_N\left[\frac{1}{3}, o(1)\right]$.

Tím je tedy ukončena první část algoritmu. Ta druhá má za úkol z obdržených relací vytvořit kongruenci čtverců modulo N, což odpovídá lineární fázi a odmocňování. Samotné řešení soustavy lineárních rovnic bude probíhat prakticky beze změny, stěžejní rozdíl bude ve volbě kvadratických charakterů.

Hrubý princip je následující. V základní verzi jsme hledali dvojice (p, r), kde p prvočíslo, $r \in \{0, \ldots, p-1\}$, $f(r) \equiv 0 \pmod{p}$. Tyto dvojice totiž dávají bijekci s prvoideály $P \subset \mathcal{O}_K$ stupně jedna. Protože $\mathcal{O}_K/P \simeq \mathbb{Z}_p$, mohli jsme to, zda je $a - b\alpha$ čtvercem v tomto faktorokruhu, rozhodnout pomocí Legendreova symbolu $\left(\frac{a-br}{p}\right)$. Podobné dvojice budeme hledat i nyní, za účelem randomizace ovšem musíme pracovat také s prvoideály vyšší normy. K tomu je potřeba rozšířit definici Legendreova symbolu (na „polynomiální“ Legendreův symbol), abychom mohli pracovat nad $K[x]/(f)$, K těleso, f ireducibilní polynom. Pak budeme náhodně volit k (do jisté meze) a hledat vhodné dvojice korespondující s prvoideály tohoto stupně. Následně pomocí příslušného rozšířeného Legendreova symbolu určíme, zda je $a - b\alpha$ čtvercem v $\mathcal{O}_K/P$.

Této problematice se nebudeme podrobněji věnovat, neboť by si to vyžadovalo spoustu dalších znalostí z teorie čísel, jejichž vysvětlení je nad rámec této práce. Důkladnou analýzu může čtenář najít v Lee a Venkatesan (2018), kapitola 6.

2.4.2 Složitost

Výpočet sestává ze tří kroků. Nejprve určíme složitost první části RNFS (tj. ekvivalent celočíselné a polynomiální fáze). Poté nastíníme výpočet lineární fáze a odmocňování. Nakonec tyto dva výsledky spojíme dohromady tak, aby celková složitost vycházela co nejmenší. Postupujeme vlastně stejně jako u všech předchozích výpočtů složitosti – spočteme ji pro jednotlivé části algoritmu a pak optimalizujeme parametry s cílem minimalizovat jejich součet.

Proveďme tedy první krok výpočtu.

Věta 2.7. *Jestliže $\beta, \gamma, \delta, \epsilon, \kappa$ splňují podmínky (A), (B), (C) ze str. 51, potom pro každé $N \in \mathbb{N}$ najde RNFS s pravděpodobností alespoň*

$$1 - L_N\left[\frac{2}{3}, \delta^{-1} - \kappa\right]$$

ireducibilní polynom f stupně d s kořenem m modulo N a

$$L_N\left[\frac{1}{3}, \max(\beta, \gamma)\right]$$

různých dvojic $(a, b) \in X_{m,f}$ v průměrném čase

$$L_N\left[\frac{1}{3}, \max(\beta, \gamma) + \frac{1}{3\delta\beta} + \frac{\epsilon\delta + \kappa}{3\gamma}\right].$$

Důkaz. Většinu potřebných argumentů jsme již viděli v předchozí subsekci. Odvodili jsme, že jeden běh stochastického prohlubování uspěje / selže s nějakou konstantní pravděpodobností. Nechť tedy p je pravděpodobnost selhání. Pak pravděpodobnost neúspěchu po logaritmickém počtu opakování je

$$p^{c \ln N} = \exp(c \ln N \ln p) = L_N[1, c \ln p],$$

což je pro vhodné $c > 0$ menší než $L_N\left[\frac{2}{3}, \delta^{-1} - \kappa\right]$. To znamená, že po logaritmickém počtu opakování dosáhneme požadované pravděpodobnosti úspěchu.

Dále víme, že otestování, zda jedna dvojice (a, b) leží v $X_{m,f}$, nám zabere $L_N\left[\frac{1}{3}, o(1)\right]$ času, přičemž pro jedno i jich máme $L_N\left[\frac{1}{3}, \max(\beta, \gamma) + \frac{1}{3\delta\beta} + \frac{\epsilon\delta+\kappa}{3\gamma}\right]$. Počet i, které procházíme, je

$$1 + \lceil \log_2 L_N\left[\frac{1}{3}, \frac{1}{3\delta\beta} + \frac{\epsilon\delta + \kappa}{3\gamma}\right] \rceil = L_N\left[\frac{1}{3}, o(1)\right].$$

Tento proces opakujeme $c \ln N = L_N[0, 1] = L_N\left[\frac{1}{3}, o(1)\right]$ krát. Výsledná průměrná časová složitost je proto

$$L_N\left[\frac{1}{3}, \max(\beta, \gamma) + \frac{1}{3\delta\beta} + \frac{\epsilon\delta + \kappa}{3\gamma}\right],$$

jak jsme chtěli dokázat. □

Ve všech doposud prezentovaných verzích číselného síta vycházela složitost celočíselné a polynomiální fáze dohromady $L_N\left[\frac{1}{3}, 2\epsilon\right]$. Podmínka (C) nám zaručuje, že nyní nedosáhneme horšího výsledku. Samozřejmě, dá se očekávat (a záhy to potvrdíme), že optimalizace nastaví parametry tak, že se z podmínky (C) stane rovnost, čím se složitost opět dostane na stejnou úroveň. Nesmíme ale zapomenout, že teď jsme k ní přišli pomocí rigorózní, nikoli heuristické analýzy. To je zásadní rozdíl.

Následující tvrzení objasňuje složitost druhé části RNFS.

Věta 2.8. *Máme-li pro nějaký ireducibilní polynom f stupně d s kořenem m modulo N*

$$L_N\left[\frac{1}{3}, \max(\beta, \gamma)\right]$$

různých dvojic $(a, b) \in X_{m,f}$, pak RNFS s pravděpodobností alespoň

$$1 - L_N\left[\frac{2}{3}, \frac{\delta - \kappa^{-1}}{3}\right]$$

nalezne kongruenci čtverců modulo N v průměrném čase nejvýše

$$L_N\left[\frac{1}{3}, 2\max\left(\beta, \gamma, \frac{2\delta}{3}\right)\right].$$

Jelikož jsme zde neuváděli popis volby kvadratických charakterů, pro kompletní důkaz se opět odkážeme na Lee a Venkatesan (2018). Nastíníme pouze jeho základní kroky.

Již víme, že časová složitost řešení soustavy lineárních rovnic a odmocňování je $L_N\left[\frac{1}{3}, 2\max(\beta, \gamma)\right]$. Rozdíl oproti všem verzím, které jsme zatím viděli, spočívá ve výpočtu charakterů. Nyní je mnohem komplexnější, vyžaduje si proto nezanedbatelné množství času. Postupně se ukáže, že

- samplování páru (p, r) reprezentující vhodný charakter $\chi_{p,r}$ trvá $L_N\left[\frac{1}{3}, \frac{4\delta}{3}\right]$,
- evaluace charakteru $\chi_{p,r}$ na $a - b\alpha$ trvá $L_N\left[\frac{1}{3}, \frac{2\delta}{3}\right]$,

- počet samplovaných charakterů je $L_N\left[\frac{1}{3}, o\left(1\right)\right]$.

Z toho plyne, že

- samplování všech charakterů trvá $L_N\left[\frac{1}{3}, \frac{4\delta}{3}\right]$,
- evaluace každého charakterů na $L_N\left[\frac{1}{3}, \max\left(\beta, \gamma\right)\right]$ hodnotách dohromady trvá $L_N\left[\frac{1}{3}, \max\left(\beta, \gamma\right) + \frac{2\delta}{3}\right]$.

Společně s řešením soustavy tedy dostáváme složitost

$$L_N\left[\frac{1}{3}, \max\left(2\max\left(\beta, \gamma\right), \max\left(\beta, \gamma\right) + \frac{2\delta}{3}, \frac{4\delta}{3}\right)\right],$$

což je ekvivalentní výrazu v tvrzení.

Problém by tedy nastal, pokud by $\frac{2\delta}{3}$ bylo větší než $\max\left(\beta, \gamma\right)$. Pak by se složitost druhé části RNFS v porovnání s ostatními modifikacemi mohla výrazně zvýšit. Naštěstí se ukáže, se to není náš případ. Optimalizací parametrů dostaneme stejný výsledek jako v základní verzi číselného síta.

Věta 2.9. *RNFS při splnění podmínek (A), (B), (C) ze str. 51 pro libovolné $N \in \mathbb{N}$ s pravděpodobností alespoň*

$$1 - L_N\left[\frac{2}{3}, \delta^{-1} - \kappa\right]$$

nalezne kongruenci čtverců modulo N v průměrném čase

$$L_N\left[\frac{1}{3}, \frac{4}{3^{2/3}}\right] \approx L_N\left[\frac{1}{3}, 1.923\right].$$

Důkaz. Kombinace předchozích dvou tvrzení nás přivádí k optimalizačnímu problému tvaru

$$\begin{aligned}
\text{minimalizuj} \quad & \max\left(\max\left(\beta, \gamma\right) + \frac{1}{3\delta\beta} + \frac{\epsilon\delta + \kappa}{3\gamma}, 2\max\left(\beta, \gamma, \frac{2\delta}{3}\right)\right) \\
\text{vzhledem k} \quad & \delta^{-1} < \kappa \\
& \delta^{-1} < \frac{\epsilon\delta + \kappa}{2} \\
& 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta + \kappa}{3\gamma} \geq \max\left(\beta, \gamma\right) \\
& \beta, \gamma, \delta, \epsilon, \kappa \geq 0,
\end{aligned}$$

jehož limitním optimálním řešením (nalezené hodnoty nejsou řešením zformulovaného problému, leží na hranici množiny přípustných řešení) je

$$\begin{aligned}
\beta = \gamma = \epsilon = \frac{2}{3^{2/3}} &\approx 0.9615, \\
\delta = 3^{1/3} &\approx 1.4423, \\
\kappa = \delta^{-1} = \frac{1}{3^{1/3}} &\approx 0.6934,
\end{aligned}$$

což určuje celkovou složitost RNFS (i obou jeho částí) na

$$L_N\left[\frac{1}{3}, \frac{4}{3^{2/3}}\right] \approx L_N\left[\frac{1}{3}, 1.923\right].$$

□

2.4.3 Randomizované číselné síto s více polynomy

Podobnosti mezi RNFS a Coppersmithovou verzí číselného síta s více polynomy (sekce 2.3.2) jsme se zatím dotkli jenom okrajově. Viděli jsme, že volba polynomu f v randomizovaném číselném sítě je zobecněním způsobu volby ve verzi s více polynomy. Toto pozorování může pro nás být odrazovým můstkem k randomizaci Coppersmithovy verze tak, abychom (stejně jako u klasického číselného síta) nahradili heuristickou analýzu rigorózní při zachovaní asymptotické časové složitosti. Ukažme si návrh takové modifikace.

Můžeme klidně pracovat s obecnější verzí, kde počet polynomů

$$P = L_N\left[\frac{1}{3}, \rho\right]$$

je libovolný. Cílem je tedy najít jedno m a P ireducibilních polynomů f_i s kořenem m modulo N a $L_N\left[\frac{1}{3}, \max(\beta, \gamma+\rho)\right]$ dvojic (a,b) takových, že $(a,b) \in X_{m,f_i}$ pro nějaké f_i.

Předně je potřeba říct, že se mírně změní jedna z podmínek fungování algoritmu. Parametry nyní musí splňovat

(A) $\delta^{-1} < \kappa$,

(B) $\delta^{-1} < \frac{\epsilon\delta+\kappa}{2}$,

(C) $2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta+\kappa}{3\gamma} + \rho \geq \max(\beta, \gamma+\rho)$.

Podmínky (A) a (B) potřebujeme ze stejných důvodů jako v RNFS. Podmínka (C) zaručuje, že nalezneme dostatečný počet hladkých relací. Tento nový tvar vychází z jiného způsobu jejich generování a jiného počtu sloupců matice soustavy. Podrobnější diskuze byla provedena na konci sekce 2.3.2.

Jednotlivé polynomy budeme generovat úplně stejně jako v RNFS. Žádné změny nenastanou ani ve volbě kvadratických charakterů. Největší rozdíl je proto v stochastickém prohlubování. Doposud jsme podle Věty 2.6 volili dvojice (m,f) tak, abychom pro nějakou z nich našli potřebný počet hladkých relací. Nyní budeme Větu 2.6 aplikovat dvakrát. Poprvé k nalezení vhodného m, podruhé vhodné množiny polynomů velikosti P, k nimž budeme generovat dvojice (a,b) a testovat, zda leží v X_{m,f_i} pro nějaké f_i ze zmíněné množiny polynomů.

Pojďme si to tedy popsat podrobněji. Z Věty 2.5 (teď máme trochu jinou podmínku (C), ta je ovšem silnější) plyne, že

$$\mathbb{E}_m\left(\mathbb{E}_f\left(|X_{m,f}|\right)\right) \geq L_N\left[\frac{1}{3}, \tau\right],$$

kde $\tau = 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta+\kappa}{3\gamma} + o(1)$, přičemž $\mathbb{E}_f\left(|X_{m,f}|\right) \leq L_N\left[\frac{1}{3}, 2\epsilon\right]$, protože stejné omezení platí pro každé $X_{m,f}$. Takže podle Věty 2.6 existuje $i \in \{0, \ldots, \lceil \log_2 K \rceil\}$, kde $K \leq L_N\left[\frac{1}{3}, 2\epsilon - \tau\right]$, splňující

$$\mathbb{P}_m\left[\mathbb{E}_f\left(|X_{m,f}|\right) \geq 2^i L_N\left[\frac{1}{3}, \tau\right]\right] \geq \frac{1}{2^{i+1}}.$$

Ke každému i z uvedeného intervalu náhodně zvolíme 2^{i+1} hodnot m. Pak tedy alespoň pro jednu dvojici (m^*, i^*) očekáváme, že bude platit

$$\mathbb{E}_f\left(|X_{m^*,f}|\right) \geq 2^{i^*} L_N\left[\frac{1}{3}, \tau\right],$$

tudíž taky

$$\mathbb{E}_{f^{(P)}}\left(\sum_{f^{(P)}}\left|X_{m^*,f^{(P)}}\right|\right) \geq 2^{i^*} L_N\left[\frac{1}{3}, \tau+\rho\right],$$

kde $f^{(P)}$ značí množinu polynomů velikosti P (každý s kořenem m modulo N). Neúspěch, tj. nenalezení žádné takové dvojice nastane s nějakou konstantní pravděpodobností.

Opětovnou aplikací Věty 2.6 dostáváme, že existuje $j \in \{0, \ldots, \lceil \log_2 K' \rceil\}$, kde $K' \leq L_N\left[\frac{1}{3}, 2\epsilon - \tau\right]$, splňující

$$\mathbb{P}_{f^{(P)}}\left[\sum_{f^{(P)}}\left|X_{m^*,f^{(P)}}\right| \geq 2^j 2^{i^*} L_N\left[\frac{1}{3}, \tau+\rho\right]\right] \geq \frac{1}{2^{j+1}}.$$

Ke každému j z tohoto intervalu náhodně zvolíme 2^{j+1} množin $f^{(P)}$. Potom tedy minimálně pro jednu trojici (m^*, i^*, j^*) očekáváme, že bude platit

$$\sum_{f^{(P)}}\left|X_{m^*,f^{(P)}}\right| \geq 2^{j^*} 2^{i^*} L_N\left[\frac{1}{3}, \tau+\rho\right].$$

To, že nenajdeme ani jednu takovou trojici, opět nastane s nějakou konstantní pravděpodobností.

Protože vybíráme z $L_N\left[\frac{1}{3}, 2\epsilon\right]$ dvojic (a, b), tak očekávaná hodnota počtu polynomů $f \in f^{(P)}$, pro které jeden náhodně vybraný pár (a, b) leží v $X_{m^*,f}$, je

$$2^{j^*} 2^{i^*} L_N\left[\frac{1}{3}, \tau+\rho-2\epsilon\right].$$

Abychom tedy mohli očekávat $L_N\left[\frac{1}{3}, \max(\beta, \gamma+\rho)\right]$ nalezených hladkých relací, potřebujeme otestovat $2^{-j^*}2^{-i^*} L_N\left[\frac{1}{3}, \max(\beta, \gamma+\rho)+2\epsilon-\tau-\rho\right]$ dvojic (a, b). I do třetice je pravděpodobnost selhání nějaká konstanta.

Shrňme si tedy celý postup:

1) pro každé $i \in \{0, \ldots, \lceil \log_2 K \rceil\}$, kde $K \leq L_N\left[\frac{1}{3}, 2\epsilon - \tau\right]$, náhodně volíme 2^{i+1} hodnot m,

2) ke každému m vygenerujeme 2^{j+1} množin $f^{(P)}$, kde $j \in \{0, \ldots, \lceil \log_2 K' \rceil\}$, přičemž $K' \leq L_N\left[\frac{1}{3}, 2\epsilon - \tau - \rho\right]$,

3) náhodně zvolíme $2^{-j}2^{-i} L_N\left[\frac{1}{3}, \max(\beta, \gamma+\rho)+2\epsilon-\tau-\rho\right]$ dvojic (a, b) pro každou dvojici $\left(m, f^{(P)}\right)$ a otestujeme $a - bm$ hladkost,

4) o každé ze zbylých $2^{-j}2^{-i} L_N\left[\frac{1}{3}, \max(\beta, \gamma+\rho)+\frac{\epsilon\delta+\kappa}{3\gamma}-\rho\right]$ dvojic (pravděpodobnost že $a - bm$ je B_1-hladké je $L_N\left[\frac{1}{3}, -\frac{1}{3\delta\beta}\right]$) rozhodneme, zda patří do $X_{m,f}$ pro všechny $f \in f^{(P)}$.

Pravděpodobnost, že nedostaneme $L_N\left[\frac{1}{3}, \max(\beta, \gamma+\rho)\right]$ hladkých relací, je nějaká konstanta. Jak už víme z RNFS, logaritmickým počtem $(L_N\left[\frac{1}{3}, o(1)\right])$ opakování dokážeme pravděpodobnost selhání srazit pod $L_N\left[\frac{2}{3}, \kappa - \delta^{-1}\right]$.

Nyní už není problém ani spočítat časovou náročnost tohoto procesu:

- otestování $a - mb$ hladkosti jedné dvojice si vyžaduje $L_N\left[\frac{1}{3}, o\left(1\right)\right]$ času, přičemž pro každé $i \in \{0, \ldots, \lceil \log_2 K \rceil\}$, $j \in \{0, \ldots, \lceil \log_2 K' \rceil\}$ jich máme $2^{i+1}2^{j+1}2^{-i}2^{-j}L_N\left[\frac{1}{3}, \max\left(\beta, \gamma+\rho\right) + 2\epsilon - \tau - \rho\right]$,
- následné otestování náležení do $X_{m,f}$ pro každou čtveřici (m, f, a, b), kde $a - bm$ je hladké a $f \in f^{(P)}$: $PL_N\left[\frac{1}{3}, o\left(1\right)\right] = L_N\left[\frac{1}{3}, \rho\right]$,
- čtveřic $\left(m, f^{(P)}, a, b\right)$ pro každé $i \in \{0, \ldots, \lceil \log_2 K \rceil\}$, $j \in \{0, \ldots, \lceil \log_2 K' \rceil\}$ máme $2^{i+1}2^{j+1}2^{-i}2^{-j}L_N\left[\frac{1}{3}, \max\left(\beta, \gamma+\rho\right) + \frac{\epsilon\delta+\kappa}{3\gamma} - \rho\right]$,
- $\lceil \log_2 K \rceil, \lceil \log_2 K' \rceil = L_N\left[\frac{1}{3}, o\left(1\right)\right]$.

Tudíž celková složitost stochastického prohlubování, neboli průměrný čas potřebný k nalezení požadovaného počtu hladký relací, je

$$L_N\left[\frac{1}{3}, \max\left(\max\left(\beta, \gamma+\rho\right) + \frac{1}{3\delta\beta} + \frac{\epsilon\delta+\kappa}{3\gamma} - \rho, \max\left(\beta, \gamma+\rho\right) + \frac{\epsilon\delta+\kappa}{3\gamma}\right)\right],$$

což je rovno

$$L_N\left[\frac{1}{3}, \max\left(\beta, \gamma+\rho\right) + \frac{\epsilon\delta+\kappa}{3\gamma} + \max\left(\frac{1}{3\delta\beta} - \rho, 0\right)\right] =: L_N\left[\frac{1}{3}, T_1\right].$$

Druhou část randomizovaného číselného síta s více polynomy, tj. převod relací na kongruenci čtverců, dokážeme s pravděpodobností alespoň $1 - L_N\left[\frac{2}{3}, \frac{\kappa-\delta^{-1}}{3}\right]$ zvládnout v průměrném čase nanejvýš

$$L_N\left[\frac{1}{3}, 2\max\left(\beta, \gamma+\rho, \frac{2\delta}{3}\right)\right] =: L_N\left[\frac{1}{3}, T_2\right].$$

(ekvivalent Věty 2.8).

Zbývá už jenom optimalizovat parametry tak, aby byl součet těchto dvou částkových složitostí co nejmenší. Jedná se tedy o optimalizační problém tvaru

$$\begin{aligned}
\text{minimalizuj} \quad & \max\left(T_1, T_2\right) \\
\text{vzhledem k} \quad & \delta^{-1} < \kappa \\
& \delta^{-1} < \frac{\epsilon\delta+\kappa}{2} \\
& 2\epsilon - \frac{1}{3\delta\beta} - \frac{\epsilon\delta+\kappa}{3\gamma} + \rho \geq \max\left(\beta, \gamma+\rho\right) \\
& \beta, \gamma, \delta, \epsilon, \kappa, \rho \geq 0,
\end{aligned}$$

jehož limitním optimálním řešením je

$$\begin{aligned}
\beta = \epsilon = \left(\frac{46+13\sqrt{13}}{108}\right)^{1/3} &\approx 0.95094, \\
\gamma = \frac{1}{3}\left(2\left(4+\sqrt{13}\right)\right)^{1/3} &\approx 0.82591, \\
\delta = \frac{\left(2\left(16-\sqrt{13}\right)\right)^{1/3}}{3^{2/3}} &\approx 1.40175, \\
\kappa = \delta^{-1} &\approx 0.71339 \\
\rho = \beta - \gamma &\approx 0.12503.
\end{aligned}$$

Randomizované číselné síto s více polynomy tedy s pravděpodobností alespoň

$$1 - L_N\left[\frac{2}{3}, \kappa - \delta^{-1}\right]$$

nalezne kongruenci čtverců modulo N v průměrném čase

$$L_N\left[\frac{1}{3}, \left(\frac{92 + 26\sqrt{13}}{27}\right)^{1/3}\right] \approx L_N\left[\frac{1}{3}, 1.90188\right],$$

což je stejný výsledek jako v klasické verzi číselného síta s více polynomy.

Závěr

Postupně jsme si předvedli výpočet časové složitosti základních verzí a různých modifikací kvadratického a číselného síta. U prvního zmíněného algoritmu žádná z představených variant nepřinesla asymptotické zlepšení. Vysvětlili jsme si však, že i přesto v praxi přináší značné výhody a staly se proto spíše jeho standardní součástí.

Naproti tomu se ukázalo, že situace v číselném sítu je přesně opačná. Viděli jsme několik modifikací, které snížili asymptotickou časovou složitost. Lehký pokles díky opakovanému použití některých výpočtů přinesla verze s více polynomy. Ten ovšem nebyl dostatečný na to, aby v řádech současně faktorizovaných čísel převážil složitost přidaných částí, které se asymptoticky neprojevily. O výraznější zlepšení se postaralo zahrnutí části prosívací fáze do předvýpočtu. Tady pak ale do hry vstoupily praktické problémy spojené s ukládáním velkého množství dat v paměti. Naše verze s více počítači zas přinesla pouze zlepšení, které nemělo vliv na časově nejnáročnější část algoritmu, tudíž ani na celkovou asymptotickou složitost.

Zjistili jsme tedy, že menší asymptotická složitost nemusí nutně znamenat časovou úsporu v praxi. Ta se naopak může objevit u dvou teoreticky časově rovnocenných algoritmů. Snažili jsme se proto při analýze a porovnávaní jednotlivých schémat zohlednit obě hlediska, aby byly naše výsledky reálně aplikovatelné.

Úplně jiný typ modifikace nabídlo randomizované číselné síto (případně jeho navrhovaná verze s více polynomy). To přeneslo analýzu nejhoršího případu (neboli worst-case scenario) na průměrný, díky čemu bylo možné určit časovou složitost na základě rigorózních výpočtů. Dostali jsme tedy mnohem přesvědčivější důkaz, že faktorizace proběhne v uvedených časech.

Seznam použité literatury

Brillhart, J., Filaseta, M. a Odlyzko, A. (1981). On an irreducibility theorem of a. cohn. *Canadian Journal of Mathematics*, **33**(5), 1055–1059.

Buhler, J. P., Lenstra, H. W. a Pomerance, C. (1993). Factoring integers with the number field sieve. In *The development of the number field sieve*, pages 50–94. Springer.

Canfield, E. R., Erdös, P. a Pomerance, C. (1983). On a problem of oppenheim concerning "factorisatio numerorum". *Journal of number theory*, **17**(1), 1–28.

Conrad, K. (2016). Stirling's formula. URL `https://kconrad.math.uconn.edu/blurbs/analysis/stirling.pdf`.

Coppersmith, D. (1993). Modifications to the number field sieve. *Journal of Cryptology*, **6**(3), 169–180.

Crandall, R. a Pomerance, C. (2001). *Prime numbers*. Springer.

Lee, J. D. a Venkatesan, R. (2018). Rigorous analysis of a randomised number field sieve. *Journal of Number Theory*, **187**, 92–159.

Lenstra Jr, H. W. (1987). Factoring integers with elliptic curves. *Annals of mathematics*, pages 649–673.

Pejlová, A. (2016). Generování polynomů pro číselné síto. Master's thesis, Univerzita Karlova v Praze.

Pesiri, A. (2007). The chebotarëv density theorem applications. URL `http://www.mat.uniroma3.it/users/pappa/sintesi/16_Pesiri.pdf`.

Pomerance, C. (1984). The quadratic sieve factoring algorithm. In *Workshop on the Theory and Application of of Cryptographic Techniques*, pages 169–182. Springer.

Rosser, J. B. a Schoenfeld, L. (1962). Approximate formulas for some functions of prime numbers. *Illinois Journal of Mathematics*, **6**(1), 64–94.

Schnorr, C.-P. (1982). Refined analysis and improvements on some factoring algorithms. *Journal of Algorithms*, **3**(2), 101–127.

Wiedemann, D. (1986). Solving sparse linear equations over finite fields. *IEEE transactions on information theory*, **32**(1), 54–62.