Souvislost dobrých aproximací s řetězovými zlomky

Předpokládejme, že $\theta \in \mathbb{R} - \frac{1}{2}\mathbb{Z}$ a definujme celá čísla p_n, q_n pro $n \geq 0$ a a_n pro $n \geq 1$ jako ve větě o dobrých aproximacích. Pro každé $n \geq 1$ ještě označme

$$\theta_n = \frac{|q_n\theta - p_n|}{|q_{n-1}\theta - p_{n-1}|}.$$

Rekurentní vztahy z věty zaručují, že pro každé $n \geq 1$ platí

$$|q_{n-1}\theta - p_{n-1}| = a_n|q_n\theta - p_n| + |q_{n+1}\theta - p_{n+1}|,$$

a tedy $\theta_n^{-1} = a_n + \theta_{n+1}$, odkud

$$\theta_n = \frac{1}{a_n + \theta_{n+1}},$$

což spolu s $\theta_1 = \langle\theta\rangle$ dává

$$\theta = [\theta] + \theta_1 = [\theta] + \frac{1}{a_1 + \theta_2} = [\theta] + \frac{1}{a_1 + \frac{1}{a_2 + \theta_3}} = [\theta] + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \theta_4}}}$$

$$= [\theta] + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \theta_5}}}} = [\theta] + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \frac{1}{a_5 + \theta_6}}}}} = \dots$$

Příklad: spočítejme několik dobrých aproximací čísla $\sqrt{15}$

Je tedy $p_0 = 1$, $q_0 = 0$, $p_1 = [\sqrt{15}] = 3$, $q_1 = 1$ a platí

$$\theta_1^{-1} = \frac{1}{\sqrt{15}-3} = \frac{\sqrt{15}+3}{15-9} = 1 + \frac{\sqrt{15}-3}{6}, \qquad a_1 = 1,$$

$$\theta_2^{-1} = \frac{6}{\sqrt{15}-3} = \frac{6(\sqrt{15}+3)}{15-9} = 6 + (\sqrt{15} - 3), \qquad a_2 = 6,$$

$$\theta_3^{-1} = \theta_1^{-1}, \qquad a_3 = a_1 = 1, \qquad a_4 = a_2 = 6, \quad \text{atd.}$$

Posloupnost a_n
je periodická. Výpočet čísel p_n, q_n je výhodné uspořádat do tabulky:

n	0	1	2	3	4	5	6	7	8	9	10
p_n	1	3	4	27	31	213	244	1677	1921	13203	15124
q_n	0	1	1	7	8	55	63	433	496	3409	3905
a_n		1	6	1	6	1	6	1	6	1	6

Protože $a_1 = 1$, dostáváme dobré aproximace čísla $\sqrt{15}$ až pro $n \geq 2$, jsou to čísla

$$4, \ \frac{27}{7}, \ \frac{31}{8}, \ \frac{213}{55}, \ \frac{244}{63}, \ \frac{1677}{433}, \ \frac{1921}{496}, \ \frac{13203}{3409}, \ \frac{15124}{3905}, \ \ldots$$

Další moderní metody hledání netriviálního dělitele

Nejúčinnější metody:

- Lenstrova metoda eliptických křivek,
- metoda kvadratického síta,
- metoda síta v číselném tělese.

Základní myšlenka kvadratického síta i síta v číselném tělese je stejná jako základní myšlenka metody řetězových zlomků, která je historicky první metodou subexponenciálního času a byla na konci 60-tých let a v 70-tých letech hlavní používanou metodou.

Nechť N je (velké) složené přirozené číslo, které není dělitelné žádnými „malými“ prvočísly (tj. prvočísly $\leq B$) a které není mocninou prvočísla. Hledáme netriviálního dělitele čísla N.

Budeme hledat x, $y \in \mathbb{Z}$, aby platilo

$$x^2 \equiv y^2 \pmod{N} \quad \text{a přitom} \quad x \not\equiv \pm y \pmod{N}.$$

Protože $x^2 - y^2 = (x - y)(x + y)$, je jasné, že pak největší společný dělitel $(x + y, N)$ bude netriviální dělitel čísla N.

Jak hledat $x, y \in \mathbb{Z}$, $x^2 \equiv y^2 \pmod N$, $x \not\equiv \pm y \pmod N$

Hledáme kongruence tvaru

$$x_k^2 \equiv (-1)^{e_{0k}} p_1^{e_{1k}} p_2^{e_{2k}} \cdots p_m^{e_{mk}} \pmod N,$$

kde p_i jsou „malá“ prvočísla a $e_{ik} \in \{0, 1\}$. Nalezneme-li dostatečně mnoho takových kongruencí (tj. alespoň $n \geq m + 2$), můžeme Gaussovou eliminací nad $\mathbb{F}_2$ v $m+1$-rozměrném prostoru $\mathbb{F}_2^{m+1}$ najít lineární závislost mezi n vektory $(e_{0k}, e_{1k}, \ldots, e_{mk})$, (ztotožňujeme $\{0, 1\}$ s $\mathbb{F}_2$), tj. najít $\varepsilon_1, \ldots, \varepsilon_n \in \mathbb{F}_2$, ne všechna nulová, pro která je $\sum_{k=1}^n \varepsilon_k (e_{0k}, e_{1k}, \ldots, e_{mk})$ nulový vektor. Budeme-li nyní $\varepsilon_1, \ldots, \varepsilon_n$ považovat za celá čísla, pak pro každé $i \in \{0, 1, \ldots, m\}$ je číslo $v_i = \frac{1}{2} \sum_{k=1}^n \varepsilon_k e_{ik} \in \mathbb{Z}$, protože $\sum_{k=1}^n \varepsilon_k e_{ik}$ leží v jádře homomorfismu okruhů $\mathbb{Z} \to \mathbb{F}_2$.
Pak pro $x = \prod_{k=1}^n x_k^{\varepsilon_k}$, $y = p_1^{v_1} p_2^{v_2} \cdots p_m^{v_m}$, platí

$$x^2 = \prod_{k=1}^n x_k^{2\varepsilon_k} \equiv \prod_{k=1}^n \left((-1)^{e_{0k}} p_1^{e_{1k}} p_2^{e_{2k}} \cdots p_m^{e_{mk}}\right)^{\varepsilon_k} = y^2 \pmod N,$$

což nám dá netriviálního dělitele čísla N, pokud $x \not\equiv y \pmod N$.

Jak hledat $x,\ y \in \mathbb{Z},\ x^2 \equiv y^2 \pmod N,\ \ x \not\equiv \pm y \pmod N$

V případě, že liché N je dělitelné právě r prvočísly, je pravděpodobnost, že nastane $x \equiv \pm y \pmod N$ za předpokladu, že platí $x^2 \equiv y^2 \pmod N$ a $(N, xy) = 1$, rovna 2^{1-r}. Proto je vhodné volit n o něco větší než $m + 2$, abychom Gaussovou eliminací našli více závislostí.

Množina $\{p_1, \ldots, p_m\}$ se nazývá báze faktorizace. Způsoby, jak ji zvolit optimálně a jak hledat potřebné kongruence

$$x_k^2 \equiv (-1)^{e_{0k}} p_1^{e_{1k}} p_2^{e_{2k}} \cdots p_m^{e_{mk}} \pmod N,$$

se u jednotlivých metod liší. Vždy však je mezi exponenty na pravé straně kongruence jen několik jedniček.

Matice takové soustavy má tedy v každém řádku jen několik jedniček a zbytek tvoří nuly. Uložit celou tuto obrovskou „řídkou“ matici do paměti se nám patrně nepodaří. Proto je třeba Gaussovu eliminaci provádět jinak, než u malých matic.

Gaussova eliminace „řídké“ matice

Nemáme uloženou celou matici, ale pro každý řádek máme uloženy jen informace o poloze jedniček v tomto řádku.
Při provádění eliminace se rozlišuje mezi „řídkými“ a „hustými“ sloupci: hodnoty v „hustých“ sloupcích se neuchovávají, místo nich se uchovává pro každý řádek informace o tom, jak byl odvozen z původní matice (tj. kterých řádků původní matice je součtem).
Eliminace se provádí tak, že hledáme řádek, který má pouze jednu jedničku v „řídkých“ sloupcích. Ten pak přičteme ke všem řádkům, které v tomto sloupci mají jedničku. Poté už tento řádek nebudeme potřebovat. V případě, že žádný řádek, který by měl pouze jednu jedničku v „řídkých“ sloupcích, neexistuje, vybereme ten, který má jedniček co nejméně. Vybereme v něm jednu jedničku a sloupce, ve kterých jsou ostatní jedničky tohoto řádku, prohlásíme za husté. Skončíme v okamžiku, kdy už nemáme žádný řídký sloupec. Pomocí informací o odvozování řádků nyní sestavíme mnohem menší „hustou“ matici, v níž se provede Gaussova eliminace obvyklým způsobem.

Metoda řetězových zlomků

Potřebujeme hledat kongruence tvaru

$$x_k^2 \equiv (-1)^{e_{0k}} p_1^{e_{1k}} p_2^{e_{2k}} \cdots p_m^{e_{mk}} \pmod{N},$$

kde p_i jsou pevně zvolená prvočísla a $e_{ik} \in \{0, 1\}$.

Budeme vycházet z toho, že pokud zvolíme do naší báze faktorizace všechna prvočísla $p_1, \ldots, p_m$ menší než nějaká hranice a najdeme-li kongruenci $x^2 \equiv t \pmod{N}$ s „malým“ $|t|$, je reálná šance, že v rozkladu čísla $|t|$ se nevyskytují jiná prvočísla než $p_1, \ldots, p_m$ a tedy že získáme kongruenci požadovaného tvaru.

Metoda řetězových zlomků - základní myšlenka

Nechť $\frac{p}{q}$ je dobrá aproximace čísla $\sqrt{kN}$, kde k je nějaké nepříliš velké přirozené číslo nedělitelné druhou mocninou prvočísla. Pak

$$\left|\sqrt{kN} - \frac{p}{q}\right| < \frac{1}{q^2}.$$

Označme $t = p^2 - kNq^2$. Pak $p^2 \equiv t \pmod{N}$. Nalezněme odhad pro $|t|$. Pak

$$-\frac{1}{q} < \sqrt{p^2 - t} - p < \frac{1}{q}.$$

Přičtením p, umocněním a odečtením p^2 dostaneme

$$-\frac{2p}{q} + \frac{1}{q^2} < -t < \frac{2p}{q} + \frac{1}{q^2},$$

odkud vzhledem k $\sqrt{kN} > \frac{p}{q} - \frac{1}{q^2}$ plyne

$$|t| < \frac{2p}{q} + \frac{1}{q^2} < 2\sqrt{kN} + \frac{3}{q^2}.$$

Číslo $|t|$ tedy opravdu není „velké“ a šance na získání užitečné kongruence hledaného tvaru je.

Metoda řetězových zlomků - postup

Metoda řetězových zlomků tedy dává následující algoritmus: postupně za k volíme přirozená čísla nedělitelná druhou mocninou prvočísla a pro každé takové k počítáme jistý počet dobrých aproximací $\frac{p}{q}$. Pro každou dobrou aproximaci zkoušíme rozložit číslo $|t| = |p^2 - kNq^2|$ pomocí prvočísel z báze faktorizace. Jestliže se to podaří, získáme kongruenci požadovaného tvaru.

Pokud $|t|$ není možné rozložit pomocí prvočísel z báze faktorizace, avšak platí $|t| = F \cdot U$, kde F se pomocí prvočísel z báze faktorizace rozkládá a U je (asi) prvočíslo podle testu Millera a Rabina, je vhodné uložit i trojici (p, t, U). Získáme-li totiž později ještě jinou trojici (p', t', U) se stejným U, pak z $p^2 \equiv t \pmod{N}$ a $(p')^2 \equiv t' \pmod{N}$ získáme kongruenci požadovaného tvaru $x^2 \equiv \frac{tt'}{U^2} \pmod{N}$, kde x je řešení kongruence $Ux \equiv pp' \pmod{N}$.

Lepší metoda: metoda kvadratického síta

Jiným způsobem budeme opět hledat kongruence tvaru

$$x_k^2 \equiv (-1)^{e_{0k}} p_1^{e_{1k}} p_2^{e_{2k}} \cdots p_m^{e_{mk}} \pmod{N},$$

kde p_i jsou pevně zvolená prvočísla a $e_{ik} \in \{0, 1\}$.
Označme $d = [\sqrt{N}]$ a uvažme kvadratický polynom

$$Q(x) = (x + d)^2 - N.$$

Je jasné, že $Q(a) \equiv (a + d)^2 \pmod{N}$ a že $|Q(a)|$ nebude „velké“ pro celá čísla a s „malou“ absolutní hodnotou. Ačkoli je to jednodušší metoda generování „malých“ kvadrátů modulo N než metoda řetězových zlomků, zatím není příliš zajímavá. Rozhodující důvod, proč je tato metoda rychlejší než metoda řetězových zlomků, je tento: není nutné rozkládat „malé“ kvadráty modulo N. Vzhledem k tomu, že většinu z nich rozložit nad zvolenou bází faktorizace nelze, znamená toto marné rozkládání plýtvání časem.

Metoda kvadratického síta - postup prosívání

Předpokládejme, že pro nějaké $n \in \mathbb{N}$ víme, že $n \mid Q(a)$. Pak ovšem pro každé $k \in \mathbb{Z}$ platí $n \mid Q(a + kn)$. Hledat takové a znamená řešit kongruenci $x^2 \equiv N \pmod{n}$ a vzít $a = x - d$. Přitom řešení této kongruence pro malé n není tak obtížné (pro prvočíslo n existuje Shanksův algoritmus časové náročnosti $O(\ln^4 n)$).

Jak budeme čísla prosívat: pro každé celé číslo a z velmi dlouhého intervalu uložíme do vektoru indexovaného a přibližnou hodnotu $\log_2 |Q(a)|$ (stačí $\frac{1}{2}$ plus řád první jedničky binárního zápisu, pak je chyba menší než $\frac{1}{2}$).

Pak pro všechny mocniny prvočísel $p^k \leq B$ pro zvolené B odečteme $\log_2 p$ od všech prvků v našem vektoru, jejichž index a je kongruentní modulo p^k s předem vypočteným řešením kongruence $Q(a) \equiv 0 \pmod{p^k}$, tj. $(a + d)^2 \equiv N \pmod{p^k}$. Protože předpokládáme, že $p \nmid N$, má pro lichá p tato kongruence dvě řešení, je-li N kvadratický zbytek modulo p, a žádné, jestliže je N kvadratický nezbytek modulo p – do báze faktorizace tedy dáváme kromě 2 jen ta prvočísla, pro která je N kvadratický zbytek.

Metoda kvadratického síta - vyhodnocení prosívání

Po ukončení prosívání zjistíme, pro která a není $Q(a)$ dělitelné mocninou prvočísla větší než B. Pro tato a je totiž prvek ve vektoru indexovaný a malý (kdyby logaritmy byly přesné, byla by to nula). V opačném případě zde musí být číslo větší než $\log_2 B$ (odhlédneme-li od nepřesnosti logaritmů).

Odhadněme potřebnou přesnost ε výpočtu $\log_2 p$. Označme k největší číslo ve vektoru před započetím prosívání. Pak každé číslo $|Q(a)|$ má nejvýše k činitelů. Je-li $Q(a)$ rozložitelné pomocí naší báze faktorizace, je po provedení odčítání logaritmů ve vektoru s indexem a číslo menší než $\frac{1}{2} + k\varepsilon$. Naproti tomu pro nerozložitelné $Q(a)$ dostaneme číslo větší než $(\log_2 B) - \frac{1}{2} - (k + \frac{1}{2} - \log_2 B)\varepsilon$. Stačí tedy $\varepsilon < \frac{-1+\log_2 B}{2k+\frac{1}{2}-\log_2 B}$.

Pak pro všechna a, pro které jsme dostali ve vektoru číslo menší než $\frac{1}{2} + k\varepsilon$, spočítáme znovu $Q(a)$ a rozložíme, čímž získáme kongruenci požadovaného tvaru. Máme-li dost místa v paměti, ukládáme v průběhu prosívání u každé položky a několik největších prvočísel, jejichž logaritmy odčítáme, což pak urychlí rozkládání.

Metoda kvadratického síta - možnosti vylepšení

Podobně jako u metody řetězových zlomků i v tomto případě můžeme hledat kongruence $x^2 \equiv F \cdot U \pmod{N}$, kde F se pomocí prvočísel z báze faktorizace rozkládá a U je „nepříliš velké“ číslo. V tom případě rozkládáme $Q(a)$ pro všechna a, pro které po prosívání zůstalo ve vektoru číslo menší než nějaká předem daná mez a nerozložitelný faktor spolu s a uchováváme pro případ, že by se týž faktor objevil ještě jednou.

Nevýhodou je, že na dlouhém intervalu prosívání hodnoty polynomu $Q(x)$ značně rostou a s tím i klesají naše šance na úspěšné rozložení. Mohli bychom proto vzít ještě další polynom a prosívat i jeho hodnoty, například $Q(x) = (x + [\sqrt{\ell N}])^2 - \ell N$ pro nějaké přirozené číslo ℓ nedělitelné druhou mocninou prvočísla. V tom případě bychom však museli doplnit naši bázi faktorizace: máme v ní pouze ta prvočísla p, pro která je N kvadratický zbytek modulo p, kdežto nyní potřebujeme ta, pro která je ℓN kvadratický zbytek modulo p. Ovšem zvětšení báze faktorizace znamená potřebu více kongruencí a také Gaussovu eliminaci větší matice.