Algoritmy teorie čísel

Radan Kučera, jarní semestr 2016

Literatura: text v ISu čerpající z následujících zdrojů

1. Cassels J. W. S.: *An Introduction to Diophantine Approximation*, University Press, Cambridge, 1965.
2. Cohen H.: *A Course in Computational Algebraic Number Theory*, Graduate Texts in Mathematics 138, Springer–Verlag, Berlin, Heidelberg, New York 1993, kapitoly 8–10, (čtvrté aktualizované výdání 2000).
3. Dietzfelbinger M., *Primality Testing in Polynomial Time (From Randomized Algorithms to "Primes is in P")*, LNCS 3000, Springer–Verlag Berlin, Heidelberg, New York 2004.
4. Knuth D.E., *The Art of Computer Programming, díl 2: Seminumerical Algorithms*, (druhé vydání), Addison-Wesley, Reading, Mass., 1981.
5. Lenstra A. K., Lenstra H. W. Jr.: *Algorithms in Number Theory*, v *Handbook of Theoretical Computer Science*, kapitola 12, Elsevier Science Publishers B.V., 1990.
6. Rosický J., Algebra, 4. vydání, skriptum MU, 2002.

Pojem algoritmu

Algoritmus je metoda, která pro jistý typ vstupů dá po konečné době výstup, tedy odpověď na zadaný problém.

Při zadání algoritmu je nutno provést:

- dokázat správnost výstupu,
- odhadnout časovou náročnost,
- odhadnout paměťovou náročnost.

Časovou náročností rozumíme závislost délky výpočtu na délce vstupu, přitom délku vstupu měříme počtem bitů potřebných pro zápis zadání a délkou výpočtu rozumíme, jak dlouho trvá nejdelší výpočet pro danou délku vstupu.

Příklad. Pro vstup jednoho přirozeného čísla N je třeba $1 + [\log_2 N]$ bitů.

Paměťovou náročnost budeme také měřit v bitech (u většiny algoritmů, se kterými se setkáme, nebude nutné se jí zabývat, neboť bude konstantní a zanedbatelně malá).

Asymptotický odhad

Nechť $(a_n)_{n=1}^{\infty}$, $(b_n)_{n=1}^{\infty}$ jsou posloupnosti. Řekneme, že posloupnost $(a_n)_{n=1}^{\infty}$ je řádu $O(b_n)$, jestliže platí

$$\limsup_{n\to\infty} \left|\frac{a_n}{b_n}\right| < \infty.$$

Řekneme, že posloupnost $(a_n)_{n=1}^{\infty}$ je řádu $o(b_n)$, jestliže existuje

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

Protože se budeme zabývat algoritmy, jejichž cílem je rozložit přirozené číslo na prvočinitele, bude vstupem pro náš algoritmus jediné přirozené číslo N.

Dohodněme se, že $\ln^k N$ bude znamenat $(\ln N)^k$. Kupříkladu tedy $\ln^2 N = (\ln N)^2$, nikoli $\ln \ln N$.

Definice. Řekneme, že algoritmus je polynomiálního času, jestliže čas, po který algoritmus poběží, najde-li na vstupu přirozené číslo N, je řádu $o(\ln^k N)$ pro nějaké přirozené číslo k.

Řekneme, že algoritmus je lineárního času, je-li tento čas řádu $O(\ln N)$. Řekneme, že je kvadratického času, je-li tento čas řádu $O(\ln^2 N)$, ale není lineárního času. Řekneme, že je kubického času, je-li tento čas řádu $O(\ln^3 N)$, ale není kvadratického času.

Je-li tento čas řádu $o(N^\alpha)$ pro každé kladné reálné číslo α a přitom algoritmus není polynomiálního času, řekneme, že algoritmus je subexponenciálního času.

Konečně, jestliže existují kladná reálná čísla $\alpha > \beta$ tak, že tento čas je řádu $O(N^\alpha)$, ale není řádu $O(N^\beta)$, řekneme, že algoritmus je exponenciálního času.

Příklad. Později se setkáme s algoritmy, jejichž čas je řádu

$$O(e^{c(\ln N)^a(\ln\ln N)^b}),$$

kde a, b, c jsou kladná reálná čísla, přičemž $a + b = 1$. Tyto algoritmy jsou subexponenciálního času.

Pravděpodobnostní algoritmy

Budeme pracovat s algoritmy, jejichž průběh výpočtu závisí na jistém zdroji náhodných čísel. Je zde možnost (pravděpodobnosti nula), že jejich běh nikdy neskončí, přesto zkušenosti ukazují, že tyto algoritmy jsou často efektivnější než ostatní a mnohdy jsou jediné, které máme k dispozici.

Na druhé straně rozhodně nebudeme nazývat algoritmem metodu, produkující výsledek, který je s vysokou pravděpodobností správný. Je podstatné, že algoritmus v okamžiku zastavení dává pouze správné výsledky (odhlédneme-li od případných chyb člověka či počítače při provádění výpočtu).

Počítání s velkými čísly

Budeme předpokládat, že máme k dispozici software, ve kterém je možné provádět základní algebraické operace s čísly, majícími řekněme 1000 dekadických cifer (MATHEMATICA, MAPLE, PARI-GP a podobně).
Taková čísla jsou zapsána v poziční soustavě o vhodném základu a operace jsou prováděny podobně jako jsme to zvyklí dělat na papíře s dekadickými čísly. Vhodný základ je mocnina dvou: čas potřebný pro vstup a výstup je pouze zanedbatelná část celkové doby výpočtu a obvykle je dominován časem pro fyzický zápis.
Sčítání a **odčítání** má lineární časovou náročnost.
Násobení malou konstantou má také lineární časovou náročnost.
Obecné násobení a **dělení se zbytkem** má kvadratickou časovou náročnost. (Jsou známy algoritmy pro násobení a dělení $n$bitových čísel, které dosahují menší časové náročnosti než „metoda tužky a papíru“. Schönhage a Strassen popsali metodu jen s $O(n \cdot \ln n \cdot \ln \ln n)$ bitových operací. Jednu poměrně jednoduchou metodu násobení velkých čísel si vysvětlíme.)

Násobení velkých čísel rychleji než kvadraticky

$a \ll n = a \cdot 2^n$ značí číslo a posunuté o n pozic doleva,
$a \gg n = \lfloor a/2^n \rfloor$ označuje a posunuté o n pozic doprava,
$a \mathbin{\mathbb{V}} n$ je číslo tvořené posledními n bity čísla a, tj. zbytek po dělení čísla a číslem 2^n.

Daná dvě $2m$-bitová čísla x, y rozložme jako $x = a \ll m + b$,
$y = c \ll m + d$, kde $0 \leq a, b, c, d < 2^m$. Pak

$$x \cdot y = (a \ll m + b)(c \ll m + d) =$$
$$= (ac \ll 2m) + ((ad + bc) \ll m) + bd.$$

Součin $x \cdot y$ jsme tedy schopni spočítat pomocí čtyř součinů čtvrtinové velikosti (ac, ad, bc, bd) – to nám, zdá se, nepomohlo.

Ovšem využijeme-li identitu $(a + b)(c + d) = ac + ad + bc + bd$, dostáváme

$$x \cdot y = (ac \ll 2m) + (((a + b)(c + d) - (ac + bd)) \ll m) + bd.$$

Stačí nám tak spočítat tři součiny čtvrtinové velikosti (ac, bd, $(a + b)(c + d)$).
Tato myšlenka vede na jednoduchý rekurzivní algoritmus.

Násobení velkých čísel rychleji než kvadraticky

Předpokládali jsme, že počet bitů násobených čísel je vždy sudý. Proto předem doplníme oba činitele zleva nezbytným počtem nul tak, aby se jejich délka stala mocninou dvojky. Pak bude velikost násobených čísel vždy mocninou dvou, a tedy sudá až do doby, kdy máme vynásobit jednobitová čísla, což už zvládneme přímo.

Další komplikace: součty $a + b$ a $c + d$ mohou přetéct. Lepší než pokoušet se násobit tato (potenciálně větší) čísla tedy bude spočítat $((a + b) \mathbin{\Vdash} m) \cdot ((c + d) \mathbin{\Vdash} m)$ a případné přetečení vyřešit zvlášť: platí

$$(a + b) \mathbin{\Vdash} m = \begin{cases} a + b & \text{je-li } a + b < 2^m, \\ a + b - 2^m & \text{jinak,} \end{cases}$$

podobně pro $(c + d) \mathbin{\Vdash} m$.

Algoritmus $\mathrm{Mult}(x, y, k)$ na násobení 2^k-bitových čísel x, y

Algoritmus (Rychlejší násobení velkých čísel). *Pro daná nezáporná celá čísla* $x < 2^{2^k}$, $y < 2^{2^k}$ *algoritmus spočítá součin* $\mathrm{Mult}(x, y, k) = x \cdot y$.

1. *[Je* k *malé?] Je-li* $k = 0$, *pak vrať* $x \cdot y$ *a skonči.*
2. *[Vytvoř z daných čísel jejich horní a dolní část.] Polož* $m \leftarrow 2^{k-1}$, $a \leftarrow x \gg m$, $b \leftarrow x \mathbin{\Vdash} m$, $c \leftarrow y \gg m$, $d \leftarrow y \mathbin{\Vdash} m$. *Pak připrav* $a\_b \leftarrow a + b$, $c\_d \leftarrow c + d$.
3. *[Rekurentní aplikace algoritmu:] Získej* $ac \leftarrow \mathrm{Mult}(a, c, k-1)$, $bd \leftarrow \mathrm{Mult}(b, d, k-1)$, $r \leftarrow \mathrm{Mult}(a\_b \mathbin{\Vdash} m, c\_d \mathbin{\Vdash} m, k-1)$.
4. *[Vyřeš přetečení.] Pokud* $a\_b \geq 2^m$, *polož* $r \leftarrow r + c\_d \ll m$. *Pokud* $c\_d \geq 2^m$, *polož* $r \leftarrow r + a\_b \ll m$. *Pokud* $a\_b \geq 2^m$ *i* $c\_d \geq 2^m$, *polož* $r \leftarrow r - 2^{2m}$.
5. *[Dokonči výpočet.] Polož* $r \leftarrow r - (ac + bd)$, *pak vrať* $(ac \ll 2m) + (r \ll m) + bd$ *a skonči.*

Asymptotická časová náročnost

Nechť $T(k)$ označuje počet kroků, které náš algoritmus vykoná při násobení dvou 2^k-bitových čísel. Pro $k = 0$ algoritmus skončí okamžitě. V opačném případě vykoná několik konstantních operací ($\ll$, $\gg$, $\mathbb{V}$; i porovnání s 2^m), několik lineárních operací ($+$, $-$) a třikrát se rekurzivně zavoláme s parametrem k zmenšeným o 1. Ať už tedy „krokem“ rozumíme cokoli, určitě existuje konstanta α taková, že $T(0) \leq \alpha$ a $T(k) \leq \alpha \cdot 2^k + 3 \cdot T(k-1)$ pro $k \geq 1$. Po k-násobném dosazení tak dostaneme následující odhad shora:

$$T(k) \leq \alpha(2^k + 3 \cdot 2^{k-1} + 3^2 \cdot 2^{k-2} + \cdots + 3^k) = \alpha(3^{k+1} - 2^{k+1}) \leq 3\alpha \cdot 3^k,$$

takže $T \in O(3^k)$.
Vzhledem k velikosti n obou činitelů je tak náš algoritmus časové náročnosti v $O(3^{\log n})$, přičemž

$$3^{\log n} = (2^{\log 3})^{\log n} = 2^{\log 3 \cdot \log n} = 2^{\log n \cdot \log 3} = (2^{\log n})^{\log 3} = n^{\log 3},$$

$\log 3 \doteq 1{,}58$.

Praktické využití

Máme algoritmus, jehož asymptotická časová složitost je lepší než kvadratická, ale zatím jsme neřešili, co to znamená pro jeho aplikovatelnost. Praktické výsledky ukazují, že tento přístup se stává vhodnější až pro 16-ciferná čísla (v 2^{32}-ární soustavě, tedy čísla mající alespoň 155 dekadických cifer).
Následující tabulka ukazuje, jak dobře si některé algoritmy vedly při násobení velkých čísel (n je počet bitů obou činitelů). Použité algoritmy byly:

- A1: přímočará implementace klasického školského násobení bit po bitu
- A2: násobení bit po bitu s používáním libovolně vysokých číslic v mezivýsledcích
- A3: jako A1, ovšem v (použitému stroji nejpřirozenější) 2^{32}-ární soustavě
- A4: náš rychlý algoritmus v 2^{32}-ární soustavě
- A5: A4 pro aspoň 16-ciferná (512-bitová) čísla, A3 pro menší

Naměřená doba výpočtu

n	128	256	512	1024	2048
A1	.11 ms	.43 ms	1.7 ms	6.9 ms	27. ms
A2	44. μs	.17 ms	.67 ms	2.7 ms	11. ms
A3	.77 μs	1.5 μs	4.2 μs	14. μs	54. μs
A4	1.6 μs	4.1 μs	11. μs	34. μs	.10 ms
A5	.77 μs	1.5 μs	4.0 μs	12. μs	35. μs

n	4096	8192	16384	32768	65536
A1	.11 s	.45 s	1.8 s	7.1 s	29. s
A2	43. ms	.18 s	.71 s	2.9 s	11. s
A3	.21 ms	.85 ms	3.4 ms	13. ms	54. ms
A4	.31 ms	.92 ms	2.8 ms	8.4 ms	25. ms
A5	.11 ms	.33 ms	1.0 ms	3.0 ms	9.1 ms

Poměr dvou sousedních časů (u velkých vstupů, kde je již vliv dalších členů přesného vzorce časové složitosti nepatrný), je u prvních tří algoritmů přibližně $(2n)^2/n^2 = 2^2 = 4$, zatímco u posledních dvou $(2n)^{\log 3}/n^{\log 3} = 2^{\log 3} = 3$.

Výpočet největšího společného dělitele

Největšího společného dělitele dvou celých čísel a, b budeme značit (a, b). Z kontextu bude jistě vždy jasné, zda máme na mysli uspořádanou dvojici anebo největší společný dělitel.

Potřebě spočítat největší společný dělitel dvou daných přirozených čísel budeme čelit často.

Naivní řešení: rozlož obě čísla na součin prvočísel a poté vynásob společné činitele.

Tento postup je vhodný jen pro velmi malá čísla (řekněme do 100) nebo v případě, že víme, že některé z daných čísel je prvočíslo (pak stačí provést jen jedno dělení se zbytkem).

Mnohem výhodnější je výpočet největšího společného dělitele pomocí Euklidova algoritmu, který je nejen nejstarší, ale asi i nejdůležitější algoritmus teorie čísel.

Euklidův algoritmus

Algoritmus (Euklidův). *Pro daná nezáporná celá čísla a, b algoritmus najde jejich největší společný dělitel.*

1. *[Jsi hotov?] Je-li $b = 0$, pak vytiskni a jako odpověď a skonči.*
2. *[Euklidovský krok] Polož $r \leftarrow a \bmod b$, $a \leftarrow b$, $b \leftarrow r$ a jdi na 1.*

Věta. 1. Je-li $1 \leq a \leq N$, $1 \leq b \leq N$, pak počet Euklidovských kroků v předchozím algoritmu je roven nejvýše

$$\left\lceil \frac{\ln(\sqrt{5}N)}{\ln \frac{1+\sqrt{5}}{2}} \right\rceil - 2 \approx 2,078 \ln N + 1,672.$$

2. Průměrný počet Euklidovských kroků v předchozím algoritmu pro $a,\ b \in \{1, \ldots, N\}$ je roven přibližně

$$\frac{12 \ln 2}{\pi^2} \ln N + 0,14 \approx 0,843 \ln N + 0,14.$$

Časová náročnost Euklidova algoritmu

Podle věty je počet kroků algoritmu lineární v $\ln N$.
Každý krok vyžaduje dlouhé dělení, které je kvadratického času.
Proto se zdá, že je tento algoritmus kubického času.
V průběhu výpočtu jsou však a, b stále menší a menší, proto je možné průběžně snižovat potřebný počet cifer v poziční soustavě.
Při výpočtu Euklidovského kroku $a = bq + r$ je časová náročnost $O((\ln a)(1 + \ln q))$, tedy celkový čas je omezen řádem

$$O((\ln N)((\sum \ln q) + O(\ln N))).$$

Ale

$$\sum \ln q = \ln \prod q \leq \ln N,$$

a tedy při pečlivém naprogramování jde o algoritmus kvadratického času.

Binární verze Euklidova algoritmu

Místo dlouhého dělení je užito odčítání a dělení 2 (realizované posunem). Základem poziční soustavy musí být mocnina 2.

Algoritmus (Binární NSD). *Pro daná nezáporná celá čísla a, b algoritmus najde jejich největší společný dělitel.*

1. *[Jednou zredukuj velikost] Je-li $a < b$, vyměň a s b. Je-li $b = 0$, pak vytiskni a jako odpověď a skonči. Jinak (tj. pro $b \neq 0$) polož $r \leftarrow a \bmod b$, $a \leftarrow b$, $b \leftarrow r$.*
2. *[Spočítej mocninu 2] Je-li $b = 0$, pak vytiskni a jako odpověď a skonči. Jinak polož $k \leftarrow 0$ a dokud budou a i b sudá, opakuj $k \leftarrow k + 1$, $a \leftarrow a/2$, $b \leftarrow b/2$.*
3. *[Odstraň přebytečné mocniny 2] Je-li a sudé, opakuj $a \leftarrow a/2$, dokud bude a sudé. Jinak, je-li b sudé, opakuj $b \leftarrow b/2$, dokud bude b sudé.*
4. *[Odečti] (Nyní jsou obě a i b lichá.) Polož $t \leftarrow \frac{a-b}{2}$. Je-li $t = 0$, vytiskni $2^k \cdot a$ jako odpověď a skonči.*
5. *[Cyklus] Dokud bude t sudé, opakuj $t \leftarrow t/2$. Pak, je-li $t > 0$, polož $a \leftarrow t$, jinak polož $b \leftarrow -t$ a jdi na 4.*

Rozšířená verze Euklidova algoritmu

Označme d největší společný dělitel celých čísel a, b, pak existují celá čísla u, v tak, že $d = ua + vb$ (tzv. Bezoutova rovnost).
V některých aplikacích budeme potřebovat spočítat nejen d, ale i čísla u, v, proto si uvedeme algoritmus pro jejich výpočet.

Algoritmus (Rozšířený Euklidův). *Pro daná nezáporná celá čísla a, b algoritmus najde trojici celých čísel (u, v, d) takovou, že d je největší společný dělitel čísel a, b a platí $d = ua + vb$.*

1. *[Inicializace] Polož $u \leftarrow 1$, $d \leftarrow a$. Je-li $b = 0$, polož $v \leftarrow 0$, vytiskni (u, v, d) jako odpověď a skonči. Jinak polož $v_1 \leftarrow 0$ a $v_3 \leftarrow b$.*
2. *[Jsi hotov?] Je-li $v_3 = 0$, pak polož $v \leftarrow \frac{d-au}{b}$, vytiskni (u, v, d) jako odpověď a skonči.*
3. *[Euklidovský krok] Současně $q \leftarrow [\frac{d}{v_3}]$, $t_3 \leftarrow d \bmod v_3$. Pak polož $t_1 \leftarrow u - qv_1$, $u \leftarrow v_1$, $d \leftarrow v_3$, $v_1 \leftarrow t_1$, $v_3 \leftarrow t_3$ a jdi na 2.*

Hodnoty proměnných d, v_3, t_3 nezávisí na hodnotách ostatních proměnných. Přeznačíme-li je a, b, r, dostaneme původní Euklidův algoritmus, tedy tento algoritmus vždy skončí, a to se správným d.

Důkaz správnosti rozšířené verze Euklidova algoritmu

Zavedeme proměnné v_2, t_2, v, které nebudou nikdy použity pro výpočet hodnot původních proměnných. Před krokem 2 vždy platí

$$at_1 + bt_2 = t_3, \qquad au + bv = d, \qquad av_1 + bv_2 = v_3.$$

Algoritmus (Upravený rozšířený Euklidův). *Pro daná nezáporná celá čísla a, b algoritmus najde trojici celých čísel (u, v, d) takovou, že d je největší společný dělitel čísel a, b a platí $d = ua + vb$.*

1. *[Inicializace] Polož $u \leftarrow 1$, $d \leftarrow a$. Je-li $b = 0$, polož $v \leftarrow 0$, vytiskni (u, v, d) jako odpověď a skonči. Jinak polož $v_1 \leftarrow 0$, $v_3 \leftarrow b$, $t_1 \leftarrow 0$, $t_2 \leftarrow 0$, $t_3 \leftarrow 0$, $v \leftarrow 0$, $v_2 \leftarrow 1$.*
2. *[Jsi hotov?] Je-li $v_3 = 0$, pak polož $v \leftarrow \frac{d-au}{b}$, vytiskni (u, v, d) jako odpověď a skonči.*
3. *[Euklidovský krok] Současně $q \leftarrow [\frac{d}{v_3}]$, $t_3 \leftarrow d \bmod v_3$. Pak polož $t_1 \leftarrow u - qv_1$, $u \leftarrow v_1$, $d \leftarrow v_3$, $v_1 \leftarrow t_1$, $v_3 \leftarrow t_3$, $t_2 \leftarrow v - qv_2$, $v \leftarrow v_2$, $v_2 \leftarrow t_2$ a jdi na 2.*

Nezbytný aparát z algebry a elementární teorie čísel
Kongruence

Definice. Nechť $m \in \mathbb{N}$, $a, b \in \mathbb{Z}$. Řekneme, že a je kongruentní s b podle modulu m, píšeme $a \equiv b \pmod{m}$, jestliže $m \mid a - b$.

Věta 1. *Nechť $m \in \mathbb{N}$, $a, b, c, d \in \mathbb{Z}$. Jestliže $a \equiv b \pmod{m}$, $c \equiv d \pmod{m}$, pak platí $a + c \equiv b + d \pmod{m}$, $a - c \equiv b - d \pmod{m}$, $ac \equiv bd \pmod{m}$.*

Věta 2. *Nechť $m, k \in \mathbb{N}$, $a, b \in \mathbb{Z}$. Pak platí $a \equiv b \pmod{m}$ právě tehdy, když $ak \equiv bk \pmod{mk}$.*

Věta 3. *Nechť $m \in \mathbb{N}$, $a, b, k \in \mathbb{Z}$. Jestliže $ak \equiv bk \pmod{m}$ a navíc $(m, k) = 1$, pak platí $a \equiv b \pmod{m}$.*

Věta 4. *Nechť $a, b \in \mathbb{Z}$. Pak existuje $x \in \mathbb{Z}$ splňující kongruenci $ax \equiv b \pmod{m}$ právě tehdy, když $(a, m) | b$.*

Věta 5 (Čínská zbytková věta). *Nechť $m_1, m_2 \in \mathbb{N}$, $(m_1, m_2) = 1$. Pak pro libovolná $x_1, x_2 \in \mathbb{Z}$ existuje $x \in \mathbb{Z}$ splňující $x \equiv x_1 \pmod{m_1}$, $x \equiv x_2 \pmod{m_2}$.*

Okruh zbytkových tříd modulo m

Definice. Pro libovolné $m \in \mathbb{N}$ a libovolné $a \in \mathbb{Z}$ definujeme zbytkovou třídu modulo m obsahující číslo a předpisem

$$[a]_m = \{b \in \mathbb{Z};\ b \equiv a \pmod{m}\},$$

jde tedy o množinu všech celých čísel dávajících stejný zbytek po dělení číslem m jako číslo a. Množinu všech těchto tříd značíme

$$\mathbb{Z}/m\mathbb{Z} = \{[a]_m;\ a \in \mathbb{Z}\}.$$

Z věty 1 plyne, že na $\mathbb{Z}/m\mathbb{Z}$ lze definovat operace $+$ a $\cdot$ pomocí reprezentantů, tj.

$$[a]_m + [b]_m = [a+b]_m, \qquad [a]_m \cdot [b]_m = [a \cdot b]_m,$$

a že vůči těmto operacím tvoří $\mathbb{Z}/m\mathbb{Z}$ komutativní okruh o m prvcích, který nazýváme **okruh zbytkových tříd modulo** m.

Eulerova funkce φ

Definice. Je-li R okruh, označme $R^\times$ jeho (multiplikativní) grupu jednotek (neboli invertibilních prvků), tj.

$$R^\times = \{a \in R;\ \exists b \in R :\ ab = 1\},$$

kde 1 značí jedničku okruhu R. Charakteristika okruhu R je nejmenší $n \in \mathbb{N}$ splňující $n \cdot 1 = \underbrace{1 + 1 + \cdots + 1}_{n} = 0$ (tj. součet n kopií $1 \in R$ je roven $0 \in R$), pokud alespoň jedno takové n existuje. V opačném případě řekneme, že R je okruh charakteristiky nula.

Definice. Pro libovolné $m \in \mathbb{N}$ je $\varphi(m)$ definováno jako počet čísel z množiny $\{1, 2, \ldots, m\}$, která jsou nesoudělná s m. Tato funkce $\varphi : \mathbb{N} \to \mathbb{N}$ se nazývá Eulerova.

Příklad. Charakteristika okruhu $\mathbb{Z}/m\mathbb{Z}$ je m. Podle věty 4 platí

$$(\mathbb{Z}/m\mathbb{Z})^\times = \{[a]_m;\ a \in \mathbb{Z},\ (a, m) = 1\},$$

je tedy $\varphi(m) = |(\mathbb{Z}/m\mathbb{Z})^\times|$.

Vlastnosti Eulerovy funkce φ

Věta 6. *Pro libovolná $m_1, m_2 \in \mathbb{N}$ taková, že $(m_1, m_2) = 1$, platí $\varphi(m_1 m_2) = \varphi(m_1)\varphi(m_2)$.*

Důkaz. Zobrazení $\{1, 2, \ldots, m_1\} \times \{1, 2, \ldots, m_2\} \to \{1, 2, \ldots, m_1 m_2\}$ přiřadí dvojici $(x_1, x_2) \in \{1, 2, \ldots, m_1\} \times \{1, 2, \ldots, m_2\}$ číslo $y \in \{1, 2, \ldots, m_1 m_2\}$ splňující $y \equiv x_1 \pmod{m_1}$, $y \equiv x_2 \pmod{m_2}$ (číslo y je kongruentní s číslem x z věty 5 modulo $m_1 m_2$). Toto zobrazení je bijekce a platí $(y, m_1 m_2) = 1$ právě když $(x_1, m_1) = 1$ a $(x_2, m_2) = 1$.

Věta 7. *Pro libovolné $m \in \mathbb{N}$ platí*

$$\varphi(m) = m \prod_{p|m} \left(1 - \tfrac{1}{p}\right)$$

kde p probíhá v součinu všechna prvočísla dělící m.

Důkaz. Zřejmé, je-li m mocninou prvočísla. Pro obecné m plyne z věty 6 indukcí vzhledem k počtu prvočísel dělících m.

Lagrangeova, Eulerova a malá Fermatova věta

Definice. Nechť G je grupa, $a \in G$. Pokud neexistuje žádné $n \in \mathbb{N}$ s vlastností $a^n = 1$, řekneme, že řád prvku a je ∞. V opačném případě nejmenší $n \in \mathbb{N}$ s touto vlastností se nazývá řád prvku a. Naproti tomu řádem grupy G rozumíme počet $|G|$ jejích prvků (je-li konečná).

Věta 8 (Lagrangeova věta). Je-li G konečná grupa, pak řád libovolného prvku $a \in G$ je přirozené číslo, které je dělitelem řádu $|G|$ grupy G. Platí tedy $a^{|G|} = 1$.

Důsledek (Eulerova věta). Pro libovolná $m \in \mathbb{N}$, $a \in \mathbb{Z}$, taková, že $(a, m) = 1$, platí

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

Důsledek (malá Fermatova věta). Pro libovolné prvočíslo p a libovolné $a \in \mathbb{Z}$ nedělitelné p platí

$$a^{p-1} \equiv 1 \pmod{p}.$$

Řád čísla a modulo m

Věta 9. *Nechť jsou $m \in \mathbb{N}$, $a \in \mathbb{Z}$, taková, že $(a, m) = 1$. Označme*

$$e = \min\{n \in \mathbb{N};\ a^n \equiv 1 \pmod{m}\}.$$

Pak pro libovolná r, $s \in \mathbb{N} \cup \{0\}$ platí

$$a^r \equiv a^s \pmod{m}, \quad \textit{právě když} \quad r \equiv s \pmod{e}.$$

Důkaz. Lze předpokládat, že $r > s$. Vydělme $r - s$ číslem e se zbytkem: $r - s = qe + z$ pro q, $z \in \mathbb{Z}$, $0 \leq z < e$. Pak $a^{r-s} = (a^e)^q \cdot a^z \equiv a^z \pmod{m}$. Odtud plyne $a^{r-s} \equiv 1 \pmod{m}$, právě když $z = 0$.

Definice. Číslo e z předchozí věty se nazývá řád čísla a modulo m. Je to vlastně řád prvku $[a]_m$ v grupě $(\mathbb{Z}/m\mathbb{Z})^\times$.

Konečná tělesa

Věta 10. *Charakteristika konečného tělesa je prvočíslo.*

Věta 11. *Buď R konečné těleso charakteristiky p. Pak počet prvků tělesa R je mocninou prvočísla p.*

Věta 12. *Nechť p je prvočíslo a $n \in \mathbb{N}$. Pak existuje těleso o p^n prvcích.*

Věta 13. *Libovolná dvě konečná tělesa o stejném počtu prvků jsou izomorfní.*

Věta 14. *Buď R konečné těleso o p^n prvcích. Pak $R^\times$ je cyklická grupa o $p^n - 1$ prvcích. Každý prvek $r \in R$ je jednoduchým kořenem polynomu $x^{p^n} - x \in \mathbb{F}_p[x]$.*

Definice. Pro libovolné prvočíslo p a libovolné $n \in \mathbb{N}$ označme $\mathbb{F}_{p^n}$ těleso o p^n prvcích.

Poznámka. Pro libovolné prvočíslo p je $\mathbb{Z}/p\mathbb{Z}$ těleso, můžeme tedy položit $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. Avšak $\mathbb{Z}/p^n\mathbb{Z}$ je těleso pouze pro $n = 1$. Proto pro žádné $n > 1$ nejsou $\mathbb{F}_{p^n}$ a $\mathbb{Z}/p^n\mathbb{Z}$ izomorfní!

Konstrukce konečných těles $\mathbb{F}_{p^n}$ pro prvočíslo p a $n > 1$

Zvolíme libovolný normovaný ireducibilní polynom $h \in \mathbb{F}_p[x]$ stupně n. To, že takový polynom existuje pro každé prvočíslo p a každé přirozené číslo n, lze dokázat pomocí vět 12 a 14; to, že není podstatné, který z nich vybereme, plyne z věty 13. Pak $\mathbb{F}_{p^n}$ konstruujeme jako faktorokruh okruhu polynomů $\mathbb{F}_p[x]$ podle ideálu generovaného polynomem h. Prvky tohoto faktorokruhu jsou třídy rozkladu množiny polynomů $\mathbb{F}_p[x]$. V každé třídě leží právě jeden polynom stupně menšího než n. Proto, označíme-li α třídu obsahující polynom x, lze psát

$$\mathbb{F}_{p^n} = \{g(\alpha);\ g(x) \in \mathbb{F}_p[x], \operatorname{st} g < n\}.$$

V tělese $\mathbb{F}_{p^n}$ pak sčítáme a násobíme prvky jako polynomiální výrazy v α s tím, že po násobení musíme někdy eliminovat vyšší mocniny α. To děláme tak, že α^n vyjádříme pomocí nižších mocnin α využitím toho, že $h(\alpha) = 0$.

Grupa jednotek okruhu zbytkových tříd $(\mathbb{Z}/m\mathbb{Z})^\times$

Je-li p prvočíslo, je $\mathbb{Z}/p\mathbb{Z}$ těleso, a tedy podle věty 14 je $(\mathbb{Z}/p\mathbb{Z})^\times$ grupa cyklická. Pro $n > 1$ však není $\mathbb{Z}/p^n\mathbb{Z}$ těleso, a proto nelze věty 14 použít.

Věta 15. *Je-li p liché prvočíslo a $n \in \mathbb{N}$ libovolné, pak $(\mathbb{Z}/p^n\mathbb{Z})^\times$ je cyklická grupa.*

Poznámka. Pro $p = 2$ a $n > 2$ cyklickou grupu nedostáváme: například $(\mathbb{Z}/8\mathbb{Z})^\times$ je necyklická čtyřprvková grupa.

Definice. Nechť p je liché prvočíslo, $n \in \mathbb{N}$, $[g]_{p^n}$ generátor grupy $(\mathbb{Z}/p^n\mathbb{Z})^\times$. Pak g nazýváme **primitivní kořen modulo** p^n.

Poznámka. Libovolné $g \in \mathbb{Z}$ je primitivní kořen modulo p^n právě tehdy, když platí: pro každé $a \in \mathbb{Z}$ nedělitelné p existuje jediné $k \in \{1, 2, \ldots, (p-1)p^{n-1}\}$ tak, že $a \equiv g^k \pmod{p^n}$.

Věta 16 (Wilsonova věta). *Nechť $n \in \mathbb{N}$, $n > 1$. Pak n je prvočíslo, právě když platí $(n-1)! \equiv -1 \pmod{n}$.*

Rozklad přirozeného čísla na součin prvočísel

Připomeňme, že hledáme co nejrychlejší algoritmus, který dané přirozené číslo N rozloží na součin prvočísel.
Rozdělme náš problém na tři úkoly:

1. **Test na složenost:** Pro dané $N \in \mathbb{N}$ rychle rozhodnout, zda N splňuje nějakou podmínku, která je splněna každým prvočíslem, a tedy rozhodnout, zda je to *určitě číslo složené* anebo zda je to *asi prvočíslo*.
2. **Test na prvočíselnost:** Je-li N asi prvočíslo, *dokázat,* že N skutečně prvočíslem je, nebo to vyvrátit.
3. **Nalezení dělitele:** Je-li N složené, nalézt netriviálního dělitele d čísla N.

Celé rozkládání je pak rekurzivní proces: máme-li dělitele d čísla N, který splňuje $1 < d < N$, opakujeme celý postup pro čísla d a $\frac{N}{d}$.

Metoda pokusného dělení

Zkoušíme dělit N postupně všemi prvočísly 2, 3, 5, 7, 11, ... až do jisté hranice. Pokud jsme schopni to provést pro všechna prvočísla $p \leq \sqrt{N}$, provedeme tím všechny tři úkoly současně.

V případě, kdy N je natolik velké, že dělení N všemi prvočísly $p \leq \sqrt{N}$ by bylo příliš zdlouhavé, můžeme N dělit všemi prvočísly až do jisté hranice, abychom se zbavili malých faktorů (čím je prvočíslo menší, tím větší je pravděpodobnost, že dělí náhodně zvolené přirozené číslo).

Budeme předpokládat, že máme uloženu tabulku prvočísel $p[1] = 2$, $p[2] = 3$, $p[3] = 5$, $p[4] = 7$, ..., $p[k]$. Po vyčerpání této tabulky budeme pokračovat v dělení N čísly z jistých zbytkových tříd (např. čísly dávajícími zbytek 1 nebo 5 po dělení 6, resp. čísly dávajícími zbytek 1, 7, 11, 13, 17, 19, 23 nebo 29 po dělení 30 apod.). Tím sice některá dělení provedeme zbytečně, ale výsledek bude stále správný (testovat, zda číslo, kterým hodláme dělit, je prvočíslo, nemá smysl).

Zvolme horní hranici B, v níž přestaneme dělit, abychom algoritmem neztratili příliš mnoho času.

Algoritmus (Pokusné dělení). *Je dána tabulka prvočísel* $p[1] = 2$, $p[2] = 3$, $p[3] = 5$, $p[4] = 7$, $\ldots$, $p[k]$ *(kde* $k > 3$*), číslo* $B \geq p[k]$*, vektor* $t = [6, 4, 2, 4, 2, 4, 6, 2]$*, číslo* j *tak, že* $j = 0$ *(resp. 1, 2, 3, 4, 5, 6, 7), právě když* $p[k] \equiv 1 \pmod{30}$ *(resp. 7, 11, 13, 17, 19, 23, 29). Pro dané* $N \in \mathbb{N}$ *algoritmus hledá rozklad čísla* N*. Jen největší činitel tohoto rozkladu nemusí být prvočíslo, ale v tom případě může být dělitelný jen prvočísly většími než* B*.*

1. *[Inicializace] Je-li* $N \leq 5$*, pak vytiskni odpovídající rozklad* N *a skonči. Jinak polož* $i \leftarrow -1$, $m \leftarrow 0$.
2. *[Další prvočíslo] Polož* $m \leftarrow m + 1$*. Je-li* $m > k$*, polož* $i \leftarrow j - 1$ *a jdi na 5, jinak polož* $d \leftarrow p[m]$.
3. *[Zkus dělit] Současně spočítej* $r \leftarrow N \bmod d$, $q \leftarrow [\frac{N}{d}]$*. Je-li* $r = 0$*, vytiskni* d *jako činitele v hledaném rozkladu a polož* $N \leftarrow \frac{N}{d}$ *a opakuj krok 3.*
4. *[Prvočíslo?] Je-li* $d \geq q$*, vytiskni* N *jako posledního činitele a zprávu, že rozklad je úplný a skonči. Jinak, je-li* $i < 0$*, jdi na 2.*
5. *[Další dělitel] Polož* $i \leftarrow (i + 1) \bmod 8$, $d \leftarrow d + t[i]$*. Je-li* $d > B$*, vytiskni* N *jako posledního činitele a zprávu, že poslední činitel v rozkladu nemusí být prvočíselný a skonči. Jinak jdi na 3.*

Metoda pokusného dělení

Jestliže v kroku 4 nastane $d \geq q$, už víme, že N není dělitelné žádným prvočíslem $p \leq d$. Navíc

$$N = qd + r < (q+1)d \leq (d+1)d,$$

a tedy $\sqrt{N} < d + 1$. Proto už musí být N prvočíslo.

Pro úplný rozklad N se metoda pokusného dělení hodí jen pro malá N (řekněme $N < 10^8$), protože pro větší N existují lepší metody. Každopádně je užitečná pro odstranění malých faktorů.

Vhodnou tabulkou prvočísel by mohla být tabulka prvočísel menších než 500 000, máme-li na ni dost místa v paměti (je to 41 538 prvočísel). Vhodnější než uložení vlastních prvočísel může být uložení diferencí mezi nimi nebo dokonce poloviny diferencí (diferenci $p[k] - p[k-1]$ můžeme uložit do jednoho bytu pro $p[k] \leq 1\,872\,851\,947$, její polovinu dokonce pro $p[k] \leq 1\,999\,066\,711\,391$).

Obsahuje-li naše tabulka prvočísla aspoň do 500 000, je asi lepší po vyčerpání tabulky v dělení nepokračovat, ale užít jinou metodu.

Testy na složenost

Zvolme nějakou podmínku, které vyhovuje každé prvočíslo a které složená čísla většinou nevyhovují, přičemž je třeba, aby bylo možné podmínku pro dané přirozené číslo rychle ověřit.

Na první pohled se zdá být vhodnou podmínkou Wilsonova věta:

$$\forall n > 1: \qquad n \text{ je prvočíslo } \Leftrightarrow \ (n-1)! \equiv -1 \pmod{n}$$

To je dokonce nutná a dostatečná podmínku prvočíselnosti a byla by tedy nejen testem na složenost, ale také testem na prvočíselnost. Avšak nikdo neví, jak spočítat pro velká N dostatečně rychle číslo $(N-1)! \mod N$.

Výhodnější podmínku dává Fermatova věta, neboť je možné rychle počítat mocniny prvků v libovolné grupě.

Předpokládejme, že je dána grupa $(G, \cdot)$ s neutrálním prvkem 1 a že umíme prvky této grupy uchovávat v paměti a také s nimi počítat (násobit je a počítat jejich inverzní prvky).

Výpočet mocniny v grupě

Algoritmus (Binární umocňování zprava doleva). *Pro dané $g \in G$ a dané celé číslo n algoritmus počítá g^n v grupě $(G, \cdot)$. Proměnné y a z slouží k uchovávání prvků grupy G.*

1. *[Inicializace] Polož $y \leftarrow 1$. Je-li $n = 0$, pak vytiskni y a skonči. Je-li $n < 0$, polož $N \leftarrow -n$, $z \leftarrow g^{-1}$. Jinak polož $N \leftarrow n$, $z \leftarrow g$.*
2. *[Násob?] Je-li N liché, polož $y \leftarrow z \cdot y$.*
3. *[Poloviční N] Polož $N \leftarrow [\frac{N}{2}]$. Je-li $N = 0$, vytiskni y a skonči. Jinak polož $z \leftarrow z \cdot z$ a jdi na 2.*

Důkaz správnosti algoritmu. Vždy před započetím kroku 2 platí $y \cdot z^N = g^n$. Jistě to platilo při prvním vstupu na krok 2.
Označme N', y' a z' nové hodnoty proměnných N, y a z po provedení kroků 2 a 3.
Je-li N sudé, pak $y' = y$, $N' = \frac{N}{2}$, $z' = z^2$, tedy

$$y' \cdot (z')^{N'} = y \cdot z^{2\frac{N}{2}} = y \cdot z^N.$$

Je-li N liché, pak $y' = y \cdot z$, $N' = \frac{N-1}{2}$, $z' = z^2$, tedy

$$y' \cdot (z')^{N'} = y \cdot z \cdot z^{2\frac{N-1}{2}} = y \cdot z^N.$$

Odhad časové náročnosti algoritmu

Grupové násobení se provádí $a + b - 1$ krát, kde a je počet cifer ve dvojkovém zápise čísla n a b je počet jedniček v tomto zápise. Jistě platí $a + b - 1 \leq 2[\log_2 |n|] + 1$.
Je-li například $G = (\mathbb{Z}/m\mathbb{Z})^\times$, je jedno násobení časové náročnosti $O(\ln^2 m)$, proto celý algoritmus je časové náročnosti $O(\ln^2 m \ln |n|)$.

V předchozím algoritmu jsme procházeli cifry dvojkového zápisu čísla n zprava doleva. Zcela analogicky můžeme tyto cifry procházet ovšem zleva doprava. Musíme však znát polohu „nejlevější“ jedničky v tomto zápise, tj. znát $e \in \mathbb{Z}$ s vlastností $2^e \leq |n| < 2^{e+1}$.

Algoritmus (Binární umocňování zleva doprava). *Pro dané $g \in G$ a dané celé číslo n algoritmus počítá g^n v grupě $(G, \cdot)$. Je-li $n \neq 0$, musí být dáno $e \in \mathbb{Z}$ s vlastností $2^e \leq |n| < 2^{e+1}$.*

1. *[Inicializace] Je-li $n = 0$, pak vytiskni 1 a skonči. Je-li $n < 0$, polož $N \leftarrow -n$, $z \leftarrow g^{-1}$. Jinak polož $N \leftarrow n$, $z \leftarrow g$. Konečně (tj. v obou případech) polož $y \leftarrow z$, $E \leftarrow 2^e$, $N \leftarrow N - E$.*
2. *[Konec?] Je-li $E = 1$, vytiskni y a skonči. Jinak polož $E \leftarrow \frac{E}{2}$.*
3. *[Násob] Polož $y \leftarrow y \cdot y$. Je-li $N \geq E$, polož $N \leftarrow N - E$, $y \leftarrow y \cdot z$. Jdi na 2.*

Důkaz správnosti algoritmu. Vždy před započetím kroku 2 platí $y^E \cdot z^N = g^n$. Jistě to platilo při prvním vstupu na krok 2, kdy $y^E \cdot z^N = z^{2^e} \cdot z^{N-2^e} = z^N = g^n$.

Označme E', N' a y' nové hodnoty proměnných E, N a y po provedení kroků 2 a 3.

Je-li $N < E'$, pak $E' = \frac{E}{2}$, $y' = y^2$, $N' = N$, tedy

$$(y')^{E'} \cdot z^{N'} = (y^2)^{\frac{E}{2}} \cdot z^N = y^E \cdot z^N.$$

Je-li $N \geq E'$, pak $E' = \frac{E}{2}$, $y' = y^2 \cdot z$, $N' = N - E'$, tedy

$$(y')^{E'} \cdot z^{N'} = (y^2 \cdot z)^{\frac{E}{2}} \cdot z^{N-\frac{E}{2}} = y^E \cdot z^N.$$

Vždy před krokem 2 je $0 \leq N < E$. Proto při skončení je $N = 0$.

Porovnání obou algoritmů

Nevýhody oproti předchozímu algoritmu: je třeba předem spočítat e, to však (je-li základ naší poziční soustavy pro uchovávání „velkých“ čísel mocnina 2) je velmi rychlé. Patrně totiž budeme znát pozici nejvyšší nenulové cifry v naší poziční soustavě a pak určení e zabere čas ohraničený konstantou. Zdánlivou nevýhodou je i uchovávání velkého čísla E a výpočet rozdílu $N - E$. Avšak při implementaci budeme uchovávat e a test $N \geq E$ i výpočet $N - E$ provedeme manipulací s bitem obsahujícím e-tou dvojkovou cifru čísla N (zde je podstatné, aby skutečně byl základ naší poziční soustavy mocnina 2).

Výhoda: jedno ze dvou násobení, které se provádí v kroku 3, je vždy proměnnou z, ve které je v průběhu celého výpočtu g (nebo g^{-1} je-li $n < 0$). Je-li tedy například $G = (\mathbb{Z}/m\mathbb{Z})^\times$, v případě, že $g = a + m\mathbb{Z}$ pro $|a|$ ohraničené konstantou (třeba základem naší poziční soustavy), je násobení g pouze lineárního času a ne kvadratického. Pak se tedy v kroku 3 provede jedno násobení řádu $O(\ln^2 m)$ a nejvýše jedno řádu $O(\ln m)$. Celý algoritmus je sice stále časové náročnosti $O(\ln^2 m \ln |n|)$, ale s menší O-konstantou.

Využití umocňování pro test na složenost

Pro test na složenost užijeme malou Fermatovu větu: máme tedy dáno přirozené číslo N, o kterém chceme vědět, zda je to číslo složené. Budeme to vědět jistě, nalezneme-li celé číslo a, $1 \leq a < N$, pro které platí $a^{N-1} \not\equiv 1 \pmod{N}$.
Takové a se nazývá svědek složenosti čísla N. Pokud však pro takové a platí $a^{N-1} \equiv 1 \pmod{N}$, nemůžeme z toho usoudit nic.

Celý algoritmus tedy bude vypadat takto: budeme náhodně volit $a \in \mathbb{Z}$, $1 < a < N$, a počítat $a^{N-1} \mod N$. Pokud pro některé a bude splněno $a^{N-1} \not\equiv 1 \pmod{N}$, jsme hotovi a víme, že N je opravdu složené číslo (zmíněné a si můžeme zapamatovat pro případ, že bychom chtěli přesvědčit někoho dalšího o složenosti N). Pokud pro všechna a budeme dostávat $a^{N-1} \equiv 1 \pmod{N}$, po jistém počtu pokusů algoritmus ukončíme a usoudíme, že patrně je N prvočíslo. Jestli je však opravdu N prvočíslo, takto zjistit nemůžeme.

Nevýhodou popsaného algoritmu je, že téměř jistě neodhalí jistý typ složených čísel, nazývaných Carmichaelova čísla.

Carmichaelova čísla

Definice. Složené číslo N se nazývá Carmichaelovo číslo, jestliže pro všechna celá čísla a, která jsou nesoudělná s N, platí $a^{N-1} \equiv 1 \pmod{N}$.

Carmichaelovo číslo by náš algoritmus označil za složené pouze tehdy, kdyby za a zvolil číslo soudělné s N, což je však velmi nepravděpodobné. Přitom platí:

Věta (Alford, Granville, Pomerance). Existuje nekonečně mnoho Carmichaelových čísel.

Příklad. $N = 561 = 3 \cdot 11 \cdot 17$ je Carmichaelovo číslo.

Důkaz. Pro libovolné celé číslo a nesoudělné s 561 z Fermatovy věty dostáváme $a^2 \equiv 1 \pmod{3}$, $a^{10} \equiv 1 \pmod{11}$ a $a^{16} \equiv 1 \pmod{17}$. Protože 2, 10 i 16 jsou dělitelé čísla 560, je 561 Carmichaelovo číslo.

Výhodnější než testovat Fermatovu větu je proto testování následujícího zesílení Fermatovy věty.

Využití zesílení malé Fermatovy věty

Věta. *Pro libovolné liché prvočíslo p a libovolné celé číslo a nedělitelné p platí*

$$a^{\frac{p-1}{2}} \equiv \pm 1 \pmod{p}.$$

Důkaz. Z Fermatovy věty

$$p \mid (a^{p-1} - 1) = (a^{\frac{p-1}{2}} - 1) \cdot (a^{\frac{p-1}{2}} + 1).$$

Protože je p prvočíslo, musí dělit některého z uvedených činitelů.

Příklad. Test, zda $a^{\frac{N-1}{2}} \equiv \pm 1 \pmod{N}$, by mohl vyloučit i 561, neboť

$$5^{280} \equiv 5^{16 \cdot 17 + 8} \equiv 5^8 = 25^4 \equiv 8^4 = 16^3 \equiv (-1)^3 = -1 \pmod{17}$$

a zároveň

$$5^{280} \equiv (5^2)^{140} \equiv 1 (\mathrm{mod}\, 3),$$

a proto 5^{280} není kongruentní modulo 561 ani s 1 ani s -1.

Další zesílení malé Fermatovy věty

Příklad. Test, zda $a^{\frac{N-1}{2}} \equiv \pm 1 \pmod{N}$, neodhalí například $N = 1729 = 7 \cdot 13 \cdot 19$, neboť $\frac{N-1}{2} = 864 = 2^5 \cdot 3^3$ je dělitelné 6, 12 i 18 a tedy z Fermatovy věty plyne, že pro všechna celá čísla a nesoudělná s N platí $a^{\frac{N-1}{2}} \equiv 1 \pmod{N}$.

Věta. *Nechť p je liché prvočíslo. Pišme $p - 1 = 2^t \cdot q$, kde t je přirozené číslo a q je liché. Pak pro každé celé číslo a nedělitelné p buď platí $a^q \equiv 1 \pmod{p}$ nebo existuje $e \in \{0, 1, \ldots, t-1\}$ splňující $a^{2^e q} \equiv -1 \pmod{p}$.*

Důkaz. Z Fermatovy věty

$$p \mid (a^{p-1} - 1) = (a^q - 1) \cdot \prod_{e=0}^{t-1} (a^{2^e q} + 1).$$

Protože je p prvočíslo, musí dělit některého z uvedených činitelů.

Teoretický základ testu Millera a Rabina

Věta. *Nechť $N > 10$ je liché složené číslo. Pišme $N - 1 = 2^t \cdot q$, kde t je přirozené číslo a q je liché. Pak nejvýše čtvrtina z čísel množiny $\{a \in \mathbb{Z};\ 1 \leq a < N,\ (a, N) = 1\}$ splňuje následující podmínku:*

$$a^q \equiv 1 \pmod{N}$$

nebo existuje $e \in \{0, 1, \ldots, t-1\}$ splňující

$$a^{2^e q} \equiv -1 \pmod{N}.$$

Pro dané N algoritmus otestuje podmínku věty pro 20 náhodně zvolených a. Pokud pro některé takové a není podmínka splněna, vytiskne zprávu, že N je složené. Pokud je splněna podmínka pro každé takové a, vytiskne zprávu, že N je asi prvočíslo.

Podle předchozí věty je pravděpodobnost, že bude vytištěna tato zpráva, ačkoli je N složené, menší než 4^{-20}.

Algoritmus Millera a Rabina

Algoritmus (Miller - Rabin). *Pro dané liché $N \geq 3$ algoritmus s vysokou pravděpodobností objeví, že N je složené. Pokud se mu to nepodaří, vytiskne zprávu, že N je asi prvočíslo.*

1. *[Inicializace] Polož $q \leftarrow N-1$, $t \leftarrow 0$. Dokud je q sudé, opakuj $q \leftarrow \frac{q}{2}$, $t \leftarrow t+1$. Polož $c \leftarrow 20$.*
2. *[Zvol a] Pomocí generátoru náhodných čísel zvol náhodně $a \in \mathbb{Z}$, $1 < a < N$. Pak polož $e \leftarrow 0$, $b \leftarrow a^q \mod N$. Je-li $b = 1$, jdi na 4.*
3. *[Umocňuj na druhou] Dokud je $b \neq N-1$ a $e \leq t-2$, opakuj $b \leftarrow b^2 \mod N$, $e \leftarrow e+1$. Je-li $b \neq N-1$, vytiskni zprávu, že N je složené a vytiskni svědka složenosti a. Skonči.*
4. *[Už proběhlo 20 pokusů?] Polož $c \leftarrow c-1$. Je-li $c > 0$, jdi na 2. Jinak vytiskni zprávu, že N je asi prvočíslo a skonči.*

Odhad časové náročnosti. Algoritmus je řádově stejné časové náročnosti jako umocňování v něm použité (předpokládáme, že generování nového a je řádově rychlejší), proto jde o kubickou časovou náročnost.

Testy na prvočíselnost

Víme, že N je asi prvočíslo (například prošlo testem Millera a Rabina).
Chceme *dokázat*, že N skutečně prvočíslem je, nebo to vyvrátit.
Na následující větě je založen $N-1$ test Pocklingtona a Lehmera.

Věta. *Nechť N je přirozené číslo, $N > 1$. Nechť p je prvočíslo dělící $N-1$. Předpokládejme dále, že existuje $a_p \in \mathbb{Z}$ tak, že*

$$a_p^{N-1} \equiv 1 \pmod{N} \quad a \quad \left(a_p^{\frac{N-1}{p}} - 1, N\right) = 1.$$

Nechť p^{α_p} je nejvyšší mocnina p dělící $N-1$. Pak pro každý kladný dělitel d čísla N platí

$$d \equiv 1 \pmod{p^{\alpha_p}}.$$

Důkaz věty Pocklingtona a Lehmera

Věta. Nechť N je přirozené číslo, $N > 1$. Nechť p je prvočíslo dělící $N-1$. Předpokládejme dále, že existuje $a_p \in \mathbb{Z}$ tak, že

$$a_p^{N-1} \equiv 1 \pmod{N} \quad a \quad \left(a_p^{\frac{N-1}{p}} - 1, N\right) = 1.$$

Nechť p^{α_p} je nejvyšší mocnina p dělící $N-1$. Pak pro každý kladný dělitel d čísla N platí $d \equiv 1 \pmod{p^{\alpha_p}}$.

Důkaz. Každý kladný dělitel d čísla N je součinem prvočíselných dělitelů čísla N, větu dokažme pouze pro d prvočíslo. Podle Fermatovy věty platí $a_p^{d-1} \equiv 1 \pmod{d}$, neboť $(a_p, N) = 1$.
Protože $\left(a_p^{\frac{N-1}{p}} - 1, N\right) = 1$, platí $a_p^{\frac{N-1}{p}} \not\equiv 1 \pmod{d}$.
Pro $e = \min\{n \in \mathbb{N};\ a_p^n \equiv 1 \pmod{d}\}$ platí $e \mid d-1$, $e \mid N-1$ a $e \nmid \frac{N-1}{p}$. Kdyby $p^{\alpha_p} \nmid e$, z $e \mid N-1$ by plynulo $e \mid \frac{N-1}{p}$, spor. Je tedy $p^{\alpha_p} \mid e$, a tedy i $p^{\alpha_p} \mid d-1$.

Užití věty Pocklingtona a Lehmera

Věta. Nechť N je přirozené číslo, $N > 1$. Nechť p je prvočíslo dělící $N - 1$. Předpokládejme dále, že existuje $a_p \in \mathbb{Z}$ tak, že

$$a_p^{N-1} \equiv 1 \pmod{N} \quad a \quad \left(a_p^{\frac{N-1}{p}} - 1, N\right) = 1. \tag{1}$$

Nechť p^{α_p} je nejvyšší mocnina p dělící $N - 1$. Pak pro každý kladný dělitel d čísla N platí $d \equiv 1 \pmod{p^{\alpha_p}}$.

Důsledek. Nechť $N \in \mathbb{N}$, $N > 1$. Předpokládejme, že můžeme psát $N - 1 = F \cdot U$, kde $(F, U) = 1$ a $F > \sqrt{N}$, přičemž známe rozklad čísla F na prvočinitele. Pak platí:

- jestliže pro každé prvočíslo $p \mid F$ můžeme najít $a_p \in \mathbb{Z}$ splňující (1) z předchozí věty, pak je N prvočíslo;
- je-li N prvočíslo, pak pro libovolné prvočíslo $p \mid N - 1$ existuje $a_p \in \mathbb{Z}$ splňující (1).

Zesílení užití věty Pocklingtona a Lehmera

Důsledek. *Nechť $N \in \mathbb{N}$, $N > 1$. Předpokládejme, že můžeme psát $N - 1 = F \cdot U$, kde $(F, U) = 1$, přičemž známe rozklad čísla F na prvočinitele. Dále předpokládejme, že všechna prvočísla dělící U jsou větší než $B \in \mathbb{N}$ a že platí $B \cdot F \geq \sqrt{N}$.*
Pak platí: jestliže pro každé prvočíslo $p \mid F$ můžeme najít $a_p \in \mathbb{Z}$ splňující (1) z předchozí věty a jestliže navíc existuje $a_U \in \mathbb{Z}$ splňující

$$a_U^{N-1} \equiv 1 \pmod{N} \quad a \quad (a_U^F - 1, N) = 1,$$

pak je N prvočíslo.
Je-li naopak N prvočíslo a $U > 1$, pak požadovaná $a_p, a_U \in \mathbb{Z}$ existují.

Důkaz. Pro každé prvočíslo $d \mid N$ víme, že $d \equiv 1 \pmod{F}$. Protože $(a_U, N) = 1$, existuje $e = \min\{n \in \mathbb{N};\ a_U^n \equiv 1 \pmod{d}\}$. Odtud $e \mid d - 1$, $e \mid N - 1$ a $e \nmid F$. Kdyby $(e, U) = 1$, z $e \mid N - 1 = FU$ by plynulo $e \mid F$. Je tedy $(e, U) > 1$ a protože U je dělitelné pouze prvočísly většími než B, platí $(e, U) > B$. Protože $(F, U) = 1$, z $d \equiv 1 \pmod{e}$ a $d \equiv 1 \pmod{F}$ plyne $d \equiv 1 \pmod{F \cdot (e, U)}$ a tedy $d > F \cdot (e, U) > FB \geq \sqrt{N}$.

Příklad užití věty Pocklingtona a Lehmera – Pépinův test

Pro $k \in \mathbb{Z}$, $k > 0$, se $F_k = 2^{2^k} + 1$ nazývá k-té Fermatovo číslo.
Pépinův test: F_k je prvočíslo, právě když $3^{(F_k-1)/2} \equiv -1 \pmod{F_k}$.

Důkaz. „$\Leftarrow$“ z důsledku věty Pocklingtona a Lehmera.
„$\Rightarrow$“ z kvadratického zákona reciprocity: $3^{(F_k-1)/2} \equiv \left(\frac{3}{F_k}\right) =$
$= \left(\frac{F_k}{3}\right) \cdot (-1)^{\frac{3-1}{2}\frac{F_k-1}{2}} = \left(\frac{F_k}{3}\right) = \left(\frac{2}{3}\right) = -1 \pmod{F_k}$.

- $F_0 = 3$, $F_1 = 5$, $F_2 = 17$, $F_3 = 257$, $F_4 = 65537$ jsou prvočísla.
- Rozklad čísla F_k na prvočinitele je znám jen pro $k \leq 11$.
- F_k jsou složená pro každé $k = 5, 6, \ldots, 32$ (ačkoli u F_{20} ani F_{24} není znám žádný prvočíselný dělitel).
- Největší F_k, pro které je znám prvočíselný dělitel, je $F_{2543548}$ dělitelné prvočíslem $9 \cdot 2^{2543551} + 1$ (objeveno 22. 6. 2011).
- Otevřené problémy: Existuje nekonečně mnoho složených Fermatových čísel? Existuje nekonečně mnoho Fermatových čísel, která jsou prvočísla?

Implementace algoritmu (tzv. $N-1$ test)

Vstupem je číslo N, které již prošlo testem Millera - Rabina, tedy číslo, o kterém s vysokou pravděpodobností platí, že je to prvočíslo. Je třeba to však dokázat.

V první části algoritmu rozkládáme číslo $N-1$ na součin $F \cdot U$ a to tak, že podrobíme $N-1$ algoritmu pokusného dělení, ukládáme získané dělitele a skončíme, až platí $BF \geq \sqrt{N}$, nebo až je B „dost velké“, abychom si byli jisti zastavením v „rozumném“ čase (zde B, F, U značí totéž, co v předchozí větě).

Pak náhodně volíme celá čísla a_p v intervalu $1 < a_p < N$ a počítáme $b_p = a_p^{\frac{N-1}{p}} \bmod N$ a $c_p = b_p^p \bmod N$ do té doby, než $c_p \equiv 1 \pmod N$ a $(b_p - 1, N) = 1$.

Je-li N opravdu prvočíslo, podmínku $(b_p - 1, N) = 1$ splňuje většina z čísel a_p, přesněji právě $\frac{p-1}{p}(N-1)$ čísel z $N-1$ čísel $1, 2, \ldots, N-1$. Můžeme tedy očekávat, že takové a_p brzy najdeme.

Pokud by však N bylo „velké“ Carmichaelovo číslo, algoritmus by se s velkou pravděpodobností nezastavil (zastaví se jen když náhodně volené a_p bude soudělné s N).

Časová náročnost algoritmu

Není-li N prvočíslo, algoritmus se nemusí zastavit.

Ani pro prvočísla nelze stanovit odhad: záleží na tom, jak snadno lze rozkládat číslo $N - 1$. Následné hledání čísel a_p je velmi rychlé (kontrola, zda zvolené a_p splňuje podmínku (1) je kvadratické časové náročnosti, navíc lze volit a_p „malá“).

Je možné nerozloženou část U podrobit testu Millera a Rabina a v případě, že test zjistí, že U je asi prvočíslo, dokázat nejprve prvočíselnost U (a tedy pracovat rekurzivně).

Zobecnění algoritmu

Je-li N prvočíslo, pak existuje těleso $\mathbb{F}_{N^2}$ o N^2 prvcích. Jeho multiplikativní grupa je cyklická řádu $N^2 - 1 = (N-1)(N+1)$. Existuje tedy $\alpha \in \mathbb{F}_{N^2}$ řádu $N+1$, tj. splňující $\alpha^{N+1} = 1$, avšak $\alpha^{\frac{N+1}{p}} \neq 1$ pro libovolné prvočíslo p dělící $N+1$.

Tuto myšlenku je možno využít pro tzv. $N+1$ test analogický $N-1$ testu. V něm vystupuje faktorizace čísla $N+1$ místo $N-1$.

Pro důkaz prvočíselnosti čísla N lze pak využít informace o dělitelích čísla N, získané z obou testů.

Podobně lze využít těleso $\mathbb{F}_{N^3}$ (a tedy faktorizovat $\frac{N^3-1}{N-1} = N^2 + N + 1$), těleso $\mathbb{F}_{N^4}$ (a faktorizovat $\frac{N^4-1}{N^2-1} = N^2 + 1$) nebo těleso $\mathbb{F}_{N^6}$ (a faktorizovat $\frac{N^6-1}{(N^3-1)(N+1)} = N^2 - N + 1$).

Vždy nám však už vycházejí čísla podstatně větší než N a tedy pravděpodobně obtížně rozložitelná.

Některé nezbytnosti z algebraické geometrie

Nechť K je těleso.

Definice. n-rozměrným afinním prostorem nad K rozumíme kartézskou mocninu K^n. Budeme jej značit $A^n(K)$, tj.

$$A^n(K) = \{(x_1, \ldots, x_n);\ x_1, \ldots, x_n \in K\}.$$

Definice. n-rozměrným projektivním prostorem nad K rozumíme rozklad na množině $K^{n+1} - \{(0, \ldots, 0)\}$ příslušný ekvivalenci $\sim$, kterou definujeme takto: pro libovolné $(n+1)$-tice $(x_1, \ldots, x_{n+1})$, $(y_1, \ldots, y_{n+1}) \in K^{n+1}$ položíme $(x_1, \ldots, x_{n+1}) \sim (y_1, \ldots, y_{n+1})$ právě tehdy, když existuje $\lambda \in K^\times$, které pro každé $i \in \{1, \ldots, n+1\}$ splňuje podmínku $x_i = \lambda y_i$. Tento n-rozměrný projektivní prostor nad K budeme značit $P^n(K)$, třídu rozkladu (tj. bod projektivního prostoru) obsahující $(n+1)$-tici $(x_1, \ldots, x_{n+1})$ budeme značit $[x_1, \ldots, x_{n+1}]$.

Afinní část projektivního prostoru

Nechť $x_1, \ldots, x_{n+1}$ jsou z tělesa K, přičemž alespoň jedno z nich je různé od nuly.

Jestliže $x_{n+1} \neq 0$, pak platí $[x_1, \ldots, x_{n+1}] = [\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}, 1]$, čímž je pevně dán bod $(\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}) \in A^n(K)$.

Jestliže naopak $x_{n+1} = 0$, určuje $[x_1, \ldots, x_{n+1}]$ jednoznačně bod $[x_1, \ldots, x_n] \in P^{n-1}(K)$.

Lze tedy n-rozměrný projektivní prostor „rozdělit“ na n-rozměrný afinní prostor, který považujeme za množinu „vlastních bodů“ a na množinu „nevlastních bodů“, která tvoří $(n-1)$-rozměrný projektivní prostor.

Můžeme si představovat, že nevlastní body „leží v nekonečnu.“ Toto rozdělení však *není* kanonické – lze to provést mnoha způsoby. Tedy to, zda je bod vlastní nebo ne, je věc naší volby.

Nadplochy projektivního prostoru

Máme-li homogenní polynom $F(t_1, \ldots, t_{n+1}) \in K[t_1, \ldots, t_{n+1}]$ o $n+1$ proměnných nad K stupně k a bod $[x_1, \ldots, x_{n+1}] \in P^n(K)$, má smysl se ptát, zda $F(x_1, \ldots, x_{n+1}) = 0$. Je-li totiž $[x_1, \ldots, x_{n+1}] = [y_1, \ldots, y_{n+1}]$, pak existuje $\lambda \in K^\times$, které pro každé $i \in \{1, \ldots, n+1\}$ splňuje podmínku $x_i = \lambda y_i$. Pak ovšem $F(x_1, \ldots, x_{n+1}) = F(\lambda y_1, \ldots, \lambda y_{n+1}) = \lambda^k \cdot F(y_1, \ldots, y_{n+1})$, a tedy $F(x_1, \ldots, x_{n+1}) = 0$, právě když $F(y_1, \ldots, y_{n+1}) = 0$.

Definice. Nechť $F(t_1, \ldots, t_{n+1}) \in K[t_1, \ldots, t_{n+1}]$ je homogenní polynom stupně k. Množina

$$\mathcal{C} = \{[x_1, \ldots, x_{n+1}] \in P^n(K);\ F(x_1, \ldots, x_{n+1}) = 0\}$$

se nazývá nadplocha stupně k v $P^n(K)$. Je-li $n = 2$, hovoříme také o křivce stupně k v projektivní rovině $P^2(K)$.

Singulární bod nadplochy projektivního prostoru

Parciální derivací homogenního mnohočlenu je opět homogenní mnohočlen. Má proto smysl následující definice.

Definice. Nechť $F(t_1, \ldots, t_{n+1}) \in K[t_1, \ldots, t_{n+1}]$ je homogenní polynom stupně k a

$$\mathcal{C} = \{[x_1, \ldots, x_{n+1}] \in P^n(K);\ F(x_1, \ldots, x_{n+1}) = 0\}$$

příslušná nadplocha. Bod $[x_1, \ldots, x_{n+1}] \in \mathcal{C}$ se nazývá singulární, jestliže pro každé $i \in \{1, \ldots, n+1\}$ platí

$$\frac{\partial F}{\partial x_i}(x_1, \ldots, x_{n+1}) = 0.$$

Nadplocha $\mathcal{C}$ se nazývá singulární, existuje-li alespoň jeden její singulární bod.

Příklad

Uvažme reálnou projektivní rovinu $P^2(\mathbb{R})$.
Abychom se vyhnuli indexům, budeme psát x, y, z místo t_1, t_2, t_3.

Kubický mnohočlen $F_1(x, y, z) = x^3 + x^2z - y^2z$ nám definuje kubickou křivku $\mathcal{C}_1$ (tj. křivku stupně 3)

$$\mathcal{C}_1 = \{[x, y, z] \in P^2(\mathbb{R});\ F_1(x, y, z) = 0\}.$$

Jistě $[0, 0, 1] \in \mathcal{C}_1$. Tento bod je singulární, neboť

$$\frac{\partial F_1}{\partial x} = 3x^2 + 2xz, \quad \frac{\partial F_1}{\partial y} = -2yz, \quad \frac{\partial F_1}{\partial z} = x^2 - y^2.$$

Je tedy $\mathcal{C}_1$ singulární křivka.

Další příklad

Opět pracujme s reálnou projektivní rovinu $P^2(\mathbb{R})$.

Uvažme nyní mnohočlen $F_2(x, y, z) = x^3 + xz^2 - y^2z$ a příslušnou kubickou křivku

$$\mathcal{C}_2 = \{[x, y, z] \in P^2(\mathbb{R});\ F_2(x, y, z) = 0\}.$$

Hledejme singulární body na $\mathcal{C}_2$. Platí

$$\frac{\partial F_2}{\partial x} = 3x^2 + z^2, \quad \frac{\partial F_2}{\partial y} = -2yz, \quad \frac{\partial F_2}{\partial z} = 2xz - y^2.$$

Z $\frac{\partial F_2}{\partial x} = 0$ plyne $x = 0$ a $z = 0$, pak ale z $\frac{\partial F_2}{\partial z} = 0$ plyne i $y = 0$.

Ale trojice nul nedává žádný bod projektivní roviny. Singulární bod na $\mathcal{C}_2$ tedy neexistuje a proto $\mathcal{C}_2$ není singulární křivka.

Eliptické křivky

Definice. Eliptická křivka nad tělesem K je uspořádaná dvojice $(\mathcal{E}, O)$, kde $\mathcal{E}$ je nesingulární kubická křivka v $P^2(K)$ a $O \in \mathcal{E}$.

Poznámka. Je možné zavést pojem biracionální ekvivalence dvou křivek, spočívající v tom, že existují transformace prostoru převádějící jednu křivku na druhou a obráceně, přičemž tyto transformace jsou „pěkné“ v tom smyslu, že transformační rovnice jsou dány homogenními polynomy téhož stupně nad K.

Věta. *Libovolná eliptická křivka nad K je biracionálně ekvivalentní s nějakou eliptickou křivkou $(\mathcal{E}, O)$ následujícího tvaru (přičemž transformace převádějí vyznačený bod jedné křivky na vyznačený bod druhé křivky)*

$$\mathcal{E} = \{[x, y, z] \in P^2(K);\ F(x, y, z) = 0\},$$

kde

$$F(x, y, z) = y^2z + a_1xyz + a_2yz^2 - x^3 - a_3x^2z - a_4xz^2 - a_5z^3,$$

$a_1, \ldots, a_5 \in K$ *a* $O = [0, 1, 0]$.

Eliptické křivky dané Weierstrassovou rovnicí

V projektivní rovině zvolme za afinní část množinu těch bodů, které mají nenulovou třetí souřadnici, tedy bodů $[x, y, 1]$.

Každá eliptická křivka ve tvaru z předchozí věty má jeden nevlastní bod (totiž $O = [0, 1, 0]$) a v afinní části je dána rovnicí

$$y^2 + a_1xy + a_2y = x^3 + a_3x^2 + a_4x + a_5.$$

Tato rovnice se nazývá **Weierstrassova rovnice**.

V dalším textu budeme předpokládat, že charakteristika tělesa K není ani 2 ani 3, tj. že 2 i 3 jsou invertibilní prvky v K.

Důvodem je to, že pro naše účely eliptické křivky nad tělesy charakteristiky 2 a 3 nejsou zapotřebí a že tento předpoklad dále zjednodušuje Weierstrassovu rovnici.

Můžeme pak totiž předpokládat, že $a_1 = a_2 = a_3 = 0$, a tedy Weierstrassova rovnice je tvaru $y^2 = x^3 + a_4x + a_5$.

Kdy Weierstrassova rovnice zadává eliptickou křivku?

Věta. *Nechť K je těleso charakteristiky různé od 2 a 3, $a, b \in K$. Rovnice $y^2 = x^3 + ax + b$ je Weierstrassovou rovnicí nějaké eliptické křivky, právě když platí $4a^3 + 27b^2 \neq 0$.*

Důkaz. Položme $F(x, y, z) = y^2z - x^3 - axz^2 - bz^3$. Platí

$$\frac{\partial F}{\partial x} = -3x^2 - az^2, \quad \frac{\partial F}{\partial y} = 2yz, \quad \frac{\partial F}{\partial z} = y^2 - 2axz - 3bz^2.$$

Předpokládejme, že $[x, y, z]$ je singulární bod. Pak $z = 0$ implikuje $x = y = 0$, spor. Je tedy $z \neq 0$. Proto $y = 0$ a pro $\gamma = \frac{x}{z}$ platí $3\gamma^2 = -a$, $2a\gamma = -3b$. Jestliže $a = 0$, pak také $b = 0$. Naopak pro $a = b = 0$ je bod $[0, 0, 1]$ singulární. Zabývejme se dále případem $a \neq 0$. Platí $\gamma = -\frac{3b}{2a}$, $\gamma^2 = -\frac{a}{3} = \frac{9b^2}{4a^2}$, tj. $4a^3 + 27b^2 = 0$. Naopak, je-li $4a^3 + 27b^2 = 0$, $a \neq 0$, ověřme, že pro $\gamma = -\frac{3b}{2a}$ je $[\gamma, 0, 1]$ singulární bod, což je snadné, například $\gamma^2 = \frac{9b^2}{4a^2} = -\frac{a}{3}$, dále

$$\gamma^3 + a\gamma + b = \left(-\tfrac{3b}{2a}\right)\left(-\tfrac{a}{3}\right) + a\left(-\tfrac{3b}{2a}\right) + b = \tfrac{b}{2} - \tfrac{3b}{2} + b = 0.$$

Eliptická křivka daná Weierstrassovou rovnicí

Nechť K je těleso charakteristiky různé od 2 a 3, $a, b \in K$, $4a^3 + 27b^2 \neq 0$. Pak Weierstrassova rovnice

$$y^2z = x^3 + axz^2 + bz^3$$

spolu s význačným bodem $O = [0, 1, 0]$ zadává v projektivní rovině $P^2(K)$ eliptickou křivku $\mathcal{E}$.

Tento význačný bod O je jediným bodem na nevlastní přímce $z = 0$. Platí dokonce, že nevlastní přímka $z = 0$ má s eliptickou křivkou $\mathcal{E}$ trojnásobný bod dotyku O, neboť dosazením $z = 0$ do rovnice křivky dostaneme $x^3 = 0$.

Ostatní body eliptické křivky jsou vlastní a jsou v afinní rovině $A^2(K) = K^2$ určeny rovnicí $y^2 = x^3 + ax + b$.

Je-li $A = [\alpha, \beta, 1] \in \mathcal{E}$, pak i $B = [\alpha, -\beta, 1] \in \mathcal{E}$. Přímka AB má v $P^2(K)$ rovnici $x = \alpha z$ a obsahuje ještě třetí bod na $\mathcal{E}$, totiž O.

Eliptická křivka $\mathcal{E}: \; y^2z = x^3 + axz^2 + bz^3$, $O = [0,1,0]$

Jsou-li $A = [\alpha, \beta, 1] \in \mathcal{E}$, $B = [\gamma, \delta, 1] \in \mathcal{E}$, přičemž $\alpha \neq \delta$, přímka AB má v $P^2(K)$ rovnici $y = \beta z + (x - \alpha z)k$, kde $k = \frac{\delta - \beta}{\gamma - \alpha}$.
Hledejme průsečíky přímky AB s eliptickou křivkou $\mathcal{E}$.

Dosazením této rovnice za y do rovnice $y^2z = x^3 + axz^2 + bz^3$ a vydělením z^3 dostaneme kubickou rovnici pro $\frac{x}{z}$ s koeficienty z K:

$$\left(\tfrac{x}{z}\right)^3 + a\tfrac{x}{z} + b - \left(\beta + (\tfrac{x}{z} - \alpha)k\right)^2 = 0.$$

Jde o normovaný kubický polynom v $\frac{x}{z}$, jehož dva kořeny α a γ už známe. Proto má ještě třetí kořen $\sigma \in K$ a z Viétových vztahů zjistíme, že platí $\alpha + \gamma + \sigma = k^2$.

Přímka AB a eliptická křivka $\mathcal{E}$ mají tedy ještě třetí průsečík $C = [\sigma, \tau, 1]$, kde $\sigma = k^2 - \alpha - \gamma$, $\tau = \beta + k(\sigma - \alpha)$.

Někdy může bod C splynout s některým z bodů A, B, v tom případě mluvíme o dvojnásobném průsečíku.

Eliptická křivka $\mathcal{E}: \; y^2z = x^3 + axz^2 + bz^3, \; O = [0,1,0]$

Podobně se odvodí, že pokud sestrojíme křivce $\mathcal{E}$ v jejím bodě A tečnu, protne tato tečna křivku $\mathcal{E}$ ještě v jednom bodě. Máme tedy operaci: pro libovolnou dvojici bodů $A, B \in \mathcal{E}$ je jejím výsledkem třetí průsečík, který nazveme $A \star B$. Tato operace však není „pěkná“: nemá neutrální prvek, není asociativní.

Proto operaci ještě trochu pozměníme pomocí pevně zvoleného bodu O. Definujeme součet bodů $A, B \in \mathcal{E}$ předpisem

$$A + B = (A \star B) \star O.$$

Tato operace sčítání bodů je zřejmě komutativní, $(\mathcal{E}, +)$ má neutrální prvek O a libovolný bod $A = [\alpha, \beta, 1] \in \mathcal{E}$ má opačná bod $-A = [\alpha, -\beta, 1] \in \mathcal{E}$. Je možné dokázat, že operace sčítání bodů je také asociativní, je tedy $(\mathcal{E}, +)$ komutativní grupa. Důkaz asociativity je mimo možnosti této přednášky.

Explicitní popis operace sčítání bodů

Věta. *Nechť K je těleso charakteristiky různé od 2 a 3, $a, b \in K$, $4a^3 + 27b^2 \neq 0$. Nechť $\mathcal{E}$ je eliptická křivka daná Weierstrassovou rovnicí $y^2z = x^3 + axz^2 + bz^3$ s význačným bodem $O = [0, 1, 0]$. Operaci $+$ na $\mathcal{E}$ je možné popsat takto:*

1. *Pro libovolné $A \in \mathcal{E}$ klademe $A + O = O + A = A$.*
2. *Pro libovolné $A = [\alpha, \beta, 1] \in \mathcal{E}$ je také $B = [\alpha, -\beta, 1] \in \mathcal{E}$ a klademe $A + B = O$. (Tento bod B pak označujeme $-A$.)*
3. *Pro libovolné $A = [\alpha, \beta, 1] \in \mathcal{E}$, $B = [\gamma, \delta, 1] \in \mathcal{E}$ takové, že $B \neq -A$, položme*

$$k = \begin{cases} \frac{\beta-\delta}{\alpha-\gamma} & \text{je-li } A \neq B, \\ \frac{3\alpha^2+a}{2\beta} & \text{je-li } A = B, \end{cases}$$
$$\sigma = k^2 - \alpha - \gamma,$$
$$\tau = -\beta + k(\alpha - \sigma),$$

pak platí $[\sigma, \tau, 1] \in \mathcal{E}$ a klademe $A + B = [\sigma, \tau, 1] \in \mathcal{E}$.

Věty o eliptických křivkách nad konečnými tělesy

Projektivní rovina nad konečným tělesem má konečně mnoho bodů, proto eliptická křivka nad konečným tělesem je konečná grupa.

Věta. (Hasse)

1. *Nechť p je prvočíslo a $(\mathcal{E}, O)$ je eliptická křivka nad $\mathbb{F}_p$. Pak $|\mathcal{E}| = p + 1 - a_p$, kde celé číslo a_p splňuje $|a_p| < 2\sqrt{p}$.*
2. *Označme $\alpha_p \in \mathbb{C}$ kořen rovnice $x^2 - a_p x + p = 0$. Pro libovolné $n \in \mathbb{N}$ nechť $(\mathcal{E}_n, O)$ je eliptická křivka nad $\mathbb{F}_{p^n}$ určená stejnou Weierstrassovou rovnicí jako $(\mathcal{E}, O)$ (to má smysl, neboť $\mathbb{F}_p \subseteq \mathbb{F}_{p^n}$). Pak platí $|\mathcal{E}_n| = p^n + 1 - 2\Re(\alpha_p^n)$, kde $\Re$ značí reálnou část komplexního čísla.*

Věta. Nechť $(\mathcal{E}, O)$ je eliptická křivka nad konečným tělesem $\mathbb{F}_q$, kde q je mocnina prvočísla. Pak $(\mathcal{E}, +)$ je cyklická grupa nebo součin dvou cyklických grup. Navíc, ve druhém případě, je-li $(\mathcal{E}, +)$ izomorfní se součinem cyklických grup o d_1 a d_2 prvcích, přičemž $d_1 \mid d_2$, pak platí $d_1 \mid q - 1$.

Věty o eliptických křivkách nad $\mathbb{Q}$

Věta. (Mordell) Nechť $(\mathcal{E}, O)$ je eliptická křivka nad $\mathbb{Q}$. Pak $(\mathcal{E}, O)$ je konečně generovaná grupa. Jinými slovy: označme $(\mathcal{E}', +)$ podgrupu prvků konečného řádu v grupě $(\mathcal{E}, +)$ (tzv. torzní podgrupa); pak existuje (jednoznačně určené) nezáporné celé číslo r tak, že $(\mathcal{E}, +)$ je izomorfní se součinem $(\mathcal{E}', +) \times (\mathbb{Z}, +)^r$.

Věta. (Mazur) Nechť $(\mathcal{E}, O)$ je eliptická křivka nad $\mathbb{Q}$. Pak její torzní podgrupa je izomorfní s některou z následujících 15 grup:

$$(\mathbb{Z}/m\mathbb{Z}, +) \qquad \text{pro } 1 \leq m \leq 10 \text{ nebo } m = 12$$
$$(\mathbb{Z}/2\mathbb{Z}, +) \times (\mathbb{Z}/2m\mathbb{Z}, +) \qquad \text{pro } 1 \leq m \leq 4$$

(a každá z uvedených grup je torzní grupa některé eliptické křivky nad $\mathbb{Q}$).

Proč si povídáme o eliptických křivkách?

Eliptické křivky se využívají v některých testech na prvočíselnost i v algoritmech hledání netriviálního dělitele.

Za tím účelem je třeba pracovat také s „eliptickými křivkami“ nad okruhem $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$ zbytkových tříd modulo N i v případě, že přirozené číslo N není prvočíslo. Ovšem projektivní prostor je definován jen nad tělesem, což v tomto případě $\mathbb{Z}_N$ není (proto ty uvozovky).

Proto budeme definovat pojem projektivního prostoru i nad okruhem $\mathbb{Z}_N$ pro libovolné přirozené číslo N.

Projektivní prostor nad okruhem $\mathbb{Z}_N$

Definice. Nechť N je přirozené číslo (ne nutně prvočíslo). Pak n-rozměrným projektivním prostorem nad okruhem $\mathbb{Z}_N$ rozumíme rozklad na následující množině (n+1)-tic zbytkových tříd modulo N

$$M = \{([a_1]_N, \ldots, [a_{n+1}]_N);\ a_1, \ldots, a_{n+1} \in \mathbb{Z},\ (N, a_1, \ldots, a_{n+1}) = 1\}$$

příslušný ekvivalenci $\sim$, kterou definujeme takto: pro libovolné $(n+1)$-tice $([a_1]_N, \ldots, [a_{n+1}]_N)$, $([b_1]_N, \ldots, [b_{n+1}]_N) \in M$ položíme $([a_1]_N, \ldots, [a_{n+1}]_N) \sim ([b_1]_N, \ldots, [b_{n+1}]_N)$ právě tehdy, když existuje $\lambda \in \mathbb{Z}$, $(\lambda, N) = 1$, které pro každé $i \in \{1, \ldots, n+1\}$ splňuje podmínku $[a_i]_N = [\lambda b_i]_N$.
V tomto n-rozměrném projektivním prostoru $P^n(\mathbb{Z}_N)$ nad $\mathbb{Z}_N$ budeme třídu rozkladu (tj. bod projektivního prostoru) obsahující $(n+1)$-tici $([a_1]_N, \ldots, [a_{n+1}]_N)$ značit $[[a_1]_N, \ldots, [a_{n+1}]_N]$.

Poznámka. Pro libovolné $d \mid N$ homomorfismus okruhů $\mathbb{Z}_N \to \mathbb{Z}_d$ určený předpisem $[a]_N \mapsto [a]_d$ pro každé $a \in \mathbb{Z}$ indukuje zobrazení n-rozměrných projektivních prostorů $P^n(\mathbb{Z}_N) \to P^n(\mathbb{Z}_d)$.

Test na prvočíselnost

Dáno přirozené číslo $N > 1$, o kterém jsme testem Millera a Rabina zjistili, že N je asi prvočíslo. Můžeme také předpokládat, že víme, že N není dělitelné malými prvočísly. Test na prvočíselnost má dokázat, že N skutečně prvočíslem je, anebo to vyvrátit.

Známe už $N - 1$ test Pocklingtona a Lehmera. Ten pracuje dobře, pokud jsme schopni dostatečně rozložit číslo $N - 1$. Pokud však neexistuje dost velký dělitel $F|N - 1$, který jsme schopni rozložit na prvočinitele, tato metoda neuspěje. Pak můžeme ještě zkusit $N + 1$ test, ten však vyžaduje rozložit dost velkého dělitele čísla $N + 1$, což se však často také nemusí podařit a skončíme nezdarem.

Řešení nabízí teorie eliptických křivek: je-li N skutečně prvočíslo, máme spoustu eliptických křivek nad $\mathbb{Z}_N$. Jejich řády jsou rovny přirozeným číslům v intervalu $(N + 1 - 2\sqrt{N}, N + 1 + 2\sqrt{N})$. Je pravděpodobné, že nezanedbatelnou část z těchto čísel budeme schopni rozložit na prvočinitele.

Síla metody eliptických křivek je v jejich počtu: pokud nevyhovuje několik konkrétních křivek, nevadí, vezmeme další.

Opakování $N-1$ testu Pocklingtona a Lehmera

Předpokládáme, že známe prvočíslo p dělící $N-1$, přitom p^{α_p} je nejvyšší mocnina p dělící $N-1$.
Dále označme d libovolné neznámé prvočíslo dělící N.
Máme homomorfismus okruhů $f : \mathbb{Z}_N \to \mathbb{Z}_d$, kde $f([a]_N) = [a]_d$ pro každé $a \in \mathbb{Z}$. Homomorfismus f je dobře definován, neboť $d \mid N$. Protože je d prvočíslo, je druhý okruh těleso $\mathbb{F}_d = \mathbb{Z}_d$.
Předpokládáme existenci $a_p \in \mathbb{Z}$, které splňuje

$$a_p^{N-1} \equiv 1 \pmod{N} \quad \text{a} \quad (a_p^{\frac{N-1}{p}} - 1, N) = 1.$$

Označme $b = f([a_p]_N) \in \mathbb{F}_d$. Pak $b^{N-1} = 1$, $b^{\frac{N-1}{p}} \neq 1$, a tedy řád prvku b je dělitelný p^{α_p}, odkud $p^{\alpha_p} \mid |\mathbb{F}_d^\times| = d-1$, tedy $d \equiv 1 \pmod{p^{\alpha_p}}$. Získali jsme tím informaci o neznámém d.

Klíčem k úspěchu zde byl homomorfismus $f : \mathbb{Z}_N \to \mathbb{Z}_d$.
Přestože jsme neznali d, a tedy nebyli schopni v $\mathbb{Z}_d$ pracovat, počítali jsme ve známém okruhu $\mathbb{Z}_N$ a výsledky výpočtů jsme do $\mathbb{Z}_d$ zobrazili homomorfismem f.

Test na prvočíselnost pomocí eliptických křivek

Přejděme k eliptickým křivkám, opět d značí libovolné neznámé prvočíslo dělící dané N, $(N, 6) = 1$. Zvolme libovolně $a, b \in \mathbb{Z}$ taková, že $(4a^3 + 27b^2, N) = 1$. Rovnice $y^2z = x^3 + axz^2 + bz^3$ nám dává „eliptickou křivku“ $\mathcal{E}_N$, na níž máme definovánu částečnou operaci, a eliptickou křivku $\mathcal{E}_d$, což je komutativní grupa. Z Hasseho věty víme, že $||\mathcal{E}_d| - d - 1| < 2\sqrt{d}$.
Přestože v $\mathcal{E}_d$ nejsme schopni počítat (vždyť neznáme d), máme částečný homomorfismus $f : \mathcal{E}_N \to \mathcal{E}_d$, kterým můžeme výpočet provedený v $\mathcal{E}_N$ zobrazit do $\mathcal{E}_d$. Víme, že $f([[0]_N, [1]_N, [0]_N]) = O$ a že pro libovolný $P = [[u]_N, [v]_N, [1]_N] \in \mathcal{E}_N$ platí $f(P) \neq O$.
Je-li q prvočíslo a bod $P = [[u]_N, [v]_N, [1]_N] \in \mathcal{E}_N$ takový, že máme definované $q \cdot P = P + P + \cdots + P = [[0]_N, [1]_N, [0]_N]$, pak řád bodu $f(P)$ v grupě $\mathcal{E}_d$ je q, a tedy $(\sqrt{d} + 1)^2 > |\mathcal{E}_d| \geq q$.
Najdeme-li takový bod P pro prvočíslo $q > (\sqrt[4]{N} + 1)^2$, plyne odtud $d > \sqrt{N}$, a tedy N je prvočíslo.

Problém je, jak volit čísla a, b a jak najít prvočíslo q a bod $P \in \mathcal{E}_N$ s potřebnými vlastnostmi...

Goldwasser - Kilian, 1986

Řešení navržené Goldwasserem a Kilianem má spíše teoretický význam; je možné dokázat, že platí-li jistá hypotéza o rozložení prvočísel v krátkých intervalech, pak očekávaný čas výpočtu je $O(\ln^{12} N)$, tedy polynomiální.

Existuje algoritmus Schoofa, který pro prvočíslo p počítá řád (tj. počet bodů) dané eliptické křivky nad $\mathbb{F}_p$ v čase $O(\ln^8 p)$.

Zvolíme náhodně a, $b \in \mathbb{Z}$ tak, aby $(4a^3 + 27b^2, N) = 1$. Pomocí Schoofova algoritmu určíme pro křivku $(\mathcal{E}, O)$ určenou rovnicí $y^2 = x^3 + ax + b$ a pro $p = N$ její řád m (jestliže N není prvočíslo, nemá m žádný význam). Získané m zkoušíme dělit malými prvočísly s nadějí, že poté, co odstraníme malé faktory, zůstane nám $q > (\sqrt[4]{N} + 1)^2$, $q < \frac{N}{2}$, o kterém test Millera a Rabina zjistí, že q je asi prvočíslo. Pokud se nám to nepodaří, začneme znovu s jinými a, $b \in \mathbb{Z}$.

Existuje algoritmus, který pro prvočíslo p a celé číslo e hledá v čase $O(\ln^4 p)$ řešení kongruence $x^2 \equiv e \pmod p$ a to, že takové řešení neexistuje, zjistí dokonce v čase $O(\ln^2 p)$.

Goldwasser - Kilian, 1986, pokračování

Najdeme bod P na křivce: náhodně zvolíme $c \in \mathbb{Z}_N$ a hledáme $d \in \mathbb{Z}_N$ tak, aby $d^2 = c^3 + ac + b$ (jde o kongruenci modulo N; d hledáme jako by bylo N prvočíslo, pak uděláme zkoušku, pokud nevyjde, nebylo N prvočíslo a jsme zcela hotovi). Neexistuje-li takové d, zkusíme jiné c. Pak za P zvolíme $\frac{m}{q}$-násobek bodu $[c, d, 1]$ v $(\mathcal{E}, +)$. Je-li $P = [0, 1, 0]$, zvolíme jiné c atd. Je-li $P \neq [0, 1, 0]$, pak platí $P = [x, y, 1]$ pro nějaké x, $y \in \mathbb{Z}_N$. Spočítáme q-násobek bodu P v $(\mathcal{E}, +)$. Není-li definován, našli jsme netriviálního dělitele čísla N. Jestliže nedostaneme $[0, 1, 0]$, není m řád křivky $(\mathcal{E}, O)$, Schoofův algoritmus tedy nedal správný výsledek a proto N není prvočíslo. Jestliže q-násobek bodu P je $[0, 1, 0]$, pak je N prvočíslo, pokud q je prvočíslo. To zjistíme rekurzívně ($N_0 = N$, N_1 je q pro N_0, N_2 je q pro N_1, ...). S rekurzí skončíme v okamžiku, kdy N_i je dost malé na to, abychom ověřili jeho prvočíselnost pokusným dělením (to nastane v $O(\ln N)$ krocích vzhledem k $N_{i+1} < \frac{1}{2} N_i$). Je třeba si uvědomit, že není-li N_i prvočíslo, skončíme jen v případě $i = 0$, pro $i < 0$ je třeba se vrátit k $i - 1$ a najít nové N_i.

Atkin, 1990

Tato metoda je založena na teoretických výsledcích, které bohužel notně převyšují možnosti naší přednášky. Nevolí křivky náhodně, ale volí speciální případ eliptických křivek, tzv. eliptické křivky s komplexním násobením. Výhoda metody je v tom, že je možné snadněji spočítat řád těchto křivek (vyhne se Schoofově algoritmu, který byl na předchozí metodě časově nejnáročnější).
Atkinův test byl implementován Atkinem a Morainem v roce 1990 a byl schopen dokazovat prvočíselnost čísel o zhruba 1000 dekadických cifrách v řádově týdnech strojového času na Sparc station (při tehdejší rychlosti počítačů, nyní by šlo o hodiny).
I v tomto případě je očekávaný čas výpočtu polynomiální (přesněji $O(\ln^6 N)$). Nejhorší možný čas výpočtu není možno stanovit, protože jde o pravděpodobnostní algoritmus.

Deterministický algoritmus AKS polynomiálního času objevili v roce 2002 pánové Agrawal, Kayal a Saxena z Kanpuru v Indii. Jejich algoritmus je založen na poměrně jednoduché myšlence a nepracuje s eliptickými křivkami. Avšak důkaz jeho polynomiálnosti vyžaduje výsledky analytické teorie čísel.

Funkce $\pi(x)$

Pro libovolné kladné reálné číslo x označme $\pi(x)$ počet prvočísel nepřevyšujících x. Je tedy
$\pi(x) = 0$ pro $x \in (0, 2)$,
$\pi(x) = 1$ pro $x \in [2, 3)$,
$\pi(x) = 2$ pro $x \in [3, 5)$,
$\pi(x) = 3$ pro $x \in [5, 7)$, atd.

Následující důležitou, hlubokou a slavnou větu uvedeme bez důkazu. Její formulaci objevil Gauss v 18. století, avšak důkaz nenašel.
Byla dokázána až na konci 19. století (v roce 1896 objevili důkaz nezávisle na sobě Hadamard a de la Vallée Poussin).
Připomeňme, že $\ln x$ značí přirozený logaritmus.

Věta.

$$\lim_{x \to \infty} \frac{\pi(x)}{\frac{x}{\ln x}} = 1.$$

Čebyševova věta

Pro účely důkazu polynomiálnosti algoritmu AKS bude stačit následující výsledek, který už budeme schopni dokázat.
Větu tohoto typu dokázal poprvé Čebyšev v roce 1852.

Věta (Čebyšev). Pro libovolné celé číslo $N \geq 2$ platí

$$\frac{N}{\log_2 N} - 2 < \pi(N) < \frac{3N}{\log_2 N}.$$

Pro reálné číslo x značí $[x]$ jeho celou část, která je jednoznačně určena podmínkami $[x] \in \mathbb{Z}$, $0 \leq x - [x] < 1$.

Dále pro libovolné přirozené číslo n a libovolné prvočíslo p je $\nu_p(n)$ počet prvočinitelů v rozkladu čísla n, které jsou rovny p, neboli platí $p^{\nu_p(n)} \mid n$ a $p^{1+\nu_p(n)} \nmid n$.

Je zřejmé, že pro libovolné $m, n \in \mathbb{N}$ platí $\nu_p(mn) = \nu_p(m) + \nu_p(n)$.

Lemma 1. *Pro libovolné přirozené číslo n a libovolné prvočíslo p platí*

$$\nu_p(n!) = \sum_{k=1}^{\infty} \left[\frac{n}{p^k} \right].$$

Důkaz. Nejprve si všimněme, že suma na pravé straně je jen formálně nekonečná: je-li $p^k > n$, platí $\left[\frac{n}{p^k}\right] = 0$.

Dále je třeba si uvědomit, že $\left[\frac{n}{p^k}\right]$ značí počet těch čísel z množiny $\{1, 2, \ldots, n\}$, která jsou dělitelná číslem p^k.

A odtud plyne i důkaz: nejprve (pro $k = 1$) započítáme jednou všechny ty činitele v $n! = 1 \cdot 2 \cdot \cdots \cdot n$, kteří jsou dělitelní p.
Pak (pro $k = 2$) započítáme podruhé všechny ty činitele, kteří jsou dělitelní p^2.
Poté (pro $k = 3$) započítáme potřetí všechny ty činitele, kteří jsou dělitelní p^3 atd.
Libovolný činitel s součinu $n! = 1 \cdot 2 \cdot \cdots \cdot n$ je tedy započítán právě $\nu_p(s)$krát a tedy pravá strana dokazované rovnosti je rovna $\sum_{s=1}^{n} \nu_p(s) = \nu_p(n!)$.

Lemma 2. *Pro libovolné přirozené číslo n a libovolné prvočíslo p platí: je-li $\ell = \nu_p\left(\binom{2n}{n}\right)$, pak $p^\ell \leq 2n$.*

Důkaz. Podle lemmatu 1 platí

$$\ell = \nu_p\left(\frac{(2n)!}{(n!)^2}\right) = \nu_p((2n)!) - 2\nu_p((n!)^2) = \sum_{k=1}^{\infty}\left(\left[\frac{2n}{p^k}\right] - 2\left[\frac{n}{p^k}\right]\right).$$

Pro libovolné reálné x takové, že $x - [x] < \frac{1}{2}$, platí $[2x] = 2[x]$. Je-li naopak $x - [x] \geq \frac{1}{2}$, platí $[2x] = 2[x] + 1$. Libovolný sčítanec v předchozí sumě je tedy 0 nebo 1. Přítom sčítance pro k takové, že $p^k > 2n$, jsou zřejmě nulové. Je tedy $\ell \leq \max\{k \in \mathbb{N};\ p^k \leq 2n\}$ a proto $p^\ell \leq 2n$.

Lemma 3. *Pro libovolná přirozená čísla n, k taková, že $1 \leq k \leq \frac{n}{2}$ platí $\binom{n}{k-1} < \binom{n}{k}$.*

Důkaz. Platí

$$\frac{\binom{n}{k}}{\binom{n}{k-1}} = \frac{n!}{k!(n-k)!} \cdot \frac{(k-1)!(n-k+1)!}{n!} = \frac{n-k+1}{k} \geq \frac{n/2+1}{n/2} > 1.$$

Lemma 4. *Pro libovolné přirozené číslo n platí $\binom{2n}{n} \leq (2n)^{\pi(2n)}$.*

Důkaz. Rozložme uvažovaný binomický koeficient na prvočinitele $\binom{2n}{n} = \frac{(2n)!}{(n!)^2} = p_1^{k_1} \dots p_r^{k_r}$. Libovolné prvočíslo p_i, které se zde vyskytuje, dělí $(2n)!$ a je tedy menší než $2n$. Proto $r \leq \pi(2n)$ a podle lemmatu 2 každé $p_i^{k_i} \leq 2n$. Odtud plyne lemma.

Lemma 5. *Pro libovolné přirozené číslo n platí $\frac{2^{2n}}{2n} \leq \binom{2n}{n} < 2^{2n}$.*

Důkaz. Z binomické věty víme, že $\sum_{i=0}^{2n} \binom{2n}{i} = (1+1)^{2n} = 2^{2n}$, odkud plyne pravá nerovnost.
Ukážeme-li, že v tomto součtu je sčítanec $\binom{2n}{n}$ největší, dostaneme i levou nerovnost, neboť $\frac{2^{2n}}{2n}$ je aritmetický průměr $2n$ čísel $\binom{2n}{0} + \binom{2n}{2n} = 2$, $\binom{2n}{1}$, $\binom{2n}{2}$, $\dots$, $\binom{2n}{2n-1}$.
Ale to je snadné: platí $\binom{2n}{2n-i} = \binom{2n}{i}$ a pro libovolné $1 \leq i \leq n$ platí $\binom{2n}{i-1} < \binom{2n}{i}$ podle lemmatu 3.

Dolní odhad z Čebyševovy věty: $\frac{N}{\log_2 N} - 2 < \pi(N)$

Z lemmat 4 a 5 plyne

$$(2n)^{\pi(2n)} \geq \binom{2n}{n} \geq \frac{2^{2n}}{2n},$$

odkud zlogaritmováním a vydělením $\log_2(2n)$ dostaneme

$$\pi(2n) \geq \frac{2n}{\log_2(2n)} - 1$$

a dolní odhad Čebyševovy věty je dokázán pro sudá $N = 2n$.
Je-li naopak $N = 2n + 1$ liché, užijeme odvozený odhad pro $\pi(2n)$:

$$\pi(2n+1) \geq \pi(2n) \geq \frac{2n}{\log_2(2n)} - 1 > \frac{2n}{\log_2(2n+1)} - 1 > \frac{2n+1}{\log_2(2n+1)} - 2,$$

což je dolní odhad Čebyševovy věty pro $N = 2n + 1$.

Lemma 6. *Pro libovolné přirozené číslo $N > 1$ platí*

$$\prod_{p \leq N} p < 4^{N-1},$$

kde v součinu p probíhá všechna prvočísla nepřevyšující N.

Důkaz. Pro přirozené číslo m označme
$b_m = \binom{2m+1}{m} = \frac{(2m+1)(2m)\ldots(m+2)}{m!}$. Je tedy b_m dělitelné všemi prvočísly p splňujícími $m+2 \leq p \leq 2m+1$, neboť tato prvočísla se vyskytují v čitateli a nedělí jmenovatele.
Proto $b_m \geq \prod_{m+2 \leq p \leq 2m+1} p$.
V součtu $\sum_{i=1}^{2m} \binom{2m+1}{i} = 2^{2m+1} - 2$ se sčítanec
$b_m = \binom{2m+1}{m} = \binom{2m+1}{m+1}$ objeví dvakrát, proto $b_m < 2^{2m}$.
Celkem tedy

$$\prod_{m+2 \leq p \leq 2m+1} p < 2^{2m}.$$

Dokazujeme: $\prod_{p\leq N} p < 4^{N-1}$

Víme: $\prod_{m+2\leq p\leq 2m+1} p < 2^{2m}$.

Nyní můžeme lemma dokázat indukcí: lemma zřejmě platí pro $N = 2$. Předpokládejme tedy, že $N \geq 3$ a že lemma bylo dokázáno pro všechna $2 \leq m < N$. Je-li N sudé, není N prvočíslo a z indukčního předpokladu pro $m = N - 1$ plyne

$$\prod_{p\leq N} p = \prod_{p\leq N-1} p < 4^{N-2} < 4^{N-1}.$$

Je-li naopak $N = 2m + 1$ liché, užijme indukční předpoklad pro $m + 1$ (vždyť $2 \leq m + 1 < N$) a odvozenou nerovnost

$$\prod_{p\leq N} p = \prod_{p\leq m+1} p \cdot \prod_{m+2\leq p\leq 2m+1} p < 4^m \cdot 4^m = 4^{N-1}.$$

Lemma 7. *Nechť $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, ... je rostoucí posloupnost všech prvočísel. Pak pro každé $k \geq 9$ platí $p_1 \dots p_k \geq 2^k \cdot k!$.*

Důkaz. Přímým výpočtem lze ověřit, že $p_1 \dots p_9 = 2 \cdot 3 \cdot 5 \cdots 19 \cdot 23 = 233092870 > 185794560 = 2^9 \cdot 9!$. Pro $k > 9$ užijeme indukci: předpokládejme, že $k \geq 9$ a že pro k lemma platí. Zřejmě $p_{k+1} > 2(k+1)$, a tedy

$$p_1 \dots p_{k+1} > 2^k \cdot k! \cdot 2(k+1) = 2^{k+1} \cdot (k+1)!,$$

což jsme měli dokázat.

Lemma 8. *Pro libovolné přirozené číslo k platí $k! > (k/e)^k$.*

Důkaz. Vzpomeňme si z analýzy na Taylorův rozvoj funkce e^x v nule:

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}.$$

Proto platí $\frac{k^k}{k!} < \sum_{i=0}^{\infty} \frac{k^i}{i!} = e^k$, odkud plyne lemma.

Horní odhad z Čebyševovy věty: $\pi(N) < \frac{3N}{\log_2 N}$

Ukážeme nyní sporem, že $\pi(N) < 2N/\ln N$. Pak totiž $3/\log_2 N = 3\ln 2/\ln N > 2,07/\ln N > 2/\ln N > \pi(N)/N$, což chceme ukázat. Předpokládejme, že $N \geq 27$ (případ $2 \leq N \leq 26$ se rychle ověří výpočtem) a že platí $\pi(N) \geq 2N/\ln N$.
Nechť $k = \pi(N)$, pak $p_1, \ldots, p_k$ jsou právě všechna prvočísla nepřevyšující N. Lemmata 6, 7 a 8 dávají

$$4^N > \prod_{p \leq N} p = p_1 \ldots p_k \geq 2^k \cdot k! > 2^k \cdot (\tfrac{k}{e})^k.$$

Zlogaritmováním

$$(2\ln 2) \cdot N > k \cdot ((\ln k) + (\ln 2) - 1).$$

Dosazením předpokladu $k \geq 2N/\ln N$ do předchozí nerovnosti dostaneme

$$(2\ln 2) \cdot N > \frac{2N}{\ln N} \cdot ((\ln 2) + (\ln N) - (\ln\ln N) + (\ln 2) - 1),$$

a tedy

$$(1 - \ln 2)\ln N - (\ln\ln N) + (2\ln 2) - 1 < 0.$$

Dostali jsme

$$(1 - \ln 2) \ln N - (\ln \ln N) + (2 \ln 2) - 1 < 0.$$

přičemž $N \geq 27$.

Ovšem funkce $f(x) = (1 - \ln 2) \ln x - (\ln \ln x) + (2 \ln 2) - 1$, která je definovaná pro $x > 1$, splňuje $f(27) > \frac{1}{5}$ a má derivaci $f'(x) = \frac{1-\ln 2}{x} - \frac{1}{x \ln x}$. Zřejmě $f'(x_0) = 0$ jedině pro $x_0 = e^{1/(1-\ln 2)} \doteq 26,02$ a platí $f'(x) > 0$ pro $x > x_0$.
Platí tedy $f(N) > 0$, ale to je spor a Čebyševova věta je dokázána.

Věta o rozložení prvočísel

Věta. *Pro libovolné přirozené číslo $n \geq 2$ platí $\prod_{p \leq 2n} p > 2^n$, kde v součinu p probíhá všechna prvočísla nepřevyšující $2n$.*

Důkaz. Jako v důkaze lemmatu 4 rozložme binomický koeficient $\binom{2n}{n}$ na prvočinitele $\binom{2n}{n} = p_1^{k_1} \dots p_r^{k_r}$. Víme, že libovolné prvočíslo, které se zde vyskytuje, je menší než $2n$. Je-li $p_i \leq \sqrt{2n}$, užijeme odhad $p_i^{k_i} \leq 2n$ z lemmatu 2. Je-li naopak $p_i > \sqrt{2n}$, platí $p_i^2 > 2n$, a odhad $p_i^{k_i} \leq 2n$ z lemmatu 2 dává $k_i = 1$. Užitím lemmatu 5

$$\frac{2^{2n}}{2n} \leq \binom{2n}{n} \leq \prod_{p \leq \sqrt{2n}} 2n \prod_{\sqrt{2n} < p \leq 2n} p.$$

Označme $s_n = \prod_{p \leq 2n} p$. Pak předchozí nerovnost spolu s Čebyševovovu větou dávají

$$\frac{2^{2n}}{2n} \leq (2n)^{\pi(\sqrt{2n})} \cdot s_n < (2n)^{3\sqrt{2n}/\log_2 \sqrt{2n}} \cdot s_n.$$

Dostali jsme

$$\frac{2^{2n}}{2n} \leq (2n)^{\pi(\sqrt{2n})} \cdot s_n < (2n)^{3\sqrt{2n}/\log_2 \sqrt{2n}} \cdot s_n.$$

Protože $(2n)^{1/\log_2 \sqrt{2n}} = (2n)^{2/\log_2 2n} = 2^2$, z poslední nerovnosti plyne

$$s_n > 2^{2n}/(2n \cdot 2^{6\sqrt{2n}}).$$

Abychom dokázali větu, musíme ukázat, že $2^n \geq 2n \cdot 2^{6\sqrt{2n}}$, neboli po zlogaritmování

$$n - 1 - \log_2 n - 6\sqrt{2n} \geq 0.$$

Uvažme funkci $f(x) = x - 1 - \log_2 x - 6\sqrt{2x}$. Platí $f(100) = 99 - \log_2 100 - 6\sqrt{200} > 7$ a derivace $f'(x) = 1 - \frac{1}{x \ln 2} - \frac{6}{\sqrt{2x}}$ je větší než $1 - \frac{1}{100 \ln 2} - \frac{6}{10\sqrt{2}} > 0$ pro $x \geq 100$. Tím jsme dokázali lemma pro $n \geq 100$. Nerovnost $s_n > 2^n$ pro hodnoty $2 \leq n < 100$ je možné ověřit numericky.

AKS test na prvočíselnost

M. Agrawal, N. Kayal a N. Saxena, Indian Institute of Technology Kanpur, Indie (2002)

První test na prvočíselnost, který je

- **obecný** - na vstupu může být libovolné přirozené číslo, ne jen čísla speciálního tvaru,
- **polynomiální** - čas výpočtu (nikoliv jen pravděpodobný, ale skutečný) je omezen polynomem v počtu cifer vstupu,
- **deterministický** - není pravděpodobnostní, v jeho průběhu se nic náhodně nevolí,
- **nepodmíněný** - správnost výstupu i polynomiálnost času výpočtu jsou dokázány, nejsou na ničem závislé (například na platnosti obecné Riemannovy hypotézy a podobně).

Základní myšlenka

Věta 1. *Nechť $a, n \in \mathbb{Z}$, $n > 1$, $(a, n) = 1$. Pak n je prvočíslo, právě když v okruhu $\mathbb{Z}_n[x]$, tj. okruhu polynomů nad okruhem zbytkových tříd modulo n, platí $(x + a)^n = x^n + a$.*

Důkaz. Z binomické věty $(x + a)^n = x^n + a^n + \sum_{i=1}^{n-1} \binom{n}{i} a^i x^{n-i}$.
Je-li n prvočíslo, pak z Fermatovy věty plyne $a^n \equiv a \pmod{n}$.
Dále pro libovolné $i = 1, 2, \ldots, n-1$ má binomický koeficient $\binom{n}{i} = \frac{n(n-1)\ldots(n-i+1)}{i!}$ prvočíslo n v čitateli a $n \nmid i!$, tedy $\binom{n}{i} \equiv 0 \pmod{n}$. Proto $(x + a)^n = x^n + a$.
Je-li naopak n složené číslo, zvolme prvočíslo p dělící n. Nechť $s = \nu_p(n)$, tj. přirozené číslo s je určené podmínkami $p^s \mid n$, $p^{s+1} \nmid n$. Pak koeficient u x^{n-p} v $(x + a)^n$ je

$$\binom{n}{p} a^p = \frac{n(n-1)\ldots(n-p+1)}{p!} \cdot a^p,$$

což není dělitelné p^s (vždyť $p \nmid a$ a $p \nmid (n-1)\ldots(n-p+1)$), a tedy ani n. To znamená $(x + a)^n \neq x^n + a$.

Využití věty 1

Věta 1 nabízí jednoduchou metodu na testování, zda je celé číslo n prvočíslo: zvolíme celé číslo a nesoudělné s n (například $a = 1$) a spočítáme pomocí rychlého umocňování v okruhu polynomů $\mathbb{Z}_n[x]$ mocninu $(x + a)^n$.

Tato metoda však není tak rychlá, jak se zdá na první pohled: v průběhu umocňování vzniká u polynomů, které jsou mezivýsledky, mnoho nenulových koeficientů. Vždyť stupeň polynomu, který má být naposledy umocňován na druhou, je nejméně $\frac{n-1}{2}$, a tedy může mít až $\frac{n+1}{2}$ nenulových koeficientů. To znamená, že počet prováděných operací nemůže být omezen shora ničím lepším než $O(n)$, a tedy tato metoda je horší než metoda pokusného dělení.

Efektivní využití věty 1

Místo rovnosti $(x+a)^n = x^n + a$ budeme kontrolovat jen kongruenci $(x+a)^n \equiv x^n + a \pmod{x^r - 1}$ pro vhodné r. Zbytek po dělení mocniny $(x+a)^n$ polynomem $x^r - 1$ pak spočítáme algoritmem rychlého umocňování, ale po každém násobení polynomů bude každá mocnina x^s nahrazena mocninou $x^{s'}$, kde s' je zbytek po dělení čísla s číslem r. Přitom pracujeme v $\mathbb{Z}_n[x]$ takto: počítáme s polynomy ze $\mathbb{Z}[x]$ a po každém provedeném výpočtu redukujeme celočíselné koeficienty modulo n. Složitost výpočtu bude polynomiální, jestliže $r = O((\log_2 n)^c)$ pro nějaké c. Je-li n prvočíslo, dávají $(x+a)^n$ a $x^n + a$ stejné zbytky po dělení polynomem $x^r - 1$, ať je r jakékoli.

Obtížné bylo dokázat, že pro libovolné n, které není mocninou prvočísla, existuje prvočíslo r (ohraničené polynomiálně), pro které $(x+a)^n$ a $x^n + a$ dávají různé zbytky po dělení $x^r - 1$ pro alespoň jednu hodnotu čísla a v jistém intervalu (jehož délka je opět ohraničena polynomiálně). To, že metoda „nepozná“ mocniny prvočísel, nevadí: tato n pozná jednoduchý polynomiální algoritmus, který provedeme hned na začátku metody.

Test na mocninu

Tímto testem bude AKS test začínat:

Algoritmus (Test na mocninu). *Pro dané celé číslo $n \geq 3$ algoritmus rozhodne, zda $n = a^b$, kde $a, b \in \mathbb{N}$, $b > 1$.*

1. *[Inicializace] Polož $b \leftarrow 2$, $a \leftarrow 1$, $c \leftarrow n$.*
2. *[Výpočet mocniny] Polož $m \leftarrow [\frac{a+c}{2}]$ a rychlým umocňováním spočti $d \leftarrow \min\{m^b, n+1\}$.*
3. *[Aktualizace mezí a, c] Je-li $d = n$, vytiskni zprávu, že $n = m^b$ je mocninou a skonči. Jinak, je-li $d < n$, polož $a \leftarrow m$, v opačném případě polož $c \leftarrow m$. Je-li $c - a \geq 2$, pokračuj bodem 2, jinak bodem 4.*
4. *[Zvýšení exponentu b] Nejmenší prvočíslo větší než b ulož do b. Je-li $2^b > n$, vytiskni zprávu, že n není mocninou a skonči. Jinak polož $a \leftarrow 1$, $c \leftarrow n$ a pokračuj bodem 2.*

Algoritmus je jistě správný: v průběhu výpočtu neustále platí $a^b < n < c^b$ a rozdíl $c - a$ se zmenšuje, dokud není $c - a = 1$.

Test na mocninu - odhad časové náročnosti

Výpočet mocniny v kroku 2 se provádí binárním umocňováním, jakmile se však v průběhu výpočtu objeví čísla větší než n, výpočet se přeruší a vrací se hodnota $n + 1$.

Protože pro dané b se rozdíl $c - a$ půlí při každém průchodu kroky 2 a 3, provedou se tyto kroky zhruba $\log_2 n$ krát. Rovněž počet kontrolovaných b je možné omezit shora číslem $\log_2 n$ (tato malá prvočísla budou uložena v tabulce, takže čas pro provedení kroku 4 je konstantní, jakmile se jednou provždy spočítá horní hranice $[\log_2 n]$ pro b).

V průběhu celého algoritmu je tedy třeba provést $O((\log_2 n)^2 \log_2 \log_2 n)$ násobení čísel menších než n, počet potřebných bitových operací lze odhadnout shora $O((\log_2 n)^4 \log_2 \log_2 n)$.

Algoritmus AKS

Algoritmus (Agrawal, Kayal, Saxena). *Pro dané přirozené číslo* $n > 1$ *algoritmus rozhodne, zda je* n *prvočíslo nebo složené.*

1. *[Mocniny] Pokud je* $n = a^b$, *kde* $a, b \in \mathbb{N}$, $b > 1$, *vytiskni, že* n *je složené a skonči. Jinak polož* $r \leftarrow 2$.
2. *[První cyklus] Jestliže* $r \geq n$, *pak vytiskni, že* n *je prvočíslo a skonči. Jestliže* $r|n$, *pak vytiskni, že* n *je složené a skonči. Jinak pro každé* i *od* 1 *do* $[4(\log_2 n)^2]$ *prověřuj: jestliže pro všechna taková* i *platí* $n^i \not\equiv 1 \pmod{r}$, *pokračuj krokem 3, jestliže naopak pro nějaké takové* i *platí* $n^i \equiv 1 \pmod{r}$, *pak nejmenší prvočíslo větší než* r *ulož do* r *a znovu prováděj krok 2.*
3. *[Druhý cyklus] Pro* a *od* 1 *do* $[2\sqrt{r}\log_2 n]$ *prováděj: jestliže pro některé takové* a *platí*

$$(x + a)^n \not\equiv (x^n + a) \pmod{x^r - 1} \text{ v } \mathbb{Z}_n[x],$$

   pak vytiskni, že n *je složené a skonči.*
4. *[Závěr] Vytiskni, že* n *je prvočíslo a skonči.*

Algoritmus AKS - důkaz správnosti algoritmu

Nejprve si promysleme, že nikdy na začátku kroku 2 nemůže být $r > n$. Protože r prochází postupně všechna prvočísla, znamenalo by to, že n je složené, ale pak by se algoritmus musel zastavit již dříve, když r se rovnalo nejmenšímu prvočíslu, které dělí n. Je tedy jasné, že pokud algoritmus skončí v kroku 1, 2 nebo 3, jistě odpoví správně. Zbývá dokázat, že i při zastavení v kroku 4 je odpověď správná.

Ve druhém kroku jsme hledali nejmenší prvočíslo r, pro které je řád čísla n modulo r větší než $4(\log_2 n)^2$.
Pokud jsme se dostali až do kroku 4, musí pro každé přirozené $a \leq 2\sqrt{r}\log_2 n$ platit $(x+a)^n \equiv (x^n + a) \pmod{x^r - 1}$ v $\mathbb{Z}_n[x]$.
Protože proběhl krok 2, víme, že n není dělitelné žádným prvočíslem menším nebo rovným r. Pak je n podle následující věty mocnina prvočísla.

Vzhledem k tomu, že proběhl krok 1, víme, že n není druhou nebo vyšší mocninou přirozeného čísla, a tedy n je prvočíslo a odpověď ve kroku 4 je správná.

Algoritmus AKS - důkaz správnosti algoritmu - věta

Věta 2. Nechť n a r jsou celá čísla splňující všechny následující podmínky:

(α) *r je prvočíslo a $r < n$;*

(β) *pro každé a splňující $2 \leq a \leq r$ platí $a \nmid n$;*

(γ) *řád čísla n modulo r je větší než $4(\log_2 n)^2$;*

(δ) *$(x+a)^n \equiv (x^n + a) \pmod{x^r - 1}$ v $\mathbb{Z}_n[x]$ pro všechna $1 \leq a \leq 2\sqrt{r}\log_2 n$.*

Pak n je mocninou prvočísla.

Důkaz provedeme později.

Pro důkaz polynomiální časové náročnosti algoritmu AKS potřebujeme pro každé celé $n \geq 2$ dokázat existenci „malého“ prvočísla r takového, že buď $r \mid n$ anebo (pokud $r \nmid n$) číslo n má modulo r řád větší než $4(\log_2 n)^2$.

Zde „malé“ znamená, že prvočíslo $r < f(\log_2 n)$ pro nějaký vhodný polynom f nezávisející na n. Následující věta ukáže, že tuto podmínku splní $f(x) = 20x^5$.

Algoritmus AKS - odhad časové náročnosti - věta

Věta 3. *Pro libovolné přirozené číslo $n \geq 2$ existuje prvočíslo $r \leq 20(\log_2 n)^5$ takové, že buď $r \mid n$ anebo platí $r \nmid n$ a současně řád čísla n modulo r je větší než $4(\log_2 n)^2$.*

Důkaz. Můžeme předpokládat, že $n \geq 4$, neboť pro menší n věta zřejmě platí. Označme $L = \log_2 n$ a $P = \prod_{i=1}^{[4L^2]}(n^i - 1)$. Zřejmě

$$P < \prod_{i=1}^{[4L^2]} n^i = n^{[4L^2][4L^2+1]/2} \leq 2^{L(4L^2)(4L^2+1)/2} \leq 2^{8L^5+2L^3}.$$

Z věty z úvodu přednášky plyne dolní odhad pro součin všech prvočísel p nepřevyšujících $20L^5$

$$\prod_{p \leq [20L^5]} p \geq \prod_{p \leq 2[10L^5]} p > 2^{[10L^5]} > 2^{10L^5-1}.$$

Ovšem $L \geq 2$ a tedy $2L^5 - 1 > 2L^3$, odkud $P < \prod_{p \leq [20L^5]} p$. Existuje tedy prvočíslo $r \leq [20L^5]$ takové, že $r \nmid P$, a tedy pro všechna přirozená čísla $i \leq 4L^2$ platí $r \nmid n^i - 1$. Pokud $r \nmid n$, je řád čísla n modulo r větší než $4L^2$ a jsme hotovi.

Odhad časové náročnosti vytvoření tabulky prvočísel

Potřebujeme tabulku prvočísel nepřevyšujících $20(\log_2 n)^5$. Máme-li připravit tabulku prvočísel menších než m pomocí Eratosthenova síta, sestavíme tabulku všech přirozených čísel od 2 do m a opakujeme toto: první neškrtlé číslo p vyznačíme jako prvočíslo a všechny jeho násobky počínaje $p \cdot p$ až po $p \cdot [\frac{m}{p}]$ škrtneme. To děláme až do doby, kdy je první neškrtlé číslo větší než $\sqrt{m}$; pak všechna zbylá neškrtlá čísla jsou prvočísla. Zřejmě $\int_{i-1}^{i} \frac{dx}{x} > 1/i$ (stačí funkci $1/x$ nahradit jejím minimem na tomto intervalu). Počet škrtání (a tedy i aritmetických operací) lze proto odhadnout shora číslem

$$\sum_{p \le \sqrt{m}} \frac{m}{p} < m \sum_{i=2}^{[\sqrt{m}]} \int_{i-1}^{i} \frac{dx}{x} = m \int_{1}^{[\sqrt{m}]} \frac{dx}{x} = m \ln [\sqrt{m}] \le \frac{m}{2} \ln m.$$

Počet bitových operací potřebných k tvorbě této tabulky je tedy $O(m(\log_2 m)^2)$. V našem případě je $m = 20(\log_2 n)^5$, a tedy časová náročnost tvorby tabulky v bitových operacích je $O((\log_2 n)^5 (\log_2 \log_2 n)^2)$.

Algoritmus AKS - odhad časové náročnosti

V kroku 2 pro každé r, kterých je $O((\log_2 n)^5)$, provádíme $O((\log_2 n)^2)$ násobení čísel nepřevyšujících r, časová náročnost kroku 2 v bitových operacích je proto $O((\log_2 n)^7(\log_2 \log_2 n)^2)$. V kroku 3 pro výpočet n-té mocniny v okruhu $\mathbb{Z}_n[x]/(x^r - 1)$ je zapotřebí $O(\log_2 n)$ okruhových násobení, která jsou prováděna jako násobení polynomů, jejichž stupeň je menší než r; každé takové okruhové násobení znamená $O(r^2)$ násobení a sčítání v $\mathbb{Z}_n$. Existují dokonce složitější algoritmy, které potřebují jen $O(r(\log_2 r)(\log_2 \log_2 r))$ operací (s větší O-konstantou). Časová náročnost obyčejného umocnění polynomu v bitových operacích je proto $O(r^2(\log_2 n)^2)$, těchto umocnění musíme provést celkem $O(\sqrt{r}\log_2 n)$. Časová náročnost kroku 3 v bitových operacích je $O(r^{5/2}(\log_2 n)^3)$, po dosazení $O((\log_2 n)^{31/2})$. Časová náročnost celého algoritmu v bitových operacích je proto $O((\log_2 n)^{31/2})$.

Pokud bychom užili v kroku 3 složitější algoritmus pro násobení polynomů, dosáhli bychom ještě lepšího výsledku $O((\log_2 n)^{21/2}(\log_2 \log_2 n)(\log_2 \log_2 \log_2 n))$.

Důkaz věty 2

Předpokládejme tedy, že celá čísla n a r splňují podmínky věty, a zvolme libovolné prvočíslo p dělící n. Je-li $p = n$, není co dokazovat, proto předpokládejme, že $p < n$, odkud plyne $p \leq \frac{n}{2}$. Označme $\ell = [2\sqrt{r}\log_2 n]$. Z podmínky (γ) ihned plyne $r > 4(\log_2 n)^2$, tj. $\sqrt{r} > 2\log_2 n$ a tedy z (β) dostáváme

$$p > r > \ell \qquad \text{a} \qquad r \nmid n. \tag{2}$$

Budeme se zabývat součiny mocnin polynomů $x + a \in \mathbb{F}_p[x]$ pro $1 \leq a \leq \ell$, zaveďme proto označení

$$P = \left\{ \prod_{a=1}^{\ell} (x+a)^{b_a};\ b_a \in \mathbb{Z},\ b_a \geq 0 \right\} \subseteq \mathbb{F}_p[x].$$

Pro stručnost vyjadřování zaveďme výrok $I(u, f)$ znamenající

$$u \in \mathbb{N},\ f \in \mathbb{F}_p[x],\ (f(x))^u \equiv f(x^u) \pmod{x^r - 1} \text{ v } \mathbb{F}_p[x].$$

Například pro $f = x + a$, kde $1 \leq a \leq \ell$, platí $I(n, f)$ díky $p \mid n$ a podmínce (δ) a současně platí též $I(p, f)$ díky větě 1.

$$I(u,f) \Leftrightarrow u \in \mathbb{N}, f \in \mathbb{F}_p[x], (f(x))^u \equiv f(x^u) \pmod{x^r - 1}$$

Lemma 1. *Z $I(u,f)$ a $I(v,f)$ plyne $I(uv,f)$.*

Důkaz. Umocněním kongruence z $I(u,f)$ dostáváme

$$(f(x))^{uv} \equiv (f(x^u))^v \pmod{x^r - 1}.$$

Dosazením x^u za x do kongruence z $I(v,f)$ dostáváme

$$(f(x^u))^v \equiv (f(x^{uv})) \pmod{x^{ur} - 1}.$$

Protože $x^r - 1 \mid x^{ur} - 1$, platí tato kongruence i modulo $x^r - 1$, a proto odtud plyne $I(uv,f)$.

Lemma 2. *Z $I(u,f)$ a $I(u,g)$ plyne $I(u,fg)$.*

Důkaz. Stačí vynásobit obě kongruence, které dostáváme z $I(u,f)$ a $I(u,g)$ a využít toho, že $(f \cdot g)(x^u) = f(x^u) \cdot g(x^u)$.

Důsledek. *Označme $U = \{n^i p^j;\ i,j \in \mathbb{Z},\ i \geq 0,\ j \geq 0\}$. Pak $I(u,f)$ platí pro všechna $f \in P$ a všechna $u \in U$.*

Konstrukce tělesa F

Polynom $x^{r-1} + x^{r-2} + \cdots + x + 1 \in \mathbb{F}_p[x]$ rozložme v $\mathbb{F}_p[x]$ na normované ireducibilní faktory; jeden z nich označme h.
Je tedy $h \in \mathbb{F}_p[x]$ normovaný ireducibilní polynom dělící $x^{r-1} + x^{r-2} + \cdots + x + 1$ a tedy i $x^r - 1$. Označme d stupeň polynomu h. Těleso $F = \mathbb{F}_p[x]/(h)$ má tedy p^d prvků a jeho prvek $\zeta = x + (h)$ je kořenem polynomu h, a tedy i polynomu $x^r - 1$.
Protože $p \nmid r$, není 1 kořenem polynomu $x^{r-1} + x^{r-2} + \cdots + x + 1$, a tedy $\zeta \neq 1$. Proto řád ζ v $F^\times$ je r.
Označme G množinu hodnot polynomů z P v ζ, tj.

$$G = \{f(\zeta);\ f \in P\} = \left\{ \prod_{a=1}^{\ell} (\zeta + a)^{b_a};\ b_a \in \mathbb{Z},\ b_a \geq 0 \right\} \subseteq F.$$

Lemma 3. *Pro* $1 \leq a \leq \ell$ *jsou* $x + a$ *různé polynomy z* $\mathbb{F}_p[x]$.

Důkaz. Je-li $1 \leq a < a' \leq \ell$, pak $0 < a' - a \leq \ell < p$ podle (2) a tedy skutečně a a a' jsou různé prvky tělesa $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$.

Lemma 4. *Pro každé $f \in P$ a každé $u \in U$ platí $f(\zeta)^u = f(\zeta^u)$.*

Důkaz. Z důsledku lemmat 1 a 2 víme, že existuje polynom $q \in \mathbb{F}_p[x]$ splňující

$$(f(x))^u = f(x^u) + (x^r - 1) \cdot q(x).$$

Dosazením ζ za x dostáváme dokazované.

Označme $T = \{\zeta^u;\ u \in U\} \subseteq F^\times$ a $t = |T|$.

Lemma 5. *Platí $r > t > 4(\log_2 n)^2$.*

Důkaz. Protože ζ má řád r, platí $T \subseteq \{1, \zeta, \ldots, \zeta^{r-1}\}$. Ovšem pro každé $u \in U$ platí $r \nmid u$ dle definice U a (2), a tedy $1 \notin T$. Proto $t < r$. Jistě $\zeta^{n^i} \in T$ pro každé $i \geq 0$. Protože ζ má řád r, platí $\zeta^{n^i} = \zeta^{n^j}$ právě tehdy, když $n^i \equiv n^j \pmod{r}$, což je ekvivalentní s $i \equiv j \pmod{e}$, kde e je řád čísla n modulo r. Proto $\zeta^{n^0}, \zeta^{n^1}, \ldots, \zeta^{n^{e-1}}$ jsou různé prvky T a předpoklad (γ) dává $t \geq e > 4(\log_2 n)^2$.

$T = \{\zeta^u;\ u \in U\} \subseteq F^\times,\ \ t = |T|$

Lemma 4. Pro každé $f \in P$ a každé $u \in U$ platí $f(\zeta)^u = f(\zeta^u)$.

Lemma 6. Jsou-li f_1 a f_2 různé polynomy z P a oba mají stupeň menší než t, pak $f_1(\zeta) \neq f_2(\zeta)$.

Důkaz. Předpokládejme naopak, že $f_1(\zeta) = f_2(\zeta)$. Pak pro každé $u \in U$ z lemmatu 4 plyne $f_1(\zeta^u) = f_1(\zeta)^u = f_2(\zeta)^u = f_2(\zeta^u)$, a tedy libovolný prvek z T je kořenem polynomu $f_1 - f_2$. Tento polynom má tedy alespoň t kořenů a stupeň menší než t, proto $f_1 = f_2$.

$P = \{\prod_{a=1}^{\ell}(x+a)^{b_a};\ b_a \geq 0\}$, $G = \{f(\zeta);\ f \in P\} \subseteq F$

Lemma 5. *Platí $r > t > 4(\log_2 n)^2$.*

Lemma 6. *Jsou-li f_1 a f_2 různé polynomy z P a oba mají stupeň menší než t, pak $f_1(\zeta) \neq f_2(\zeta)$.*

Lemma 7. *Platí $|G| > \frac{1}{2}n^{2\sqrt{t}}$.*

Důkaz. Nechť $\mu = \min\{\ell, t-1\}$. Z věty o jednoznačném rozkladu polynomů v $\mathbb{F}_p[x]$ na ireducibilní faktory a z lemmatu 3 plyne, že $\prod_{a=1}^{\mu}(x+a)^{b_a}$, kde $b_a \in \{0,1\}$, jsou různé polynomy z P stupně menšího než t. Podle lemmatu 6 jsou jejich funkční hodnoty v ζ různé a z toho plyne odhad $|G| \geq 2^{\mu}$. Jsou dvě možnosti: je-li $\mu = t-1$, platí díky odhadu $t > 4(\log_2 n)^2$ z lemmatu 5

$$\mu = t - 1 > 2\sqrt{t}\log_2 n - 1.$$

Je-li naopak $\mu = \ell$, platí díky odhadu $r > t$ z lemmatu 5

$$\mu = [2\sqrt{r}\log_2 n] > 2\sqrt{r}\log_2 n - 1 > 2\sqrt{t}\log_2 n - 1.$$

V obou případech dostáváme $|G| \geq 2^{\mu} > 2^{2\sqrt{t}\log_2 n - 1} = \frac{1}{2}n^{2\sqrt{t}}$ a lemma je dokázáno.

$G = \{f(\zeta);\ f \in P\} \subseteq F$

Lemma 4. Pro každé $f \in P$ a každé $u \in U$ platí $f(\zeta)^u = f(\zeta^u)$.

Lemma 7. Platí $|G| > \frac{1}{2}n^{2\sqrt{t}}$.

Označme $U_0 = \{n^i p^j;\ i, j \in \mathbb{Z},\ 0 \leq i \leq [\sqrt{t}],\ 0 \leq j \leq [\sqrt{t}]\} \subseteq U$.

Lemma 8. Pro různá $u, v \in U_0$ platí $\zeta^u \neq \zeta^v$.

Důkaz. Z $p \leq \frac{n}{2}$ plyne $np \leq \frac{1}{2}n^2$, a tedy pro každé $u \in U_0$ je $u \leq (\frac{1}{2}n^2)^{\sqrt{t}} \leq \frac{1}{2}n^{2\sqrt{t}} < |G|$ podle lemmatu 7.
Sporem: předpokládejme, že pro různá $u, v \in U_0$ platí $\zeta^u = \zeta^v$. Libovolné $g \in G$ je tvaru $g = f(\zeta)$ pro nějaké $f \in P$. Podle lemmatu 4 platí $g^u = f(\zeta)^u = f(\zeta^u) = f(\zeta^v) = f(\zeta)^v = g^v$ a tedy každé $g \in G$ je kořenem polynomu $x^u - x^v$. Na začátku tohoto důkazu jsme ukázali, že u a v jsou menší než $|G|$. Ovšem $u \neq v$, a tedy nenulový polynom $x^u - x^v$ má více kořenů než je jeho stupeň. To je spor.

Dokončení důkazu věty 2

$T = \{\zeta^u;\ u \in U\} \subseteq F^\times, \quad t = |T|$

$U_0 = \{n^i p^j;\ i, j \in \mathbb{Z},\ 0 \le i \le [\sqrt{t}],\ 0 \le j \le [\sqrt{t}]\} \subseteq U$

Lemma 8. Pro různá $u, v \in U_0$ platí $\zeta^u \neq \zeta^v$.

Počet dvojic (i, j), kde $i, j \in \mathbb{Z}$, $0 \le i \le [\sqrt{t}]$, $0 \le j \le [\sqrt{t}]$, je roven $([\sqrt{t}] + 1)^2 > \sqrt{t}^2 = t$, na druhou stranu z lemmatu 8 plyne $|U_0| \le |T| = t$. Znamená to, že existují různé dvojice (i, j) a (k, m) takové, že $i, j, k, m \in \{0, 1, \ldots, [\sqrt{t}]\}$ a že $n^i p^j = n^k p^m$. Lze navíc předpokládat, že $i \ge k$. Kdyby $i = k$, muselo by platit i $j = m$ a dvojice by nebyly různé. Je tedy $i > k$ a platí $n^{i-k} = p^{m-j}$. Odtud plyne, že v rozkladu čísla n na prvočinitele se nevyskytují jiná prvočísla než p, a tedy n je mocninou prvočísla p. Věta 2 je dokázána.

Hledání netriviálního dělitele

Předpokládejme, že máme dáno přirozené číslo N, o němž víme, že je složené. Naším úkolem je nalézt netriviálního dělitele čísla N.

Odhadněme nejprve časovou náročnost metody pokusného dělení: je třeba číslo N postupně vydělit všemi prvočísly nepřevyšujícími $\sqrt{N}$. Každé takové dělení zabere čas řádu $O(\ln^2 N)$, celá metoda je tedy řádu $O(N^{\frac{1}{2}} \ln^2 N)$.

První metoda, jejíž čas je lepší než právě uvedený, byla navržena Lehmannem. Je založena na následující větě.

Lehmannova metoda

Věta (Lehmann). *Nechť N je liché přirozené číslo, $N = pq$, kde $\sqrt[3]{N} < p \leq q$ jsou prvočísla. Pak existují $x, y, k \in \mathbb{Z}$ splňující*

(1) $x^2 - y^2 = 4kN, \quad 1 \leq k \leq [\sqrt[3]{N}];$

(2) $2|k \;\Rightarrow\; x \equiv 1 \pmod 2; \qquad 2 \nmid k \;\Rightarrow\; x \equiv k + N \pmod 4;$

(3) $0 \leq x^2 - 4kN < \frac{N}{[\sqrt[3]{N}]}, \quad x > 0.$

Jestliže naopak pro dané liché přirozené číslo $N = pq$, kde p, q jsou prvočísla, máme celá čísla x, y, k splňující podmínky (1), (2) a (3), pak jeden z největších společných dělitelů $(x + y, N)$ a $(x - y, N)$ je roven p a druhý q.

Rovněž platí, že je-li N liché prvočíslo, pak žádná trojice celých čísel x, y, k splňujících podmínky (1), (2) a (3) neexistuje.

Důkaz. Později si ukážeme důkaz pomocí teorie dobrých aproximací, který objevil Don Zagier.

Použití věty

Mějme dáno liché přirozené číslo N, o kterém je známo, že to není prvočíslo. Metodou pokusného dělení ověříme, že N není dělitelné prvočísly nepřevyšujícími $\sqrt[3]{N}$, anebo najdeme netriviálního dělitele. Tato část algoritmu je tedy řádu $O(N^{\frac{1}{3}} \ln^2 N)$. Pokud N nemá prvočíselného dělitele menšího než $\sqrt[3]{N}$, musí být tvaru $N = pq$, kde p, q jsou prvočísla.
Budeme pak postupně volit $k \in \{1, 2, \ldots, [\sqrt[3]{N}]\}$ a pro každé takové k necháme x proběhnout všechna celá čísla splňující podmínky (2) a (3) z předchozí věty. Pro každé takové x pak testujeme, zda $x^2 - 4kN$ je druhá mocnina přirozeného čísla. Pokud ano, označíme $y = \sqrt{x^2 - 4kN}$ a spočítáme $(x + y, N)$, což je p nebo q. Je jasné, že časová náročnost algoritmu závisí na tom, jak rychle jsme schopni rozhodnout, zda přirozené číslo je nebo není druhou mocninou. Cesta vedoucí přes výpočet reálné odmocniny, zaokrouhlení a zkoušku jistě není ta pravá.

Algoritmus (Celočíselná druhá odmocnina). *Pro dané přirozené číslo n algoritmus najde přirozené číslo m splňující $m^2 \leq n < (m+1)^2$.*

1. *[Inicializace] Polož $x \leftarrow n$ (viz též diskusi za algoritmem).*
2. *[Krok] Pomocí celočíselného dělení a posunu spočítej $y \leftarrow [(x + [\frac{n}{x}])/2]$.*
3. *[Konec?] Je-li $y < x$, polož $x \leftarrow y$ a jdi na 2. Jinak vytiskni x a skonči.*

Důkaz algoritmu. Podle kroku 3 hodnota proměnné x klesá, algoritmus se tedy zastaví. Ukažme, že výsledek, který dává, je správný. Protože $x \in \mathbb{Z}$, je $[(x + [\frac{n}{x}])/2] \leq (x + [\frac{n}{x}])/2 \leq (x + \frac{n}{x})/2 < (x + [\frac{n}{x}] + 1)/2 \leq [(x + [\frac{n}{x}])/2] + 1$, a tedy platí $[(x + [\frac{n}{x}])/2] = [(x + \frac{n}{x})/2]$. Označme $q = [\sqrt{n}]$. Protože $\frac{1}{t}(t - \sqrt{n})^2 \geq 0$ pro libovolné $t > 0$, platí $\frac{1}{2}(t + \frac{n}{t}) \geq \sqrt{n}$, tedy $x \geq q$ je splněno v průběhu celého algoritmu. Předpokládejme, že se algoritmus zastavil, tj. že $y = [(x + \frac{n}{x})/2] \geq x$ a dokažme $x = q$. Předpokládejme $x \geq q + 1$. Pak $x > \sqrt{n}$ a platí

$$y - x = \left[\tfrac{1}{2}(x + \tfrac{n}{x})\right] - x = \left[\tfrac{1}{2}(\tfrac{n}{x} - x)\right] = \left[\tfrac{1}{2x}(n - x^2)\right] < 0, \quad \text{spor.}$$

Časová náročnost celočíselné odmocniny

V kroku 1 je jistě výhodnější místo n zvolit číslo bližší $\sqrt{n}$. Vhodné může být např. zjistit řád e nejvyšší dvojkové cifry n, tj. přirozené číslo e splňující $2^e \leq n < 2^{e+1}$ a položit $x \leftarrow 2^{1+[\frac{e}{2}]}$. Pak totiž $x^2 \leq 2^{e+2} \leq 4n$, $x^2 \geq 2^{e+1} > n$, tj. $\sqrt{n} < x \leq 2\sqrt{n}$. Po provedení kroku 2 pak platí

$$\begin{aligned} x - y &= -\left[\tfrac{1}{2x}(n - x^2)\right] \geq -\tfrac{1}{2x}(n - x^2) = \\ &= \tfrac{1}{2x}(x + \sqrt{n})(x - \sqrt{n}) \geq \tfrac{1}{2x}(x + \tfrac{x}{2})(x - \sqrt{n}) = \tfrac{3}{4}(x - \sqrt{n}). \end{aligned}$$

V každém dalším provedení kroku 3 se hodnota $x - \sqrt{n}$ zmenší alespoň čtyřikrát, neboť $y - \sqrt{n} = (x - \sqrt{n}) - (x - y) \leq \frac{1}{4}(x - \sqrt{n})$ a tedy krok 3 provádíme řádově $O(\ln n)$-krát. Protože celočíselné dělení je řádu $O(\ln^2 n)$, je celý algoritmus řádu $O(\ln^3 n)$.

Zkrácení času výpočtu

Pokud nás, podobně jako v případě Lehmannova algoritmu, zajímá jen to, zda n je či není druhou mocninou přirozeného čísla, je možné rozhodování zrychlit: zjistíme, zda je n kvadratickým zbytkem modulo nějaké zvolené číslo m (tj. zda má řešení kongruence $x^2 \equiv n \pmod{m}$ – pokud n je druhou mocninou přirozeného čísla, tato kongruence řešení mít musí). Budeme postupovat takto: vydělíme číslo n číslem m se zbytkem a získaný zbytek porovnáme s tabulkou všech kvadratických zbytků modulo m, kterou budeme mít předem spočítánu v paměti. Vhodným modulem může být například číslo $1989 = 3^2 \cdot 13 \cdot 17$ nebo $1925 = 5^2 \cdot 7 \cdot 11$. Pravděpodobnost, že náhodně zvolené přirozené číslo je kvadratický zbytek modulo 1925, je $\frac{11}{25} \cdot \frac{4}{7} \cdot \frac{6}{11} = \frac{24}{175}$, pro modul 1989 dokonce je $\frac{4}{9} \cdot \frac{7}{13} \cdot \frac{9}{17} = \frac{28}{221}$. Provedeme-li test pro oba moduly, poběží předchozí algoritmus jen s pravděpodobností $\frac{96}{5525}$, tedy jen asi v 1,7% případů.

Toto vylepšení nebude mít vliv na asymptotickou časovou náročnost, sníží však významně O-konstantu.

Algoritmus (Naplnění tabulek kvadratických zbytků).
Algoritmus sestaví vektory T_1 *o délce* 1989 *a* T_2 *o délce* 1925 *tak, že pro každé* $0 \leq i \leq 1988$ *platí* $T_1[i] = 1$, *právě když kongruence* $x^2 \equiv i \pmod{1989}$ *má řešení, a pro každé* $0 \leq i \leq 1924$ *platí* $T_2[i] = 1$, *právě když kongruence* $x^2 \equiv i \pmod{1925}$ *má řešení.*

1. *[Naplň* T_1*] Pro* i *od* 0 *po* 1988 *polož* $T_1[i] \leftarrow 0$. *Pak pro* i *od* 0 *po* 994 *polož* $T_1[i^2 \bmod 1989] \leftarrow 1$.
2. *[Naplň* T_2*] Pro* i *od* 0 *po* 1924 *polož* $T_2[i] \leftarrow 0$. *Pak pro* i *od* 0 *po* 962 *polož* $T_2[i^2 \bmod 1925] \leftarrow 1$.

Algoritmus (Test na čtverec). *Pro dané přirozené číslo n algoritmus zjistí, zda je n druhá mocnina přirozeného čísla, a pokud ano, vytiskne $\sqrt{n}$.*

1. *[Test na 1989] Polož $r \leftarrow n \bmod 1989$. Je-li $T_1[r] = 0$, odpověz, že n není druhá mocnina přirozeného čísla a skonči.*
2. *[Test na 1925] Polož $r \leftarrow n \bmod 1925$. Je-li $T_2[r] = 0$, odpověz, že n není druhá mocnina přirozeného čísla a skonči.*
3. *[Spočítej odmocninu] Algoritmem celočíselné druhé odmocniny spočítej $m = \lfloor \sqrt{n} \rfloor$. Je-li $n \neq m^2$, odpověz, že n není druhá mocnina přirozeného čísla a skonči. Jinak odpověz, že n je druhá mocnina přirozeného čísla m a skonči.*

Časová náročnost Lehmannova algoritmu

Odhadněme počet hodnot, které musíme za x dosazovat pro pevně zvolené $k \in \{1, 2, \ldots, [\sqrt[3]{N}]\}$. Odhadneme-li $x + 2\sqrt{kN} \geq 4\sqrt{kN}$ pomocí (3) ve výrazu

$$x - 2\sqrt{kN} = \frac{x^2 - 4kN}{x + 2\sqrt{kN}} < \frac{N}{[\sqrt[3]{N}]} \cdot \frac{1}{4\sqrt{kN}} = \frac{1}{4[\sqrt[3]{N}]} \cdot \sqrt{\frac{N}{k}},$$

dostáváme, že x splňuje

$$2\sqrt{kN} \leq x < 2\sqrt{kN} + \frac{1}{4[\sqrt[3]{N}]} \cdot \sqrt{\frac{N}{k}},$$

patří tedy x do intervalu délky $\frac{1}{4[\sqrt[3]{N}]}\sqrt{\frac{N}{k}}$. Délka intervalu je řádu $O(k^{-\frac{1}{2}} N^{\frac{1}{6}})$, pro pevné k je tedy časová náročnost algoritmu řádu $O(k^{-\frac{1}{2}} N^{\frac{1}{6}} \ln^3 N)$.

Sečtením přes všechna k dostáváme, že celková časová náročnost je řádu

$$O\Big(N^{\frac{1}{6}} \ln^3 N \sum_{k=1}^{[\sqrt[3]{N}]} k^{-\frac{1}{2}}\Big).$$

Přitom $\int_1^r k^{-\frac{1}{2}}\,dk = [2k^{\frac{1}{2}}]_1^r = 2\sqrt{r} - 2$, volbou $r = \sqrt[3]{N}$ upravíme řád časové náročnosti hledání čísel k, x, y do tvaru

$$O\Big(N^{\frac{1}{6}} \ln^3 N \cdot \sqrt{N^{\frac{1}{3}}}\Big) = O(N^{\frac{1}{3}} \ln^3 N).$$

Protože časová náročnost první části algoritmu, totiž metody pokusného dělení čísly nepřevyšujícími $\sqrt[3]{N}$, je řádu $O(N^{\frac{1}{3}} \ln^2 N)$, je celková časová náročnost Lehmannova algoritmu řádu $O(N^{\frac{1}{3}} \ln^3 N)$. Lehmannova metoda je tedy asymptoticky výrazně lepší než algoritmus pokusného dělení, jehož časová náročnost je řádu $O(N^{\frac{1}{2}} \ln^2 N)$.

Další metoda hledání netriviálního dělitele: Pollardova ρ metoda

Předpokládejme, že M je konečná množina a $f : M \to M$ zobrazení. Zvolme $x_0 \in M$ a pro každé $n \in \mathbb{N}$ položme $x_n = f(x_{n-1})$. Protože je M konečná, v posloupnosti $(x_n)_{n=0}^{\infty}$ nemohou být všechny prvky různé.

Nechť $i \in \mathbb{N} \cup \{0\}$ je nejmenší index, pro který existuje nějaký index $n > i$ s vlastností $x_i = x_n$. Dále označme j nejmenší takové n. Pak i nazýváme předperioda a $j - i$ perioda posloupnosti $(x_n)_{n=0}^{\infty}$.

Je možné dokázat, že střední hodnota předperiody i periody (mají-li všechny dvojice $(x_0, f) \in M \times M^M$ stejnou pravděpodobnost) je řádu $O(\sqrt{|M|})$.

Základní myšlenka Pollardovy ρ metody

Nechť $f(x)$ je mnohočlen s celými koeficienty. Hledáme (neznámého) prvočíselného dělitele přirozeného čísla N, o kterém víme, že je složené. Zvolme celé číslo x_0 a počítejme $x_n = f(x_{n-1}) \mod N$. Pak ovšem $y_n = x_n \mod p$ vyhovuje téže rekurzi modulo p. Pokud se f chová jako náhodné zobrazení (což nevíme, ale budeme to předpokládat), je předperioda a perioda posloupnosti $(y_n)_{n=0}^{\infty}$ řádu $O(\sqrt{p})$, kdežto předperioda a perioda posloupnosti $(x_n)_{n=0}^{\infty}$ je řádu řádu $O(\sqrt{N})$. Dá se tedy čekat, že existují $i < j$ tak, že $y_i = y_j$, ale $x_i \neq x_j$. Pak ovšem je $(x_i - x_j, N)$ netriviální dělitel čísla N.

Je nutné nějak zvolit x_0 a f. Volba x_0 se zdá být nepodstatná, ne však volba f. Je vhodné, aby f byl jednoduchý polynom pro výpočet, konstantní ani lineární však nejsou vhodné. Budeme tedy volit f jako co nejjednodušší kvadratický polynom. Je ověřeno experimentálně, že polynomy $f = x^2$ a $f = x^2 - 2$ nejsou vhodné, kdežto $f = x^2 + c$, kde $c \neq 0$ a $c \neq -2$, pracuje docela dobře, i když nejsme schopni určit ani periodu ani předperiodu.

Implementace Pollardovy ρ metody

Je jasné, že uchovávání všech již vypočtených členů posloupnosti $(x_n)_{n=0}^{\infty}$ a jejich neustálé porovnávání s nově vypočtenou hodnotou by bylo velmi zdlouhavé. Jednoduchou metodou, jak se tomuto zdlouhavému výpočtu vyhnout, je porovnávat postupně x_n a x_{2n}. Pak totiž prvočíselného dělitele p čísla N objevíme nejpozději po k krocích, kde k je součet předperiody a periody posloupnosti modulo p. Znamená to počítat iterace dvou posloupností: položit $z_0 = x_0$, iterovat $x_n = f(x_{n-1}) \mod N$ a $z_n = f(f(z_{n-1})) \mod N$ a počítat $(x_n - z_n, N)$.

Za (nedokázaného) předpokladu, že f se chová jako náhodné zobrazení, je počet nutných kroků $O(\sqrt{p})$. V každém kroku počítáme třikrát f, dvakrát zbytek po dělení N a jednou největší společný dělitel, vše je $O(\ln^2 N)$ Celková časová náročnost je tedy $O(\sqrt{p} \ln^2 N)$, což vzhledem k $p \leq \sqrt{N}$ dává $O(\sqrt[4]{N} \ln^2 N)$. Je vhodné si uvědomit, že podobně jako metoda postupného dělení je i tato metoda citlivá k velikosti prvočíselných dělitelů – „malé“ dělitele čísla N odstraňuje rychleji než „velké“.

Další metoda hledání netriviálního dělitele: Pollardova $p-1$ metoda

Tato metoda je schopna najít i značně velké prvočíselné dělitele p čísla N, pokud $p-1$ není dělitelné příliš velkou mocninou prvočísla.

Definice. Nechť B je přirozené číslo. Řekneme, že přirozené číslo n je B-hladké, jestliže pro libovolné prvočíslo p a libovolné přirozené číslo k platí

$$p^k \mid n \quad \Rightarrow \quad p^k \leq B.$$

Příklad. Víme, že pro každé $n \in \mathbb{N}$ platí, že $\binom{2n}{n}$ je $2n$-hladké. Dokázali jsme totiž už dříve, že platí

Lemma 2. Pro libovolné přirozené číslo n a libovolné prvočíslo p platí: je-li $\ell = \nu_p(\binom{2n}{n})$, pak $p^\ell \leq 2n$.

Základní myšlenka Pollardovy $p-1$ metody

Přepokládejme, že pro nějaký prvočíselný dělitel p čísla N platí, že číslo $p-1$ je B-hladké pro nějaké nepříliš velké přirozené číslo B. Zvolme libovolně $1 < a < N$. Je-li $(a, N) > 1$, jsme hotovi. Budeme proto předpokládat, že $(a, N) = 1$. Pak podle definice číslo $p-1$ dělí nejmenší společný násobek L_B čísel $1, 2, 3, \ldots, B$. Z Fermatovy věty pak plyne $a^{L_B} \equiv 1 \pmod{p}$ a tedy $(a^{L_B} - 1, N) > 1$. Budeme tedy testovat poslední podmínku pro zvyšující se hodnoty exponentu $e \mid L_B$ (budeme postupně umocňovat na faktory z kanonického rozkladu čísla L_B). Je velmi nepravděpodobné, že poprvé, kdy platí $(a^e - 1, N) > 1$, je tento největší společný dělitel roven N. Může se ovšem stejně stát, že metoda selže, jestliže pro žádné prvočíslo $p \mid N$ číslo $p-1$ není B-hladké.

Při výpočtu zabere nejvíce času výpočet největšího společného dělitele, proto budeme postupovat tak, že budeme uchovávat součiny a počítat největší společný dělitel jen čas od času.

Algoritmus (Pollardova $p-1$ metoda, první stadium). *Dáno složené N a hranice B, hledáme netriviálního dělitele N. Máme tabulku $p[1], p[2], \ldots, p[k]$ všech prvočísel menších nebo rovných B.*

1. *[Inicializace] Polož $x \leftarrow 2$, $y \leftarrow x$, $P \leftarrow 1$, $c \leftarrow 0$, $i \leftarrow 0$, $j \leftarrow i$.*
2. *[Další prvočíslo] Polož $i \leftarrow i+1$. Je-li $i > k$, spočti největší společný dělitel $g \leftarrow (P, N)$. Je-li $g = 1$, napiš, že jsi neuspěl a skonči, jinak polož $i \leftarrow j$, $x \leftarrow y$ a jdi na 5. Jinak (tj. pro $i \leq k$) polož $q \leftarrow p[i]$, $q_1 \leftarrow q$, $\ell \leftarrow [\frac{B}{q}]$.*
3. *[Spočti mocninu] Dokud $q_1 \leq \ell$, dělej $q_1 \leftarrow q_1 \cdot q$. Pak polož $x \leftarrow x^{q_1} \bmod N$, $P \leftarrow P \cdot (x-1) \bmod N$, $c \leftarrow c+1$ a je-li $c < 20$, jdi na 2.*
4. *[Největší společný dělitel] Polož $g \leftarrow (P, N)$. Je-li $g = 1$, polož $c \leftarrow 0$, $j \leftarrow i$, $y \leftarrow x$ a jdi na 2. Jinak polož $i \leftarrow j$, $x \leftarrow y$.*
5. *[Počítej znovu] Polož $i \leftarrow i+1$, $q \leftarrow p[i]$, $q_1 \leftarrow q$.*
6. *[Skončil jsi?] Polož $x \leftarrow x^q \bmod N$, $g \leftarrow (x-1, N)$. Je-li $g = 1$, polož $q_1 \leftarrow q \cdot q_1$ a je-li $q_1 < B$, jdi na 6, jinak jdi na 5. Jinak (tj. pro $g > 1$), je-li $g < N$, vytiskni g a skonči. Konečně, je-li $g = N$ (velmi nepravděpodobné), napiš, že jsi neuspěl a skonči.*

Pokud algoritmus selhal v bodě 6, znamená to, že všechna prvočísla p dělící N byla nalezena současně, což je značně nepravděpodobné. Může proto mít smysl zkusit tentýž algoritmus s jinou počáteční hodnotou (např. $x \leftarrow 3$).

I v této jednoduché formě jsou výsledky algoritmu působivé. Samozřejmě, jsou-li $p < q$ prvočísla zhruba stejně velká taková, že i $2p + 1$ a $2q + 1$ jsou prvočísla, pro $N = (2p + 1)(2q + 1)$ by algoritmus rozložil N jen pro $B \geq p$. Uspěl by tedy za dobu srovnatelnou s algoritmem pokusného dělení.

Obvyklé hodnoty B jsou mezi 10^5 a 10^6.

Druhé stadium Pollardovy $p-1$ metody

Požadavek, aby existovalo prvočíslo $p \mid N$ takové, že $p-1$ je B-hladké, je poměrně silný. Má proto smysl jej zeslabit a požadovat jen, aby bylo $p-1$ zcela rozloženo po pokusném dělení do hranice B, tj. požadovat, aby $p-1 = f \cdot q$, kde f je B-hladké a q je prvočíslo větší než B (ale zase ne příliš velké). Pro naše účely budeme předpokládat, že f je B_1-hladké a prvočíslo q splňuje $B_1 < q \leq B_2$, kde B_1 je naše staré B a B_2 je o dost větší konstanta. Samozřejmě, že bychom p objevili i předchozím algoritmem pro $B = B_2$, ale to by trvalo příliš dlouho.

Podobně jako předtím nyní platí $(a^{qL_B} - 1, N) > 1$. Budeme postupovat takto: po ukončení prvního stadia (tj. předchozího algoritmu) máme spočítáno $b = a^{L_B} \bmod N$. Předpokládejme, že máme uloženy rozdíly prvočísel od B_1 do B_2. Tyto rozdíly jsou malé a je jich nemnoho. Můžeme proto snadno předpočítat b^d pro všechny možné rozdíly d a získat b^q postupným donásobováním původní mocniny b předpočítanými hodnotami b^d. Znamená to, že pro každé prvočíslo mezi B_1 a B_2 nahradíme umocňování pouhým násobením, které je samozřejmě mnohem rychlejší.

Algoritmus (Pollardova $p-1$ metoda, druhé stadium). *Dáno složené N a hranice B_1 a B_2, hledáme netriviálního dělitele N. Máme tabulku $p[1], p[2], \ldots, p[k_1]$ všech prvočísel menších nebo rovných B_1 a tabulku $d[1], d[2], \ldots, d[k_2]$ všech diferencí prvočísel mezi B_1 a B_2 tak, že $d[1] = p[k_1+1] - p[k_1]$ atd.*

1. *[První stadium] Pro $B = B_1$ (a $k = k_1$) zkus rozložit N pomocí předchozího algoritmu. Jestliže tento algoritmus uspěje, skonči. Jinak tento algoritmus dal x. Polož $b \leftarrow x$, $P \leftarrow 1$, $c \leftarrow 0$, $i \leftarrow 0$, $j \leftarrow i$.*
2. *[Předpočítání] Pro všechny hodnoty rozdílů $d[i]$ (které jsou malé a je jich málo) spočítej a ulož $b^{d[i]}$. Polož $x \leftarrow x^{p[k_1]} \bmod N$, $y \leftarrow x$.*
3. *[Vpřed] Polož $i \leftarrow i+1$, $x \leftarrow x \cdot b^{d[i]}$ (pomocí předpočítané hodnoty $b^{d[i]}$), $P \leftarrow P \cdot (x-1) \bmod N$, $c \leftarrow c+1$. Je-li $i \geq k_2$, jdi na 6. Jinak, je-li $c < 20$, jdi na 3.*
4. *[Největší společný dělitel] Polož $g \leftarrow (P, N)$. Je-li $g = 1$, polož $c \leftarrow 0$, $j \leftarrow i$, $y \leftarrow x$ a jdi na 3.*
5. *[Počítej znovu] Polož $i \leftarrow j$, $x \leftarrow y$. Pak opakuj $x \leftarrow x \cdot b^{d[i]}$, $i \leftarrow i+1$, $g \leftarrow (x-1, N)$ dokud nenastane $g > 1$ (což musí nastat). Je-li $g < N$, vytiskni g a skonči. Jinak (tj. je-li $g = N$, což je velmi nepravděpodobné), napiš, že jsi neuspěl a skonči.*
6. *[Neuspěl jsi?] Polož $g \leftarrow (P, N)$. Je-li $g > 1$, jdi na 5. V opačném případě (tj. je-li $g = 1$), napiš, že jsi neuspěl a skonči.*

V případě nepravděpodobného neúspěšného konce v kroku 5 by také bylo možné začít znovu první stadium pro $x \leftarrow 3$ místo $x \leftarrow 2$ v kroku 1.

V této formě je algoritmus mnohem efektivnější než ve formě pouze prvního stadia. Obvyklé hodnoty konstant jsou $B_1 = 2 \cdot 10^6$, $B_2 = 10^8$.

Algoritmus je založen na aritmetice grupy $\mathbb{F}_p^\times$. Podobně lze pracovat i v $\mathbb{F}_{p^2}^\times / \mathbb{F}_p^\times$, v tomto případě je požadována B-hladkost čísla $p + 1$ místo $p - 1$.

Bylo by možné samozřejmě pracovat i v $\mathbb{F}_{p^4}^\times / \mathbb{F}_{p^2}^\times$ nebo $\mathbb{F}_{p^3}^\times / \mathbb{F}_p^\times$ nebo $\mathbb{F}_{p^6}^\times / (\mathbb{F}_{p^2}^\times \cdot \mathbb{F}_{p^3}^\times)$ s požadavkem B-hladkosti čísla $p^2 + 1$ nebo $p^2 + p + 1$ nebo $p^2 - p + 1$. To už jsou ale mnohem větší čísla a splnění požadavku B-hladkosti těchto čísel je méně pravděpodobné. Potřebujeme proto další grupy, jejichž řád je zhruba p, ve kterých jsme schopni pracovat (aniž známe prvočíslo p). Takovými grupami jsou grupy eliptických křivek nad $\mathbb{F}_p$.

Hledání netriviálního dělitele pomocí eliptických křivek

Mějme opět dáno složené přirozené číslo N, které chceme rozložit. Je přirozené předpokládat, že $(N, 6) = 1$. Zvolme a, $b \in \mathbb{Z}$ tak, aby $(4a^3 + 27b^2, N) = 1$. Pak rovnice

$$y^2 z = x^3 + axz^2 + bz^3$$

nám určuje „eliptickou křivku“ $(\mathcal{E}, O)$ nad $\mathbb{Z}_N$, přičemž $O = [0, 1, 0] \in P^2(\mathbb{Z}_N)$. Nechť p je nějaké (neznámé) prvočíslo dělící N. Předchozí rovnicí je zadána eliptická křivka $(\mathcal{E}_p, O_p)$, přičemž $O_p = [0, 1, 0] \in P^2(\mathbb{F}_p)$.

Připomeňme, že $(\mathcal{E}_p, +)$ je komutativní grupa a podle Hasseovy věty platí $|\mathcal{E}_p| = p + 1 - a_p$, kde celé číslo a_p splňuje $|a_p| < 2\sqrt{p}$.

Na množině $\mathcal{E}$ máme definovánu částečnou operaci $+$, přičemž kdykoli známe nějaké body $P = [\alpha_1, \beta_1, 1]$, $Q = [\alpha_2, \beta_2, 1] \in \mathcal{E}$ takové, že $P + Q$ není definováno, snadno najdeme netriviálního dělitele čísla N.

Navíc existuje částečný homomorfismus $f_p : \mathcal{E} \to \mathcal{E}_p$ takový, že jestliže je pro P, $Q \in \mathcal{E}$ definováno $P + Q$, pak platí $f_p(P + Q) = f_p(P) + f_p(Q)$.

Lenstrova metoda eliptických křivek

Představme si, že známe nějaký bod $P = [\alpha, \beta, 1] \in \mathcal{E}$ a že $|\mathcal{E}_p|$ je B-hladké pro nějaké nepříliš velké přirozené číslo B.
Označme L_B nejmenší společný násobek čísel $1, 2, \ldots, B$.
Pak ovšem $|\mathcal{E}_p| \mid L_B$ a platí tedy $L_B \cdot f_p(P) = O_p$.
Předpokládejme, že je definováno $L_B \cdot P$ (přitom si při provádění algoritmu budeme přát samozřejmě opak).
Pak musí pro $L_B \cdot P = [\alpha', \beta', \gamma']$ platit $p \mid \alpha'$, $p \mid \beta' - 1$, $p \mid \gamma'$.
Protože naše vzorce pro sčítání bodů ve třetí složce dávají vždy 0 nebo 1, musí platit $L_B \cdot P = O$. To ale znamená, že $L_B \cdot f_q(P) = O_q$ pro každé prvočíslo $q \mid N$.
Přitom budeme $L_B \cdot P$ počítat postupně „donásobováním“ jednotlivými prvočísly z rozkladu L_B, a tedy každý mezivýsledek musí mít ve třetí složce buď 0 nebo 1.
Protože donásobování prvočísly dělícími L_B provádíme podle velikosti od nejmenších k největším, pokud je $L_B \cdot P$ definováno, musí být největší prvočíslo dělící řád r_q bodu $f_q(P)$ v grupě $(\mathcal{E}_q, +)$ pro všechna prvočísla $q|N$ stejné.
To je ale značně nepravděpodobné.

Lenstrova metoda eliptických křivek - volba parametrů

Proto lze čekat, že pokud pro zvolené nepříliš velké přirozené číslo B platí, že $|\mathcal{E}_p|$ je B-hladké pro nějaké prvočíslo p dělící N, s velkou pravděpodobností najdeme zmíněným postupem netriviálního dělitele čísla N.

Problémem zůstává, že pro zvolené číslo B nemusí $|\mathcal{E}_p|$ být B-hladké pro žádné prvočíslo p dělící N, což objevíme až poté, co spočítáme $L_B \cdot P$. V tomto případě zvolíme jiná a, b a celý postup znovu zopakujeme.

Zbývá vyjasnit několik věcí: jak volit a, b, jak najít $P \in \mathcal{E}$ a jak zvolit přirozené číslo B.

Nelze zvolit a, b náhodně a bod P najít jako nějaké řešení kongruence $y^2 \equiv x^3 + ax + b \pmod{N}$, tj. pro zvolené x nalézt y. Protože N není prvočíslo, řešit kvadratickou kongruenci modulo N je příliš obtížné (ze znalosti všech řešení takové kongruence bychom snadno spočítali netriviálního dělitele čísla N).

Proto zvolíme jiný postup: položíme $b = 1$, $P = [0, 1, 1]$ a volíme pouze a. Jistě potom $P \in \mathcal{E}$.

Lenstrova metoda eliptických křivek - volba parametrů

Otázkou zůstává jak volit B. Protože pro menší p je také menší $|\mathcal{E}_p|$, vzhledem k tomu, že menší čísla jsou s větší pravděpodobností B-hladká než velká, je metoda citlivá spíše na velikost p než na velikost N. Proto je nutno zvolit B tak velké, jak velká prvočísla jsme ještě ochotni hledat (nebo lépe, kolik času jsme ochotni hledání věnovat). Analýza pomocí odhadu pravděpodobnosti toho, že číslo jisté velikosti je B-hladké, ukazuje, že pro hledání prvočísel do velikosti v je vhodné volit B tak, aby $\ln B \doteq \sqrt{\frac{1}{2} \ln v \ln \ln v}$. Speciálně tedy, pro hledání prvočísel menších než 10^{20} je vhodnou hodnotou $B = 12\,000$ (přičemž očekáváme, že bude potřeba projít zhruba 12 000 eliptických křivek, než najdeme p).

Podobně jako u Pollardovy $p - 1$ metody je vhodné i zde doplnit druhé stadium spočívající v tom, že předpokládáme, že $|\mathcal{E}_p|$ je B_1-hladký násobek prvočísla menšího než B_2. Toto druhé stadium je zcela analogické jako u Pollardovy metody, proto si uvedeme algoritmus jen pro první stadium.

Algoritmus (Lenstrova metoda el. křivek, 1. stadium). *Dáno složené N nesoudělné s 6 a hranice B, hledáme netriviálního dělitele N. Máme tabulku $p[1], p[2], \ldots, p[k]$ všech prvočísel $\leq B$.*

1. *[Inicializace] Polož $a \leftarrow 0$.*
2. *[Inicializace křivky] Označme $(\mathcal{E}, O)$ křivku danou rovnicí $y^2z = x^3 + axz^2 + z^3$, kde $O = [0, 1, 0]$. Polož $P = [0, 1, 1]$, $i \leftarrow 0$.*
3. *[Další prvočíslo] Polož $i \leftarrow i + 1$. Je-li $i > k$, polož $a \leftarrow a + 1$ a jdi na 2. Jinak polož $q \leftarrow p[i]$, $r \leftarrow q$, $\ell \leftarrow [\frac{B}{q}]$, $R \leftarrow P$.*
4. *[Násob bod na křivce] Dokud $r \leq \ell$, opakuj $r \leftarrow q \cdot r$. Pak zkus spočítat $P \leftarrow r \cdot P$ na křivce $(\mathcal{E}, O)$. Pokud se to nepodařilo (tj. v průběhu výpočtu byl objeven nenulový prvek okruhu $\mathbb{Z}_N$, který není invertibilní), vytiskni získaného netriviálního dělitele N a skonči. Jinak (tj. P byl vypočten), je-li $P \neq O$, jdi na 3.*
5. *[Počítej znovu] Dokud nebude $R = O$, opakovaně zkoušej spočítat $R \leftarrow q \cdot R$ (pokud se to nepodaří, vytiskni získaného netriviálního dělitele N a skonči). Nakonec polož $a \leftarrow a + 1$ a jdi na 2.*

Dobré aproximace reálných čísel

Definice. Pro libovolné reálné číslo α nechť $\langle\alpha\rangle$ značí necelou část čísla α, to znamená $\alpha - \langle\alpha\rangle \in \mathbb{Z}$ a $0 \leq \langle\alpha\rangle < 1$.

Pro celou část $[\alpha]$ reálného čísla α tedy platí $[\alpha] = \alpha - \langle\alpha\rangle$.

Definice. Pro libovolné reálné číslo α nechť $\|\alpha\|$ je vzdálenost α od nejbližšího celého čísla, tj.

$$\|\alpha\| = \min\{|\alpha - n|;\ n \in \mathbb{Z}\}.$$

Definice. Nechť $\theta \in \mathbb{R}$, $p \in \mathbb{Z}$, $q \in \mathbb{N}$, přičemž $(p, q) = 1$.
Racionální číslo $\frac{p}{q}$ se nazývá dobrá aproximace čísla θ, jestliže $\|q\theta\| = |q\theta - p|$ a pro všechna $q' \in \mathbb{N}$, $q' < q$ platí $\|q'\theta\| > \|q\theta\|$.

Příklad. Platí $\|\pi\| \doteq 0.141593$, $\|2\pi\| \doteq 0.283185$,
$\|3\pi\| \doteq 0.424778$, $\|4\pi\| \doteq 0.433629$, $\|5\pi\| \doteq 0.292037$,
$\|6\pi\| \doteq 0.150444$, $\|7\pi\| \doteq 0.008851$, $\|8\pi\| \doteq 0.132741$,
$\|9\pi\| \doteq 0.274334$. Proto jsou 3 a $\frac{22}{7}$ dobré aproximace čísla π.
Další dobrá aproximace $\frac{333}{106}$ pochází z $\|106\pi\| \doteq 0.008821$.

Věta 1. *Nechť $\theta \in \mathbb{R}$, $Q \in \mathbb{R}$, $Q > 1$. Pak existuje $q \in \mathbb{N}$, $q < Q$ s vlastností $\|q\theta\| \leq \frac{1}{Q}$. Jestliže navíc $\theta \notin \mathbb{Q}$ anebo $Q \notin \mathbb{N}$, existuje $q \in \mathbb{N}$ tak, že $q < Q$, $\|q\theta\| < \frac{1}{Q}$.*

Důkaz. Nejprve budeme předpokládat $Q \in \mathbb{N}$. Uvažme $Q+1$ čísel 0, 1, $\langle\theta\rangle$, $\langle 2\theta\rangle$, $\ldots$, $\langle(Q-1)\theta\rangle$ a rozdělme je do Q intervalů $[0, \frac{1}{Q})$, $[\frac{1}{Q}, \frac{2}{Q})$, $\ldots$, $[\frac{Q-1}{Q}, 1]$. Z Dirichletova principu plyne, že alespoň jeden interval obsahuje aspoň dvě čísla, tedy existují $r_1, r_2, s_1, s_2 \in \mathbb{Z}$ taková, že $0 \leq r_1 < r_2 < Q$ s vlastností $|(r_1\theta - s_1) - (r_2\theta - s_2)| \leq \frac{1}{Q}$. Položme $q = r_2 - r_1$, pak $\|q\theta\| \leq \frac{1}{Q}$. Jestliže $Q \notin \mathbb{N}$, plyne věta z platnosti věty pro $[Q]+1$.

Poznámka. Ukážeme si, že z předchozí věty plyne, že pro libovolné $\theta \in \mathbb{R} - \mathbb{Q}$ existuje nekonečně mnoho $q \in \mathbb{N}$ splňujících

$$q \cdot \|q\theta\| < 1. \tag{3}$$

Ve skutečnosti je možné dokonce dokázat více: číslo 1 na pravé straně může být nahrazeno číslem $\frac{1}{\sqrt{5}}$. Toto silnější tvrzení však nebudeme dokazovat.

Konstrukce posloupnosti dobrých aproximací čísla θ

Zvolme pevně $\theta \in \mathbb{R} - \mathbb{Q}$. Sestrojme indukcí posloupnost všech dobrých aproximací čísla θ. Jistě $q_1 = 1$ dává dobrou aproximaci $\frac{p_1}{q_1}$ čísla θ spolu s nějakým $p_1 \in \mathbb{Z}$ a platí $|q_1\theta - p_1| = \|\theta\| < \frac{1}{2}$. Protože $2\theta \notin \mathbb{Z}$, je touto rovností p_1 určeno jednoznačně. Zřejmě $(q_1, p_1) = 1$.

Předpokládejme, že pro nějaké $n \in \mathbb{N}$ máme dobrou aproximaci $\frac{p_n}{q_n}$ čísla θ. Protože $\theta \notin \mathbb{Q}$, je $\|q_n\theta\| \neq 0$ a věta 1 s $Q = \|q_n\theta\|^{-1}$ zaručuje existenci $q \in \mathbb{N}$, které splňuje $\|q\theta\| < \|q_n\theta\|$. Nechť q_{n+1} je nejmenší q s touto vlastností a $p_{n+1} \in \mathbb{Z}$ je určeno podmínkou $\|q_{n+1}\theta\| = |q_{n+1}\theta - p_{n+1}|$. Je tedy $\|q_{n+1}\theta\| < \|q_n\theta\|$ a pro všechna pro všechna $q \in \mathbb{N}$, $q < q_{n+1}$ platí $\|q\theta\| \geq \|q_n\theta\|$; je tedy $\frac{p_{n+1}}{q_{n+1}}$ dobrá aproximace čísla θ. Protože $\frac{p_n}{q_n}$ je také dobrá aproximace čísla θ, platí $q_{n+1} > q_n$. Kdyby $t = (q_{n+1}, p_{n+1}) > 1$, pak by $p' = \frac{p_{n+1}}{t} \in \mathbb{Z}$, $q' = \frac{q_{n+1}}{t} \in \mathbb{N}$, $q' < q_{n+1}$, přitom
$\|q_{n+1}\theta\| = |q_{n+1}\theta - p_{n+1}| = t|q'\theta - p'| \geq t\|q'\theta\| > \|q'\theta\|$,
což by byl spor s definicí q_{n+1}. Je tedy $(q_{n+1}, p_{n+1}) = 1$.

Vlastnosti posloupnosti dobrých aproximací čísla θ

Dostali jsme posloupnost přirozených čísel

$$1 = q_1 < q_2 < q_3 < \dots \tag{4}$$

a celých čísel $p_1,\ p_2,\ p_3,\ \dots$ splňujících $(p_n, q_n) = 1$ a

$$\|q_n\theta\| = |q_n\theta - p_n|, \tag{5}$$

$$\|q_{n+1}\theta\| < \|q_n\theta\|, \tag{6}$$

$$\|q\theta\| \geq \|q_n\theta\| \quad \text{pro všechna } q \in \mathbb{N},\ q < q_{n+1}. \tag{7}$$

Z věty 1 pro $Q = q_{n+1}$ dostaneme existenci $q \in \mathbb{N}$, $q < q_{n+1}$, takového, že $\|q\theta\| \leq \frac{1}{q_{n+1}}$. Podle (7) platí

$$q_n\|q_n\theta\| < q_{n+1}\|q_n\theta\| \leq 1. \tag{8}$$

Kdyby čísla $q_{n+1}\theta - p_{n+1}$ a $q_n\theta - p_n$ měla stejná znaménka, pro $p' = p_{n+1} - p_n$, $0 < q' = q_{n+1} - q_n < q_{n+1}$, bychom dostali $|q'\theta - p'| < |q_n\theta - p_n| = \|q_n\theta\|$, což by byl spor s (7). Proto

$$(q_n\theta - p_n)(q_{n+1}\theta - p_{n+1}) < 0. \tag{9}$$

Lemma 1. $\{\frac{p_n}{q_n};\, n \in \mathbb{N}\}$ *je množina všech dobrých aproximací a platí* $\lim_{n\to\infty} \frac{p_n}{q_n} = \theta$.

Důkaz. První část plyne z konstrukce, druhá z (8), neboť $|\theta - \frac{p_n}{q_n}| < \frac{1}{q_n^2}$.

Lemma 2. $q_{n+1}p_n - q_n p_{n+1} = \pm 1$.

Důkaz. Levá strana je celé číslo a platí

$$q_{n+1}p_n - q_n p_{n+1} = q_n(q_{n+1}\theta - p_{n+1}) - q_{n+1}(q_n\theta - p_n), \quad (10)$$

odkud spolu s (5), (6), (8) a (9) plyne

$$0 < q_n\|q_{n+1}\theta\| + q_{n+1}\|q_n\theta\| = |q_{n+1}p_n - q_n p_{n+1}| < 2q_{n+1}\|q_n\theta\| \le 2.$$

Důsledek. *Číslo* $q_{n+1}p_n - q_n p_{n+1}$ *má opačné znaménko než* $q_n\theta - p_n$ *a platí* $q_{n+1}p_n - q_n p_{n+1} = -(q_n p_{n-1} - q_{n-1}p_n)$.

Důkaz. Plyne z (10) s přihlédnutím k (5), (6) a $q_{n+1} > q_n$, druhá část z (9) a lemmatu 2.

Lemma 3. *Pro libovolné $n \geq 2$ existuje $a_n \in \mathbb{N}$ tak, že*

$$q_{n+1} = a_n q_n + q_{n-1}, \tag{11}$$

$$p_{n+1} = a_n p_n + p_{n-1}, \tag{12}$$

$$|q_{n-1}\theta - p_{n-1}| = a_n |q_n\theta - p_n| + |q_{n+1}\theta - p_{n+1}|. \tag{13}$$

Důkaz. Z důsledku dostáváme
$p_n(q_{n+1} - q_{n-1}) = q_n(p_{n+1} - p_{n-1})$. Protože $(q_n, p_n) = 1$, plyne odtud existence celého čísla a_n s vlastností $q_{n+1} - q_{n-1} = a_n q_n$, $p_{n+1} - p_{n-1} = a_n p_n$. Protože $q_{n+1} > q_{n-1}$, je $a_n > 0$. Konečně, (13) plyne z (11) a (12) díky (9).

Poznámka. Z (13) pro každé $n \geq 2$ plyne

$$a_n = \left[\frac{|q_{n-1}\theta - p_{n-1}|}{|q_n\theta - p_n|}\right] = \left[\frac{\|q_{n-1}\theta\|}{\|q_n\theta\|}\right]. \tag{14}$$

Známe-li tedy p_1, p_2, q_1, q_2 a θ, můžeme pomocí (14), (11) a (12) dopočítat všechny dobré aproximace iracionálního čísla θ.
Pro $\theta \in \mathbb{Q}$, $2\theta \notin \mathbb{Z}$, probíhá celý proces stejně až do okamžiku, kdy $\|q_n\theta\| = 0$, tj. $\frac{p_n}{q_n} = \theta$, kdy se proces konstrukce dobrých aproximací zastaví. Pro $\theta \in \mathbb{Q} - \frac{1}{2}\mathbb{Z}$ tedy dostáváme konečně mnoho dobrých aproximací, z nichž poslední je rovna θ.

Věta. *Nechť $\theta \in \mathbb{R}$, $2\theta \notin \mathbb{Z}$. Generujme rekurentně celá čísla p_n, q_n, a_n; nejprve položme*

$$p_0 = 1, \qquad q_0 = 0,$$
$$p_1 = [\theta], \qquad q_1 = 1.$$

Pro každé $n \in \mathbb{N}$ takové, že $q_n\theta \neq p_n$, pokračujme

$$\begin{aligned} a_n &= \left[\frac{|q_{n-1}\theta - p_{n-1}|}{|q_n\theta - p_n|}\right] \\ p_{n+1} &= a_n p_n + p_{n-1}, \\ q_{n+1} &= a_n q_n + q_{n-1}. \end{aligned}$$

Pak všechny dobré aproximace čísla θ jsou právě získaná čísla $\frac{p_n}{q_n}$ pro $n \geq 1$, je-li $a_1 > 1$, resp. pro $n \geq 2$, je-li $a_1 = 1$. Navíc platí

$$(-1)^n(q_n\theta - p_n) \leq 0, \tag{15}$$
$$q_{n+1}p_n - q_n p_{n+1} = (-1)^n. \tag{16}$$

Ve studijním textu je uveden *důkaz této věty* i její použití pro *důkaz Lehmannovy věty*.

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.

Legendreův symbol

Nechť p je liché prvočíslo, $a \in \mathbb{Z}$. Legendreův symbol $(\frac{a}{p})$ (čti a vzhledem k p) definujeme takto:

$$(\tfrac{a}{p}) = \begin{cases} 0 & \text{jestliže } p \mid a, \\ 1 & \text{jestliže } p \nmid a \text{ a kongruence } x^2 \equiv a \pmod{p} \text{ má řešení}, \\ -1 & \text{jestliže } p \nmid a \text{ a kongruence } x^2 \equiv a \pmod{p} \text{ nemá řešení}. \end{cases}$$

Jestliže $(\frac{a}{p}) = 1$, nazývá se a kvadratický zbytek modulo p, jestliže $(\frac{a}{p}) = -1$, nazývá se a kvadratický nezbytek modulo p.

Zřejmě platí

$$a, b \in \mathbb{Z},\ a \equiv b \pmod{p} \quad \Rightarrow \quad (\tfrac{a}{p}) = (\tfrac{b}{p}).$$

Proto můžeme definici ekvivalentně přepsat také takto:

$$(\tfrac{a}{p}) = \begin{cases} 0 & \text{jestliže } [a]_p = [0]_p, \\ 1 & \text{jestliže } [a]_p \in \mathbb{Z}_p^\times \text{ je druhou mocninou v této grupě}, \\ -1 & \text{jestliže } [a]_p \in \mathbb{Z}_p^\times \text{ není druhou mocninou v této grupě}. \end{cases}$$

Lemma 1. *Nechť p je liché prvočíslo, $a \in \mathbb{Z}$. Pak*

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}.$$

Důkaz. Případ $p \mid a$ je zřejmý. Dále předpokládejme $p \nmid a$.

Protože grupa $\mathbb{Z}_p^\times$ je cyklická sudého řádu $p-1$, jsou druhými mocninami prvků právě mocniny generátoru se sudým exponentem. Je zde tedy $\frac{p-1}{2}$ prvků, které jsou druhé mocniny, a $\frac{p-1}{2}$ prvků, které nejsou druhé mocniny.
Každý prvek grupy $\mathbb{Z}_p^\times$ je kořenem polynomu $x^{p-1}-1$ v tělese $\mathbb{Z}_p$. Protože $x^{p-1}-1 = (x^{(p-1)/2}-1)(x^{(p-1)/2}+1)$, je každý z prvků $\mathbb{Z}_p^\times$ kořenem alespoň jednoho z obou polynomů stupně $\frac{p-1}{2}$.
Protože polynom nad tělesem nemůže mít víc kořenů, než je jeho stupeň, má každý z obou polynomů $x^{(p-1)/2}-1$, $x^{(p-1)/2}+1$ právě $\frac{p-1}{2}$ kořenů.
Zřejmě každá sudá mocnina generátoru je kořenem polynomu $x^{(p-1)/2}-1$. Množina jeho kořenů se tedy skládá právě ze sudých mocnin generátoru, zatímco liché tvoří množinu kořenů polynomu $x^{(p-1)/2}+1$. Odtud plyne dokazovaná kongruence.

Důsledek 1. *Pro každé liché prvočíslo p platí* $(\frac{-1}{p}) = (-1)^{(p-1)/2}$.

Důkaz. Podle lemma 1 je $(\frac{-1}{p}) \equiv (-1)^{(p-1)/2} \pmod{p}$. Na obou stranách kongruence je ± 1, jejich rozdíl je tedy -2, 0, nebo 2. Ovšem $p \nmid 2$.

Důsledek 2. *Pro každé liché prvočíslo p a každé $a, b \in \mathbb{Z}$ platí* $(\frac{ab}{p}) = (\frac{a}{p})(\frac{b}{p})$.

Důkaz. Podle lemma 1 je
$(\frac{ab}{p}) \equiv (ab)^{(p-1)/2} = a^{(p-1)/2} b^{(p-1)/2} \equiv (\frac{a}{p})(\frac{b}{p}) \pmod{p}$.
Na obou stranách kongruence je opět ± 1, proto rovnost.

Poznámka. Každé celé číslo je kongruentní modulo p s právě jedním z čísel 0, ± 1, ± 2, …, $\pm \frac{p-1}{2}$.

Lemma 2. *Nechť p je liché prvočíslo, $a \in \mathbb{Z}$, $p \nmid a$. Pro každé $i = 1, \ldots, \frac{p-1}{2}$ nechť $a \cdot i \equiv (-1)^{e_i} \cdot b_i \pmod{p}$, kde $e_i \in \{0, 1\}$, $b_i \in \{1, \ldots, \frac{p-1}{2}\}$. Pak $(\frac{a}{p}) = (-1)^e$, kde $e = \sum_{i=1}^{(p-1)/2} e_i$.*

Důkaz. Víme, že $\{b_1, \ldots, b_{(p-1)/2}\} \subseteq \{1, \ldots, \frac{p-1}{2}\}$. Ukažme sporem, že zde platí rovnost. Jestliže zde není rovnost, existují $1 \leq i < j \leq \frac{p-1}{2}$ tak, že $b_i = b_j$. Pak $a \cdot i \equiv \pm a \cdot j \pmod{p}$, což vzhledem k $p \nmid a$ dává $i \equiv \pm j \pmod{p}$, a to je spor. Vynásobením kongruencí $a^{(p-1)/2} \cdot (\frac{p-1}{2})! \equiv (-1)^e \cdot (\frac{p-1}{2})! \pmod{p}$. Protože $p \nmid (\frac{p-1}{2})!$, plyne odtud $a^{(p-1)/2} \equiv (-1)^e \pmod{p}$ a lemma 1 dává $(\frac{a}{p}) \equiv (-1)^e \pmod{p}$. Na obou stranách je ± 1, proto rovnost.

Lemma 3. *Nechť p je liché prvočíslo, $a \in \mathbb{Z}$, $p \nmid a$. Pak $(\frac{a}{p}) = (-1)^{\sum_{i=1}^{(p-1)/2} [\frac{2ai}{p}]}$.*

Důkaz. Platí $\frac{ai}{p} = [\frac{ai}{p}] + \langle \frac{ai}{p} \rangle$, a tedy $a \cdot i \equiv p \cdot \langle \frac{ai}{p} \rangle \pmod{p}$. V lemma 2 je tedy $e_i = 0$, právě když $\langle \frac{ai}{p} \rangle < \frac{1}{2}$. Odtud $e_i = [2\langle \frac{ai}{p} \rangle]$. Platí $[\frac{2ai}{p}] = [2[\frac{ai}{p}] + 2\langle \frac{ai}{p} \rangle] = 2[\frac{ai}{p}] + [2\langle \frac{ai}{p} \rangle] = 2[\frac{ai}{p}] + e_i$. Proto $(-1)^{[\frac{2ai}{p}]} = (-1)^{e_i}$. Lemma 2 dává dokazované.

Důsledek 3. *Pro každé liché prvočíslo p platí* $\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}$.

Důkaz. Pro $1 \leq i \leq \frac{p-1}{2}$ platí $0 < \frac{4i}{p} < 2$, a tedy $[\frac{4i}{p}] \in \{0, 1\}$.
Přitom $[\frac{4i}{p}] = 1 \Leftrightarrow \frac{4i}{p} \geq 1 \Leftrightarrow i \geq \frac{p}{4} \Leftrightarrow i > [\frac{p}{4}]$. Proto
$\sum_{i=1}^{(p-1)/2} [\frac{4i}{p}] = \frac{p-1}{2} - [\frac{p}{4}]$.
Nechť $p = 4k \pm 1$ pro $k \in \mathbb{N}$. Pak $\frac{p-1}{2} = 2k + \frac{\pm 1 - 1}{2}$,
$[\frac{p}{4}] = k + [\frac{\pm 1}{4}]$, a tedy $\frac{p-1}{2} - [\frac{p}{4}] = k$.
Současně platí $\frac{p^2-1}{8} = \frac{16k^2 \pm 8k}{8} = 2k^2 \pm k$.
Užitím lemma 3 dostáváme $\left(\frac{2}{p}\right) = (-1)^{\sum_{i=1}^{(p-1)/2} [\frac{4i}{p}]} = (-1)^{(p^2-1)/8}$.

Lemma 4. *Nechť p je liché prvočíslo, $a \in \mathbb{Z}$, $p \nmid a$, $2 \nmid a$. Pak*
$\left(\frac{a}{p}\right) = (-1)^{\sum_{i=1}^{(p-1)/2} [\frac{ai}{p}]}$.

Důkaz. Platí $\sum_{i=1}^{(p-1)/2} i = \frac{p^2-1}{8}$. Protože je a liché, je $\frac{a+p}{2} \in \mathbb{Z}$.
Užitím lemma 3 a důsledků 3 a 2 dostáváme $(-1)^{\sum_{i=1}^{(p-1)/2} [\frac{ai}{p}]} =$
$= (-1)^{\sum_{i=1}^{(p-1)/2} [\frac{(a+p)i}{p}] - i} = \left(\frac{\frac{a+p}{2}}{p}\right) \cdot \left(\frac{2}{p}\right) = \left(\frac{a+p}{p}\right) = \left(\frac{a}{p}\right)$.

Kvadratický zákon reciprocity

Věta 1. *Pro lichá prvočísla $p \neq q$ platí* $\left(\frac{q}{p}\right) \cdot \left(\frac{p}{q}\right) = (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}}$.

Důkaz. V kartézské soustavě souřadnic si představme obdélník, jehož strany leží na osách a na přímkách $x = \frac{p}{2}$, $y = \frac{q}{2}$. Uvnitř tohoto obdélníku leží právě $\frac{p-1}{2} \cdot \frac{q-1}{2}$ mřížových bodů (tedy bodů, jejichž obě souřadnice jsou celá čísla).
Jeho úhlopříčka leží na přímce $y = \frac{q}{p}x$, žádný z mřížových bodů uvnitř obdélníku neobsahuje a rozděluje obdélník na dva trojúhelníky.
Pro pevně zvolené $i \in \{1, \ldots, \frac{p-1}{2}\}$ je uvnitř „dolního“ trojúhelníku právě $[\frac{qi}{p}]$ mřížových bodů s x-ovou souřadnicí i. Proto je uvnitř „dolního“ trojúhelníku právě $\sum_{i=1}^{(p-1)/2}[\frac{qi}{p}]$ mřížových bodů.
Ze symetrie je uvnitř „horního“ trojúhelníku právě $\sum_{i=1}^{(q-1)/2}[\frac{pi}{q}]$ mřížových bodů.
Dostali jsme $\frac{p-1}{2} \cdot \frac{q-1}{2} = \sum_{i=1}^{(p-1)/2}[\frac{qi}{p}] + \sum_{i=1}^{(q-1)/2}[\frac{pi}{q}]$.
Věta nyní plyne z lemma 4.

Jacobiho symbol

Pro usnadnění počítání Legendreova symbolu $(\frac{a}{p})$ pro konkrétní hodnoty a, p tento symbol nyní zobecníme.

Nechť $b \in \mathbb{N}$ je liché číslo, $a \in \mathbb{Z}$. Je-li $b = 1$, klademe $(\frac{a}{b}) = 1$. Je-li $b > 1$, rozložíme b na součin prvočísel $b = p_1 \cdot p_2 \dots p_s$. Jacobiho symbol $(\frac{a}{b})$ pak definujeme rovností

$$(\tfrac{a}{b}) = (\tfrac{a}{p_1}) \cdot (\tfrac{a}{p_2}) \dots (\tfrac{a}{p_s}).$$

Lemma 5. *Jsou-li $b, d \in \mathbb{N}$ lichá čísla, $a, c \in \mathbb{Z}$, pak* $(\frac{ac}{bd}) = (\frac{a}{b})(\frac{c}{b})(\frac{a}{d})(\frac{c}{d})$.

Důkaz. Plyne z definice a důsledku 2.

Lemma 6. *Jsou-li $a, b \in \mathbb{N}$ lichá čísla, pak* $(-1)^{(a-1)/2}(-1)^{(b-1)/2} = (-1)^{(ab-1)/2}$.

Důkaz. Máme dokázat $\frac{a-1}{2} + \frac{b-1}{2} \equiv \frac{ab-1}{2} \pmod 2$. Ekvivalentně $(a-1) + (b-1) \equiv ab - 1 \pmod 4$, neboli $4 \mid (a-1)(b-1)$. To však zřejmě platí.

Důsledek 4. *Pro každé liché číslo* $b \in \mathbb{N}$ *platí* $(\frac{-1}{b}) = (-1)^{(b-1)/2}$.

Důkaz. Plyne z definice, lemma 6 a důsledku 1 indukcí.

Lemma 7. *Jsou-li* $a, b \in \mathbb{N}$ *lichá čísla, pak* $(-1)^{(a^2-1)/8}(-1)^{(b^2-1)/8} = (-1)^{(a^2b^2-1)/8}$.

Důkaz. Máme dokázat $\frac{a^2-1}{8} + \frac{b^2-1}{8} \equiv \frac{a^2b^2-1}{8} \pmod 2$.
Ekvivalentně $(a^2-1)+(b^2-1) \equiv a^2b^2-1 \pmod{16}$, neboli $16 \mid (a^2-1)(b^2-1)$. Platí dokonce $64 \mid (a^2-1)(b^2-1)$.

Důsledek 5. *Pro každé liché číslo* $b \in \mathbb{N}$ *platí* $(\frac{2}{b}) = (-1)^{(b^2-1)/8}$.

Důkaz. Plyne z definice, lemma 7 a důsledku 3 indukcí.

Věta 2. *Pro lichá nesoudělná* $a, b \in \mathbb{N}$ *platí* $(\frac{b}{a}) \cdot (\frac{a}{b}) = (-1)^{\frac{a-1}{2} \cdot \frac{b-1}{2}}$.

Důkaz. Rozložme na součin prvočísel $a = p_1 \cdot p_2 \dots p_s$, $b = q_1 \cdot q_2 \dots q_t$. Z lemma 5 a věty 1 plyne užitím lemma 6

$$(\tfrac{b}{a}) \cdot (\tfrac{a}{b}) = \prod_{i=1}^{s} \prod_{j=1}^{t} (\tfrac{p_i}{q_j}) \cdot (\tfrac{q_j}{p_i}) = \prod_{i=1}^{s} \prod_{j=1}^{t} (-1)^{\frac{p_i-1}{2} \cdot \frac{q_j-1}{2}} =$$

$$= \prod_{j=1}^{t} \Big(\prod_{i=1}^{s} (-1)^{\frac{p_i-1}{2}} \Big)^{\frac{q_j-1}{2}} = \prod_{j=1}^{t} \big((-1)^{\frac{a-1}{2}} \big)^{\frac{q_j-1}{2}} =$$

$$= \Big(\prod_{j=1}^{t} (-1)^{\frac{q_j-1}{2}} \Big)^{\frac{a-1}{2}} = \big((-1)^{\frac{b-1}{2}} \big)^{\frac{a-1}{2}} = (-1)^{\frac{a-1}{2} \cdot \frac{b-1}{2}}.$$

Zjednodušení vzorců

Větu 2 lze formulovat také jako rovnost

$$\left(\tfrac{a}{b}\right) = \begin{cases} \left(\tfrac{b}{a}\right) & \text{pro } a \equiv 1 \pmod 4 \text{ nebo } b \equiv 1 \pmod 4, \\ -\left(\tfrac{b}{a}\right) & \text{pro } a \equiv b \equiv 3 \pmod 4, \end{cases}$$

která platí pro každá lichá čísla $a, b \in \mathbb{N}$ (i pro soudělná, kdy je na obou stranách 0). Důsledky 4 a 5 lze formulovat takto:

$$\left(\tfrac{-1}{b}\right) = \begin{cases} 1 & \text{pro } b \equiv 1 \pmod 4, \\ -1 & \text{pro } b \equiv 3 \pmod 4, \end{cases} \qquad \left(\tfrac{2}{b}\right) = \begin{cases} 1 & \text{pro } b \equiv \pm 1 \pmod 8, \\ -1 & \text{pro } b \equiv \pm 3 \pmod 8. \end{cases}$$

Výhoda výše uvedených rovností oproti původním je v tom, že je vidět, že není nutné počítat hodnoty zlomků $\frac{a-1}{2}$, $\frac{b-1}{2}$ a $\frac{b^2-1}{8}$.

Výpočet hodnoty Jacobiho, a tedy i Legendreova symbolu

Algoritmus (Jacobiho symbol). *Pro daná* $a \in \mathbb{Z}$, $b \in \mathbb{N}$, $2 \nmid b$, $(a, b) = 1$, *algoritmus najde hodnotu Jacobiho symbolu* $(\frac{a}{b})$.

1. *[Inicializace] Je-li* $a > 0$, *polož* $k \leftarrow 1$. *Je-li* $a < 0$, *polož* $k \leftarrow (\frac{-1}{b})$, $a \leftarrow -a$.
2. *[Jsi hotov?] Je-li* $b = 1$, *pak vytiskni* k *jako odpověď a skonči.*
3. *[Euklidovský krok] Polož* $r \leftarrow a \bmod b$, $u \leftarrow 0$. *Dokud je* r *sudé, opakuj* $r \leftarrow \frac{r}{2}$, $u \leftarrow u + 1$. *Je-li* u *liché, polož* $k \leftarrow k \cdot (\frac{2}{b})$.
4. *[Zde je již* r *liché] Polož* $a \leftarrow b$, $b \leftarrow r$. *Jestliže platí* $a \equiv b \equiv 3 \pmod 4$, *polož* $k \leftarrow -k$. *Jdi na 2.*

Vzhledem k podobnosti s Euklidovým algoritmem víme, že se tento algoritmus vždy zastaví a že je kvadratické časové náročnosti (avšak s jinou O-konstantou než Euklidův algoritmus). Zbývá dokázat, že dává správný výsledek. Ukažme, že vždy na začátku kroku 3 je $k \cdot (\frac{a}{b})$ rovno hledané hodnotě Jacobiho symbolu. To zřejmě platí, když jsme na začátku kroku 3 poprvé.

Důkaz správnosti algoritmu

Algoritmus (Jacobiho symbol). *Pro daná $a \in \mathbb{Z}$, $b \in \mathbb{N}$, $2 \nmid b$, $(a, b) = 1$, algoritmus najde hodnotu Jacobiho symbolu $(\frac{a}{b})$.*

1. *[Inicializace] Je-li $a > 0$, polož $k \leftarrow 1$. Je-li $a < 0$, polož $k \leftarrow (\frac{-1}{b})$, $a \leftarrow -a$.*
2. *[Jsi hotov?] Je-li $b = 1$, pak vytiskni k jako odpověď a skonči.*
3. *[Euklidovský krok] Polož $r \leftarrow a \bmod b$, $u \leftarrow 0$. Dokud je r sudé, opakuj $r \leftarrow \frac{r}{2}$, $u \leftarrow u + 1$. Je-li u liché, polož $k \leftarrow k \cdot (\frac{2}{b})$.*
4. *[Zde je již r liché] Polož $a \leftarrow b$, $b \leftarrow r$. Jestliže platí $a \equiv b \equiv 3 \pmod 4$, polož $k \leftarrow -k$. Jdi na 2.*

Po provedení kroku 3 je nová hodnota $r' \equiv a \pmod b$, poté je spočítáno $r'' = r' \cdot 2^{-u}$, $k' = k \cdot (\frac{2}{b})^u$.

V kroku 4 je $a' = b$, $b' = r''$, $k'' = k' \cdot (-1)^{\frac{a'-1}{2} \cdot \frac{b'-1}{2}}$. Pak platí $k'' \cdot (\frac{a'}{b'}) = k' \cdot (-1)^{\frac{a'-1}{2} \cdot \frac{b'-1}{2}} \cdot (\frac{a'}{b'}) = k' \cdot (-1)^{\frac{b-1}{2} \cdot \frac{r''-1}{2}} \cdot (\frac{b}{r''}) = k' \cdot (\frac{r''}{b}) =$ $= k \cdot (\frac{2}{b})^u \cdot (\frac{r''}{b}) = k \cdot (\frac{2^u r''}{b}) = k \cdot (\frac{r'}{b}) = k \cdot (\frac{a}{b})$. Hodnota $k \cdot (\frac{a}{b})$ je tedy na začátku kroku 3 skutečně stejná, jako byla minule.

Metoda kvadratického síta s více polynomy

Pro obecný kvadratický polynom $Q(x) = Ax^2 + 2Bx + C$ takový, že $A \in \mathbb{N}$, $B, C \in \mathbb{Z}$, platí $AQ(x) = (Ax + B)^2 - (B^2 - AC)$. Pokud bude splněno $N \mid B^2 - AC$, pro každé $a \in \mathbb{Z}$ dostaneme kongruenci tvaru $AQ(a) \equiv (Aa + B)^2 \pmod{N}$.

Zvolíme délku $2M$ intervalu; protože chceme, aby maximum funkce $|Q(x)|$ na intervalu prosívání bylo co nejmenší, zvolíme interval $I = (-\frac{B}{A} - M, -\frac{B}{A} + M)$ a chceme $Q(-\frac{B}{A} + M) \doteq -Q(-\frac{B}{A})$, tj. $A^2M^2 \doteq 2(B^2 - AC)$, tedy $A \doteq \frac{\sqrt{2(B^2-AC)}}{M}$. Potom platí

$$\max_{x \in I} |Q(x)| \doteq |Q(-\tfrac{B}{A})| = \tfrac{B^2-AC}{A} \doteq M\sqrt{\tfrac{B^2-AC}{2}}.$$

Protože toto číslo potřebujeme mít co nejmenší, ale současně má být $B^2 - AC$ dělitelné číslem N, je vhodné volit A, B, C tak, aby $B^2 - AC = N$, kdy maximum $|Q(x)|$ na I bude zhruba $M\sqrt{\frac{N}{2}}$.

Volba koeficientů polynomu $Q(x) = Ax^2 + 2Bx + C$

Nejdříve zvolíme délku prosívání M. Pak zvolíme A blízko $\frac{\sqrt{2N}}{M}$ tak, aby A bylo prvočíslo a N byl kvadratický zbytek modulo A.

Pak nalezneme B tak, aby $B^2 \equiv N \pmod{A}$.

Nakonec položíme $C = \frac{B^2 - N}{A}$. Pak tedy skutečně $N = B^2 - AC$.

Dále pokračujeme stejně jako v metodě kvadratického síta – pro každou mocninu p^k prvočísla p menší než nějaká předem daná hranice určíme kořen a_{p^k} kongruence $x^2 \equiv N \pmod{p^k}$, má-li tato kongruence řešení (pro lichá p to znamená, že N je kvadratický zbytek modulo p), ostatní prvočísla ignorujeme. Čísla a_{p^k} spočítáme pro všechny polynomy jen jednou a uschováme.

Protože $AQ(x) = (Ax + B)^2 - N$, pak kořeny polynomu $Q(x)$ modulo p^k vyhovují kongruenci $Ax \equiv -B \pm a_{p^k} \pmod{p^k}$.

V bázi faktorizace pak máme -1, 2, všechna lichá prvočísla p až do zvolené hranice taková, že N je kvadratický zbytek modulo p, a konečně pro každý použitý polynom $Q(x)$ jeho koeficient A.

Vlastní algoritmus

Postupně prosíváme hodnoty jednoho polynomu $Q(x)$ po druhém, dokud nezískáme dostatek kongruencí pro Gaussovu eliminaci.

Protože malá prvočísla dělí hodně hodnot $Q(x)$, trvá prosívání malými prvočísly nejdéle, přičemž jejich logaritmus je malý. Proto se v některých implementacích prosívání malými prvočísly (řekněme menšími než 100) vynechává, jen je nutné zvýšit hranici, používanou po skončení prosívání pro rozhodování, zda dotyčnou hodnotu polynomu $Q(x)$ budeme rozkládat nebo ne. Přitom strategie je taková: raději zkusit rozkládat nerozložitelné $Q(x)$, než ztratit některé rozložitelné, a tedy nějakou užitečnou kongruenci.

Vzhledem k tomu, že získané kongruence je snadné kontrolovat, je možné do generování kongruencí zapojit více lidí tak, že pomocí e-mailu je jim distribuován program s daty, který nechají běžet ve volném čase na svém počítači, a získané výsledky opět vracejí e-mailem.

Příklad použití metody distribuovaného počítání

Metoda distribuovaného počítání e-maily s následnou kontrolou vrácených výsledků byla s úspěchem použita při rozkládání devátého Fermatova čísla $N = 2^{2^9} + 1$ v roce 1990 (toto N má 155 dekadických cifer).

A. K. Lenstra, H. W. Lenstra, M. S. Manasse a J. M. Pollard tímto způsobem získali matici o 226 688 řádcích a 199 203 sloupcích. Po „zahuštění“ této matice získali matici o 72 413 řádcích a 72 213 sloupcích. Gaussovou eliminací této matice pak získali kongruenci, která jim určila netriviálního dělitele čísla N.

Nepoužili metodu kvadratického síta s více polynomy, ale metodu síta v číselném tělese. Tato metoda je založena na výsledcích algebraické teorie čísel, je tedy z námi studovaných metod teoreticky nejnáročnější, a proto se jí budeme věnovat až do konce semestru.

Algebraická čísla

Definice. Komplexní číslo α se nazývá algebraické, existuje-li normovaný polynom $f(x) \in \mathbb{Q}[x]$, jehož je α kořenem. V opačném případě se α nazývá transcendentní.

Příklad. Všechna racionální čísla jsou algebraická, pro každé $a \in \mathbb{Q}$, $n \in \mathbb{N}$ je $\sqrt[n]{a}$ kořen polynomu $x^n - a$, a tedy číslo algebraické. Čísla $\pi = 3,14159\ldots$, $e = 2,71828\ldots$ jsou transcendentní (to není vidět na první pohled, naopak je to věta, kterou je docela těžké dokázat).

Definice. Nechť α je algebraické číslo, pak ze všech normovaných polynomů s racionálními koeficienty, jejichž je α kořenem, vyberme polynom $f(x) \in \mathbb{Q}[x]$ co nejmenšího stupně. Tento polynom nazýváme minimální polynom čísla α.

Poznámka. Minimální polynom algebraického čísla α je určen jednoznačně (pokud jsou $g_1(x)$ a $g_2(x)$ dva různé normované polynomy stejného stupně mající kořen α, pak je α kořenem i nenulového rozdílu $g_1(x) - g_2(x)$ majícího menší stupeň, který je možné vydělením vedoucím koeficientem normovat).

Vlastnosti minimálního polynomu

Věta 1. Nechť $f(x)$ je minimální polynom algebraického čísla α. Pak $f(x)$ je ireducibilní nad $\mathbb{Q}$ a pro libovolný polynom $h(x) \in \mathbb{Q}[x]$ platí $h(\alpha) = 0$, právě když $f(x) \mid h(x)$ v $\mathbb{Q}[x]$.

Důkaz. Sporem: je-li $f(x) = g_1(x) \cdot g_2(x)$ rozklad $f(x)$ na součin nekonstantních polynomů s racionálními koeficienty, pak $g_1(\alpha) = 0$ nebo $g_2(\alpha) = 0$. Po vydělení vedoucím koeficientem dostaneme normovaný polynom s racionálními koeficienty s kořenem α menšího stupně než je stupeň $f(x)$, spor.
Vydělme polynom $h(x)$ polynomem $f(x)$ se zbytkem: $h(x) = q(x)f(x) + r(x)$ pro $q(x), r(x) \in \mathbb{Q}[x]$, $\operatorname{st} r(x) < \operatorname{st} f(x)$. Dosazením α za x dostaneme $h(\alpha) = r(\alpha)$. Je-li $r(x)$ nulový polynom, pak $f(x) \mid h(x)$ v $\mathbb{Q}[x]$ a současně α je kořenem $h(x)$. Jestliže $r(x)$ není nulový polynom, pak by $r(\alpha) = 0$ vedlo ke sporu (vydělením vedoucím koeficientem bychom dostali normovaný polynom s racionálními koeficienty s kořenem α menšího stupně než je stupeň $f(x)$), a tedy $h(\alpha) = r(\alpha) \neq 0$ a $f(x) \nmid h(x)$ v $\mathbb{Q}[x]$.

Celá algebraická čísla

Definice. Algebraické číslo α se nazývá celé algebraické, má-li jeho minimální polynom $f(x)$ celočíselné koeficienty.

Příklad. Libovolné $a \in \mathbb{Q}$ má minimální polynom $x - a$, a tedy racionální čísla jsou celá algebraická, právě když jsou celá. Číslo $\sqrt[3]{2}$ je celé algebraické (jeho minimální polynom je $x^3 - 2$), číslo $\sqrt{\frac{2}{3}}$ není celé algebraické (jeho minimální polynom je $x^2 - \frac{2}{3}$).

Definice. Nenulový polynom $f(x) \in \mathbb{Z}[x]$ se nazývá primitivní, je-li největší společný dělitel jeho koeficientů roven 1.

Lemma (Gaussovo). Součin libovolných dvou primitivních polynomů je primitivní polynom.

Důkaz. Sporem: předpokládejme, že každý koeficient součinu primitivních polynomů $f(x)$, $g(x)$ je dělitelný nějakým prvočíslem p. Máme homomorfismus okruhů $\psi : \mathbb{Z}[x] \to \mathbb{Z}_p[x]$ (každý koeficient je nahrazen zbytkovou třídou). Z primitivnosti $\psi(f(x)) \neq 0$, $\psi(g(x)) \neq 0$. Přitom $\psi(f(x)) \cdot \psi(g(x)) = \psi(f(x) \cdot g(x)) = 0$. Ovšem $\mathbb{Z}_p[x]$ je obor integrity, spor.

Celá algebraická čísla

Věta 2. *Algebraické číslo α je celé algebraické, právě když existuje normovaný polynom $h(x) \in \mathbb{Z}[x]$, jehož je α kořenem.*

Důkaz. Je-li α celé algebraické, je tímto polynomem jeho minimální polynom.
Naopak, předpokládejme, že existuje normovaný polynom $h(x) \in \mathbb{Z}[x]$, $h(\alpha) = 0$. Označme $f(x)$ minimální polynom čísla α. Z věty 1 víme, že existuje $g(x) \in \mathbb{Q}[x]$ tak, že $h(x) = f(x) \cdot g(x)$. Protože $f(x)$, $g(x)$ jsou normované, existují přirozená čísla n, m tak, že $nf(x)$, $mg(x)$ jsou primitivní (n, m jsou nejmenší společné násobky jmenovatelů koeficientů polynomů $f(x)$, $g(x)$). Podle Gaussova lemmatu je $mn \cdot h(x) = (nf(x)) \cdot (mg(x))$ také primitivní. Protože polynom $h(x) \in \mathbb{Z}[x]$, znamená to, že $mn = 1$, tedy $n = 1$, odkud $f(x) \in \mathbb{Z}[x]$, a tedy α je celé algebraické.

Celá algebraická čísla

Věta 3. *Nechť $\omega_1, \ldots, \omega_n \in \mathbb{C}$. Nechť M je aditivní grupa, generovaná $\omega_1, \ldots, \omega_n$, tj.*

$$M = \{a_1\omega_1 + \cdots + a_n\omega_n;\ a_1, \ldots, a_n \in \mathbb{Z}\}.$$

Jestliže pro každé $\alpha, \beta \in M$ platí $\alpha \cdot \beta \in M$, pak je libovolný prvek M celé algebraické číslo.

Důkaz. Bez újmy na obecnosti můžeme předpokládat, že $\omega_1 \ldots \omega_n \neq 0$. Buď $\alpha \in M$ libovolné. Protože pro každé $i = 1, \ldots, n$ platí $\alpha\omega_i \in M$, existují celá čísla a_{ij} splňující

$$\alpha\omega_i = \sum_{j=1}^{n} a_{ij}\omega_j$$

pro každé $i = 1, \ldots, n$. Odtud plyne, že $\det(\alpha E - (a_{ij})) = 0$, kde E je jednotková matice řádu n. Proto je α kořenem normovaného polynomu $f(x) = \det(xE - (a_{ij})) \in \mathbb{Z}[x]$.

Celá algebraická čísla

Věta 4. *Označme A množinu všech celých algebraických čísel. Pak A je obor integrity.*

Důkaz. Abychom ověřili, že A je obor integrity, stačí ukázat, že je podokruhem tělesa $\mathbb{C}$. Víme, že $\mathbb{Z} \subseteq A$. Musíme tedy dokázat, že pro libovolná $\alpha, \beta \in A$ jsou $\alpha+\beta$, $\alpha-\beta$ i $\alpha\beta$ celá algebraická čísla. Protože α a β jsou celá algebraická čísla, existují polynomy s celými koeficienty $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, $g(x) = x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$ tak, že $f(\alpha) = 0$ a $g(\beta) = 0$. Pak ovšem platí

$$\alpha^n = -a_{n-1}\alpha^{n-1} - \cdots - a_1\alpha - a_0, \ \beta^m = -b_{m-1}\beta^{m-1} - \cdots - b_1\beta - b_0,$$

a tedy podgrupa M aditivní grupy tělesa K generovaná součiny

$$\alpha^i\beta^j, \qquad \text{kde } 0 \le i < n,\ 0 \le j < m, \tag{17}$$

je uzavřená na násobení, neboť libovolný součin $\alpha^u\beta^v$ pro $u \ge 0$, $v \ge 0$ je možné vyjádřit jako $\mathbb{Z}$-lineární kombinaci prvků (17). Podle věty 3 jsou $\alpha+\beta$, $\alpha-\beta$, $\alpha\beta \in M$ celá algebraická čísla.

Těleso algebraických čísel

Definice. Jsou-li K, L tělesa a je-li K podokruhem L, řekneme, že L je rozšířením tělesa K. Pak je L vektorový prostor nad K (sčítání vektorů i násobení vektorů skaláry je určeno operacemi $+$, $\cdot$ v L). Je-li navíc L konečněrozměrný vektorový prostor nad K, hovoříme o konečném rozšíření, jeho dimenzi značíme $[L : K]$ a nazýváme stupněm rozšíření.

Poznámka. Je-li K podtěleso tělesa $\mathbb{C}$, pak K obsahuje $\mathbb{Q}$, a tedy je rozšířením tělesa $\mathbb{Q}$. Je-li toto rozšíření konečné, říkáme, že K je těleso algebraických čísel stupně $[K : \mathbb{Q}]$.

Věta 5. Nechť K je těleso algebraických čísel, pak každé $\alpha \in K$ je algebraické.

Důkaz. Označme $n = [K : \mathbb{Q}]$. Pak α^n, α^{n-1}, $\ldots$, α, 1 je $n+1$ vektorů v n-rozměrném vektorovém prostoru nad $\mathbb{Q}$, proto jsou lineárně závislé, tj. existují a_n, a_{n-1}, $\ldots a_1$, $a_0 \in \mathbb{Q}$, ne všechna nulová, tak, že $a_n\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 = 0$, tedy α je kořen nenulového polynomu s racionálními koeficienty $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$.

Okruh R celých algebraických čísel v tělese K

Věta 6. Nechť K je těleso algebraických čísel, pak množina R všech celých algebraických čísel z tělesa K tvoří obor integrity, jehož podílovým tělesem je K.

Důkaz. Stejně jako ve větě 4 označme A množinu všech celých algebraických čísel. Platí $R = K \cap A$, přičemž A i K jsou podokruhy tělesa $\mathbb{C}$. Proto i R je podokruh tělesa $\mathbb{C}$, tedy obor integrity. Zbývá dokázat, že K je podílové těleso okruhu R. Nechť $\beta \in K$ je libovolné. Podle věty 5 existuje normovaný polynom $f(x) = x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0 \in \mathbb{Q}[x]$ tak, že $f(\beta) = 0$. Nechť n je nejmenší společný násobek jmenovatelů koeficientů polynomu $f(x)$. Pak polynom $g(x) = x^k + na_{k-1}x^{k-1} + \cdots + n^{k-1}a_1 x + n^k a_0 \in \mathbb{Z}[x]$ má kořen $\alpha = n\beta$, neboť $g(\alpha) = g(n\beta) = n^k \cdot f(\beta) = 0$. Je tedy $\alpha \in R$, rovněž $n \in R$. Je tedy $\beta = \frac{\alpha}{n}$ podílem dvou čísel z R. Dokázali jsme, že K je podílové těleso okruhu R.

Opakování z algebry: dělitelnost v oborech integrity

Nechť R je obor integrity, $a, b \in R$.

Definice. Řekneme, že a dělí b v R, píšeme $a|b$, jestliže existuje $c \in R$ tak, že $b = a \cdot c$.

Definice. Řekneme, že a a b jsou asociované v R, píšeme $a \sim b$, jestliže $a|b$ a současně $b|a$.

Poznámka. Platí, že $a \sim b$, právě když existuje jednotka $c \in R^\times$ tak, že $b = a \cdot c$.

Definice. Prvek a se nazývá ireducibilní prvek v R, jestliže $a \neq 0$, $a \notin R^\times$, a kdykoli $a = c \cdot d$ pro $c, d \in R$, pak $c \in R^\times$ nebo $d \in R^\times$.

Příklad. V $\mathbb{Z}$ jsou ireducibilními prvky právě prvočísla a čísla k nim opačná. Je-li K těleso, ireducibilními prvky v $K[x]$ jsou ireducibilní polynomy (například pro $K = \mathbb{C}$ jsou to právě lineární polynomy, pro $K = \mathbb{R}$ jsou to lineární polynomy a kvadratické polynomy se záporným diskriminantem).

Opakování z algebry: okruh s jednoznačným rozkladem

Definice. Říkáme, že okruh R je okruh s jednoznačným rozkladem, jestliže

- R je obor integrity;
- každý $a \in R$, $a \neq 0$, $a \notin R^\times$, je možné napsat jako součin ireducibilních prvků, a to jednoznačně až na pořadí činitelů a jejich asociovanost.

Poznámka. Jednoznačností až na pořadí činitelů a jejich asociovanost znamená toto: jsou-li $a = p_1 \cdots p_n$ a $a = q_1 \cdots q_m$ rozklady prvku a na součiny ireducibilních prvků v R, pak $n = m$ a případnou změnou pořadí činitelů v součinech lze docílit toho, že platí $p_1 \sim q_1, \ldots p_n \sim q_n$.

Příklad. Okruhem s jednoznačným rozkladem je například $\mathbb{Z}$ nebo $K[x]$, kde K je libovolné těleso.

Příklad: $K = \mathbb{Q}(i\sqrt{15}) = \{a + bi\sqrt{15};\ a,\ b \in \mathbb{Q}\}$

Snadno se ukáže, že K je těleso algebraických čísel a že $[K : \mathbb{Q}] = 2$. Označme R okruh všech celých algebraických čísel v K. Je-li $a,\ b \in \mathbb{Z}$, pak $\alpha = a + b\frac{1+i\sqrt{15}}{2}$ je kořenem polynomu

$$(x-a-b\tfrac{1+i\sqrt{15}}{2})(x-a-b\tfrac{1-i\sqrt{15}}{2}) = x^2-(2a+b)x+(a^2+ab+4b^2),$$

a tedy $\alpha \in R$.

Předpokládejme naopak, že pro nějaké $a,\ b \in \mathbb{Q}$ platí $\alpha = a + b\frac{1+i\sqrt{15}}{2} \in R$ a dokažme, že $a,\ b \in \mathbb{Z}$. Je-li $b = 0$, je $\alpha = a \in \mathbb{Q}$, jeho minimální polynom je $x - a$, a tedy $a \in \mathbb{Z}$. Nechť dále $b \neq 0$, tj. $\alpha \notin \mathbb{Q}$. Pak je minimálním polynomem čísla α polynom $f(x) = x^2 - (2a + b)x + (a^2 + ab + 4b^2)$, tedy $c = 2a + b \in \mathbb{Z}$, $d = a^2 + ab + 4b^2 \in \mathbb{Z}$. Proto $-15b^2 = c^2 - 4d \in \mathbb{Z}$, tj. $b \in \mathbb{Z}$. Pak ovšem $2a = c - b \in \mathbb{Z}$, a tedy $2a^2 = 2d - (2a)b - 8b^2 \in \mathbb{Z}$, odkud $a \in \mathbb{Z}$. Dokázali jsme, že $a,\ b \in \mathbb{Z}$, tj.

$$R = \{a + b\tfrac{1+i\sqrt{15}}{2};\ a,\ b \in \mathbb{Z}\}.$$

Aritmetika okruhu $R = \{a + b\frac{1+i\sqrt{15}}{2};\ a,\ b \in \mathbb{Z}\}$

Definujme zobrazení (tzv. normu) $\mathcal{N} : R \to \mathbb{Z}$ předpisem $\mathcal{N}(\alpha) = \alpha \cdot \overline{\alpha} = |\alpha|^2$ pro libovolné $\alpha \in R$. Pro $a, b \in \mathbb{Z}$ tedy $\mathcal{N}(a + b\frac{1+i\sqrt{15}}{2}) = a^2 + ab + 4b^2$. Pak platí, že $\mathcal{N}(\alpha\beta) = |\alpha\beta|^2 = |\alpha|^2 \cdot |\beta|^2 = \mathcal{N}(\alpha) \cdot \mathcal{N}(\beta)$ pro každé $\alpha,\ \beta \in R$. Ukažme, že grupa $R^\times$ všech jednotek okruhu R je rovna $R^\times = \{\alpha \in R;\ \mathcal{N}(\alpha) = \pm 1\}$. Skutečně, je-li $\alpha \in R^\times$, existuje $\beta \in R$ tak, že $\alpha\beta = 1$, odkud plyne $1 = \mathcal{N}(\alpha\beta) = \mathcal{N}(\alpha)\mathcal{N}(\beta)$, a tedy $\mathcal{N}(\alpha) = \pm 1$. Naopak, je-li $\mathcal{N}(\alpha) = \pm 1$, pak $\alpha \cdot (\pm\overline{\alpha}) = 1$, a tedy $\alpha \in R^\times$. Protože $a^2 + ab + 4b^2 = \frac{1}{4}(2a + b)^2 + \frac{15}{4}b^2$, platí pro $a, b \in \mathbb{Z}$, že $\mathcal{N}(a + b\frac{1+i\sqrt{15}}{2}) = \pm 1$, právě když $(2a + b)^2 + 15b^2 = \pm 4$, což nastává právě když $b = 0$ a $a = \pm 1$. Je tedy $R^\times = \{1, -1\}$. Rozložme $4 = 2 \cdot 2 = \frac{1+i\sqrt{15}}{2} \cdot \frac{1-i\sqrt{15}}{2}$ na součin ireducibilních prvků. Kdyby totiž některý z těchto činitelů nebyl ireducibilní, z $\mathcal{N}(2) = \mathcal{N}(\frac{1+i\sqrt{15}}{2}) = \mathcal{N}(\frac{1-i\sqrt{15}}{2}) = 4$ by plynula existence $\alpha \in R$ tak, že $\mathcal{N}(\alpha) = \pm 2$, tedy existence $a, b \in \mathbb{Z}$ tak, že $(2a + b)^2 + 15b^2 = \pm 8$. Tato rovnice však nemá řešení v $\mathbb{Z}$. Proto R **není okruh s jednoznačným rozkladem**.

Další příklad: $K = \mathbb{Q}(\sqrt{10}) = \{a + b\sqrt{10};\ a,\ b \in \mathbb{Q}\}$

Opět je snadné ukázat, že K je těleso algebraických čísel a že $[K : \mathbb{Q}] = 2$. Označme R okruh všech celých algebraických čísel v K. Je-li $a,\ b \in \mathbb{Z}$, pak $\alpha = a + b\sqrt{10}$ je kořenem polynomu

$$(x - a - b\sqrt{10})(x - a + b\sqrt{10}) = x^2 - 2ax + (a^2 - 10b^2),$$

a tedy $a + b\sqrt{10} \in R$. Předpokládejme naopak, že pro nějaké $a,\ b \in \mathbb{Q}$ platí $\alpha = a + b\sqrt{10} \in R$ a dokažme, že $a,\ b \in \mathbb{Z}$. Je-li $b = 0$, je $\alpha = a \in \mathbb{Q}$, jeho minimální polynom je $x - a$, a tedy $a \in \mathbb{Z}$. Nechť dále $b \neq 0$, tj. $\alpha \notin \mathbb{Q}$. Pak je minimálním polynomem čísla α polynom $f(x) = x^2 - 2ax + (a^2 - 10b^2)$, tedy $c = 2a \in \mathbb{Z}$, $a^2 - 10b^2 \in \mathbb{Z}$. Proto $40b^2 = c^2 - 4(a^2 - 10b^2) \in \mathbb{Z}$, odkud $d = 2b \in \mathbb{Z}$. Pak ovšem $4 \mid 4(a^2 - 10b^2) = c^2 - 10d^2$ a tedy c^2 je sudé číslo. Je tedy sudé i samo c a proto je sudé i d^2 a tedy i d. Dokázali jsme, že $a,\ b \in \mathbb{Z}$, tj.

$$R = \{a + b\sqrt{10};\ a,\ b \in \mathbb{Z}\}.$$

Aritmetika okruhu $R = \{a + b\sqrt{10};\ a,\ b \in \mathbb{Z}\}$

Definujme normu $\mathcal{N} :\ R \to \mathbb{Z}$, pro libovolné $a, b \in \mathbb{Z}$ položme $\mathcal{N}(a + b\sqrt{10}) = (a + b\sqrt{10}) \cdot (a - b\sqrt{10}) = a^2 - 10b^2$. Opět platí $\mathcal{N}(\alpha\beta) = \mathcal{N}(\alpha) \cdot \mathcal{N}(\beta)$ pro každé $\alpha,\ \beta \in R$. Odtud plyne, že grupa všech jednotek okruhu R je $R^\times = \{\alpha \in R;\ \mathcal{N}(\alpha) = \pm 1\}$.
Zřejmě $3 + \sqrt{10} \in R^\times$. Odtud $R^\times \supseteq \{\pm(3 + \sqrt{10})^n;\ n \in \mathbb{Z}\}$.
Dokažme, že platí rovnost: budeme předpokládat existenci nějaké jednotky $\eta \in R^\times$, pro kterou $\pm\eta$ není mocninou $3 + \sqrt{10}$ a dojdeme ke sporu. Můžeme předpokládat, že $\eta > 0$ (jinak vezmeme $-\eta$), dokonce že $\eta > 1$ (jinak vezmeme $\frac{1}{\eta}$). Navíc můžeme předpokládat $\eta < 3 + \sqrt{10}$ (jinak vydělíme η největší mocninou čísla $3 + \sqrt{10}$ menší než η). Je tedy $\eta = a + b\sqrt{10}$ pro nějaké $a,\ b \in \mathbb{Z}$ a platí $1 < a + b\sqrt{10} < 3 + \sqrt{10}$,
$\mathcal{N}(a + b\sqrt{10}) = (a + b\sqrt{10})(a - b\sqrt{10}) = \pm 1$. Tudíž $a - b\sqrt{10} = \frac{\pm 1}{\eta}$, a proto $-1 < a - b\sqrt{10} < 1$. Sečtením odtud plyne $0 < 2a < 4 + \sqrt{10}$, což vzhledem k tomu, že a je celé číslo, znamená $a \in \{1, 2, 3\}$. Protože b je rovněž celé číslo a platí $b^2 = \frac{1}{10}(a^2 \mp 1)$, dostali jsme, že $\eta = 1$ nebo $\eta = 3 \pm \sqrt{10}$, spor.

Aritmetika okruhu $R = \{a + b\sqrt{10};\ a,\ b \in \mathbb{Z}\}$

Rozložme

$$9 = 3 \cdot 3 = (1 + \sqrt{10})(-1 + \sqrt{10}).$$

Přitom $\mathcal{N}(3) = 9$ a $\mathcal{N}(1 + \sqrt{10}) = \mathcal{N}(-1 + \sqrt{10}) = -9$.
Dokážeme-li, že v R neexistují čísla s normou ± 3, budeme vědět, že všechna čtyři čísla uvedená v rozkladu čísla 9 jsou ireducibilní, tj. není možné je zapsat ve tvaru součinu dvou čísel z R, které nejsou jednotkami. To je ale snadné: z $a^2 - 10b^2 = \pm 3$ plyne $a^2 \equiv \pm 3 \pmod 5$, spor.
Zbývá vysvětlit, že činitelé nejsou asociovaní: kdyby platilo $3 \sim 1 + \sqrt{10}$ v R, bylo by $\frac{1}{3} + \frac{1}{3}\sqrt{10} \in R$, spor.
Proto R **není okruh s jednoznačným rozkladem**.

Opakování z algebry: ideály v komutativních okruzích

Nechť R je komutativní okruh (jako vždy s jedničkou).

Definice. Řekneme, že $I \subseteq R$ je ideál okruhu R, jestliže $I \neq \emptyset$, pro každé $a, b \in I$ a každé $r \in R$ platí $a + b \in I$, $r \cdot a \in R$.

Příklad. Pro libovolné $a \in R$ je $aR = \{r \cdot a;\ r \in R\}$ ideál okruhu R. Ideál $\{0\}$ se nazývá nulový.

Definice. Ideály aR pro $a \in R$ se nazývají hlavní.

Poznámka. Je-li R obor integrity, pak pro $a, b \in R$ platí $aR = bR$, právě když $a \sim b$, tj. právě když existuje jednotka $c \in R^\times$ tak, že $b = a \cdot c$.

Definice. Jsou-li I, J ideály okruhu R, definujeme
$I + J = \{a + b;\ a \in I,\ b \in J\}$ jejich součet a
$I \cdot J = \{\sum_{i=1}^{n} a_i b_i;\ n \in \mathbb{N},\ a_1, \ldots, a_n \in I,\ b_1, \ldots, b_n \in J\}$ jejich součin.

Příklad. Pro libovolné $a, b \in R$ platí $aR \cdot bR = (a \cdot b)R$. Pozor, nic podobného pro sčítání hlavních ideálů neplatí!

Opakování z algebry: ideály v komutativních okruzích

Stále R je komutativní okruh.

Poznámka. Součet a součin libovolných ideálů okruhu R je ideálem okruhu R. Součet $I + J$ je nejmenší ze všech ideálů obsahujících $I \cup J$. Operace $+$ a $\cdot$ jsou asociativní a komutativní, pro libovolné ideály I_1, I_2, J platí $(I_1 + I_2) \cdot J = I_1 \cdot J + I_2 \cdot J$.

Definice. Ideál I okruhu R se nazývá prvoideál, jestliže $I \neq R$ a pro každé $a, b \in R$ z $ab \in I$ plyne $a \in I$ nebo $b \in I$.

Příklad. Nulový ideál $\{0\}$ je prvoideál, právě když R je obor integrity.

Definice. Ideál I okruhu R se nazývá maximální ideál, jestliže $I \neq R$ a neexistuje žádný ideál J okruhu R splňující $I \subsetneq J \subsetneq R$.

Příklad. V okruhu $\mathbb{Z}$ jsou všechny ideály hlavní, maximální ideály jsou právě ideály $p\mathbb{Z}$, kde p je prvočíslo. Prvoideály okruhu $\mathbb{Z}$ jsou právě tyto maximální ideály a také nulový ideál.

Opakování z algebry: faktorokruh komutativního okruhu

Věta 7. *Nechť R je komutativní okruh, I jeho ideál. Pro libovolné $a \in R$ označme $a + I = \{a + j;\ j \in I\}$. Pak $R/I = \{a + I;\ a \in R\}$ tvoří rozklad na množině R, na kterém lze definovat operace $+$ a $\cdot$ „pomocí reprezentantů“, tj. předpisem*

$$\begin{aligned} (a + I) + (b + I) &= (a + b) + I, \\ (a + I) \cdot (b + I) &= (a \cdot b) + I. \end{aligned}$$

Pak R/I s těmito operacemi tvoří komutativní okruh (tzv. faktorokruh okruhu R podle ideálu I).

Věta 8. *Nechť R je komutativní okruh, I jeho ideál. Pak I je prvoideál okruhu R, právě když R/I je obor integrity. Podobně I je maximální ideál okruhu R, právě když R/I je těleso.*

Důkazy obou vět lze najít ve skriptech J. Rosický: Algebra.

Důsledek. *Každý maximální ideál okruhu R je prvoideálem okruhu R.*

Aritmetika okruhů R celých algebraických čísel

Nechť K je těleso algebraických čísel (tj. $K \subseteq \mathbb{C}$, $[K : \mathbb{Q}] < \infty$), nechť R je okruh celých algebraických čísel v tělese K.

Poznámka. Je-li $K = \mathbb{Q}$, pak $R = \mathbb{Z}$ je okruh s jednoznačným rozkladem. Viděli jsme však, že pro $K = \mathbb{Q}(\sqrt{10})$ dostaneme $R = \{a + b\sqrt{10};\ a,\ b \in \mathbb{Z}\}$ a pro $K = \mathbb{Q}(i\sqrt{15})$ máme $R = \{a + b\frac{1+i\sqrt{15}}{2};\ a,\ b \in \mathbb{Z}\}$, což nejsou okruhy s jednoznačným rozkladem. Kummer v polovině 19. století objevil způsob, jak jednoznačné rozkládání v okruzích celých algebraických čísel zachránit: platí zde následující věta o jednoznačném rozkladu ideálů (takto Kummerovy výsledky přeformulovat Dedekind).

Věta 9. Nechť R je okruh celých algebraických čísel v nějakém tělese algebraických čísel K. Nechť I je ideál R, $I \neq R$, $I \neq \{0\}$. Pak existuje jednoznačně určené $n \in \mathbb{N}$ a jednoznačně (až na pořadí) určené prvoideály $P_1, \ldots, P_n$ takové, že platí

$$I = P_1 \ldots P_n.$$

Důkaz je mimo možnosti této přednášky.

Aritmetika okruhů R celých algebraických čísel

Věta 10. *Nechť R je okruh celých algebraických čísel v nějakém tělese algebraických čísel K. Je-li každý ideál okruhu R hlavní, pak R je okruh s jednoznačným rozkladem.*

Náznak důkazu. Protože je každý ideál okruhu R hlavní, pro každý ideál A existuje prvek $a \in R$ tak, že $A = aR$. Přitom je prvek a určen ideálem A jednoznačně až na asociovanost a platí, že A je prvoideál, právě když a je ireducibilní.
Nechť $a \in R$, $a \neq 0$, $a \notin R^\times$. Pak existence a jednoznačnost rozkladu prvku a na součin ireducibilních prvků plyne z existence a jednoznačnosti rozkladu ideálu aR na součin prvoideálů.

Poznámka. Míru toho, nakolik se okruh celých algebraických čísel R nějakého tělesa algebraických čísel K liší od okruhu s jednoznačným rozkladem, nám vlastně udává to, kolik ze všech ideálů okruhu R je hlavních. Ovšem všech ideálů je spočetně mnoho, hlavních ideálů je také spočetně mnoho, proto slovu „kolik“ v předchozí větě je nutno rozumět správně.

Grupa tříd ideálů okruhu R celých algebraických čísel

Nechť K je těleso algebraických čísel (tj. $K \subseteq \mathbb{C}$, $[K : \mathbb{Q}] < \infty$), nechť R je okruh celých algebraických čísel v tělese K. Uvažme pologrupu $(\mathcal{I}, \cdot)$ všech nenulových ideálů okruhu R a jeho podpologrupu všech nenulových hlavních ideálů. Můžeme uvážit faktorizaci této pologrupy podle zmíněné podpologrupy, což odpovídá následující ekvivalenci mezi ideály: položíme $I \approx J$, právě když existují nenulová $a, b \in R$ splňující $aR \cdot I = bR \cdot J$. Pro libovolný nenulový ideál I označme $[I] = \{J \in \mathcal{I};\ J \approx I\}$ třídu všech ideálů ekvivalentních s I.
Nechť $\mathcal{I}/\approx = \{[I]; I \in \mathcal{I}\}$ je rozklad příslušný této ekvivalenci.
Na $\mathcal{I}/\approx$ lze zavést operaci pomocí reprezentantů: $[I] \cdot [J] = [I \cdot J]$.

Věta 11. $(\mathcal{I}/\approx, \cdot)$ *je konečná komutativní grupa.*
Důkaz je mimo možnosti této přednášky.

Definice. Grupa z věty 11 se nazývá grupa tříd ideálů okruhu R (nebo také tělesa K) a je jednou z nejdůležitějších charakteristik aritmetiky v okruhu R. Počet jejích prvků se nazývá počet tříd ideálů okruhu R (též tělesa K).

Příklad: $R = \{a + b\frac{1+i\sqrt{15}}{2};\ a,\ b \in \mathbb{Z}\}$

Zjistili jsme, že $4 = 2 \cdot 2 = \frac{1+i\sqrt{15}}{2} \cdot \frac{1-i\sqrt{15}}{2}$ jsou podstatně různé rozklady na součin ireducibilních prvků v R. Podívejme se, jak situace vypadá, rozkládáme-li na prvoideály.

Označme ideály $I = 2R + \frac{1+i\sqrt{15}}{2}R$, $J = 2R + \frac{1-i\sqrt{15}}{2}R$. Zřejmě $R/I \cong \mathbb{Z}_2 \cong R/J$, tedy I a J jsou prvoideály okruhu R.
Pak platí $I \cdot J = 4R + (1 + i\sqrt{15})R + (1 - i\sqrt{15})R \subseteq 2R$, protože $4 = 2 \cdot 2$, $1 + i\sqrt{15} = 2 \cdot \frac{1+i\sqrt{15}}{2}$, $1 - i\sqrt{15} = 2 \cdot \frac{1-i\sqrt{15}}{2}$.
Na druhou stranu je $2R \subseteq I \cdot J$, protože $2 = (1 + i\sqrt{15}) + (1 - i\sqrt{15})$, dohromady $I \cdot J = 2R$.
Podobně $I^2 = 4R + (1 + i\sqrt{15})R + (\frac{1+i\sqrt{15}}{2})^2R =$
$4R + (1 + i\sqrt{15})R + (\frac{-7+i\sqrt{15}}{2})R \subseteq \frac{1+i\sqrt{15}}{2}R$, neboť
$4 = \frac{1+i\sqrt{15}}{2} \cdot \frac{1-i\sqrt{15}}{2}$. Na druhou stranu je $\frac{1+i\sqrt{15}}{2} = 4 + \frac{-7+i\sqrt{15}}{2}$, dohromady $I^2 = \frac{1+i\sqrt{15}}{2}R$. Podobně $J^2 = \frac{1-i\sqrt{15}}{2}R$.
Oba podstatně různé rozklady čísla 4 na součin ireducibilních prvků dávají stejný rozklad ideálu $4R$ na součin prvoideálů:
$4R = (I \cdot J)^2 = I^2 \cdot J^2$.

Příklad: $R = \{a + b\sqrt{10};\ a,\ b \in \mathbb{Z}\}$

Zjistili jsme, že $9 = 3 \cdot 3 = (1 + \sqrt{10})(-1 + \sqrt{10})$ jsou podstatně různé rozklady na součin ireducibilních prvků v R. Ukažme si opět, že dostaneme stejné rozklady, rozkládáme-li na prvoideály.

Označme ideály $I = 3R + (1 + \sqrt{10})R$, $J = 3R + (1 - \sqrt{10})R$. Zřejmě $R/I \cong \mathbb{Z}_3 \cong R/J$, tedy I a J jsou prvoideály okruhu R. Platí $I \cdot J = 3R$, $I^2 = (1 + \sqrt{10})R$, $J^2 = (1 - \sqrt{10})R$. Dostáváme stejný rozklad ideálu $9R$ na součin prvoideálů:
$9R = (I \cdot J)^2 = I^2 \cdot J^2$.

Platí $(1 + \sqrt{10})R \cdot J = I^2 \cdot J = (I \cdot J) \cdot I = 3R \cdot I$, a tedy $I \approx J$. V tomto příkladě je možné ukázat, že ideály I a J nejsou hlavní, dokonce platí, že libovolné dva nehlavní ideály jsou ekvivalentní. Proto grupa tříd ideálů okruhu $R = \{a + b\sqrt{10};\ a,\ b \in \mathbb{Z}\}$ má právě dva prvky: třídu všech hlavních ideálů a třídu všech nehlavních ideálů.

I pro druhý námi studovaný příklad $R = \{a + b\frac{1+i\sqrt{15}}{2};\ a,\ b \in \mathbb{Z}\}$ platí, že grupa tříd ideálů má právě dva prvky.

Shrnutí obou předchozích příkladů

Pro $K = \mathbb{Q}(i\sqrt{15}) = \{a + bi\sqrt{15};\ a,\ b \in \mathbb{Q}\}$ máme
$R = \{a + b\frac{1+i\sqrt{15}}{2};\ a,\ b \in \mathbb{Z}\}$ a platí $R^\times = \{1, -1\}$. Normou čísla $\alpha \in R$ je $\mathcal{N}(\alpha) = \alpha \cdot \overline{\alpha} = |\alpha|^2$, obecněji pro libovolné $a, b \in \mathbb{Q}$ definujeme
$\mathcal{N}(a + bi\sqrt{15}) = (a + bi\sqrt{15}) \cdot (a - bi\sqrt{15}) = a^2 + 15b^2$.
Všimněme si, že zobrazení $a + bi\sqrt{15} \mapsto a - bi\sqrt{15}$ pro každé $a, b \in \mathbb{Q}$ dává vnoření $K \to \mathbb{C}$.

Pro $K = \mathbb{Q}(\sqrt{10}) = \{a + b\sqrt{10};\ a,\ b \in \mathbb{Q}\}$ máme
$R = \{a + b\sqrt{10};\ a,\ b \in \mathbb{Z}\}$ a platí $R^\times = \{\pm(3 + \sqrt{10})^n;\ n \in \mathbb{Z}\}$.
Normou čísla $a + b\sqrt{10}$, kde $a, b \in \mathbb{Q}$, je
$\mathcal{N}(a + b\sqrt{10}) = (a + b\sqrt{10}) \cdot (a - b\sqrt{10}) = a^2 - 10b^2$.
Všimněme si, že opět zobrazení $a + b\sqrt{10} \mapsto a - b\sqrt{10}$ pro každé $a, b \in \mathbb{Q}$ dává vnoření $K \to \mathbb{C}$.

V obou případech platí $R^\times = \{\alpha \in R;\ \mathcal{N}(\alpha) = \pm 1\}$.

Zobecnění předchozích příkladů

Nechť K je těleso algebraických čísel, nechť R je okruh celých algebraických čísel v tělese K. Je tedy $K \subseteq \mathbb{C}$, $n = [K : \mathbb{Q}] \in \mathbb{N}$. Platí, že existuje právě n různých vnoření $K \to \mathbb{C}$ (včetně identického vnoření $\alpha \mapsto \alpha$). Označme je $\sigma_1, \ldots, \sigma_n$.
Normu pak definujeme pomocí těchto vnoření: normou libovolného $\alpha \in K$ je číslo $\mathcal{N}(\alpha) = \prod_{i=1}^{n} \sigma_i(\alpha)$.

Pak pro každé $\alpha, \beta \in K$ platí $\mathcal{N}(\alpha\beta) = \mathcal{N}(\alpha) \cdot \mathcal{N}(\beta)$, pro $\alpha \in R$ je $\mathcal{N}(\alpha) \in \mathbb{Z}$ a platí $R^\times = \{\alpha \in R; \mathcal{N}(\alpha) = \pm 1\}$.

Složíme-li $\sigma_i : K \to \mathbb{C}$ s komplexní konjugovaností $\mathbb{C} \to \mathbb{C}$, dostaneme opět některé z vnoření $K \to \mathbb{C}$. Pokud $\sigma_i(K) \subseteq \mathbb{R}$, pak je tímto vnořením opět σ_i. V opačném případě dostaneme nějaké σ_j, $j \neq i$, přičemž složením σ_j s komplexní konjugovaností dostaneme opět σ_i. Vnoření $\sigma_i : K \to \mathbb{C}$ s vlastností $\sigma_i(K) \subseteq \mathbb{R}$ nazýváme reálná. Vnoření, která nejsou reálná, se vyskytují ve dvojicích; označme t počet těchto dvojic a s počet reálných vnoření. Platí tedy $s + 2t = n$.

Dirichletova věta o jednotkách

Nechť K je těleso algebraických čísel, nechť R je okruh celých algebraických čísel v tělese K. Víme, že $n = [K : \mathbb{Q}] = s + 2t$, kde s je počet reálných vnoření $K \to \mathbb{R}$ a t počet dvojic nereálných vnoření $K \to \mathbb{C}$.
Označme $W = \{\alpha \in R;\ \exists m \in \mathbb{N} : \alpha^m = 1\}$ podgrupu všech odmocnin z jedné ležících v K (v případě $s > 0$ je nutně $W = \{1, -1\}$). Vždy platí, že W je konečná cyklická grupa.
Následující Dirichletova věta o jednotkách říká, že platí $R^\times / W \cong \mathbb{Z}^{s+t-1}$.

Věta 12. *Existují jednotky $\varepsilon_1, \ldots, \varepsilon_{s+t-1}$ tak, že každou jednotku $\eta \in R^\times$ můžeme vyjádřit jediným způsobem ve tvaru*

$$\eta = \rho \prod_{i=1}^{s+t-1} \varepsilon_i^{c_i},$$

kde $\rho \in W$ a $c_1, \ldots, c_{s+t-1} \in \mathbb{Z}$.

Důkaz je mimo možnosti této přednášky.

Zpět k našim příkladům

Věta 12. Existují jednotky $\varepsilon_1, \ldots, \varepsilon_{s+t-1}$ tak, že každou jednotku $\eta \in R^\times$ můžeme vyjádřit jediným způsobem ve tvaru

$$\eta = \rho \prod_{i=1}^{s+t-1} \varepsilon_i^{c_i},$$

kde $\rho \in W$ a $c_1, \ldots, c_{s+t-1} \in \mathbb{Z}$.

Příklad. Pro $K = \mathbb{Q}(i\sqrt{15})$ je $s = 0$, $t = 1$, tedy $s + t - 1 = 0$ a $R^\times = W$ je konečná, přičemž $W = \{1, -1\}$.

Příklad. Pro $K = \mathbb{Q}(\sqrt{10})$ je $s = 2$, $t = 0$, tedy $s + t - 1 = 1$. Dokázali jsme, že Dirichletova věta v tomto případě platí pro $\varepsilon_1 = 3 + \sqrt{10}$, přičemž opět $W = \{1, -1\}$.

Nový příklad. Pro $K = \mathbb{Q}(i)$ je $R = \mathbb{Z}[i] = \{a + bi;\ a, b \in \mathbb{Z}\}$, platí $s = 0$, $t = 1$, tedy $s + t - 1 = 0$ a $R^\times = W$ je konečná, přičemž $W = \{1, i, -1, -i\}$. V tomto případě je každý ideál okruhu R hlavní, tedy grupa tříd ideálů okruhu R je triviální a R je okruh s jednoznačným rozkladem.

Metoda síta v číselném tělese

Vraťme se k našemu původnímu problému: máme dáno velké přirozené číslo N, o kterém víme, že je složené, ale není mocninou prvočísla, a hledáme jeho netriviálního dělitele.
Sestrojíme normovaný polynom $f(x) \in \mathbb{Z}[x]$ tak, aby pro nějaké $m \in \mathbb{Z}$ platilo $N \mid f(m)$. Je vhodné, aby absolutní hodnota koeficientů polynomu f nebyla moc velká. Jedna z možností je následující: zvolíme nevelké $n \in \mathbb{N}$ (obvykle asi 5 nebo 6) a zvolíme $m \in \mathbb{N}$ tak, aby bylo o trochu menší než $\sqrt[n]{N}$, tedy aby platilo $m^n < N < 2m^n$. Protože n je malé a N hodně velké, bude takové m existovat. Pak vyjádříme N v poziční soustavě o základu m a získané cifry užijeme jako koeficienty normovaného polynomu $f(x)$, pak tedy bude platit $f(m) = N$ a koeficienty polynomu f budou nezáporné a menší než m (tento postup je možné ještě vylepšit, lze vzít $m \doteq \sqrt[n]{N}$ a brát „cifry“ v rozmezí od $-\frac{m}{2}$ do $\frac{m}{2}$).

Metoda síta v číselném tělese

Nyní tedy máme normovaný polynom $f(x) \in \mathbb{Z}[x]$ a číslo $m \in \mathbb{Z}$ tak, že $N \mid f(m)$. Pravděpodobně je $f(x)$ ireducibilní nad $\mathbb{Z}$, a tedy i nad $\mathbb{Q}$. Pokud však není, rozložíme jej na normované ireducibilní činitele. Z nich vybereme ten, jehož hodnota v m je dělitelná N (kdyby takový neexistoval, dostali bychom netriviální rozklad N a byli bychom hotovi). Tímto ireducibilním činitelem pak nahradíme f. Nechť dále $n = \operatorname{st} f$.

Předchozím postupem jsme získali normovaný ireducibilní polynom $f(x) \in \mathbb{Z}[x]$ a číslo $m \in \mathbb{Z}$ tak, že $N \mid f(m)$. Zvolme kořen $\theta \in \mathbb{C}$ polynomu $f(x)$. Označme $K = \mathbb{Q}(\theta)$ těleso generované číslem θ v $\mathbb{C}$. Platí $K \cong \mathbb{Q}[x]/(f\mathbb{Q}[x])$, tedy $[K : \mathbb{Q}] = \operatorname{st} f = n$. Protože θ je celé algebraické číslo, platí

$$\mathbb{Z}[\theta] = \{a_{n-1}\theta^{n-1} + \cdots + a_1\theta + a_0;\ a_0, \ldots, a_{n-1} \in \mathbb{Z}\} \subseteq R,$$

kde R je okruh celých algebraických čísel tělesa K. Pak $(\mathbb{Z}[\theta], +)$ je podgrupou grupy $(R, +)$ a faktorgrupa $R/\mathbb{Z}[\theta]$ je konečná; označme $r = |R/\mathbb{Z}[\theta]|$. Předpokládejme, že $N \nmid r$ (což je velmi pravděpodobné), a tedy že $(N, r) = 1$ (jinak jsme hotovi).

Shrnutí

N dané složené velké přirozené číslo, není mocninou prvočísla
normovaný ireducibilní polynom $f(x) \in \mathbb{Z}[x]$ stupně $n = \operatorname{st} f$
$m \in \mathbb{Z}$ takové, že $N \mid f(m)$
$\theta \in \mathbb{C}$ takové, že $f(\theta) = 0$
$K = \mathbb{Q}(\theta)$, $[K : \mathbb{Q}] = n$, R okruh celých algebraických čísel v K
$r = |R/\mathbb{Z}[\theta]| \in \mathbb{N}$, $(N, r) = 1$, máme tedy $u, v \in \mathbb{Z}$, že $uN + vr = 1$

Protože $[f(m)]_N = [0]_N$, předpis

$$\varphi(a_{n-1}\theta^{n-1} + \cdots + a_1\theta + a_0) = [a_{n-1}m^{n-1} + \cdots + a_1 m + a_0]_N$$

dává homomorfismus okruhů $\varphi : \mathbb{Z}[\theta] \to \mathbb{Z}_N$.

Pro libovolné $\alpha \in R$ je $r\alpha \in \mathbb{Z}[\theta]$, můžeme tedy rozšířit φ na homomorfismus $\psi : R \to \mathbb{Z}_N$ předpisem $\psi(\alpha) = \varphi(vr\alpha)$, neboť
$\psi(1) = [vr]_N = [1 - uN]_N = [1]_N$ a pro každé $\alpha, \beta \in R$ platí
$\psi(\alpha + \beta) = \varphi(vr\alpha + vr\beta) = \varphi(vr\alpha) + \varphi(vr\beta) = \psi(\alpha) + \psi(\beta)$,
$\psi(\alpha \cdot \beta) = \varphi(vr\alpha\beta) = \varphi(vr\alpha vr\beta) = \varphi(vr\alpha) \cdot \varphi(vr\beta) = \psi(\alpha) \cdot \psi(\beta)$.

Základní myšlenka metody

Rádi bychom našli $a, b \in \mathbb{Z}$ tak, aby $a + b\theta = \alpha^2$ pro vhodné $\alpha \in R$ a současně aby $[a + bm]_N = [z]_N^2$ pro vhodné $z \in \mathbb{Z}$. Pak by totiž pro reprezentanta $y \in \mathbb{Z}$ zbytkové třídy $[y]_N = \psi(\alpha)$ platilo $[y]_N^2 = \psi(\alpha)^2 = \psi(\alpha^2) = \psi(a + b\theta) = [a + bm]_N = [z]_N^2$, což by byla kongruence $y^2 \equiv z^2 \pmod{N}$, která by nám mohla dát hledaného netriviálního dělitele čísla N.

Takovou dvojici (a, b) však až na nezajímavé triviální případy (jako $a = 1$, $b = 0$) nenajdeme. Budeme tedy hledat několik takových dvojic (a, b), aby součinem příslušných $a + b\theta$ byla druhá mocnina v R a současně aby součinem odpovídajících $[a + bm]_N$ byla druhá mocnina v $\mathbb{Z}_N$.

Prosívání

Takové dvojice $a, b \in \mathbb{Z}$ budeme hledat postupně: prosíváním z mnoha dvojic nesoudělných $a, b \in \mathbb{Z}$ vybereme ty, pro které je možné $a + bm$ rozložit nad zvolenou bází faktorizace, do níž dáme -1 a všechna „malá“ prvočísla (tj. prvočísla menší než zvolená mez). Pro vybrané dvojice a, b pak budeme vybírat ty, pro které lze nad vhodnou bází faktorizace rozložit $a + b\theta$. To je už delikátnější záležitost: v R obecně není rozklad na ireducibilní prvky jednoznačný, musíme tedy rozkládat místo prvků ideály na prvoideály; do báze faktorizace dáme vhodně zvolené prvoideály. Ovšem tím jsme schopni postihnout jen to, aby hlavní ideál generovaný součinem příslušných $a + b\theta$ byl druhou mocninou ideálu, kdežto my potřebujeme, aby byl dokonce druhou mocninou hlavního ideálu. A nejen to, i když by to byla druhá mocnina hlavního ideálu, znamenalo by to, že takový ideál lze generovat nějakou druhou mocninou prvku z R, nikoli že náš součin příslušných $a + b\theta$ je druhou mocninou: podílem těchto generátorů musí být jednotka okruhu R, avšak nemusí být druhou mocninou jiné jednotky.

Báze faktorizace v okruhu R

Pro každé „malé“ prvočíslo $p \nmid r$ vyřešíme kongruenci $f(x) \equiv 0 \pmod{p}$, tj. najdeme všechna $c \in \{0, 1, \ldots, p-1\}$ taková, že $p \mid f(c)$ v $\mathbb{Z}$. Pro každé takové c pak ideál $\mathcal{P}_{p,c} = pR + (\theta - c)R$ je prvoideálem okruhu R. Tyto prvoideály jsou výhodné v tom, že snadno zjistíme, zda vystupují v rozkladu ideálu $(a + b\theta)R$ na součin prvoideálů: platí $\mathcal{P}_{p,c} \mid (a + b\theta)R$ v pologrupě ideálů, právě když $p \mid a + bc$ v $\mathbb{Z}$.

Pro zjednodušení úvah předpokládejme, že **grupa tříd ideálů okruhu R je triviální**. Pak každý prvoideál $\mathcal{P}_{p,c}$ je hlavní, tedy $\mathcal{P}_{p,c} = \wp_{p,c}R$ pro nějaké číslo $\wp_{p,c} \in R$, a $\mathcal{N}(\wp_{p,c}) = p$. Jestliže

$$(a + b\theta)R = \mathcal{P}_{p_1,c_1} \cdot \mathcal{P}_{p_2,c_2} \cdots \mathcal{P}_{p_s,c_s},$$

pak existuje jednotka $\varepsilon \in R^\times$ tak, že

$$a + b\theta = \varepsilon \cdot \wp_{p_1,c_1} \cdot \wp_{p_2,c_2} \cdots \wp_{p_s,c_s}.$$

Do báze faktorizace tedy dáme generátory grupy jednotek $R^\times$, o které víme, že má nejvýše n generátorů. Pro každé „malé“ prvočíslo $p \nmid r$ a každé $c \in \{0, 1, \ldots, p-1\}$ splňující $p \mid f(c)$ v $\mathbb{Z}$ pak ještě přidáme do báze faktorizace čísla $\wp_{p,c}$.

Prosívání dvojic $a, b \in \mathbb{Z}$, $|a|, |b|$ „malá“, $b \geq 0$

1. Pro každé prvočíslo p z 1. báze odstraníme dvojice (a, b) splňující $p|a$, $p|b$.
2. (První inicializace) Ke každé zbylé dvojici (a, b) uložme přibližnou hodnotu $\log_2 |a + bm|$.
3. (První prosívání) Pro každou mocninu p^k prvočísla p z 1. báze menší než jistá mez odečteme $\log_2 p$ od hodnot uložených těm zbylým dvojicím (a, b), pro které $p^k \mid a + bm$.
4. Odstraníme všechny dvojice (a, b) s příliš velkou uloženou hodnotou.
5. (Druhá inicializace) Ke každé zbylé dvojici (a, b) uložme přibližnou hodnotu $\log_2 |\mathcal{N}(a + b\theta)|$.
6. (Druhé prosívání) Pro každé $\wp_{p,c}$ z 2. báze faktorizace odečteme $\log_2 p$ od hodnot uložených těm zbylým dvojicím (a, b), pro které $p \mid a + bc$.
7. Pro všechny dvojice (a, b), jejichž uložená hodnota zůstala menší než jistá mez, zjistíme, jestli se $a + b\theta$ rozkládá v 2. bázi faktorizace.

Další postup ve speciálním případě

Pro každou dvojici, pro kterou jsme v 7. kroku ověřili, že se $a + b\theta$ rozkládá v 2. bázi faktorizace, rozložíme v 1. bázi faktorizace $a + bm$. Tím získáme pro tuto dvojici (a, b) z exponentů obou rozkladů vektor nad dvouprvkovým tělesem $\mathbb{F}_2$.

Až máme těchto vektorů o několik více než kolik je celkem prvků v obou bázích faktorizace, provedeme Gausovu eliminaci (nejprve řídké a pak husté matice), abychom našli jejich lineární závislosti. Každá lineární závislost nám dá jednu kongruenci

$$y^2 \equiv z^2 \pmod{N}.$$

Budeme-li mít těchto kongruencí několik, je reálná šance, že pro některou z nich platí $y \not\equiv \pm z \pmod{N}$, a tedy $(N, y + z)$ je netriviální dělitel čísla N.

Obecný případ

V obecném případě, kdy **grupa tříd ideálů okruhu R není triviální**, je celá situace komplikovanější. Má-li tato grupa sudý řád, může se totiž stát, že přestože je ideál, který získáme z nalezené lineární závislosti vynásobením vhodných ideálů $(a + b\theta)R$ druhou mocninou nějakého ideálu I, nemusí být tento ideál I hlavní. Omezme se zde jen na konstatování, že se tento problém dá řešit například tím, že pro každou takto nalezenou lineární závislost uložíme informaci o tom, ve které třídě grupy tříd ideálů leží ideál I. Pak znovu provedením Gaussovy eliminace nalezneme lineární závislost mezi těmito vektory a ta nám dá lineární závislost ideálů $(a + b\theta)R$, pro kterou je odpovídající ideál I hlavní. Ovšem další ještě větší komplikace spočívá v tom, jak najít generátor tohoto hlavního ideálu (jde o druhou odmocninu z celého algebraického čísla, které je součinem tisíců činitelů tvaru $(a + b\theta)$, a tedy lze čekat, že vyjádříme-li toto číslo jako hodnotu v θ polynomu stupně menšího než n, koeficienty tohoto polynomu mohou mít několik stovek tisíc dekadických cifer). Vysvětlit triky, pomocí kterých se tato komplikace překonává, už v této přednášce nestihneme...

Odhad časové náročnosti

Metoda síta v číselném tělese je nejnovější a potenciálně nejrychlejší známá metoda rozkládání velkých přirozených čísel. Na základě některých heuristických argumentů lze odhadovat, že metoda řetězových zlomků i metoda kvadratického síta jsou časové náročnosti

$$O\left(e^{\sqrt{\ln N \ln\ln N}(1+o(1))}\right).$$

Proto před objevením metody síta v číselném tělese panovalo přesvědčení, že lepší časové náročnosti už patrně nepůjde dosáhnout. Bylo překvapením, že na základě podobných argumentů lze odhadovat, že metoda síta v číselném tělese je časové náročnosti

$$O\left(e^{\sqrt[3]{(\ln N)(\ln\ln N)^2}(c+o(1))}\right)$$

pro poměrně malé c (menší než $\sqrt[3]{\frac{64}{9}}$), což je asymptoticky mnohem lepší než jakákoli jiná známá metoda.