ALGORITMY TEORIE ČÍSEL

Radan Kučera

Literatura

[Ca] Cassels J. W. S.: *An Introduction to Diophantine Approximation*, University Press, Cambridge, 1965.

[C] Cohen H.: *A Course in Computational Algebraic Number Theory*, Graduate Texts in Mathematics 138, Springer–Verlag, Berlin, Heidelberg, New York 1993, kapitoly 8–10.

[K] Knuth D.E., *The Art of Computer Programming, díl 2: Seminumerical Algorithms*, (druhé vydání), Addison-Wesley, Reading, Mass., 1981.

[L] 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.

[R] Rosický J., Algebra, Grupy a okruhy, skriptum MU, 2000.

Úvod

Cílem této přednášky je ukázat, jak silné a účinné jsou prostředky, založené na výsledcích teorie čísel, při řešení některých matematických problémů. Protože jsem se nechtěl této problematice věnovat jen povrchně, ale přitom jsem měl k disposici jen jeden semestr, bylo třeba se zaměřit pouze na jeden problém a v tomto problému se dostat dost hluboko. Při rozhodování, kterému problému se věnovat, jsem vyloučil ty problémy, které se objevují při stavbě teorie čísel samotné, neboť jsem chtěl dokumentovat její užitečnost i pro ostatní matematiku. Proto jsem si nakonec zvolil problém na první pohled banálně jednoduchý, na který se však narazí na samém počátku budování přirozených čísel a který se objevuje už ve školském učivu: rozkládání přirozeného čísla na prvočinitele. Jde o problém skutečně velmi jednoduchý, máme-li na mysli „malá“ přirozená čísla, avšak pro „velká“ přirozená čísla (mající řekněme 50 – 100 dekadických cifer) všechny „naivní“ metody selhávají. Přesto však byly objeveny „rafinované“ metody založené právě na výsledcích teorie čísel, které jsou schopny i tato „velká“ přirozená čísla rozložit. A právě těmto metodám bude tato přednáška věnována. Přitom budeme využívat hlavně knihu [C].

1. Algoritmy

Obsahem této přednášky není definovat, co to je algoritmus, a ani to nebudeme potřebovat. Přesto si však blíže určeme, co budeme algoritmem rozumět: je to metoda, která pro jistý typ vstupů dává po *konečné* době odpověď. Jestliže zadáváme algoritmus, je třeba provést několik dalších věcí:

1. Dokázat jeho správnost, tj. ukázat, že dává skutečně požadovaný výsledek poté, co se zastaví.
2. Protože se zajímáme o praktické implementace, je třeba dát odhad, jak dlouho algoritmus poběží, je-li to možné, odhadnout čas pro nejhorší případ a také v průměru.

Vysázeno systémem $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Zde je třeba být opatrný: čas výpočtu budeme měřit vždy v *bitových* operacích, tj. počtem logických a aritmetických operacích prováděných na nulách a jedničkách. To je totiž nejrealističtější model, předpokládáme-li, že používáme skutečné počítače.

3. Je třeba odhadnout i paměťovou náročnost algoritmu (měřenou v bitech). U většiny algoritmů je tato náročnost zanedbatelná a není nutné se jí zabývat, mnohdy to však může být důležité.

Pro potřeby časové náročnosti budeme velikost vstupů měřit počtem bitů, které jsou zapotřebí pro jejich zápis (např. pro vstup přirozeného čísla N je třeba $1 + [\log_2 N]$ bitů).

Definice. *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*

$$0 < \limsup_{n\to\infty} \left|\frac{a_n}{b_n}\right| < \infty.$$

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, můžeme předpokládat, že vstupem pro náš algoritmus je jediné přirozené číslo N. V obecnějším případě by bylo třeba nahradit v následující definici $\ln N$ počtem bitů potřebných pro zapsání celého vstupu.

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 (resp. kvadratického, resp. kubického) času, je-li tento čas řádu $O(\ln N)$, (resp. $O(\ln^2 N)$, resp. $O(\ln^3 N)$).

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, je-li tento čas řádu $O(N^\alpha)$ pro nějaké kladné reálné číslo α, 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 $a + b = 1$. Ověřte si, že tyto algoritmy jsou subexponenciálního času.

Uvedená „definice“ algoritmu, ačkoli je značně vágní, je přesto často příliš striktní pro praktické účely. Budeme také potřebovat „pravděpodobnostní algoritmy“, jejichž běh závisí na jistém zdroji náhodných čísel. Takový „algoritmus“ by vlastně ani algoritmem nazýván být neměl, protože je zde možnost (pravděpodobnosti nula), že nikdy neskončí. Zkušenosti však ukazují, že tyto „pravděpodobnostní algoritmy“ jsou často efektivnější než ostatní a mnohdy jsou jediné, které máme k disposici. Na druhé straně rozhodně nebudeme nazývat algoritmem metodu, produkující výsledek, který je s vysokou pravděpodobností správný. Je podstatné, že algoritmus dává správné výsledky (odhlédneme-li od chyb člověka či počítače při jeho provádění).

Příklad. V posluchárně si nechám nadiktovat od několika posluchačů postupně 100 cifer. Pak odpovím, že vzniklé stociferné číslo není druhou mocninou přirozeného čísla. Asi budu mít pravdu, protože mezi $9 \cdot 10^{99}$ stociferným čísly je jen asi $10^{50} - 10^{49} \cdot \sqrt{10} \doteq 6,84 \cdot 10^{49}$ druhých mocnin přirozených čísel, tedy pravděpodobnost neúspěchu je menší než 10^{-50}. Přesto však nemůže být řeči o algoritmu.

Je vhodné si uvědomit, že u pravděpodobnostních algoritmů nemá smysl hovořit o nejdelším možném času výpočtu, ale o očekávaném času výpočtu.

2. Počítání s velkými čísly

Velice často budeme potřebovat provádět výpočty s celými čísly, jejichž absolutní hodnota je značně velká. Proto budeme předpokládat, že máme k disposici software, ve kterém je možné provádět základní algebraické operace s čísly, majícími řekněme 1000 dekadických cifer. Nejznámější systémy, které takové výpočty umožňují, jsou asi MATHEMATICA a MAPLE, jejichž nevýhodou je, že jsou komerční a jsou poměrně drahé. Méně známý systém je PARI-GP, který je zaměřen na teorii čísel. Je volně šiřitelný a je ho tedy možné získat zdarma.

Pro získání představy o časové náročnosti jednotlivých operacích je však vhodné si představit, že systém umožňující operace s libovolně velkými celými čísly máme vytvořit. Taková čísla budeme zapisovat v poziční soustavě o vhodném základu a operace budeme provádět podobně jako jsme to zvyklí dělat na papíře s dekadickými čísly. Je zřejmé, že lineární časovou náročnost bude mít sčítání a odčítání, dále pak násobení „malým" přirozeným číslem a násobení či dělení se zbytkem mocninou zvoleného základu. Naproti tomu kvadratickou časovou náročnost bude mít obecné násobení a dělení se zbytkem. Pokud se týká vstupu a výstupu, časová závislost je lineární nebo kvadratická v závislosti na tom, zda zvolený základ je či není mocninou deseti. Protože pro aritmetické operace je výhodnější, je-li základ zvolené poziční sopustavy mocnina dvou, je tato volba nejvhodnější. Obvykle totiž čas potřebný pro vstup a výstup je pouze zanedbatelná část celkové doby výpočtu a obvykle je dominován časem pro fyzický zápis. Pro zajímavost si uveďme, že v PARI-GP je základem mocnina dvou, kdežto MAPLE používá mocninu deseti.

Uveďme si ještě, že jsou známy algoritmy pro násobení a dělení $n$bitových čísel, které dosahují menší časové náročnosti než „metoda tužky a papíru". Nejlepší z nich, kterou objevili Schönhage a Strassen, vyžaduje jen $O(n \cdot \ln n \cdot \ln\ln n)$ bitových operací. Avšak tyto rafinované metody jsou rychlejší až pro čísla mající několik set dekadických cifer a více. S takovými čísly je však třeba pracovat jen zřídka a proto se patrně implementace těchto metod nevyplatí.

3. Největší společný dělitel

Často budeme potřebovat spočítat největší společný dělitel dvou celých čísel. Naivní řešení by bylo: rozlož obě čísla na součin prvočísel a poté vynásob společné činitele. Ale to je postup, který se osvědčí jen pro čísla s velmi malou absolutní hodnotou (řekněme do 100) nebo v případě, že víme, že některé z daných čísel je prvočíslo a stačí tedy provést jen jedno dělení se zbytkem. Mnohem výhodnější je výpočet největšího společného dělitele pomocí Euklidova algoritmu, který je patrně nejstarší a nejdůležitější algoritmus teorie čísel.

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.*

Pro určení časové náročnosti je třeba vědět, kolikrát se provede Euklidovský krok. Platí následující věta[1]

Věta. *1. Je-li $1 \le a \le N$, $1 \le b \le N$, 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, $b \in \{1, \dots, N\}$ je roven přibližně

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

Je tedy počet kroků algoritmu konstanta vynásobená $\ln N$. Každý krok vyžaduje dlouhé dělení, které je kvadratického času. Proto je tento algoritmus kubického času. Existuje však trik, jak algoritmus výrazně urychlit: v průběhu výpočtu jsou a, b stále menší a menší, proto je možné průběžně snižovat potřebný počet cifer v naší poziční soustavě. Všimněme si, že při výpočtu Euklidovského kroku $a = bq + r$ je časová náročnost $O((\ln a)(1 + \ln q))$, tedy celkový čas je omezen řádem

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

Ale

$$\textstyle\sum \ln q = \ln \prod q \le \ln N$$

a tedy při pečlivém naprogramování jde o algoritmus kvadratického času.

Jiná varianta Euklidova algoritmu, která je také užitečná v praxi, je tzv. binární algoritmus, ve kterém se nepoužívá dlouhé dělení, ale jen odčítání a dělení dvěma (realizované posunem). Počet kroků je větší, ale užité operace jsou rychlejší a tedy je zde určité zrychlení pro naše „velká“ čísla. Je ale podstatné, aby základem naší poziční soustavy byla 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 $2^k a$ 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.*

---
[1] Důkaz je možno najít v knize [K].

Je jasné, že počet opakování kroků 4 a 5 je $O(\ln ab)$ (po každém průchodu se součin ab zmenší na méně než polovinu), odčítání i posuv jsou lineárního času, tedy opět jde o algoritmus kvadratického času. Je ovšem nezbytné všechna dělení dvěma provádět jako posuvy.

Označíme-li d největší společný dělitel celých čísel a, b, pak existují celá čísla u, v tak, že $d = ua + vb$ (Bezoutova rovnost). V některých aplikacích budeme potřebovat spočítat nejen d, ale i čísla u, v, proto si uveďme 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 spočítej $q \leftarrow [\frac{d}{v_3}]$ a $t_3 \leftarrow d \bmod v_3$. 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.*

Poznamenejme, že „současně“ v kroku 3 poukazuje na to, že obvykle instrukce „dělení se zbytkem“ dává jak podíl tak i zbytek, ať už je to instrukce z assembleru nebo z nějakého softwaru pracujícího s velkými čísly. Hodnota zlomku v kroku 2 je celočíselná.

Důkaz algoritmu. Uvědomme si, že výpočty v proměnných d, v_3, t_3 nezávisí na hodnotách ostatních proměnných. Přeznačíme-li je a, b, r, dostaneme právě Euklidův algoritmus, proto náš algoritmus musí skončit a po skončení je v d hledaný největší společný dělitel. Zbývá dokázat, že pak také platí $au + bv = d$. Za tím účelem rozšiřme daný algoritmus tak, že zavedeme další proměnné v_2, t_2, v (které nebudou ovšem nikdy použity pro výpočet hodnot původních proměnných). Rozšiřme inicializační krok 1 o $t_1 \leftarrow 0$, $t_2 \leftarrow 0$, $t_3 \leftarrow 0$, $v \leftarrow 0$, $v_2 \leftarrow 1$ a krok 3 o $t_2 \leftarrow v - qv_2$, $v \leftarrow v_2$, $v_2 \leftarrow t_2$. Po skončení kroku 1 platí

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

Dokážeme indukcí, že $(*)$ platí vždy, když začínáme krok 2. Zbývá ověřit, že platilo-li $(*)$ před krokem 3, platí $(*)$ i po něm. Abychom zabránili zmatku, budeme v důkaze nově získané hodnoty proměnných označovat čárkami. Pro $q = [\frac{d}{v_3}]$ platí:

$$\begin{aligned}
t_3' &= d - qv_3\\
t_1' &= u - qv_1\\
u' &= v_1\\
d' &= v_3\\
v_1' &= u - qv_1\\
v_3' &= d - qv_3\\
t_2' &= v - qv_2\\
v' &= v_2\\
v_2' &= v - qv_2
\end{aligned}$$

Předpokládáme tedy, že $(*)$ platí pro nečárkované proměnné a dokážeme ji pro čárkované:

$$\begin{aligned}
at_1' + bt_2' &= a(u - qv_1) + b(v - qv_2) = au + bv - q(av_1 + bv_2) = d - qv_3 = t_3' \\
au' + bv' &= av_1 + bv_2 = v_3 = d' \\
av_1' + bv_2' &= a(u - qv_1) + b(v - qv_2) = (au + bv) - q(av_1 + bv_2) = d - qv_3 = v_3'
\end{aligned}$$

Dokázali jsme, že algoritmus je správný. Snadno se vidí, že oproti původnímu Euklidovu algoritmu v cyklu přibylo 1 odčítání, 1 (dlouhé) násobení a 2 přiřazení, cyklus nyní trvá asi dvakrát déle než v předchozím algoritmu. Proto i celý rozšířený Euklidův algoritmus trvá asi dvojnásobek času potřebného pro Euklidův algoritmus. Jde tedy opět o algoritmus kvadratického času.

4. Nezbytný aparát z algebry a elementární teorie čísel

V této kapitole zopakujeme aritmetiku okruhu zbytkových tříd $\mathbb{Z}/m\mathbb{Z}$, kde $m \in \mathbb{N}$ (užíváme standardního značení: $\mathbb{Z}$ značí množinu či okruh celých čísel, $\mathbb{N}$ množinu či polookruh čísel přirozených, (a, b) je označení největšího společného dělitele celých čísel a, b).

Přestože byla v algebře teorie faktorizace okruhů podle ideálů jistě probrána a mohl jsem tedy okruh zbytkových tříd popisovat jako speciální případ faktorokruhu, rozhodl jsem se pro elementárnější výklad pomocí kongruencí s tím, že se občas zmíníme o ekvivalentním tvrzení pro okruhy zbytkových tříd. Zdá se mi totiž jednodušší hovořit například o kongruencích vzhledem ke dvěma modulům než o dvou okruzích zbytkových tříd a vztazích mezi nimi (porovnej např. obsah věty 2). Důkazy vět, které jsou zřejmé, budu vynechávat.

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 $m | a - b$.*

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 $a + c \equiv b + d \pmod{m}$, $a - c \equiv b - d \pmod{m}$, $ac \equiv bd \pmod{m}$.*

Je jasné, že $a \equiv b \pmod{m}$ znamená právě to, že celá čísla a, b jsou obsažena v téže třídě rozkladu $\mathbb{Z}/m\mathbb{Z}$. Kongruence modulo m jsou tedy totéž co rovnosti v okruhu zbytkových tříd $\mathbb{Z}/m\mathbb{Z}$ (předchozí věta neříkala nic jiného než to, že na rozkladu $\mathbb{Z}/m\mathbb{Z}$ bylo možno zavést operace sčítání, odčítání a násobení pomocí reprezentantů).

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 $a \equiv b \pmod{m}$.*

Vzhledem k tomu, že $(\mathbb{Z}/m\mathbb{Z})$ je konečný okruh, z předchozí věty snadno plyne, že v $\mathbb{Z}/m\mathbb{Z}$ jsou jednotkami (tj. invertibilními prvky) právě třídy obsahující čísla nesoudělná s m.

Důkaz. Podle Bezoutovy rovnosti existují $t, r \in \mathbb{Z}$ tak, že platí $tm + rk = 1$. Vynásobte tuto rovnost číslem $a - b$ a dokažte, že m dělí levou stranu.

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

Důkaz. Jestliže $(a, m) | b$, vynásobte Bezoutovu rovnost pro (a, m) číslem $\frac{b}{(a,m)}$. Opačný směr je zřejmý.

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 $x \equiv x_1 \pmod{m_1}$, $x \equiv x_2 \pmod{m_2}$.*

Důkaz. Podle Bezoutovy rovnosti existují $a, b \in \mathbb{Z}$ tak, že platí $am_1 + bm_2 = 1$. Položte $x = x_2am_1 + x_1bm_2$.

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

Věta 6. *Pro libovolná $m_1, m_2 \in \mathbb{N}$ taková, že $(m_1, m_2) = 1$, platí $\varphi(m_1m_2) = \varphi(m_1)\varphi(m_2)$.*

Důkaz. Uvažme zobrazení $\{1, 2, \ldots, m_1\} \times \{1, 2, \ldots, m_2\} \to \{1, 2, \ldots, m_1m_2\}$ přiřazující dvojici $(x_1, x_2) \in \{1, 2, \ldots, m_1\} \times \{1, 2, \ldots, m_2\}$ číslo $y \in \{1, 2, \ldots, m_1m_2\}$ 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_1m_2). Je zřejmé, že toto zobrazení je bijekce a že $(y, m_1m_2) = 1$ 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. Je-li m mocninou prvočísla, je tvrzení zřejmé. Pro obecné m odvoďte tvrzení z věty 6 indukcí vzhledem k počtu prvočísel dělících m.

Definice. *Je-li R okruh s jedničkou 1, značíme $R^\times$ jeho (multiplikativní) grupu invertibilních prvků, tj. $R^\times = \{a \in R; \exists b \in R : ab = 1\}$. Charakteristika okruhu R je nejmenší $n \in \mathbb{N}$ splňující $n \cdot 1 = 0$ (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.*

Z poznámky za větou 3 plyne, že $\varphi(m)$ je rovno počtu prvků $(\mathbb{Z}/m\mathbb{Z})^\times$.

Věta 8. **(Eulerova věta)** *Pro libovolná $m \in \mathbb{N}$, $a \in \mathbb{Z}$, taková, že $(a, m) = 1$, platí*

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

Důkaz. Věta plyne z předchozí poznámky a z toho, že v konečné grupě $(\mathbb{Z}/m\mathbb{Z})^\times$ řád libovolného prvku dělí řád grupy.

Důsledek. **(Fermatova věta)** *Pro libovolné prvočíslo p a libovolné $a \in \mathbb{Z}$ nedělitelné p platí $a^{p-1} \equiv 1 \pmod{p}$.*

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. Bez újmy na obecnosti lze předpokládat, že $r > s$. Je-li $r = s + ke$ pro vhodné $k \in \mathbb{N}$, platí $a^r = a^s \cdot (a^e)^k \equiv a^s \pmod{m}$. Naopak, předpokládejme $a^r \equiv a^s \pmod{m}$. Protože je $(a, m) = 1$, plyne odtud $a^{r-s} \equiv 1 \pmod{m}$. Vydělme $r - s$ číslem e se zbytkem: existují q, $z \in \mathbb{Z}$, $0 \le z < e$, splňující $r - s = qe + z$. Pak platí $a^z \equiv (a^e)^q \cdot a^z = a^{r-s} \equiv 1 \pmod{m}$ a tedy $z = 0$ podle definice čísla e. Odtud $r \equiv s \pmod{e}$.

Věta 10. *Charakteristika konečného tělesa je prvočíslo.*

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

Důkaz. Viz [R], str. 121.

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

Důkaz. Viz [R], str. 121.

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

Důkaz. Viz [R], str. 122.

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

Poznamenejme, že pro libovolné prvočíslo p je $\mathbb{Z}/p\mathbb{Z}$ těleso o p prvcích, můžeme tedy přímo položit $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. Naproti tomu $\mathbb{Z}/p^n\mathbb{Z}$ pro $n > 1$ není těleso, proto $\mathbb{F}_{p^n}$ není izomorfní s $\mathbb{Z}/p^n\mathbb{Z}$.

Definice. *Nechť $m \in \mathbb{N}$. Existuje-li $g \in \mathbb{Z}$ s vlastností $g^i \not\equiv 1 \pmod{m}$ pro všechna $i = 1, 2, \dots, \varphi(m) - 1$, nazývá se g primitivní kořen modulo m.*

Je tedy g primitivní kořen modulo m, právě když třída obsahující g je generátor grupy $(\mathbb{Z}/m\mathbb{Z})^\times$. Podle věty 9 pak pro libovolná nezáporná celá čísla i, j platí $g^i \equiv g^j \pmod{m}$ právě když $i \equiv j \pmod{\varphi(m)}$.

Věta 14. *Buď R konečné těleso. Pak $R^\times$ je cyklická grupa.*

Důkaz. Viz [R], str. 90.

Věta 15. *Nechť p je liché prvočíslo a $n \in \mathbb{N}$. Pak existuje primitivní prvek modulo p^n.*

Důkaz. Podle věty 14 existuje primitivní kořen g_1 modulo p. Ukážeme, že také existuje primitivní kořen g modulo p, splňující $g^{p-1} \equiv 1 + p \pmod{p^2}$. Zmíněné g hledáme ve tvaru $g = g_1 + xp$. Jistě je g primitivní kořen modulo p a platí

$$g^{p-1} = (g_1 + xp)^{p-1} \equiv g_1^{p-1} + (p-1)xpg_1^{p-2} \equiv g_1^{p-1} - xpg_1^{p-2} \pmod{p^2}.$$

Protože $g_1^{p-1} \equiv 1 \pmod{p}$, je $c = \frac{1}{p}(g_1^{p-1} - 1) \in \mathbb{Z}$, tj. $g^{p-1} \equiv 1 + p(c - xg_1^{p-2}) \pmod{p^2}$. Hledáme $x \in \mathbb{Z}$ s vlastností $c - xg_1^{p-2} \equiv 1 \pmod{p}$, tj. $xg_1^{p-2} \equiv c - 1 \pmod{p}$ a takové x existuje podle věty 4.

Dokážeme indukcí vzhledem k $n \ge 2$, že toto g je primitivní kořen modulo p^n a že platí $g^{(p-1)p^{n-2}} \equiv 1 + p^{n-1} \pmod{p^n}$. Pro $n = 2$ jsme hotovi. Předpokládejme, že $n > 2$

a že dokazované tvrzení platí pro $n-1$. Označme k řád prvku g v $(\mathbb{Z}/p^n\mathbb{Z})^\times$. Pak platí $g^k \equiv 1 \pmod{p^n}$, tedy i $g^k \equiv 1 \pmod{p^{n-1}}$, odkud $(p-1)p^{n-2} = \varphi(p^{n-1}) | k$. Naopak $k|\varphi(p^n) = (p-1)p^{n-1}$. V jedné z předchozích relací tedy nastává rovnost. Z indukčního předpokladu víme, že

$$g^{(p-1)p^{n-3}} \equiv 1 + p^{n-2} \pmod{p^{n-1}},$$

odkud

$$g^{(p-1)p^{n-2}} \equiv (1 + p^{n-2})^p \pmod{p^n}.$$

Skutečně, platí-li $a \equiv b \pmod{p^{n-1}}$ pro nějaká $a, b \in \mathbb{Z}$, pak

$$a^p - b^p = (a-b)(a^{p-1} + a^{p-2}b + \cdots + b^{p-1})$$

je dělitelné p^n, neboť první závorka je dělitelná p^{n-1} a druhá p, protože z $a \equiv b \pmod{p}$ plyne

$$a^{p-1} + a^{p-2}b + \cdots + b^{p-1} \equiv pa^{p-1} \equiv 0 \pmod{p}.$$

Proto platí

$$g^{(p-1)p^{n-2}} \equiv (1 + p^{n-2})^p \equiv (1 + p^{n-1}) \pmod{p^n}$$

a odtud $k \neq \varphi(p^{n-1})$. Je tedy $k = \varphi(p^n)$ a g je primitivní kořen modulo p^n.

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

Důkaz. Je-li n složené nebo $n = 2$, je tvrzení zřejmé. Nechť je tedy n liché prvočíslo a označme g primitivní kořen modulo n. Pak platí $g^{\frac{1}{2}(n-1)} \equiv -1 \pmod{n}$, neboť $n | g^{(n-1)} - 1 = (g^{\frac{1}{2}(n-1)} - 1)(g^{\frac{1}{2}(n-1)} + 1)$ a $n \nmid (g^{\frac{1}{2}(n-1)} - 1)$. Odtud plyne

$$(n-1)! \equiv \prod_{i=0}^{n-2} g^i = g^{\frac{1}{2}(n-1)(n-2)} \equiv (-1)^{n-2} = -1 \pmod{n}.$$

5. Rozklad přirozeného čísla na součin prvočísel

Přistupme nyní k základnímu problému celé naší přednášky, totiž k rozkládání přirozeného čísla na prvočinitele. Celý problém můžeme rozdělit na tři úkoly:

1. Je-li dáno přirozené číslo 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 (tzv. *test na složenost*).
2. Víme-li, že N je asi prvočíslo, dokázat, že N skutečně prvočíslem je, nebo to vyvrátit. (tzv. *test na prvočíselnost*).
3. Víme-li, že je 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 $\frac{N}{d}$.

Než přejdeme k rafinovanějším metodám, je vhodné se zastavit u naivní metody, ve které zkoušíme dělit N postupně všemi prvočísly 2, 3, 5, 7, 11, … až do jisté hranice. Pokud jsme schopni to provést pro všechna prvočísla nepřevyšující $\sqrt{N}$, provedeme tím všechny tři úkoly současně. Avšak i v případě, kdy N je natolik velké, že dělení N všemi prvočísly nepřevyšujícími $\sqrt{N}$ by bylo příliš zdlouhavé, bývá výhodné zkoušet dělit N všemi prvočísly až do jisté hranice, abychom se zbavili malých faktorů (uvědomte si, že čím je prvočíslo menší, tím větší je pravděpodobnost, že dělí náhodně zvolené přirozené číslo). Budeme předpokládat, že máme uloženu tabulku prvočísel $p[1] = 2$, $p[2] = 3$, $p[3] = 5$, $p[4] = 7$, $\ldots$, $p[k]$. 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, pochopitelně nemá smysl).

Algoritmus (Pokusné dělení). *Nechť je dána tabulka prvočísel* $p[1] = 2, p[2] = 3, p[3] = 5, p[4] = 7, \ldots, p[k]$ *(kde* $k > 3$*), vektor* $t = [6, 4, 2, 4, 2, 4, 6, 2]$*, a číslo* j *takové, že* $j = 0$ *(resp. 1, 2, 3, 4, 5, 6, 7) právě tehdy, když* $p[k]$ *po dělení 30 dává zbytek 1 (resp. 7, 11, 13, 17, 19, 23, 29). Konečně, mějme zvolenu horní hranici* $B \geq p[k]$ *(v podstatě proto, abychom algoritmem neztratili příliš mnoho času).*

Pro dané přirozené číslo N *algoritmus hledá úplný rozklad* N *a jestliže se mu to nepodaří, v nalezeném rozkladu největší činitel může být dělitelný pouze prvočísly většími než* B *a všichni ostatní činitelé jsou prvočísla.*

1. *[Inicializace] Je-li* $N \leq 5$*, pak vytiskni odpovídající rozklad* N *a skonči. Jinak polož* $i \leftarrow -1$, $m \leftarrow 0$, $l \leftarrow [\sqrt{N}]$.
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] Polož* $r \leftarrow N \bmod d$*. Je-li* $r = 0$*, vytiskni* d *jako činitele v hledaném rozkladu a polož* $N \leftarrow \frac{N}{d}$, $l \leftarrow [\sqrt{N}]$ *a opakuj krok 3.*
4. *[Prvočíslo?] Je-li* $d \geq l$*, 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, avšak může být dělitelný jen prvočísly většími než* B *a skonči. Jinak jdi na 3.*

Poznámky k algoritmu. V průběhu výpočtu je $i = -1$, dokud používáme naši tabulku prvočísel, pak už je stále $i \geq 0$.

Důkaz algoritmu neuvádíme pro jeho jednoduchost.

Tento algoritmus by neměl být používán pro úplný rozklad N, pokud N není malé (řekněme $N < 10^8$), protože pro větší N existují lepší metody. Na druhé straně je užitečný 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 je 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.

Není nutné počítat $[\sqrt{N}]$, stačí test v kroku 4 nahradit testem zda $d \geq q$, kde q je podíl po dělení čísla N číslem d se zbytkem, který byl získán jako vedlejší produkt při dělení v kroku 3.

6. Testy na složenost

Musíme si zvolit 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, která dává dokonce nutnou a dostatečnou 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 $(N-1)! \bmod N$.

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 a počítat inverzní prvky).

Algoritmus (Binární umocňování zprava doleva). *Pro dané $g \in G$ a dané celé číslo n algoritmus počítá g^n v grupě $(G,\cdot)$.*

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; označíme-li N', y' a z' nové hodnoty proměnných N, y a z po provedení kroků 2 a 3, platí

a) pro N sudé: $y' = y$, $N' = \frac{N}{2}$, $z' = z^2$, tedy $y' \cdot (z')^{N'} = y \cdot z^{2\frac{N}{2}} = y \cdot z^N$;

b) pro N liché: $y' = y \cdot z$, $N' = \frac{N-1}{2}$, $z' = z^2$, tedy $y' \cdot (z')^{N'} = y \cdot z \cdot z^{2\frac{N-1}{2}} = y \cdot z^N$.

Je zřejmé, že při skončení algoritmu je tato hodnota v y.

Odhad časové náročnosti algoritmu. Je jasné, že 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 $a + b - 1 \leq 2[\log_2 |n|] + 1$. Například, je-li $G = (\mathbb{Z}/m\mathbb{Z})^\times$, 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 $n \neq 0$, předpokládáme, že je 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, platí

a) pro $N < E'$: $E' = \frac{E}{2}$, $y' = y^2$, $N' = N$, tedy $(y')^{E'} \cdot z^{N'} = (y^2)^{\frac{E}{2}} \cdot z^N = y^E \cdot z^N$;

b) pro $N \geq E'$: $E' = \frac{E}{2}$, $y' = y^2 \cdot z$, $N' = N - E'$, tedy $(y')^{E'} \cdot z^{N'} = (y^2 \cdot z)^{\frac{E}{2}} \cdot z^{N - \frac{E}{2}} = y^E \cdot z^N$.

V průběhu celého výpočtu platí $N \geq 0$ a vždy před provedením kroku 2 je $N < E$. Proto při skončení algoritmu při $E = 1$ nutně platí $N = 0$ a tedy $g^n = y$. Je jasné, že kroky 2 a 3 se provedou právě e krát a tedy algoritmus se jistě zastaví.

Nevýhody oproti předchozímu algoritmu: je třeba před začátkem 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 je to, že 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 $|a|$ menší než je základ naší poziční soustavy, je násobení g pouze lineárního času a ne kvadratického, jak je tomu u obecného násobení v grupě G. 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.

Nyní, když vidíme, že umocňování v okruhu zbytkových tříd můžeme opravdu provádět rychle, si rozmysleme, jak užít Fermatovu větu pro test na složenost. 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 $a \in \mathbb{Z}$, $1 \leq a < N$, a počítat $a^{N-1} \bmod N$. 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.

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 následující věta, kterou zde uvádím bez důkazu.

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.

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.

Test založený na výpočtu $a^{\frac{N-1}{2}} \bmod N$ a kontrole, zda je výsledek 1 nebo $N-1$ 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

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

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

Na druhou stranu, tento test 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

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

Proto je třeba podmínku, kterou budeme testovat, ještě více zesílit. Algoritmus, navržený Millerem a Rabinem, užívá následujícího tvrzení:

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.

Test Millera a Rabina, založený na předchozí větě, s vysokou pravděpodobností objeví každé složené číslo, neboť platí následující věta.

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:*

$$a^q \equiv 1 \pmod N \text{ nebo existuje } e \in \{0, 1, \ldots, t-1\} \text{ splňující } a^{2^e q} \equiv -1 \pmod N.$$

Důkaz. Pro značnou délku prozatím důkaz neuvádím.

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 $e \leftarrow 0$, $b \leftarrow a^q \bmod N$. Je-li $b = 1$, jdi na 4.*
3. *[Umocňuj na druhou] Dokud je $b \neq N - 1$ a $e \leq t - 2$, opakuj $b \leftarrow b^2 \bmod N$, $e \leftarrow e + 1$. 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.*

Důkaz správnosti algoritmu. Algoritmus testuje podmínku z předchozí 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}.

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.

7. Testy na prvočíselnost

V této kapitole si uvedeme tzv. $N - 1$ test Poclingtona a Lehmera. Je založen na následující větě.

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*

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

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. Protože každý kladný dělitel d čísla N je součinem prvočíselných dělitelů čísla N, stačí větu dokázat pouze pro případ, kdy je d prvočíslo. Podle Fermatovy věty platí $a_p^{d-1} \equiv 1 \pmod d$, neboť $(a_p, N) = 1$. Protože $(a_p^{\frac{N-1}{p}} - 1, N) = 1$, platí $a_p^{\frac{N-1}{p}} \not\equiv 1 \pmod d$.

Označme

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

Podle věty 9 čtvrté kapitoly 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 $p^{\alpha_p} \mid d-1$.

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*

(a) jestliže pro každé prvočíslo $p \mid F$ můžeme najít $a_p \in \mathbb{Z}$ splňující $(*)$ z předchozí věty, pak je N prvočíslo;

(b) je-li N prvočíslo, pak pro libovolné prvočíslo $p \mid N-1$ existuje $a_p \in \mathbb{Z}$ splňující $(*)$.

Důkaz. (a) Z předchozí věty plyne, že libovolný kladný dělitel čísla N splňuje $d \equiv 1 \pmod F$, tedy mezi čísly $2, 3, \ldots, F$ není žádný dělitel čísla N. Protože $F > \sqrt{N}$, jsme hotovi.

(b) Stačí za a_p zvolit primitivní kořen modulo N.

Nevýhodou předchozí věty je to, že musíme dostatečně rozložit číslo $N - 1$. Jistě to bude číslo sudé, ale rozložit na prvočinitele číslo $\frac{N-1}{2}$ může být notně obtížné. Poměrně snadno lze ale získat všechny prvočíselné dělitele tohoto čísla až po vhodnou hranici B (např. algoritmem pokusného dělení). Této informace pak můžeme s výhodou využít.

Věta. *Nechť $N \in \mathbb{N}$, $N > 1$. Předpokládejme, že můžeme psát $N - 1 = F \cdot U$, kde $(F, U) = 1$ a 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 $(*)$ 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, pak požadovaná $a_p, a_U \in \mathbb{Z}$ vždy existují.

Důkaz. Nechť d je prvočíselný dělitel čísla N. Z předchozí věty dostáváme $d \equiv 1 \pmod F$. Označme

$$e = \min\{n \in \mathbb{N};\ a_U^n \equiv 1 \pmod d\}$$

(takové e existuje, protože $(a_U, N) = 1$). Z věty 9 čtvrté kapitoly vyplývá $e \mid d-1$, $e \mid N-1$ 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 $d > F \cdot (e, U) > FB \geq \sqrt{N}$. Je tedy N prvočíslo.

Naopak, je-li N prvočíslo, stačí za a_U zvolit primitivní kořen modulo N.

Příklad. Aplikujme důsledek na k-té Fermatovo číslo $N = 2^{2^k} + 1$, kde $k \in \mathbb{N}$. Platí tedy: N je prvočíslo právě když existuje $a \in \mathbb{Z}$ splňující

$$a^{2^{2^k}} \equiv 1 \pmod N \quad \text{a} \quad (a^{2^{2^k-1}} - 1, N) = 1.$$

Je možné dokázat, že je-li N prvočíslo, platí $3^{2^{2^k-1}} \equiv -1 \pmod N$ (důkaz vyžaduje hlubší znalosti, uvádím jej pro ty, kteří znají kvadratický zákon reciprocity):

$$\left(\tfrac{3}{N}\right) = \left(\tfrac{N}{3}\right) \cdot (-1)^{\frac{3-1}{2}\frac{N-1}{2}} = \left(\tfrac{N}{3}\right) = \left(\tfrac{2}{3}\right) = -1.$$

Je zřejmé, že jestliže naopak platí $3^{2^{2^k-1}} \equiv -1 \pmod N$, pak $a = 3$ splňuje stanovenou podmínku a tedy N je prvočíslo.

Poznámka k implementaci algoritmu. 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 $(b_p - 1, N) = 1$. (Pokud by snad $c_p \not\equiv 1 \pmod N$), znamenalo by to, že N není prvočíslo.) 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 pravděpodobně nezastavil.

Poznámka k časové náročnosti algoritmu. Je obtížné hovořit o časové náročnosti – předně, není-li N prvočíslo, algoritmus se nemusí zastavit. Ale i pro prvočísla je těžké stanovit jakýkoli odhad, protože záleží na tom, jak snadno lze rozkládat číslo $N-1$. Je také 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).

Poznámka o možném zobecnění algoritmu. Je-li N prvočíslo, podle věty 12 čtvrté kapitoly existuje těleso $\mathbb{F}_{N^2}$ o N^2 prvcích. Podle věty 14 stejné kapitoly je jeho multiplikativní grupa 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 test analogický testu Poclingtona a Lehmera, kde však bude vystupovat faktorizace čísla $N+1$ místo $N-1$. Je dokonce možné informace, získané z obou testů, zkombinovat.

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 $\frac{N^6-1}{(N^3-1)(N+1)} = N^2 - N + 1$) a podobně. Vždy nám však už vycházejí čísla podstatně větší než N a tedy pravděpodobně obtížně rozložitelná.

8. Některé nezbytnosti z algebraické geometrie

V celé kapitole předpokládáme, že 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 $\sim$ definované 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 $[x_1, \ldots, x_{n+1}]$.*

Poznámka. Nechť $x_1, \ldots, x_{n+1} \in K$, 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 $[x_1, \ldots, x_n] \in P^{n-1}(K)$. Můžeme tedy n-rozměrný projektivní prostor „rozdělit" na n-rozměrný afinní prostor a na „část nevlastních bodů", kterou je $(n-1)$-rozměrný projektivní prostor. Toto rozdělení *není* kanonické – lze to provést mnoha způsoby.

Poznámka. 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 $P^2(K)$.

Poznámka. 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 $F_1(x, y, z) = x^3 + x^2 z - y^2 z$ 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. 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 $y = 0$. Singulární bod na $\mathcal{C}_2$ tedy neexistuje a proto $\mathcal{C}_2$ není singulární křivka.

Definice. *Eliptická křivka nad 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. Precizní zavedení tohoto pojmu je však časově náročné a proto od něj upouštím. Tento pojem je zde zapotřebí pouze proto, abychom si ukázali, že vlastně neztrácíme nic na obecnosti, omezíme-li se na eliptické křivky speciálního tvaru. Nebudeme tedy ani dokazovat následující větu.

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 $O = [0,1,0]$.

Poznámka. Každá eliptická křivka ve výše uvedeném tvaru má jeden nevlastní bod (totiž O) 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.

Omezení. 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 $y^2 = x^3 + a_4x + a_5$.

9. Aritmetika eliptických křivek

V celé kapitole předpokládáme, že K je těleso charakteristiky různé od 2 a 3 a že je dána eliptická křivka $(\mathcal{E}, O)$, kde $O = [0,1,0]$ a $\mathcal{E}$ je dána Weierstrassovou rovnicí

$$y^2 = x^3 + ax + b,$$

kde $a, b \in K$. Jak plyne z následující věty, důsledkem našich předpokladů je, že $4a^3 + 27b^2 \neq 0$.

Věta. *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. $4a^3 + 27b^2 = 0$. Zbývá ověřit, že $[\gamma, 0, 1]$ vyhovuje rovnici, což je snadné:

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

Poznámka. Naším cílem je definovat na $\mathcal{E}$ grupovou operaci $+$. Je třeba tedy najít nějaký předpis, jak dvěma bodům z $\mathcal{E}$ přiřadit třetí. Máme-li dány dva různé body z $\mathcal{E}$, můžeme jimi vést přímku. Dosazením rovnice této přímky do Weierstrassovy rovnice získáme kubickou rovnici, jejíž dva kořeny známe. Existuje proto třetí kořen, který lze snadno spočítat. Tento třetí kořen odpovídá třetímu průsečíku přímky s eliptickou křivkou (který může popřípadě i splynout s některým z daných bodů).

Podobně můžeme postupovat i v případě, kdy vezmeme dvakrát týž bod z $\mathcal{E}$: sestrojíme v tomto bodě tečnu k $\mathcal{E}$. Protože K nemusí být těleso reálných čísel, je možná vhodné upřesnit, co rozumíme touto tečnou: je to taková přímka, že po dosazení její rovnice do rovnice eliptické křivky dostaneme kubickou rovnici, ve které bod dotyku odpovídá kořenu alespoň dvojnásobnému. Zbylý kořen pak odpovídá dalšímu průsečíku přímky s eliptickou křivkou (který by opět mohl splynout s daným bodem).

V obou případech nám dvojice bodů z $\mathcal{E}$ určila další bod z $\mathcal{E}$. Tato binární operace by nám však nevytvořila z $\mathcal{E}$ grupu (je zřejmé, že tato operace obecně nemá neutrální prvek).

Operaci $+$ na $\mathcal{E}$ definujeme takto: pro libovolné body A, $B \in \mathcal{E}$ označme C bod z $\mathcal{E}$ jimi určený. Součtem $A + B$ pak nazveme bod z $\mathcal{E}$ určený body C a O.

Příklad. Nevlastní přímka $z = 0$ má s $\mathcal{E}$ trojnásobný bod dotyku O: dosazením $z = 0$ do rovnice $y^2z = x^3 + axz^2 + bz^3$ dostaneme rovnici $x^3 = 0$, která má trojnásobný kořen $x = 0$. Proto pro $A = B = O$ je i $C = O$ a tedy i $A + B = O$. Je tedy $O + O = O$.

Uvažme případ $A = O$, $B \neq O$. Pak $B = [\alpha, \beta, 1]$ pro vhodné α, $\beta \in K$. Přímka určená body O, B má rovnici $x = \alpha z$ (nevlastním bodem této přímky je O, vlastní body jsou $[\alpha, y, 1]$ pro všechna $y \in K$). Je zřejmé, že $C = [\alpha, -\beta, 1]$ a že třetí bod na přímce určené C a O je B. Ověřili jsme tedy, že platí $O + B = B$.

Je zřejmé, že operace $+$ je komutativní. Víme tedy, že $(\mathcal{E}, +)$ je komutativní grupoid s neutrálním prvkem O a že pro každý bod $A \in \mathcal{E}$ existuje bod $B \in \mathcal{E}$ splňující $A+B = O$ (je-li $A = O$, vezmeme $B = O$; je-li $A = [\alpha, \beta, 1]$, vezmeme $B = [\alpha, -\beta, 1]$).

Poznámka. K důkazu tvrzení, že $(\mathcal{E}, +)$ je komutativní grupa, je třeba dokázat, že $+$ je asociativní operace. To ale není snadné a omezíme se pouze na konstatování tohoto faktu bez důkazu.

Věta. *Na eliptické křivce $(\mathcal{E}, O)$ nad K definujeme operaci $+$ 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*

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

pak platí $[\sigma, \tau, 1] \in \mathcal{E}$ a klademe $A + B = [\sigma, \tau, 1] \in \mathcal{E}$.
Pak $(\mathcal{E}, +)$ je komutativní grupa.

Poznámka. Důkaz toho, že $+$ je asociativní operace, je mimo možnosti naší přednášky. Doporučuji si ale samostatně ověřit, že vzorce uvedené ve větě odpovídají výše uvedené geometrické konstrukci.

Eliptické křivky tvoří komutativní grupu i nad tělesy charakteristiky 2 a 3. Je však třeba uvažovat obecnější tvar Weierstrassovy rovnice a proto i vzorce popisující sčítání bodů jsou komplikovanější.

Následující věty budeme potřebovat v dalším textu. Jde o velmi hluboká tvrzení, které mohu uvádět jen bez důkazu.

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 $|\mathcal{E}_n| = p^n + 1 - 2\operatorname{Re}\alpha_p^n$, 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 $d_1 \mid q - 1$.*

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:*

$$\begin{aligned} (\mathbb{Z}/m\mathbb{Z}, +) \qquad & \textit{pro } 1 \le m \le 10 \textit{ nebo } m = 12 \\ (\mathbb{Z}/2\mathbb{Z}, +) \times (\mathbb{Z}/2m\mathbb{Z}, +) \qquad & \textit{pro } 1 \le m \le 4 \end{aligned}$$

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

10. Opět testy na prvočíselnost

Předpokládáme, že $N > 1$ je liché přirozené číslo, o kterém jsme testem Millera a Rabina zjistili, že N je asi prvočíslo. Naším cílem je vyložit test na prvočíselnost, využívající eliptické křivky. Začněme ale zopakováním $N-1$ testu Poclingtona a Lehmera. Pracujeme v něm se známým prvočíslem p dělícím $N-1$ (přičemž p^{α_p} je nejvyšší mocnina p dělící $N-1$) a s jistým neznámým dělitelem d čísla N (budeme předpokládat, že i d je prvočíslo). Pak můžeme uvážit homomorfismus okruhů

$$f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/d\mathbb{Z}, \quad \text{kde} \quad a + N\mathbb{Z} \mapsto a + d\mathbb{Z}$$

(homomorfismus f je dobře definován, neboť $d \mid N$). Protože je d prvočíslo, je druhý okruh těleso $\mathbb{F}_d = \mathbb{Z}/d\mathbb{Z}$. 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 $b = f(a_p + N\mathbb{Z}) \in \mathbb{F}_d$. Pak $b^{N-1} = 1$, $b^{\frac{N-1}{p}} \neq 1$ a tedy řád prvku b je dělitelný p^{α_p}, odkud $p^{\alpha_p} \mid d-1$, tedy $d \equiv 1 \pmod{p^{\alpha_p}}$.

Promysleme si nyní $N+1$ test. Potřebujeme znát prvočíslo p, které (v tomto případě) dělí $N+1$. Označme opět p^{α_p} nejvyšší mocninu p dělící $N+1$. Zvolme pevně $q \in \mathbb{Z}$, $(q, N) = 1$, takové, že $x^2 \equiv q \pmod N$ nemá řešení. (Takové q jistě existuje: zvolíme-li libovolné prvočíslo r dělící N a primitivní kořen g modulo r, pak tuto vlastnost mají všechna čísla nesoudělná s N, která jsou modulo r kongruentní s lichými mocninami g. Podle věty 5 čtvrté kapitoly je takových čísel alespoň polovina ze všech přirozených čísel menších než N.) Zkonstruujme okruh R takto:

$$(R, +) = (\mathbb{Z}/N\mathbb{Z}, +) \times (\mathbb{Z}/N\mathbb{Z}, +)$$

a pro libovolné $(a, b), (c, d) \in R$ položme $(a, b) \cdot (c, d) = (ac + qbd, ad + bc)$ (můžeme si místo (a, b) „představovat“ $a + \sqrt{q} \cdot b$). Snadno se ukáže, že R je komutativní okruh s jedničkou $(1, 0)$ (přesněji $(1 + N\mathbb{Z}, N\mathbb{Z})$). Opět budeme pracovat s jistým (neznámým) prvočíselným dělitelem d čísla N. Uvažme konečné těleso $\mathbb{F}_{d^2}$ a označme g generátor grupy $\mathbb{F}_{d^2}^\times$ (ten existuje podle věty 14 čtvrté kapitoly). Protože $\mathbb{F}_d \subset \mathbb{F}_{d^2}$, je $\mathbb{F}_d^\times$ podgrupa řádu $d-1$ cyklické grupy $\mathbb{F}_{d^2}^\times$ řádu d^2-1. Protože $(q,N)=1$, je i $(q,d)=1$ a tedy má smysl uvažovat $q \in \mathbb{F}_d^\times$ (přesněji $q+d\mathbb{Z} \in (\mathbb{Z}/d\mathbb{Z})^\times = \mathbb{F}_d^\times$). Proto $q = g^{(d+1)c}$ pro vhodné $c \in \mathbb{Z}$. Označme $s = g^{\frac{d+1}{2}\cdot c}$, pak $s^2 = q$. Uvažme dvojici homomorfismů $f_{1,2} : R \to \mathbb{F}_{d^2}$ definovanou takto: $f_1((a,b)) = a + sb$, $f_2((a,b)) = a - sb$. (Sami ověřte, že jde o homomorfismy okruhů). Předpokládejme, že jsme našli $\alpha \in R$ s vlastností $\alpha^{N+1} = 1$, $\alpha^{\frac{N+1}{p}} \neq 1$. (Jak takové α hledat: zvolíme nějaké $\beta = (a,b) \in R$ splňující $b \neq 0$ a položíme $\alpha = \beta^{N-1}$. Jestliže pak $\alpha^{N+1} \neq 1$, není R těleso a proto není N prvočíslo a jsme hotovi. Pokud $\alpha^{\frac{N+1}{p}} = 1$, což v případě prvočíselného N nastává s pravděpodobností $\frac{1}{p}$, zvolíme jiné β.) Můžeme předpokládat, že $\alpha^{\frac{N+1}{p}} = (x + N\mathbb{Z}, y + N\mathbb{Z})$, kde $(x-1,N) = 1$ nebo $(y,N) = 1$ (v opačném případě bychom dostali netriviálního dělitele N). Označme $\gamma_1 = f_1(\alpha)$, $\gamma_2 = f_2(\alpha)$. Platí $\gamma_1^{N+1} = 1$ a současně $\gamma_2^{N+1} = 1$.

Dokážeme sporem, že $\gamma_1^{\frac{N+1}{p}} \neq 1$ nebo $\gamma_2^{\frac{N+1}{p}} \neq 1$. Jestliže totiž $\gamma_1^{\frac{N+1}{p}} = 1$ a současně $\gamma_2^{\frac{N+1}{p}} = 1$, znamená to, že v tělese $\mathbb{F}_{d^2}$ platí $x+sy = x-sy = 1$, odkud plyne $2sy = 0$, a tedy $y = 0$, neboť $2s \in \mathbb{F}_{d^2}^\times$, a $x = 1$. Tyto rovnosti v tělese $\mathbb{F}_{d^2}$ znamenají $y \equiv 0 \pmod d$ a $x \equiv 1 \pmod d$, což je spor.

Existuje tedy $i \in \{1,2\}$ tak, že $\gamma_i^{N+1} = 1$ a $\gamma_i^{\frac{N+1}{p}} \neq 1$. Označme r řád prvku γ_i. Pak $p^{\alpha_p} \mid r$. Současně $r \mid d^2-1$, tedy $d^2 \equiv 1 \pmod{p^{\alpha_p}}$. Odtud lze odvodit $d \equiv \pm 1 \pmod{p^{\alpha_p}}$, je-li $p \neq 2$, popřípadě $d \equiv \pm 1 \pmod{2^{\alpha_2 - 1}}$, je-li $p = 2$.

V případě $N-1$ i $N+1$ testu jsme byli schopni odvodit jistou informaci o potenciálních dělitelích d čísla N za předpokladu, že jsme byli schopni nalézt alespoň některé prvočinitele čísla $N-1$ (resp. $N+1$). Obecně ale ve hledání dělitelů těchto čísel nemusíme být vůbec úspěšní (odhlédneme-li od zřejmého faktu, že čísla $N-1$ a $N+1$ jsou sudá a jedno z nich je dokonce dělitelné čtyřmi). Je proto potřeba provést analogické úvahy i v jiných situacích, tj. pro grupy o jiném počtu prvků než mají $\mathbb{F}_d^\times$ a $\mathbb{F}_{d^2}^\times$. Potřebujeme, aby počet prvků takové grupy nebyl o mnoho větší než d, aby byla reálná šance nalezení prvočíselného dělitele. Navíc je nutné, abychom byli schopni v této grupě počítat, přestože neznáme číslo d, ale jen jeho násobek N.

Výše popsaným požadavkům vyhovují eliptické křivky nad $\mathbb{F}_d$. Protože však d neznáme, výpočty v této grupě budeme provádět pomocí nějakého homomorfismu z grupy, ve které jsme schopni počítat. Tím by mohla být „eliptická křivka“ nad $\mathbb{Z}/N\mathbb{Z}$. Uvozovky zde stály proto, že pro složené N není $\mathbb{Z}/N\mathbb{Z}$ těleso a tedy tento pojem nebyl zaveden. Pokusme se to napravit. Pro stručnost označme $R = \mathbb{Z}/N\mathbb{Z}$.

Uvažme množinu M všech $(n+1)$-tic $(x_1,\dots,x_{n+1}) \in R^{n+1}$ takových, že zvolíme-li reprezentanty zbytkových tříd $x_1' \in x_1, \dots, x_{n+1}' \in x_{n+1}$, čísla $x_1',\dots,x_{n+1}',N$ jsou nesoudělná. Podobně jako pro tělesa nazveme n-rozměrným projektivním prostorem $P^n(R)$ nad R rozklad množiny M, příslušný ekvivalenci $\sim$ definované takto: pro libovolné $(n+1)$-tice $(x_1,\dots,x_{n+1})$, $(y_1,\dots,y_{n+1}) \in R^{n+1}$ je $(x_1,\dots,x_{n+1}) \sim (y_1,\dots,y_{n+1})$ právě tehdy, když existuje $\lambda \in R^\times$, které pro každé $i \in \{1,\dots,n+1\}$ splňuje podmínku $x_i = \lambda y_i$.

Bod projektivního prostoru $P^n(R)$ (tj. třídu rozkladu $M/\sim$) obsahující $(n+1)$-tici $(x_1,\dots,x_{n+1})$ budeme značit $[x_1,\dots,x_{n+1}]$.

Máme-li homogenní polynom $F(t_1,\dots,t_{n+1}) \in R[t_1,\dots,t_{n+1}]$ o $n+1$ proměnných nad R stupně k a $[x_1, \dots, x_{n+1}] \in P^n(R)$, má opět smysl položit

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

Pro libovolné $a, b \in R$ takové, že $4a^3 + 27b^2 \in R^\times$, můžeme definovat „eliptickou křivku“ $(\mathcal{E}, O)$ nad $\mathbb{Z}/N\mathbb{Z}$ takto: $O = [0, 1, 0]$ a

$$\mathcal{E} = \{[x, y, z] \in P^2(R);\ y^2 z = x^3 + axz^2 + bz^3\}.$$

Chceme nyní definovat na $\mathcal{E}$ operaci $+$. Uvažme proto vzorce ze druhé věty deváté kapitoly. Pro složené N nastanou komplikace v bodě 3, kdy $A = [\alpha, \beta, 1] \in \mathcal{E}$, $B = [\gamma, \delta, 1] \in \mathcal{E}$ a $B \neq -A$. Může se totiž stát, že v případě $\alpha \neq \gamma$ je $\alpha - \gamma \notin R^\times$ nebo že v případě $\alpha = \gamma$ je $\beta \notin R^\times$. Další komplikací je, že může existovat bod $[x, y, z] \in \mathcal{E}$ takový, že $z \neq 0$ a současně $z \notin R^\times$. Pro body tohoto typu žádné vzorce pro sčítání nemáme.

Proto budeme definovat sčítání bodů z $\mathcal{E}$ jen jako částečnou operaci: někdy sčítat lze (tj. při užívání zmíněných vzorců dělíme invertibilními prvky) a součtem je opět bod na $\mathcal{E}$ (což vyžaduje ověření, které si zde odpouštím); jindy sčítat nelze, neboť by bylo třeba dělit nenulovým prvkem z R, který není invertibilní (to však vzhledem k tomu, že $R = \mathbb{Z}/N\mathbb{Z}$, znamená nalezení netriviálního dělitele čísla N).

Poznamenejme, že je možná i druhá cesta, totiž definovat grupovou operaci pomocí projektivních souřadnic na celém $\mathcal{E}$. Tím se však definice operace značně komplikuje (je třeba rozlišit devět různých případů). Pro naše účely je to zcela zbytečné, neboť v okamžiku, kdy se dostaneme do situace, kdy nelze sčítat, znamená to, že jsme nalezli netriviálního dělitele čísla N a jsme hotovi.

Uvažujme nyní prvočíselný dělitel d čísla N. Pak máme homomorfismus okruhů $f: R \to \mathbb{F}_d$, přičemž $f(s + N\mathbb{Z}) = s + d\mathbb{Z}$. Protože $4a^3 + 27b^2 \in R^\times$, je

$$y^2 z = x^3 + f(a)xz^2 + f(b)z^3$$

rovnicí eliptické křivky $(\mathcal{E}_d, O_d)$ nad $\mathbb{F}_d$, kde $O_d = [0, 1, 0]$. Podstatné je, že nám f indukuje zobrazení $f': P^2(R) \to P^2(\mathbb{F}_d)$ určené předpisem

$$f'([\alpha, \beta, \gamma]) = [f(\alpha), f(\beta), f(\gamma)]$$

(zde jsme užili to, že $f(R^\times) \subset \mathbb{F}_d^\times$). Zúžení tohoto zobrazení na $\mathcal{E}$ je částečný homomorfismus $f': \mathcal{E} \to \mathcal{E}_d$ v tomto smyslu: jsou-li $A, B \in \mathcal{E}$ takové, že je $A + B$ definováno, pak platí $f'(A + B) = f'(A) + f'(B)$.

Věta. *Nechť $N > 1$ je přirozené číslo nesoudělné s 6, $R = \mathbb{Z}/N\mathbb{Z}$, $(\mathcal{E}, O)$ „eliptická křivka“ nad R. Předpokládejme, že jsme našli bod $P = [x, y, z] \in \mathcal{E}$ a prvočíslo q splňující*

1. *$q > (\sqrt[4]{N} + 1)^2$,*
2. *$q \cdot P$ (tj. součet q kopií bodu P) je definováno a platí $q \cdot P = O = [0, 1, 0]$,*
3. *$z \in R^\times$.*

Pak je N prvočíslo.

Důkaz. Předpokládejme, že N není prvočíslo. Označme d nejmenší prvočíslo dělící N. Pak platí $d < \sqrt{N}$. Užijeme-li částečný homomorfismus f', dostaneme z podmínek 2 a 3, že $f'(P)$ je řádu q v grupě $\mathcal{E}_d$, odkud

$$|\mathcal{E}_d| \geq q > (\sqrt[4]{N} + 1)^2 \geq (\sqrt{d} + 1)^2,$$

což je ale spor s Hasseho větou, podle které

$$\big||\mathcal{E}_d| - (d + 1)\big| < 2\sqrt{d}.$$

Na předchozí větě je založen test na prvočíselnost pomocí eliptických křivek. Otázkou zůstává, jak zvolit eliptickou křivku (tedy jak zvolit a, $b \in R$) a jak na ní najít bod P a prvočíslo q splňující předpoklady věty. Je jasné, že při hledání a, b, P, q můžeme postupovat, jako kdyby bylo N prvočíslo (o čemž jsme přesvědčeni, vždyť prošlo testem Millera a Rabina). I kdybychom je uhádli ze skleněné koule, výsledek je týž: větou dokážeme, že N skutečně prvočíslo je. Pro hledání a, b, P, q se užívají následující metody.

1. Goldwasser – Kilian. Existuje algoritmus Schoofa, který pro prvočíslo p počítá řád (tj. počet bodů) eliptické křivky nad $\mathbb{F}_p$ v čase $O(\ln^8 p)$. Postupujeme takto: zvolíme náhodně a, $b \in R$ tak, aby $4a^3 + 27b^2 \in R^\times$. 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, doufajíce, 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, $b \in R$.

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 $O(\ln^2 p)$. Bod P na křivce hledáme takto: náhodně zvolíme $c \in R$ a hledáme $d \in R$ tak, aby $d^2 = c^3 + ac + b$ (jde o kongruenci modulo N; d hledáme jako by bylo N prvočíslo, pak uděláme zkoušku, pokud nevýjde, 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 $P \neq [0, 1, 0]$, vzhledem k tomu, že při výpočtu používáme jen vzorců ze druhé věty deváté kapitoly, platí $P = [x, y, 1]$ pro nějaké x, $y \in R$. Spočítáme q-násobek bodu P v $(\mathcal{E}, +)$. 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 $[0, 1, 0]$, podle věty 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.

2. Atkin. Jeho metoda je založena na teoretických výsledcích, které bohužel notně převyšují možnosti naší přednášky, proto jen informačně: 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).

Poznámky k časové náročnosti. Test Golwassera a Kiliana (objevený v roce 1986) má spíše teoretický význam; je možné dokázat (za jakýchsi rozumných předpokladů o rozložení prvočísel v krátkých intervalech, že očekávaný čas výpočtu je $O(\ln^{12} N)$, tedy polynomiální (nejhorší možný čas výpočtu není možno stanovit, protože jde o pravděpodobnostní algoritmus). Atkinův test byl implementován Atkinem a Morainem v roce 1990 a je schopen dokazovat prvočíselnost čísel o zhruba 1000 dekadických cifrách v řádově týdnech strojového času na Sparc station. I v tomto případě je očekávaný čas výpočtu polynomiální (přesněji $O(\ln^6 N)$).

Další moderní test na prvočíselnost je tzv. metoda Jacobiho sum, která byla objevena v roce 1980 (Adleman, Pomerance a Rumely) a dále zjednodušena a implementována v roce 1981 (Cohen a Lenstra). Existuje jak v pravděpodobnostní, tak v deterministické verzi, která je však méně praktická. Pomerance a Odlyzko dokázali, že čas výpočtu tohoto algoritmu je $O((\ln N)^{C \ln\ln\ln N})$ pro jistou konstantu C. Tato časová náročnost se velmi blíží polynomiální (uvědomme si, že funkce $\ln\ln\ln N$ roste velmi pomalu: $\ln\ln\ln(10^{10}) \doteq 1.143145$, $\ln\ln\ln(10^{100}) \doteq 1.693632$, $\ln\ln\ln(10^{1000}) \doteq 2.046633$, $\ln\ln\ln(10^{10000}) \doteq 2.307013$).

11. Hledání netriviálního dělitele – Lehmannova metoda

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 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.

Věta (Lehmann). *Nechť N je liché přirozené číslo tvaru $N = pq$, kde p a q jsou prvočísla. Nechť celé číslo r splňuje*

$$1 \le r < \sqrt{N} \quad a \quad \sqrt{\frac{N}{r+1}} \le p \le \sqrt{N}.$$

Pak existují celá čísla x, y, k taková, že

1. $x^2 - y^2 = 4kN$, *kde* $1 \le k \le r$*;*
2. $x \equiv 1 \pmod 2$, *je-li k sudé, a* $x \equiv k + N \pmod 4$, *je-li k liché;*
3. $0 \le x - \sqrt{4kN} \le \dfrac{1}{4(r+1)}\sqrt{\dfrac{N}{k}}$

a navíc

$$p = \min\{(x+y, N), (x-y, N)\}.$$

Jestliže je N prvočíslo, pak celá čísla x, y, k splňující podmínky 1, 2 a 3 neexistují pro žádné přirozené $r < \sqrt{N}$.

Důkaz prozatím neuvádím.

Použití věty. Pro dané přirozené číslo N označme $r = \left[\sqrt[3]{N}\right]$. Metodou pokusného dělení ověříme, že N není dělitelné prvočísly nepřevyšujícími r (nebo nalezneme netriviálního dělitele). Pak existují prvočísla p, q tak, že $N = pq$ (víme, že N není prvočíslo). Předpokládáme-li, že $p \le q$, pak skutečně

$$\sqrt{\frac{N}{r+1}} \le \sqrt{\frac{N}{\sqrt[3]{N}}} = \sqrt[3]{N} < p \le \sqrt{N}.$$

Budeme pak postupně volit $k \in \{1, 2, \dots, r\}$ 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 p. 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 $m^2 \le n < (m+1)^2$.*
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 $x \in \mathbb{Z}$, platí $[(x+[\frac{n}{x}])/2] = [(x + \frac{n}{x})/2]$. Označme $q = [\sqrt{n}]$. Protože $\frac{1}{t}(t - \sqrt{n})^2 \ge 0$ pro libovolné $t > 0$, platí $\frac{1}{2}(t + \frac{n}{t}) \ge \sqrt{n}$, tedy $x \ge q$ je splněno v průběhu celého algoritmu. Předpokládejme, že se algoritmus zastavil, tj. že $y = [(x + \frac{n}{x})/2] \ge x$ a dokažme $x = q$. Předpokládejme $x \ge q + 1$. Pak $x > \sqrt{n}$ a platí

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

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 $2^e \le n < 2^{e+1}$ a položit $x \leftarrow 2^{1+[\frac{e}{2}]}$. Pak totiž $x^2 \le 2^{e+2} \le 4n$, $x^2 \ge 2^{e+1} > n$, tj. $\sqrt{n} < x \le 2\sqrt{n}$. Po provedení kroku 2 pak platí

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

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

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 $1925 = 5^2 \cdot 7 \cdot 11$. Podle věty 15 a věty 5 čtvrté kapitoly je pravděpodobnost, že náhodně zvolené přirozené číslo je kvadratický zbytek modulo 1925 rovna $\frac{11}{25} \cdot \frac{4}{7} \cdot \frac{6}{11} = \frac{24}{175}$, pro modul 1989 je dokonce rovna $\frac{4}{9} \cdot \frac{7}{13} \cdot \frac{9}{17} = \frac{28}{221}$. Provedeme-li test pro oba moduly, poběží předchozí algoritmus jen s pravděpodobností $\frac{96}{5525}$, tedy jen asi v 1,7% případů.

Algoritmus (Naplnění tabulek kvadratických zbytků). *Algoritmus sestaví vektory T_1 o délce 1989 a T_2 o délce 1925 tak, že pro každé $1 \le i \le 1988$ platí $T_1[i] = 1$*

právě když kongruence $x^2 \equiv i \pmod{1989}$ *má řešení a pro každé* $1 \le i \le 1924$ *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 $m = [\sqrt{n}]$. 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. Pro pevné $k \in \{1, 2, \dots, r\}$ probíhá x celá čísla z intervalu délky $\frac{1}{4(r+1)}\sqrt{\frac{N}{k}}$, přičemž $r = [\sqrt[3]{N}]$. Platí tedy, že $\frac{1}{4(r+1)}\sqrt{\frac{N}{k}}$ je řádu $O(k^{-\frac{1}{2}} N^{\frac{1}{6}})$ a časová náročnost pro pevné k je $O(k^{-\frac{1}{2}} N^{\frac{1}{6}} \ln^3 N)$. Sečtením přes všechna k dostáváme celkovou časovou náročnost

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

Přitom $\int_1^r k^{-\frac{1}{2}} dk = [2k^{\frac{1}{2}}]_1^r = 2\sqrt{r} - 2$, tedy časová náročnost je řádu

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

Počáteční pokusné dělení čísla N všemi prvočísly nepřevyšujícími r je řádu

$$O(N^{\frac{1}{3}} \ln^2 N),$$

celková časová náročnost metody je tedy $O(N^{\frac{1}{3}} \ln^3 N)$, což je výrazně lepší oproti časové náročnosti algoritmu pokusného dělení, která je $O(N^{\frac{1}{2}} \ln^2 N)$.

12. 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 je následující: 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. 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, lineární se však nezdá být vhodný. Promysleme si situaci s lineárním polynomem $f(x) = ax + b$. Vzhledem k tomu, že celá čísla a, b budeme volit, je rozumné očekávat, že se nepodaří zvolit a ani x_0 soudělné s N; je-li $(a, N) = 1$, je $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$ bijekce a tedy předperioda je nulová, periodu je kromě triviálních případů ($b = 0$ nebo $a = \pm 1$) obtížné určit. Budeme tedy volit f jako co nejjednodušší kvadratický polynom. Volba $f = x^2$ vhodná není (promyslete si sami, že pro $\varphi(N) = 2^e \cdot l$ s lichým l bude v případě $(x_0, N) = 1$ předperioda menší nebo rovna e a perioda dělitelem čísla r, kde r je nejmenší přirozené číslo splňující $2^r \equiv 1 \pmod{l}$). Podobně polynom $f = x^2 - 2$ není vhodný, neboť pokud bychom náhodou zvolili x_0 ve tvaru $x_0 = u + u^{-1}$, bylo by $x_1 = f(x_0) = (u + u^{-1})^2 - 2 = u^2 + u^{-2}$ atd. Perioda by tedy byla dělitelem periody pro $f = x^2$ a $x_0 = u$ (je možné dokázat, že podíl těchto period je tvaru 2^e, kde e je menší nebo rovno počtu prvočíselných dělitelů čísla N). Je ověřeno experimentálně, že polynom $f = x^2 + c$, kde $c \neq 0$ a $c \neq -2$, pracuje docela dobře, i když nejsme schopni určit ani periodu ani předperiodu.

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“.

13. 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*

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

Celá medoda je založena na následující myšlence: 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). *Nechť N je složené číslo, B předem daná hranice. Algoritmus zkouší najít netriviálního dělitele N. Má naději na úspěch, pokud existuje prvočíslo $p \mid N$, pro které $p-1$ je B-hladké. Předpokládáme, že 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 $g = 1$, vydej zprávu, že algoritmus neuspěl a skonči, jinak polož $i \leftarrow j$, $x \leftarrow y$ a jdi na 5. V opačném případě (tj. pro $i \le k$) polož $q \leftarrow p[i]$, $q_1 \leftarrow q$, $l \leftarrow [\frac{B}{q}]$.*
3. *[Spočti mocninu] Dokud $q_1 \le l$, 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. V opačném případě (tj. pro $g > 1$), je-li $g < N$, vytiskni g a skonči. Konečně, je-li $g = N$ (což nastane s velmi malou pravděpodobností), vytiskni zprávu, že algoritmus neuspěl a skonči.*

Poznamenejme, že 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 $B \ge p$. Uspěl by tedy za dobu srovnatelnou s algoritmem pokusného dělení.

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

Druhé stadium. 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 $B_1 < q \le B_2$, 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). *Nechť N je složené číslo, B_1 a B_2 předem dané hranice. Algoritmus zkouší najít netriviálního dělitele N. Má naději na úspěch, pokud existuje prvočíslo $p \mid N$, pro které $p-1$ je B_1-hladké nebo je to B_1-hladký násobek prvočísla mezi B_1 a B_2. Předpokládáme, že 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. V opačném případě jsme tímto algoritmem získali 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 nastane s velmi malou pravděpodobností), vytiskni zprávu, že algoritmus neuspěl (nebo zkus znovu pro $x \leftarrow 3$ místo $x \leftarrow 2$ v kroku 1 prvního stadia) a skonči.*
6. *[Neuspěl jsi?] Polož $g \leftarrow (P,N)$. Je-li $g>1$, jdi na 5. V opačném případě (tj. je-li $g=1$), vytiskni zprávu, že algoritmus neuspěl a skonči.*

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 $p-1$. Je 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 $\mathbb{F}_p$.

14. Hledání netriviálního dělitele – Lenstrova metoda 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

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

(kde bychom správně měli psát $a+N\mathbb{Z}$, $b+N\mathbb{Z}$ místo a, b) nám určuje „eliptickou křivku" $(\mathcal{E}, O)$ nad $\mathbb{Z}/N\mathbb{Z}$, přičemž $O = [0,1,0] \in P^2(\mathbb{Z}/N\mathbb{Z})$. Nechť p je nějaké (neznámé) prvočíslo dělící N. Předchozí rovnicí (kde a, b chápeme jako $a+p\mathbb{Z}$, $b+p\mathbb{Z} \in \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$) 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 vzorci z druhé věty deváté kapitoly 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 $f_p(P + Q) = f_p(P) + f_p(Q)$.

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, …, 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 musel mít ve třetí složce buď 0 nebo 1. Znamená to, že řád bodu $f_q(P)$ v grupě $(\mathcal{E}_q, +)$ musí být stejný pro všechna prvočísla q dělící N. To je ale značně nepravděpodobné. 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 ale 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. První nápad je 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. Avšak N není prvočíslo a řešit kvadratickou kongruenci modulo N je stejně obtížné jako najít netriviálního dělitele N. Proto zvolíme jiný postup: položíme $b = 1$, $P = [0,1,1]$ a volíme pouze a. Jistě potom $P \in \mathcal{E}$.

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 eliptických křivek, první stadium). *Nechť N je složené číslo nesoudělné s 6, B předem daná hranice. Algoritmus zkouší najít netriviálního dělitele N. Předpokládáme, že máme tabulku $p[1], p[2], \ldots, p[k]$ všech prvočísel menších nebo rovných 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 $q \leftarrow p[i]$, $q_1 \leftarrow q$, $l \leftarrow [\frac{B}{q}]$, $R \leftarrow P$.*
4. *[Násob bod na křivce] Dokud $q_1 \leq l$, opakuj $q_1 \leftarrow q \cdot q_1$. Pak zkus spočítat $P \leftarrow q_1 \cdot P$ na křivce $(\mathcal{E}, O)$. Pokud se to nepodařilo (tj. v průběhu výpočtu byl objeven nějaký nenulový prvek okruhu $\mathbb{Z}/N\mathbb{Z}$, 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.*

15. Další moderní metody hledání netriviálního dělitele

V současnosti se používají tři nejúčinnější metody hledání netriviálních dělitelů velkých čísel: Lenstrova metoda eliptických křivek, metoda kvadratického síta a metoda síta v číselném tělese. Všechny tři uvedené metody jsou subexponenciálního času.

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. Proto jsem i tuto metodu zařadil do našeho přehledu.

Základní myšlenka. Nechť N je (velké) složené přirozené číslo, jehož netriviálního dělitele hledáme. Budeme předpokládat, že jsme ověřili, že N není dělitelné žádnými „malými" prvočísly (tj. prvočísly menšími než jistá hranice) a také, že N není mocnina prvočísla (jednoduchý test, jak zjistit, že N není mocnina prvočísla, uvedeme na konci této kapitoly). Budeme hledat taková celá čísla x, y, 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.

Avšak náhodné hledání takových čísel x, y je beznadějný úkol. Trik, který je společný pro všechny tři zmíněné metody, spočívá v tom, že místo toho 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

$\sum_{k=1}^n \varepsilon_k(e_{0k}, e_{1k}, \dots, e_{mk})$ nulový vektor. Budeme-li nyní $\varepsilon_1, \dots, \varepsilon_n$ považovat za celá čísla, pak pro každé $i \in \{0, 1, \dots, m\}$ je číslo $v_i = \frac{1}{2}\sum_{k=1}^n \varepsilon_k e_{ik}$ celé (uvažte homomorfismus okruhů $\mathbb{Z} \to \mathbb{F}_2$, jehož jádrem je množina všech sudých čísel). Položíme-li pak

$$x = \prod_{k=1}^n x_k^{\varepsilon_k}, \qquad y = p_1^{v_1} p_2^{v_2} \cdots p_m^{v_m},$$

platí

$$x^2 = \prod_{k=1}^n x_k^{2\varepsilon_k} \equiv \prod_{k=1}^n \left((-1)^{e_{0k}} p_1^{e_{1k}} p_2^{e_{2k}} \cdots p_m^{e_{mk}}\right)^{\varepsilon_k} = (-1)^{2v_0} p_1^{2v_1} p_2^{2v_2} \cdot p_m^{2v_m} = y^2 \pmod N,$$

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

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 $\{p_1, \dots, p_m\}$ se nazývá báze faktorizace. Způsob, jak ji zvolit optimálně, se u jednotlivých metod liší.

Zmiňme se ještě o Gaussově eliminaci. Hledáme $\mathbb{F}_2$-lineární závislosti mezi řádky obrovské matice, která má však v každém řádku jen několik jedniček a zbytek tvoří nuly. Uložit celou matici do paměti by se nám patrně nepodařilo. Proto se u těchto „řídkých“ matic pro každý řádek uchovávají pouze indexy 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.

Uveďme ještě slíbený test, jak zjistit, že N není mocnina prvočísla. Je-li $B > 1$ takové, že N není dělitelné žádným prvočíslem menším než B, pak z toho, že $N = p^k$ pro nějaké prvočíslo p a nějaké $k \in \mathbb{N}$, $k > 1$, plyne $p \geq B$ a tedy $k \leq \frac{\log_2 N}{\log_2 B}$. Stačí tedy pro každé $k = 2, 3, \dots, \left[\frac{\log_2 N}{\log_2 B}\right]$ spočítat $x = [\sqrt[k]{N}]$ a zjistit, zda $x^k = N$. Snadno se vidí, že přitom můžeme vynechávat ta k, která nejsou prvočísla.

16. Některé nezbytnosti o řetězových zlomcích

Při výkladu této kapitoly jsem užil knihu [Ca].

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 $||q'\theta|| > ||q\theta||$.*

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

Důkaz. Nejprve budeme předpokládat $Q \in \mathbb{N}$. Uvažme $Q+1$ čísel $0, 1, \langle\theta\rangle, \langle 2\theta\rangle, \dots, \langle (Q-1)\theta\rangle$ a rozdělme je do Q intervalů $[0, \frac{1}{Q}), [\frac{1}{Q}, \frac{2}{Q}), \dots, [\frac{Q-1}{Q}, 1]$. 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 $[Q] + 1$.

Zafixujme do konce kapitoly číslo $\theta \in \mathbb{R} - \mathbb{Q}$. Ukážeme si, že z předchozí věty plyne, že existuje nekonečně mnoho $q \in \mathbb{N}$ splňujících

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

(Poznamenejme, že je možné dokonce dokázat více: číslo 1 na pravé straně může být nahrazeno číslem $\frac{1}{\sqrt{5}}$.)

Sestrojme posloupnost všech dobrých aproximací čísla θ. Jistě $q_1 = 1$ dává dobrou aproximaci čísla θ spolu s nějakým $p_1 \in \mathbb{Z}$ a platí $|q_1\theta - p_1| = ||\theta|| < \frac{1}{2}$. Protože $\theta \notin \mathbb{Z}$, je $||\theta|| \neq 0$ a věta 1 s $Q = ||q_1\theta||^{-1}$ zaručuje existenci $q \in \mathbb{N}$, které splňuje $||q\theta|| < ||q_1\theta||$. Nechť q_2 je nejmenší s touto vlastností a $p_2 \in \mathbb{Z}$ splňuje $||q_2\theta|| = |q_2\theta - p_2|$. Pak $(q_2, p_2) = 1$: pokud by totiž existovalo $t \in \mathbb{N}$, $t > 1$, splňující $p_2 = tp'$, $q_2 = tq'$, pro celá čísla p', q', pak by $||q_2\theta|| = |q_2\theta - p_2| = t|q'\theta - p'| \geq t||q'\theta|| > ||q'\theta||$, což by byl spor s definicí q_2. Protože $\theta \notin \mathbb{Q}$, je $||q_2\theta|| \neq 0$ a věta 1 s $Q = ||q_2\theta||^{-1}$ zaručuje existenci $q \in \mathbb{N}$, které splňuje $||q\theta|| < ||q_2\theta||$. Nechť q_3 je nejmenší s touto vlastností a $p_3 \in \mathbb{Z}$ splňuje $||q_3\theta|| = |q_3\theta - p_3|$. Opět $(q_3, p_3) = 1$. Tímto procesem dostáváme posloupnost přirozených čísel

$$1 = q_1 < q_2 < q_3 < \dots$$

a celých čísel $p_1, p_2, p_3, \dots$ splňujících

$$||q_n\theta|| = |q_n\theta - p_n|, \tag{2}$$

$$||q_{n+1}\theta|| < ||q_n\theta||, \tag{3}$$

$$||q\theta|| \geq ||q_n\theta|| \quad \text{pro všechna } q \in \mathbb{N},\ q < q_{n+1}. \tag{4}$$

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 (4) platí

$$q_n||q_n\theta|| < q_{n+1}||q_n\theta|| \leq 1 \tag{5}$$

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 se (4). Proto

$$(q_n\theta - p_n)(q_{n+1}\theta - p_{n+1}) < 0. \tag{6}$$

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 (5), neboť $|\theta - \frac{p_n}{q_n}| < \frac{1}{q_n^2}$.

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

Důkaz. Levá strana je celé číslo a platí

$$q_{n+1}p_n - q_np_{n+1} = q_n(q_{n+1}\theta - p_{n+1}) - q_{n+1}(q_n\theta - p_n), \tag{7}$$

odkud spolu s (5) a (6) plyne

$$0 < q_n||q_{n+1}\theta|| + q_{n+1}||q_n\theta|| = |q_{n+1}p_n - q_np_{n+1}| < 2q_{n+1}||q_n\theta|| \leq 2. \tag{7'}$$

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

$$q_{n+1}p_n - q_np_{n+1} = -(q_np_{n-1} - q_{n-1}p_n).$$

Důkaz. Plyne z (7) s přihlédnutím k (2), (3) a $q_{n+1} > q_n$, druhá část z (6) a lemmatu 2.

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

$$q_{n+1} = a_nq_n + q_{n-1}, \tag{8}$$

$$p_{n+1} = a_np_n + p_{n-1}, \tag{9}$$

$$|q_{n-1}\theta - p_{n-1}| = a_n|q_n\theta - p_n| + |q_{n+1}\theta - p_{n+1}|. \tag{10}$$

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 $q_{n+1} - q_{n-1} = a_nq_n$, $p_{n+1} - p_{n-1} = a_np_n$. Protože $q_{n+1} > q_{n-1}$, je $a_n > 0$. Konečně, (10) plyne z (8) a (9) díky (6).

Poznamenejme, že (10) dává jednoduchý vzorec pro výpočet a_n:

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

Poznámka. Vysvětleme si, odkud se vzal termín „řetězové zlomky". Předpokládejme, že $0 < \theta < \frac{1}{2}$. Pak $q_1 = 1$, $p_1 = 0$, a $q_1\theta - p_1 > 0$. Položíme-li $q_0 = 0$, $p_0 = 1$, $a_1 = q_2$, zůstane pro dodefinované hodnoty v platnosti lemma 2 i jeho důsledek, proto i lemma 3. Pak pro $n = 1$ z (10) dostáváme $1 = a_1\theta + ||q_2\theta||$, tedy $a_1 = [\frac{1}{\theta}]$. Označme $\theta_0 = 1$, $\theta_1 = \theta$ a pro $n \geq 2$ nechť $\theta_n = \frac{||q_n\theta||}{||q_{n-1}\theta||}$. Pak podle (10) platí $\theta_n^{-1} = a_n + \theta_{n+1}$ pro libovolné $n \in \mathbb{N}$. Odtud dostáváme

$$\theta = \frac{1}{\theta^{-1}} = \frac{1}{a_1 + \theta_2} = \frac{1}{a_1 + \frac{1}{\theta_2^{-1}}} = \frac{1}{a_1 + \frac{1}{a_2 + \theta_3}} = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\theta_3^{-1}}}} = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \theta_4}}} = \dots$$

Ukažme si ještě, jak se výpočet čísel a_n zjednoduší, je-li $\theta = \sqrt{D} \notin \mathbb{Q}$ pro $D \in \mathbb{N}$. Označme $d = [\theta]$. Podle poznámky za lemmatem 3 vzhledem k (6) a (2) platí

$$a_n = \left[-\frac{q_{n-1}\theta - p_{n-1}}{q_n\theta - p_n}\right] = \left[-\frac{(q_{n-1}\theta - p_{n-1})(q_n\theta + p_n)}{q_n^2 D - p_n^2}\right].$$

Podle důsledku lemmatu 2 je

$$a_n = \left[-\frac{\pm(-1)^n\sqrt{D} + q_{n-1}q_n D - p_{n-1}p_n}{q_n^2 D - p_n^2}\right],$$

přičemž znaménko $\pm$ určíme podmínkou $\pm(q_1\theta - p_1) < 0$, tj. platí-li $d \leq \theta < d + \frac{1}{2}$, je $q_1 = 1$, $p_1 = d$ a platí znaménko $-$, kdežto je-li $d + \frac{1}{2} \leq \theta < d + 1$, je $q_1 = 1$, $p_1 = d + 1$ a platí znaménko $+$. Dále je zřejmé, že pro výpočet čísla a_n můžeme ve výše uvedeném vzorci nahradit číslo $\sqrt{D}$ např. číslem $d + \frac{1}{2}$ a zcela se vyhnout reálné aritmetice. Při výpočtu čísel p_{n+1}, q_{n+1} můžeme pak užívat (8) a (9), jen je třeba dodefinovat p_0, q_0 tak, aby platilo lemma 2 i jeho důsledky, tj. $q_0 = 0$, $p_0 = \pm 1$, přičemž podmínka $0 > (q_0\theta - p_0)(q_1\theta - p_1) = -p_0(q_1\theta - p_1)$ určí vhodné znaménko pro p_0.

17. Metoda řetězových zlomků

Jak už bylo zmíněno v patnácté kapitole, budeme 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 $p_1, \dots, p_m$ 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 $p_1, \dots, p_m$ a tedy že získáme kongruenci požadovaného tvaru.

Předpokládejme, že jsme pomocí řetězových zlomků našli dobrou aproximaci $\frac{p}{q}$ čísla $\sqrt{kN}$, kde k je nějaké nepříliš velké přirozené číslo nedělitelné druhou mocninou prvočísla. Označme $t = p^2 - kNq^2$. Pak $p^2 \equiv t \pmod{N}$. Nalezněme odhad pro $|t|$. Podle (2), (5) a Lemmatu 1 předchozí kapitoly platí

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

tedy

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

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 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), pak z $p^2 \equiv t \pmod{N}$ a $(p')^2 \equiv t' \pmod{N}$ získáme kongruenci $x^2 \equiv \frac{tt'}{U^2} \pmod{N}$, kde x je řešení kongruence $Ux \equiv pp' \pmod{N}$.

18. Metoda kvadratického síta

Opět budeme 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 $e_{ik} \in \{0, 1\}$. Rozdíl oproti metodě řetězových zlomků je ve způsobu, jakým jsou tyto kongruence hledány.

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.

Jak tedy budeme postupovat: 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 (je-li dokonce n prvočíslo, lze použít Shanksův algoritmus, který je časové náročnosti $O(\ln^4 n)$).

Jak budeme čísla prosívat: na velmi dlouhém intervalu pro všechna celá čísla a spočítáme velmi zhruba $\ln |Q(a)|$ (ukážeme si, že stačí s chybou menší než 1, proto určitě nepoužijeme algoritmus pro logaritmus, ale nějaký jiný postup, například uvažujeme logaritmy o základu 2 a počítáme řád první jedničky binárního zápisu). Tyto hodnoty uložíme do vektoru indexovaného a. Pak pro všechny mocniny prvočísel p^k menší nebo rovny jisté hranici B, odečteme přibližnou hodnotu $\log_2 p$ 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 jen ta prvočísla, pro něž je N kvadratický zbytek).

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 $k = [\max_a \log_2 |Q(a)|]$, 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 $1 + k\varepsilon$. Naproti tomu pro nerozložitelné $Q(a)$ dostaneme číslo větší než $(\log_2 B) - 1 - (k - \log_2 B)\varepsilon$. Pro $\varepsilon < \frac{-2+\log_2 B}{2k-\log_2 B}$ je tedy druhé číslo větší než první. Místo k sem přitom můžeme dosadit číslo o jedna větší než bylo největší číslo ve vektoru před započetím prosívání.

Pak pro všechna a, pro které je $Q(a)$ rozložitelné, spočítáme znovu $Q(a)$ a rozložíme, čímž získáme kongruenci požadovaného tvaru. Máme-li dost místa v paměti, můžeme také ukládat v průběhu prosívání pro každou položku a prvočísla, jejichž logaritmy jsme odčítali (nebo alespoň několik největších z nich), což nám urychlí rozkládání.

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 $Q(x) = (x+[\sqrt{lN}\,])^2 - lN$ pro nějaké přirozené číslo l nedělitelné druhou mocninou prvočísla. V tom případě bychom však museli doplnit naši bázi faktorizace: připomeňme, že v ní máme 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 provádění Gaussovy eliminace větší matice.

19. Metoda kvadratického síta s více polynomy

Uvažujme obecný kvadratický polynom $Q(x) = Ax^2+2Bx+C$ takový, že $A, B, C \in \mathbb{Z}$ a $A > 0$. 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 $AQ(a) \equiv (Aa + B)^2 \pmod N$. Zvolme délku $2M$ intervalu, na kterém budeme prosívat a pokusme se optimálně zvolit konstanty A, B, C. Abychom měli šanci na rozložení čísla $|Q(a)|$ nad naší bází faktorizace, je vhodné, aby maximum funkce $|Q(x)|$ na intervalu prosívání bylo co možná nejmenší, proto interval zvolíme tak, aby minimum funkce $Q(x)$ padlo doprostřed, tj. prosívat budeme na intervalu $I = (-\frac{B}{A} - M, -\frac{B}{A} + M)$. Dále budeme požadovat, aby $Q(-\frac{B}{A} + M) \doteq -Q(-\frac{B}{A})$, tj. $A^2M^2 \doteq 2(B^2 - AC)$, odkud plyne

$$A \doteq \frac{\sqrt{2(B^2 - AC)}}{M}.$$

Je tedy

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

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 $M\sqrt{\frac{N}{2}}$.

Budeme tedy postupovat tak, že nejdříve zvolíme délku prosívání M. Pak budeme koeficienty jednotlivých polynomů $Q(x)$ generovat takto: zvolíme A blízko $\frac{\sqrt{2N}}{M}$, navíc tak, že A je prvočíslo a N je kvadratický zbytek modulo A. Pak nalezneme B tak, že $B^2 \equiv N \pmod A$ a konečně položíme $C = \frac{B^2-N}{A}$. Pak 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. To spočítáme pro všechny polynomy jednou a uschováme. Pak totiž kořeny polynomu $Q(x)$ modulo p^k vyhovují kongruenci $Ax \equiv -B \pm a_{p^k} \pmod{p^k}$. 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. Tato metoda byla s úspěchem použita při rozkládání devátého Fermatova čísla $N = 2^{2^9} + 1$ pomocí metody síta v číselném tělese v roce 1990. 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 (viz poznámku na konci patnácté kapitoly) získali matici o 72 413 řádcích a 72 213 sloupcích. Gaussovou eliminací této matice pak získali kongruenci, která jim určila netriviálního dělitele čísla N.

20. Základy algebraické teorie čísel

Definice. *Jsou-li K, L tělesa a je-li $K \subseteq L$, řekneme, že L je rozšířením tělesa K. Je-li navíc L jakožto vektorový prostor nad K konečněrozměrné, hovoříme o konečném rozšíření. Dimenzi L nad K značíme $[L : K]$.*

Definice. *Podtěleso komplexních čísel, které je konečným rozšířením $\mathbb{Q}$, se nazývá těleso algebraických čísel.*

Definice. *Nechť K je těleso algebraických čísel, $\alpha \in K$. Pokud existuje normovaný polynom $f(x) \in \mathbb{Z}[x]$, jehož je α kořenem, nazývá se α celé algebraické číslo.*

Poznámka. V předchozí definici je podstatný požadavek normovanosti. Je-li totiž $[K : \mathbb{Q}] = n$, pak $1, \alpha, \alpha^2, \dots, \alpha^n$ musí být $\mathbb{Q}$-lineárně závislé, a tedy existuje polynom $f(x) \in \mathbb{Z}[x]$ stupně nejvýše n tak, že $f(\alpha) = 0$.

Lemma 1. *Nechť K je těleso algebraických čísel, $\omega_1, \dots, \omega_n \in K$. Nechť M je aditivní grupa, generovaná $\omega_1, \dots, \omega_n$, tj.*

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

Je-li M okruh, pak je libovolný prvek M celé algebraické číslo.

Důkaz. Bez újmy na obecnosti můžeme předpokládat, že $\omega_1 \dots \omega_n \neq 0$. Buď $\alpha \in M$ libovolné. Protože pro každé $i = 1, \dots, n$ 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 $i = 1, \dots, n$. Odtud plyne, že $\det(\alpha E - (a_{ij})) = 0$, kde E je jednotková matice řádu n. Proto je α kořenem normovaného polynomu $f(x) = \det(xE - (a_{ij})) \in \mathbb{Z}[x]$.

Věta 1. *Nechť K je těleso algebraických čísel. Označme R množinu všech celých algebraických čísel $\alpha \in K$. Pak R je obor integrity a K je jeho podílové těleso.*

Důkaz. Abychom ověřili, že R je obor integrity, musíme dokázat, že pro libovolná $\alpha, \beta \in R$ jsou $\alpha+\beta$, $\alpha-\beta$ i $\alpha\beta$ celá algebraická čísla. Protože α a β jsou celá algebraická čísla, existují polynomy s celými koeficienty

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0, \qquad g(x) = x^m + b_{m-1}x^{m-1} + \cdots + b_1x + b_0$$

tak, že $f(\alpha) = 0$ a $g(\beta) = 0$. Pak ovšem platí

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

a tedy podgrupa M aditivní grupy tělesa K generovaná všemi součiny

$$(*) \qquad \alpha^i\beta^j, \qquad \text{kde } 0 \le i < n,\ 0 \le j < m,$$

tvoří okruh, neboť libovolný součin $\alpha^k\beta^l$ pro $k \ge 0$, $l \ge 0$ je možné vyjádřit jako $\mathbb{Z}$-lineární kombinaci prvků (*). Podle lemmatu 1 jsou $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta \in M$ celá algebraická čísla.

Zbývá dokázat, že K je podílové těleso okruhu R. Nechť $\alpha \in K$ je libovolné. Podle poznámky před lemmatem 1 existuje polynom $f(x) = a_kx^k + a_{k-1}x^{k-1} + \cdots + a_1x + a_0 \in \mathbb{Z}[x]$ tak, že $f(\alpha) = 0$ a $a_k \ne 0$. Pak ovšem číslo $\beta = a_k\alpha$ je kořenem polynomu

$$x^k + a_{k-1}x^{k-1} + \cdots + a_1a_k^{k-2}x + a_0a_k^{k-1} \in \mathbb{Z}[x],$$

a proto číslo β je celé algebraické. Je tedy $\alpha = \frac{\beta}{a_k}$ podílem dvou čísel z R. Dokázali jsme, že K je podílové těleso okruhu R.

Příklad. Označme

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

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 $\alpha = a + b\sqrt{10}$ je kořenem polynomu

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

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 tedy α kořenem nějakého normovaného polynomu $f(x) \in \mathbb{Z}[x]$. Je-li $b = 0$, je $\alpha \in \mathbb{Q}$ a podle věty 8.5 na straně 102 skript [R] platí $\alpha \in \mathbb{Z}$. Nechť tedy $b \ne 0$, tj. $\alpha \notin \mathbb{Q}$. Vydělme polynom $f(x)$ polynomem $g(x) = x^2 - 2a + (a^2 - 10b^2)$ se zbytkem: existují tedy polynomy $q(x)$, $r(x)$ tak, že $f(x) = q(x)g(x) + r(x)$ a přitom je $r(x)$ buď nulový nebo stupně nejvýše jedna. Dosazením α za x dostaneme $r(\alpha) = 0$, odkud vzhledem k $\alpha \notin \mathbb{Q}$ plyne, že $r(x)$ je nulový polynom. Připomeňme (viz [R], důkaz věty 8.7 na straně 103), že polynom s celočíselnými koeficienty takový, že největší společný dělitel jeho koeficientů je roven 1, se nazývá primitivní. Platí (viz tamtéž), že součin primitivních polynomů je opět primitivní polynom. Nechť u a v jsou racionální čísla taková, že polynomy $uq(x)$ a $vg(x)$ jsou primitivní (tím jsou čísla u a v určena jednoznačně až na znaménko). Pak je tedy primitivní i polynom $uvf(x) = uq(x) \cdot vg(x)$. Ovšem polynom $f(x)$ je normovaný s celými koeficienty a tedy primitivní. Je proto $uv = \pm 1$. Protože polynom $g(x)$ je normovaný, je normovaný i

$q(x)$ a tedy z definice u a v plyne, že $u, v \in \mathbb{Z}$. Proto $v = \pm 1$ a tedy $g(x)$ má celočíselné koeficienty: $c = 2a \in \mathbb{Z}$, $a^2 - 10b^2 \in \mathbb{Z}$. 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.

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

Poznámka. Zajímá nás, nakolik je aritmetika v okruhu R podobná aritmetice v $\mathbb{Z}$. Aritmetika v $\mathbb{Z}$ je poměrně jednoduchá díky větě o jednoznačném rozkladu na součin prvočísel: každé nenulové celé číslo je možno zapsat ve tvaru součinu vhodné jednotky (tj. invertibilního prvku, v tomto případě 1 nebo -1) a konečně mnoha ireducibilních prvků (tj. prvočísel), přičemž je tento rozklad jednoznačný až na pořadí činitelů.

Příklad. Ukažme si, že aritmetika R může být i odlišná: nechť K a R jsou jako v předchozím příkladě a definujme zobrazení (tzv. normu) $\mathcal{N} : R \to \mathbb{Z}$ předpisem

$$\mathcal{N}(a + b\sqrt{10}) = (a + b\sqrt{10})(a - b\sqrt{10}) = a^2 - 10b^2.$$

Snadno se nahlédne, že $\mathcal{N}(\alpha\beta) = \mathcal{N}(\alpha)\mathcal{N}(\beta)$ pro každé $\alpha, \beta \in R$. Dokažme, že množina všech jednotek okruhu R je rovna

$$E = \{\alpha \in R; \mathcal{N}(\alpha) = \pm 1\}.$$

Skutečně, je-li α jednotka okruhu R, existuje $\beta \in R$ tak, že $\alpha\beta = 1$, odkud plyne $1 = \mathcal{N}(\alpha\beta) = \mathcal{N}(\alpha)\mathcal{N}(\beta)$, a tedy $\mathcal{N}(\alpha) = \pm 1$. Opačná implikace je zřejmá.

V okruhu R můžeme rozložit

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

Přitom $\mathcal{N}(3) = 9$ a $\mathcal{N}(1 + \sqrt{10}) = -9$. 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. V R tedy neplatí věta o jednoznačném rozkladu čísel na součin ireducibilních faktorů.

Je zřejmé, že $3 + \sqrt{10}$ je jednotka okruhu R a proto i všechny její mocniny. Odtud plyne, že E je nekonečná. Dokažme, že platí

$$E = \{\pm(3 + \sqrt{10})^n;\ n \in \mathbb{Z}\}.$$

Budeme předpokládat existenci nějaké jednotky η okruhu R, pro kterou platí $\eta \notin E$ 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.

Poznámka. Kummer v polovině minulého století objevil způsob, jak jednoznačné rozkládání zachránit. Jak jsme viděli, věta o jednoznačném rozkladu prvků v okruzích celých algebraických čísel neplatí, platí zde ale věta o jednoznačném rozkladu ideálů. Každý nenulový ideál okruhu celých algebraických čísel je možné psát ve tvaru součinu prvoideálů a to jednoznačně až na pořadí činitelů.

Připomeňme některé použité pojmy: nulový ideál je ideál $\{0\}$, prvoideál je ideál I okruhu R, pro který platí: $I \neq R$ a pro každé α, $\beta \in R$ z $\alpha\beta \in I$ plyne $\alpha \in I$ nebo $\beta \in I$ (ekvivalentně: ideál I je prvoideál, právě když je faktorokruh R/I oborem integrity, viz [R], strany 110 a 111). Je ještě třeba definovat, co je to součin ideálů: jsou-li I, J ideály okruhu R, jejich součinem je ideál

$$IJ = \Big\{ \textstyle\sum_{i=1}^{n} \alpha_i \beta_i;\ n \in \mathbb{N},\ \alpha_i \in I, \beta_i \in J \ \text{ pro všechna } i = 1, \dots, n \Big\}.$$

Věta 2. *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 $P_1, \dots, P_n$ takové, že platí*

$$I = P_1 \dots P_n.$$

Důkaz je mimo možnosti naší přednášky.

Poznámka. Jak lze větu použít na rozkládání nenulového prvku $\alpha \in R$, který není jednotka? Rozložíme hlavní ideál

$$(\alpha) = \{\alpha\beta;\ \beta \in R\}.$$

(Užíváme následujícího označení: $(\alpha_1, \dots, \alpha_n)$ je ideál generovaný čísly $\alpha_1, \dots, \alpha_n$, tj. nejmenší ideál, který je obsahuje.) Pokud je R okruh hlavních ideálů (tj. každý ideál okruhu R je tvaru (α) pro vhodné $\alpha \in R$), dostaneme podle výše uvedené věty ireducibilní prvky $\pi_1, \dots \pi_n \in R$ tak, že

$$(\alpha) = (\pi_1) \dots (\pi_n) = (\pi_1 \dots \pi_n),$$

odkud plyne, že $\alpha = \varepsilon \cdot \pi_1 \dots \pi_n$ pro vhodnou jednotku ε. Odvodili jsme tedy, že je-li okruh celých algebraických čísel R okruhem hlavních ideálů, platí v něm věta o jednoznačném rozkladu na prvočinitele (což je fakt platný obecně pro libovolný okruh hlavních ideálů).

Příklad. Vraťme se k našemu předchozímu příkladu: označme I, resp. J ideály generované 3 a $1 + \sqrt{10}$, resp. 3 a $1 - \sqrt{10}$, tj.

$$I = (3, 1 + \sqrt{10}) = \{3\alpha + (1 + \sqrt{10})\beta;\ \alpha, \beta \in R\},$$
$$J = (3, 1 - \sqrt{10}) = \{3\alpha + (1 - \sqrt{10})\beta;\ \alpha, \beta \in R\}.$$

Pak I, J jsou prvoideály (platí $R/I \simeq R/J \simeq \mathbb{F}_3$) a platí $IJ = (3)$ (snadno je vidět, že $IJ \subseteq (3)$, opačná inkluze plyne z $3 = 3 \cdot 3 - 3(1 - \sqrt{10}) - (1 + \sqrt{10})3$). Dále platí

$$I^2 = (9, 3(1 + \sqrt{10}), (1 + \sqrt{10})^2) = (9, 3 + 3\sqrt{10}, 11 + 2\sqrt{10}),$$

proto $1+\sqrt{10} = 9+(3+3\sqrt{10})-(11+2\sqrt{10}) \in I^2$. Na druhou stranu jistě 9, $3(1+\sqrt{10})$ a $(1+\sqrt{10})^2$ jsou prvky ideálu $(1+\sqrt{10})$, tedy $I^2 = (1+\sqrt{10})$. Analogicky $J^2 = (1-\sqrt{10})$. Obě strany identity

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

udávají tedy týž rozklad ideálu (9) na prvoideály: $(9) = I^2J^2$.

Poznámka. Míru toho, nakolik se okruh celých algebraických čísel R nějakého tělesa algebraických čísel K liší od okruhu s jednoznačným rozkladem prvků, nám vlastně udává to, kolik ze všech ideálů okruhu R je hlavních. Uvažme pologrupu všech nenulových ideálů okruhu R a jeho podpologrupu všech hlavních ideálů. Můžeme uvážit faktorizaci této pologrupy podle zmíněné podpologrupy (což odpovídá následující ekvivalenci mezi ideály: $I \sim J$, právě když existují $\alpha, \beta \in R$ splňující $(\alpha)\cdot I = (\beta)\cdot J$). Je možné dokázat, že faktorstrukturou je konečná komutativní grupa. Počet jejích prvků se nazývá počet tříd ideálů okruhu R (nebo také tělesa K) a je jednou z nejdůležitějších charakteristik aritmetiky v okruhu R.

Poznámka. V našem příkladě jsme odvodili, že grupa jednotek okruhu celých algebraických čísel tělesa $\mathbb{Q}(\sqrt{10})$ je

$$E = \{\pm(3+\sqrt{10})^n;\ n \in \mathbb{Z}\},$$

tedy komutativní grupa se dvěma generátory -1 a $3+\sqrt{10}$. Tento fakt platí obecně: grupa jednotek okruhu celých algebraických čísel libovolného tělesa algebraických čísel je konečně generovaná komutativní grupa a pro minimální počet jejích generátorů platí, že nepřevýší stupeň $[K:\mathbb{Q}]$.

21. Metoda síta v číselném tělese

Vraťme se k našemu původnímu problému: máme velké přirozené číslo N, o kterém víme, že je složené, a hledáme jeho netriviálního dělitele.

Zvolme celé algebraické číslo θ a označme $T(x) \in \mathbb{Z}[x]$ normovaný mnohočlen nejmenšího stupně, jehož je θ kořenem. Nechť $K = \mathbb{Q}(\theta)$ je nejmenší těleso algebraických čísel, které obsahuje číslo θ. Označme

$$L = \{a_0 + a_1\theta + \cdots + a_{d-1}\theta^{d-1};\ a_0, \ldots, a_{d-1} \in \mathbb{Q}\},$$

kde d je stupeň polynomu $T(x)$.

Dokážeme, že platí $K = L$. Protože $T(\theta) = 0$, je jasné, že L je okruh. Dokažme, že je L dokonce těleso, pak bude zřejmé, že je L nejmenší těleso obsahující θ a tedy, že $K = L$. Nejprve ukážeme, že $T(x)$ je ireducibilní nad $\mathbb{Z}$. Kdyby existovaly nekonstantní polynomy $f(x)$, $g(x) \in \mathbb{Z}[x]$ takové, že $T(x) = f(x)g(x)$, musely by být oba normované a stupně menšího než d. Dosazením θ za x bychom pak dostali spor s definicí $T(x)$. Podle věty 8.7 na straně 102 v [R] (popřípadě užitím techniky užité v příkladu v předchozí kapitole) dostáváme, že $T(x)$ je ireducibilní nad $\mathbb{Q}$. Nechť $\alpha \in L$, $\alpha \neq 0$ je libovolné. Pak existuje polynom $f(x) \in \mathbb{Q}[x]$ stupně menšího než d tak, že $\alpha = f(\theta)$. Protože $T(x)$ je ireducibilní nad $\mathbb{Q}$ a $f(x)$ má stupeň menší než $T(x)$, je jejich největší společný dělitel (který existuje podle věty 5.18 na straně 85 v [R]) roven 1. Podle věty 5.20 na straně 86 v [R] existují polynomy $u(x)$, $v(x) \in \mathbb{Q}[x]$ tak, že $f(x)u(x)+T(x)v(x) = 1$. Dosazením θ za x dostaneme $u(\theta) = \frac{1}{\alpha} \in L$, a tedy L je těleso. Stejným postupem můžeme dokázat, že neexistuje nenulový polynom s racionálními koeficienty stupně menšího než d, jehož by byl θ kořenem, a proto je $[K : \mathbb{Q}] = d$.

Označme R okruh všech celých algebraických čísel v K. Protože $\theta \in R$, pro

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

platí $\mathbb{Z}[\theta] \subseteq R$. Obecně zde rovnost platit nemusí. Je však možné dokázat, že faktorgrupa $R/\mathbb{Z}[\theta]$ je konečná. Označme f počet jejích prvků.

Budeme předpokládat, že jsme schopni spočítat všechno potřebné o tělese K, tj. počet tříd ideálů okruhu R, generátory grupy jednotek okruhu R, výše uvedenou konstantu f, všechny „malé“ prvoideály až do zvolené hranice („velikost“ prvoideálu P je dána počtem prvků oboru integrity R/P, o kterém je možné dokázat, že je to konečné těleso) a v případě, že R není okruh hlavních ideálů, i reprezentanty jednotlivých tříd ideálů (v tomto případě je situace o něco složitější, metodu síta v číselném tělese proto popíšeme jen pro případ, kdy je R okruh hlavních ideálů; podrobný popis metody v obecné situaci lze najít v [C]). Dodejme, že pro obecné těleso algebraických čísel K jde o velmi obtížný úkol, jehož náročnost silně stoupá se stupněm tělesa a který nebyl dosud zvládnut ani pro některá poměrně jednoduchá tělesa. Vzhledem k tomu, že však číslo θ volíme, bude možné předpokladům tohoto odstavce vyhovět.

Protože polynom $T(x)$ je ireducibilní nad $\mathbb{Q}$, kořeny polynomu $T(x)$ v $\mathbb{C}$ jsou po dvou různé (vícenásobný kořen by byl kořenem derivace $T'(x)$ a tedy největší společný dělitel $(T(x), T'(x))$ by byl vlastním dělitelem polynomu $T(x)$). Nechť $\theta_1 = \theta$, θ_2, $\ldots$, θ_d jsou všechny komplexní kořeny polynomu $T(x)$. Z Viétovych vztahů a z hlavní věty o symetrických polynomech plyne, že hodnota libovolného symetrického polynomu o d proměnných v θ_1, θ_2, $\ldots$, θ_d bude celé číslo. Proto můžeme definovat normu $\mathcal{N} : \mathbb{Z}[\theta] \to \mathbb{Z}$ následujícím způsobem. Pro libovolný prvek $\alpha \in \mathbb{Z}[\theta]$ existuje polynom $g(x) \in \mathbb{Z}[x]$ tak, že $\alpha = g(\theta)$. Pak klademe

$$\mathcal{N}(\alpha) = g(\theta_1) \ldots g(\theta_d) \in \mathbb{Z}.$$

Snadno se nahlédne, že tato definice nezávisí na volbě polynomu g a že platí $\mathcal{N}(\alpha\beta) = \mathcal{N}(\alpha)\mathcal{N}(\beta)$. Poznamenejme, že je možné definici normy rozšířit na celý okruh R: pro libovolné $\beta \in R$ je $f\beta \in \mathbb{Z}[\theta]$, položíme pak $\mathcal{N}(\beta) = f^{-d}\mathcal{N}(f\beta)$. Je možné dokázat, že pro každé $\beta \in R$ je $\mathcal{N}(\beta)$ celé číslo a že $|\mathcal{N}(\beta)|$ je počet prvků faktorokruhu $R/(\beta)$, kde (β) značí hlavní ideál generovaný prvkem β. (Odtud plyne, že β je jednotka okruhu R, právě když $\mathcal{N}(\beta) = 1$.)

Představme si, že známe nějaké celé číslo m takové, že $T(m) = \pm kN$ pro nějaké „malé“ přirozené číslo k. Pak můžeme sestrojit homomorfismus okruhů $\varphi : \mathbb{Z}[\theta] \to \mathbb{Z}/N\mathbb{Z}$ tak, že položíme $\varphi(\theta) = m + N\mathbb{Z}$. (Je jasné, že jde o homomorfismus aditivních grup. To, že φ zachovává i násobení, plyne z toho, že $N \mid T(m)$.)

Pro každé $\alpha \in R$ platí $f\alpha \in \mathbb{Z}[\theta]$. Vzhledem k tomu, že f pravděpodobně nebude dělitelné číslem N, můžeme předpokládat $(f, N) = 1$ (jinak máme netriviálního dělitele N). Nechť $u \in \mathbb{Z}$ splňuje $uf \equiv 1 \pmod{N}$. Pak můžeme φ dodefinovat na celém R

takto: pro libovolné $\alpha \in R$ klademe $\varphi(\alpha) = u \cdot \varphi(f\alpha)$. Je jasné, že pro $\alpha \in \mathbb{Z}[\theta]$ jsme tím $\varphi(\alpha)$ nezměnili a že φ je skutečně homomorfismus okruhů $\varphi : R \to \mathbb{Z}/N\mathbb{Z}$.

Nyní k čemu chceme dojít: rádi bychom našli a, $b \in \mathbb{Z}$ tak, aby $a + bm$ byla druhá mocnina v $\mathbb{Z}$ a současně aby $a + b\theta$ byla druhá mocnina v R. Je-li totiž $a + bm = x^2$ pro $x \in \mathbb{Z}$ a současně $a + b\theta = \alpha^2$ pro $\alpha \in R$, máme šanci, že se nám podaří rozložit N: protože je φ homomorfismus, platí

$$\varphi(\alpha)^2 = \varphi(\alpha^2) = \varphi(a + b\theta) = (a + bm) + N\mathbb{Z} = x^2 + N\mathbb{Z}$$

a je-li y nějaký reprezentant třídy $\varphi(\alpha)$, tj. $\varphi(\alpha) = y + N\mathbb{Z}$, platí $x^2 \equiv y^2 \pmod N$ a $(x + y, N)$ by mohl být netriviální dělitel čísla N podobně jako u metody řetězových zlomků či metody kvadratického síta.

Nalezení takových a, b je však velmi nesnadný úkol a jistě se nám je nepodaří najít rovnou. Použijeme proto podobné techniky jako v předchozích metodách a budeme rozkládat nad nějakými bazemi faktorizace. Jasná je situace pro čísla typu $a + bm \in \mathbb{Z}$, která budeme rozkládat nad „malými“ prvočísly. Větší problém je s čísly typu $a + b\theta \in \mathbb{Z}[\theta]$. Jak jsme se už zmínili, budeme pro jednoduchost předpokládat, že R je okruh hlavních ideálů. Pak je možné pro každý „malý“ prvoideál P nalézt $\pi \in R$ tak, že $P = (\pi)$. Budeme tedy ideály $(a + b\theta)$ rozkládat nad bazí faktorizace složené z těchto „malých“ prvoideálů a vždy, když existuje nějaký rozklad

$$(a + b\theta) = (\pi_1)^{k_1} \dots (\pi_n)^{k_n},$$

musí existovat jednotka η okruhu R, splňující

$$a + b\theta = \eta \pi_1^{k_1} \dots \pi_n^{k_n}.$$

Protože grupa jednotek okruhu R je konečně generovaná grupa (viz poznámku na konci předchozí kapitoly), můžeme zvolit její generátory a získanou jednotku η pomocí nich vyjádřit. Celkem tedy čísla $a + b\theta$ rozkládáme nad bází faktorizace, složené z generátorů jednotlivých „malých“ prvoideálů a z generátorů grupy jednotek okruhu R.

Je možné dokázat, že pro libovolné prvočíslo $p \nmid f$ jsou ideály tvaru $\wp = (p, \theta - c_p)$, kde c_p probíhá všechna (navzájem nekongruentní modulo p) řešení kongruence $T(x) \equiv 0 \pmod p$, navzájem různé (ne však nutně všechny) prvoideály okruhu R dělící hlavní ideál (p). Přitom $a + b\theta \in \wp$ pro nějaká a, $b \in \mathbb{Z}$ nastane právě tehdy, když $a + bc_p \in \wp$, tj. právě když $p \mid a + bc_p$ (celá čísla jsou prvky $\wp$, právě když jsou dělitelná p, to plyne z toho, že $\wp \cap \mathbb{Z}$ musí být prvoideál v $\mathbb{Z}$ obsahující p). Označme v nějaké celé číslo splňující $vf \equiv 1 \pmod p$. Nechť $\beta \in R$ je libovolné. Pak $f \cdot \beta \in \mathbb{Z}[\theta]$, tedy existuje polynom $g(x) \in \mathbb{Z}[x]$ tak, že $f \cdot \beta = g(\theta)$. Protože $x - c_p \mid g(x) - g(c_p)$ v $\mathbb{Z}[x]$, platí $\theta - c_p \mid g(\theta) - g(c_p)$ v R, proto $vf\beta - vg(c_p) = vg(\theta) - vg(c_p) = v(g(\theta) - g(c_p)) \in \wp$. Protože $p \mid 1 - vf$ a $p \in \wp$, je $(1 - vf)\beta \in \wp$, a tedy $\beta - vg(c_p) \in \wp$. Je-li r zbytek po dělení celého čísla $vg(c_p)$ prvočíslem p, z $p \mid (vg(c_p) - r)$ plyne $\beta - r \in \wp$. V každé třídě rozkladu $R/\wp$ tedy leží (jediné) z čísel $0, 1, \dots, p - 1$, je tedy $|R/\wp| = p$. Proto platí: jestliže $\pi \in R$ splňuje $(\pi) = \wp$, pak $\mathcal{N}(\pi) = p$.

Podobně jako u metody kvadratického síta budeme dvojice celých čísel (a, b), pro které máme šanci rozložit $a + bm$ i $a + b\theta$ nad zvolenými bázemi faktorizace, hledat prosíváním. Uvažujme tedy všechny dvojice celých čísel (a, b), pro která $|a|$ i $|b|$ jsou menší než zvolená mez.

1. Pro všechna malá prvočísla p vyznačme ty dvojice (a, b), pro které p dělí a i b.
2. (První inicializace) Pro všechny nevyznačené dvojice (a, b) inicializujme položku, obsahující přibližnou hodnotu $\log_2(a + bm)$.
3. (První prosívání) Pro každou mocninu p^k prvočísla, která je menší než zvolená mez, odečteme $\log_2 p$ od položek, příslušných těm nevyznačeným dvojicím (a, b), pro které $p \mid a + bm$.
4. Vyznačíme všechny dvojice (a, b), jejichž položka zůstala příliš velká.
5. (Druhá inicializace) Pro všechny nevyznačené dvojice (a, b) inicializujme položku, obsahující přibližnou hodnotu $\log_2(\mathcal{N}(a + b\theta))$.
6. (Druhé prosívání) Pro každé prvočíslo p, které je menší než zvolená mez a které nedělí f, nalezněme všechna (navzájem nekongruentní modulo p) řešení c_p kongruence $T(x) \equiv 0 \pmod{p}$. Pro každé takové řešení c_p odečteme $\log_2 p$ od všech položek, příslušných těm dosud nevyznačeným dvojicím (a, b), pro které $p \mid a + bc_p$.
7. Pro všechny dvojice (a, b), jejichž položka zůstala menší než jistá mez, zjistíme rozkladem čísla $\mathcal{N}(a + b\theta)$, zda opravdu všechny prvoideály dělící $a + b\theta$ jsou ve zvolené bázi faktorizace.

Poznamenejme, že v bodě 6 odčítáme $\log_2 p$ jen jednou bez ohledu na to, zda prvoideál $(p, \theta - c_p)$ vystupuje v rozkladu ideálu $(a + b\theta)$ v první nebo ve vyšší mocnině (to totiž nejsme schopni rozlišit). Aby se vyloučil případ, že by se tím na některou užitečnou dvojici zapomnělo, je „jistá mez“ užitá v bodě 7 o něco vyšší, než by bylo jinak zapotřebí. To však má zase za následek, že je nutná zde zmíněná kontrola rozkladem.

Označme $p_1, \dots, p_l$ bázi faktorizace použitou pro rozkládání čísel $a + bm$, dále pak $u_1, \dots, u_k$ generátory grupy jednotek okruhu R a konečně $\pi_1, \dots, \pi_n$ čísla, generující hlavní prvoideály použité pro rozkládání ideálů $(a + b\theta)$. Po skončení uvedených sedmi bodů máme pro každou nevyznačenou dvojici (a, b), která prošla úspěšně kontrolou v bodě 7, rozklady

$$a + bm = (-1)^{e_0} p_1^{e_1} \dots p_l^{e_l}, \qquad a + b\theta = u_1^{e_{l+1}} \dots u_k^{e_{l+k}} \pi_1^{e_{l+k+1}} \dots \pi_n^{e_{l+k+n}}.$$

Tím dostaneme vektor $(e_0, e_1, \dots, e_{l+k+n})$ přirozených čísel, který budeme interpretovat jako vektor z vektorového prostoru $\mathbb{F}_2^{1+k+l+n}$ (sudá čísla jako nuly a lichá čísla jako jedničky). Až budeme mít těchto vektorů dost (tj. alespoň o několik více než $1+k+l+n$), provedeme Gaussovu eliminaci nad $\mathbb{F}_2$ (viz poznámku o „řídkých“ maticích na konci patnácté kapitoly), abychom našli $\mathbb{F}_2$-lineární závislosti mezi získanými vektory. Každá taková závislost nám určí, která čísla $a + bm$, resp. $a + b\theta$ máme mezi sebou vynásobit, abychom dostali kýženou kongruenci $x^2 \equiv y^2 \pmod{N}$. Přitom místo $a + b\theta$ budeme rovnou násobit $\varphi(a + b\theta)$.

Poznámka. 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. Bohužel výpočty v této metodě jsou notně komplikované (jak jsme již konec konců sami viděli) a tedy asymptotická výhodnost se začne projevovat až pro značně velká čísla. Odhaduje se, že tato metoda by měla být rychlejší až pro čísla mající aspoň 130 dekadických cifer, jenže tak velká čísla jsou stejně mimo dnešní možnosti. Na druhé straně, pro čísla ve speciálním tvaru, například pro Mersenova čísla $2^p - 1$, kde p je prvočíslo, nebo Fermatova čísla $2^{2^k} + 1$ může být tato metoda zjednodušena (je možné snížit c až pod $\sqrt[3]{\frac{32}{9}}$) a stává se praktickou už i pro N do 120 cifer.

Příklad použití metody. Předpokládejme, že naše velké přirozené číslo je tvaru $N = r^e - s$, kde r a s jsou celá čísla s malou absolutní hodnotou. Zvolme vhodné d (ukazuje se, že pro N o 70 a více cifrách je vhodné $d = 5$) a položme $k = -[-\frac{e}{d}]$. Je tedy $-k \leq -\frac{e}{d} < -k + 1$, tj. $0 \leq kd - e < d$. Protože $|r|$ a $|s|$ jsou malé, je malé i $|sr^{kd-e}|$. Uvažme mnohočlen

$$T(x) = x^d - sr^{kd-e}.$$

Zvolíme-li $m = r^k$, je jasné, že $T(m) = r^{kd-e}N$ je malý násobek N. Předpokládejme, že polynom $T(x)$ je ireducibilní nad $\mathbb{Z}$ (to je velmi přirozený předpoklad, rozkladem polynomu $T(x)$ v $\mathbb{Z}[x]$ bychom patrně dostali netriviálního dělitele čísla N). Budeme pracovat v tělese $K = \mathbb{Q}(\theta)$, kde θ je kořen polynomu $T(x)$. Protože $d = 5$ a $|sr^{kd-e}|$ je malé, asi nebude těžké najít vše potřebné o aritmetice okruhu celých čísel tělesa K.

První případ, kdy metoda síta v číselném tělese slavila úspěch, bylo úplné rozložení devátého Fermatova čísla $N = 2^{512}+1$, které má 155 cifer, na prvočinitele. Již dříve byl znám sedmiciferný dělitel čísla N, avšak rozložit podíl, o kterém se vědělo, že je složený, se nedařilo. Až v roce 1990 byl tento podíl rozložen metodou síta v číselném tělese (viz poznámku na konci devatenácté kapitoly). Je zajímavé, že znalost sedmiciferného dělitele vůbec nepomohla, bylo výhodnější rozkládat celé N a využít tak jeho speciálního tvaru: $N = r^e - s$ pro $r = 2$, $s = -1$ a $e = 512$. Užijeme-li postup z předchozího odstavce, je $d = 5$, $k = 103$, $T(x) = x^5 + 8$, tj. $K = \mathbb{Q}(\sqrt[5]{8}) = \mathbb{Q}(\sqrt[5]{2})$, což je těleso, jehož okruh celých algebraických čísel je dokonce okruh hlavních ideálů. Za pomoci mnoha dobrovolníků prostřednictvím e-mailu byl vynaložen ekvivalent několika let CPU času na jednom počítači. Získaná data byla zpracována Gaussovou eliminací na superpočítači a tak byl získán rozklad N na tři prvočísla o 7, 49 a 99 cifrách.