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

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

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

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

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

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

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

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

Shrnutí obou předchozích příkladů

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

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

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

Zobecnění předchozích příkladů

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

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

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

Dirichletova věta o jednotkách

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

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

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

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

Důkaz je mimo možnosti této přednášky.

Zpět k našim příkladům

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

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

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

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

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

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

Metoda síta v číselném tělese

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

Metoda síta v číselném tělese

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

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

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

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

Shrnutí

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

Protože $[f(m)]_N = [0]_N$, předpis

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

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

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

Základní myšlenka metody

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

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

Prosívání

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

Báze faktorizace v okruhu R

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

Pro zjednodušení úvah předpokládejme, že **grupa tříd ideálů okruhu R je triviální**. Pak každý prvoideál $\mathcal{P}_{p,c}$ je hlavní, tedy $\mathcal{P}_{p,c} = \wp_{p,c} R$ pro nějaké číslo $\wp_{p,c} \in R$, a $\mathcal{N}(\wp_{p,c}) = p$. Jestliže

$$(a + b\theta)R = \mathcal{P}_{p_1,c_1} \cdot \mathcal{P}_{p_2,c_2} \cdots \mathcal{P}_{p_s,c_s},$$

pak existuje jednotka $\varepsilon \in R^\times$ tak, že

$$a + b\theta = \varepsilon \cdot \wp_{p_1,c_1} \cdot \wp_{p_2,c_2} \cdots \wp_{p_s,c_s}.$$

Do báze faktorizace tedy dáme generátory grupy jednotek $R^\times$, o které víme, že má nejvýše n generátorů. Pro každé „malé“ prvočíslo $p \nmid r$ a každé $c \in \{0, 1, \ldots, p-1\}$ splňující $p \mid f(c)$ v $\mathbb{Z}$ pak ještě přidáme do báze faktorizace čísla $\wp_{p,c}$.

Prosívání dvojic $a, b \in \mathbb{Z}$, $|a|, |b|$ „malá“, $b \geq 0$

1. Pro každé prvočíslo p z 1. báze odstraníme dvojice (a, b) splňující $p|a$, $p|b$.
2. (První inicializace) Ke každé zbylé dvojici (a, b) uložme přibližnou hodnotu $\log_2 |a + bm|$.
3. (První prosívání) Pro každou mocninu p^k prvočísla p z 1. báze menší než jistá mez odečteme $\log_2 p$ od hodnot uložených těm zbylým dvojicím (a, b), pro které $p^k \mid a + bm$.
4. Odstraníme všechny dvojice (a, b) s příliš velkou uloženou hodnotou.
5. (Druhá inicializace) Ke každé zbylé dvojici (a, b) uložme přibližnou hodnotu $\log_2 |\mathcal{N}(a + b\theta)|$.
6. (Druhé prosívání) Pro každé $\wp_{p,c}$ z 2. báze faktorizace odečteme $\log_2 p$ od hodnot uložených těm zbylým dvojicím (a, b), pro které $p \mid a + bc$.
7. Pro všechny dvojice (a, b), jejichž uložená hodnota zůstala menší než jistá mez, zjistíme, jestli se $a + b\theta$ rozkládá v 2. bázi faktorizace.

Další postup ve speciálním případě

Pro každou dvojici, pro kterou jsme v 7. kroku ověřili, že se $a + b\theta$ rozkládá v 2. bázi faktorizace, rozložíme v 1. bázi faktorizace $a + bm$. Tím získáme pro tuto dvojici (a, b) z exponentů obou rozkladů vektor nad dvouprvkovým tělesem $\mathbb{F}_2$.

Až máme těchto vektorů o několik více než kolik je celkem prvků v obou bázích faktorizace, provedeme Gausovu eliminaci (nejprve řídké a pak husté matice), abychom našli jejich lineární závislosti. Každá lineární závislost nám dá jednu kongruenci

$$y^2 \equiv z^2 \pmod{N}.$$

Budeme-li mít těchto kongruencí několik, je reálná šance, že pro některou z nich platí $y \not\equiv \pm z \pmod{N}$, a tedy $(N, y + z)$ je netriviální dělitel čísla N.

Obecný případ

V obecném případě, kdy **grupa tříd ideálů okruhu R není triviální**, je celá situace komplikovanější. Má-li tato grupa sudý řád, může se totiž stát, že přestože je ideál, který získáme z nalezené lineární závislosti vynásobením vhodných ideálů $(a + b\theta)R$ druhou mocninou nějakého ideálu I, nemusí být tento ideál I hlavní. Omezme se zde jen na konstatování, že se tento problém dá řešit například tím, že pro každou takto nalezenou lineární závislost uložíme informaci o tom, ve které třídě grupy tříd ideálů leží ideál I. Pak znovu provedením Gaussovy eliminace nalezneme lineární závislost mezi těmito vektory a ta nám dá lineární závislost ideálů $(a + b\theta)R$, pro kterou je odpovídající ideál I hlavní. Ovšem další ještě větší komplikace spočívá v tom, jak najít generátor tohoto hlavního ideálu (jde o druhou odmocninu z celého algebraického čísla, které je součinem tisíců činitelů tvaru $(a + b\theta)$, a tedy lze čekat, že vyjádříme-li toto číslo jako hodnotu v θ polynomu stupně menšího než n, koeficienty tohoto polynomu mohou mít několik stovek tisíc dekadických cifer). Vysvětlit triky, pomocí kterých se tato komplikace překonává, už v této přednášce nestihneme...

Odhad časové náročnosti

Metoda síta v číselném tělese je nejnovější a potenciálně nejrychlejší známá metoda rozkládání velkých přirozených čísel. Na základě některých heuristických argumentů lze odhadovat, že metoda řetězových zlomků i metoda kvadratického síta jsou časové náročnosti

$$O\left(e^{\sqrt{\ln N \ln\ln N}(1+o(1))}\right).$$

Proto před objevením metody síta v číselném tělese panovalo přesvědčení, že lepší časové náročnosti už patrně nepůjde dosáhnout. Bylo překvapením, že na základě podobných argumentů lze odhadovat, že metoda síta v číselném tělese je časové náročnosti

$$O\left(e^{\sqrt[3]{(\ln N)(\ln\ln N)^2}(c+o(1))}\right)$$

pro poměrně malé c (menší než $\sqrt[3]{\frac{64}{9}}$), což je asymptoticky mnohem lepší než jakákoli jiná známá metoda.