Shorův faktorizační algoritmus

Shorův faktorizační algoritmus je nejvýznamnější aplikací kvantové Fourierovy transformace a jeden z hlavních důvodů zájmu o kvantové počítače, které by umožnily pravděpodobnostní polynomiální faktorizaci velkých čísel.

Z číselně teoretického hlediska se přitom nejedná o žádnou novinku: základem Shorova algoritmu je Fermatův faktorizační algoritmus, ve kterém se ze znalosti dvou čísel a, b, splňujících $a^2 \equiv b^2 \mod N$ získá faktorizace N díky vztahu

$$(a+b)(a-b) \equiv 0 \mod N.$$

Fermatův postup lze použít mimo jiné tehdy, pokud známe nějaký prvek a a jeho sudý řád r v multiplikativní grupě $\mathbb{Z}_N$. Pak platí

$$(a^{\frac{r}{2}}+1)(a^{\frac{r}{2}}-1) \equiv 0 \mod N,$$

což poskytuje faktorizaci N právě když $a^{\frac{r}{2}}$ není rovno $-1 \mod N$. Shorův fraktorizační algoritmus pro složené liché N tedy vypadá následovně:

- zvol náhodně $a \in \mathbb{Z}_N^*$ (volba neinveribilního prvku vede k faktorizaci okamžitě)
- najdi řád r prvku a v $\mathbb{Z}_N^*$
- je-li r liché nebo je-li $a^{\frac{r}{2}} \equiv -1 \mod N$, skonči selháním
- jinak vrať faktor $\mathrm{NSD}(N, a^{\frac{r}{2}}-1)$

Z teorie čísel víme, že počet prvků a, která nevedou k selhání, je dostatečný (nejméně jedna polovina). Nepraktičnost tohoto algoritmu ale plyne z toho, že je obtížné zjistit řád prvku v grupě $\mathbb{Z}_N^*$. Kvantovou podstatou Shorova algoritmu je tedy hledání řádu prvku, k čemuž je vhodná Fourierova transformace, a ta je na kvantovém počítači polynomiální.

Hledání řádu prvku. Umocňování prvku a modulo N, tedy $k \mapsto a^k \mod N$, je zobrazení $f : \mathbb{N} \to \mathbb{Z}_N^*$ s periodou r. To dává základní představu, proč může být Fourierova transformace pro hledání řádu užitečná.

Kvantová realizace umocňování se musí odehrávat na konečných binárních registrech. Nechť tedy $n = \lceil \log N \rceil$ je počet bitů v zápisu čísla N, zvolme nějaké $M = 2^m$ dostatečně velké (velikost m bude mít vliv na pravděpodobnost úspěchu algoritmu).

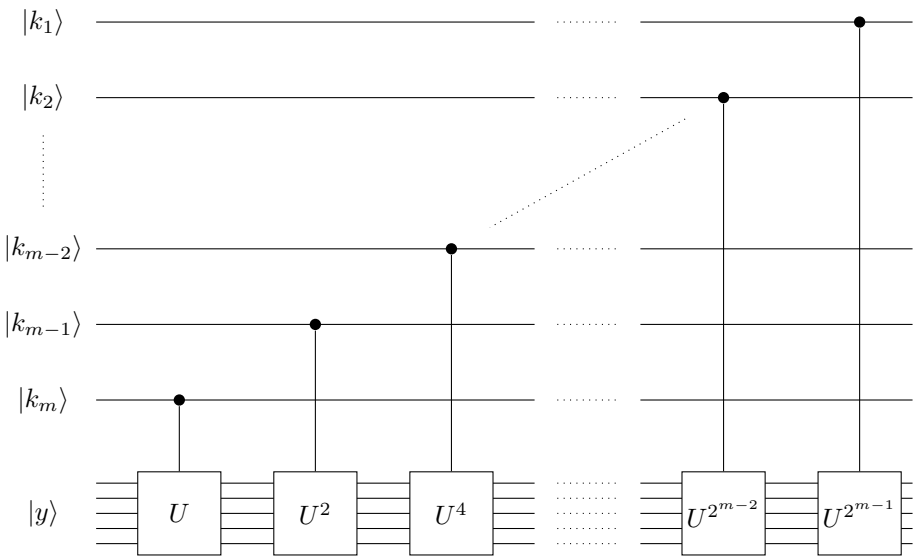
Umocňování nyní aproximuje operátor

$$\begin{aligned} W : \mathbb{H}_2^m \otimes \mathbb{H}_2^n &\to \mathbb{H}_2^m \otimes \mathbb{H}_2^n \\ |k\rangle|y\rangle &\mapsto |k\rangle|ya^k \mod N\rangle \end{aligned}$$

přičemž pro $N \leq y \leq 2^n - 1$, tedy pro prvky, pro které by se zbytek opakoval, definujeme$W|k\rangle|y\rangle := |k\rangle|y\rangle$. Protože a je nesoudělné s N, permutuje W bázové prvky, a je tedy unitární. Realizace operátoru W je možná pomocí modulárního umocňování. Je-li U nějaký operátor, pro který máme k dispozici kontrolované mocniny U^{2^j}, vypadá obvod umocňující U, tedy realizující zobrazení

$$|k\rangle|y\rangle \mapsto |k\rangle U^k|y\rangle,$$

takto:

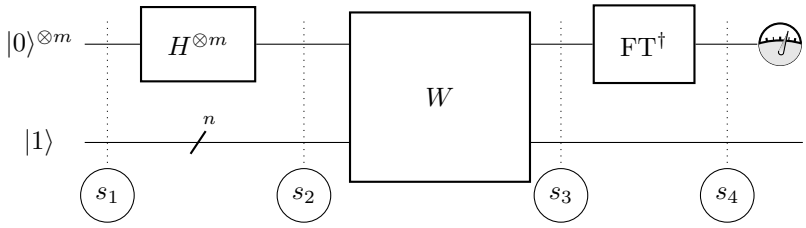


V případě operátoru W odpovídá U násobení prvkem a v grupě $\mathbb{Z}_N$, tedy transformaci

$$\begin{aligned} U : \mathbb{H}_2^n &\to \mathbb{H}_2^n \\ |y\rangle &\mapsto |ay \mod N\rangle, \end{aligned}$$

kde opět $U|y\rangle := |y\rangle$ pro $y \geq N$.

Základní myšlenka algoritmu odhalujícího řád je standardní: vyhodnotit W na všech hodnotách $|k\rangle$ současně. Protože funkce umocňování je periodická, aplikujeme na ní Fourierovu transformaci a měli bychom získat informaci o periodě. Celý algoritmus vypadá takto:



Uvědomme si, že stav $|1\rangle$ (neboli $|y\rangle$ pro $y = 1$), je bázový prvek n-kubitového registru s číslem 1, tedy $|0\rangle^{(n-1)}|1\rangle = |0\cdots01\rangle$. První tři fáze dávají

$$s_1 : \ |0\rangle^{\otimes m}|0\cdots01\rangle \qquad s_2 : \ \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} |k\rangle|0\cdots01\rangle \qquad s_3 : \ \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} |k\rangle|a^k\rangle,$$

čímž je připravena kýžená rovnoměrná superpozice hodnot funkce $k \mapsto a^k$. Aplikací Fourierovy transformace na první registr dostaneme

$$s_4 : \quad \frac{1}{M} \sum_{k}^{M-1} \sum_{z}^{M-1} \exp\left[2\pi i \frac{kz}{M}\right] |z\rangle|a^k\rangle$$

Nyní změříme první registr. Pravděpodobnost, že výsledek měření bude odpovídat nějakému zvolenému $|z\rangle$, odvodíme podle postulátu měření jako čtverec velikosti projekce na podprostor výsledku, tedy na vektor složek obsahujících $|z\rangle$. Je to tedy součet druhých mocnin velikosti amplitud pravděpodobnosti pro všechny členy, ve kterých se $|z\rangle$ vyskytuje. Těchto členů je právě r, totiž $|z\rangle|a^0\rangle$, $|z\rangle|a^1\rangle$,..., $|z\rangle|a^{r-1}\rangle$, přičemž koeficient u $|z\rangle|a^t\rangle$ je součtem koeficientů u všech $|z\rangle|a^k\rangle$, kde k je tvaru $sr+t \mod N$. Všechny členy obsahující nějaké pevné $|z\rangle$ tedy jsou

$$\frac{1}{M}\sum_{t=0}^{r-1}\left(\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{(sr+t)z}{M}\right]\right)|z\rangle|a^t\rangle$$

a příslušná pravděpodobnost je rovna

$$\begin{aligned}P(z) &= \frac{1}{M^2}\sum_{t=0}^{r-1}\left|\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{(sr+t)z}{M}\right]\right|^2 = \\ &= \frac{1}{M^2}\sum_{t=0}^{r-1}\left|\exp\left[2\pi i\frac{tz}{M}\right]\right|^2\left|\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{srz}{M}\right]\right|^2 = \\ &= \frac{1}{M^2}\sum_{t=0}^{r-1}\left|\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{rz}{M}s\right]\right|^2.\end{aligned}$$

Číslo ℓ_t je největší takové, že $\ell_t r + t$ je menší než M, tedy

$$\ell_t = \left\lfloor\frac{M-1-t}{r}\right\rfloor.$$

Hodnoty ℓ_t se mohou pro různá t lišit o jedna. Komplikace plyne z toho, že r obecně nedělí M; kdyby ho dělilo, bylo by ℓ jednoduše rovno $M/r-1$. Tato nepravidelnost má hlubší důležitost. Uvědomme si, že provádíme Fourierovu transformaci na grupě $\mathbb{Z}_M$, nikoli $\mathbb{Z}_N$! Výsledek bude mít určitou nepřesnost, protože funkce $k \mapsto a^k \mod N$ není na $\mathbb{Z}_M$ zcela periodická: v okolí nuly je periodicita porušena (pokud r nedělí M). Pro velká M ale bude tato nepřesnost zanedbatelná.

Tyto obecné úvahy se konkretizují ve výpočtu hodnoty $P(z)$. Ukážeme, že platí

$$(*)\quad \left|\sum_{s=0}^{\ell_t}\exp\left[2\pi i\frac{rz}{M}s\right]\right| \approx \begin{cases}\frac{M}{r} & \text{pokud } rz \approx pM \text{ pro nějaké přirozené } p,\\ 0 & \text{jinak.}\end{cases}$$

Ve výše zmíněném ideálním případě, kdy r dělí M, probíhá uvažovaná suma hodnoty charakteru grupy $\mathbb{Z}_M$, a vztah $(*)$ tedy platí s rovnostmi na místě $\approx$. S jistotou tedy naměříme z, které je tvaru $p\cdot\frac{M}{r}$, kde $p \in \{0,1,2,\ldots,r-1\}$. Pro každé takové z je pravděpodobnost $P(z)$ rovna $\frac{1}{r}$, jak se snadno dopočte. Ze z získáme zlomek

$$\frac{z}{M} = \frac{p}{r},$$

jehož jmenovatel je r, pokud je p s r nesoudělné. To pro $r > 19$ nastává s pravděpodobností alespoň $\frac{1}{4}\frac{1}{\log\log r}$. Pokud má p s r nějaký společný faktor, dostáváme alespoň nějakou část r. Opakováním postupu několikrát se s velkou pravděpodobností dopracujeme k r.

V obecném případě, tedy pokud r nedělí M, platí, že naměřené z je s velkou pravděpodobností nějakému násobku $\frac{M}{r}$ blízko, tedy že

$$\frac{z}{M} \approx \frac{p}{r}.$$

Vyvstává zajímavá otázka, jak najít všechny zlomky s omezeným čitatelem, které jsou blízko dané hodnotě α. Odpovědí je rozvoj do řetězového zlomku. Platí, že pokud je vzdálenost mezi α a zlomkem $\frac{p}{r}$ menší než $\frac{1}{2r^2}$, pak je tento zlomek přítomen v řetězovém rozvoji čísla α (viz přednášku o řetězových zlomcích v rámci *Teorie čísel a RSA*, http://www.karlin.mff.cuni.cz/ holub/soubory/Retez.pdf, zejména aplikaci na Shorův algoritmus na str. 8). Pokud budeme předpokládat, že z je zaokrouhlená hodnota $p\frac{M}{r}$, tedy že

$$\left| z - p\frac{M}{r} \right| \leq \frac{1}{2},$$

pak

$$\left| \frac{z}{M} - \frac{p}{r} \right| \leq \frac{1}{2M},$$

což vede k volbě M přibližně N^2 zajišťující odhalení příslušného $\frac{p}{r}$ pomocí řetězových zlomků.

Zbývá ukázat, s jakou přesností za těchto okolností platí odhad $(*)$. Označme

$$\varphi = \frac{rz}{M} - p$$

aproximační „chybu“, která podle našeho předpokladu splňuje

$$|\varphi| \leq \frac{r}{2M}.$$

Aproximujeme součet geometrické řady:

$$\left| \sum_{s=0}^{\ell} \exp\left[2\pi i \frac{rz}{M} s\right] \right|^2 = \left| \sum_{s=0}^{\ell} \exp\left[2\pi i \varphi s\right] \right|^2 = \frac{\left|\exp\left[2\pi i \varphi(\ell+1)\right] - 1\right|^2}{\left|\exp\left[2\pi i \varphi\right] - 1\right|^2} = \frac{\sin^2 \pi\varphi(\ell+1)}{\sin^2 \pi\varphi},$$

kde poslední rovnost plyne ze vztahu

$$\left|e^{ix} - 1\right|^2 = (e^{ix} - 1)(e^{-ix} - 1) = 2(1 + \cos^2 x) = 4\sin^2 \frac{x}{2}.$$

Není těžké ověřit, že hodnota klesá s rostoucím φ, což je v souladu s tím, že φ je míra nepřesnosti: maximum M/r je dosaženo v našem ideálním případě, který odpovídá $\varphi = 0$. Protože je navíc $\sin^2$ sudá funkce, dostáváme

$$\frac{\sin^2 \pi\varphi(\ell+1)}{\sin^2 \pi\varphi} \geq \frac{\sin^2 \frac{\pi}{2}\frac{r(\ell+1)}{M}}{\sin^2 \frac{\pi}{2}\frac{r}{M}}.$$

Z definice ℓ plyne, že $M - r < r(\ell+1) < M + r$. Čitatel zlomku je tedy velmi blízko jedné (pro $r/M < 1/100$ se liší od jedné o méně než tisícinu) a jmenovatel, který je naopak velmi malý, můžeme zhora dosti přesně odhadnout vztahem $\sin x < x$. Celkem dostáváme

$$\left| \sum_{s=0}^{\ell} \exp\left[2\pi i \frac{rz}{M} s\right] \right|^2 > 0.999 \cdot \frac{4}{\pi^2}\frac{M^2}{r^2} > \frac{2}{5}\frac{M^2}{r^2}$$

a

$$P(z) > \frac{2}{5}\frac{1}{r}.$$

Můžeme uzavřít, že s pravděpodobností alespoň $\frac{2}{5}$ naměříme z, pro které je $\frac{p}{r}$ přítomno v řetězovém rozvoji $\frac{z}{M}$.

Celkovou úspěšnost algoritmu shrnuje následující tabulka:

podmínka úspěchu	pravděpodobnost
volba vhodného a	$\frac{1}{2}$
z je blízko $p\frac{M}{r}$	$\frac{2}{5}$
p je nesoudělné s r	$\frac{1}{4}\frac{1}{\log\log n}$

Celková úspěšnost je tedy nejméně $\frac{1}{20}\frac{1}{\log\log n}$. Např. pro RSA modul délky 4096 je úspěšnost jednoho kola algoritmu nejméně 0.6%, takže čtyřista kol dává více než 90% pravděpodobnost úspěchu. Tento odhad je navíc zbytečně pesimistický zejména v požadavku na nesoudělnost r a p; i pokud jsou r a p soudělná, získáme část r a po několika pokusech je možné r zrekonstruovat jako nejmenší společný násobek nalezených faktorů.