《PRIMES is in **P**》论文精讲

陈威宇 黄兆鋆 周书予

信息科学技术学院

March 6, 2022



Introduction

$$\text{PRIMES} = \{w | w \text{ is binary representation of a prime number}\}$$

PRIMES 属于以下复杂度类:

- **EXP**, 比如众所周知的有 *Sieve of Eratosthenes*.
- **coNP**, 因为 $\overline{\text{PRIMES}} = \text{COMPOSITES} \in \mathbf{NP}$.
- **NP**, [Pra75].
- **BPP**, [AB99].
- **P**, [AKS02].

[Pra75]: Every Prime Has a Succinct Certificate

Lucas Primality Test

正整数 $n \geqslant 3$ 是质数当且仅当存在整数 $1 < a < n$ 满足 $a^{n-1} \equiv 1 \mod 1$ 且对于 $n-1$ 的每个质因子 p 都有 $a^{(n-1)/p} \not\equiv 1 \bmod n$.

证明.

$\Leftarrow$: 考虑 $\mathrm{ord}_n(a)$. $\Rightarrow$: 取 n 的原根作为 a. □

想验证 p 是质数, 只需要给出 $p-1$ 的质因数分解 $p-1 = \prod_{i=1}^{k} p_i$, **同时**进一步验证 p_i 都是质数.
可以归纳证明验证 p 时所需的 certificate 长度为 $O(\log p)$, 从而说明了 PRIMES $\in$ **NP**.

Generalization of Fermat's Little Theorem

Theorem

正整数 $n \geqslant 2$ 是质数当且仅当存在与 n 互质的整数 a, 满足

$$(x+a)^n \equiv x^n + a \mod n$$

这个结论常被戏称为 freshman's dream.

证明.

n 是质数 $\Rightarrow n|\binom{n}{k}$ $(\forall 1 \leqslant k < n)$, $(x+a)^n - x^n - a^n$ 系数全为 0.
n 不是质数 $\Rightarrow$ 若 $p^k \| n$, 则 $p^k \nmid \binom{n}{p} = \frac{n \times (n-1) \times \cdots \times (n-p+1)}{1 \times 2 \times \cdots \times p}$, p 次项系数非零. □

[AB99]: PRIMES is in **BPP**

前述结论用于素数测试的一大障碍在于需要计算的多项式次数太大了.

所以可以考虑随机选取一个模数多项式 $r(x) \in \mathbb{Z}[x]$, 然后检验 $(x+a)^n \equiv x^n + a \bmod (r(x), n)$ 是否成立.

Theorem

设 n 是一个含有质因子 $p \geqslant 17$ 的非质数, 且不是某个质数的整数次幂. 在 $\mathbb{Z}[x]$ 中随机选取一个 $l = \lceil \log n \rceil$ 次首一多项式 $r(x)$, 则有至少 60% 的概率, $(x+1)^n \not\equiv x^n + 1 \bmod (r(x), n)$.

[AB99]: PRIMES is in **BPP**

令 $P_n(x) = (x+1)^n - 1 - x^n$. 当 n 不是质数时, $P_n(x) \neq 0 \bmod p$, 这是因为 $[x^{p^k}]P_n(x) = \binom{n}{p^k} = \frac{n\times(n-1)\times\cdots\times(n-p^k+1)}{(p^k)!} \neq 0 \bmod p$, 其中 $p^k \| n$, 这里要求了 $n \neq p^k$.

于是我们在 $\mathbb{Z}_p[x]$ 下考虑问题. 注意到 $\mathbb{Z}_p[x]$ 下存在唯一分解, 不可约因式的概念, 记 $I(d)$ 表示其中 d 次不可约首一多项式的数量, 可以估计其下界[1]

$$I(d) = \frac{1}{d}\sum_{e|d}\mu(e)p^{d/e} \geqslant \frac{1}{d}\left(p^d - \sum_{e<d, e|d} p^{d/e}\right) \geqslant \frac{p^d}{d} - p^{d/2}$$

[1] 这个结论不太初等, 就不在此过多阐述了, 可详细参考 [LN94].

[AB99]: PRIMES is in **BPP**

记 $\mathcal{C}$ 表示 $\mathbb{Z}_p[x]$ 中含有 $> \frac{l}{2}$ 次不可约因式的 l 次首一多项式集合.
同样对 $|\mathcal{C}|$ 估计下界

$$|\mathcal{C}| = \sum_{k=\lfloor \frac{l}{2} \rfloor + 1}^{l} I(k) p^{l-k} \geqslant \sum_{k=\lfloor \frac{l}{2} \rfloor + 1}^{l} p^l \left(\frac{1}{k} - \frac{1}{p^{k/2}} \right) \geqslant \left(\ln 2 - \frac{1}{48} \right) p^l$$

(其中利用了 $p \geqslant 17$.)
注意一个 l 次多项式中不可能包含超过一个 $> \frac{l}{2}$ 次不可约因式.

[AB99]: PRIMES is in **BPP**

因为 $\deg P_n(x) < n$，$P_n(x)$ 只有不超过 $\frac{2n}{l}$ 个 $> \frac{l}{2}$ 次不可约因式，从而 $\mathcal{C}$ 中只有不超过 $\frac{2n}{l} \cdot p^{\frac{l}{2}-1}$ 个多项式可能是 $P_n(x)$ 的因式.

$$\begin{aligned}
\mathbb{P}\left(r \text{ doesn't divide } P_n\right) &\geqslant \mathbb{P}\left(r \in \mathcal{C} \wedge r \text{ doesn't divide } P_n\right) \\
&= \mathbb{P}\left(r \in \mathcal{C}\right) - \mathbb{P}\left(r \in \mathcal{C} \wedge r \text{ divides } P_n\right) \\
&\geqslant \left(\ln 2 - \frac{1}{48}\right) - \frac{2n}{l} \cdot p^{\frac{l}{2}-1}/p^l \\
&\geqslant \ln 2 - \frac{1}{48} - \frac{2n}{l4^{\frac{l}{2}+1}} > \frac{3}{5}
\end{aligned}$$

(注意到 $n > p \geqslant 17, l = \lceil \log n \rceil \geqslant 5$.)

Notations & Preliminaries

在这一部分我们引入一些记号.

当 $\gcd(a, r) = 1$ 时, 满足 $a^k \equiv 1 \bmod r$ 的正整数 k 是一定存在的, 称其中最小的为 a 模 r 的阶, 记作 $\mathrm{ord}_r(a)$. 用欧拉函数 $\varphi(r)$ 表示比 r 小的与 r 互质的数的数量, 容易验证 $\mathrm{ord}_r(a) | \varphi(r)$.

当 p 是素数, $h(x)$ 是 $\mathbb{Z}_p[x]$ 中不可约多项式, 则 $\mathbb{Z}_p[x]/h(x)$ 构成一个有限域. 下文中将不加声明地用记号 $\mathbb{F}$ 表示 $\mathbb{Z}_p[x]/h(x)$.

$\tilde{O}(f(n)) = O(f(n)\ \mathrm{poly}(\log f(n))) = O(f^{1+\varepsilon}(n))$ 表示忽略对数因子.

Natural Language Description

算法分为如下五个步骤:

1. 第一步检查 n 是否是“整数幂”形式的非素数.
2. 第二步求出最小的满足 $\mathrm{ord}_r(n) \geqslant \log^2 n$ 的正整数 r.
3. 第三步检查 n 是否有不超过 r 的非平凡因子.
4. 第四步是平凡的: 如果 $n \leqslant r$ 没有不超过 r 的非平凡因子, 此时 n 一定是质数.
5. 第五步对于 $a = 1, \cdots, \lfloor \sqrt{\varphi(r)} \log n \rfloor$, 检查 $(x+a)^n \equiv x^n + a \bmod (x^r - 1, n)$ 是否成立, 若存在不满足则 n 不是质数, **否则 n 是质数**.

只有最后一句话的正确性是非平凡的.

Pseodocode

1: **function** AKS_PRIMALITY_TESTING(n)
2: **if** $n = a^b$ for $a \in \mathbb{N}$ and $b > 1$ **then**
3: **return** COMPOSITE
4: $r \leftarrow \min\{m \in \mathbb{N} | \mathrm{ord}_m(n) > \log^2 n\}$
5: **if** $1 < \gcd(a, n) < n$ for some $a \leqslant r$ **then**
6: **return** COMPOSITE
7: **if** $n \leqslant r$ **then**
8: **return** PRIME
9: **for** $a = 1 \rightarrow \lfloor \sqrt{\varphi(r) \log n} \rfloor$ **do**
10: **if** $(x + a)^n \not\equiv x^n + a \bmod (x^r - 1, n)$ **then**
11: **return** COMPOSITE
12: **return** PRIME

Reference

[AB99] M. Agrawal and S. Biswas, *Primality and identity testing via chinese remaindering*, 40th Annual Symposium on Foundations of Computer Science (Cat. No.99CB37039), 1999, pp. 202–208.

[AKS02] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena, *Primes is in p*, Annals of Mathematics **160** (2002).

[LN94] Rudolf Lidl and Harald Niederreiter, *Introduction to finite fields and their applications*, 2 ed., Cambridge University Press, 1994.

[Pra75] Vaughan R. Pratt, *Every prime has a succinct certificate*, SIAM Journal on Computing **4** (1975), no. 3, 214–220.