Primality Testing

Hyunjae Oh

Abstract

In this essay we will discuss methods of testing if a number is prime and finding large prime numbers.

1 Introduction

Finding large primes is an extremely important part of cryptography. From generating massive fields over modulo p to creating computationally difficult problems such as the discrete logarithm problem to add security, it's no wonder that cryptographers have invented numerous methods to find large primes. In this essay, we will discuss some methods of primality testing and understand the logic of how they work.

2 Fermat Primality Testing

Theorem 1 *(Fermat's Little Theorem) If p is a prime number and a is any integer such that* $p \nmid a$*, then* $a^p \equiv a \pmod{p}$.

Proof. Let us prove the corollary to Fermat's Little Theorem which states that $a^{p-1} \equiv 1$ (mod p). Every integer can be written as 0, 1, 2, ..., p-1 (mod p). Let's ignore 0 (mod p) because those are the multiples of p, and $p \nmid a$. Let's now take the set of numbers 1, 2, ..., p-1 and multiply them by a. We get:

a, 2a, ..., (p-1)a

These are the first p-1 multiples of a. Now we want to show that these multiples of a give unique numbers for modulo p. Let's assume that two of these multiples give the same number modulo p. That means that for some unique $r < p$ and $s < p$, $ra \equiv sa$ (mod p). Let's rewrite this as

$a(r-s) \equiv 0 \pmod{p}$

since $p \nmid a$, that means that $p \mid (r-s)$ for this congruence to be true. But since $r < p$ and $s < p$, the only way for $p \mid (r-s)$ to be true is if r = s, which contradicts our assumption that r and s are unique. Therefore for each multiple of a, you get a unique number modulo p.

From this realization we get that $\prod_{i=1}^{p-1} ia \equiv \prod_{i=1}^{p-1} i \equiv (p-1)! \pmod{p}$, or the product of the multiples of a is congruent to $(p-1)!$. We can also simplify $\prod_{i=1}^{p-1} ia = a * 2a * \ldots * (p-1)a$ as

$a^{p-1}(p-1)!$.

Therefore, $a^{p-1}(p-1)! \equiv (p-1)! \pmod{p}$. Since we know that $p \nmid (p-1)!$, we can divide both sides by $(p-1)!$ and get that $a^{p-1} \equiv 1 \pmod{p}$.

So does that mean we can just check if Fermat's Little Theorem holds true for some number n, then that n is prime? Unfortunately no. Even though we know that FLT holds true for all prime numbers, it also holds true for some odd composite numbers as well. For example, $3^{91-1} \equiv 1 \pmod{91}$, but $91 = 13 * 7$. These numbers are called *pseudoprimes*. But can't we just change the base a until there is a very high probability that a number is prime or not? Unfortunately again, there are odd composite numbers n such that for any base a, $a^{n-1} \equiv 1 \pmod{n}$. These are called *Carmichael numbers*. Some examples are $561 = 3 * 11 * 17$, $1105 = 5 * 13 * 17$, and $1729 = 7 * 13 * 19$. We can see that for any $a < n$, $a^{561-1} \equiv 1 \pmod{561}$, $a^{1105-1} \equiv 1 \pmod{1105}$, and $a^{1729-1} \equiv 1 \pmod{1729}$. There is a method called the Miller-Rabin Primality [Mil76] test to determine whether or not a number is Carmichael or not, but we will not be discussing it in this paper.

3 Willan's Formula

Theorem 2 *(Wilson's theorem) If p is a prime number, then* $(p-1)! \equiv -1 \pmod{p}$. *This theorem is the main component that makes Willan's formula work.*

Proof. Let us write out $(p-1)!$ as $(p-1)(p-2)\ldots(2)(1)$. Since p is a prime number, then that means $(p-1)(p-2)\ldots(2)(1)$ is a field modulo p. That means each number n should have an inverse n^{-1} such that $n * n^{-1} \equiv 1 \pmod{p}$. That means if we pair each number with its inverse, most of them will equal 1 modulo p, except for the numbers x that are its own inverse. We can solve for these x by setting up the congruence $x * x \equiv x^2 \equiv 1 \pmod{p}$.

Subtracting 1 from both sides we get $x^2 - 1 \equiv (x-1)(x+1) \equiv 0 \pmod{p}$. We get that if $x - 1 \equiv 0 \pmod{p}$ then $x \equiv 1 \pmod{p}$ and if $x + 1 \equiv 0 \pmod{p}$ then $x \equiv -1 \pmod{p}$. We notice that $-1 \equiv p - 1 \pmod{p}$. Therefore, most of the terms match with their inverse and our congruence becomes $(p-1)! \equiv (p-1)(1)(1)\ldots(1)(1) \equiv (p-1) \equiv -1 \pmod{p}$.

Postulate 3 *(Bertrand's Postulate) For any integer $n > 1$, there exists some prime p such that $n < p < 2n$. This postulate will help with the bound for the summation in Willan's formula.*

Now let's talk about Willan's formula. [WIL64] Here it is:
$p_n = 1 + \sum_{i=1}^{2^n} \lfloor (\frac{n}{\sum_{j=1}^{i} \lfloor (cos \frac{(j-1)!+1}{j} \pi)^2 \rfloor})^{1/n} \rfloor$

where p_n is the nth prime. Let us make some sense of this formula, starting from $\frac{(j-1)!+1}{j}$ (1). $(j-1)!+1$ may seem familiar, and it is in fact just Wilson's theorem. If and only if j is a prime will $(j-1)! + 1$ be a multiple of j. Therefore, (1) will only output an integer if j is a prime number.

Next, we multiply (1) by pi, input it into a cosine function, square it, and then take the floor. For $k \in \mathbb{Z}$, $cos(k\pi) = \pm 1$. If $k \notin \mathbb{Z}$, then $-1 < cos(k\pi) < 1$. Therefore, if we square $cos(k\pi)$, then we will get 1 if k is an integer, and some value $0 \leq cos(k\pi) < 1$ otherwise. We can then take the floor of $cos(k\pi)$, which will give us either 1 or 0. Recall that if j is prime, then (1) is an integer. That means that $\lfloor (cos \frac{(j-1)!+1}{j} \pi)^2 \rfloor$ (2) will give us 1 if j is prime, and 0 otherwise.

Now let us consider $\lfloor (\frac{n}{\sum_{j=1}^{i}(2)})^{1/n} \rfloor$ (3). The sum $\sum_{j=1}^{i}(2)$ simply counts the number of prime numbers from 1 to i and then adds one, and will always be an integer greater than or equal to 1. If we look at the graph of $f(n) = (\frac{n}{s})^{1/n}$ where $s \in \mathbb{N}$, we get that $0 < f(n) < 2$. We find that $f(n) \geq 1$ for $n \geq s$. If we take the floor of $f(n)$, we get that $f(n) = 1$ for $n \geq s$, and 0 otherwise. This means that (3) will output a 1 if $n \geq p_n + 1$ where p_n is the $\#primes \leq i$.

Finally, if we take the summation of (3) from 1 to 2^n, we will get the nth prime number - 1. Clearly $2^n \geq 2n$ for an integer $n > 1$ and by Bertrand's postulate, we are guaranteed to check i's that are greater than the nth prime. The final step is to just add 1 to get the nth prime number.

So now we have a formula to calculate the nth prime number. What's the point of any other primality tests? Willan's Formula has a major flaw that makes it near useless: it takes an extremely long time to compute. Remember

that we have to calculate $(j-1)!$ for every j from 1 to i, and then iterate those i's from 1 to 2^n. Of course, we can massively improve the 2^n bound, but it still won't change the fact that $(j-1)!$ will take a factorial amount of time to compute. This makes Willan's formula only useful for very small i, which isn't very helpful. However, perhaps this formula can be used in the future if we manage to create quantum computers that can calculate large factorials in reasonable amounts of time.

4 Lucas-Lehmer Test

Definition 4 *(Mersenne Numbers) A Mersenne number is a number in the form* $2^n - 1$.

Definition 5 *(Mersenne Primes) A Mersenne Prime is a prime number that is also a Mersenne Number. If n is composite, then* $2^n - 1$ *must be composite. If n is prime, then* $2^n - 1$ *may or may not be a prime. A trivial proof is* $2^{ab} - 1$ *can be written as* $(2^a - 1)(2^a + 2^{2a} + 2^{3a} + \ldots + 2^{(b-1)a})$. *a and b are symmetrical and can be switched if one wishes to.*

The Lucas-Lehmer Test [BRU93] is meant to determine whether the nth Mersenne number is prime, if n is a prime number. Let's start by defining the Lucas-Lehmer sequence:

$s_i = 4$ if $i = 0$, and $s_i = s_{i-1}^2 - 2$ otherwise.

The Lucas-Lehmer Test says that the Mersenne number M_p is prime only when $s_{p-2} \equiv 0 \pmod{M_p}$.

Proof. Let $\omega = 2 + \sqrt{3}$ and $\bar{\omega} = 2 - \sqrt{3}$. $\omega\bar{\omega} = 2^2 - \sqrt{3}^2 = 4 - 3 = 1$. We will create a group X that has element ω then prove the Lucas-Lehmer test using a contradiction.

Lemma 6 $S_m = \omega^{2^m} + \bar{\omega}^{2^m}$, *where* S_m *is the* m*th number in the Lucas-Lehmer Sequence. Here is a simple proof by induction:* $S_1 = 4^2 - 2 = 14$. $\omega^{2^1} + \bar{\omega}^{2^1} = \omega^2 + \bar{\omega}^2 = (2+\sqrt{3})^2 + (2-\sqrt{3})^2 = 4 + 4\sqrt{3} + 3 + 4 - 4\sqrt{3} + 3 = 8 + 6 = 14$. *So* S_1 *works. Now assume that* $S_k = \omega^{2^k} + \bar{\omega}^{2^k}$ *works. Then* $S_{k+1} = \omega^{2^{k+1}} + \bar{\omega}^{2^{k+1}} = S_k^2 - 2$. $S_k^2 - 2 = (\omega^{2^k} + \bar{\omega}^{2^k})^2 - 2 = \omega^{2^{k+1}} + \bar{\omega}^{2^{k+1}} + 2\omega^{2^k}\bar{\omega}^{2^k} - 2 = \omega^{2^{k+1}} + \bar{\omega}^{2^{k+1}} + 2 * 1 - 2 = \omega^{2^{k+1}} + \bar{\omega}^{2^{k+1}}$.

If M_p *divides* $S_{p-2} = \omega^{2^{p-2}} + \bar{\omega}^{2^{p-2}}$, *then by definition of modular arithmetic* $\omega^{2^{p-2}} + \bar{\omega}^{2^{p-2}} \equiv 0 \pmod{M_p}$. *Let us write this as* $\omega^{2^{p-2}} + \bar{\omega}^{2^{p-2}} =$

RM_p for some integer R. Multiply both sides by $\omega^{2^{p-2}}$ we get that $\omega^{2^{p-1}} = RM_p\omega^{2^{p-2}} - 1$ (1). Square both sides and we get $\omega^{2^p} = \left(RM_p\omega^{2^{p-2}} - 1\right)^2$ (2).

Lemma 7 *Let X be a set with a binary operation which is associative and has an identity. If X^* is the set of elements in X that have an inverse with respect to multiplication, then X^* forms a group. Here is a simple proof: Since X^* has an identity with respect to multiplication, clearly $1 \in X^*$. If elements x_1 and x_2 have inverses x_1^{-1} and x_2^{-1} respectively, then x_1x_2 has an inverse $x_1^{-1}x_2^{-1}$ and the set is closed. This lemma will help us show that the group for our proof really is a group.*

Lemma 8 *If G is a finite group then the order of an element is less than or equal to the order of the group. If $a \in G$, and $a^b = 1$, then the order of a divides b. This lemma will help us write inequalities that will lead to our contradiction.*

Now let's begin the actual proof for the Lucas-Lehmer test. Assume that M_p is composite and M_p divides $S_{p-2} = \omega^{2^{p-2}} + \bar{\omega}^{2^{p-2}}$, and pick some prime divisor q such that $q^2 \leq M_p$ and $q \neq 2$. Now let $\mathbb{Z}/q\mathbb{Z}$ be the set of integers modulo q, and X be the set $\{a + \sqrt{3}b : a, b \in \mathbb{Z}/q\mathbb{Z}\}$. Let's define multiplication on X as $(a_1 + \sqrt{3}b_1)(a_2 + \sqrt{3}b_2) = (a_1a_2 + 3b_1b_2) + (a_1b_2 + b_1a_2)\sqrt{3}$. X is associative and has an identity 1 under multiplication. Therefore, X^* is a group according to Lemma 7. There are q possible values of a and q possible values of b. Therefore, the order of X^* is $q^2 - 1$ since we don't want to count 0 which isn't included in X^* as it does not have an inverse that can make it equal to the identity 1 under multiplication. By Lemma 8, the order of any element in X^* is $\leq q^2 - 1$. Now consider ω which is an element of X and X^*. Since $q \mid M_p$, $RM_p\omega^{2^{p-2}} = 0$ as an element of X. Therefore, (1) and (2) become $\omega^{2^{p-1}} = -1$ and $\omega^{2^p} = 1$ respectively. That means that the order of ω in X^* is 2^p. Using Lemma 8, we know that the order of ω is less than the order of X^*, written as $2^p \leq q^2 - 1$. However, $q^2 - 1 < q^2 \leq M_p = 2^p - 1$ and we have a proof by contradiction, so M_p has no prime divisors q and is a prime number.

5 AKS algorithm

The AKS test, named after its creators Manindra Agrawal; Neeraj Kayal; and Nitin Saxena, is arguably one of the best primality tests we know of

today. Unlike the previously mentioned primality tests that all have some sort of limitation, the AKS test doesn't really have any limitations and works for any prime number.

Lemma 9 *If an integer $n \geq 2$ and an integer a has $gcd(a, n) = 1$, then n is prime if and only if $(X + a)^n \equiv X^n + a$ (mod n).*

Proof. We want to show that the coefficients of $(x + 1)^n - (x^n + 1)$ are divisible by n. The binomial theorem states that $(x + 1)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$. Then $(x + 1)^n - (x^n + 1) = \sum_{k=1}^{n-1} \binom{n}{k} x^k$. If n is prime, then $n \mid \binom{n}{k}$ for $k = 1, 2, \ldots, n-1$ because n will only have factors that are 1 and itself, so any number $< n$ will not share a prime factor with n. $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, and $k < n$, as well as $n - k < n$, so a prime number n will always divide $\binom{n}{k}$. Now let's suppose that n is composite, and $n = pd$ where p is a prime and d is an integer such that $p, d \in \{1, 2, \ldots, n-1\}$. To prove that $n \nmid \binom{n}{p}$ if n is composite, let's see that $\binom{n}{p} = \frac{n*(n-1)*\ldots*(n-p+1)}{p!}$. Then $\frac{n*(n-1)*\ldots*(n-p+1)}{p!} = n*l$ for some $l \in \mathbb{N}$. $l = \frac{(n-1)*(n-2)*\ldots*(n-p+1)}{p!}$. If p does not divide the numerator, then $l \notin \mathbb{N}$. $(n-p+1) \equiv 1$ (mod p) because $n = pd$, which means $n-p = pd-p = p(d-1)$ is clearly divisble by p, so $(n - p + 1) \equiv 1$(mod p). Continue this pattern, so that $(n - p + 2) \equiv 2$ (mod p), $(n - p + 3) \equiv 3$ (mod p), all the way up to $(n - 1) \equiv (p - 1)$ (mod p). We get that $(n - 1) * (n - 2) * \ldots * (n - p + 1) \equiv 1 * 2 * \ldots * (p - 1) \equiv (p - 1)!$ (mod p). Since p is prime, $(p - 1)! \equiv -1$ (mod p) by Wilson's theorem. Since $(p - 1)! \not\equiv 0$ (mod p), p does not divide $(n - 1) * (n - 2) * \ldots * (n - p + 1)$ so l is not an integer, and $n \nmid \binom{n}{p}$.

The actual algorithm to deteremine whether a number n is prime goes like this:

1. Check if there exists integers $a, b > 1$ such that $n = a^b$. If so, then n is not prime. Otherwise, move on.

2. Find the smallest r such that the order of r modulo n is $> log^2(n)$.

3. If $1 < gcd(a, n) < n$ for some $a \leq r$, then n is composite.

4. If $n \leq r$ then n is prime.

5. For a = 1 to $\lfloor \sqrt{\phi(r)} log(n) \rfloor$, compute $(x + a)^n \equiv x^n + a$ (mod $x^r - 1$, n). If this is ever not true, then n is composite.

6. Otherwise, n is prime.

Step 1 is just a filter so that the rest of the algorithm doesn't break. The main steps are steps 2 and 5, but let's also briefly explore steps 3 and 4. If $1 < gcd(a, n) < n$, then clearly n is not prime as it shares a factor with some number $a \leq r$. In step 4, if $n \leq r$ then we have checked that all integers $\leq n$ do not share any factors with n, so n is clearly prime. Step 5 is the most clever part of the algorithm, because it greatly lessens the time to check whether Lemma 9 holds for n. For a proof, see [AKS04].

References

[Mil 76] Gary L. Miller. 1976. Riemann's hypothesis and tests for primality. *Journal of Computer and System Sciences* 13 (3): 300-317. Working papers presented at the ACMSIGACT Symposium on the Theory of Computing (Albuquerque, N.M., 1975).

[WIL64] Willans CP. On Formulae for the Nth Prime Number. The Mathematical Gazette. 1964;48(366):413-415. doi:10.2307/3611701

[BRU93] Bruce, J. W. (1993). A Really Trivial Proof of the Lucas-Lehmer Test. The American Mathematical Monthly, 100(4), 370–371. https://doi.org/10.1080/00029890.1993.11990414

[AKS04] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. 2004. PRIMES is in P. *Annals of Mathematics (2)* 160 (2): 781-793. https://doi.org/10.4007/annals/2004.160.781.