On Discovering Prime Numbers

Aaron Wang (atw9139)

September 7, 2021

1 Background

Prime numbers are an important topic in number theory, with wide ranging applications in cryptography, computing, and even recreation; There are infinitely many primes, but having discovered the largest known prime is a great honor. The current record is $2^{82,589,933} - 1 \approx 10^{24,862,047.1729}$ discovered in 2018 [3]. Written out fully, it is:

148894445742041325547806458472397916603026273992795324185271289425213239361064475310309971132180337174752834401422

It is a pity that this page is too narrow to contain it. Where does it even end?

This prime was discovered in GIMPS, the Great Internet Mersenne Prime Search. It is a distributed computing project with participants around the world donating their spare computing power to finding ever larger prime numbers. You can join them at https://www.mersenne.org/, and there are prizes ($3k - $50k) for discovering primes!

2 Proving a Number is Prime

The primality of a number is most easily checked by referring to the definition of primality itself:

Definition 1 (Primality)**.** *An integer p is prime iff* $p > 1 \wedge \neg(\exists k \in \mathbb{Z} : 1 < k < p)(k|p)$

Therefore, one can check that an integer n is prime by trial division, checking that n is not divided evenly by every smaller positive integer other than 1. Clearly, this becomes impractical for any n larger than about 10^2 if done by hand, or about 10^8 for a computer.

A readily implementable simplification — at least in terms of practical computability — is as such:

Lemma 1.

$$p > 1 \Rightarrow \left(\neg(\exists k \in \mathbb{Z} : 1 < k \leqslant \sqrt{p})(k|p) \Leftrightarrow \neg(\forall k \in \mathbb{Z} : 1 < k < p)(k|p)\right)$$

Proof. Fix p to some integer greater than 1. For the forward implication, assume for the sake of contradiction that, $\neg(\exists j \in \mathbb{Z} : 1 < j \leqslant \sqrt{p})(j|p)$ and $(\exists k \in \mathbb{Z} : \sqrt{p} < k < p)(k|p)$. That is, there is no integer factor of p less than or

equal to $\sqrt{p}$, but there is one greater than $\sqrt{p}$, namely k. Since $k|p$, write $p = kd$ for some $d \in \mathbb{Z}$. Observe that this implies d is also an integer factor of p, and that $d > 1$ since $k < p$. However, $1 < d = \frac{p}{k} < \frac{p}{\sqrt{p}} = \sqrt{p}$, which is a clear contradiction of the first assumption ($\Rightarrow\Leftarrow$).

The reverse implication is trivial. If there is no factor of p in $(1, p)$, there isn't one in the subset $(1, \sqrt{p}]$ either. □

This tightening of bounds on k from $(1, p)$ to $(1, \sqrt{p}]$, is significant, since it guarantees that the amount of computation necessary to show that a number is prime can be vastly decreased. Using this, it becomes practical to check the primality of integers all the way up to 10^4 by hand if one is very perseverant, and up to 10^{16} on a computer.

Even more can be done to reduce the amount of work needed to prove primality. Let $\mathbb{P}_n$ be the set of prime numbers less than n, thus $\mathbb{P}_7 = \{2, 3, 5\}$ and $\mathbb{P}_{26.54} = \{2, 3, 5, 7, 11, 13, 17, 19, 23\}$.

Lemma 2.

$$\mathbb{N} \ni p \text{ prime} \Leftrightarrow p > 1 \wedge \neg(\exists k \in \mathbb{P}_p)(k|p)$$

Proof. Fix p to some integer greater than 1. For the reverse implication, assume for the sake of contradiction that p is composite, and that there is no prime factor of p. Since p is composite, $(\exists d \in \mathbb{N}, d > 1)(d|p)$. By the Fundamental Theorem of Arithmetic, d has a unique prime factorization which includes at least one prime factor, $j \in \mathbb{P}_p$. Since $j|d \wedge d|p$, then $j|p$, but this is in contradiction with the assumption that p has no prime factors ($\Rightarrow\Leftarrow$).

The forward implication is trivial. If p is prime, then p has no integer factors $k > 1$, let alone prime factors. □

This time, the simplification seemingly does not actually decrease the number of divisibility checks necessary to show a number is prime. The $\sqrt{p}$ method requires $\sqrt{p}$ divisibility checks. In comparison, the Prime Number Theorem [2] implies that this prime factor method requires approximately $\frac{p}{\ln p}$ checks, which is always larger. However, combining the two approaches does lead to a more efficient primality test:

Theorem 1 (Sieve of Eratosthenes)**.**

$$\mathbb{N} \ni p \text{ prime} \Leftrightarrow p > 1 \wedge \neg(\exists k \in \mathbb{P}_{\sqrt{p}+1})(k|p)$$

Proof. Fix p to some integer greater than 1. For the reverse implication, assume for the sake of contradiction that p is composite and $\neg(\exists k \in \mathbb{P}_{\sqrt{n}+1})(k|p)$. If p is composite, then it can be written $p = cd$; WLOG let $1 < d \leqslant \sqrt{p}$. By the Fundamental Theorem of Arithmetic, d has a prime factorizationwhich includes at least one prime factor, m. Since $m|d$, $1 < m \leqslant d \leqslant \sqrt{p}$, thus $m \in \mathbb{P}_{\sqrt{p}+1}$. But, additionally $d|p$, so $m|p$, and this is a contradiction with the assumption that p has no prime factors less than or equal to $\sqrt{p}$ ($\Rightarrow\Leftarrow$).

The forward implication is proved in the same way is it is in Lemma 2. □

The algorithm that arises from this theorem is named the Sieve of Eratosthenes since it involves sifting through all the integers from 2 to n to find every prime in between [4]. The computation can be made practical by hand by using a list that contains every integer from 2 to n and starting from the first integer 2:

- Mark the current integer as prime.
- Remove every multiple of the marked integer from the list. Not all of them may still be in the list.
- If the marked integer is greater than the square root of n, then the computation is done and every remaining number is marked as prime. Otherwise, move on to the next integer.

This algorithm is rather time-efficient for determining the primality of arbitrary integers. $|\mathbb{P}_{\sqrt{p}+1}| \approx \frac{2\sqrt{p}}{\ln p}$ by the Prime Number Theorem, which is also the number of divisibility checks needed; this is only $\frac{2}{\ln p}$ as many checks as for the naive $\sqrt{p}$ method. As p approaches infinity, this ratio tends to 0. Of course, a list of the prime numbers less than p must be supplied to the computer running these checks, or it must find them by itself. Even so, if the computer's goal is, for example, to check the primality of every integer from 1 to 10^{24}, the computer can simply save the prime numbers it finds into a list that continually gets referenced to check the primality of larger integers. The cost in time of finding the smaller prime numbers is amortized over the many integers it can check all in one go. In practice, however, the requirement of storing the list of p integers becomes memory-prohibitive once p exceeds about 10^{12}, so the full sieve is not used to find large primes.

Even a partial sieve is more efficient than the naive $\sqrt{p}$ method, though. By simply skipping over even integers which are obviously composite, the naive method is improved twofold. Skipping over every multiple of 2 and 3 allows the computer to check only integer factors $n \equiv \pm 1 \pmod 6$, or a threefold decrease in time. There is a trade-off between the length of the list of prime numbers stored that increases memory cost and the necessity of checking numbers that are multiples of unstored primes, which increases time cost.

There are more complicated sieves that achieve better time and memory efficiency, and even primality tests for arbitrary integers that do not depend on divisibility checks like ECPP [9] and AKS [10] which achieve near-polynomial and polynomial asymptotic time complexity; these are more commonly used to check the primality of large integers, but they are mostly only usable for integers up to around $10^{20,000}$. This figure falls far, far short of $10^{24,862,047}$, near the record prime from above. How can such large primes be tested in reasonable time?

3 Mersenne Numbers and Mersenne Primes

Definition 2 (Mersenne Number)**.** *A Mersenne number is some M_n such that $M_n = 2^n - 1$, where $n \in \mathbb{N}$.*

They are named after Marin Mersenne, who extensively studied them in the 1600s.

Definition 3 (Mersenne Prime)**.** *A Mersenne prime is some prime M_p such that $M_p = 2^p - 1$, where $p \in \mathbb{N}$.*

If n is prime, then there is a possibility that M_n is prime. In that case, M_n would be a Mersenne prime. If n is not prime, M_n is definitely not prime either.

Theorem 2. *If n is composite, then M_n is composite.*

Proof. If $n \in \mathbb{N}$ is composite, then if can be written as $n = ab$, where $\mathbb{N} \ni a, b > 1$. Thus, $M_n = 2^{ab} - 1 = (2^a - 1)(2^{a(b-1)} + 2^{a(b-2)} + \ldots + 2^{a(2)} + 2^{a(1)} + 2^{a(0)})$, which has a factor $2^a - 1$. Since $a > 1$, $2^a - 1 > 1$. Also, since $2^{1a} - 1 < 2^{ba} - 1$, then $2^a - 1$ is a proper factor. □

The first few Mersenne primes are, (OEIS A000668) $3, 7, 31, 127, 8191, 131071, 524287, 2147483647, \ldots$ [5]. They grow quickly due to the exponential term. Evidently, the record prime from above is actually $M_{82,589,933}$, and it is probably (but is not yet verified to be) the 51^{st} Mersenne prime. Because of the rigid structure of Mersenne numbers, they can be checked extremely quickly for primality compared to other integers of similar magnitude.

4 Lucas-Lehmer Primality Test

The Lucas-Lehmer Primality Test is a powerful theorem named after Édouard Lucas and D.H. Lehmer that deterministically checks whether the n^{th} Mersenne number is prime [1]. Apocryphally, Lucas used a prototype of this test to verify by hand in a herculean effort spanning 19 years that $2^{127} - 1 = M_{127}$ is prime [8].

Theorem 3 (Lucas-Lehmer Primality Test)**.** *M_p is prime iff $s_{p-2} \equiv 0 \pmod{M_p}$, where sequence $\{s_i\}$ is defined,*

$$s_i = \begin{cases} 4, & i = 0 \\ s_{i-1}^2 - 2, & i > 0 \end{cases}$$

The first few terms of the sequence $\{s_i\}$ are, (OEIS A003010) $4, 14, 194, 37634, 1416317954, \ldots$ [6]

This primality test does not rely on divisibility checking by many integers. It only requires that *one* number be divisble by another, namely that $M_p | s_{p-2}$. Unfortunately both of these numbers grow large quickly, and the second one much faster than even the first. Nevertheless, it is remarkably simple considering what it can prove. A proof of the Lucas-Lehmer Test is omitted here since it relies on rather complicated math (group theory!!!).

This method of checking primality requires vastly less computation than methods for abritrary integers. For the case of $M_{82,589,933}$, the Sieve of Eratosthenes would require a colossal list of prime numbers up to that integer, and then approximately $\frac{2\sqrt{M_{82,589,933}}}{\ln M_{82,589,933}} \approx \frac{10^{12,431,000}}{28,500,000} \gg 10^{12,430,900}$ divisibility checks, which is completely unthinkable. Using the Lucas-Lehmer Test requires calculating the $82,589,931^{th}$ term of the sequence. While this is hard since the numbers involved are quite large, it only takes a few weeks on a single computer and days on a cluster. This is helped by the fact that each term in the sequence is not needed, just their remainder divided by $M_{82,589,933}$.

4.1 Solving for the Lucas-Lehmer sequence

Solving for a closed-form expression of the i^{th} term of the Lucas-Lehmer sequence $\{s_i\}$ is not as simple as for some other recurrence relations given in Scheinerman's Mathematics: A Discrete Introduction [11], which are either first-order affine, second-order linear, or generated by polynomials. $\{s_i\}$, in comparison, is first-order, but it is not linear because of the square and constant terms.

The first step is inspired by Scheinerman. Simply start listing out the terms of the sequence, replacing the initial term with an arbitrary integer.

$$\{s_i\} = (m, m^2 - 2, m^4 - 4m^2 + 2, m^8 - 8m^6 + 20m^4 - 16m^2 + 2, \ldots)$$

The highest degree (and thus likely most important) terms in each sequence term are $m, m^2, m^4, m^8, \ldots$ reflecting the square nature of the sequence definition. Thus, an *ansatz*, or educated initial guess, might be $s_i = m^{2^i}$. Of course, $m^{2^{i+1}} = (m^{2^i})^2 \neq (m^{2^i})^2 - 2$, but it is a start. The solution is probably some modification of the term m^{2^i} that when squared produces $(m^{2^{i+1}}) + 2$. With just one term, though, this is clearly impossible. What about two?

$$\left(m^{2^i} + n^{2^i}\right)^2 = m^{2^{i+1}} + n^{2^{i+1}} + 2mn^{2^i}$$

This equation suggests that we are almost done with the general solution! The only remaining step is to ensure that $mn^{2^i} = 1$, or equivalently $mn = 1$. Simply let $n = m^{-1}$:

$$\left(m^{2^i} + m^{-2^i}\right)^2 = m^{2^{i+1}} + m^{-2^{i+1}} + 2$$

$$s_{i+1} = m^{2^{i+1}} + m^{-2^{i+1}} = \left(m^{2^i} + m^{-2^i}\right)^2 - 2 = s_i^2 - 2$$

Therefore, the general solution to the Lucas-Lehmer Test sequence is,

$$s_i = m^{2^i} + m^{-2^i}$$

In fact, the Lucas-Lehmer test is correctly shows the primality of every Mersenne Primes for infinitely many initial values, given by (OEIS A018844) [7], and correctly shows primality of some Mersenne Primes for some other choices of inital value. thus this solution is useful for many values of m. It so happens that an initial value of 4 is the smallest that is universal, though. Finally, find the value of m that satisfies this initial condition.

$$4 = s_0 = m^1 + m^{-1}$$

$$4m = m^2 + 1$$

$$m = \frac{4 \pm \sqrt{16 - 12}}{2} = 2 \pm \sqrt{3}$$

The specific solution to the Lucas-Lehmer Test sequence with initial value 4 is,

$$s_i = \left(2 + \sqrt{3}\right)^{2^i} + \left(2 - \sqrt{3}\right)^{2^i}$$

A proof by induction is left as an exercise to the reader.

References

[1] J. W. Bruce. "A Really Trivial Proof of the Lucas-Lehmer Test". In: *American Mathematical Monthly* 100 (4 1993), pp. 370–371. ISSN: 0002-9890,1930-0972. DOI: 10.2307/2324959. URL: http://doi.org/10.2307/2324959.

[2] S Gerig. "A simple proof of the Prime Number Theorem". In: *Journal of Number Theory* 8 (2 1976), pp. 131–136. ISSN: 0022-314X,1096-1658. DOI: 10.1016/0022-314x(76)90096-2. URL: http://doi.org/10.1016/0022-314x%2876%2990096-2.

[3] Mersenne Research Inc. *Mersenne Prime Number discovery - $2^{82589933} - 1$ is Prime!* URL: https://www.mersenne.org/primes/?press=M82589933.

[4] Nicomachus. *Nicomachi Geraseni Pythagorei Introductionis arithmeticae libri II*. Ed. by Richard Hoche. 1866.

[5] *OEIS A000668: Mersenne primes (of form $2^p - 1$ where p is a prime).* URL: https://oeis.org/A000668.

[6] *OEIS A003010: A Lucas-Lehmer sequence: $a(0) = 4$; for $n > 0$, $a(n) = a(n-1)^2 - 2$.* URL: https://oeis.org/A003010.

[7] *OEIS A018844: Arises from generalized Lucas-Lehmer test for primality.* URL: https://oeis.org/A018844.

[8] *Prime Curios: 17014...05727 (39-digits).* URL: http://primes.utm.edu/curios/page.php?number_id=135.

[9] Robert Denomme; Gordan Savin. "Elliptic curve primality tests for Fermat and related primes". In: *Journal of Number Theory* 128 (8 2008), pp. 2398–2412. ISSN: 0022-314X,1096-1658. DOI: 10.1016/j.jnt.2007.12.009. URL: http://doi.org/10.1016/j.jnt.2007.12.009.

[10] Manindra Agrawal; Neeraj Kayal; Nitin Saxena. "PRIMES is in P". In: *Annals of Mathematics* 160 (2 2004), pp. 781–793. ISSN: 0003-486X. DOI: 10.4007/annals.2004.160.781. URL: http://doi.org/10.4007/annals.2004.160.781.

[11] Edward A. Scheinerman. *Mathematics: A Discrete Introduction 3^{rd} ed.* Cengage Learning, 2012. ISBN: 978-0-8400-4942-1.

5794795829753159520880719269367652178218447252664007691211435530831196948763376645782369507403795121032521790259 1

There it is!