Grade 11/12 Math Circles
March 20
Primality Testing

Prime Numbers

Definition

A prime number is an integer $p > 1$ whose only positive divisors are 1 and p.

Example

The primes less than 100 are given in this table.

2	3	5	7	11
13	17	19	23	29
31	37	41	43	47
53	59	61	67	71
73	79	83	89	97

Fundamental Theorem of Arithmetic

Every positive integer has a unique factorization into prime powers.

This theorem was apparently first stated fully by Gauss, though various weaker forms were stated and proved by Euclid, al-Farisi, Prestet, Euler, and Legendre.

Example

- $12 = 2^2 \times 3^1$
- $36 = 2^2 \times 3^2$
- $210 = 2^1 \times 3^1 \times 5^1 \times 7^1$
- $37 = 37^1$ (already prime)
- $1 =$ (empty product)

Euclid's Theorem

There are infinitely many primes.

Proof: Suppose towards contradiction that there are only finitely many prime numbers, say $p_1, \ldots, p_n$. But then $q = p_1 \times \cdots \times p_n + 1$ is divisible by none of the p_i's. By the Fundamental Theorem of Arithmetic, q has a prime factor r, but this r did not appear in the original list of primes; contradiction.

Is there a systematic way to determine the prime factorization of a number, as opposed to guessing? The most conceptually simple algorithm is trial division by primes.

Example

Let's calculate the prime factorization of 2093. 2093 is not divisible by 2, 3, or 5, but $2093 = 7 \times 299$. We continue by factoring 299. Notice that none of 2, 3, or 5 can divide 299, so our list of trial divisors starts at 7. 299 is not divisible by 7 or 11, but $299 = 13 \times 23$. Since 23 is prime, we conclude that the prime factorization of 2093 is $7 \times 13 \times 23$.

Notice that there is no point in trial dividing by composites. Indeed, since 2093 was not divisible by either 2 or 3, it cannot be divisible by 6, for example.

Example

By trial division, we find that $5687 = 11 \times 517$. We still need to check if 517 is divisible by 11, and indeed $517 = 11 \times 47$ and the prime factorization of 5687 is $11^2 \times 47$.

Exercise

Determine whether 161 is prime, and if not, factor it.

Exercise

Calculate the prime factorization of 1001.

How many primes do we need to divide by before we can conclude that the integer we are testing is prime? The following proposition gives us a clue.

Proposition

Suppose that a positive integer n has a non-trivial divisor $a \geq \sqrt{n}$ (a divisor equal to neither 1 nor n). Then n has a non-trivial divisor $b \leq \sqrt{n}$.

Exercise

Prove the proposition above.

Now suppose that we are performing trial factoring on n and have determined that for every prime $p \leq \sqrt{n}$, that $p \nmid n$ (p does not divide n). Then this implies that n is prime! Indeed, if n has any non-trivial divisor a, then n has a non-trivial divisor $b \leq \sqrt{n}$. If $a \leq \sqrt{n}$, we take $b = a$, and otherwise we use the previous proposition. In any case, b has a prime factor p, which is also a prime factor of n, and $p \leq b \leq \sqrt{n}$. The moral of the story is that we need only trial divide by primes up to $\sqrt{n}$, which is a *significant* speedup over trial dividing by primes up to n.

Exercise

Determine whether 1739 and 1741 are prime, and if not, factor them.

Example

By trial division, we find that $4981 = 17 \times 293$. Since no prime < 17 divides 4981, no prime < 17 divides 293. But 17 is the largest prime $\leq \sqrt{293}$, so 293 is prime and the prime factorization of 4981 is indeed 17×293.

Exercise

Find the prime factorization of 344929. (The calculation is not as bad as it seems).

Example

The prime factorization of $10^8 + 1 = 100000001$ is 17×5882353.

In the previous example, we would need to know beforehand that 5882353 was prime, as it could not be deduced by simply trial factoring by the primes up to 17. In general, trial factoring n requires us to have the list of primes $\leq \sqrt{n}$. Is there an efficient way to calculate the list of prime numbers up to a given number x?

Sieve of Eratosthenes

Write down the integers between 2 and x, then strike out every multiple of 2 except 2 itself. The next number not yet stricken is 3; strike out every multiple of 3 except 3 itself. Continue with 5 and so on, until the next unstricken number is $> \sqrt{x}$. The numbers in the list not stricken are the primes between 2 and x.

Example

(At this point I do an example on the board. Rather than insert a large amount of text into this document, I refer the reader to https://en.wikipedia.org/wiki/Sieve_of_Eratosthenes; there is a nice animation as of March 17, 2024).

Natural Logarithm

Define the mathematical constant

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots = 2.718\ldots$$

The natural logarithm $\ln(x)$ is the inverse function of the function e^x. That is, $e^{\ln(x)} = x$ for all $x > 0$ and $\ln(e^x) = x$ for all real numbers x.

Two useful identities related to $\ln(x)$ are $\ln(xy) = \ln(x) + \ln(y)$ and $\ln(x^y) = y\ln(x)$.

Prime Number Theorem

Let $\pi(x)$ be the number of primes which are $\leq x$. Then

$$\pi(x) \sim \frac{x}{\ln(x)}.$$

If you are familiar with calculus, this means that

$$\lim_{x\to\infty} \frac{\pi(x)}{\left(\frac{x}{\ln(x)}\right)} = 1.$$

Less rigorously, it means that $\pi(x)$ is approximately equal to $x/\ln(x)$ for real numbers x and that the ratio between these functions gets closer to 1 as x gets larger.

(What does $\pi = 3.14\ldots$ have to do with prime numbers? Rather confusingly, we are re-defining π here. This is a standard notation for the "prime counting function", and is simply used because π is the Greek equivalent of the letter "p").

Exercise

At the beginning of this talk, we listed the twenty-five primes which were ≤ 100. How many primes does this approximate formula predict?

Example

We list the values of $\pi(x)$ for several values of x and the approximations via the Prime Number Theorem:

- $\pi(10^3) = 168$, $\frac{10^3}{\ln(10^3)} \approx 145$
- $\pi(10^4) = 1229$, $\frac{10^4}{\ln(10^4)} \approx 1086$
- $\pi(10^5) = 9592$, $\frac{10^5}{\ln(10^5)} \approx 8686$
- $\pi(10^6) = 78498$, $\frac{10^6}{\ln(10^6)} \approx 72382$
- $\pi(10^7) = 664579$, $\frac{10^7}{\ln(10^7)} \approx 620421$

A question which a computer algorithm specialist would ask is "What is the time complexity of the Sieve of Eratosthenes?" Time complexity is a measure of the amount of time taken to complete an algorithm vs. the input. The input to the sieve algorithm is the value x, and to simplify things, we will assume that time is proportional to the number of strike-out operations we perform. In performing the algorithm, we strike out about $1/2$ of the numbers $\leq x$, then $1/3$, then $1/5$, and so on, till we strike out $1/p$ for the largest prime p which is $\leq \sqrt{x}$. Overall, we perform about

$$(1/2 + 1/3 + 1/5 + \cdots + 1/p)x = x \sum_{\substack{q \text{ prime} \\ q \leq \sqrt{x}}} \frac{1}{q}$$

strike-out operations (note that numbers with multiple prime factors get struck out more than once).

Mertens' Theorem

Define

$$S(x) = \sum_{\substack{p \text{ prime} \\ p \leq x}} \frac{1}{p}.$$

Then $S(x) \sim \ln(\ln(x))$.

The function $\ln(\ln(x))$ grows extremely slowly. For example, $\ln(\ln(10^{100})) < 6$. Nevertheless, $\ln(\ln(x)) > y$ for $x > e^{e^y}$, so it grows without bound as x grows.

Back to the sieve, Mertens' theorem implies that the sieve algorithm requires about

$$x \ln(\ln(\sqrt{x})) = x \ln(\ln(x)/2) = x(\ln(\ln(x)) - \ln(2)) \sim x \ln(\ln(x))$$

strike-out operations. Furthermore, if we are factoring n and sieve the primes up to $\sqrt{n}$, we should expect about $\sqrt{n}\ln(\ln(\sqrt{n})) \sim \sqrt{n}\ln(\ln(n))$ strikeout operations.

Exercise

Suggest an algorithm to sieve the primes between x and y, where x is not necessarily 0, and estimate its time complexity.

Trial factoring will yield the prime factors of a number in increasing order. Fermat discovered a factorization method for numbers of the form $n = ab$, where a and b are close to each other.

Proposition

Suppose there exist integers c, d such that $n + c^2 = d^2$. Then $n = d^2 - c^2 = (d - c)(d + c)$.

We can try factoring n by iterating through several values of c, and testing whether $n+c^2$ is a perfect square. This method will quickly detect factorizations of the form $(d - c)(d + c)$ if c is small (if the factors are close together).

Exercise

Find a factor of 999991.

Exercise

(Challenge) Find a factor of 2146681.

Next week, we will discuss modular arithmetic, and investigate more advanced primality tests such as Fermat's Little Theorem and the Miller-Rabin test.