# Prime numbers

The classical Greek mathematicians were more interested in geometry than in what we would call algebra. (One notable exception was Diophantus, whose surviving work is concerned with integer solutions of polynomial equations, and who was the first Western mathematician to develop any sort of algebraic notation.) However, one of the

**Definition 1.** A positive integer $n$ is called prime if it has exactly two factors, namely itself and 1.

Notice that 1 is not prime, because it has only one factor, not two. The first several prime numbers are 2, 3, 5, 7, 11, 13, 17, 23, ...., 1999, 2003, 2011, . . . , 999983, 1000003, 1000033, . . .

**Does the sequence of primes ever come to an end?**

In Book IX of the *Elements*, Euclid gave an elegant proof that the answer is **no**: there are infinitely many prime numbers. Euclid did not use the term "infinite", as the Greeks were not comfortable with the idea of infinity[1]. Instead, he stated the theorem in the following way.

**Theorem 2.** *No finite set of prime numbers can possibly be the complete list of all prime numbers.*

*Proof.* Suppose that we have a list of prime numbers: $p_1, p_2, \ldots, p_n$. Let

$$x = (p_1 p_2 \cdots p_n) + 1.$$

Then $x$ is a positive integer. Observe that

- $p_1$ can't be a factor of $x$ (because dividing $x$ by $p_1$ leaves a remainder of 1);
- $p_2$ can't be a factor of $x$ (because dividing $x$ by $p_2$ leaves a remainder of 1);
  ...
- $p_n$ can't be a factor of $x$ (because dividing $x$ by $p_n$ leaves a remainder of 1).

On the other hand, $p_n$ has to have at least one prime factor — either it is itself prime, or it is divisible by some smaller prime number. But whatever that prime factor is, we've seen that it can't be any of the primes $p_1, p_2, \ldots, p_n$. Therefore, that list of primes cannot be the complete list of all primes. □

There are a lot of natural questions to ask about primes (many of which are still unanswered, and are the subject of current research). Here are two.

**1. How do you test whether a number is prime?**

For example, suppose I ask you whether 416578961978323403 is prime. How would you find the right answer?

[1] Neither was anyone else; mathematicians didn't really begin to understand infinity until the work of Georg Cantor (1845–1918), more than two millennia after Euclid.

The Greeks knew a method, called the *Sieve of Eratosthenes.* to find all the primes up to a given number $N$. Rather than writing down the primes directly, the trick is to write down all the numbers up to $N$ and then systematically cross out the composite[2] numbers. For example, suppose that $N = 30$: Write down all the numbers from 2 to 30. (Ignore 1; that's a special case and we know that's not prime.)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |

The next number is 2. That has to be prime, so let's remember that by putting a box around it. On the other hand, all other multiples of 2 can't be prime, so cross them off.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\boxed{2}$ | 3 | ~~4~~ | 5 | ~~6~~ | 7 | ~~8~~ | 9 | ~~10~~ |
| 11 | ~~12~~ | 13 | ~~14~~ | 15 | ~~16~~ | 17 | ~~18~~ | 19 | ~~20~~ |
| 21 | ~~22~~ | 23 | ~~24~~ | 25 | ~~26~~ | 27 | ~~28~~ | 29 | ~~30~~ |

The first number we haven't either boxed or crossed out is 3. So 3 is prime, and we can cross off all larger multiples of 3:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\boxed{2}$ | $\boxed{3}$ | ~~4~~ | 5 | ~~6~~ | 7 | ~~8~~ | ~~9~~ | ~~10~~ |
| 11 | ~~12~~ | 13 | ~~14~~ | ~~15~~ | ~~16~~ | 17 | ~~18~~ | 19 | ~~20~~ |
| ~~21~~ | ~~22~~ | 23 | ~~24~~ | 25 | ~~26~~ | ~~27~~ | ~~28~~ | 29 | ~~30~~ |

Again, the first number we haven't either boxed or crossed out is 5. So 5 is prime, and we can cross off all larger multiples of 5 (the only one is 25):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\boxed{2}$ | $\boxed{3}$ | ~~4~~ | $\boxed{5}$ | ~~6~~ | 7 | ~~8~~ | ~~9~~ | ~~10~~ |
| 11 | ~~12~~ | 13 | ~~14~~ | ~~15~~ | ~~16~~ | 17 | ~~18~~ | 19 | ~~20~~ |
| ~~21~~ | ~~22~~ | 23 | ~~24~~ | ~~25~~ | ~~26~~ | ~~27~~ | ~~28~~ | 29 | ~~30~~ |

We could keep going, although as you can see, all of the surviving numbers are prime — we've already crossed off all the composite numbers less than or equal to $N = 30$. (In general, you can stop the process when the first number remaining — in this example, 7 — is less than $\sqrt{N}$. You might want to think about why this is so.)

---

The Sieve of Eratosthenes is a valid method; unfortunately, it is extremely slow. As $N$ gets larger, the time needed to perform the steps of the Sieve gets infeasibly large very quickly, even if you automate the process on a computer.

Is it possible to find a more efficient method of checking whether a number $N$ is prime? This was an open question for a long time, and was only recently answered by three Indian mathematicians, Agrawal, Kayal, and Saxena, in 2004, who found a brilliant new method of testing primesness. While their method still takes longer for larger numbers, it takes "much less longer": the time needed to perform their algorithm, grows much, much more slowly[3] as a function of $N$ than any other method known (including the Sieve of Eratosthenes). Therefore, for extremely large numbers $N$ (whatever "extremely large" means), the Agrawal/Kayal/Saxena method will be much more efficient.

[2] *Composite* just means "not prime".

[3] The time needed is more or less $\log(N)$, while all other methods are polynomials in $N$.

**2. How are the primes distributed?**

Here are some data about the number of primes that occur in various ranges of integers.

| Range | Number of primes in range | Range | Number of primes |
|---|---|---|---|
| $1 \leq n \leq 100$ | 25 | $10001 \leq n \leq 10100$ | 11 |
| $101 \leq n \leq 100$ | 21 | $100001 \leq n \leq 100100$ | 6 |
| $201 \leq n \leq 100$ | 16 | $1000001 \leq n \leq 1000100$ | 6 |
| $301 \leq n \leq 100$ | 16 | $10000001 \leq n \leq 10000100$ | 2 |
| $401 \leq n \leq 100$ | 17 | $100000001 \leq n \leq 100000100$ | 6 |
| $501 \leq n \leq 100$ | 14 | | |
| $601 \leq n \leq 100$ | 16 | | |
| $701 \leq n \leq 100$ | 14 | | |
| $801 \leq n \leq 100$ | 15 | | |
| $901 \leq n \leq 100$ | 14 | | |

What you can see from this table is that the primes tend to get sparser and sparser among larger and larger numbers — but the rate at which they get sparser is not regular. The distribution of primes is essentially very uneven. That's what makes it hard to attack problems like the following.

**Definition 3.** Two prime numbers are called twin primes if their difference is 2.

For example, 3 and 5 are a pair of twin primes. So are 41 and 43, and 101 and 103, and 809 and 811.

The **twin primes conjecture**[4] states that there are infinitely many twin primes. No one knows if the conjecture is true.

Another famous conjecture along the same lines is the **Goldbach conjecture**, which states that every positive even number $N$ (other than 2) can be written as the sum of two primes. For example, we can write $N = 50$ as the sum of two primes in four different ways:

$$50 \;=\; 3+47 \;=\; 7+43 \;=\; 13+37 \;=\; 19+31.$$

Again, no one knows for sure if the Goldbach conjecture is true. There is evidence to think that it is probably true: not only has it been checked for lots and lots of values of $N$ (up to $10^{18}$), but the number of different ways to express $N$ as the sum of two primes increases (albeit unevenly!) as $N$ itself increases. On the other hand, it is possible that there is some humongous even number out there that slips through the cracks and can't be expressed as the sum of two primes.

Lev Schnirelmann proved the following odd-looking result in 1930:

*Every even number greater than 2 can be written as the sum of at most 300,000 primes.*

This result is far weaker than Goldbach's conjecture, but it is nevertheless of value because it suggests a method of attack: try to refine Schnirelmann's argument to reduce the "300,000" in his proof, and hope to eventually get it down to "2". The state of the art is a theorem proved by Olivier Ramaré in 1995:

*Every even number greater than 2 can be written as the sum of at most six*[5] *primes.*

This is a common pattern in mathematics: A makes a conjecture, B finds a theorem which has the same flavor but is not enough to resolve A's conjecture, C strengthens B's theorem, and so on. Sometimes Z is able to prove (or disprove!) A's conjecture, sometimes not, but this is one way that mathematics advances.

[4] "Conjecture" is math-speak for "educated guess." It's a statement that might be true or false (that is, might or might not be a theorem), but which someone *thinks* is true (but doesn't have a proof for).

[5] In class, I think I said "three", which sounds more dramatic in light of the Goldbach conjecture but is not what Ramaré actually proved.