- $\mathbb{N}=\{1,2,3,\ldots\}$ are the **natural numbers**.
- $\mathbb{Z}=\{0,\pm1,\pm2,\ldots\}$ are the **integers** ("Zahlen").
- $\mathbb{Q}=\left\{\frac{p}{q}:\ p\in\mathbb{Z}, q\in\mathbb{N}\right\}$ are the **rationals**.
- $\mathbb{R}$ are the **reals** (limits of sequences of rationals).
- $\mathbb{C}$ are the **complex numbers**.

Advanced comment. Number theory is also very much concerned with the study of the **algebraic numbers** $\bar{\mathbb{Q}}$, which are those numbers that are the roots of polynomials with integer coefficients. For instance, $\sqrt{5}$ (a root of x^2-5) and i (a root of x^2+1) are examples of simple algebraic numbers (neither of which is rational).

Comment. The numbers π and e are probably the most fundamental mathematical constants, which are not rational. However, we understand the nature of these numbers so little that we do not even know whether $e+\pi$ is rational or not. (Overwhelming evidence suggests that $e+\pi$ is irrational but we do not have a proof.) Isn't that shocking and shameful?!

Example 1. $\sqrt{5}$ is not rational.

Proof. Assume (for contradiction) that we can write $\sqrt{5}=\frac{n}{m}$ with $n,m\in\mathbb{N}$. By canceling common factors, we can ensure that this fraction is reduced.

Then $5m^2=n^2$, from which we conclude that n is divisible by 5. Write $n=5k$ for some $k\in\mathbb{N}$. Then $5m^2=(5k)^2$ implies that $m^2=5k^2$. Hence, m is also divisible by 5. This contradicts the fact that the fraction n/m is reduced. Hence, our initial assumption must have been wrong. □

Variations. Does the same proof apply to, say, $\sqrt{7}$? Which step of the proof fails for $\sqrt{4}$?

1 Divisibility

1.1 Quotients and remainders

Theorem 2. Let $a,b\in\mathbb{Z}$, with $b\neq0$. Then there exist unique integers q and r such that

$$a=qb+r,\quad 0\leqslant r<|b| \qquad \text{(that is, } \frac{a}{b}=q+\frac{r}{b}\text{)}.$$

q is the **quotient**, and r the **remainder** in the division of a by b.

Example 3. For $a=20$, $b=6$, we have $\frac{20}{6}=3+\frac{2}{6}$. That is, $q=3$ and $r=2$.
For $a=20$, $b=5$, we have $\frac{20}{5}=4+\frac{0}{5}$. That is, $q=4$ and $r=0$.

Example 4. When $b=2$, then $r\in\{0,1\}$, and every integer is either of the form $2q$ or of the form $2q+1$. We call numbers **even** or **odd** correspondingly.

Example 5. Show that the square of an integer leaves the remainder 0 or 1 upon division by 4.

That is, none of the squares $1,4,9,16,25,36,\ldots$ leave remainder 2 or 3 when dividing by 4!!

Proof. Every integer is of the form $2q$ or $2q+1$. Upon division by 4, $(2q)^2=4q^2$ leaves remainder 0, $(2q+1)^2=4q^2+4q+1$ leaves remainder 1.

Example 6. Show that the square of an integer leaves the remainder 0 or 1 upon division by 3.

Proof. Every integer is of the form $3q$, $3q+1$ or $3q+2$. Upon division by 3, $(3q)^2=9q^2$ leaves remainder 0, while both $(3q+1)^2=9q^2+6q+1$ and $(3q+2)^2=9q^2+12q+4$ leave remainder 1.

Armin Straub
straub@southalabama.edu

1.2 Greatest common divisor

Definition 7. Let $a, b \in \mathbb{Z}$ and $a \neq 0$. We write $a|b$ (and say b is **divisible** by a) if $\frac{b}{a} \in \mathbb{Z}$.

In other words, $a|b$ if and only if there exists an integer c such that $ac = b$.

Example 8. $3|9$ but $3 \nmid 10$.

Definition 9. Let $a, b \in \mathbb{Z}$ (not both zero). The **greatest common divisor** $\gcd(a, b)$ of a and b is the largest positive integer c such that $c|a$ and $c|b$.

If $\gcd(a, b) = 1$ then we say that a and b are **coprime** (to each other).

Example 10.

(a) $\gcd(2, 4) = 2$

(b) $\gcd(15, 28) = 1$

(c) $\gcd(30, 108) = \gcd(2 \cdot 3 \cdot 5, 2^2 \cdot 3^3) = 6$

(d) $\gcd(60, 2020) = \gcd(2^2 \cdot 3 \cdot 5, 2^2 \cdot 5 \cdot 101) = 2^2 \cdot 5 = 20$

BAD?! Computing $\gcd(a, b)$ by factoring a and b is not a good approach. Though small numbers might be easy to factor, it is very hard to factor even moderately large numbers in general.

Next class, we will learn about a good way to compute the gcd, which works well even for enormous numbers (in particular, it avoids factorizing the involved numbers).

Indeed, in 1991, RSA Laboratories challenged everyone to factor several numbers including:

135066410865995223349603216278805969938881475605667027524485143851526 5\
106048595338339402871505719094417982072821644715513736804197039641917 4\
304649658927425623934102086438320211037295872576235850964311056407350 1\
508187510676594629205563685529475213500852879416377328533906109750544 3\
3499981115005697723689092756 3

Since then, nobody has been able to factor this 1024 bit number (309 decimal digits). Until 2007, cash prizes were offered up to 200,000 USD, with 100,000 USD for the number above (20,000 USD collected in 2005 for factoring a number with 193 decimal digits; 232 decimal digits factored in 2009, larger ones remain unfactored; largest one has 617 decimal digits). The reason people are very interested in factoring is that the difficulty of factoring is actually crucially used in many cryptosystems, including RSA.

https://en.wikipedia.org/wiki/RSA_Factoring_Challenge

Lemma 11. If $a = qb + r$, then $\gcd(a,b) = \gcd(b,r)$.

Proof. Let $d \in \mathbb{N}$. We need to show that $d|a$ and $d|b$ iff $d|r$ and $d|b$. [**iff** is short for "if and only if"]
Equivalently, assuming that $d|b$, we need to show that $d|a$ iff $d|r$.
Indeed, it follows from $\frac{a}{d} = \frac{qb+r}{d} = \frac{qb}{d} + \frac{r}{d}$ that $\frac{a}{d} \in \mathbb{Z}$ iff $\frac{r}{d} \in \mathbb{Z}$. □

Example 12. Using this lemma to compute gcd's is referred to as the **Euclidean algorithm**.

(a) $\underbrace{\gcd(30,108)}_{108=3\cdot 30+18} = \underbrace{\gcd(18,30)}_{30=1\cdot 18+12} = \underbrace{\gcd(12,18)}_{18=1\cdot 12+6} = \underbrace{\gcd(6,12)}_{12=2\cdot 6+0} = 6$

Alternatively, taking a shortcut by allowing negative remainders:

$\underbrace{\gcd(30,108)}_{108=4\cdot 30-12} = \underbrace{\gcd(12,30)}_{30=2\cdot 12+6} = \underbrace{\gcd(6,12)}_{12=2\cdot 6+0} = 6$

(b) $\underbrace{\gcd(16,25)}_{25=1\cdot 16+9} = \underbrace{\gcd(9,16)}_{16=1\cdot 9+7} = \underbrace{\gcd(7,9)}_{9=1\cdot 7+2} = \underbrace{\gcd(2,7)}_{7=3\cdot 2+1} = \gcd(1,2) = 1$

Alternatively, again, taking a shortcut by allowing negative remainders:

$\underbrace{\gcd(16,25)}_{25=2\cdot 16-7} = \underbrace{\gcd(7,16)}_{16=2\cdot 7+2} = \underbrace{\gcd(2,7)}_{7=3\cdot 2+1} = \gcd(1,2) = 1$

Theorem 13. (Bézout's identity) Let $a, b \in \mathbb{Z}$ (not both zero). There exist $x, y \in \mathbb{Z}$ such that

$$\gcd(a,b) = ax + by.$$

Proof. We proceed iteratively:

$$\begin{aligned}
a &= q_1 b + r_1, \quad 0 < r_1 < b \\
b &= q_2 r_1 + r_2, \quad 0 < r_2 < r_1 \\
r_1 &= q_3 r_2 + r_3, \quad 0 < r_3 < r_2 \\
&\vdots \\
r_{n-3} &= q_{n-1} r_{n-2} + r_{n-1}, \quad 0 < r_{n-1} < r_{n-2} \\
r_{n-2} &= q_n r_{n-1} + r_n, \quad 0 < r_n < r_{n-1} \\
r_{n-1} &= q_{n+1} r_n + 0
\end{aligned}$$

Along the way, we have $\gcd(a,b) = \gcd(b,r_1) = \gcd(r_1,r_2) = \ldots = \gcd(r_{n-2},r_{n-1}) = \gcd(r_{n-1},r_n) = r_n$ (why is it obvious that the last gcd is r_n?).
By the second-to-last equation, $\gcd(a,b) = r_n = r_{n-2} - q_n r_{n-1}$ is a linear combination of r_{n-2} and r_{n-1}. Then, moving one up, we replace r_{n-1} with $r_{n-3} - q_{n-1} r_{n-2}$ to write $\gcd(a,b)$ as a linear combination of r_{n-3} and r_{n-2}. Continuing in that fashion, we ultimately obtain $\gcd(a,b)$ as a linear combination of a and b. □

Let us revisit the previous example to illustrate how the Euclidean algorithm provides us with a way to write $\gcd(a,b)$ as an integer linear combination of a and b.

Armin Straub
straub@southalabama.edu

Example 14. Find $d=\gcd(30,108)$ as well as integers r,s such that $d=30r+108s$.

Solution. We apply the extended Euclidean algorithm:

$$\begin{aligned}&\gcd(30,108) && \boxed{108}=4\cdot\boxed{30}-12 && \text{or:} && \boxed{A}\ \ 12=-1\cdot\boxed{108}+4\cdot\boxed{30}\\ &=\gcd(12,30) && \boxed{30}=2\cdot\boxed{12}+6 && && \boxed{B}\ \ 6=1\cdot\boxed{30}-2\cdot\boxed{12}\\ &=\gcd(6,12) && \boxed{12}=2\cdot\boxed{6}+0\\ &=6\end{aligned}$$

Backtracking through this, we find that **Bézout's identity** takes the form

$$6 \underset{\boxed{B}}{=} 1\cdot\boxed{30}-2\cdot\boxed{12} \underset{\boxed{A}}{=} 1\cdot\boxed{30}-2(-1\cdot\boxed{108}+4\cdot\boxed{30})=-7\cdot\boxed{30}+2\cdot\boxed{108}$$

In summary, we have $-7\cdot 30+2\cdot 108=6$.

Example 15. Find $d=\gcd(16,25)$ as well as integers r,s such that $d=16r+25s$.

Solution. We apply the extended Euclidean algorithm:

$$\begin{aligned}&\gcd(16,25) && \boxed{25}=2\cdot\boxed{16}-7 && \text{or:} && \boxed{A}\ \ 7=-1\cdot\boxed{25}+2\cdot\boxed{16}\\ &=\gcd(7,16) && \boxed{16}=2\cdot\boxed{7}+2 && && \boxed{B}\ \ 2=1\cdot\boxed{16}-2\cdot\boxed{7}\\ &=\gcd(2,7) && \boxed{7}=3\cdot\boxed{2}+1 && && \boxed{C}\ \ 1=\boxed{7}-3\cdot\boxed{2}\\ &=1\end{aligned}$$

Backtracking through this, we find that **Bézout's identity** takes the form

$$1 \underset{\boxed{C}}{=} \boxed{7}-3\cdot\boxed{2} \underset{\boxed{B}}{=} 7\cdot\boxed{7}-3\cdot\boxed{16} \underset{\boxed{A}}{=} -7\cdot\boxed{25}+11\cdot\boxed{16}$$

In summary, we have $-7\cdot 25+11\cdot 16=1$.

Example 16. (extra) Find $d=\gcd(17,23)$ as well as integers r,s such that $d=17r+23s$.

Solution. We apply the extended Euclidean algorithm:

$$\begin{aligned}&\gcd(17,23) && \boxed{23}=1\cdot\boxed{17}+6 && \text{or:} && \boxed{A}\ \ 6=1\cdot\boxed{23}-1\cdot\boxed{17}\\ &=\gcd(6,17) && \boxed{17}=3\cdot\boxed{6}-1 && && \boxed{B}\ \ 1=-1\cdot\boxed{17}+3\cdot\boxed{6}\\ &=1\end{aligned}$$

Backtracking through this, we find that **Bézout's identity** takes the form

$$1 \underset{\boxed{B}}{=} -1\cdot\boxed{17}+3\cdot\boxed{6} \underset{\boxed{A}}{=} -4\cdot\boxed{17}+3\cdot\boxed{23}$$

In summary, we have $1=-4\cdot 17+3\cdot 23$.

2 Primes

Lemma 17. (Euclid's lemma) If $d|ab$ and $\gcd(a,d)=1$, then $d|b$.

Proof. Since $(a,d)=1$, we can find x,y so that $ax+dy=1$.
We then see that $b=abx+bdy$ is divisible by d (because $d|ab$). □

Definition 18. An integer $p>1$ is a **prime** if its only positive divisors are 1 and p.

Lemma 19. If p is a prime and $p|ab$, then $p|a$ or $p|b$.

Proof. If $p|a$, then we are done. Otherwise, $p\nmid a$. In that case, $\gcd(a,p)=1$ because the only positive divisors of p are 1 and p. Our claim therefore is a special case of the previous lemma. □

Corollary 20. If p is a prime and $p|a_1a_2\cdots a_r$, then $p|a_k$ for some $k\in\{1,2,\ldots,r\}$.

Example 21. This property is unique to primes. For instance, $6|8\cdot 21$ but $6\nmid 8$ and $6\nmid 21$.

Whereas, $2|8\cdot 21$ and, indeed $2|8$. Similarly, $3|8\cdot 21$ and, indeed $3|21$.

Theorem 22. (Fundamental Theorem of Arithmetic) Every integer $n>1$ can be written as a product of primes. This factorization is unique (apart from the order of the factors).

Proof. Let us first prove, by (strong) induction, that every integer $n>1$ can be written as a product of primes.

- **(base case)** $n=2$ is a prime. There is nothing to do.
- **(induction step)** Suppose that we already know that all integers less than n can be written as a product of primes. We need to show that n can be written as a product of primes, too.

  Let $d>1$ be the smallest divisor of n. Then d is necessarily a prime (because if $a>1$ divides d, then a also divides n so that $a=d$ because d is the smallest number dividing n).

  If $d=n$, then n is a prime, and we are already done.

  Otherwise, $\frac{n}{d}>1$ is an integer, which, by the induction hypothesis, can be written as the product of some primes $p_1\cdots p_r$. Then, $n=dp_1\cdots p_r$.

Finally, let us think about why this factorization is unique. Suppose we have two factorizations

$$n=p_1p_2\cdots p_r=q_1q_2\cdots q_s.$$

By the corollary, each p_i divides one of the q_j's (and vice versa), in which case $p_i=q_j$, so we can cancel common factors until we see that both factorizations are identical. □

Advanced comment. The idea of factorization into primes and the uniqueness of such factorizations should not be taken entirely for granted. For instance, when instead of integers $a\in\mathbb{Z}$ we work with "generalized integers" such as $a+bi\sqrt{5}$, with $a,b\in\mathbb{Z}$, then factorization is not unique: for example, we have two different factorizations of 6, namely,

$$6=2\cdot 3,\quad 6=(1+i\sqrt{5})(1-i\sqrt{5}),$$

and each of the numbers 2, 3, $1\pm i\sqrt{5}$ cannot be factored further.

Example 23. $140=2^2\cdot 5\cdot 7$, $2016=2^5\cdot 3^2\cdot 7$, 2017 is a prime, $2018=2\cdot 1009$, $2019=3\cdot 673$, $2020=2^2\cdot 5\cdot 101$.

How can we check that 2017 is indeed prime? Well, none of the small primes $2,3,5,7,11$ divide 2017. But how far do we need to check? Since $\sqrt{2017}\approx 44.91$, we only need to check up to prime 43. (Why?!)

Armin Straub
straub@southalabama.edu

Definition 24. Let $a, b \in \mathbb{Z}$ (both not zero). The **least common multiple** $\operatorname{lcm}(a,b)$ of a and b is the smallest positive integer m such that $a|m$ and $b|m$.

Example 25. $\operatorname{lcm}(12,42) = \operatorname{lcm}(2^2 \cdot 3, 2 \cdot 3 \cdot 7) = 2^2 \cdot 3 \cdot 7 = 84 = \frac{12 \cdot 42}{6}$

Lemma 26. For $a, b \in \mathbb{N}$, $\operatorname{lcm}(a,b) = \frac{ab}{\gcd(a,b)}$.

Proof. Write $d = \gcd(a,b)$ and $m = \frac{ab}{d}$. Note that $a|m$ because $\frac{m}{a} = \frac{b}{d}$ is an integer. Likewise, $b|m$. In other words, m is a common multiple of a and b. We still need to show that it is the smallest.

Let n be a positive integer such that $a|n$ and $b|n$. (We need to show that $m \leqslant n$. We do that by showing $m|n$.)

Recall that $d = ax + by$ for some integers x, y. Using that, we find that

$$\frac{n}{m} = \frac{nd}{ab} = \frac{n(ax+by)}{ab} = \frac{n}{b}x + \frac{n}{a}y$$

is an integer. That is, $m|n$. □

3 More on primes

Example 27. The **sieve of Eratosthenes** is an efficient way to find all primes up to some n.

Write down all numbers $2, 3, 4, \ldots, n$. We begin with 2 as our first prime. We proceed by crossing out all multiples of 2, because these are not primes. The smallest number we didn't cross out is 3, our next prime. We again proceed by crossing out all multiples of 3, because these are not primes. The smallest number we didn't cross out is 5 (note that it has to be prime because, by construction, it is not divisible by any prime less than itself).

Problem. If $n = 10^6$, at which point can we stop crossing out numbers?

We can stop when our "new prime" exceeds $\sqrt{n} = 1000$. All remaining numbers have to be primes. Why?!

Example 28. (Euclid) There are infinitely many primes.

Proof. Assume (for contradiction) there are only finitely many primes: $p_1, p_2, \ldots, p_n$.
Consider the number $N = p_1 \cdot p_2 \cdot \ldots \cdot p_n + 1$.
Each prime p_i divides $N - 1$ and so p_i does not divide N.
Thus any prime dividing N is not on our list. Contradiction. □

Historical note. This is not necessarily a proof by contradiction, and Euclid (300BC) himself didn't state it as such. Instead, one can think of it as a constructive machinery of producing more primes, starting from any finite collection of primes.

A variation. Can we replace $N = p_1 \cdot p_2 \cdot \ldots \cdot p_n + 1$ in the proof with $N = p_1 \cdot p_2 \cdot \ldots \cdot p_n - 1$? Yes! (If $n \geqslant 2$.)

Playing with numbers.

$2 + 1 = 3$ is prime. $2 \cdot 3 + 1 = 7$ is prime. $2 \cdot 3 \cdot 5 + 1 = 31$ is prime. $2 \cdot 3 \cdot 5 \cdot 7 + 1 = 211$ is prime. $2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 = 2311$ is prime. $2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 = 30031 = 59 \cdot 509$ is not prime.

Let $P_n = p_1 \cdot p_2 \cdot \ldots \cdot p_n + 1$ where p_i is the ith prime. If P_n is prime, it is called a primorial prime. We have just checked that P_1, P_2, P_3, P_4, P_5 are primes but that P_6 is not a prime.

The next primorial primes are $P_{11}, P_{75}, P_{171}, P_{172}$. It is not known whether there are infinitely P_n which are prime.

More shamefully, it is not known whether there are infinitely many P_n which are not prime.

See, for instance: http://mathworld.wolfram.com/PrimorialPrime.html

Example 29. In 12/2018, a new largest (proven) prime was found: $2^{82,589,933}-1$.

https://www.mersenne.org/primes/?press=M82589933

This is a **Mersenne prime** (like the last 17 record primes). It has a bit over 24.8 million (decimal) digits (versus 23.2 for the previous record). The prime was found as part of GIMPS (Great Internet Mersenne Prime Search), which offers a $3,000 award for each new Mersenne prime discovered.

The EFF (Electronic Frontier Foundation) is offering $150,000 (donated anonymously for that specific purpose) for the discovery of the first prime with at least 100 million decimal digits.

https://www.eff.org/awards/coop

[Prizes of $50,000 and $100,000 for primes with 1 and 10 million digits have been claimed in 2000 and 2009.]

Example 30. $(p, p+2)$ is a twin prime pair if both p and $p+2$ are primes.

Just making sure. $(2,3)$ is the only pair $(p, p+1)$ with p and $p+1$ both prime. (Why?!)

Some twin prime pairs. $(3,5)$, $(5,7)$, $(11,13)$, $(17,19)$, $(29,31)$, $(41,43)$, $(59,61)$, $(71,73)$, $(101,103)$, ...

Largest known one: $2996863034895 \cdot 2^{1290000} \pm 1$ ($388,342$ decimal digits; found 2016)

Twin prime conjecture. Euclid already conjectured in 300 BCE that there are infinitely many twin primes. Despite much effort, noone has been able to prove that in more than 20 centuries.

Recent progress. It is now known that there are infinitely many pairs of primes (p_1, p_2) such that the gap between p_1 and p_2 is at most 246 (the break-through in 2013 due to Yitang Zhang had $7 \cdot 10^7$ instead of 246).

Example 31. (Bertrand's postulate) For any $n > 1$, the interval $(n, 2n)$ contains at least one prime.

Advanced comment. Let $\pi(x)$ be the number of primes $\leqslant x$. It follows from Betrand's postulate that $\pi(2^n) \geqslant n$.

To prove that, note that 2 is a prime and that each of the (disjoint!) intervals $(2,4)$, $(4,8)$, $(8,16)$, ..., $(2^{n-1}, 2^n)$ contains at least one prime.

This is a very poor bound. For instance, we find $\pi(2^{20}) \geqslant 20$ where 2^{20} is a little bigger than 10^6. Compare that to the actual numbers in the prime number theorem below.

Historical comment. This was conjectured by Bertrand in 1845 (he checked up to $n = 3 \cdot 10^6$), and proved by Chebyshev in 1852.

The following famous and deep result quantifies the infinitude of primes.

Theorem 32. (prime number theorem) Let $\pi(x)$ be the number of primes $\leqslant x$. Then

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

In other words: Up to x, there are roughly $x/\ln(x)$ many primes.

Examples.

proportion of primes up to 10^6: $\frac{78,498}{10^6} = 7.85\%$ vs the estimate $\frac{1}{\ln(10^6)} = \frac{1}{6\ln(10)} = 7.24\%$

proportion of primes up to 10^{12}: $\frac{37,607,912,018}{10^{12}} = 3.76\%$ vs the estimate $\frac{1}{\ln(10^{12})} = \frac{1}{12\ln(10)} = 3.62\%$

An example of huge relevance for crypto. Many cryptographic schemes require us to be able to generate large random primes, where large typically means numbers with about 2048 binary digits.

By the PNT, the proportion of primes up to 2^{2048} is about $\frac{1}{\ln(2^{2048})} = 0.0704\%$.

That means, roughly, 1 in 1500 numbers of this magnitude are prime. That means we (i.e. our computer) can efficiently generate large random primes by just repeatedly generating large random numbers and discarding those that are not prime (we will discuss primality testing in cryptography).

Comment. Here, $\ln(x)$ is the logarithm with base e. Isn't it wonderful how Euler's number $e \approx 2.71828$ is sneaking up on the primes?

Historical comment. Despite progress by Chebyshev (who succeeded in 1852 in showing that the quotient in the above limit is bounded, for large x, by constants close to 1), the PNT was not proved until 1896 by Hadamard and, independently, de la Vallée Poussin, who both used new ideas due to Riemann.