Part I

Elementary Arithmetic of Conics

1. Starting with the Unit Circle

The title of this chapter is stolen from the book [Ken1981] by Hua Loo Keng. I remember that I was quite disappointed when I first picked it up because the title made me expect a book on number theory, not on analysis. Let me add that Hua Loo Keng's book (like any of his other books) is a perfectly fine and beautiful work – it's just not number theory.

In this chapter we will discuss integral solutions of the Pythagorean equation, which is a classical result and the main tool in Fermat's proof of his Last Theorem for exponent 4. We will also introduce congruences and residue classes, as well as the Euclidean algorithm and the concept of unique factorization.

1.1. Some Diophantine Problems

Where we learn how to look at arithmetic problems from a geometric point of view.

One of the oldest diophantine problems is the Pythagorean equation[1]

$$x^2 + y^2 = z^2. \tag{1.1}$$

Already the Pythagoreans must have known that there are infinitely many solutions, some of which can be found in the writings of Plato and the Elements of Euclid.

The most famous integral solution of (1.1) is certainly the classical Pythagorean triple $(x, y, z) = (3, 4, 5)$. This is the smallest solution coming from the infinite family

$$x = n^2 - 1, \quad y = 2n, \quad z = n^2 + 1. \tag{1.2}$$

The second smallest solution $(x, y, z) = (5, 12, 13)$ shows that not all integral solutions come from (1.2). Nevertheless these formulas contain the germs of the general solution. By allowing $n = \frac{a}{b}$ to attain rational values we get the set of rational solutions $(a^2/b^2 - 1, 2a/b, a^2/b^2 + 1)$ of the Pythagorean equation. Since this equation is homogeneous (all terms have the same degree, here they are all quadratic) we may multiply any rational solution by its common denominator and produce integral solutions. In this way, our set of rational solutions above produces the integral solutions

$$x = a^2 - b^2, \quad y = 2ab, \quad z = a^2 + b^2. \tag{1.3}$$

The way we have arrived at our solutions does not tell us whether we have found all of them; we will come back to this problem later on.

Over the last centuries, many methods have been invented for solving this and other diophantine equations. One such method is the following. If we divide (1.1) through by z and set $X = \frac{x}{z}$, $Y = \frac{y}{z}$, then $X^2 + Y^2 = 1$, and (X, Y) is a **rational point** on the unit circle, that is, a point (X, Y) with rational coordinates X and Y.

[1] Isn't this a beautiful anachronism? Pythagoras lived more than eight centuries before Diophantus.

But not only does every nonzero integral solution of (1.1) give a rational point on the unit circle, we also can easily convince ourselves that every rational point on the unit circle produces an integral solution of (1.1): all we have to do is write the coordinates using a common denominator z; the rational point $(\frac{x}{z}, \frac{y}{z})$ then gives us a solution (x, y, z) of (1.1).

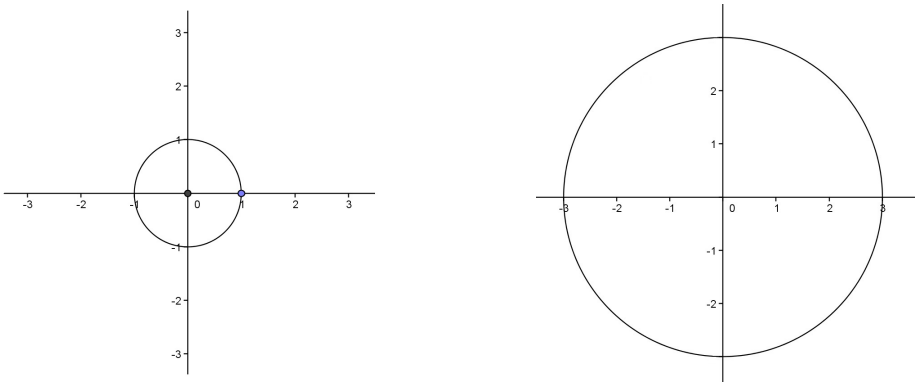


Fig. 1.1. The circles $X^2 + Y^2 = 1$ and $X^2 + Y^2 = 3$

It is not clear at all at this point whether the geometric interpretation that we have given to our problem is going to help us. In fact, the pictures of the unit circle and that of $X^2 + Y^2 = 3$ (see Fig. 1.1) do not look any different except for size, yet the unit circle has infinitely many rational points, whereas there is not a single rational point on $\mathcal{C}_3 : X^2 + Y^2 = 3$. This may be proved as follows: assume that $(\frac{x}{z}, \frac{y}{z})$ is a rational point on $\mathcal{C}_3$. Then $x^2 + y^2 = 3z^2$. We claim that at least one of x or y must be divisible by 3. If not, we can write $x = 3a \pm 1$ and $y = 3b \pm 1$ and find

$$x^2 + y^2 = 9a^2 \pm 6a + 1 + 9b^2 \pm 6b + 1 = 3(3a^2 \pm 2a + 3b^2 \pm 2b) + 2$$

contradicting the fact that $x^2 + y^2 = 3z^2$ is divisible by 3.

But if y is divisible by 3, say $y = 3u$, then $x^2 = 3z^2 - 9u^2$ is divisible by 3, and we conclude: if $x^2 + y^2 = 3z^2$ has a solution in integers, then x and y are multiples of 3.

Now we can finish off our proof by invoking the method of descent. Assume that (x, y, z) is a solution of $x^2 + y^2 = 3z^2$ with $x, y, z > 0$. Then $x = 3x_1$ and $y = 3y_1$ are multiples of 3, hence $3z^2 = 9(x_1^2 + y_1^2)$ is divisible by 9, which implies that $z = 3z_1$ is a multiple of 3. Thus we have found a smaller solution (x_1, y_1, z_1) of $x^2+y^2 = 3z^2$ in positive integers. This means that if there is a positive solution (x, y, z), then there is a smaller solution (x_1, y_1, z_1); by the same argument, there is an even smaller solution (x_2, y_2, z_2), and continuing this way we run into a contradiction since positive integral solutions cannot become smaller and smaller.

Fact 1.1. *The circle $X^2 + Y^2 = 3$ does not contain any rational points.*

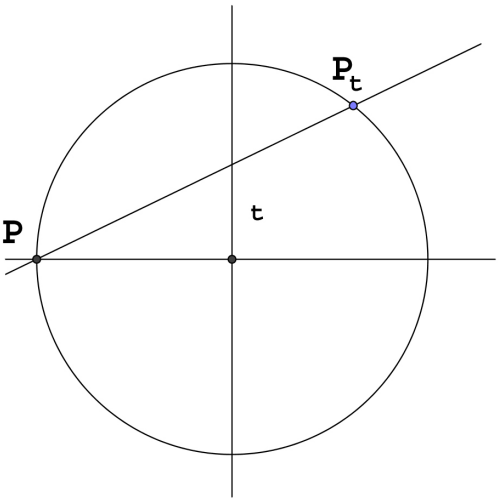
The argument using the numbers $3a \pm 1$ and their squares etc. will be formalized below using the notion of congruences. This argument can also be used for proving that $3 \mid z^2$ implies $3 \mid z$ (which we did use in our proof), and this observation will also be generalized to "Euclid's Lemma" in Prop. 1.10 below.

The main difference between the unit circle and $\mathcal{C}_3$, as we will see, is that the unit circle has some "obvious" rational points, such as $(\pm 1, 0)$ and $(0, \pm 1)$. What we will see below is that the existence of a single rational point immediately implies that there are infinitely many!

Although we have started with a circle, let us remark right away that, over the rationals, a circle like $X^2+Y^2=3$ can be "transformed" into a hyperbola as follows: Writing $X=\frac{x}{z}$, $Y=\frac{y}{z}$, the circle becomes $x^2+y^2=3z^2$. Dividing through by y and setting $T=\frac{x}{y}$, $U=\frac{z}{y}$ we end up with the equation $T^2-3U^2=-1$ (note that the points $(\pm1,0)$ have no image under this map. The transformation of the circle into a hyperbola can be explained painlessly using some projective geometry – we will come back to this in due time.

For the moment we will exploit this connection between the circle and the hyperbola for giving a second derivation of the formulas (1.3). Start from the unit circle $X^2+Y^2=1$; writing $X=\frac{x}{z}$ and $Y=\frac{y}{z}$ we get $x^2+y^2=z^2$, hence $y^2=z^2-x^2=(z-x)(z+x)$ and finally $1=\frac{z-x}{y}\cdot\frac{z+x}{y}$. Since the product on the right hand side equals 1, we may set the first factor equal to t and the second to $\frac{1}{t}$ for nonzero rational numbers $t\in\mathbb{Q}^\times$. From $\frac{z-x}{y}=t$ and $\frac{z+x}{y}=\frac{1}{t}$ we get $\frac{1}{Y}=\frac{z}{y}=\frac{1}{2}(t+\frac{1}{t})$ or $Y=\frac{2t}{t^2+1}$. Similarly we get $\frac{x}{y}=\frac{1}{2}(t-\frac{1}{t})=\frac{t^2-1}{2t}$, hence $X=\frac{x}{z}=\frac{x}{y}\cdot\frac{y}{z}=\frac{t^2-1}{t^2+1}$.

The Geometric Method: Parametrization



In order to find all rational points on the unit circle $\mathcal{C}$ we start with an obvious solution, say $P=(-1,0)$ (any rational point on $\mathcal{C}$ would do). A line through P with rational slope t will intersect the circle $\mathcal{C}$ in two points, namely P and one other point, say P_t. It is easy to see that P_t is a rational point if t is rational, and that every rational point $\neq P$ on $\mathcal{C}$ is one of these points P_t: in fact, if $Q=(x,y)\neq P$ is a rational point on $\mathcal{C}$, then $t=\frac{y}{x+1}$ (the slope of the line PQ) gives $Q=P_t$.

The actual calculation is simple: intersecting the line $Y=t(X+1)$ with the unit circle $\mathcal{C}: X^2+Y^2=1$ gives $X^2-1+t^2(X+1)^2=0$. This equation has the obvious solution $x=1$, and so it can be written in the form $(X+1)(X-1+t^2(X+1))=0$. Setting the second bracket equal to 0 gives the second solution $x=\frac{1-t^2}{1+t^2}$, and plugging this into the line equation gives the formulas

$$x=\frac{1-t^2}{1+t^2},\quad y=\frac{2t}{1+t^2}. \tag{1.4}$$

Thus we have proved

Proposition 1.2. *The rational points $P=(x,y)$ different from $(-1,0)$ are those given in (1.4), where t runs through all the rational numbers.*

The peculiar role of the point $(-1,0)$ disappears by allowing t to attain the value ∞; this will be achieved in a very natural way by switching to the projective point of view.

The Method of Diophantus

Diophantus solved a problem closely related to that of finding Pythagorean triples; it is problem 8 in Book II:

Problem 1. *To divide a given square into a sum of two squares.*

This problem asks for a solution of the equation $a^2 = x^2 + y^2$ in positive rational numbers $x, y \in \mathbb{Q}$, where $a \in \mathbb{Q}$ is given. Diophantus did not have the modern algebraic notation at his command, and so he begins by assuming that $a^4 = 16$. He calls the first square x^2 (he had symbols for one unknown and its powers, but not for two); then $16 - x^2$ is the second square. He wants $16 - x^2$ to be a square. To this end he forms the square of the difference of an arbitrary multiple of x diminished by the root of 16, that is, he considers $2x - 4$. Its square should be equal to $16 - x^2$, that is, $(2x-4)^2 = 4x^2 - 16x + 16 = 16 - x^2$. Since the constant term cancels, we find $5x^2 = 16x$, hence $x = \frac{16}{5}$. Thus $4^2 = (\frac{16}{5})^2 + (\frac{12}{5})^2$.

For us it is very easy to treat the general case: starting from $a^2 = x^2 + y^2$, we get $y^2 = a^2 - x^2$. We now substitute $y = tx - a$: then the constant terms will cancel, and we can solve for x. In fact, we find $(t^2+1)x = 2at$, thus $x = \frac{2at}{t^2+1}$ and $y = tx - a$, and finally

$$a^2 = \Big(a\frac{2t}{t^2+1}\Big)^2 + \Big(a\frac{t^2-1}{t^2+1}\Big)^2.$$

Dividing the last equation through by a we find

$$1 = \Big(\frac{2t}{t^2+1}\Big)^2 + \Big(\frac{t^2-1}{t^2+1}\Big)^2. \tag{1.5}$$

Looking carefully at the method with which we have obtained (1.5) we cannot help but notice that even the solution may be interpreted geometrically: the substitution $y = tx - a$ we have used above is the equation of a line with slope t through the point $(0, -a)$. In the case $a = 1$ of the unit circle, this line goes through the point $(0, -1)$ on the unit circle, and what we have done above was intersecting this line with the unit circle. The resulting quadratic equation had the solution $x = 0$, giving us the point $(0, -1)$ we started with, and the solution $x = \frac{2t}{t^2+1}$, giving the rational parametrization of the unit circle we have found above.

For what it's worth, here's a similar problem (Problem 9 from the second book) from Diophantus:

Problem 2. *To divide a given number which is the sum of two squares in two other squares.*

Again, Diophantus works with explicitly given numbers; in this case he starts with $13 = 2^2 + 3^2$ and sets $3^2 + 2^2 = (x+2)^2 + (mx-3)^2$. Setting $m = 2$ provides him with

$$3^2 + 2^2 = (x+2)^2 + (2x-3)^2 = 5x^2 - 8x + 2^2 + 3^2,$$

which is equivalent to $5x^2 - 8x = 0$. Thus $x = \frac{8}{5}$, and we find

$$2^2 + 3^2 = \Big(\frac{18}{5}\Big)^2 + \Big(\frac{1}{5}\Big)^2.$$

Can you extract a parametric solution for the general problem $a^2 + b^2 = x^2 + y^2$ from Diophantus' solution?

The Analytic Method: Trigonometry

It is well known that every real point (x, y) on the unit circle can be written in the form

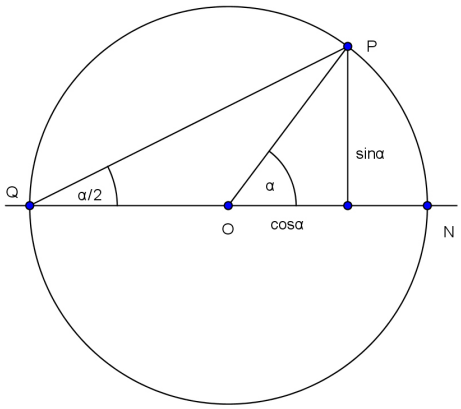
$$x = \cos\alpha, \quad y = \sin\alpha$$

for some real number $\alpha \in [0, 2\pi)$.

In general, these points will not be rational. For finding the values of α that produce rational numbers, consider the adjacent diagram. We have already seen above that the point P is rational if and only if the slope of line PQ is rational. This slope is given by $t = \tan(\sphericalangle PQN)$. Since $\sphericalangle POQ = 180^\circ - \alpha$ and since $\overline{PO} = \overline{QO}$, we have

$$\sphericalangle PQO = \frac{1}{2}(180^\circ - (180^\circ - \alpha)) = \frac{1}{2}\alpha.$$

(Observe that this relation is actually a generalization of the Theorem of Thales on right angles in semi-circles.) Thus P is rational if and only if $t = \tan\frac{\alpha}{2}$ is rational.



We also observe that if $\tan\frac{\alpha}{2}$ is rational, then the associated point $P(\cos\alpha, \sin\alpha)$ has rational coordinates, which in turn implies that $\tan\alpha$ is rational. This suggests the existence of a rational relation between $\tan\frac{\alpha}{2}$ and $\tan\alpha$, and in fact we have $\tan\alpha = \frac{2\tan\frac{\alpha}{2}}{1-\tan^2\frac{\alpha}{2}}$, as can be seen from (1.12).

We have proved

Lemma 1.3. *The point* $(\cos\alpha, \sin\alpha)$ *on the unit circle is rational if and only if* $\tan\frac{\alpha}{2}$ *is rational.*

In order to extract a rational parametrization from the analytic parametrization of the unit circle we replace the trigonometric functions of α by trigonometric functions of $\frac{\alpha}{2}$. This is accomplished by using the identities $\cos^2\alpha - \sin^2\alpha = \cos 2\alpha$ and $\cos^2\alpha + \sin^2\alpha = 1$. Now we find

$$\begin{aligned} x = \cos\alpha &= \frac{\cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}}{\cos^2\frac{\alpha}{2} + \sin^2\frac{\alpha}{2}} = \frac{1-t^2}{1+t^2}, \\ y = \sin\alpha &= \frac{2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}}{\cos^2\frac{\alpha}{2} + \sin^2\frac{\alpha}{2}} = \frac{2t}{1+t^2}, \end{aligned}$$

where we have put $t = \tan\frac{\alpha}{2}$. Every rational number t now gives us a rational point on the unit circle. Conversely, if x and $y \neq 0$ are rational, then so is $t = \frac{1-x}{y}$. Since the point $(1,0)$ corresponds to $t = 0$ (although $t = \frac{1-x}{y}$ does not make any sense here), this parametrization gives us all rational points $\neq (-1,0)$ on $\mathcal{C}$.

Substitutions for Computing Integrals

Assume that you have to find the integral of a rational function of $\sin x$ and $\cos x$, say $\int \frac{P(\sin x,\cos x)}{Q(\sin x,\cos x)}\,dx$, where $P, Q \in \mathbb{R}[X,Y]$ are polynomials in two variables X, Y with real coefficients. Then we claim that the substitution $t = \tan\frac{x}{2}$ transforms this into an integral of a rational function.

In fact, $t = \tan\frac{x}{2}$ implies, as we have seen, that $\sin x = \frac{2t}{1+t^2}$ and $\cos x = \frac{1-t^2}{1+t^2}$; in addition we see that $\frac{dt}{dx} = \frac{1}{2}(1+t^2)$, i.e., $dx = \frac{2\,dt}{1+x^2}$. This implies the claim. The resulting rational function of t can be integrated using partial fraction decomposition.

Classifying Solutions

We have seen that there are various methods for deriving the formulas giving a rational parametrization of the unit circle. We have also seen that every rational value of t gives us a rational point, and that every rational point on the unit circle has the form (1.4) for some rational t.

What we have not yet addressed is the following question, which at first perhaps may make no sense at all: how can we uniquely represent the rational numbers t? We know from our school days that every rational number t can be written in the form $t = \frac{a}{b}$, where a is an integer and b a nonzero natural number. But are these integers a and b uniquely determined by t? The answer has been known for thousands of years: fractions can be written in lowest terms, that is, given a fraction $\frac{a}{b}$ there are integers e, f with $\frac{a}{b} = \frac{e}{f}$ such that e and f have no factor in common. It turns out that these are also the minimal numbers with the property $\frac{a}{b} = \frac{e}{f}$. As Euclid knew, the claim that fractions can be written in lowest terms requires a proof.

We will prove these claims in the next section, and hope to make clear in subsequent chapters that statements of this kind actually do require a proof. To have seen that such propositions require a proof is one of the main achievements of Greek mathematics.

A simple example showing that one should not be too careless with these things is the following. Consider the monoid of numbers $H_3 = \{1, 4, 7, 10, \ldots\}$ of natural numbers of the form $3n+1$. In the set of fractions $\frac{a}{b}$ with $a, b \in H_3$, the fraction $\frac{100}{40}$ has two different reductions, namely $\frac{10}{4}$ as well as $\frac{25}{10}$; note that in H_3, neither of them can be reduced to $\frac{5}{2}$ since neither 5 nor 2 lie in H_3. It is instructive to watch the proofs below fail[2] while following this example.

Let me take this opportunity to give a modern proof of the uniqueness of reduced fractions:

Proposition 1.4. *Let e be the smallest numerator of all fractions equal to $\frac{a}{b}$ (with a, b positive), and write $\frac{a}{b} = \frac{e}{f}$. Then there is an integer m such that $a = me$ and $b = mf$.*

Proof. Let n be the maximal integer such that $ne < a$ (observe that we clearly have $n \geq 1$). Since $af = be$, we also have $nf < b$. Now it is easily verified that

$$\frac{e}{f} = \frac{a - ne}{b - nf},$$

and the minimality of e implies that $a - ne \geq e$. By the choice of n we also have $(n+1)e \geq a$, i.e., $a - ne \leq e$. But then $a - ne = e$, hence $a = (n+1)e$ is a multiple of e as desired. □

This result also shows that minimal representations of fractions are the same as representations in lowest terms:

Lemma 1.5. *The fraction $\frac{e}{f}$ equal to $\frac{a}{b}$ with minimal e is also the one with coprime numbers e and f, and conversely every coprime representation of this fraction is one with minimal numerator.*

Proof. If $\frac{a}{b} = \frac{e}{f}$ with e minimal, then e and f are coprime: otherwise $e = ng$ and $f = nh$, and we would have $\frac{a}{b} = \frac{g}{h}$ with $g < e$.

If $\frac{a}{b} = \frac{e}{f}$ with e and f coprime, then e is minimal: in fact, let $\frac{a}{b} = \frac{g}{h}$ with g minimal; then $e = ng$ and $f = nh$ for some integer $n \geq 1$ by Prop. 1.4. Since e and f were assumed to be coprime, we must have $n = 1$. □

[2] Problems arise as soon as the ring structure (addition and subtraction) is used. Later we will also discuss genuine rings where reduction of fractions fails.

1.2. Unique Factorization

Where we define irreducible numbers and prime numbers, show that these notions coincide in the ring of integers $\mathbb{Z}$, and prove that numbers cannot be factored into primes in two different ways.

We say that an integer $b \in \mathbb{Z}$ **divides** $a \in \mathbb{Z}$ (and write $b \mid a$) if there exists an integer $q \in \mathbb{Z}$ such that $a = bq$. If sums are involved we usually use brackets, but occasionally omit them if there is no possibility of misunderstanding the intended meaning, as in $2^n - 1 \mid 4^n - 1$.

Fact 1.6. *For all integers $a, b \in \mathbb{Z}$ we have*

1. *$a \mid 0$;*
2. *$1 \mid a$ and $a \mid a$;*
3. *$a \mid b$ if and only if $(-a) \mid b$;*
4. *$a \mid b$ if and only if $a \mid (-b)$;*

These claims follow immediately from the definition of divisibility.

Proposition 1.7. *For all integers a, b, c, the following assertions hold.*

1. *If $a \mid b$ and $b \mid c$, then $a \mid c$;*
2. *If $a \mid b$ and $a \mid c$, then $a \mid (b \pm c)$.*

Proof. For showing the first claim, observe that we have $b = aq$ and $c = br$ for $q, r \in \mathbb{Z}$; but then $c = br = a(qr)$, hence $a \mid c$.

The second claim is proved as follows: we have $b = aq$ and $c = ar$ for $q, r \in R$; then $b \pm c = a(q \pm r)$ implies that $a \mid (b \pm c)$. □

Elements dividing 1 are called **units**; the units in $\mathbb{Z}$ can be determined easily:

Proposition 1.8. *The unit group of $\mathbb{Z}$ is $\mathbb{Z}^\times = \{-1, +1\}$.*

Proof. Both -1 and $+1$ are units because they divide 1. Now assume that $r \in \mathbb{Z}$ is a unit; then there exists an element $s \in \mathbb{Z}$ with $rs = 1$. Clearly $r, s \neq 0$, hence $|r|, |s| \geq 1$. If $|r| > 1$, then $0 < |s| < 1$, but there are no integers strictly between 0 and 1. Thus $|r| = 1$, that is, $r = \pm 1$. □

We now give two important definitions. A non-unit $p \in \mathbb{Z}$ is called

- **irreducible** if it only has trivial factorizations, i.e. if $p = ab$ for $a, b \in \mathbb{Z}$ implies that $a = \pm 1$ or $b = \pm 1$.
- **prime** if $p \mid ab$ for $a, b \in \mathbb{Z}$ implies that $p \mid a$ or $p \mid b$.

Observe that we have only used the multiplicative structure of the ring $\mathbb{Z}$; this fact will allow us to transfer these definitions to monoid (see Exercises 71 – 75), such as the set of numbers $H_3 = \{1, 4, 7, 10, 13, \ldots\}$; in this set, the number 4 is irreducible, but not prime because $4 \mid 10 \cdot 10 = 4 \cdot 25$, yet $4 \nmid 10$.

In general it is easier to prove a number irreducible than proving it to be prime. For showing that an integer n is irreducible we only have to show that it is not divisible by the finitely many factors less than n (even less than $\sqrt{n}$), whereas for showing that a number is prime we have to look at infinitely many possible products ab divisible by p.

A trivial but occasionally useful observation is the following: an integer p is prime (resp. irreducible) if $|p|$ is prime (resp. irreducible). Thus for studying the primality of numbers we may restrict our attention without loss of generality to the natural numbers $\mathbb{N}$.

Observe also that the units ± 1 are, by definition, neither irreducible nor prime. The following observation is easy to prove:

Proposition 1.9. *Prime numbers in* $\mathbb{N}$ *are irreducible.*

Proof. Let p be a prime number. If p is not irreducible, then $p = ab$ for numbers $a, b > 1$. But then $p \mid ab$, and since p is prime, we must have $p \mid a$ or $p \mid b$. This implies $p = a$ (and $b = 1$) or $p = b$ (and $a = 1$), contradicting the assumption that $a, b > 1$. □

Proving the converse is more difficult: numbers such as 2, 3 or 5 are easily seen to be irreducible, but showing that they are prime is more involved. For proving that 2 is a prime number, assume that $2 \mid ab$. We have to show that $2 \mid a$ or $2 \mid b$. If both a and b were odd, say $a = 2m + 1$ and $b = 2n + 1$, then $ab = (2m + 1)(2n + 1) = 2(2mn + m + n) + 1$ is also odd, contradicting the fact that ab is even. Direct proofs that 3 or 5 are prime are even more technical.

We will now show that the converse of Prop. 1.9, a result sometimes called Euclid's Lemma, is also true using a beautiful idea due to Zermelo:

Proposition 1.10. *Irreducible numbers in* $\mathbb{N}$ *are prime.*

Proof. We have already shown that 2 is prime. Assume that p is irreducible and that the claim holds for all irreducible numbers less than p. Given integers a and b with $p \mid ab$ we have to show that $p \mid a$ or $p \mid b$. Write $b = kp + r$ for some number $0 \leq r < p$. If $r = 0$, then $p \mid b$. If $r \geq 1$, multiplying the last equation through by a we get $ab = akp + ar$, and since p divides the first two terms we find $p \mid ar$. If $a = 1$, then $p \mid a$ and we are done. If $r > 1$, write $ar = pb$ and let q be an irreducible factor of r. Since $q \leq r < p$, this factor q is prime, so $q \mid pb$ implies that q divides p or b. The first case is impossible since p is irreducible, hence $q \mid b$, and canceling q we get $ar' = pb'$ with $1 \leq r' < r$. After finitely many steps we end up with an equation of the form $a = pc$, which shows that $p \mid a$. □

Zermelo's proof used division with remainder as its main tool; the same holds for the proof of Prop. 1.4, and in fact we can use Prop. 1.4 to give a

Second proof of Prop. 1.10. In fact, assume that $p \mid ab$, so $ab = pr$. This implies $\frac{a}{p} = \frac{r}{b}$. If $p \nmid a$, then $\frac{a}{p}$ is reduced (remember that p is irreducible), hence is the minimal representation. Prop. 1.4 then shows that we must have $p \mid b$ as claimed. □

Now we can show that every nonzero integer can be written (in an essentially unique way) as a product of irreducibles. This is not obvious, as the monoid $M = \{1, 5, 9, \ldots\}$ of natural numbers of the form $4n + 1$ (see Exercises 71 – 75 for more details) shows: here $21 \cdot 21 = 9 \cdot 49$ are two distinct factorizations of 441 into irreducibles.

Theorem 1.11. *Every nonzero integer* $n \in \mathbb{Z}$ *can be written as the product of a unit and positive irreducible elements. This product is unique up to the order of the factors.*

Instead of saying that $\mathbb{Z}$ has unique factorization, we will often say that $\mathbb{Z}$ is **factorial**. We first remark that it is sufficient to prove everything for natural numbers, since every nonzero integer is ± 1 times a natural number.

Existence: This is trivial for $m = 1, 2, 3$. Assume we have proved the existence of a factorization into irreducibles for all numbers $< m$. If n is irreducible, there is nothing to prove; if not, then $n = ab$ for natural numbers $a, b < m$. These can be factored into irreducibles by induction assumption, hence the same is true for their product.

Uniqueness: Let $n = p_1 \cdots p_r = q_1 \cdots q_s$ denote two factorizations of $n \in \mathbb{N}$ into positive irreducible numbers. Since p_1 is prime, it must divide some factor on the right, say q_1. Since q_1 is irreducible, we must have $p_1 = q_1$. Canceling this factor and repeating this step (at most r times) we find that both factorizations actually must be the same.

Proof. Let $p \in \mathbb{N}$ be irreducible, and assume that $p \mid ab$. Write $a = up_1 \cdots p_r$ and $b = vq_1 \cdots q_s$ for units $u, v \in \{-1, +1\}$ and irreducibles p_i and q_j. Then $ab = uvp_1 \cdots p_r q_1 \cdots q_s$ is the factorization of ab into irreducibles. On the other hand, $ab = pc$, and by factoring c into irreducibles we see that p occurs in the factorization of ab into irreducibles. Thus p must be one of the p_i or one of the q_j. But then $p \mid a$ in the first, and $p \mid b$ in the second case. □

Unique factorization implies that a number $n \in \mathbb{N}$ with prime factorization $n = \prod p^{a_p}$, where the product is over all prime numbers p, and where $a_p \geq 0$ is nonzero only for finitely many primes p, is a square number if and only if all exponents a_p are even. This observation in turn implies

Corollary 1.12. *The square root $\sqrt{n}$ of an integer $n \in \mathbb{N}$ is rational if and only if n is a square number.*

Proof. Assume that $n = \frac{a^2}{b^2}$ for $a, b \in \mathbb{N}$. Then $nb^2 = a^2$ is a square number, hence all exponents in the prime factorization of nb^2 and, therefore, of n must be even. But then n is a perfect square. □

In the monoid $M = \{1, 2, 4, 5, 6, 8, \ldots\}$ of all positive integers not of the form $4n + 3$, the number 9 is the square of the fraction $\frac{6}{2}$ since $\frac{6}{2} \cdot \frac{6}{2} = \frac{36}{4} = 9$. On the other hand, 9 is not a square in M.

The following famous result along with its beautiful proof is due to Euclid:

Proposition 1.13. *There exist infinitely many prime numbers.*

Proof. Examples of primes are 2, 3, 5. Assume that there are only finitely many primes $p_1 \cdots p_n$. Consider the integer $N = p_1 \cdots p_n + 1$. Then $N \neq 0$ is not a unit in $\mathbb{Z}$ because $N > 1$. But then N must be divisible by some prime p. If $p = p_j$, then $p_j \mid N$ and $p_j \mid N-1$ (since $N - 1 = p_1 \cdots p_n$), and thus $p_j \mid 1$ (the difference between N and $N - 1$). But then p_j is a unit, and this is a contradiction. Thus the prime factor p is distinct from the primes $p_1, \ldots, p_n$, and this implies the claim. □

Note that Euclid's proof does not show that the numbers $N = p_1 \cdots p_n + 1$ are prime. With `pari`, it is easy to factor the first few such numbers: typing in `factor(2*3*5*7+1)`, for example, will yield the answer 211. In this way, the results in Table 1.1 were constructed.

$$
\begin{aligned}
2 + 1 &= 3 \\
2 \cdot 3 + 1 &= 7 \\
2 \cdot 3 \cdot 5 + 1 &= 31 \\
2 \cdot 3 \cdot 5 \cdot 7 + 1 &= 211 \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 &= 2311 \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 &= 59 \cdot 509 \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 + 1 &= 19 \cdot 97 \cdot 277 \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 + 1 &= 347 \cdot 27953 \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 + 1 &= 317 \cdot 703763, \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 29 + 1 &= 331 \cdot 571 \cdot 34231, \\
2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 + 1 &= 200560490131.
\end{aligned}
$$

Table 1.1. Factorizations of products of primes +1

The current version of `pari` has no problems with factoring the products of the first n primes for $n \leq 33$. The running time becomes longer for $n = 34$ and $n = 44$. The program I used was the following one-liner:

```
p=1; P=1; for(n=1,100,p=nextprime(p+1); P=P*p; print(n," ",factor(P+1)))
```

Pythagorean Triples via Unique Factorization

We next show how to apply unique factorization to the problem of Pythagorean triples.

Assume that (x, y, z) is a Pythagorean triple. If d divides two of these, it divides the third, and then $(x/d, y/d, z/d)$ is another Pythagorean triple. We may therefore assume that x, y and z are pairwise coprime (two integers a and b are called coprime when $p \mid a$ and $p \mid b$ implies that p is a unit); such triples are called primitive.

Now let (x, y, z) be a primitive Pythagorean triple. Clearly, x and y cannot both be even; we now claim that they cannot both be odd. In fact, if $x \equiv y \equiv 1 \bmod 2$, then $x^2 \equiv y^2 \equiv 1 \bmod 4$, hence $z^2 = x^2 + y^2 \equiv 2 \bmod 4$. But squares are always congruent to 0 or 1 mod 4, so this is a contradiction.

Interchanging x and y if necessary we may assume that y is even. Now we transfer the additive problem $x^2 + y^2 = z^2$ into a multiplicative one (if we are to use unique factorization, we need products, not sums) by writing $y^2 = z^2 - x^2 = (z - x)(z + x)$.

Here we have two factors whose product is a square. Unique factorization allows us to derive information from such a situation:

Proposition 1.14. *Let $a, b \in \mathbb{N}$ be coprime integers such that ab is a square. Then a and b are squares.*

Proof. Write down the prime factorizations of a and b as

$$a = p_1^{a_1} \cdots p_r^{a_r}, \quad b = q_1^{b_1} \cdots q_s^{b_s}.$$

Now a and b are coprime, so the set of p_i and the set of q_j are disjoint, and we conclude that the prime factorization of ab is given by

$$ab = p_1^{a_1} \cdots p_r^{a_r} q_1^{b_1} \cdots q_s^{b_s}.$$

Since ab is a square, all the exponents in the prime factorization of ab must be even. This implies that the a_i and the b_j are even, therefore a and b are squares. □

Now we claim that $\frac{z-x}{2}$ and $\frac{z+x}{2}$ are coprime. In fact, assume that $p \mid \frac{z-x}{2}$ and $p \mid \frac{z+x}{2}$. Then p divides both $\left(\frac{z+x}{2} + \frac{z-x}{2}\right) = z$ and $\left(\frac{z+x}{2} - \frac{z-x}{2}\right) = x$, and since these numbers were assumed to be coprime, p must be a unit as claimed.

Thus by Prop. 1.14 there exist integers $m, n \in \mathbb{N}$ such that $\frac{z-x}{2} = n^2$ and $\frac{z+x}{2} = m^2$. Adding and subtracting these equations gives $z = m^2 + n^2$ and $x = m^2 - n^2$, respectively, and from $y^2 = (z - x)(z + x) = 4m^2n^2$ and $y \in \mathbb{N}$ we deduce that $y = 2mn$.

Note that we must have $\gcd(m, n) = 1$: in fact, any common divisor of m and n would divide x, y and z contradicting our assumption that our triple be primitive. We have shown:

Theorem 1.15. *If (x, y, z) is a primitive Pythagorean triple with y even, then there exist coprime integers $m, n \in \mathbb{N}$ such that $x = m^2 - n^2$, $y = 2mn$, and $z = m^2 + n^2$.*

1.3. Congruences

Where we form congruence classes into one ring to rule them all.

Congruences are a very clever notation invented by Gauss (and published in 1801 in his "Disquisitiones Arithmeticae") to denote the residue of a number a upon division by a nonzero integer m. More precisely, he wrote $a \equiv b \bmod m$ (read: a is **congruent** to b modulo m) if $m \mid (a-b)$ for elements $a, b, m \in \mathbb{Z}$. We have, for example,

$$10 \equiv 3 \bmod 7, \qquad 10 \equiv 0 \bmod 5, \qquad 5 \equiv 2 \equiv -1 \bmod 3.$$

It follows from the definition that $a \equiv b \bmod m$ if and only if $a \equiv b \bmod (-m)$; this allows us to assume without loss of generality that the moduli m we are using are always positive.

We next observe

Proposition 1.16. *Congruence between integers is an equivalence relation.*

Proof. Recall that a relation is called an **equivalence relation** if it is reflexive, symmetric and transitive. In our case, we have to show that the relation $\equiv$ has the following properties:

- reflexivity: $a \equiv a \bmod m$;
- symmetry: $a \equiv b \bmod m$ implies $b \equiv a \bmod m$;
- transitivity: $a \equiv b \bmod m$ and $b \equiv c \bmod m$ imply $a \equiv c \bmod m$

for $a, b, c \in \mathbb{Z}$ and $m \in \mathbb{Z} \setminus \{0\}$.

The proofs are straightforward. In fact, $a \equiv a \bmod m$ means $m \mid (a-a)$, and every integer m divides 0. Similarly, $a \equiv b \bmod m$ is equivalent to $m \mid (a-b)$; but this implies $m \mid (b-a)$, hence $b \equiv a \bmod m$. Finally, if $a \equiv b \bmod m$ and $b \equiv c \bmod m$, then $m \mid (b-a)$ and $m \mid (c-b)$, hence m divides the sum $c-a = (c-b)+(b-a)$, and we find $a \equiv c \bmod m$ as claimed. □

Since $\equiv$ defines an equivalence relation, it makes sense to talk about equivalence classes. The equivalence class $[a]$ (or $[a]_m$ if we want to express the dependence on the modulus m) of an integer a is the set of all integers $b \in \mathbb{Z}$ such that $b \equiv a \bmod m$; in particular, every residue class contains infinitely many integers. In the special case $m = 3$, for example, we have

$$\begin{aligned}
[0] &= \{\ldots, -6, -3, 0, 3, 6, \ldots\},\\
[1] &= \{\ldots, -5, -2, 1, 4, 7, \ldots\},\\
[2] &= \{\ldots, -4, -1, 2, 5, 8, \ldots\},\\
[3] &= \{\ldots, -3, 0, 3, 6, 9, \ldots\} = [0],
\end{aligned}$$

etc. Note that $[0] = [3] = [6] = \ldots$ (in fact, $[0] = [a]$ for any $a \in [0]$), and similarly $[1] = [4] = \ldots$ etc. In general, we have $[a] = [a']$ if and only if $a \equiv a' \bmod m$, that is, if and only if $m \mid (a-a')$.

In the case $m = 3$, there were exactly 3 different residue classes modulo 3, namely [0], [1], and [2] (or, say, [0], [1], and [−1] since $[-1] = [2]$). This is easily generalized:

Lemma 1.17. *For any integer $m > 1$, there are exactly m different residue classes modulo m, namely $[0]$, $[1]$, $[2]$, $\ldots$, $[m-1]$.*

Proof. We first show that these classes are pairwise distinct. To this end, assume that $[a] = [b]$ for $0 \le a, b < m$; this implies $b \in [a]$, hence $a \equiv b \bmod m$ or $m \mid (b-a)$: but since $|b-a| < m$, this can only happen if $a = b$.

Next, there are no other residue classes: given any class $[a]$, we write $a = mq + r$ with $0 \le r < m$ (the division algorithm at work again), and then $[a] = [r]$ is one of the classes listed above. □

The set $\{0, 1, 2, \dots, m-1\}$ is often called a complete set of representatives modulo m for this reason. Sometimes we write $r + m\mathbb{Z}$ instead of $[r]$.

The one thing that makes congruences *really* useful is the fact that we can define a ring structure on the set of residue classes. This is fundamental, so let us do this in detail.

Residue Class Rings

The elements of our ring $\mathbb{Z}/m\mathbb{Z}$ will be the residue classes $[0]$, $[1]$, $\dots$, $[m-1]$ modulo m. We have to define an addition and a multiplication and then verify the ring axioms (see Sect. 1.5.).

• Addition $\oplus$: Given two classes $[a]$ and $[b]$, we put $[a] \oplus [b] = [a+b]$. We have to check that this is well defined: assume that $[a] = [a']$ and $[b] = [b']$; then we have to show that $[a+b] = [a'+b']$. But this is easy: we have $a - a' \in m\mathbb{Z}$, say $a - a' = mA$, and similarly $b - b' = mB$. But then $(a+b) - (a'+b') = m(A+B) \in m\mathbb{Z}$, hence $[a+b] = [a'+b']$.

The neutral element of addition is the residue class $[0] = m\mathbb{Z}$, and the inverse element of $[a]$ is $[-a]$, or, if you prefer, $[m-a]$. In fact, we have $[a] \oplus [0] = [a+0] = [a]$ and $[a] \oplus [-a] = [a + (-a)] = [0]$. The law of associativity and the commutativity are inherited from the corresponding properties of integers: since e.g. $(a+b)+c = a+(b+c)$, we have $([a] \oplus [b]) \oplus [c] = [a] \oplus ([b] \oplus [c])$.

• Multiplication $\odot$: of course we put $[a] \odot [b] = [ab]$. The verification that this is well defined is left as an exercise. The neutral element of multiplication is the class $[1]$.

• Distributive Law: Again, $([a] \oplus [b]) \odot [c] = [a] \odot [c] \oplus [b] \odot [c]$ follows from the corresponding properties of integers.

Theorem 1.18. *The residue classes* $[0]$, $[1]$, $\dots$, $[m-1]$ *modulo* m *form a ring* $\mathbb{Z}/m\mathbb{Z}$ *with respect to addition* $\oplus$ *and multiplication* $\odot$.

The main content of the fact that $\mathbb{Z}/m\mathbb{Z}$ is a ring is the following

Proposition 1.19. *Let* $f \in \mathbb{Z}[X]$ *be a polynomial with integral coefficients. If* $a \equiv b \bmod m$, *then* $f(a) \equiv f(b) \bmod m$.

The proof boils down to the fact that if $a \equiv b \bmod m$ and $c \equiv d \bmod m$, then $a + c \equiv b + d \bmod m$ and $ac \equiv bd \bmod m$.

Now that we have introduced the rings that we will study for some time to come, we simplify the notation by writing $+$ and $\cdot$ instead of $\oplus$ and $\odot$. Moreover, we will drop our references to classes and deal only with the integers representing them; in order to make clear that we are dealing with residue classes, we write $\equiv$ instead of $=$, and for indicating the modulus we add a "mod m" at the end.

Let us also remark that the observation that squares of numbers of the form $3n \pm 1$ have the form $9n^2 \pm 6n + 1 = 3N + 1$ for $N = 3n^2 \pm 2n$ can now be stated as a simple congruence $(\pm 1)^2 \equiv 1 \bmod 3$. Similarly it is easily seen that squares of odd numbers are $\equiv 1 \bmod 8$, and that squares of integers coprime to 6 are $\equiv 1 \bmod 24$.

1.4. Greatest Common Divisors and the Euclidean Algorithm

Where we define greatest common divisors and learn from Euclid how to compute them.

We will now introduce greatest common divisors: we say that an integer d is a greatest common divisor of $a, b \in \mathbb{Z}$ if d satisfies the following two properties:

1. $d \mid a$, $d \mid b$: d is a common divisor of a and b.
2. if $e \in \mathbb{Z}$ satisfies $e \mid a$ and $e \mid b$, then $e \mid d$: every common divisor of a and b divides d.

The following properties of greatest common divisors follow right from the definition:

Fact 1.20. *Let $a, b \in \mathbb{Z}$ be nonzero integers.*

1. *If d is a greatest common divisor of a and b, then so is $-d$.*
2. *Up to units, the greatest common divisor of a and b is unique.*

This allows us to assume that $d > 0$, and we usually write $d = \gcd(a, b)$.

We can use the unique factorization property to give a formula for the greatest common divisor of two integers. Before we do so, let us introduce some notation. We can write an $a \in \mathbb{Z}$ as a product of primes. In fact fact we can write $a = \pm \prod p_i^{a_i}$, where the product is over all irreducible elements p_1, p_2, p_3, $\ldots$, and where at most finitely many a_i are nonzero. In order to avoid the $\pm$ in our formulas, let us restrict to positive integers from now on.

Lemma 1.21. *For integers $a, b \in \mathbb{N}$ with prime factorizations $a = \prod p_i^{a_i}$ and $b = \prod p_i^{b_i}$ we have $b \mid a$ if and only if $b_i \le a_i$ for all i.*

Proof. We have $b \mid a$ if and only if there is a $c \in \mathbb{N}$ such that $a = bc$. Let $c = \prod p_i^{c_i}$ be its prime factorization. Then $c_i \ge 0$ for all i, and $a_i = b_i + c_i$, hence $b \mid a$ is equivalent to $a_i \ge b_i$ for all i. □

Here's our formula for greatest common divisors:

Theorem 1.22. *The gcd of two nonzero integers*

$$a = \prod p_1^{a_i} \quad \text{and } b = \prod p_i^{b_i}$$

is given by

$$d = \prod p_i^{\min\{a_i, b_i\}}.$$

For a proof we have to show that the two properties characterizing greatest common divisors are satisfied:

1. $d \mid a$ and $d \mid b$. But this follows immediately from Lemma 1.21.
2. If $e \mid a$ and $e \mid b$, then $e \mid d$. In fact, write down the prime factorization $e = \prod p_i^{e_i}$ of e. Then $e \mid a$ and $e \mid b$ imply $e_i \le \min(a_i, b_i) = d_i$, hence $e \mid d$.

Observe that these formulas are mostly of a theoretical interest. In practice, they can only be used if the prime factorizations of a and b are known; as we will see, finding the prime factors of large integers is usually quite challenging.

For the ring $\mathbb{Z}$ of integers, we have much more than the mere existence of greatest common divisors: the gcd of two integers $a, b \in \mathbb{Z}$ has a "Bezout representation", that is, if $d = \gcd(a, b)$, then there exist integers $m, n \in \mathbb{Z}$ such that $d = am + bn$.

Theorem 1.23 (Bezout's Lemma)**.** *Assume that $d = \gcd(a, b)$ for $a, b \in \mathbb{Z}$; then d has a Bezout representation: there exist integers m, n such that $am + bn = d$.*

Proof. Consider the set $D = a\mathbb{Z} + b\mathbb{Z} = \{am + bn : m, n \in \mathbb{Z}\}$. Clearly D contains a nonzero number. If $c \in D$ then we also have $-c \in D$. Thus D contains positive integers.

Let d be the smallest positive integer in D; we claim that $d = \gcd(a, b)$. There are two things to show:

Claim 1: d is a common divisor of a and b. By symmetry, it is sufficient to show that $d \mid a$. Write $a = rd + s$ with $0 \leq s < d$; from $d = am + bn$ we get $s = a - rd = a - r(am + bn) = a(1 - rm) + b(-rn)$, hence $s \in D$. The minimality of d implies $s = 0$, hence $d \mid m$.

Claim 2: if e is a common divisor of a and b, then $e \mid d$. Assume that $e \mid a$ and $e \mid b$. Since $d = am + bn$, we conclude that $e \mid d$.

The existence of the Bezout representation is a simple consequence of the fact that $d \in D$. □

Note that the key to this proof (as well as to Zermelo's proof of unique factorization) is the existence of a division with remainder. The proof we have given is not constructive; in the next section we will show how Bezout representations can be computed with the Euclidean algorithm.

Bezout's Lemma can be used to give an important generalization of the Euclidean property $p \mid ab \Longrightarrow p \mid a$ or $p \mid b$ of primes p:

Proposition 1.24. *If $m \mid ab$ and $\gcd(m, b) = 1$, then $m \mid a$.*

Proof. Write $ab = mn$; by Bezout, there are $x, y \in \mathbb{Z}$ such that $mx + by = 1$. Multiplying through by a gives $a = max + aby = max + mny = m(ax + ny)$, that is, $m \mid a$. □

Second Proof of Unique Factorization

Note that our proof of Bezout's result did not use unique factorization (in particular, it did not use Thm. 1.22). In fact, we can use Bezout to give another proof of unique factorization. First we show that in $\mathbb{Z}$, irreducibles are prime.

Second Proof of Cor. 1.10. In fact, let $p \in \mathbb{N}$ be irreducible, and assume that $p \mid ab$. Let $d = \gcd(p, b)$; since $d \mid p$ and p is irreducible, we either have $d = 1$ or $d = p$. If $d = p$, then $p \mid b$. If $d = 1$, then Bezout gives us $x, y \in \mathbb{Z}$ with $1 = px + by$. Multiplying through by a shows $a = pax + aby$. Since p divides both terms on the right hand side, we conclude that $p \mid a$. □

Now we can give our

Second Proof of Unique Factorization. Assume that unique factorization fails; then let n be the smallest positive integer with two distinct factorizations (the smallest criminal, as such elements are called in group theoretic circles) into positive irreducible elements. Write $n = p_1 \cdots p_r = q_1 \cdots q_s$. Then $p_1 \mid q_1 \cdots q_s$, and since irreducibles are prime, p_1 must divide a factor of the right hand side, say $p_1 \mid q_1$. Since q_1 is irreducible we find that $q_1 = p_1 u$ for some unit u, and since $p_1, q_1 > 0$ by assumption we must have $u = 1$ and $p_1 = q_1$. But then canceling p_1 gives $p_2 \cdots p_r = q_1 \cdots q_s$. Since n was the smallest integer with two distinct factorizations, we must have $r = s$ and, after reordering the factors if necessary, $p_2 = q_2, \ldots, p_r = q_r$. Thus the two factorizations were the same after all. □

There are a couple of simple properties of greatest common divisors that can easily be proved from the definition or using Bezout or unique factorization:

Proposition 1.25. *For integers $a, m, n \in \mathbb{N}$ we have*

1. $\gcd(am, an) = a \gcd(m, n)$;
2. *If $a = \gcd(m, n)$, then $\gcd(\frac{m}{a}, \frac{m}{b}) = 1$;*

3. *If* $a \mid m$ *and* $a \mid n$, *then* $a \mid \gcd(m,n)$;
4. $\gcd(a,mn) \mid \gcd(a,m) \cdot \gcd(a,n)$;
5. *If* $\gcd(a,m) = 1$, *then* $\gcd(a,mn) = \gcd(a,n)$.

Proof. For proving the last claim, set $d = \gcd(a,n)$. Then $d \mid a$, $d \mid n$, hence $d \mid mn$, and this shows that $d \mid \gcd(a,mn)$.

Conversely, assume that $e \mid \gcd(a,mn)$; we have to show that $e \mid d$. Since a and m are coprime, there exist integers r,s with $1 = ar + ms$. Then $n = arn + mns$, and since $e \mid a$ and $e \mid mn$ we conclude that $e \mid n$. Thus $e \mid \gcd(a,n) = d$ as claimed. □

Finally, observe that canceling factors in congruences is dangerous: we have $2 \equiv 8 \bmod 6$, but not $1 \equiv 4 \bmod 6$. Here's what we're allowed to do:

Proposition 1.26. *If* $ac \equiv bc \bmod m$, *then* $a \equiv b \bmod \frac{m}{\gcd(m,c)}$.

Proof. We have $m \mid (ac - bc) = c(a-b)$. Write $d = \gcd(m,c)$, $m = dm'$, $c = dc'$, and note that $\gcd(m',c') = 1$. From $dm' \mid dc'(a-b)$ we deduce immediately that $m' \mid c'(a-b)$; since $\gcd(m',c') = 1$, we even have $m' \mid (a-b)$ by Prop 1.24, i.e. $a \equiv b \bmod \frac{m}{\gcd(m,c)}$. □

The Euclidean Algorithm

In many modern textbooks, unique factorization for integers is proved using the Euclidean algorithm; it has the advantage that a similar proof can also be used for other rings, e.g. polynomial rings $K[X]$ over fields K. The Euclidean algorithm is a procedure that computes the gcd of integers without using their prime factorization (which may be difficult to obtain if the numbers involved are large). Moreover, it allows us to compute a Bezout representation of this gcd (note that our proof of Thm. 1.23 was an existence proof, giving no hint at how to compute such a representation).

Given integers m and n, there are uniquely determined integers q_1 and r_1 such that $m = q_1 n + r_1$ and $0 \le r_1 < n$. Repeating this process with n and r_1, we get $n = r_1 q_2 + r_2$ with $0 \le r_2 < r_1$, etc. Since $n > r_1 > r_2 > \ldots \ge 0$, one of the r_i, say r_{n+1}, must eventually be 0:

$$\begin{aligned} m &= q_1 n + r_1 && (1.6)\\ n &= q_2 r_1 + r_2 && (1.7)\\ r_1 &= q_3 r_2 + r_3 && (1.8)\\ &\ldots \\ r_{n-2} &= q_n r_{n-1} + r_n && (1.9)\\ r_{n-1} &= q_{n+1} r_n && (1.10) \end{aligned}$$

Example. For finding the greatest common divisor of $m = 56$ and $n = 35$, we compute

$$\begin{aligned} 56 &= 1 \cdot 35 + 21\\ 35 &= 1 \cdot 21 + 14\\ 21 &= 1 \cdot 14 + 7\\ 14 &= 2 \cdot 7 \end{aligned}$$

Note that the last r_i that does not vanish (namely $r_3 = 7$) is the gcd of m and n. This is no accident: we claim that $r_n = \gcd(m,n)$ in general. For a proof, we have to verify two things:

Claim 1: r_n is a common divisor of m and n. Equation (1.10) shows $r_n \mid r_{n-1}$; plugging this into (1.9) we find $r_n \mid r_{n-2}$, and going backwards we eventually find $r_n \mid r_1$ from (1.8),

$r_n \mid n$ from (1.7) and finally $r_n \mid m$ from (1.6). In particular, r_n is a common divisor of m and n.

Claim 2: if e is a common divisor of m and n, then $e \mid r_n$. This is proved by reversing the argument above: (1.6) shows that $e \mid r_1$, (1.7) then gives $e \mid r_2$, and finally we find $e \mid r_n$ from (1.10) as claimed.

The Euclidean algorithm does more than just compute the gcd: take our example $m = 56$ and $n = 35$; writing the third line as $\gcd(m, n) = 7 = 21 - 1 \cdot 14$ and replacing the 14 by $14 = 35 - 1 \cdot 21$ coming from the second line we get $\gcd(m, n) = 21 - 1 \cdot (35 - 1 \cdot 21) = 2 \cdot 21 - 1 \cdot 35$. Now $21 = 56 - 1 \cdot 35$ gives $\gcd(m, n) = 2 \cdot (56 - 1 \cdot 35) - 1 \cdot 35 = 2 \cdot 56 - 3 \cdot 35$, and we have found a Bezout representation of the gcd of 56 and 35.

This works in complete generality: (1.9) says $r_n = r_{n-2} - q_n r_{n-1}$; the line before, which reads $r_{n-1} = r_{n-3} - q_{n-1} r_{n-2}$, allows us to express r_n as a $\mathbb{Z}$-linear combination of r_{n-2} and r_{n-3}, and going backwards we eventually find an expression of r_n as a $\mathbb{Z}$-linear combination of a and b.

1.5. Rings and Domains

Where we point out the abstract algebra behind elementary number theory.

A lot of the notions we have defined so far make sense not only for the ring of integers in which we are mainly interested, but also for quite general rings. In this section we will define divisibility in general rings, which will lead us to irreducible and prime elements in general domains. We will also enlarge our list of examples of domains.

Rings

First recall that a **ring** is a set R with two compositions that are traditionally denoted by $+$ (addition) and $\cdot$ (multiplication). This means that for all elements $a, b \in R$ we have $a + b \in R$ and $ab = a \cdot b \in R$. Addition and multiplication satisfy the following axioms:

- associativity of addition: $(a + b) + c = a + (b + c)$.
- existence of a neutral element $0 \in R$ with $a + 0 = a$ for all $a \in R$.
- existence of an additive inverse: for every $a \in R$ there is an element b with $a + b = 0$. This element is usually denoted by $b = -a$.
- associativity of multiplication: $(ab)c = a(bc)$ for all $a, b, c \in R$;
- distributivity: $a(b + c) = ab + ac$ for all $a, b, c \in R$.

If there is an element $1 \in R$ with $1 \cdot a = a \cdot 1 = a$ for all $a \in R$, then the ring is called **unital**. If multiplication is commutative, the ring is called a commutative ring.

The notion of divisibility may be extended to arbitrary rings R: for $a, b \in R$ we say that b **divides** a (as for integers, this is denoted as $b \mid a$) if there is an element $c \in R$ such that $a = bc$. Proposition 1.7 can now be extended without problems to arbitrary rings. Similarly, we can define congruences in the usual way by saying that $a \equiv b \bmod m$ if $m \mid (a - b)$.

An element $r \in R \setminus \{0\}$ is called a **zero divisor** if there is an $s \neq 0$ such that $rs = 0$ (in other words: if there is a nonzero element $s \in R$ with $s \mid 0$). A ring without zero divisors is called a **domain**.

The prime example of a ring is the ring of integers, which is also a domain. In Thm. 1.18 we have constructed residue class rings modulo m; the ring $\mathbb{Z}/6\mathbb{Z}$ is not a domain since $[2] \cdot [3] = [0]$.

A **ring homomorphism** $R \longrightarrow S$ is a map from a ring R to another ring S that respects addition and multiplication, i.e., that satisfies $f(a+b) = f(a)+f(b)$ and $f(ab) = f(a)f(b)$. A **homomorphism of unital rings** is a homomorphism $R \longrightarrow S$ of rings containing 1 with the additional property that the unit element $1_R \in R$ is mapped to the unit element $1_S \in S$. An **isomorphism** of rings is a ring homomorphism that is both injective and surjective.

An element $u \in R$ is called a **unit** if it divides 1_R, that is, if there is an element $v \in R$ with $uv = 1_R$. The units in R form a group $R^\times$ called the **unit group** of R, which means that products and quotients of units are units.

For prime numbers p, the unit groups of the rings $\mathbb{Z}/p\mathbb{Z}$ are particularly simple: every nonzero element in $\mathbb{Z}/p\mathbb{Z}$ is a unit. In fact, if $p \nmid a$ then there exist, by Bezout, integers x, y with $ax + py = 1$, and we have $[a]_p[x]_p = [1]_p$. But when every element $\neq 0$ of a ring has an inverse, then this ring is a field, and we have proved the following

Proposition 1.27. *If $p \in \mathbb{N}$ is a prime, then the residue class ring $\mathbb{Z}/p\mathbb{Z}$ is a field.*

The field $\mathbb{Z}/p\mathbb{Z}$ is called a finite field because it has finitely many elements. As we have seen, there are finite fields with p elements for every prime p. Later we will see that there exist finite fields with $m > 1$ elements if and only if m is a prime power.

The fact that $\mathbb{Z}/p\mathbb{Z}$ is a field means that expressions like $\frac{1}{7} \bmod 11$ make sense. To compute such 'fractions', you can choose one of the following[3] methods:

1. Change the numerator mod 11 until the division becomes possible:
$$\frac{1}{7} \equiv \frac{12}{7} \equiv \frac{23}{7} \equiv \frac{34}{7} \equiv \frac{45}{7} \equiv \frac{56}{7} = 8 \bmod 11,$$
and in fact $7 \cdot 8 = 56 \equiv 1 \bmod 11$. This method only works well if p is very small.
2. Apply the Euclidean algorithm to the pair $(7, 11)$, and compute a Bezout representation; you will find that $1 = 2 \cdot 11 - 3 \cdot 7$, and reducing mod 11 gives $1 \equiv (-3) \cdot 7 \bmod 11$, hence the multiplicative inverse of 7 mod 11 is $-3 \equiv 8 \bmod 11$.

Computing the inverse of $[2]$ in $\mathbb{Z}/p\mathbb{Z}$ is easy: note that we want an integer b such that $[2b] = [1]$; but $[1] = [p+1]$, hence we can simply take $b = \frac{p+1}{2}$.

We call a nonzero nonunit $a \in R$ **irreducible** if every factorization $a = bc$ for $b, c \in R$ has the property that b or c is a unit. Finally, a nonzero nonunit $p \in R$ is called **prime** if $p \mid ab$ for elements $a, b \in R$ implies that $p \mid a$ or $p \mid b$. As for the ring of integers $\mathbb{Z}$, prime numbers in arbitrary rings R are irreducible, and the proof in general rings is the same we have given for $\mathbb{Z}$ in Prop. 1.9. The converse, however, fails in general: there are rings in which not every irreducible element is prime.

A domain R is called **atomic** if if every nonzero nonunit in R can be written as a product of irreducible elements, and **factorial** or a **unique factorization domain** if every nonzero nonunit in R can be written uniquely (up to units and order) as a product of irreducible elements: in a factorial domain, $p_1 \cdots p_r = q_s \cdots q_s$ for irreducible elements p_i, q_j implies that $r = s$ and that, possibly after permuting the q_j, $p_i = u_i q_i$ for units u_i with $u_1 \cdots u_r = 1$.

Our second proof of unique factorization shows

Proposition 1.28. *Let R be a domain. Then the following statements are equivalent:*

- *R is a unique factorization domain.*
- *R is atomic, and every irreducible element is prime.*

[3] See p. 64 for a third method.

As an example showing that our intuition gained from studying the ring of integers may fail in general rings let us look at the ring $R = \mathbb{Z}/6\mathbb{Z}$. The unit group of R is $R^\times = \{[1], [5]\}$. The nonzero nonunits are [2], [3] and [4]. The element [2] is not irreducible since $[2] = [4]\cdot[2]$. For the same reason, the factorizations $[3] = [3]\cdot[3]$ and $[4] = [2]\cdot[2]$ show that [3] and [4] are not irreducible, so R does not possess any irreducible elements. This implies in particular that R does not possess the property of factorization into irreducibles.

Given two rings R and S we can give the cartesian product $R \times S$ a ring structure by demanding that

$$(r_1, s_1) + (r_2, s_2) = (r_1 + r_2, s_1 + s_2), \qquad (r_1, s_1)(r_2, s_2) = (r_1 r_2, s_1 s_2).$$

The zero element of $R \times S$ is $(0, 0)$, the unit element is $(1, 1)$.

Proposition 1.29. *The unit group of $R \times S$ is $(R \times S)^\times = \{(r, s) : r \in R^\times,\ s \in S^\times\}$.*

The simple proof is left as an exercise. You have to show that (r, s) is a unit in $R \times S$ if and only if r and s are units in R and S, respectively. The proof consists in little more than writing down the definitions involved. Such proofs are called "purely formal proofs", and will eventually become trivial.

Prop. 1.29 will be used for decomposing "complex" rings into a product of simpler rings (and even fields). As an example, let us show that

$$\mathbb{Z}/6\mathbb{Z} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}.$$

Any residue class $[a]_6$ determines the two residue classes $[a]_2$ and $[a]_3$:

$[a]_6$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$
$[a]_2$	$[0]_2$	$[1]_2$	$[0]_2$	$[1]_2$	$[0]_2$	$[1]_2$
$[a]_3$	$[0]_3$	$[1]_3$	$[2]_3$	$[0]_3$	$[1]_3$	$[2]_3$

The map $f : \mathbb{Z}/6\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ defined by $f([a]_6) = ([a]_2, [a]_3)$ is a homomorphism of unital rings because

$$\begin{aligned} f([a]_6)f([b]_6) &= ([a]_2, [a]_3) \cdot ([b]_2, [b]_3) && \text{by definition of } f \\ &= ([a]_2[b]_2, [a]_3[b]_3) && \text{by definition of } \cdot \text{ in } \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \\ &= ([ab]_2, [ab]_3) && \text{by multiplication in } \mathbb{Z}/2\mathbb{Z} \text{ and } \mathbb{Z}/3\mathbb{Z} \\ &= f([ab]_6) && \text{by definition of } f, \end{aligned}$$

and since $f([1]_6) = ([1]_2, [1]_3)$ is the unit element of $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. The map f is easily seen to be both surjective and injective, hence it is an isomorphism.

One of our main goals in Chap. 3 will be generalizing this result from $\mathbb{Z}/6\mathbb{Z}$ to arbitrary residue class rings $\mathbb{Z}/m\mathbb{Z}$.

We will also have to review a few basic facts about the characteristic of rings and fields. Let R be an arbitrary ring. If there exists an integer n such that $n \cdot 1 = \underbrace{1 + \ldots + 1}_{n \text{ terms}} = 0$, then the smallest such integer is called the **characteristic** of the ring; if no such n exists, we say that R has characteristic 0.

If R is a finite ring, the characteristic of R is always nonzero: in fact, consider the ring elements $1 \cdot 1$, $2 \cdot 1$, $3 \cdot 1$, ...; since R is finite, eventually elements must repeat, that is, there are integers $m < m'$ such that $m \cdot 1 = m' \cdot 1$. But then $n \cdot 1 = 0$ for $n = m' - m$.

The rings $\mathbb{Z}/n\mathbb{Z}$ have characteristic n, so every integer $n \geq 1$ occurs as the characteristic of a ring. This is not true for fields (or domains):

Proposition 1.30. *Let R be an integral domain; then* char R *is* 0 *or a prime.*

Proof. Let p denote the characteristic of R. Assume that $p \cdot 1 = 0$ and $p = mn$. Then $(m \cdot 1)(n \cdot 1) = p \cdot 1 = 0$; but since R is an integral domain, a product can be 0 only if one of its factors is 0, so we can conclude that $m \cdot 1 = 0$ or $n \cdot 1 = 0$. Since p was chosen to be the minimal integer with $p \cdot 1 = 0$, this implies $m \ge p$ or $n \ge p$, that is, $m = p$ or $n = p$. In other words: all factorizations of p are trivial, hence p is prime. □

Given any prime p, Prop. 1.27 shows that the ring $\mathbb{Z}/p\mathbb{Z}$ of residue classes modulo p is a finite field with p elements and characteristic p. Thus all primes occur as the characteristic of some finite field.

There are other differences between finite rings and fields that one has to be aware of, in particular since we are so used to working over $\mathbb{Z}$, $\mathbb{Q}$ or $\mathbb{F}_p$ that we sometimes fall into traps.

Fact 1.31. *If E is a subfield of F, then E and F have the same zero and identity elements.*

Isn't this obvious? Well, it is not: consider the ring $R = M_2(F)$ of 2×2-matrices over some field F. The set of all matrices $S = \left(\begin{smallmatrix} a & 0 \\ 0 & 0 \end{smallmatrix}\right)$ forms a subring of R with zero $\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and identity $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$. The unit element of $M_2(F)$, on the other hand, is the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

Proof of Fact 1.31. Let 0 be the zero element of F and $0'$ that of E. Then $0' + 0' = 0'$ in E (and therefore in F). Moreover, we have $0 + 0' = 0'$ in F, so comparing yields $0 + 0' = 0' + 0'$, and canceling $0'$ gives $0 = 0'$. Observe that this proof works even when E and F are arbitrary rings.

Similarly, let 1 and $1'$ denote the identities in F and in E, respectively. Then $1' \cdot 1' = 1'$ in E and therefore in F, as well as $1 \cdot 1' = 1'$ in F. Comparing these equations gives $1 \cdot 1' = 1' \cdot 1'$; but in arbitrary rings, we cannot simply cancel $1'$. If F is a field, however, then $1' \ne 0' = 0$ by the field axioms, and canceling the nonzero factor $1'$ yields $1 = 1'$. □

1.6. Linear Diophantine Equations

Where we study linear diophantine equations from a geometric point of view.

The simplest linear diophantine equation is given by $aX = b$, where a and b are given integers. This equation has a rational solution whenever $a \ne 0$, and it has an integral solution (by definition of divisibility) if and only if $a \mid b$.

The next simplest case is that of the equation

$$\ell : \qquad aX + bY = c, \tag{1.11}$$

where $a, b, c \in \mathbb{Z}$ are given integers. In geometry, such equations represent lines unless $a = b = 0$; we will always exclude this case in the following. Here we can also easily give a complete solution:

Proposition 1.32. *The diophantine equation (1.11) is solvable in integers if and only if* $\gcd(a, b) \mid c$. *In this case, let (x_0, y_0) denote an arbitrary integral point on the line ℓ, and set $d = \gcd(a, b)$. Then all integral solutions of (1.11) are given by*

$$x = x_0 + r \cdot \frac{b}{d}, \quad y = y_0 - r \cdot \frac{a}{d}, \qquad \text{where } r \in \mathbb{Z}.$$

Proof. If (x, y) is an integral solution and $d = \gcd(a, b)$, then $ax + by = c$ implies that $d \mid c$. Conversely, if $d \mid c$, then we can write $a = d\alpha$, $b = d\beta$ and $c = d\gamma$, and the equation $aX + bY = c$ becomes $\alpha X + \beta Y = \gamma$. Since α and β are coprime, $\alpha X + \beta Y = 1$ has integral solutions (Bezout); multiplying this solution through by γ yields integral solutions of (1.11).

Now let (x_0, y_0) denote an integral solution of (1.11). If (x, y) is an arbitrary integral solution, then $ax_0 + bx_0 - c = ax + by - c$ shows $a(x - x_0) + b(y - y_0) = 0$, which in turn implies $\alpha(x - x_0) + \beta(y - y_0) = 0$. Since α and β are coprime, the last equation implies $\alpha \mid (y - y_0)$ and $\beta \mid (x - x_0)$. Thus there is an integer r such that $x - x_0 = r\beta$, which yields $y - y_0 = r\alpha$. But this is exactly what we wanted to prove. □

The Frobenius Postage Stamp Problem

Suppose that you are given an unlimited supply of postage stamps with denominations a and b, where $a, b > 1$ are coprime. We will show that a simple geometric argument provides us with the fact that every integer amount $> ab$ can be made using these stamps.

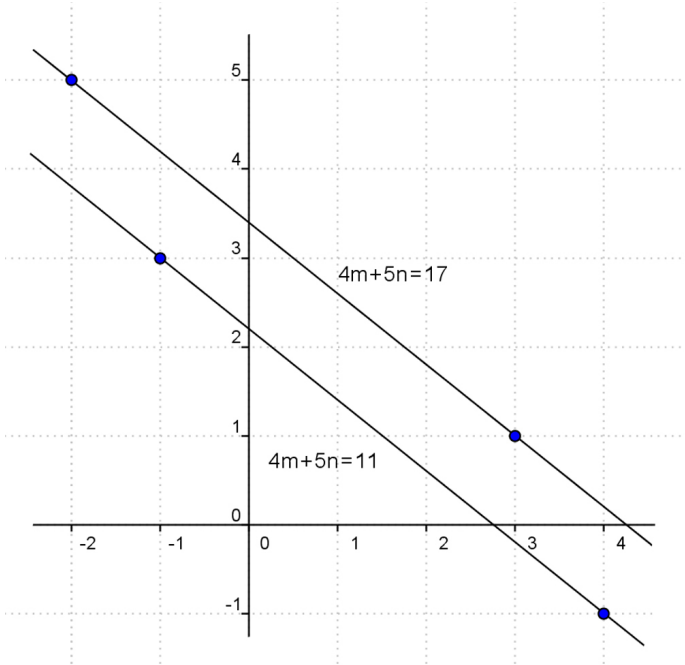
We have to look at the problem $c = ma + nb$, where $m, n \geq 0$ are positive integers. By Prop. 1.32, these solutions are given by the line

$$\left(\begin{smallmatrix} m \\ n \end{smallmatrix}\right) = \left(\begin{smallmatrix} m_0 \\ n_0 \end{smallmatrix}\right) + r\left(\begin{smallmatrix} b \\ -a \end{smallmatrix}\right),$$

which also can be written in the form

$$a(m_0 c + br) + b(n_0 c - ar) = c,$$

where $am_0 + bn_0 = 1$ is a Bezout representation of $1 = gcd(a, b)$. The vector $\left(\begin{smallmatrix} -a \\ b \end{smallmatrix}\right)$ has length $\ell = \sqrt{a^2 + b^2}$, hence there is an integral point on the line in every line segment whose length is greater than ℓ.



This line intersects the coordinate axes in $P(0,\frac{c}{b})$ and $Q(\frac{c}{a},0)$. Thus if the distance between P and Q is $\ge \ell$, there will be an integral solution in the first quadrant. Now

$$\overline{PQ} = \sqrt{\left(\frac{c}{a}\right)^2 + \left(\frac{c}{b}\right)^2} = \frac{c}{ab}\sqrt{a^2+b^2} = \frac{c}{ab}\cdot\ell.$$

Thus there must be a solution of $c = ma + nb$ in nonnegative integers m, n whenever $c \ge ab$ as claimed.

It is easy to see that $c = ab - a - b$ cannot be represented in the form $c = ma + bn$ for nonnegative integers m, n. In fact, if we have $ma + nb = ab - a - b$, then clearly $m \equiv -1 \bmod b$ and $n \equiv -1 \bmod a$. This implies $m \ge b-1$ and $n \ge a-1$, hence $ma + nb \ge (b-1)a + (a-1)b = 2ab - a - b$, which is a contradiction.

It remains to study the representation of integers $ab-a-b < c < ab$. We will find that all of them have a representation of the required form. For seeing this, write $c = ab-a-b+d$ for some integer d with $0 < d < a+b$. The equation $ma + nb = ab - a - b + d$ provides us with the congruences $ma \equiv d - a \bmod b$ and $nb \equiv d - b \bmod a$. Since a and b are coprime, we can write them in the form $m \equiv \frac{d}{a} - 1 \bmod b$ and $n \equiv \frac{d}{b} - 1 \bmod a$. The equation $m_0 a + n_0 b = 1$, on the other hand, shows that $n_0 b \equiv 1 \bmod a$ and $m_0 a \equiv 1 \bmod b$, hence $\frac{1}{b} \equiv n_0 \bmod a$ and $\frac{1}{a} \equiv m_0 \bmod b$. Thus we have $m \equiv \frac{d}{a} - 1 \equiv m_0 d - 1 \bmod a$ and $n \equiv \frac{d}{b} - 1 \equiv n_o d - 1 \bmod b$.

This suggests that we write $m = m_0 d - 1 + kb$ and $n = n_0 d - 1 + la$ for integers k, l, which we choose in such a way that $0 \le m < b$ and $0 \le n < a$. Then we get

$$\begin{aligned} ma + nb &= (m_0 d - 1 + kb)a + (n_0 d - 1 + la)b = m_0 ad - a + kab + n_0 bd - b + lab \\ &= (m_0 a + n_0 b)d - a - b + (k+l)ab = d - a - b + (k+l)ab. \end{aligned}$$

Since $0 \le m_0 d - 1 + kb \le b-1$ and $0 \le n_0 d - 1 + la \le a-1$ we have $0 \le ma + nb \le 2ab - a - b$. If $k + l \le 0$, the equation above would imply $ma + nb < 0$, which is impossible. On the other hand $ma + nb = d - a - b + (k+l)ab \le 2ab - a - b$ shows that we must have $k + l < 2$ since $d > 0$. But then $k + l = 1$, hence $ma + nb = ab - a - b + d$. We have proved:

Proposition 1.33. *Given two coprime integers $a, b > 1$, the equation $ma + nb = c$ has solutions in nonnegative integers m, n for all integers $c > ab - a - b$.*

Notes

Pythagorean triples were studied already by the Babylonians: the clay tablet catalogued as Plimpton 322 dating from about 1900 – 1600 BC contains a list of 15 such triples. Pythagorean triples were, according to Proclus (see Dickson [Dic1920, vol. II, Chap. IV]), known to Pythagoras, who had the 'formulas' $x = 2a + 1$, $y = 2a(a+1)$ and $z = y + 1$. Specific solutions were known to the Hindus about the fifth century BC. Rules for forming Pythagorean triples can also be found in Euclid, or the writings of Nipsus.

Berkhan [Ber1853] presented 19 methods of finding Pythagorean triples. Ono [Ono1994] gives five proofs. For other methods, see Kahle [Kah1970] and Jakob [Jak1995].

The parametrization of the unit circle modulo p was found in the unpublished papers of Gauss [Gau1900].

The parametrization of the unit circle using trigonometric functions appears in Kronecker's 'Vorlesungen über Zahlentheorie'; the connection with the fact that the circle has genus 0 (see Dickson [Dic1920, vol. II, p. 169]) was actually added by the editor Hensel in 1901. See also Stratton [Str1979].

The most prominent among the figurate numbers studied by the Pythagoreans are the squares. Right from their definition as figurate numbers (see Fig. 1.2) we deduce the formula

$$1 + 3 + 5 + \ldots + (2n - 1) = n^2,$$

as well as

$$n^2 + (2n + 1) = (n + 1)^2.$$

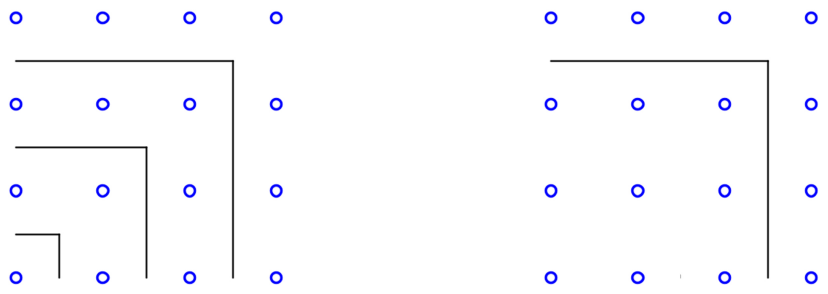


Fig. 1.2. Squares as figurate numbers

The last formula gives a simple construction[4] of Pythagorean triples: all we have to do is make $2n + 1$ a square, say $2n + 1 = m^2$. Then $n = \frac{m^2-1}{2}$, and we find

$$\Big(\frac{m^2 - 1}{2}\Big)^2 + m^2 = \Big(\frac{m^2 + 1}{2}\Big)^2$$

for odd numbers $m \geq 3$, a formula that Proclus ascribes to the Pythagoreans. Proclus also credits Plato with the identity

$$(n^2 - 1)^2 + (2n)^2 = (n^2 + 1)^2$$

for arbitrary numbers $n \geq 2$. Euclid gives a construction of Pythagorean triples in Book X, Lemma 1 of Prop. 29, which, when translated into formulas, reads

$$(pq)^2 + \Big(\frac{p^2 - q^2}{2}\Big)^2 = \Big(\frac{p^2 + q^2}{2}\Big)^2.$$

Euclid's Elements, which consisted of 13 books, are a compilation of the basic propositions known to the Greeks; some more advanced results were not included even though some of them were already known to Euclid. The most spectacular results in Greek mathematics were obtained after Euclid, in particular by Archimedes (areas of segments of parabolas, or the volume of spheres) and Appolonius (who wrote books on geometric properties of conics). Book II is an early example of connections between algebra and geometry: there, Euclid gives geometric proofs of what we perceive as algebraic formulas, such as $a(b + c) = ab + ac$. Books VII, VIII and IX deal with what we would call number theory. Proposition VII.2 explains how to compute the greatest common divisor of two numbers using what is now known as the Euclidean algorithm.

Many of the results in this chapter have their origin in the Elements. Euclid distinguished between the unit 1 and numbers 2, 3, 4, ..., which were proper multiples of 1. Thus 1 not only was not a prime number, it wasn't even a number. Euclid defined prime numbers as numbers that do not possess nontrivial factors.

[4] This construction was used heavily by Leonardo Pisano (Fibonacci) in his Book of Squares.

Perhaps we would expect that the essential parts of Prop. 1.7 can be found in Euclid's Elements, but they are not. The Greeks did not know 0 or negative numbers, so 1.7.1, 1.7.3 and 1.7.4 were not missed by anyone. The fact that $1 \mid a$ was more or less the Greeks' definition of a number as a proper multiple of the unit, and $a \mid a$ did not hold since the quotient $a/a = 1$ wasn't a number. Finally, 1.7.5 and 1.7.6 were used without proof.

The fact that irreducibles are prime (Prop. 1.10) is proved in Proposition VII.30:

> If two numbers multiplied by one another make some number, and any prime number measures the product, then it also measures one of the original numbers.

Euclid stated and proved a special case of unique factorization in Proposition IX.14:

> If a number be the least that is measured by prime numbers, it will not be measured by any other prime number except those originally measuring it.

This states that if n is the lowest common multiple of numbers a, b, c, ..., then the only primes dividing n are those that divide the individual factors. Although this is unique factorization except that repeated factors are not taken care of, it must be remarked that there is nothing fundamental about this proposition: Euclid used this and other results of a similar type only for studying the prime divisors of numbers $2^{n-1}(2^n-1)$ in connection with perfect numbers.

Proposition 1.14 is contained in Euclid's theory of plane numbers. A plane number is a number that can be represented as a product. Two plane numbers $a \cdot b$ and $c \cdot d$ are called similar if $a : c = d : b$. Finally Euclid's propositions VII.24 and VII.26 state that two numbers m and n are similar if and only if $m : n$ is equal to the ratio of square numbers.

Diophantus, as we have seen, solved his problems using clever substitutions. The fact that the majority of his methods have a geometric interpretation was not noticed before the end of the 19th century. This geometric interpretation was made possible through the invention of coordinates by Descartes and Fermat, and in fact Newton [New1971] did parametrize conics using geometric means, although he did so only in a paper that remained unpublished until the 1970s. The idea of using lines (and other simple curves) for parametrizing algebraic curves occurred in the work of Clebsch, whose aim was making Riemann's methods from complex analysis available in a more algebraic context. Clebsch [Cle1865] showed that plane algebraic curves admit a parametrization by rational functions if and only if their genus is 0. This result was then put into a more arithmetic context by Hilbert & Hurwitz [HH1891] as well as by Poincaré [Poi1900]. For more background, see Lemmermeyer [Lem2012].

Although there was hardly anyone interested in number theory between Fermat and Euler, the connection between diophantine analysis and the integrability of certain functions was pointed out by various mathematicians. For finding $\int \frac{dx}{\sqrt{x^2-1}}$, for example, Jacob Bernoulli would write $x^2 - 1 = (x-t)^2$, which is easily solved for x; in fact $x = \frac{t^2+1}{2t}$, so $x^2 - 1 = y^2$ for $y = \frac{t^2-1}{2t}$. The substitution $x = \frac{t^2+1}{2t}$ gives $dx = \frac{t^2-1}{2t^2}dt$, hence $\int \frac{dx}{\sqrt{x^2-1}} = \int \frac{dt}{t}$. Bernoulli's remark in [Ber1909, p.95] concerning the intgegration of $\int \frac{x^2\,dx}{\sqrt{a^4-x^4}}$ is very interesting; he writes

> If someone would plan to remove the irrationality using the method of Diophantus that we have used in the preceding part, then he would use up a whole life. Because the geometers have observed that a sum or a difference of two fourth powers such as $a^4 - x^4$ is never a square.

Indeed the claim that $x^4 \pm y^4 = z^2$ only has trivial solutions (in which $xyz = 0$) was proved by Fermat and Frenicle.

Congruences as we know them were introduced by Gauss in his masterpiece Disquisitiones Arithmeticae. Traces of the calculus of congruences can be found in many ancient

sources (let me mention e.g. the technique of casting out nines). The idea of using congruences in number theoretic calculations is described in a letter by Christian Goldbach to Euler dated May 22, 1730. Goldbach had told Euler about Fermat's claim that all numbers of the form $F_n = 2^{2^n} + 1$ are prime, and now Goldbach[5] shows that nontrivial factors of Fermat numbers cannot be too small:

> Regarding Fermat's observation I agree with you it does not seem credible that he got as far as expressing six terms in that series of his; but that much work is not necessary either for the likelihood of his observation, for it is easy to notice that the remainders of the terms in their natural order, with regard to any fixed divisor, repeat in a cycle. Take, e.g., the term $2^{2^x} + 1$, where $x = 2$, and divide by 7, then the remainder is 3; therefore the next term leaves the same remainder that is left when the number $(3-1)^2+1$ is divided by 7, viz. 5; the term following this will give the same remainder as $(5-1)^2+1$ divided by 7, viz. 3; therefore all possible remainders of all the terms of the series upon division by 7 (where, of course, the quotient is to be positive) are either 3 or 5. By a similar argument it follows easily that no term of Fermat's series can be divided by any number less than 100.

Here Goldbach explains to Euler how to use congruences for computing remainders: according to Goldbach, $2^4+1 \equiv 3 \bmod 7$ implies $2^8+1 = (2^4)^2+1 \equiv (3-1)^2+1 \equiv 5 \bmod 7$. Continuing in this way he finds that the sequence of remainders is periodic: we have $F_{2n} \equiv 3 \bmod 7$ and $F_{2n+1} \equiv 5 \bmod 7$. In particular, no Fermat number is divisible by 7. Had Goldbach continued his calculations, he would have found that the prime 641 does divide a Fermat number, namely $F_5 = 2^{32}+1$. This was later done by Euler (see his letter to Goldbach dated June 30, 1742).

Legendre also felt the need to introduce congruences (as a concept, but not as a notation) in his Recherches d'Analyse Indéterminée, a memoir on diophantine equations that would grow into the first published textbook in number theory. In [Leg1788, p. 516–517], he wrote

> Nous avertissons aussi que cette expression $d^{\frac{c-1}{2}} = 1$ ou -1, suppose qu'on a rejeté les multiples de c dans le premier membre.[6]

Gauss brought home his point that the notion of a congruence is fundamental by opening his Disquisitiones with the definition of congruence:

> If the number a divides the difference of the numbers b and c, then we will call b and c congruent modulo a, and incongruent modulo a otherwise.

In the second article he introduces the notation $b \equiv c \pmod a$ and gives the examples $-16 \equiv 9 \pmod 5$ and $-7 \equiv 15 \pmod{11}$.

The fact that an equation $ax - by = 1$ is solvable in integers x, y if and only if a and b are coprime numbers was proved by Bachet de Méziriac (Problèmes plaisans et délectables qui se font par les nombres; 1624), but was already known to Indian mathematicians such as Bhascara and Brahmagupta. The more general result that $\gcd(a, b)$ can be written as a linear combination of a and b is due to Bezout. This may very well be the first susbtantial result in elementary number theory going beyond Euclid.

The Theorem of Unique Factorization was first stated explicitly and proved by Gauss in his Disquisitiones. In the literature before Gauss, more or less complete statements of the main content can be found in Euclid's Elements, and in the work of many writers

[5] An English translation of the correspondence between Euler and Goldbach can be found in [EG2014].

[6] We also remark that this expression $d^{\frac{c-1}{2}} = 1$ or -1 assumes that on the left hand side multiples of c have been removed.

who studied perfect and amicable numbers; results such as the following, that the only factors of numbers such as $n = p^a q^b r^c$ with p, q, r prime are the numbers $p^\alpha q^\beta r^\gamma$ with $0 \le \alpha \le a$, $0 \le \beta \le b$ and $0 \le \gamma \le c$ were needed for computing the sum of the divisors of n, but the result did not play a fundamental role in any number theoretic investigation before Gauss. In a similar vein, the theorem of unique factorization for polynomials in $\mathbb{C}[X]$ (or $\mathbb{R}[X]$) is an immediate consequence of the fundamental theorem of algebra, and was used heavily e.g. for decomposing rational functions into partial fractions, which was an important technique for integrating rational functions in calculus.

Gauss himself not only proved that the ring of integers is factorial, but also established unique factorization in polynomial rings $(\mathbb{Z}/p\mathbb{Z})[T]$ and in the ring of Gaussian integers $\mathbb{Z}[i]$. Dirichlet later showed that Euclidean rings are always factorial. Kronecker proved that polynomial rings $R[X_1, \ldots, X_n]$ in finitely many variables are factorial if $R = \mathbb{Z}$ or if R is a field, and Hensel finally proved that if R is a factorial domain, then so is $R[X]$.

The fact that $\sin\alpha$ and $\cos\alpha$ are rational expressions in $\tan\frac{\alpha}{2}$ is the content of Lambert's formulas (1765)

$$\sin\alpha = \frac{2\tan\frac{\alpha}{2}}{1+\tan^2\frac{\alpha}{2}}, \qquad \cos\alpha = \frac{1-\tan^2\frac{\alpha}{2}}{1+\tan^2\frac{\alpha}{2}}. \tag{1.12}$$

The analytic parametrization of the rational points on the unit circle appears in Šimerka [Sim1870, p. 205]: the line[7]

$$\tan\frac{1}{2}A = \frac{t}{u}, \quad \text{i.e.,} \quad \sin A = \frac{2tu}{t^2+u^2}, \quad \cos A = \frac{u^2-t^2}{t^2+u^2}$$

gives, in light of $x = \sin A$ and $y = \cos A$, the usual parametrization.

The geometric approach to the Frobenius Postage Stamp Problem was taken from Erickson's book [Eri2011]; the second part of the proof there used generating functions instead of congruences. The problem was raised repeatedly in lectures by Frobenius.

The example of the monoid $H_3 = \{1, 4, 7, 10, \ldots\}$ of natural numbers was used by Hilbert in his lectures on number theory for showing the need of proving unique factorization. The proof of Prop. 1.4 was taken from Mazur's [Maz2007] (see also Pengelley [Pen2012]).

Exercises

1.1 Use infinite descent to show that the equation $x^3+3y^3+9z^3 = 0$ only has the trivial solution $(0,0,0)$ in integers.

1.2 Show that a number $m \in \mathbb{N}$ is an n-th power if and only if the exponents in its prime factorization are multiples of n. Deduce that $\sqrt[n]{m}$ is rational if and only if n is an m-th power of an integer.

1.3 (Cayley, Stieltjes) Let $p_1, p_2, \ldots, p_n$ be a list of n primes. Let $P = \prod p_k$ denote their product, and write $P = AB$ for some choice of numbers $A, B \ge 1$. Show that $A + B$ is coprime to P, and use this to deduce the infinitude of primes.

1.4 Let r, s, t, u be four consecutive Fibonacci numbers (these are defined recursively by $F_0 = 0$, $F_1 = 1$, and $F_{n+1} = F_n + F_{n-1}$). Show that $(ru, 2st, tu - rs)$ is a Pythagorean triple. Does every Pythagorean triple arise in this way? See Raine [Rai1948], Horadam [Hor1961], Umansky & Tallman [UT1968], and Boulger [Bou1989].

1.5 Generalize Exercise 1.4 to conics $X^2 + bY^2 = Z^2$. Hint: see Shannon & Horadam [SH1973].

[7] Šimerka uses the notation $\operatorname{tg}\alpha$ for $\tan\alpha$.

1.6 Find all rational points on the conic $X^2 + 2Y^2 = 1$ using the geometric, the arithmetic, and the analytic method.

1.7 The diagram in the section on the trigonometric parametrization of the unit circle on p. 11 shows that $\tan \frac{\alpha}{2} = \frac{\sin \alpha}{1+\cos \alpha}$.
By squaring this identity and using $\sin^2 x + \cos^2 x = 1$, deduce the duplication formulas

$$\cos^2 \frac{\alpha}{2} = \frac{1 + \cos \alpha}{2} \quad \text{and} \quad 2 \sin \frac{\alpha}{2} \cos \frac{\alpha}{2} = \sin \alpha.$$

Similarly show how to deduce the identities (1.12) from the duplication formulas.

1.8 Let (a, b) be a rational point on the circle $X^2 + Y^2 = c$, where $c \in \mathbb{Q}$. Show that $X = a \sin t + b \cos t$ and $Y = b \sin t - a \cos t$ parametrize the circle.

1.9 Show that $X = \frac{1}{\cos t}$ and $Y = \tan t$ parametrize the hyperbola $X^2 - Y^2 = 1$.

1.10 Show that $X = \cosh t$ and $Y = \sinh t$ parametrize the hyperbola $X^2 - Y^2 = 1$.

1.11 Show that $X = \cos t$ and $Y = \frac{1}{2}(\cos 2t + 1)$ parametrize the parabola $Y = X^2$.

1.12 Given four natural numbers a_0, b_0, c_0, d_0, form four new integers $a_1 = |a_0 - b_0|$, $b_1 = |b_0 - c_0|$, $c_1 = |c_0 - d_0|$, and $d_1 = |d_0 - a_0|$, and continue this way. Here is an example:

$$\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 3 \\ 0 & 0 & 2 & 2 \\ 0 & 2 & 0 & 2 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{array}$$

Show that eventually all numbers become 0.

1.13 Show that in $\mathbb{Z}$, we have $0 \mid a$ if and only if $a = 0$.

1.14 Give an explicit example of non-unique factorization in $M = \{1, 5, 9, 13, \ldots\}$.

1.15 Why does Zermelo's proof of unique factorization not work in

$$M = \{1, 5, 9, 13, \ldots\}?$$

1.16 Prove or disprove: if $n \mid ab$ and $n \nmid a$, then $n \mid b$.

1.17 Prove that $2 \mid n(n + 1)$ for all $n \in \mathbb{N}$.

1.18 Prove that $3 \mid n(n^2 - 1)$ for all $n \in \mathbb{N}$. Generalizations?

1.19 Prove that $8 \mid (n^2 - 1)$ for all odd $n \in \mathbb{N}$.

1.20 Prove or disprove: if $n \mid ab$ and $n \nmid a$, then $n \mid b$.

1.21 Show that there are arbitrary long sequences of composite numbers (Hint: observe that $2 \cdot 3 + 2$ and $2 \cdot 3 + 3$ can be seen to be composite without performing any division; generalize!)

1.22 Show that divisibility defines a *partial order* on $\mathbb{Z}$ by writing $a \leq b$ if $b \mid a$.

1.23 Show that, for integers $a, b, c, d, m \in \mathbb{Z}$ with $m > 0$, we have
- $a \equiv b \bmod m \Longrightarrow a \equiv b \bmod n$ for every $n \mid m$;
- $a \equiv b \bmod m$ and $c \equiv d \bmod m \Longrightarrow a + c \equiv b + d \bmod m$ and $ac \equiv bd \bmod m$;
- $a \equiv b \bmod m \Longrightarrow ac \equiv bc \bmod m$ for any $c \in \mathbb{Z}$.

1.24 Compute $d = \gcd(77, 105)$ and write d as a $\mathbb{Z}$-linear combination of 77 and 105.

1.25 Find a formula for the lowest common multiple $\operatorname{lcm}(a, b)$ of two integers analogous to that in Thm. 1.22.

1.26 Show that $\gcd(a, b) \cdot \operatorname{lcm}(a, b) = ab$ for all $a, b \in \mathbb{N}$.

1.27 Check the addition and multiplication table for the ring $\mathbb{Z}/3\mathbb{Z}$:

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

·	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

1.28 Compute addition and multiplication tables for the rings $\mathbb{Z}/5\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}$.

1.29 Compute the multiplicative inverse of [17] in $\mathbb{Z}/101\mathbb{Z}$.

1.30 Compute $\gcd(2^m-1, 2^n-1)$ for small values of $m, n \ge 1$ until you discover a general formula for d.

1.31 Let $U_1 = U_2 = 1$, and $U_{n+1} = U_n + U_{n-1}$ denote the Fibonacci numbers. Show that $\gcd(U_m, U_n) = U_k$ for $k = \gcd(m, n)$.

1.32 (Goldbach) Show that the Fermat numbers $F_n = 2^{2^n} + 1$ are pairwise coprime.

1.33 Show that there are infinitely many primes of the form $p = 4n + 3$. Hint: if $p_1, \ldots, p_r$ are such primes, set $N = 4p_1 \cdots p_r - 1$ and imitate Euclid's proof.

1.34 Show that there are infinitely many primes of the form $p = 3n + 2$.

1.35 Try to extend the above proof to the case of primes of the form $3n + 1$ (and $5n - 1$). What goes wrong?

1.36 Show that primes $p = c^2 + 2d^2$ satisfy $p = 2$ or $p \equiv 1, 3 \bmod 8$.

1.37 Show that primes $p = c^2 - 2d^2$ satisfy $p = 2$ or $p \equiv 1, 7 \bmod 8$.

1.38 Show that primes $p = c^2 + 3d^2$ satisfy $p = 3$ or $p \equiv 1 \bmod 3$.

1.39 Compute $\gcd(x^2 + 2x + 2, x^2 - x - 2)$ over $\mathbb{Z}/m\mathbb{Z}$ for $m = 2, 3, 5$ and 7, and find its Bezout representation.

1.40 Let $a, b \in \mathbb{N}$ be coprime, and let $r \in \mathbb{N}$ be a divisor of ab. Put $u = \gcd(a, r)$ and $v = \gcd(b, r)$, and show that $r = uv$.

1.41 Assume that $M_p = 2^p - 1$ is a prime. List the complete set of (positive) divisors of $N_p = 2^{p-1}M_p$, and compute their sum. Conclude that if M_p is prime, then N_p is a perfect number (a number n is called perfect if the sum of its (positive) divisors equals $2n$).
Euler later proved that every even perfect number has the form $2^{p-1}M_p$ for some Mersenne prime M_p. It is conjectured (but not known) that odd perfect numbers do not exist.

1.42 Compute the last two digits of 27^{19}.

1.43 For primes $p \in \{3, 5, 7, 11, 13\}$, compute $A \equiv (\frac{p-1}{2})! \bmod p$. Can you find a pattern? If not, compute $B \equiv A^2 \bmod p$. Formulate a conjecture.

1.44 Check which of the primes $p \in \{3, 5, 7, 11, 13\}$ can be written as $p = a^2 + b^2$ with integers $a, b \in \mathbb{N}$ (e.g. $5 = 1^2 + 2^2$). Formulate a conjecture.

1.45 For some small primes $p = 4n + 1$, compute the smallest residue (in absolute value) of $a \bmod p$, where $a = \binom{2n}{n}$. (Example: for $p = 5$, we have $n = 1$ and $\binom{2}{1} = 2 \equiv 2 \bmod 5$.) Compare with the results from the preceding Exercise. Formulate a conjecture and test it for a few more primes.

1.46 a) Given a 5-liter jar and a 3-liter jar and an unlimited supply of water, how do you measure out 4 liters exactly?
b) Can you also measure out 1, 2 and 3 liters?
c) Which quantities can you measure out if you are given a 6-liter and a 9-liter jar?
d) Formulate a general conjecture. Can you prove it (at least partially)?

1.47 Show that a number $n = d_n \ldots d_1 d_0 = d_n \cdot 10^n + \ldots + d_1 \cdot 10 + d_0$ satisfies the congruence $n \equiv d_n + \ldots + d_1 + d_0 \bmod 9$: the residue class modulo 9 of any integer is congruent to the sum of the digits of n.

1.48 Show that a number $d_n \ldots d_1 d_0 = d_n 10^n + \ldots + d_1 \cdot 10 + d_0$ satisfies the congruence $n \equiv (-1)^n d_n + \ldots + d_2 - d_1 + d_0 \bmod 11$.

1.49 Invent a simple method to compute the residue class of $n = d_n \ldots d_1 d_0 = d_n 10^n + \ldots + d_1 \cdot 10 + d_0$ modulo 7.

1.50 Compute the last digit of 7^{100}. Compute the last two digits of 3^{65}.

1.51 Solve the diophantine equation $x^2 + 2y^2 = z^2$.

1.52 Solve the diophantine equation $x^2 - 2y^2 = z^2$.

1.53 Solve the diophantine equation $x^2 + y^2 = 2z^2$.

1.54 Solve the diophantine equation $x^2 - y^2 = 3$.

1.55 Prove that each odd prime p can be written as a difference of squares of natural numbers ($p = y^2 - x^2$ for $x, y \in \mathbb{N}$) in a unique way.

1.56 Fermat repeatedly challenged English mathematicians by sending them problems he claimed to have solved and asking for proofs. Two of them were the following that he sent to Wallis:

- Prove that the only solution of $x^2 + 2 = y^3$ in positive integers is given by $x = 5$ and $y = 3$;
- Prove that the only solution of $x^2 + 4 = y^3$ in positive integers is given by $x = 11$ and $y = 5$.

In a letter to his English colleague Digby, Wallis called these problems trivial and useless, and mentioned a couple of problems that he claimed were of a similar nature:

- $x^2 + 12 = y^4$ has unique solution $x = 2$, $y = 2$ in integers;
- $x^4 + 9 = y^2$ has unique solution $x = 2$, $y = 5$ in integers;
- $x^3 - y^3 = 20$ has no solution in integers;
- $x^3 - y^3 = 19$ has unique solution $x = 3$, $y = 2$ in integers.

When Fermat learned about Wallis's comments, he called Wallis's problems mentioned above "amusements for a three-day arithmetician" in a letter to Digby. In fact, while Fermat's problems were hard (and maybe not even solvable using the mathematics known in his times), Wallis's claims are easy to prove. Do this.

1.57 Prove Prop. 1.14 by induction. To this end, change the statement into the following: Let $m \geq 1$ be an integer, and assume that $a \leq m$ and b are positive coprime integers; if ab is a square, then so are a and b. Now use induction on m and show that if a prime p divides m, then $p^2 \mid m$.

1.58 Assume that $ab = rx^n$ for $a, b, r, x \in \mathbb{N}$ and $\gcd(a, b) = 1$. Show that there exist $u, v, y, z \in \mathbb{N}$ such that $a = uy^n$, $b = vz^n$, and $uv = r$.

1.59 Use infinite descent to prove that $\sqrt{3}$ is irrational.

1.60 Let p be a prime; show that $x^3 + py^3 + p^2z^3 = 0$ does not have a solution.

1.61 The employees of a big company are represented in a computer by 5-digit numbers. A check digit c is introduced to minimize errors. The company uses the formula $c \equiv 1 \cdot d_1 + 2 \cdot d_2 + 3 \cdot d_3 + 4 \cdot d_4 + 5 \cdot d_5 \bmod 10$ to compute the check digit of $d_1d_2d_3d_4d_5$.

1. Compute the check digits of 01716 and 01718. What do you observe? Is the formula for the check digit a good choice? Why (not)?
2. Would the formula $c \equiv d_1 + 3 \cdot d_2 + d_3 + 3 \cdot d_4 + d_5 \bmod 10$ be a better choice? What about $c \equiv d_1 + 2 \cdot d_2 + d_3 + 2 \cdot d_4 + d_5 \bmod 10$?

1.62 Show that the rational points on the unit circle $\mathcal{C} : X^2 + Y^2 = 1$ are dense: given any point $P \in \mathcal{C}(\mathbb{R})$ on the real unit circle and some $\varepsilon > 0$ there is a rational point $Q \in \mathcal{C}(\mathbb{Q})$ with $|P - Q| < \varepsilon$.

1.63 Parametrize the unit sphere: find all rational solutions of the diophantine equation

$$X^2 + Y^2 + Z^2 = 1.$$

Show in addition that there are no rational points on the sphere $X^2 + Y^2 + Z^2 = 7$.

1.64 International Standard Book Numbers. From the 1970s onward books were assigned an ISBN consisting of four parts: the first block specifies the country (or rather the language of the country), the second block gives information about the publishing company, the third about the book within that company, and the last digit is a check digit that is computed as follows: multiply the digits of the ISBN by 1, 2, 3, ..., 10, starting on the left; the check digit is the integer ≤ 10 for which the sum of these products is $\equiv 0 \bmod 11$. The check 'digit' X stands for 10.

1. Show that the check digit of the ISBN 0-387-94225-?. is 4.
2. Show that if you type in an ISBN and make a single error, then the check digit will catch it; thus the ISBN is an example of a 1-error detecting code.

1.65 When the 10-digit ISBNs became too small, the new system ISBN-13 was introduced. The prefix 978 or 979 is common for all books. Then follow the country code, the code for the publishing company, and the actual book number. The last digit z_{13} is a check digit, which is computed from the digits z_1, z_2, ..., z_{12} in such a way that

$$z_1 + 3z_2 + z_3 + 4z_4 + \ldots + 3z_{12} + z_{13} \equiv 0 \bmod 10.$$

Verify the check digit of the book with the ISBN 978-2-7483-5855-1.

1.66 (Poinsot [Poi1845, p. 36]) Use the Euclidean algorithm for showing that $\gcd(ab, m) = 1$ if $\gcd(a, m) = \gcd(b, m) = 1$.
Hint: Apply the Euclidean algorithm to the pair a, n, and then multiply everything through by b.

1.67 Prove, using the Euclidean algorithm, that $\gcd(am, an) = a\gcd(m, n)$. Hint: apply the Euclidean algorithm to the pair (m, n). What can you say about the remainders when you apply the algorithm to (am, an) instead?

1.68 Using the last exercise, prove that if $a \mid mn$ and $\gcd(a, m) = 1$, then $a \mid n$. Hint: we have $\gcd(an, mn) = n$; now observe that a divides both an and mn.

1.69 Let R and S be unital rings. Show that the map $\pi_1 : R \times S \longrightarrow R$ defined by $\pi((r, s)) = r$ (called the projection onto the first coordinate) is a ring homomorphism.
Show that the "embedding" $\iota : R \longrightarrow R \times S$ satisfies $\iota(r_1 + r_2) = \iota(r_1) + \iota(r_2)$ and $\iota(r_1 r_2) = \iota(r_1)\iota(r_2)$, but that ι does not necessarily send the unit element of R to that of $R \times S$.

1.70 Let R and S be unital rings. Show that $(r_1, s_1) \mid (r_2, s_2)$ in $R \times S$ if and only if $r_1 \mid r_2$ in R and $s_1 \mid s_2$ in S.
Show that an element $(r, s) \in R \times S$ is irreducible (prime) if and only if r and s are irreducible (prime) in R and S, respectively.

1.71 Monoids are sets M with a (multiplicatively written) composition $M \times M \longrightarrow M$ satisfying the following properties:
1. multiplication is commutative and associative.
2. M contains a neutral element $1 \in M$ with $1m = m$ for all $m \in M$.

A monoid is called cancellative if $ma = mb$ for $a, b, m \in M$ implies $a = b$.
Show that if R is an arbitrary ring, then $M = R \setminus \{0\}$ is a monoid. Show that if R is a domain, then M is cancellative.
Also show that for $a, b, c \in M$ we have
1. $1 \mid a$ and $a \mid a$;
2. if $a \mid b$ and $b \mid c$, then $a \mid c$;
3. if $a \mid b$ then $a \mid bc$.

1.72 Show that the set $M^\times$ of units (elements dividing 1) in some monoid M is a group.

1.73 Let M be a monoid; then a non-unit $p \in M$ is called
- irreducible if it only has trivial factorizations, i.e. if $p = ab$ for $a, b \in M$ implies that $a \in M^\times$ or $b \in M^\times$.
- prime if $p \mid ab$ for $a, b \in M$ implies that $p \mid a$ or $p \mid b$.

Show that if p and q are irreducible elements with $p \mid q$, then $q = pu$ for some unit $u \in M^\times$.
Also show that in cancellative monoids, primes are irreducible.

1.74 Show that the monoid $\{1, 2, 4, 8, \ldots\}$ consisting of powers of 2 contain only one irreducible element, namely 2.

1.75 Consider the monoid $M = \{1\} \cup 2\mathbb{N} = \{1, 2, 4, 6, 8, \ldots\}$. Show that 1 is a unit, and that the numbers $2n$ are irreducible if and only if n is odd. Show also that M does not contain prime numbers.