UMN Spring 2019 Math 4281 notes

Darij Grinberg

April 12, 2019

Contents

1. Introduction ......................................................... 6
   1.1. Organisation .................................................. 6
   1.2. Literature .................................................... 6
   1.3. The plan ...................................................... 7
   1.4. Motivation: \( n = x^2 + y^2 \) .......................... 8
   1.5. Motivation: Algebraic numbers ............................ 11
   1.6. Motivation: Shamir’s Secret Sharing Scheme ............. 14
       1.6.1. The problem ............................................. 14
       1.6.2. The \( k = 1 \) case .................................... 14
       1.6.3. The \( k = n \) case: what doesn’t work ................. 15
       1.6.4. The XOR operations .................................... 15
       1.6.5. The \( k = n \) case: an answer .......................... 19
       1.6.6. The \( k = 2 \) case ...................................... 20
       1.6.7. The \( k = 3 \) case ...................................... 20

2. Elementary number theory ........................................ 22
   2.1. Notations ..................................................... 22
   2.2. Divisibility .................................................. 23
   2.3. Congruence modulo \( n \) ..................................... 26
   2.4. Chains of congruences ..................................... 30
   2.5. Substitutivity for congruences ............................ 32
   2.6. Division with remainder ................................... 36
   2.7. Even and odd numbers ..................................... 43
   2.8. The floor function ......................................... 43
   2.9. Common divisors, the Euclidean algorithm and the Bezout theorem 45
       2.9.1. Divisors .................................................. 45
       2.9.2. Common divisors ....................................... 46
       2.9.3. Greatest common divisors ............................. 48
       2.9.4. Bezout’s theorem ....................................... 55
2.9.5. First applications of Bezout’s theorem ........................................ 58
2.9.6. gcds of multiple numbers .......................................................... 62
2.9.7. On converses of Bezout’s theorem ............................................. 69
2.10. Coprime integers ........................................................................... 70
   2.10.1. Definition ................................................................................ 70
   2.10.2. Properties of coprime integers ................................................ 72
   2.10.3. An application to sums of powers ............................................ 75
   2.10.4. More properties of gcds and coprimality ................................. 77
2.11. Lowest common multiples ............................................................. 79
2.12. The Chinese remainder theorem (elementary form) ......................... 85
2.13. Primes ............................................................................................ 89
   2.13.1. Definition and the Sieve of Eratosthenes ................................. 89
   2.13.2. Basic properties of primes ....................................................... 92
   2.13.3. Prime factorization I ............................................................... 94
   2.13.4. Permutations ......................................................................... 97
   2.13.5. p-valuations .......................................................................... 100
   2.13.6. Prime factorization II ........................................................... 107
   2.13.7. The canonical factorization .................................................... 109
   2.13.8. Coprimality through prime factors ........................................ 121
   2.13.9. There are infinitely many primes ............................................ 123
2.14. Euler’s totient function (φ-function) ................................................ 124
   2.14.1. Definition and some formulas ................................................ 124
   2.14.2. The totient sum theorem ....................................................... 126
2.15. Fermat, Euler, Wilson .................................................................... 132
   2.15.1. Fermat and Euler: statements ................................................ 132
   2.15.2. Proving Euler and Fermat ...................................................... 134
   2.15.3. The Pigeonhole Principles ...................................................... 137
   2.15.4. Wilson .................................................................................. 138
2.16. The Chinese Remainder Theorem as a bijection ............................... 145
   2.16.1. The bijection \( K_{m,n} \) ............................................................ 145
   2.16.2. Coprime remainders .............................................................. 147
   2.16.3. Proving the formula for \( \phi \) .................................................. 150
2.17. Binomial coefficients ..................................................................... 153
   2.17.1. Definitions and basics ............................................................ 153
   2.17.2. Combinatorial interpretation .................................................. 157
   2.17.3. Binomial formula and Vandermonde convolution .................... 162
   2.17.4. Some divisibilities and congruences ....................................... 168
   2.17.5. Integer-valued polynomials ................................................... 170
2.18. Counting divisors .......................................................................... 172

3. Equivalence relations and residue classes ......................................... 184
3.1. Relations ......................................................................................... 184
3.2. Equivalence relations .................................................................... 186
3.3. Equivalence classes 192
3.3.1. Definition of equivalence classes 192
3.3.2. Basic properties 193
3.3.3. More examples 195
3.3.4. The “is a permutation of” relation on tuples 195
3.3.5. The “is a cyclic rotation of” relation on tuples 196
3.3.6. Definition of the quotient set and the projection map 197
3.4. \( \mathbb{Z}/n \) (“integers modulo \( n \”) 198
3.4.1. Definition of \( \mathbb{Z}/n \) 198
3.4.2. What \( \mathbb{Z}/n \) looks like 199
3.4.3. Making choices that don’t matter: The universal property of quotient sets 204
3.4.4. Projecting from \( \mathbb{Z}/n \) to \( \mathbb{Z}/d \) 208
3.4.5. Addition, subtraction and multiplication in \( \mathbb{Z}/n \) 209
3.4.6. Scaling by \( r \in \mathbb{Z} \) 212
3.4.7. \( k \)-th powers for \( k \in \mathbb{N} \) 214
3.4.8. Rules and properties for the operations 214
3.5. Modular inverses revisited 219
3.6. The Chinese Remainder Theorem as a bijection between residue classes 227
3.7. Substitutivity and chains of congruences revisited 233
3.8. A couple of applications of elementary number theory 234
3.8.1. The RSA cryptosystem 235
3.8.2. Computing using the Chinese Remainder Theorem 238
3.9. Primitive roots: an introduction 240
3.9.1. Definition and examples 240

4. Complex numbers and Gaussian integers 242
4.1. Complex numbers 242
4.1.1. An informal introduction 242
4.1.2. Rigorous definition of the complex numbers 244
4.1.3. Rules for +, − and \cdot 245
4.1.4. Finite sums and finite products 247
4.1.5. Embedding \( \mathbb{R} \) into \( \mathbb{C} \) 248
4.1.6. Inverses and division of complex numbers 250
4.1.7. Powers of complex numbers 254
4.1.8. The Argand diagram 255
4.1.9. Norms and conjugates 259
4.1.10. \( \text{Re} \), \( \text{Im} \) and the \( 2 \times 2 \)-matrix representation 264
4.1.11. The fundamental theorem of algebra 266
4.2. Gaussian integers 266
4.2.1. Definitions and basics 267
4.2.2. Divisibility and congruence 273
4.2.3. Division with remainder 276
4.2.4. Common divisors 279
1. Introduction

This file will contain the notes from the Math 4281 class (“Introduction to Modern Algebra”) I am teaching at UMN in Spring 2019. I will type the first draft directly in the classroom, and subsequently expand it into proper writing. Occasionally, I will also add extra sections not covered in class.

The website of the class is [http://www-users.math.umn.edu/~dgrinber/19s/index.html](http://www-users.math.umn.edu/~dgrinber/19s/index.html); you will find homework sets there.

1.1. Organisation

See the syllabus for the organization of this class and for the homework.

1.2. Literature

Many books have been written about abstract algebra. I have only a passing familiarity with most of them. Some of the “bibles” of the subject (bulky texts covering lots of material) are Dummit/Foote [DumFoo04], Knapp [Knapp16a] and [Knapp16b] (both freely available), van der Waerden [Waerde91a] and [Waerde91b] (one of the oldest texts on modern algebra, thus rather dated, but still as readable as ever).

Of course, any book longer than 200 pages likely goes further than our course will (unless it is full of details or solved exercises or printed in really large letters). Thus, let me recommend some more introductory sources. Siksek’s lecture notes [Siksek15] are a readable introduction that is a lot more amusing than I had ever expected an algebra text to be. Goodman’s free book [Goodma16] combines introductory material with geometric motivation and applications, such as the classification
of regular polyhedra and 2-dimensional crystals. In a sense, it is a great comple-
ment to our ungeometric course. Pinter’s [Pinter10] often gets used in classes like
ours. Armstrong’s notes [Armstr18] cover a significant part of what we do (and he
will likely have notes for a second course written up by the end of this semester).
Childs’s [Childs00] comes the closest to what we are setting out to do here, that is,
give an example-grounded introduction to basic abstract algebra.

Keith Conrad’s blurbs [Conrad*] are not a book, as they only cover selected
topics. But at pretty much every topic they cover, they are one of the best sources
(clear, full of examples, and often going fairly deep). We shall follow one of them
particularly closely: the one on Gaussian integers [ConradG].

We will use some basic linear algebra, all of which can be found in Hefferon’s
book [Heffer17] (but we won’t need all of this book). As far as determinants are
concerned, we will briefly build up their theory; we refer to [Strick13, Section 12
& Appendix B] for proofs (and to [Grinbe15, Chapter 6] for a really detailed and
formal treatment).

This course will begin (after some motivating questions) with a survey of ele-
mentary number theory. This is in itself a deep subject (despite the name) with a
long history (perhaps as old as mathematics), and of course we will just scratch the
surface. Books like [NiZuMo91], [Burton10] and [UspHea39] cover a lot more than
we can do. The Gallier/Quaintance survey [GalQua17] covers a good amount of
basics and more.

We assume that the reader is familiar with the commonplaces of mathematical ar-
gumentation, such as induction (including strong induction), “WLOG” arguments,
proof by contradiction, summation signs (∑) and polynomials (a vague notion of
polynomials will suffice; we will give a precise definition when it becomes nec-
necessary). If not, several texts can be helpful in achieving such familiarity: e.g.,
[LeLeMe18, particularly Chapters 1–5], [Hammac18], [Day16].

I thank the students of the Math 4281 class for discovering and reporting errors
in previous versions of these notes. Some of the discussion of variants of Gaussian
integers (and the occasional correction) is due to Keith Conrad; the discussion of
Gaussian integers itself owes much to his [ConradG].

These notes include some excerpts from [Grinbe16] and slightly rewritten sec-
tions of [Grinbe15].

2019-01-23 lecture

1.3. The plan

The material I am going to cover is mostly standard. However, the order in which
I will go through it is somewhat unusual: I will spend a lot of time studying the
basic examples before defining abstract notions such as “group”, “monoid”, “ring”
and “field”. This way, once I come to these notions, you’ll already have many
examples to work with. (Don’t be fooled by the word “example”: We will prove a
lot about them, much of which is neither straightforward nor easy.)
First, I will show some motivating questions that are easy to state yet require abstract algebra to answer. We will hopefully see their answers by the end of this class. (Some of them can also be answered elementarily, without using abstract algebra, but such answers usually take more work and are harder to find.)

1.4. Motivation: \( n = x^2 + y^2 \)

A perfect square means the square of an integer. Thus, the perfect squares are

\[
0^2 = 0, \quad 1^2 = 1, \quad 2^2 = 4, \quad 3^2 = 9, \quad 4^2 = 16, \quad \ldots
\]

Here is an old problem (first solved by Pierre de Fermat in 1640, but apparently already studied by Diophantus in the 3rd Century):

<table>
<thead>
<tr>
<th>Question 1.4.1. What integers can be written as sums of two perfect squares?</th>
</tr>
</thead>
</table>

For example, 5 can be written in this way, since \( 5 = 2^2 + 1^2 \).

So can 4, since \( 4 = 2^2 + 0^2 \). (Keep in mind that 0 is a perfect square.)

However, 7 cannot be written in this way. In fact, if we had \( 7 = a^2 + b^2 \) for two integers \( a \) and \( b \), then \( a^2 \) and \( b^2 \) would have to be \( \leq 7 \) (since \( a^2 \) and \( b^2 \) are always \( \geq 0 \), no matter what sign \( a \) and \( b \) have); but the only perfect squares that are \( \leq 7 \) are 0, 1, 4, and there is no way to write 7 as a sum of two of these perfect squares (just check all the possibilities).

For a similar but simpler reason, no negative number can be written as a sum of two perfect squares.

We can of course approach Question 1.4.1 using a computer: It is easy to check, for a given integer \( n \), whether \( n \) is a sum of two perfect squares. (Just check all possibilities for \( a \) and \( b \) for the validity of the equation \( n = a^2 + b^2 \). You only need to try \( a \) and \( b \) belonging to \( \{0, 1, \ldots, \lfloor \sqrt{n} \rfloor \} \), where \( \lfloor y \rfloor \) (for a real number \( y \)) denotes the largest integer that is less or equal than \( y \) (also known as “\( y \) rounded down”).) If you do this, you will see that among the first 101 nonnegative integers, the ones that can be written as sums of two perfect squares are precisely

\[
0, 1, 2, 4, 5, 8, 9, 10, 13, 16, 17, 18, 20, 25, 26, 29, \\
32, 34, 36, 37, 40, 41, 45, 49, 50, 52, 53, 58, 61, 64, \\
65, 68, 72, 73, 74, 80, 81, 82, 85, 89, 90, 97, 98, 100.
\]

Having this data, you can look up the sequence in the Online Encyclopedia of Integer Sequences (short OEIS) and see that the sequence of these integers is known as OEIS Sequence A001481. In the “Comments” field, you can read a lot of what is known about it (albeit in telegraphic style).

For example, one of the comments says “Closed under multiplication”. This is short for “if you multiply two entries of the sequence, then the product will again be an entry of the sequence”. In other words, if you multiply two integers that
are sums of two perfect squares, then you get another sum of two perfect squares. Why is this so?

It turns out that there is a “simple” reason for this: the identity

\[(a^2 + b^2)(c^2 + d^2) = (ad + bc)^2 + (ac - bd)^2,\] (1)

which holds for arbitrary reals \(a, b, c, d\) (and thus, in particular, for integers). This is known as the Brahmagupta-Fibonacci identity, and of course can easily be proven by expanding both sides. But how would you come up with such an identity?

If you stare at the above sequence long enough, you may also discover another pattern: An integer of the form \(4k + 3\) with integer \(k\) (that is, an integer that is larger by 3 than a multiple of 4) cannot be written as a sum of two perfect squares. (Thus, 3, 7, 11, 15, 19, 23, \ldots cannot be written in this way.) This does not account for all integers that cannot be written in this way, but it does provide some clues to the answer that we will later see. In order to prove this observation, we shall need basic modular arithmetic (or at least division with remainder); we will see this proof very soon (see Exercise 2.7.2 (c)).

Further questions can be asked. One of them is: Given an integer \(n\), how many ways are there to represent \(n\) as a sum of two perfect squares? This is actually several questions masquerading as one, since it is not so clear what a “way” is. Do \(5 = 1^2 + 2^2\) and \(5 = 2^2 + 1^2\) count as two different ways? What about \(5 = 1^2 + 2^2\) versus \(5 = (-1)^2 + 2^2\) (here, the perfect squares are the same, but do we really want to count the squares or rather the numbers we are squaring?).

Let me formalize the question as follows:

**Question 1.4.2.** Let \(n\) be an integer.

(a) How many pairs \((a, b)\) \(\in \mathbb{N}^2\) are there that satisfy \(n = a^2 + b^2\)? Here, and in the following, \(\mathbb{N}\) denotes the set \(\{0, 1, 2, \ldots\}\) of all nonnegative integers.

(b) How many pairs \((a, b)\) \(\in \mathbb{Z}^2\) are there that satisfy \(n = a^2 + b^2\)? Here, and in the following, \(\mathbb{Z}\) denotes the set \(\{\ldots, -2, -1, 0, 1, 2, \ldots\}\) of all integers.

(c) How do these counts change if we count unordered pairs instead (i.e., count \((a, b)\) and \((b, a)\) as one only)?

Note that when I say “pair”, I always mean “ordered pair” by default, unless I explicitly say “unordered pair”.

Again, a little bit of programming easily yields answers to all three parts of this question for small values of \(n\), and the resulting data can be plugged into the OEIS and yields lots of information.

**First steps toward answering Question 1.4.2** (a) I claim that the number of such pairs is even unless \(n\) is twice a perfect square (i.e., unless \(n = 2m^2\) for some integer \(m\)); in the latter case, this number is odd instead.

Why? Let me define a solution to be a pair \((a, b)\) such that \(n = a^2 + b^2\). So I want to know whether the number of solutions is even or odd. But we have \(a^2 + b^2 = b^2 + a^2\) for all \(a\) and \(b\). Thus, if \((a, b)\) is a solution, then so is \((b, a)\).
Hence, the solutions themselves “come in pairs”, with each solution \((a, b)\) being matched to the solution \((b, a)\), unless there is a solution \((a, b)\) with \(a = b\) (because such a solution would be matched to itself, and thus not form an actual pair). But solutions \((a, b)\) with \(a = b\) are easy to classify: If \(n\) is twice a perfect square, then there is exactly one such solution (namely, \(\sqrt{n/2}, \sqrt{n/2}\)); otherwise there is none (because \(n = a^2 + b^2\) with \(a = b\) leads to \(n = b^2 + b^2 = 2b^2\), which can only happen when \(n\) is twice a perfect square). Since we know that all the other solutions “come in pairs”, we thus conclude that the number of solutions is odd if \(n\) is twice a perfect square and even otherwise. This proves our claim.

Of course, we have not made much headway into Question 1.4.2 knowing whether a number is even or odd is far from knowing the number itself. But I think the argument above was worth showing; similar reasoning is used a lot in algebra.

(b) By reasoning analogous to the one we used in part (a), we can see that the number of such pairs will be divisible by 8 whenever \(n\) is neither a perfect square nor twice a perfect square. Indeed, this relies on the fact that

\[
a^2 + b^2 = b^2 + a^2 = (-a)^2 + b^2 = (-a)^2 + (-a)^2 = a^2 + (-b)^2 = (-b)^2 + a^2
\]

for all \(a\) and \(b\). Thus the pairs \((a, b)\) \(\in\ \mathbb{Z}^2\) that satisfy \(n = a^2 + b^2\) don’t just come in pairs; they come in sets of 8 (namely, each \((a, b)\) comes in a set with \((b, a), (-a, b), (b, -a), (a, -b), (-b, a), (-a, -b)\) and \((-b, -a)\)). These sets of 8 can “degenerate” to smaller sets when some of their elements coincide, but this can only happen when \(n\) is a perfect square (in which case we can have \((a, b) = (-a, b)\) for example) or twice a perfect square (in which case we can have \((a, b) = (b, a)\) or \((a, b) = (-b, -a)\) or other such coincidences). (Check this!)

(c) We can reduce this to parts (a) and (b). Indeed,

- When \(n\) is not twice a perfect square, the number of unordered pairs will be half the number of ordered pairs, since each unordered pair \((u, v)\) corresponds to precisely two ordered pairs \((u, v)\) and \((v, u)\).
- When \(n\) is twice a perfect square, we have

\[
\frac{(\text{the number of unordered pairs})}{2} = \frac{(\text{the number of ordered pairs}) + (\text{the number of pairs with } a = b)}{2}.
\]

Indeed, each unordered pair \((u, v)\) corresponds to precisely two ordered pairs \((u, v)\) and \((v, u)\) unless \(u = v\), in which case it corresponds to only one ordered pair. Thus, if we multiply the number of unordered pairs by 2, then we overcount the number of ordered pairs, because we are counting the pairs \((u, v)\) with \(u = v\) (that is, the pairs with \(a = b\)) twice. So we

\[1\text{In the rest of this argument, “pair” will always mean “pair } (a, b) \text{ satisfying } n = a^2 + b^2\text{“.} \]
get (the number of ordered pairs) + (the number of pairs with \( a = b \)). This proves our above formula.

What is the number of pairs with \( a = b \)? If \( n = 0 \), then it is 1 (and the only such pair is \((0,0)\)). Otherwise, it is 1 if we are counting pairs in \( \mathbb{N}^2 \) (and the only such pair is \((\sqrt{n}/2, \sqrt{n}/2)\)), and is 2 if we are counting pairs in \( \mathbb{Z}^2 \) (and the only two such pairs are \((\sqrt{n}/2, \sqrt{n}/2)\) and \((-\sqrt{n}/2, -\sqrt{n}/2)\)).

Note that sums of squares have a geometric meaning (going back to Pythagoras): Two real numbers \( a \) and \( b \) satisfy \( a^2 + b^2 = n \) (for a given integer \( n \geq 0 \) if and only if the point with Cartesian coordinates \((a, b)\) lies on the circle with center 0 and radius \( \sqrt{n} \). This will actually prove a valuable insight that will lead us to the answers to the above questions.

Just as a teaser: There are formulas for all three parts of Question 1.4.2 in terms of divisors of \( n \) of the forms \( 4k + 1 \) and \( 4k + 3 \). We will see these formulas after we have properly understood the concept of Gaussian integers.

1.5. Motivation: Algebraic numbers

A real number \( z \) is said to be algebraic if there exists a nonzero polynomial \( P \) with rational coefficients such that \( P(z) = 0 \). In other words, a real number \( z \) is algebraic if and only if it is a root of a nonzero polynomial with rational coefficients.

(If you know the complex numbers, you can replace “real” by “complex” in this definition; but we shall only see real numbers in this little motivational section.)

Examples:

- Each rational number \( a \) is algebraic (being a root of the nonzero polynomial \( x - a \) with rational coefficients).
- The number \( \sqrt{2} \) is algebraic (being a root of the nonzero polynomial \( x^2 - 2 \)).
- The number \( \sqrt[3]{5} \) is algebraic (being a root of \( x^3 - 5 \)).
- All the roots of the polynomial \( f(x) := \frac{3}{2}x^4 + 17x^3 - 12x + \frac{9}{4} \) (whatever they are) are algebraic.

Speaking of these roots, what are they? Using a computer, one can show that this polynomial \( f(x) \) has 4 real roots \((-11.269 \ldots, -0.960 \ldots, 0.198 \ldots, 0.697 \ldots)\), which can be written as complicated expressions with radicals (i.e., \( \sqrt[3]{ \text{signs} } \)), though complex numbers appear in these expressions (despite the roots being real!). All this does not matter to the fact that they are algebraic :)

- All the roots of the polynomial \( g(x) := x^7 - x^5 + 1 \) are algebraic.

This polynomial has only one real root. This root cannot be written as an expression with radicals (as can be proven using Galois theory – indeed, the discovery of this theory greatly motivated the development of abstract algebra). Nevertheless, it is algebraic, by definition. (The same holds for the
remaining 6 complex roots of $g$ – we are working with real numbers here only for the sake of familiarity.

- The most famous number that is not algebraic is $\pi$. This is a famous result of Lindemann, but it belongs to analysis, not to algebra, because $\pi$ is not defined algebraically in the first place (it is defined as the length of a curve or as an area of a curved region – but either of these definitions boils down to a limit of a sequence).

- The second most famous number that is not algebraic is Euler’s number $e$ (the basis of the natural logarithm). Again, analysis is needed to define $e$, and thus also to prove its non-algebraicity.

Numbers that are not algebraic are called transcendental. We shall not study them much, since most of them do not come from algebra. Instead, we shall try our hands at the following question:

**Question 1.5.1.**

(a) Is the sum of two (or, more generally, finitely many) algebraic numbers always algebraic?

(b) What if we replace “sum” by “difference” or “product”?

Let me motivate why this is a natural question to ask. The sum of two integers is still an integer; the sum of two rational numbers is still a rational number. These facts are fundamental; without them we could hardly work with integers and rational numbers. If a similar fact would not hold for algebraic numbers, it would mean that the algebraic numbers are not a good “number system” to work in; on a practical level, it would mean that (e.g.) if we defined a function on the set of all algebraic numbers, then we could not plug a sum of algebraic numbers into it.

Attempts at answering **Question 1.5.1 (a)**. Let us try a particularly simple example of a sum of two algebraic numbers: Let $w$ be $\sqrt{2} + \sqrt{3}$. Is $w$ algebraic?

To answer this question affirmatively, we need to find a nonzero polynomial $f(x)$ with rational coefficients that has $w$ as a root.

Just looking at the equality $w = \sqrt{2} + \sqrt{3}$, we cannot directly eyeball such an $f$. The problem, in a sense, is that there are too many (namely, two) square roots in this equality.

However, if we square this equality, then we obtain

$$w^2 = \left(\sqrt{2} + \sqrt{3}\right)^2 = 2 + 2\sqrt{2} \cdot \sqrt{3} + 3 = 5 + 2\sqrt{6},$$

which is an equality with only one square root (a sign of progress). Subtracting 5 from this equality (in order to “isolate” this remaining square root), we obtain $w^2 - 5 = 2\sqrt{6}$. If we now square this equality, then we obtain $(w^2 - 5)^2 = \left(2\sqrt{6}\right)^2 = 24$.

At this point all square roots are gone, and we are left with an equality that contains rational numbers and $w$ only! We can further rewrite it as $(w^2 - 5)^2 - 24 = 0$. Thus,
$w$ is a root of the polynomial $f(x) := (x^2 - 5)^2 - 24 = x^4 - 10x^2 + 1$. This means that $w$ is algebraic (since $f$ is nonzero).

Let us try a more complicated example: Let $z$ be the number $\sqrt{2} + 3\sqrt{2}$. Is $z$ algebraic? The squaring trick no longer works, since squaring $\sqrt{2} + 3\sqrt{2}$ does not reduce the number of radicals (= root signs). Let’s instead try rewriting $z = \sqrt{2} + \sqrt[3]{2}$ as $z - \sqrt{2} = \sqrt[3]{2}$. Cubing this equality, we obtain $(z - \sqrt{2})^3 = 2$. In view of

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

(this is a particular case of the identity $(a - b)^3 = a^3 - 3a^2b + 3ab^2 - b^3$, which is one form of the Binomial Theorem for exponent 3), this rewrites a

$$z^3 - 3z^2\sqrt{2} + 3z\left(\sqrt{2}\right)^2 - \left(\sqrt{2}\right)^3 = 2.$$

This simplifies to

$$z^3 - 3\sqrt{2}z^2 + 6z - 2\sqrt{2} = 2.$$ Let us transform this inequality in such a way that all terms with a $\sqrt{2}$ in them end up on the right hand side while all the remaining terms end up on the left. We thus obtain

$$z^3 + 6z - 2 = \sqrt{2}\left(3z^2 + 2\right).$$

Now, squaring this equality yields

$$\left(z^3 + 6z - 2\right)^2 = 2\left(3z^2 + 2\right)^2.$$ Hence, $z$ is a root of the polynomial

$$g(x) := \left(x^3 + 6x - 2\right)^2 - 2\left(3x^2 + 2\right)^2 = x^6 - 6x^4 - 4x^3 + 12x^2 - 24x - 4.$$ This is a nonzero polynomial with rational coefficients; hence, $z$ is algebraic.

We thus have verified that the sum of two algebraic numbers is algebraic in two cases. What about more complicated cases, such as

$$\sqrt{2} + \sqrt{3} + \sqrt{11}?$$

This is a sum of two algebraic numbers (since we already know that $\sqrt{2} + \sqrt{3} = w$ is algebraic). Is it algebraic? Neither of our above two methods properly works here; do we have to come up with new ad-hoc tricks? 

2019-01-25 lecture
1.6. Motivation: Shamir’s Secret Sharing Scheme

1.6.1. The problem

Adi Shamir is one of the founders of modern mathematical cryptography (famous in particular for the RSA cryptosystem, which we will discuss in Subsection 3.8.1). Shamir’s Secret Sharing Scheme is a way in which a secret \( a \) (a piece of data – e.g., nuclear launch codes) can be distributed among \( n \) people in such a way that

- any \( k \) of them can (if they come together) reconstruct it uniquely, but
- any \( k - 1 \) of them (if they come together) cannot gain any insight about it (i.e., not only cannot they reconstruct it, but they cannot even tell that some values are more likely than others to be \( a \)).

Here \( n \) and \( k \) are fixed positive integers.

Understanding this scheme completely will require some abstract algebra, but we can already start thinking about the problem and get reasonably far.

So we have \( n \) people \( 1, 2, \ldots, n \), a positive integer \( k \in \{1, 2, \ldots, n\} \) and a secret piece of data \( a \). We assume that this data \( a \) is encoded as a bitstring – i.e., a finite sequence of bits. A bit is an element of the set \( \{0, 1\} \). Thus, examples of bitstrings are \( (0,1,1,0) \) and \( (1,0) \) and \( (1,1,0,1,0,0,0) \) as well as the empty sequence \( () \). When writing bitstring, we shall usually omit both the commas and the parentheses; thus, e.g., the bitstring \( (1,1,0,1,0,0,0) \) will become 1101000. Make sure you don’t mistake it for a number. Our goal is to give each of the \( n \) people \( 1, 2, \ldots, n \) some bitstring in such a way that:

- **Requirement 1**: Any \( k \) of the \( n \) people can (if they come together) reconstruct \( a \) uniquely.
- **Requirement 2**: Any \( k - 1 \) of the \( n \) people are unable to gain any insight about \( a \) (even if they collaborate).

We denote the bitstrings given to the people \( 1, 2, \ldots, n \) by \( a_1, a_2, \ldots, a_n \), respectively.

We assume that the length of our secret bitstring \( a \) is known in advance to all parties; i.e., it is not a secret. Thus, when we say “\( k - 1 \) persons cannot gain any insight about \( a \)”, we do not mean that they don’t know the length; and when we say “some values are more likely than others to be \( a \)”, we only mean values that fit this length.

1.6.2. The \( k = 1 \) case

One simple special case of our problem is when \( k = 1 \). In this case, it suffices to give each of the \( n \) people the full secret \( a \) (that is, we set \( a_i = a \) for all \( i \)). Then, Requirement 1 is satisfied (since any 1 of the \( n \) people already knows \( a \)), while Requirement 2 is satisfied as well (0 people know nothing).
1.6.3. The \( k = n \) case: what doesn’t work

Let us now consider the case when \( k = n \). This case will not help us solve the general problem, but it will show some ideas that we will encounter again and again in abstract algebra.

We want to ensure that all \( n \) people needed to reconstruct the secret \( a \), while any \( n - 1 \) of them will be completely clueless.

It sounds reasonable to split \( a \) into \( n \) parts, and give each person one of these parts\(^2\) (i.e., we let \( a_i \) be the \( i \)-th part of \( a \) for each \( i \in \{1, 2, \ldots, n\} \)). This method satisfies Requirement 1 (indeed, all \( n \) people together can reconstruct \( a \) simply by fusing the \( n \) parts back together), but fails Requirement 2 (indeed, any \( n - 1 \) people know \( n - 1 \) parts of the secret \( a \), which is a far from being clueless about \( a \)). So this method doesn’t work. It is not that easy.

1.6.4. The XOR operations

One way to solve the \( k = n \) case is using the XOR operation.

Let us first define some basic language. A binary operation on a set \( S \) is (informally speaking) a function that takes two elements of \( S \) and assigns a new element of \( S \) to them. More formally:

**Definition 1.6.1.** A binary operation on a set \( S \) is a map \( f \) from \( S \times S \) to \( S \). When \( f \) is a binary operation on \( S \) and \( a \) and \( b \) are two elements of \( S \), we shall write \( a \circ b \) for the value \( f(a, b) \).

**Example 1.6.2.** Addition, subtraction and multiplication of integers are three binary operations on the set \( \mathbb{Q} \) (the set of all rational numbers). For example, addition is the map from \( \mathbb{Q} \times \mathbb{Q} \) to \( \mathbb{Q} \) that sends each pair \((a, b) \in \mathbb{Q} \times \mathbb{Q} \) to \( a + b \).

Division is not a binary operation on the set \( \mathbb{Q} \). Indeed, if it was, then it would send the pair \((1, 0) \) to some integer called \( 1/0 \); but there is no such integer.

There are myriad more complicated binary operations around waiting for someone to name them. For example, you could define a binary operation \( \odot \) on the set \( \mathbb{Q} \) by \( a \odot b = \frac{a - b}{1 + a^2 + b^2} \). Indeed, you can do this because \( 1 + a^2 + b^2 \) is always nonzero when \( a, b \in \mathbb{Q} \) (after all, squares are nonnegative, so that \( 1 + a^2 \geq 0 \) and \( b^2 \geq 0 \)). I am not saying that you should...

Now, we define some specific binary operations on the set \( \{0, 1\} \) of all bits, and on the set \( \{0, 1\}^n \) of all length-\( n \) bitstrings (for a given \( n \)).

\(^2\)assuming that \( a \) is long enough for that
Definition 1.6.3. We define a binary operation XOR on the set \{0, 1\} by setting
\[
\begin{align*}
0 \text{ XOR } 0 &= 0, \\
0 \text{ XOR } 1 &= 1, \\
1 \text{ XOR } 0 &= 1, \\
1 \text{ XOR } 1 &= 0.
\end{align*}
\]
This is a valid definition, because there are only four pairs \((a, b) \in \{0, 1\} \times \{0, 1\}\), and we have just defined \(a \text{ XOR } b\) for each of these four options. We can also rewrite this definition as follows:
\[
a \text{ XOR } b = \begin{cases} 
1, & \text{if } a \neq b; \\
0, & \text{if } a = b
\end{cases} = \begin{cases} 
1, & \text{if exactly one of } a \text{ and } b \text{ is } 1; \\
0, & \text{otherwise}.
\end{cases}
\]
For lack of a better name, we refer to \(a \text{ XOR } b\) as the “XOR of \(a\) and \(b\)”.

The name “XOR” is short for “exclusive or”. In fact, if you identify bits with boolean truth values (so the bit 0 stands for “False” and the bit 1 stands for “True”), then \(a \text{ XOR } b\) is precisely the truth value for “exactly one of \(a\) and \(b\) is True”, which is also known as “\(a\) exclusive-or \(b\)”.

Definition 1.6.4. Let \(m\) be a nonnegative integer. We define a binary operation XOR on the set \(\{0, 1\}^m\) (this is the set of all length-\(m\) bitstrings) by
\[
(a_1, a_2, \ldots, a_m) \text{ XOR } (b_1, b_2, \ldots, b_m) = (a_1 \text{ XOR } b_1, a_2 \text{ XOR } b_2, \ldots, a_m \text{ XOR } b_m).
\]
In other words, if \(a\) and \(b\) are two length-\(m\) bitstrings, then \(a \text{ XOR } b\) is obtained by taking the XOR of each entry of \(a\) with the corresponding entry of \(b\), and packing these \(m\) XORs into a new length-\(m\) bitstring.

For example,
\[
\begin{align*}
(1001) \text{ XOR } (1100) &= 0101; \\
(11011) \text{ XOR } (10101) &= 01110; \\
(11010) \text{ XOR } (01011) &= 10001; \\
(1) \text{ XOR } (0) &= 1; \\
(\ ) \text{ XOR } (\ ) &= (\ ).
\end{align*}
\]
Note that if \(a\) and \(b\) are two length-\(m\) bitstrings, then the 0’s in the bitstring \(a \text{ XOR } b\) are at the positions where \(a\) and \(b\) have equal entries, and the 1’s in \(a \text{ XOR } b\) are at the positions where \(a\) and \(b\) have different entries. Thus, \(a \text{ XOR } b\) essentially pinpoints the differences between \(a\) and \(b\).
We observe the following simple properties of these operations XOR on bits and on bitstrings:

- We have \( a \oplus 0 = a \) for any bit \( a \). (This can be trivially checked by considering both possibilities for \( a \).)
- Thus, \( a \oplus 0 = a \) for any bitstring \( a \), where \( 0 \) denotes the bitstring \( 00 \cdots 0 = (0, 0, \ldots, 0) \) (of appropriate length – i.e., of the same length as \( a \)).
- We have \( a \oplus a = 0 \) for any bit \( a \). (This can be trivially checked by considering both possibilities for \( a \).)
- Thus, \( a \oplus a = 0 \) for any bitstring \( a \). We shall refer to this as the self-cancellation law.
- We have \( a \oplus b = b \oplus a \) for any bits \( a, b \). (Again, this is easy to check by going through all four options for \( a \) and \( b \).)
- Thus, \( a \oplus b = b \oplus a \) for any bitstrings \( a, b \).
- We have \( a \oplus (b \oplus c) = (a \oplus b) \oplus c \) for any bits \( a, b, c \). (Again, this is easy to check by going through all eight options for \( a, b, c \).)
- Thus, \( a \oplus (b \oplus c) = (a \oplus b) \oplus c \) for any bitstrings \( a, b, c \).
- Thus, for any bitstrings \( a \) and \( b \), we have
  \[
  (a \oplus b) \oplus b = a \oplus \underbrace{b \oplus b}_0 = a \oplus 0 = a.
  \]
  (by the self-cancellation law)

This observation gives rise to a primitive cryptosystem (known as a one-time pad): If you have a secret bitstring \( a \) that you want to encrypt, and another secret bitstring \( b \) that can be used as a key, then you can encrypt \( a \) by XORing it with \( b \) (that is, you transform it into \( a \oplus b \)). Then, you can decrypt it again by XORing it with \( b \) again; indeed, if you do this, you will obtain \( (a \oplus b) \oplus b = a \). This is a highly safe cryptosystem as long as you can safely communicate the key \( b \) to whomever needs to be able to decrypt (or encrypt) your secrets, and as long as you are able to generate uniformly random keys \( b \) of sufficient length. Its only weakness is its impracticality (in many situations): If the secret you want to encrypt is long (say, a whole book), your key will need to be equally long. Even storing such keys can become difficult.

\[3\] As a mnemonic, we shall try to use boldfaced letters like \( a \) and \( b \) for bitstrings and regular italic letters like \( a \) and \( b \) for single bits.
We shall refer to the properties $a \text{ XOR } b = b \text{ XOR } a$ and $a \text{ XOR } b = b \text{ XOR } a$ as laws of commutativity, and we shall refer to the properties $a \text{ XOR } (b \text{ XOR } c) = (a \text{ XOR } b) \text{ XOR } c$ and $a \text{ XOR } (b \text{ XOR } c) = (a \text{ XOR } b) \text{ XOR } c$ as laws of associativity. These are, of course, similar to well-known facts like $\alpha + \beta = \beta + \alpha$ and $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ for numbers $\alpha, \beta, \gamma$ (which is why we are giving them the same names). This similarity is not coincidental. Just as for addition or multiplication of numbers, these laws lead to a notion of “XOR-products”:

**Proposition 1.6.5.** Let $m$ be a positive integer. Let $a_1, a_2, \ldots, a_m$ be $m$ bitstrings. Then, the “XOR-product” expression

$$a_1 \text{ XOR } a_2 \text{ XOR } a_3 \text{ XOR } \cdots \text{ XOR } a_m$$

is well-defined, in the sense that it does not depend on the parenthesization.

What do we mean by “parenthesization”? To clarify things, let us set $m = 4$. In this case, we want to make sense of the expression $a_1 \text{ XOR } a_2 \text{ XOR } a_3 \text{ XOR } a_4$. This expression does not make sense a priori, since it is a XOR of four bitstrings, whereas we have defined only the XOR of two bitstrings. But there are five ways to put parentheses around some of its sub-expressions such that the expression becomes meaningful:

$$(a_1 \text{ XOR } a_2) \text{ XOR } (a_3 \text{ XOR } a_4),$$
$$(a_1 \text{ XOR } a_2) \text{ XOR } a_3 \text{ XOR } a_4,$$
$$a_1 \text{ XOR } (a_2 \text{ XOR } a_3) \text{ XOR } a_4,$$
$$a_1 \text{ XOR } (a_2 \text{ XOR } a_3),$$
$$(a_1 \text{ XOR } a_2) \text{ XOR } a_3 \text{ XOR } a_4.$$

Each of these five parenthesizations (= placements of parentheses) turns our expression $a_1 \text{ XOR } a_2 \text{ XOR } a_3 \text{ XOR } a_4$ into a combination of XOR’s of two bitstrings each, and thus gives it meaning. The question is: Do these five parenthesizations give it the same meaning?

Well, let us calculate:

$$(a_1 \text{ XOR } a_2) \text{ XOR } (a_3 \text{ XOR } a_4)$$

$$= a_1 \text{ XOR } (a_2 \text{ XOR } (a_3 \text{ XOR } a_4))$$

$$= (a_2 \text{ XOR } a_3) \text{ XOR } a_4$$

$$= a_1 \text{ XOR } ((a_2 \text{ XOR } a_3) \text{ XOR } a_4)$$

$$= (a_1 \text{ XOR } (a_2 \text{ XOR } a_3)) \text{ XOR } a_4$$

$$= (a_1 \text{ XOR } a_2) \text{ XOR } a_3 \text{ XOR } a_4,$$

where we used the law of associativity in each step. This shows that our five parenthesizations yield the same result. Thus, they all give our “XOR-product”
expression $a_1 \text{ XOR } a_2 \text{ XOR } a_3 \text{ XOR } a_4$ the same meaning; so we can say that this expression is well-defined. This confirms Proposition [1.6.5] for $m = 4$.

Of course, proving Proposition [1.6.5] is less simple. Such a proof will appear in Exercise 4 on homework set #0.

1.6.5. The $k = n$ case: an answer

Let us now return to our problem. We have $n$ persons $1, 2, \ldots, n$ and a secret $a$ (encoded as a bitstring). We want to give each person $i$ some bitstring $a_i$ such that only all $n$ of them can recover $a$ but any $n - 1$ of them cannot gain any insight about $a$.

We let $a_1, a_2, \ldots, a_{n-1}$ be $n - 1$ uniformly random bitstrings of the same length as $a$. (Think of them as random gibberish.) Set

$$a_n = a \text{ XOR } a_1 \text{ XOR } a_2 \text{ XOR } \cdots \text{ XOR } a_{n-1}.$$  

(This expression makes sense because of Proposition [1.6.5])

Then,

$$a_n \text{ XOR } a_{n-1} \text{ XOR } a_{n-2} \text{ XOR } \cdots \text{ XOR } a_1$$

$$= (a \text{ XOR } a_1 \text{ XOR } a_2 \text{ XOR } \cdots \text{ XOR } a_{n-1}) \text{ XOR } a_{n-1} \text{ XOR } a_{n-2} \text{ XOR } \cdots \text{ XOR } a_1$$

$$= a \text{ XOR } a_1 \text{ XOR } a_2 \text{ XOR } \cdots \text{ XOR } a_{n-2} \text{ XOR } 0 \text{ XOR } a_{n-2} \text{ XOR } \cdots \text{ XOR } a_1$$

$$= a \text{ XOR } a_1 \text{ XOR } a_2 \text{ XOR } \cdots \text{ XOR } a_{n-2} \text{ XOR } 0 \text{ XOR } a_{n-2} \cdots \text{ XOR } a_1$$

$$= \cdots$$

$$= a$$

(here, we have been unravelling the big XOR-product from the middle on, by cancelling equal bitstrings using the self-cancellation law and then removing the resulting 0 using the $a \text{ XOR } 0 = a$ law). Hence, the $n$ people together can decrypt the secret $a$.

Can $n - 1$ people gain any insight about it? The $n - 1$ people $1, 2, \ldots, n - 1$ certainly cannot, since all they know are the random bitstrings $a_1, a_2, \ldots, a_{n-1}$. But the $n - 1$ people $2, 3, \ldots, n$ cannot gain any insight about $a$ either: In fact, all they know are the random bitstrings $a_2, a_3, \ldots, a_{n-1}$ and the bitstring

$$a_n = a \text{ XOR } a_1 \text{ XOR } a_2 \text{ XOR } \cdots \text{ XOR } a_{n-1};$$

therefore, all the information they have about $a$ and $a_1$ comes to them through $a \text{ XOR } a_1$, which says nothing about $a$ as long as they know nothing about $a_1$.

(We used a bit of handwaving in this argument, but then again we never formally
defined what it means to “gain no insight”; this is done in courses on cryptography and information theory.) Similar arguments show that any other choice of \( n - 1 \) persons remains equally clueless about \( a \). So we have solved the problem in the case \( k = n \).

1.6.6. The \( k = 2 \) case

The next simple case is when \( k = 2 \). So we want to ensure that any 2 of our \( n \) people can together recover the secret, but no 1 person can learn anything about it alone.

A really nice approach was suggested by Nathan in class: We pick \( n \) random bitstrings \( x_1, x_2, \ldots, x_{n-1} \) of the same length as \( a \). Set

\[
x_n = a \text{ XOR } x_1 \text{ XOR } x_2 \text{ XOR } \cdots \text{ XOR } x_{n-1};
\]

thus, as in the \( k = n \) case, we have

\[
x_n \text{ XOR } x_{n-1} \text{ XOR } x_{n-2} \text{ XOR } \cdots \text{ XOR } x_1 = a.
\]  

(2)

Each person \( i \) now receives the bitstring

\[
a_i = x_1 x_2 \cdots x_{i-1} x_{i+1} x_{i+2} \cdots x_n,
\]

where the product stands for concatenation (i.e., the bitstring \( a_i \) is formed by writing down all of the bitstrings \( x_1, x_2, \ldots, x_n \) one after the other but skipping \( x_i \)). Thus, each person \( i \) can recover all the \( n - 1 \) bitstrings \( x_1, x_2, \ldots, x_{i-1}, x_{i+1}, x_{i+2}, \ldots, x_n \) (because their lengths are the length of \( a \), which is known), but knows nothing about \( x_i \) (his “blind spot”). Hence, 2 people together can recover all the \( n \) bitstrings \( x_1, x_2, \ldots, x_n \) and therefore recover the secret \( a \) (by (2)). On the other hand, each single person has no insight about \( a \) (this is proven similarly to the \( k = n \) case). So again, the problem is solved in this case.

1.6.7. The \( k = 3 \) case

Now, let us come to the case when \( k = 3 \). Now I think the usefulness of the XOR approach has come to its end: at least I don’t know how to make it work here. Instead, out of the blue, I will invoke something completely different: polynomials (let’s say with rational coefficients).

Recall a fact you might have heard in high school: A polynomial \( p(x) = cx^2 + bx + a \) of degree \( \leq 2 \) is uniquely determined by any three of its values. More precisely: If \( u, v, w \) are three fixed distinct numbers, then a polynomial \( p(x) = cx^2 + bx + a \) of degree \( \leq 2 \) is uniquely determined by the values \( p(u), p(v), p(w) \). We will put this to use now, and sort-of solve the problem.

Also recall that any bitstring of given length \( N \) can be encoded as an integer in \( \{0, 1, \ldots, 2^N - 1\} \); just read it as a number in binary. More precisely, any bitstring \( a_{N-1} a_{N-2} \cdots a_0 \) of length \( N \) becomes the integer \( a_{N-1} \cdot 2^{N-1} + a_{N-2} \cdot 2^{N-2} + \cdots + \)
\(a_0 \cdot 2^0 \in \{0, 1, \ldots, 2^N - 1\}\). For example, the bitstring 010110 of length 6 becomes the integer
\[
0 \cdot 2^5 + 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0 = 22 \in \{0, 1, \ldots, 2^6 - 1\}.
\]

Choose two uniformly random bitstrings \(c\) and \(b\) (of the same length as \(a\)) and encode them as numbers \(c\) and \(b\) (as just explained). Encode the secret \(a\) as a number as well (in the same way). Define the polynomial \(p(x) = cx^2 + bx + a\). Reveal to each person \(i \in \{1, 2, \ldots, n\}\) the value \(p(i)\) – or, rather, a bitstring that encodes it in binary – as \(a_i\).

As we know, any three of the values \(p(i)\) uniquely determine the polynomial \(p\). Thus, any three people can use their bitstrings \(a_i\) to recover three values \(p(i)\) and therefore \(p\) and therefore \(a\) (as the constant term of \(p\)) and therefore \(a\) (by decoding \(a\)). So our method satisfies Requirement 1.

Now, let us see whether it satisfies Requirement 2. Any 2 people can recover two values \(p(i)\), which generally do not determine \(p\) uniquely. It is not hard to show that they do not even determine \(a\) uniquely; thus, they do not determine \(a\) uniquely. What’s better: If you know just two values of \(p\), there are infinitely many possible choices for \(p\), and all of them have distinct constant terms (unless one of the two values you know is \(p(0)\), which of course pins down the constant term). So we get infinitely many possible values for \(a\), and thus infinitely many possible values for \(a\). This means that our 2 people don’t gain any insight about \(a\), right?

Not so fast! We cannot really have “infinitely many possible values for \(a\)”, since \(a\) is bound to be a bitstring of a given length – there are only finitely many of those! You can only get infinitely many possible values for \(p\) if you forget how \(p\) was constructed (from \(c\) and \(b\) and \(a\)) and pretend that \(p\) is just a “uniformly random” polynomial (whatever this means). But no one can force the 2 people to do this; it is certainly not in their interest! Here are some things they might do with this knowledge:

- Let \(N\) be the length of \(a\) (which, as we said, is known). Thus, \(c\) and \(b\) are bitstrings of length \(N\), so that \(c\) and \(b\) are integers in \(\{0, 1, \ldots, 2^N - 1\}\). Assume that one of the 2 people is person 2. Now, person 2 knows \(p(2) = c2^2 + b2 + a = 4c + 2b + a\), and thus knows whether \(a\) is even or odd (because \(a\) is even resp. odd if and only if \(4c + 2b + a\) is even resp. odd). This means she knows the last bit of the secret \(a\). This is not “clueless”.

- You might try to fix this by picking \(c\) and \(b\) to be uniformly random rational numbers instead (rather than using uniformly random bitstrings \(c\) and \(b\)). Unfortunately, there is no such thing as a “uniformly random rational number” (in the sense that, e.g., larger numbers aren’t less likely to be picked than smaller numbers). Any probability distribution will make some numbers more likely than others, and this will usually cause information about \(a\) to “leak”. For example, if \(c\) and \(b\) are chosen from the interval \([0, 2^N - 1]\),
then person 1’s knowledge of $p(1) = c1^2 + b1 + a = c + b + a$ will sometimes reveal to person 1 that $a \geq 0.5 \cdot (2^N - 1)$ (namely, this will happen when $p(1) \geq 2.5 \cdot (2^N - 1)$, which occasionally happens). This, again, is nontrivial information about the secret $a$, which a single person (or even two people) should not be having.

So we cannot make Requirement 2 hold, and the culprit is that there are too many numbers (namely, infinitely many). What would help is a finite “number system” in which we can add, subtract, multiply and divide (so that we can define polynomials over it, and a polynomial of degree $\leq 2$ is still uniquely determined by any 3 values). Assuming that this “number system” is large enough that we can encode bitstrings using “numbers” of this system (instead of integers), we can then play the above game using this “number system” and obtain actually uniformly random numbers.

It turns out that such “number systems” exist. They are called finite fields, and we will construct them later in this course.

Assuming that they can be constructed, we thus obtain a method of solving the problem for $k = 3$. A similar method works for arbitrary $k$, using polynomials of degree $\leq k - 1$. This is called Shamir’s secret sharing scheme.

2019-01-30 lecture (virtual)

2. Elementary number theory

Let us now begin a systematic introduction to algebra. We start with studying integers and their divisibility properties – the beginnings of number theory. Part of these will be used directly in what will follow; part of these will inspire more general results and proofs.

2.1. Notations

**Definition 2.1.1.** Let $\mathbb{N} = \{0, 1, 2, \ldots \}$ be the set of nonnegative integers.

Let $\mathbb{P} = \{1, 2, 3, \ldots \}$ be the set of positive integers.

Let $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots \}$ be the set of integers.

Let $\mathbb{Q}$ be the set of rational numbers.

Let $\mathbb{R}$ be the set of real numbers.

Be careful with the notation $\mathbb{N}$: While I use it for $\{0, 1, 2, \ldots \}$, various other authors use it for $\{1, 2, 3, \ldots \}$ instead. There is no consensus in sight on what $\mathbb{N}$ should mean.

Same holds for the word “natural number” (which I will avoid): It means “element of $\mathbb{N}$”, so again its ultimate meaning depends on the author.
2.2. Divisibility

We now go through the basics of divisibility of integers.

**Definition 2.2.1.** Let $a$ and $b$ be two integers. We say that $a \mid b$ (or “$a$ divides $b$” or “$b$ is divisible by $a$” or “$b$ is a multiple of $a$”) if there exists an integer $c$ such that $b = ac$.

We furthermore say that $a \nmid b$ if $a$ does not divide $b$.

Some authors define the “divisibility” relation a bit differently, in that they forbid $a = 0$. From the viewpoint of abstract algebra, this feels like an unnecessary exception, so we don’t follow them.

**Example 2.2.2.** (a) We have $4 \mid 12$, since $12 = 4 \cdot 3$.

(b) We have $a \mid 0$ for any $a \in \mathbb{Z}$, since $0 = a \cdot 0$.

(c) An integer $b$ satisfies $0 \mid b$ only when $b = 0$, since $0 \mid b$ implies $b = 0c = 0$ (for some $c \in \mathbb{Z}$).

(d) We have $a \mid a$ for any $a \in \mathbb{Z}$, since $a = a \cdot 1$.

(e) We have $1 \mid b$ for each $b \in \mathbb{Z}$, since $b = 1 \cdot b$.

I apologize in advance for the next proposition, in which vertical bars stand both for the “divides” relation and for the absolute value of a number. Unfortunately, both of these uses are standard notation. Confusion is possible, but hopefully will not happen often.

**Proposition 2.2.3.** Let $a$ and $b$ be two integers.

(a) We have $a \mid b$ if and only if $|a| \mid |b|$. (Here, “$|a| \mid |b|$” means “$|a|$ divides $|b|$”.)

(b) If $a \mid b$ and $b \neq 0$, then $|a| \leq |b|$.

(c) Assume that $a \neq 0$. Then, $a \mid b$ if and only if $\frac{b}{a} \in \mathbb{Z}$.

Before we prove this proposition, let us recall a well-known fact: We have

$$|xy| = |x| \cdot |y|$$

for any two integers $x$ and $y$. (This can be easily proven by case distinction: $x$ is either nonnegative or negative, and so is $y$.)

---

4Unfortunately, the use of vertical bars for absolute values alone suffices to generate confusion! Just think of the meaning of “$|a| \mid |b|c$” when $a$, $b$ and $c$ are three numbers. Does it stand for “$(|a|) \cdot b \cdot (|c|)$” (where I am using parentheses to make the ambiguity disappear) or for “$(|a| \cdot |b| \cdot |c|)$”? If you see any expressions in my notes that allow for more than one meaningful interpretation, please let me know!

5or real numbers
Proof of Proposition 2.2.3. (a) \(\Rightarrow\) Assume that \(a \mid b\). Thus, there exists an integer \(d\) such that \(b = ad\) (by Definition 2.2.1). Consider this \(d\). We have \(b = ad\) and thus
\[
|b| = |ad| = |a| \cdot |d| \quad \text{(by (3))}.
\]
Thus, there exists an integer \(c\) such that \(|b| = |a| \cdot c\) (namely, \(c = |d|\)). In other words, \(|a| \mid |b|\). This proves the “\(\Rightarrow\)” direction of Proposition 2.2.3 (a).

\(\Leftarrow\): Assume that \(|a| \mid |b|\). Thus, there exists an integer \(f\) such that \(|b| = |a| \cdot f\) (by Definition 2.2.1). Consider this \(f\).

The definition of \(|b|\) shows that \(|b|\) equals either \(b\) or \(-b\). In other words, \(|b|\) equals either 1\(b\) or \(-1\)\(b\) (since \(b = 1b\) and \(-b = (-1)b\)). In other words, \(|b| = qb\) for some \(q \in \{1, -1\}\). Similarly, \(|a| = ra\) for some \(r \in \{1, -1\}\). Consider these \(q\) and \(r\).

From \(q \in \{1, -1\}\), we obtain \(q^2 \in \left\{ \frac{1}{1}, (-1)^2 \right\} = \{1, 1\} = \{1\}\). In other words, \(q^2 = 1\).

Now, \(qb = qg\) \(\iff \quad b = q \mid b\), so that \(b = q \mid b\) \(\implies |a| \cdot f = qra \cdot f = a \cdot qfr\).

Hence, there exists an integer \(c\) such that \(b = ac\) (namely, \(c = qfr\)). In other words, \(a \mid b\). This proves the “\(\Leftarrow\)” direction of Proposition 2.2.3 (a).

Thus, the proof of Proposition 2.2.3 (a) is complete.

(b) Assume that \(a \mid b\) and \(b \neq 0\).

From \(a \mid b\), we conclude that there exists an integer \(c\) such that \(b = ac\). Consider this \(c\). We have \(ac = b \neq 0\), thus \(c \neq 0\). Hence, \(|c| > 0\), and thus \(|c| \geq 1\) (since \(|c|\) is an integer). We can multiply this inequality by \(|a|\) (since \(|a| \geq 0\), and obtain \(|a| \cdot |c| \geq |a| \cdot 1 = |a|\).

From \(b = ac\), we obtain \(|b| = |ac| = |a| \cdot |c|\) (by (3)). Hence, \(|b| = |a| \cdot |c| \geq |a|\).

This proves Proposition 2.2.3 (b).

(c) \(\Longrightarrow\): Assume that \(a \mid b\). We must prove that \(\frac{b}{a} \in \mathbb{Z}\).

We have \(a \mid b\). In other words, there exists an integer \(d\) such that \(b = ad\). Consider this \(d\). We can divide the equality \(b = ad\) by \(a\) (since \(a \neq 0\)), and thus obtain \(\frac{b}{a} = d \in \mathbb{Z}\). This proves the “\(\Longrightarrow\)” direction of Proposition 2.2.3 (c).

\(\iff\): Assume that \(\frac{b}{a} \in \mathbb{Z}\). We must prove that \(a \mid b\).

---

6If you are unfamiliar with the shorthand notation “:\(\Longrightarrow\)”, let me explain it. Our goal is to prove that \(a \mid b\) if and only if \(|a| \mid |b|\). In other words, we need to prove the equivalence \((a \mid b) \iff (|a| \mid |b|)\). In order to prove this equivalence, it suffices to prove the two implications \((a \mid b) \implies (|a| \mid |b|)\) (called the “forward implication” or the “\(\Longrightarrow\)” direction of the equivalence) and \((a \mid b) \impliedby (|a| \mid |b|)\) (called the “backward implication” or the “\(\iff\)” direction”). The shorthand “:\(\Longrightarrow\)” simply marks the beginning of the proof of the forward implication; similarly, the symbol “:\(\iff\)” heralds in the proof of the backward implication.

7Me saying “Consider this \(d\)” means that I am picking some integer \(d\) such that \(b = ad\) (this can be done, since we have just proven that such a \(d\) exists), and will be referring to it as \(d\) from now on.
We have $\frac{b}{a} \in \mathbb{Z}$ and $b = a \cdot \frac{b}{a}$. Thus, there exists an integer $c$ such that $b = ac$ (namely, $c = \frac{b}{a}$). In other words, $a \mid b$. This proves the “$\iff$” direction of Proposition 2.2.3 (c). Hence, the proof of Proposition 2.2.3 (c) is complete. \(\square\)

Proposition 2.2.3 (a) shows that both $a$ and $b$ in the statement “$a \mid b$” can be replaced by their absolute values. Thus, when we talk about divisibility of integers, the sign of the integers does not really matter – it usually suffices to work with nonnegative integers. We will often use this (tacitly, after a couple times) in proofs.

The next proposition shows some basic properties of the divisibility relation:

**Proposition 2.2.4. (a)** We have $a \mid a$ for every $a \in \mathbb{Z}$. (This is called the **reflexivity** of divisibility.)

(b) If $a, b, c \in \mathbb{Z}$ satisfy $a \mid b$ and $b \mid c$, then $a \mid c$. (This is called the **transitivity** of divisibility.)

(c) If $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ satisfy $a_1 \mid b_1$ and $a_2 \mid b_2$, then $a_1a_2 \mid b_1b_2$.

**Proof.** (a) Let $a \in \mathbb{Z}$. Then, there exists an integer $c$ such that $a = ac$ (namely, $c = 1$). In other words, $a \mid a$. This proves Proposition 2.2.4 (a).

(b) Let $a, b, c \in \mathbb{Z}$ satisfy $a \mid b$ and $b \mid c$. From $a \mid b$, we conclude that there exists an integer $d$ such that $b = ad$. Consider this $d$.

From $b \mid c$, we conclude that there exists an integer $e$ such that $c = be$. Consider this $e$.

We have $c = \overbrace{b}^{=ad} \overbrace{e}^{=de} = ade$. Hence, there exists an integer $f$ such that $c = af$ (namely, $f = de$). In other words, $a \mid c$ (by Definition 2.2.1). This proves Proposition 2.2.4 (b).

(c) Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ satisfy $a_1 \mid b_1$ and $a_2 \mid b_2$.

From $a_1 \mid b_1$, we conclude that there exists an integer $d$ such that $b_1 = a_1d$. Consider this $d$.

From $a_2 \mid b_2$, we conclude that there exists an integer $e$ such that $b_2 = a_2e$. Consider this $e$.

We have $\overbrace{b_1}^{=a_1d} \overbrace{b_2}^{=a_2e} = a_1da_2e = a_1a_2de$. Hence, there exists an integer $f$ such that $b_1b_2 = a_1a_2f$ (namely, $f = de$). In other words, $a_1a_2 \mid b_1b_2$ (by Definition 2.2.1). This proves Proposition 2.2.4 (c). \(\square\)

**Exercise 2.2.1.** Let $a \in \mathbb{Z}$.

(a) Prove that $a \mid |a|$. (This means “$a$ divides $|a|$”.)

(b) Prove that $|a| \mid a$. (This means “$|a|$ divides $a$”.)

**Exercise 2.2.2.** Let $a$ and $b$ be two integers such that $a \mid b$ and $b \mid a$. Prove that $|a| = |b|$. 
Exercise 2.2.3. Let $a, b, c$ be three integers such that $c \neq 0$. Prove that $a \mid b$ holds if and only if $ac \mid bc$.

Exercise 2.2.4. Let $n \in \mathbb{Z}$. Let $a, b \in \mathbb{N}$ be such that $a \leq b$. Prove that $n^a \mid n^b$.

Exercise 2.2.5. Let $g$ be a nonnegative integer such that $g \mid 1$. Prove that $g = 1$.

Exercise 2.2.6. Let $a, b \in \mathbb{Z}$ be such that $a \mid b$. Let $k \in \mathbb{N}$. Prove that $a^k \mid b^k$.

2.3. Congruence modulo $n$

The next definition is simple but crucial:

**Definition 2.3.1.** Let $n, a, b \in \mathbb{Z}$. We say that $a$ is congruent to $b$ modulo $n$ if and only if $n \mid a - b$. We shall use the notation “$a \equiv b \mod n$” for “$a$ is congruent to $b$ modulo $n$”.

We furthermore shall use the notation “$a \not\equiv b \mod n$” for “$a$ is not congruent to $b$ modulo $n$”.

**Example 2.3.2.** (a) Is $3 \equiv 7 \mod 2$? Yes, since $2 \mid 3 - 7 = -4$.
(b) Is $3 \equiv 6 \mod 2$? No, since $2 \nmid 3 - 6 = -3$. So we have $3 \not\equiv 6 \mod 2$.
Now, let $a$ and $b$ be two integers.
(c) We have $a \equiv b \mod 0$ if and only if $a = b$. (Indeed, $a \equiv b \mod 0$ is defined to mean $0 \mid a - b$, but the latter divisibility happens only when $a - b = 0$, which is tantamount to saying $a = b$.)
(d) We have $a \equiv b \mod 1$ always, since $1 \mid a - b$ always holds (remember: $1$ divides everything).

Note that being congruent modulo 2 means having the same parity: i.e., two even numbers will be congruent modulo 2, and two odd numbers will be, but an even number will never be congruent to an odd number modulo 2. (To be rigorous: This is not quite obvious at this point yet; but it will be easy once we have properly introduced division with remainder. See Exercise 2.7.1(i) below for the proof.)

The word "modulo" in the phrase “$a$ is congruent to $b$ modulo $n$” is due to Gauss and means something like “with respect to”. You should think of “$a$ is congruent to $b$ modulo $n$” as a relation between all three of the numbers $a$, $b$ and $n$, but $a$ and $b$ are the “main characters” and $n$ sets the scene.

**Exercise 2.3.1.** Let $a, b \in \mathbb{Z}$. Prove that $a + b \equiv a - b \mod 2$.

We begin with a proposition so fundamental that we will always use it without saying:
Proposition 2.3.3. Let \( n \in \mathbb{Z} \) and \( a \in \mathbb{Z} \). Then, \( a \equiv 0 \mod n \) if and only if \( n \mid a \).

Proof of Proposition 2.3.3. We have the following chain of equivalences:

\[
(a \equiv 0 \mod n) \iff (n \mid a - 0) \quad \text{(by Definition 2.3.1)} \\
\iff (n \mid a) \quad \text{(since } a - 0 = a) .
\]

This proves Proposition 2.3.3. \qed

Next come some staple properties of congruences:

Proposition 2.3.4. Let \( n \in \mathbb{Z} \).

(a) We have \( a \equiv a \mod n \) for every \( a \in \mathbb{Z} \).

(b) If \( a, b, c \in \mathbb{Z} \) satisfy \( a \equiv b \mod n \) and \( b \equiv c \mod n \), then \( a \equiv c \mod n \).

(c) If \( a, b \in \mathbb{Z} \) satisfy \( a \equiv b \mod n \), then \( b \equiv a \mod n \).

(d) If \( a_1, a_2, b_1, b_2 \in \mathbb{Z} \) satisfy \( a_1 \equiv b_1 \mod n \) and \( a_2 \equiv b_2 \mod n \), then

\[
\begin{align*}
a_1 + a_2 & \equiv b_1 + b_2 \mod n; \\
a_1 - a_2 & \equiv b_1 - b_2 \mod n; \\
a_1a_2 & \equiv b_1b_2 \mod n.
\end{align*}
\]

(e) Let \( m \in \mathbb{Z} \) be such that \( m \mid n \). If \( a, b \in \mathbb{Z} \) satisfy \( a \equiv b \mod n \), then \( a \equiv b \mod m \).

Proof. (a) Let \( a \in \mathbb{Z} \). Recall that \( a \equiv a \mod n \) is defined to mean \( n \mid a - a \). Since \( n \mid a - a \) holds (because \( a - a = 0 = n \cdot 0 \)), we thus see that \( a \equiv a \mod n \) holds. This proves Proposition 2.3.4 (a).

(b) Let \( a, b, c \in \mathbb{Z} \) satisfy \( a \equiv b \mod n \) and \( b \equiv c \mod n \).

We have \( a \equiv b \mod n \). In other words, \( n \mid a - b \) (by Definition 2.3.1). In other words, there exists an integer \( p \) such that \( a - b = np \) (by Definition 2.2.1). Consider this \( p \).

We have \( b \equiv c \mod n \). In other words, \( n \mid b - c \) (by Definition 2.3.1). In other words, there exists an integer \( q \) such that \( b - c = nq \) (by Definition 2.2.1). Consider this \( q \).

Now,

\[
a - c = (a - b) + (b - c) = np + nq = n(p + q) .
\]

Hence, there exists an integer \( r \) such that \( a - c = nr \) (namely, \( r = p + q \)). In other words, \( n \mid a - c \) (by Definition 2.2.1). In other words, \( a \equiv c \mod n \) (by Definition 2.3.1). This proves Proposition 2.3.4 (b).

(c) Let \( a, b \in \mathbb{Z} \) satisfy \( a \equiv b \mod n \).

We have \( a \equiv b \mod n \). In other words, \( n \mid a - b \) (by Definition 2.3.1). In other words, there exists an integer \( p \) such that \( a - b = np \) (by Definition 2.2.1). Consider this \( p \). Now,

\[
b - a = - (a - b) = -np = n(-p) .
\]


Hence, there exists an integer $c$ such that $b - a = nc$ (namely, $c = -p$). In other words, $n \mid b - a$ (by Definition 2.2.1). In other words, $b \equiv a \pmod{n}$ (by Definition 2.3.1). This proves Proposition 2.3.4 (c).

(d) Let $a_1, a_2, b_1, b_2 \in \mathbb{Z}$ satisfy $a_1 \equiv b_1 \pmod{n}$ and $a_2 \equiv b_2 \pmod{n}$.

We have $a_1 \equiv b_1 \pmod{n}$. In other words, $n \mid a_1 - b_1$ (by Definition 2.3.1). In other words, there exists an integer $p$ such that $a_1 - b_1 = np$ (by Definition 2.2.1). Consider this $p$.

We have $a_2 \equiv b_2 \pmod{n}$. In other words, $n \mid a_2 - b_2$ (by Definition 2.3.1). In other words, there exists an integer $q$ such that $a_2 - b_2 = nq$ (by Definition 2.2.1). Consider this $q$.

We have

$$(a_1 + a_2) - (b_1 + b_2) = (a_1 - b_1) + (a_2 - b_2) = np + nq = n(p + q).$$

Hence, there exists an integer $c$ such that $(a_1 + a_2) - (b_1 + b_2) = nc$ (namely, $c = p + q$). In other words, $n \mid (a_1 + a_2) - (b_1 + b_2)$ (by Definition 2.2.1). In other words, $a_1 + a_2 \equiv b_1 + b_2 \pmod{n}$ (by Definition 2.3.1). A similar argument (using $p - q$ instead of $p + q$) shows that $a_1 - a_2 \equiv b_1 - b_2 \pmod{n}$. It thus remains to show that $a_1a_2 \equiv b_1b_2 \pmod{n}$.

Let us first show that $a_1a_2 \equiv a_1b_2 \pmod{n}$. Indeed, $a_1a_2 - a_1b_2 = a_1(a_2 - b_2) = a_1nq = n(a_1q)$. Hence, there exists an integer $c$ such that $a_1a_2 - a_1b_2 = nc$ (namely, $c = a_1q$). In other words, $n \mid a_1a_2 - a_1b_2$ (by Definition 2.2.1). In other words, $a_1a_2 \equiv a_1b_2 \pmod{n}$ (by Definition 2.3.1).

Next, let us show that $a_1b_2 \equiv b_1b_2 \pmod{n}$. Indeed, $a_1b_2 - b_1b_2 = b_2(a_1 - b_1) = b_2np = n(b_2p)$. Hence, there exists an integer $c$ such that $a_1b_2 - b_1b_2 = nc$ (namely, $c = b_2p$). In other words, $n \mid a_1b_2 - b_1b_2$ (by Definition 2.2.1). In other words, $a_1b_2 \equiv b_1b_2 \pmod{n}$ (by Definition 2.3.1).

From $a_1a_2 \equiv a_1b_2 \pmod{n}$ and $a_1b_2 \equiv b_1b_2 \pmod{n}$, we now conclude that $a_1a_2 \equiv b_1b_2 \pmod{n}$ (by Proposition 2.3.4 (c), applied to $a = a_1a_2$, $b = a_1b_2$ and $c = b_1b_2$). This completes the proof of Proposition 2.3.4 (d).

(e) Let $a, b \in \mathbb{Z}$ satisfy $a \equiv b \pmod{n}$.

We have $a \equiv b \pmod{n}$. In other words, $n \mid a - b$ (by Definition 2.3.1). From $m \mid n$ and $n \mid a - b$, we obtain $m \mid a - b$ (by Proposition 2.2.4 (b), applied to $m, n$ and $a - b$ instead of $a, b$ and $c$). In other words, $a \equiv b \pmod{m}$ (by Definition 2.3.1). This proves Proposition 2.3.4 (e).

In the above proof, we took care to explicitly cite Definition 2.2.1 and Definition 2.3.1 whenever we used them; in the following, we will omit references like this.

Proposition 2.3.4 (d) is saying that congruences modulo $n$ (for a fixed integer $n$) can be added, subtracted and multiplied together. This does not mean that you can do everything with them that you can do with equalities. The next exercise shows
that dividing congruences and taking a congruence to the power of another does not generally work:

**Exercise 2.3.2.** Let \( n, a_1, a_2, b_1, b_2 \in \mathbb{Z} \) satisfy \( a_1 \equiv b_1 \mod n \) and \( a_2 \equiv b_2 \mod n \). Then, in general, neither \( a_1/a_2 \equiv b_1/b_2 \mod n \) nor \( a_1^{a_2} \equiv b_1^{b_2} \mod n \) is necessarily true. Of course, this is partly due to the fact that \( a_1/a_2, b_1/b_2 \) and \( a_1^{a_2}, b_1^{b_2} \) are not always integers in the first place (and being congruent modulo \( n \) only makes sense for integers, at least for now). But even when \( a_1/a_2, b_1/b_2 \) and \( a_1^{a_2}, b_1^{b_2} \) are integers, the congruences \( a_1/a_2 \equiv b_1/b_2 \mod n \) and \( a_1^{a_2} \equiv b_1^{b_2} \mod n \) are often false. Find examples of \( n, a_1, a_2, b_1, b_2 \) such that \( a_1/a_2, b_1/b_2 \) and \( a_1^{a_2}, b_1^{b_2} \) are integers but the congruences \( a_1/a_2 \equiv b_1/b_2 \mod n \) and \( a_1^{a_2} \equiv b_1^{b_2} \mod n \) are false.

However, we can divide a congruence \( a \equiv b \mod n \) by a nonzero integer \( d \) when all of \( a, b, n \) are divisible by \( d \):

**Exercise 2.3.3.** Let \( n, d, a, b \in \mathbb{Z} \), and assume that \( d \neq 0 \). Assume that \( d \) divides each of \( a, b, n \), and assume that \( a \equiv b \mod n \). Prove that \( a/d \equiv b/d \mod n/d \).

We can also take a congruence to the \( k \)-th power when \( k \in \mathbb{N} \):

**Exercise 2.3.4.** Let \( n, a, b \in \mathbb{Z} \) be such that \( a \equiv b \mod n \). Prove that \( a^k \equiv b^k \mod n \) for each \( k \in \mathbb{N} \).

(We can add not just two, but any number of congruences (where “number” means “finite number”):

**Exercise 2.3.5.** Let \( n \) be an integer. Let \( S \) be a finite set. For each \( s \in S \), let \( a_s \) and \( b_s \) be two integers. Assume that

\[
a_s \equiv b_s \mod n \quad \text{for each } s \in S. \tag{7}
\]

(a) Prove that

\[
\sum_{s \in S} a_s \equiv \sum_{s \in S} b_s \mod n. \tag{8}
\]

(b) Prove that

\[
\prod_{s \in S} a_s \equiv \prod_{s \in S} b_s \mod n. \tag{9}
\]

(Keep in mind that if the set \( S \) is empty, then \( \sum_{s \in S} a_s = \sum_{s \in S} b_s = 0 \) and \( \prod_{s \in S} a_s = \prod_{s \in S} b_s = 1 \); this holds by the definition of empty sums and of empty products.)
Exercise 2.3.6. Is it true that if \( a_1, a_2, b_1, b_2, n_1, n_2 \in \mathbb{Z} \) satisfy \( a_1 \equiv b_1 \mod n_1 \) and \( a_2 \equiv b_2 \mod n_2 \), then \( a_1a_2 \equiv b_1b_2 \mod n_1n_2 \)?

Exercise 2.3.7. Let \( a, b, n \in \mathbb{Z} \). Prove that \( a \equiv b \mod n \) if and only if there exists some \( d \in \mathbb{Z} \) such that \( b = a + nd \).

Exercise 2.3.8. Let \( a, b, c, n \in \mathbb{Z} \). Prove that we have \( a - b \equiv c \mod n \) if and only if \( a \equiv b + c \mod n \).

2.4. Chains of congruences

For this whole Section 2.4, we fix an integer \( n \).

Chains of equalities are a fundamental piece of notation used throughout mathematics. For example, here is a chain of equalities:

\[
(ad + bc)^2 + (ac - bd)^2 \\
= (ad)^2 + 2ad \cdot bc + (bc)^2 + (ac)^2 - 2ac \cdot bd + (bd)^2 \\
= a^2d^2 + 2abcd + b^2c^2 + a^2c^2 - 2abcd + b^2d^2 \\
= a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2 \\
= (a^2 + b^2)(c^2 + d^2)
\]

(where \( a, b, c, d \) are arbitrary numbers). This chain proves the equality \( (1) \). But why does it really? If we look closely at this chain of equalities, we see that it has the form “\( A = B = C = D = E \)”, where \( A, B, C, D, E \) are five numbers (namely, \( A = (ad + bc)^2 + (ac - bd)^2 \) and \( B = (ad)^2 + 2ad \cdot bc + (bc)^2 + (ac)^2 - 2ac \cdot bd + (bd)^2 \) and so on). This kind of statement is called a “chain of equalities”, and, a priori, it simply means that any two adjacent numbers in this chain are equal: \( A = B \) and \( B = C \) and \( C = D \) and \( D = E \). Without as much as noticing it, we have concluded that any two numbers in this chain are equal; thus, in particular, \( A = E \), which is precisely the equality \( (1) \) we wanted to prove.

That this kind of “chaining” is possible is one of the most basic facts in mathematics. Let us define a chain of equalities formally:

**Definition 2.4.1.** If \( a_1, a_2, \ldots, a_k \) are \( k \) objects, then the statement “\( a_1 = a_2 = \cdots = a_k \)” shall mean that

\[ a_i = a_{i+1} \text{ holds for each } i \in \{1, 2, \ldots, k-1\}. \]

(In other words, it shall mean that \( a_1 = a_2 \) and \( a_2 = a_3 \) and \( a_3 = a_4 \) and \( \cdots \) and \( a_{k-1} = a_k \). This is vacuously true when \( k \leq 1 \). If \( k = 2 \), then it simply means that \( a_1 = a_2 \).

Such a statement will be called a chain of equalities.
Proposition 2.4.2. Let \( a_1, a_2, \ldots, a_k \) be \( k \) objects such that \( a_1 = a_2 = \cdots = a_k \). Let \( u \) and \( v \) be two elements of \( \{1, 2, \ldots, k\} \). Then, \( a_u = a_v \).

So we have defined a chain of equalities to be true if and only if any two adjacent terms in this chain are equal (i.e., if “each equality sign in the chain is satisfied”). Proposition 2.4.2 shows that in such a chain, any two terms are equal (i.e., if “each equality sign in the chain is satisfied”).

This is intuitively rather clear, but can also be formally proven by induction using the basic properties of equality (transitivity\(^9\), reflexivity\(^10\), and symmetry\(^11\)).

But our goal is to understand basic number theory, not to scrutinize the foundations of mathematics. So let us recall that we have fixed an integer \( n \), and consider congruences modulo \( n \). We claim that these can be chained just as equalities:

**Definition 2.4.3.** If \( a_1, a_2, \ldots, a_k \) are \( k \) integers, then the statement “\( a_1 \equiv a_2 \equiv \cdots \equiv a_k \mod n \)” shall mean that

\[
 a_i \equiv a_{i+1} \mod n \text{ holds for each } i \in \{1, 2, \ldots, k-1\}.
\]

(In other words, it shall mean that \( a_1 \equiv a_2 \mod n \) and \( a_2 \equiv a_3 \mod n \) and \( a_3 \equiv a_4 \mod n \) and \( \cdots \) and \( a_{k-1} \equiv a_k \mod n \). This is vacuously true when \( k \leq 1 \). If \( k = 2 \), then it simply means that \( a_1 \equiv a_2 \mod n \).

Such a statement will be called a **chain of congruences modulo \( n \)**.

**Proposition 2.4.4.** Let \( a_1, a_2, \ldots, a_k \) be \( k \) integers such that \( a_1 \equiv a_2 \equiv \cdots \equiv a_k \mod n \). Let \( u \) and \( v \) be two elements of \( \{1, 2, \ldots, k\} \). Then, \( a_u \equiv a_v \mod n \).

Proposition 2.4.4 shows that any two terms in a chain of congruences modulo \( n \) must be congruent to each other modulo \( n \). Again, this can be formally proven by induction; see [Grinbe15, proof of Proposition 2.16]. The ingredients of the proof are basic properties of congruence modulo \( n \): transitivity, reflexivity and symmetry. These are fancy names for parts (b), (a) and (c) of Proposition 2.3.4.

We will use Proposition 2.4.4 tacitly (just as you would use Proposition 2.4.2): i.e., every time we prove a chain of congruences like \( a_1 \equiv a_2 \equiv \cdots \equiv a_k \mod n \), we assume that the reader will automatically conclude that any two of its terms are congruent to each other modulo \( n \) (and will remember this conclusion). For instance, if we show that \( 1 \equiv 4 \equiv 34 \equiv 334 \equiv 304 \mod 3 \), then we automatically get the congruences \( 1 \equiv 304 \mod 3 \) and \( 334 \equiv 1 \mod 3 \) and \( 4 \equiv 334 \mod 3 \) and several others out of this chain.

Chains of congruences can also include equality signs. For example, if \( a, b, c, d \) are integers, then “\( a \equiv b = c \equiv d \mod n \)” means that \( a \equiv b \mod n \) and \( b = c \) and \( c \equiv d \mod n \). Such a chain is still a chain of congruences, because \( b = c \) implies \( b \equiv c \mod n \) (by Proposition 2.3.4(a)).

---

8“Objects” can be numbers, sets, tuples or any other well-defined things in mathematics.

9Transitivity of equality says that if \( a, b, c \) are three objects satisfying \( a = b \) and \( b = c \), then \( a = c \).

10Reflexivity of equality says that every object \( a \) satisfies \( a = a \).

11Symmetry of equality says that if \( a, b \) are two objects satisfying \( a = b \), then \( b = a \).
Just as there are chains of equalities and chains of congruences, there are chains of divisibilities:

**Definition 2.4.5.** If \( a_1, a_2, \ldots, a_k \) are \( k \) integers, then the statement “\( a_1 \mid a_2 \mid \cdots \mid a_k \)” shall mean that

\[
a_i \mid a_{i+1}
\]

holds for each \( i \in \{1, 2, \ldots, k - 1\} \).

(In other words, it shall mean that \( a_1 \mid a_2 \) and \( a_2 \mid a_3 \) and \( a_3 \mid a_4 \) and \( \cdots \) and \( a_{k-1} \mid a_k \). This is vacuously true when \( k \leq 1 \). If \( k = 2 \), then it simply means that \( a_1 \mid a_2 \).

Such a statement will be called a *chain of divisibilities*.

**Proposition 2.4.6.** Let \( a_1, a_2, \ldots, a_k \) be \( k \) integers such that \( a_1 \mid a_2 \mid \cdots \mid a_k \). Let \( u \) and \( v \) be two elements of \( \{1, 2, \ldots, k\} \) such that \( u \leq v \). Then, \( a_u \mid a_v \).

Note that we had to require \( u \leq v \) in this proposition, unlike the analogous propositions for chains of equalities and chains of congruences, because there is no “symmetry of divisibility” (i.e., if \( a \mid b \), then we don’t generally have \( b \mid a \)). The proof of Proposition 2.4.6 relies on the reflexivity of divisibility (Proposition 2.2.4 (a)) and on the transitivity of divisibility (Proposition 2.2.4 (b)).

Again, chains of divisibilities can include equality signs. For example, \( 4 \mid 3 \cdot 4 = 12 = 2 \cdot 6 \mid 4 \cdot 6 = 24 \).

### 2.5. Substitutivity for congruences

In Section 2.4, we have learnt that congruences modulo an integer \( n \) can be chained together like equalities. A further important feature of congruences is the principle of *substitutivity for congruences*. This is yet another way in which congruences behave like equalities. We are not going to state it fully formally (as it is a metamathematical principle), but merely explain its meaning. Later on, once we understand what the rings \( \mathbb{Z}/n \) (for integer \( n \)) are, we will no longer need this principle, since it will just boil down to “equal things can be substituted for one another” (the whole point of \( \mathbb{Z}/n \) is to “make congruent numbers equal”); but for now, we cannot treat “congruent modulo \( n \)” as “equal”, so we have to state it.

You are probably used to making computations like these:

\[
(a + b)^2 + (a - b)^2 = \left( a^2 + 2ab + b^2 \right) + \left( a^2 - 2ab + b^2 \right)
\]

\[
= a^2 + a^2 + b^2 + b^2 = 2a^2 + 2b^2
\]

(for any two numbers \( a \) and \( b \)). What is going on in these underbraces (like “\((a + b)^2\)”)? Something pretty simple is going on: You are replacing a num-
ber (in this case, \((a + b)^2\)) by an equal number (in this case, \(a^2 + 2ab + b^2\)). This relies on a fundamental principle of mathematics (called the principle of substitutivity for equalities), which says that an object in an expression can indeed be replaced by any object equal to it (without changing the value of the expression). (This is also known as Leibniz’s equality law.) To be precise, we are using this principle twice in some of our equality signs above, since we are making several replacements at the same time; but this is fine (we can just do the replacement one by one instead).

We would like to have a similar principle for congruences modulo \(n\): We would like to be able to replace any integer by an integer congruent to it modulo \(n\). For example, we would like to be able to say that if seven integers \(a, a', b, b', c, c', n\) satisfy \(a \equiv a' \mod n\) and \(b \equiv b' \mod n\) and \(c \equiv c' \mod n\), then

\[
\begin{align*}
    b \equiv b' \mod n & \quad \equiv c' \mod n \\
    c \equiv c' \mod n & \quad \equiv a' \mod n \\
    a \equiv a' \mod n & \quad \equiv b' \mod n
\end{align*}
\]

We have to be careful with this: For example, we run into troubles if division is involved in our expressions. For example, we have \(6 \equiv 9 \mod 3\), but we do not have \(\frac{6}{3} \equiv \frac{9}{3} \equiv 3 \mod 3\). Similarly, exponentiation can be problematic. So we need to state the principle we are using here in clearer terms, so that we know what we can do.

For this whole Section 2.5 we fix an integer \(n\).

The principle of substitutivity for equalities says the following:

**Principle of substitutivity for equalities (PSE):** If two objects \(x\) and \(x'\) are equal, and if we have any expression \(A\) that involves the object \(x\), then we can replace this \(x\) (or, more precisely, any arbitrary appearance of \(x\) in \(A\)) in \(A\) by \(x'\); the value of the resulting expression \(A'\) will equal the value of \(A\).

Here are two examples of how this principle can be used:

- If \(a, b, c, d, e, c'\) are numbers such that \(c = c'\), then the PSE says that we can replace \(c\) by \(c'\) in the expression \(a \ (b - (c + d) \ e)\), and the value of the resulting expression \(a \ (b - (c' + d) \ e)\) will equal the value of \(a \ (b - (c + d) \ e)\); that is, we have

  \[
  a \ (b - (c + d) \ e) = a \ (b - (c' + d) \ e). \tag{10}
  \]

- If \(a, b, c, a'\) are numbers such that \(a = a'\), then

  \[
  (a - b) \ (a + b) = (a' - b) \ (a + b), \tag{11}
  \]

because the PSE allows us to replace the first \(a\) appearing in the expression \((a - b) \ (a + b)\) by an \(a'\). (We can also replace the second \(a\) by \(a'\), of course.)
More generally, we can make several such replacements at the same time.
The PSE is one of the headstones of mathematical logic; it is the essence of what
it means for two objects to be equal.
The principle of substitutivity for congruences is similar, but far less fundamental; it
says the following:

**Principle of substitutivity for congruences (PSC):** If two numbers \(x\) and \(x'\)
are congruent to each other modulo \(n\) (that is, \(x \equiv x' \mod n\)), and if we
have any expression \(A\) that involves only integers, addition, subtraction
and multiplication, and involves the object \(x\), then we can replace this \(x\)
(or, more precisely, any arbitrary appearance of \(x\) in \(A\)) in \(A\) by \(x'\); the
value of the resulting expression \(A'\) will be congruent to the value of \(A\)
modulo \(n\).

This principle is less general than the PSE, since it only applies to expressions
that are built from integers and certain operations (note that division is not one of
these operations). But it still lets us prove analogues of our above examples (10)
and (11):

- If \(a, b, c, d, e, c'\) are integers such that \(c \equiv c' \mod n\), then the PSC says that
we can replace \(c\) by \(c'\) in the expression \(a (b - (c + d) e)\), and the value of
the resulting expression \(a (b - (c' + d) e)\) will be congruent to the value of
\(a (b - (c + d) e)\) modulo \(n\); that is, we have
\[
  a (b - (c + d) e) \equiv a (b - (c' + d) e) \mod n.
\]
  (12)

- If \(a, b, c, a'\) are integers such that \(a \equiv a' \mod n\), then
\[
  (a - b) (a + b) \equiv (a' - b) (a + b) \mod n,
\]
  (13)
because the PSC allows us to replace the first \(a\) appearing in the expression
\((a - b) (a + b)\) by an \(a'\). (We can also replace the second \(a\) by \(a'\), of course.)

We shall not prove the PSC, since we have not formalized it (after all, we have not
defined what an “expression” is). But we shall prove the specific congruences (12)
and (13) using Proposition 2.3.4: the way in which we prove these congruences is
symptomatic: Every congruence obtained from the PSC can be proven in a manner
like these. Thus, the proofs of (12) and (13) given below can serve as templates
which can easily be adapted to any other situation in which an application of the
PSC needs to be justified.

**Proof of (12).** Let \(n\) be any integer, and let \(a, b, c, d, e, c'\) be integers such that \(c \equiv c' \mod n\).

Adding the congruence \(c \equiv c' \mod n\) with the congruence \(d \equiv d \mod n\) (which
follows from Proposition 2.3.4 (a)), we obtain \(c + d \equiv c' + d \mod n\). Multiplying


\[\text{Proposition 2.3.4 (d) shows that we can add, subtract and multiply congruences modulo } n \text{ at will.}\]

\[\text{We are using this freedom here and will use it many times below.}\]
this congruence with the congruence \( e \equiv e \mod n \) (which follows from Proposition 2.3.4 (a)), we obtain \((c + d) e \equiv (c' + d) e \mod n\). Subtracting this congruence from the congruence \( b \equiv b \mod n \) (which, again, follows from Proposition 2.3.4 (a)), we obtain \( b - (c + d) e \equiv b - (c' + d) e \mod n \). Multiplying the congruence \( a \equiv a \mod n \) (which follows from Proposition 2.3.4 (a)) with this congruence, we obtain \( a (b - (c + d) e) \equiv a (b - (c' + d) e) \mod n \). This proves (12).

**Proof of (13).** Let \( n \) be any integer, and let \( a, b, c, a' \) be integers such that \( a \equiv a' \mod n \).

Subtracting the congruence \( b \equiv b \mod n \) (which follows from Proposition 2.3.4 (a)) from the congruence \( a \equiv a' \mod n \), we obtain \( a - b \equiv a' - b \mod n \). Multiplying this congruence with the congruence \( a + b \equiv a + b \mod n \) (which follows from Proposition 2.3.4 (a)), we obtain \((a - b)(a + b) \equiv (a' - b)(a + b) \mod n \). This proves (13).

As we said, these two proofs are exemplary: Any congruence obtained from the PSC can be proven in such a way (starting with the congruence \( x \equiv x' \mod n \), and then “wrapping” it up in the expression \( A \) by repeatedly adding, multiplying and subtracting congruences that follow from Proposition 2.3.4 (a)).

When we apply the PSC, we shall use underbraces to point out which integers we are replacing. For example, when deriving (12) from this principle, we shall write

\[
\begin{align*}
\left( b - \left( \begin{array}{c}
\equiv c' \mod n \n\end{array} \right) + d \right) e & \equiv a \left( b - (c' + d) e \right) \mod n,
\end{align*}
\]

in order to stress that we are replacing \( c \) by \( c' \). Likewise, when deriving (13) from the PSC, we shall write

\[
\left( \begin{array}{c}
\equiv a' \mod n \n\end{array} \right) - b \left( a + b \right) \equiv (a' - b)(a + b) \mod n,
\]

in order to stress that we are replacing the first \( a \) (but not the second \( a \)) by \( a' \).

The PSC allows us to replace a **single** integer \( x \) appearing in an expression by another integer \( x' \) that is congruent to \( x \) modulo \( n \). Applying this principle many times, we thus conclude that we can also replace **several** integers at the same time (because we can get to the same result by performing these replacements one at a time, and Proposition 2.3.4 shows that the value of the final result will be congruent to the value of the original result). For example, if seven integers \( a, a', b, b', c, c', n \) satisfy \( a \equiv a' \mod n \) and \( b \equiv b' \mod n \) and \( c \equiv c' \mod n \), then

\[
bc + ca + ab \equiv b'c' + c'a' + a'b' \mod n,
\]

because we can replace all the six integers \( b, c, c, a, a, b \) in the expression \( bc + ca + ab \) (listed in the order of their appearance in this expression) by \( b', c', c', a', a', b' \),
respectively. If we want to derive this from the PSC, then we must perform the replacements one at a time, e.g., as follows:

\[
\begin{align*}
\text{b} & \equiv b' \mod n \\
\text{c} + ca + ab & \equiv b'c' + c' \mod n \\
+ab & \equiv b'c' + c'a' \mod n \\
\equiv b'c' + c'a' + a'b' \mod n.
\end{align*}
\]

Of course, we shall always just show the replacements as a single step:

\[
\begin{align*}
\text{b} & \equiv b' \mod n \\
\text{c} & \equiv c' \mod n \\
\text{a} & \equiv a' \mod n \\
\text{b} & \equiv b' \mod n \\
\end{align*}
\]

The PSC can be extended: The expression \( A \) can be allowed to involve not only integers, addition, subtraction, multiplication and \( x \), but also \( k \)-th powers for \( k \in \mathbb{N} \) (as long as \( k \) remains unchanged in our replacement) as well as finite sums and products (as long as the bounds of the sums and products are unchanged). This follows from Exercise 2.3.4 and Exercise 2.3.5.

**Exercise 2.5.1.** Let \( n \in \mathbb{N} \). Show that \( 7 \mid 3^{2n+1} + 2^{n+2} \).

2019-02-01 lecture

### 2.6. Division with remainder

The following fact you likely remember from high school:

**Theorem 2.6.1.** Let \( n \) be a positive integer. Let \( u \in \mathbb{Z} \). Then, there exists a unique pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \( u = qn + r \).

We shall refer to this as the “division-with-remainder theorem for integers”. Before we prove this theorem, let us introduce the notations that it justifies:

**Definition 2.6.2.** Let \( n \) be a positive integer. Let \( u \in \mathbb{Z} \). Theorem 2.6.1 shows that there exists a unique pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \( u = qn + r \). Consider this pair.

(a) We denote the integer \( q \) by \( u / n \), and refer to it as the **quotient of the division of \( u \) by \( n \)**.

(b) We denote the integer \( r \) by \( u \% n \), and refer to it as the **remainder of the division of \( u \) by \( n \)**.

The words “quotient” and “remainder” are standard, but the notations “\( u / n \)” and “\( u \% n \)” are not (I have taken them from the Python programming language); be prepared to see other notations in the literature (e.g., the notations “\( \text{quo}(u, n) \)” and “\( \text{rem}(u, n) \)” for \( u / n \) and \( u \% n \), respectively).
Proof of Lemma 2.6.5. Let \( u \) such that \( u \) we can use the notations \( \left\lfloor \frac{u}{r} \right\rfloor \). Also, \( r = 2 \) is the unique pair \((q,r)\) such that \( r \). Similarly, \( q \) integers. We begin by proving Lemma 2.6.5 (which is the easier one):

Lemma 2.6.4. Let \( n \) be a positive integer. Let \( u \in \mathbb{Z} \). Then, there exists at least one pair \((q,r)\) such that \( u = qn + r \).

Lemma 2.6.5. Let \( n \) be a positive integer. Let \( u \in \mathbb{Z} \). Then, there exists at most one pair \((q,r)\) such that \( u = qn + r \).

We begin by proving Lemma 2.6.5 (which is the easier one):

Proof of Lemma 2.6.5. Let \((q_1, r_1)\) and \((q_2, r_2)\) be two pairs \((q,r)\) such that \( u = qn + r \). We shall show that \((q_1, r_1) = (q_2, r_2)\).

We know that \((q_1, r_1)\) is a pair \((q,r)\) such that \( u = qn + r \). In other words, \((q_1, r_1) \in \mathbb{Z} \times \{0,1,\ldots,n-1\}\) and \( u = q_1n + r_1 \). Similarly, \((q_2, r_2) \in \mathbb{Z} \times \{0,1,\ldots,n-1\}\) and \( u = q_2n + r_2 \).

From \((q_1, r_1) \in \mathbb{Z} \times \{0,1,\ldots,n-1\}\), we obtain \( q_1 \in \mathbb{Z} \) and \( r_1 \in \{0,1,\ldots,n-1\}\). Similarly, \( q_2 \in \mathbb{Z} \) and \( r_2 \in \{0,1,\ldots,n-1\}\). Thus, in particular, \((q_1,q_2,r_1,r_2)\) are integers.

From \( r_1 \in \{0,1,\ldots,n-1\} \) and \( r_2 \in \{0,1,\ldots,n-1\} \), we can easily derive

\[
|r_2 - r_1| \leq n - 1. \tag{15}
\]

Proof of (15): Intuitively, this should be clear: Both \( r_1 \) and \( r_2 \) belong to the integer interval \( \{0,1,\ldots,n-1\} \), and thus the unsigned distance between \( r_1 \) and \( r_2 \) is at most \( n - 1 \) (with the worst case being when \( r_1 \) and \( r_2 \) are at opposite ends of this interval).

Here is a formal restatement of this argument: We have \( r_1 \in \{0,1,\ldots,n-1\} \), thus \( r_1 \geq 0 \). Also, \( r_2 \in \{0,1,\ldots,n-1\} \), hence \( r_2 \leq n - 1 \). Hence, \( r_1 - r_2 \leq (n-1)-0 = n-1 \).

Similarly, \( r_1 - r_2 \leq n - 1 \). But recall that \( |x| \in \{x,-x\} \) for each \( x \in \mathbb{Z} \). Applying this to \( x = r_2 - r_1 \), we obtain

\[
|r_2 - r_1| \in \left\{ r_2 - r_1 - (r_2 - r_1) \right\} = \left\{ r_2 - r_1, r_1 - r_2 \right\}.
\]
In other words, \( |r_2 - r_1| \) is one of the two numbers \( r_2 - r_1 \) and \( r_1 - r_2 \). Since both of these numbers \( r_2 - r_1 \) and \( r_1 - r_2 \) are \( \leq n - 1 \) (as we have just shown), we thus conclude that \( |r_2 - r_1| \leq n - 1 \). This proves (15).

We have \( q_1 n + r_1 = u = q_2 n + r_2 \), thus \( q_1 n - q_2 n = r_2 - r_1 \). Hence,

\[
\begin{align*}
r_2 - r_1 &= q_1 n - q_2 n = (q_1 - q_2) n. \\
\end{align*}
\]

(16)

Assume (for the sake of contradiction) that \( q_1 \neq q_2 \). Thus, \( q_1 - q_2 \neq 0 \), so that \( |q_1 - q_2| > 0 \) and therefore \( |q_1 - q_2| \geq 1 \) (since \( |q_1 - q_2| \) is an integer). We can multiply this inequality by \( n \) (since \( n \) is positive) and thus obtain \( |q_1 - q_2| n \geq 1 \). But from (16), we obtain

\[
|r_2 - r_1| = |(q_1 - q_2) n| = |q_1 - q_2| \cdot |n| = n \quad \text{(by (3))}
\]

(since \( n \) is positive)

\[
|q_1 - q_2| n \geq n > n - 1.
\]

This contradicts (15). This contradiction shows that our assumption (that \( q_1 \neq q_2 \)) was false. Hence, we have \( q_1 = q_2 \). Thus, \( q_1 - q_2 = 0 \), so that (16) becomes \( r_2 - r_1 = (q_1 - q_2) n = 0 \) and thus \( r_2 = r_1 \), so that \( r_1 = r_2 \). Combining this with \( q_1 = q_2 \), we obtain \((q_1, r_1) = (q_2, r_2)\).

Now, forget that we have fixed \((q_1, r_1)\) and \((q_2, r_2)\). We thus have proven that if \((q_1, r_1)\) and \((q_2, r_2)\) are two pairs \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \( u = qn + r \), then \((q_1, r_1) = (q_2, r_2)\). In other words, any two pairs \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \( u = qn + r \) must be equal. In other words, there exists at most one such pair. This proves Lemma 2.6.5.

But we also need to prove Lemma 2.6.4. This lemma can be proven by induction on \( u \), but not without some complications: Since it is stated for all integers \( u \) (rather than just for nonnegative or positive integers), the classical induction principle (with an induction base and a “\( u \) to \( u + 1 \)” step) cannot prove it directly. Instead, we have to either add a “\( u \) to \( u - 1 \)” step to our induction (resulting in a “two-sided induction” or “up- and down-induction” argument), or to treat the case of negative \( u \) separately. A proof using the first of these two methods can be found in [Grinbe15 proof of Proposition 2.150] (where \( n \) and \( u \) are denoted by \( N \) and \( n \)). We shall instead give a proof using the second method; thus, we first state the particular case of Lemma 2.6.4 when \( u \) is nonnegative:

**Lemma 2.6.6.** Let \( n \) be a positive integer. Let \( u \in \mathbb{N} \). Then, there exists at least one pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \( u = qn + r \).

This lemma can be proven by induction on \( u \) as in [Grinbe15 proof of Proposition 2.150]. Let us instead prove it by **strong** induction on \( u \). See [Grinbe15 §2.8] for an introduction to strong induction; in particular, recall that a strong induction needs no induction base (but often contains a case distinction in its “induction step” that,
in some way, does give the first few values a special treatment). The proof of Lemma 2.6.6 that we give below follows a stupid but valid method of finding the pair \((q, r)\): Keep subtracting \(n\) from \(u\) until \(u\) becomes \(< n\); then \(r\) will be the resulting number, whereas \(q\) will be the number of times you have subtracted \(n\).

**Proof of Lemma 2.6.6** We proceed by strong induction on \(u\).

Let \(U \in \mathbb{N}\). Assume (as the induction hypothesis) that Lemma 2.6.6 holds for every \(u \in \mathbb{N}\) satisfying \(u < U\). We must prove that Lemma 2.6.6 also holds for \(u = U\). In other words, we must prove that there exists at least one pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \(U = qn + r\).

We are in one of the following two cases:

**Case 1:** We have \(U < n\).

**Case 2:** We have \(U \geq n\).

Let us first consider Case 1. In this case, we have \(U < n\). Thus, \(U \leq n - 1\) (since \(U\) and \(n\) are integers), so that \(U \in \{0, 1, \ldots, n - 1\}\) (since \(U \in \mathbb{N}\)). Combining this with \(0 \in \mathbb{Z}\), we obtain \((0, U) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\). Hence, \((0, U)\) is a pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \(U = qn + r\) (since \(U = 0n + U\)). Thus, there exists at least one pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \(U = qn + r\) (namely, \((q, r) = (0, U)\)).

Let us now consider Case 2. In this case, we have \(U \geq n\). Hence, \(U - n \geq 0\), so that \(U - n \in \mathbb{N}\) (remember that \(\mathbb{N} = \{0, 1, 2, \ldots\}\)). Also, \(U - n < U\) (since \(n\) is positive). But our induction hypothesis said that Lemma 2.6.6 holds for every \(u \in \mathbb{N}\) satisfying \(u < U\). Hence, in particular, Lemma 2.6.6 holds for \(u = U - n\) (since \(U - n \in \mathbb{N}\) and \(U - n < U\)). In other words, there exists at least one pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \(U - n = qn + r\). Fix such a pair and denote it by \((q_0, r_0)\). Thus, \((q_0, r_0) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) and \(U - n = q_0 n + r_0\).

From \(U - n = q_0 n + r_0\), we obtain \(U = n + (q_0 n + r_0) = (q_0 + 1) n + r_0\). Also, from \((q_0, r_0) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\), we obtain \(q_0 \in \mathbb{Z}\) and \(r_0 \in \{0, 1, \ldots, n - 1\}\), and thus \((q_0 + 1, r_0) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\). Thus, the pair \((q_0 + 1, r_0)\) is a pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \(U = qn + r\) (since \(U = (q_0 + 1) n + r_0\)). Therefore, there exists at least one pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \(U = qn + r\) (namely, \((q, r) = (q_0 + 1, r_0)\)).

Now, in each of the two Cases 1 and 2, we have shown that there exists at least one pair \((q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\) such that \(U = qn + r\). Hence, this holds always. In other words, Lemma 2.6.6 holds for \(u = U\). This completes the induction step; thus, Lemma 2.6.6 is proven by strong induction.

In order to derive Lemma 2.6.4 from Lemma 2.6.6 (that is, to extend Lemma 2.6.6 to the case of negative \(u\)), we shall need a simple but important trick:

**Lemma 2.6.7.** Let \(n\) be a positive integer. Let \(u \in \mathbb{Z}\). Then, there exists a \(v \in \mathbb{N}\) such that \(u \equiv v \pmod{n}\).

**Proof of Lemma 2.6.7** We are in one of the following two cases:

**Case 1:** We have \(u \geq 0\).
Case 2: We have $u < 0$.

Let us first consider Case 1. In this case, we have $u \geq 0$. Thus, $u \in \mathbb{N}$. Also, $u \equiv u \mod n$ (by Proposition 2.3.4 (a)). Thus, there exists a $v \in \mathbb{N}$ such that $u \equiv v \mod n$ (namely, $v = u$). This proves Lemma 2.6.7 in Case 1.

Let us now consider Case 2. In this case, we have $u < 0$. Hence, $-u > 0$. Now, $u - (n-1)(-u) = nu$ is divisible by $n$ (since $u \in \mathbb{Z}$). In other words, $n \mid u - (n-1)(-u)$. In other words, $u \equiv (n-1)(-u) \mod n$. Moreover, $n \geq 1$ (since $n$ is a positive integer), so that $n - 1 \geq 0$. We can multiply this inequality with $-u$ (since $-u > 0$), and thus obtain $(n-1)(-u) \geq 0 (-u) = 0$. In other words, $(n-1)(-u) \in \mathbb{N}$. Thus, there exists a $v \in \mathbb{N}$ such that $u \equiv v \mod n$ (namely, $v = (n-1)(-u)$). This proves Lemma 2.6.7 in Case 2.

We have now proven Lemma 2.6.7 in both Cases 1 and 2; hence, Lemma 2.6.7 always holds.

Proof of Lemma 2.6.4 Lemma 2.6.7 shows that there exists a $v \in \mathbb{N}$ such that $u \equiv v \mod n$. Consider this $v$.

Note that $n \mid u - v$ (since $u \equiv v \mod n$). In other words, there exists an integer $c$ such that $u - v = nc$. Consider this $c$. From $u - v = nc$, we obtain $u = v + nc$.

Lemma 2.6.6 (applied to $v$ instead of $u$) yields that there exists at least one pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n-1\}$ such that $v = qn + r$. Fix such a pair, and denote it by $(q_0, r_0)$. Thus, $(q_0, r_0) \in \mathbb{Z} \times \{0, 1, \ldots, n-1\}$ and $v = q_0n + r_0$. Now,

$$u = v + nc = (q_0n + r_0) + nc = (q_0 + c)n + r_0.$$  

Also, from $(q_0, r_0) \in \mathbb{Z} \times \{0, 1, \ldots, n-1\}$, we obtain $q_0 \in \mathbb{Z}$ and $r_0 \in \{0, 1, \ldots, n-1\}$, and thus $(q_0 + c, r_0) \in \mathbb{Z} \times \{0, 1, \ldots, n-1\}$. Thus, the pair $(q_0 + c, r_0)$ is a pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n-1\}$ such that $u = qn + r$ (since $u = (q_0 + c)n + r_0$). Therefore, there exists at least one pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n-1\}$ such that $u = qn + r$ (namely, $(q, r) = (q_0 + c, r_0)$). This proves Lemma 2.6.4.

Proof of Theorem 2.6.1 Theorem 2.6.1 follows by combining Lemma 2.6.4 with Lemma 2.6.5.

Remark 2.6.8. We can visualize Theorem 2.6.1 as follows: Mark all the multiples of $n$ on the real line. These multiples are evenly spaced points, with a distance of $n$ between any two neighboring multiples. Thus, they subdivide the real line into infinitely many intervals of length $n$. More precisely, for each $a \in \mathbb{Z}$, let $I_a$ be the interval $[an, (a+1)n) = \{x \in \mathbb{R} \mid an \leq x < (a+1)n\}$; then, every real belongs to exactly one of these intervals $I_a$. (This is intuitively clear – I am not saying this is a rigorous proof.) Thus, in particular, $u$ belongs to $I_q$ for some $q \in \mathbb{Z}$. This $q$ is precisely the $q$ in the unique pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n-1\}$ satisfying $u = qn + r$. Moreover, the $r$ from this pair specifies the relative position of $u$ in the interval $I_q$.

(Unfortunately, it is not clear to me whether this intuition can be turned into a proper proof of Theorem 2.6.1, since it relies on the fact that every real number
belongs to exactly one of the intervals $I_a$, which fact may well require Theorem 2.6.1 for its proof.)

The following properties of the quotient and the remainder are simple but will be used all the time:

**Corollary 2.6.9.** Let $n$ be a positive integer. Let $u \in \mathbb{Z}$.

(a) Then, $u \% n \in \{0, 1, \ldots, n - 1\}$ and $u \% n \equiv u \mod n$.

(b) We have $n \mid u$ if and only if $u \% n = 0$.

(c) If $c \in \{0, 1, \ldots, n - 1\}$ is such that $c \equiv u \mod n$, then $c = u \% n$.

(d) We have $u = (u \div n) n + (u \% n)$.

Before we prove this corollary, let us explain its purpose. Corollary 2.6.9 (a) says that $u \% n$ is a number in the set $\{0, 1, \ldots, n - 1\}$ that is congruent to $u$ modulo $n$. Corollary 2.6.9 (c) says that $u \% n$ is the only such number (as it says that any further such number $c$ must be equal to $u \% n$). Corollary 2.6.9 (b) gives an algorithm to check whether $n \mid u$ holds (namely, compute $u \% n$ and check whether $u \% n = 0$). Corollary 2.6.9 (d) is a trivial consequence of the definition of quotient and remainder.

**Proof of Corollary 2.6.9.** Theorem 2.6.1 says that there is a unique pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}$ such that $u = qn + r$. Consider this pair $(q, r)$. The uniqueness of this pair can be restated as follows: If $(q', r') \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}$ is any further pair such that $u = q'n + r'$, then

$$(q', r') = (q, r).$$

(17)

Recall that $u \% n$ was defined to be $r$ (in Definition 2.6.2 (b)). Thus, $u \% n = r$. Now, $n \mid qn = u - r$ (since $u = qn + r$). In other words, $u \equiv r \mod n$. Hence, $r \equiv u \mod n$ (by Proposition 2.3.4 (c)). This rewrites as $u \% n \equiv u \mod n$ (since $r = u \% n$).

Furthermore, $u \% n = r \in \{0, 1, \ldots, n - 1\}$ (since $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}$).

This completes the proof of Corollary 2.6.9 (a).

Also, $u \div n$ was defined to be $q$ (in Definition 2.6.2 (a)). Hence, $u \div n = q$. Now,

$$u = \underbrace{q n + r}_{u \div n} = (u \div n) n + (u \% n).$$

This proves Corollary 2.6.9 (d).

(b) $\Longrightarrow$: Assume that $n \mid u$. We must prove that $u \% n = 0$.

We have $n \mid u$. In other words, there exists some integer $w$ such that $u = nw$. Consider this $w$.

We have $n - 1 \in \mathbb{N}$ (since $n$ is a positive integer), thus $0 \in \{0, 1, \ldots, n - 1\}$. Hence, $(w, 0) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}$ (since $w \in \mathbb{Z}$). Also, $u = nw = wn = wn + 0$. Hence, (17) (applied to $(q', r') = (w, 0)$) yields $(w, 0) = (q, r)$. In other words, $w = q$ and $0 = r$. Hence, $r = 0$, so that $u \% n = r = 0$. This proves the “$\Longrightarrow$” implication of Corollary 2.6.9 (b).
\(\iff\): Assume that \(u \% n = 0\). We must prove that \(n \mid u\).

We have \(u = qn + r\) with \(q = \floor{\frac{u}{n}}\) and \(r = u \mod n\). Thus, \(n \mid u\). This proves the "\(\iff\)" implication of Corollary 2.6.9(b).

(c) Let \(c \in \{0, 1, \ldots, n - 1\}\) be such that \(c \equiv u \mod n\).

We have \(c \equiv u \mod n\). In other words, \(n \mid c - u\). In other words, there exists some integer \(w\) such that \(c - u = nw\). Consider this \(w\).

From \(-w \in \mathbb{Z}\) and \(c \in \{0, 1, \ldots, n - 1\}\), we obtain \((\pm w, c) \in \mathbb{Z} \times \{0, 1, \ldots, n - 1\}\). Also, from \(c - u = nw\), we obtain \(u = c - nw = (-w) n + c\). Hence, (17) (applied to \((q', r') = (-w, c)\)) yields \((-w, c) = (q, r)\). In other words, \(-w = q\) and \(c = r\).

Hence, \(c = r = u \% n\). This proves Corollary 2.6.9(c).

\[
\text{Exercise 2.6.1. Let } n \text{ be a positive integer. Let } u \text{ and } v \text{ be integers. Prove that } u \equiv v \mod n \text{ if and only if } u \% n = v \% n. 
\]

The following exercise provides an analogue of Theorem 2.6.1 in which \(r\) is required to be an integer satisfying \(|r| \leq n/2\) rather than an element of \(\{0, 1, \ldots, n - 1\}\). Note, however, that \(r\) is not always unique in this case.

\[
\text{Exercise 2.6.2. Let } n \text{ be a positive integer. Let } u \in \mathbb{Z}.
\]

(a) Prove that there exists a pair \((q, r) \in \mathbb{Z} \times \mathbb{Z}\) such that \(u = qn + r\) and \(|r| \leq n/2\).

(b) Prove that this pair is not unique in general (i.e., find \(n\) and \(u\) for which it is not unique).

Remark 2.6.10. There is a simple visualization that makes Exercise 2.6.2(a) intuitively obvious: Mark all the multiples of \(n\) on the real line. These multiples are evenly spaced points, with a distance of \(n\) between any two neighboring multiples. Hence, every point on the real line is at most a distance of \(n/2\) away from the closest multiple of \(n\). Applying this to the point \(u\), we conclude that \(u\) is at most a distance of \(n/2\) away from the closest multiple of \(n\). In other words, if \(qn\) is the closest multiple of \(n\) to \(u\) (or one of the two closest multiples of \(n\), if \(u\) is in the middle between two multiples), then \(|u - qn| \leq n/2\). Thus, if we set \(r = u - qn\), then \(u = qn + r\) and \(|r| \leq n/2\). This proves Exercise 2.6.2(a) intuitively.

This point of view also makes Exercise 2.6.2(b) evident: When the point \(u\) is exactly in the middle of one of the length-\(n\) intervals between multiples of \(n\), then there are two multiples of \(n\) equally close to \(u\), and we can pick either of them; hence, the pair \((q, r)\) is not unique.

Convention 2.6.11. The symbols \(/\) and \(\%\) will be granted higher precedence (in the sense of operator precedence) than addition. This means that an expression of the form \(c + a /\%n + b\) will always be interpreted as \(c + (a /\%n) + b\), rather than as \((c + a) /\%(n + b)\) (or in any other way). Likewise, an expression of the form \(c + a /\%n + b\) will always be interpreted as \(c + (a /\%n) + b\), rather than as \((c + a) /\%(n + b)\).
Exercise 2.6.3. Let $u$ and $v$ be two integers. Let $n$ be a positive integer.
(a) Prove that $u\%n + v\%n - (u + v)\%n \in \{0, n\}$.
(b) Prove that $(u + v) \div n - u \div n - v \div n \in \{0, 1\}$.

2.7. Even and odd numbers

Recall the following:

Definition 2.7.1. Let $u$ be an integer.
(a) We say that $u$ is even if $u$ is divisible by 2.
(b) We say that $u$ is odd if $u$ is not divisible by 2.

So an integer is either even or odd (but not both at the same time). The following exercise collects various properties of even and odd integers:

Exercise 2.7.1. Let $u$ be an integer.
(a) Prove that $u$ is even if and only if $u \% 2 = 0$.
(b) Prove that $u$ is odd if and only if $u \% 2 = 1$.
(c) Prove that $u$ is even if and only if $u \equiv 0 \mod 2$.
(d) Prove that $u$ is odd if and only if $u \equiv 1 \mod 2$.
(e) Prove that $u$ is odd if and only if $u + 1$ is even.
(f) Prove that exactly one of the two numbers $u$ and $u + 1$ is even.
(g) Prove that $u (u + 1) \equiv 0 \mod 2$.
(h) Prove that $u^2 \equiv -u \equiv u \mod 2$.
(i) Let $v$ be a further integer. Prove that $u \equiv v \mod 2$ holds if and only if $u$ and $v$ are either both odd or both even.

Exercise 2.7.2. (a) Prove that each even integer $u$ satisfies $u^2 \equiv 0 \mod 4$.
(b) Prove that each odd integer $u$ satisfies $u^2 \equiv 1 \mod 4$.
(c) Prove that no two integers $x$ and $y$ satisfy $x^2 + y^2 \equiv 3 \mod 4$.
(d) Prove that if $x$ and $y$ are two integers satisfying $x^2 + y^2 \equiv 2 \mod 4$, then $x$ and $y$ are both odd.

Exercise 2.7.2 (c) establishes our previous experimental observation that an integer of the form $4k + 3$ with integer $k$ (that is, an integer that is larger by 3 than a multiple of 4) can never be written as a sum of two perfect squares.

2019-02-04 lecture

2.8. The floor function

We shall now briefly introduce the floor function (following [Grinbe16]), as it is closely connected to division with remainder.
Definition 2.8.1. Let \( x \) be a real number. Then, \( \lfloor x \rfloor \) is defined to be the unique integer \( n \) satisfying \( n \leq x < n + 1 \). This integer \( \lfloor x \rfloor \) is called the floor of \( x \), or the integer part of \( x \).

Remark 2.8.2. (a) Why is \( \lfloor x \rfloor \) well-defined? I mean, why does the unique integer \( n \) in Definition 2.8.1 exist, and why is it unique? This question is trickier than it sounds and relies on the construction of real numbers. However, in the case when \( x \) is rational, the well-definedness of \( \lfloor x \rfloor \) follows from Proposition 2.8.3 below.

(b) What we call \( \lfloor x \rfloor \) is typically called \([x]\) in older books (such as [NiZuMo91]). I suggest avoiding the notation \([x]\) wherever possible; it has too many different meanings (whereas \( \lfloor x \rfloor \) almost always means the floor of \( x \)).

(c) The map \( \mathbb{R} \to \mathbb{Z} \), \( x \mapsto \lfloor x \rfloor \) is called the floor function or the greatest integer function.

There is also a ceiling function, which sends each \( x \in \mathbb{R} \) to the unique integer \( n \) satisfying \( n - 1 < x \leq n \); this latter integer is called \( \lceil x \rceil \). The two functions are connected by the rule \( \lceil x \rceil = -\lfloor -x \rfloor \) (for all \( x \in \mathbb{R} \)).

The floor and the ceiling functions are some of the simplest examples of discontinuous functions.

(d) Here are some examples of floors:

\[
\begin{align*}
\lfloor n \rfloor &= n & \text{for every } n \in \mathbb{Z}; \\
\lfloor 1.32 \rfloor &= 1; & \lfloor \pi \rfloor &= 3; & \lfloor 0.98 \rfloor &= 0; \\
\lfloor -2.3 \rfloor &= -3; & \lfloor -0.4 \rfloor &= -1.
\end{align*}
\]

(e) You might have the impression that \( \lfloor x \rfloor \) is “what remains from \( x \) if the digits behind the comma are removed”. This impression is highly imprecise. For one, it is completely broken for negative \( x \) (for example, \( \lfloor -2.3 \rfloor \) is \(-3\), not \(-2\)). But more importantly, the operation of “removing the digits behind the comma” from a number is not well-defined; in fact, the periodic decimal representations 0.999\ldots and 1.000\ldots belong to the same real number (1), but removing their digits behind the comma leaves us with different integers.

(f) A related map is the map \( \mathbb{R} \to \mathbb{Z} \), \( x \mapsto \lfloor x + \frac{1}{2} \rfloor \). It sends each real \( x \) to the integer that is closest to \( x \), choosing the larger one in the case of a tie. This is one of the many things that are commonly known as “rounding” a number.

Proposition 2.8.3. Let \( a \) and \( b \) be integers such that \( b > 0 \). Then, \( \left\lfloor \frac{a}{b} \right\rfloor \) is well-defined and equals \( a \div b \).

Proof of Proposition 2.8.3. This is a rather easy and neat exercise. A full proof can be found in [Grinbe16, proof of Proposition 1.1.3].

See [Grinbe16] and [NiZuMo91, §4.1] for further properties of the floor function.
2.9. Common divisors, the Euclidean algorithm and the Bezout theorem

2.9.1. Divisors

**Definition 2.9.1.** Let \( b \in \mathbb{Z} \). The **divisors** of \( b \) are defined as the integers that divide \( b \).

Be aware that some authors use a mildly different definition of “divisors”; namely, they additionally require them to be positive. We don’t make such a requirement.

For example, the divisors of 6 are \(-6, -3, -2, -1, 1, 2, 3, 6\). Of course, the negative divisors of an integer \( b \) are merely the reflections of the positive divisors through the origin\(^{13}\) (this follows easily from Proposition 2.2.3 (a)); thus, the positive divisors are usually the only ones of interest.

Here are some basic properties of divisors:

**Proposition 2.9.2.**

(a) If \( b \in \mathbb{Z} \), then 1 and \( b \) are divisors of \( b \).

(b) The divisors of 0 are all the integers.

(c) Let \( b \in \mathbb{Z} \) be nonzero. Then, all divisors of \( b \) belong to the set \( \{-|b|, -|b| + 1, \ldots, |b|\} \setminus \{0\} \).

**Proof of Proposition 2.9.2.**

(a) Clearly, 1 | \( b \) (since \( b = 1 \cdot b \)), so that 1 is a divisor of \( b \).

Also, \( b | b \) (since \( b = b \cdot 1 \)), so that \( b \) is a divisor of \( b \).

(b) Each integer \( a \) divides 0 (since \( 0 = a \cdot 0 \)) and thus is a divisor of 0. This proves Proposition 2.9.2 (b).

(c) Let \( a \) be a divisor of \( b \). Then, \( a \) divides \( b \). In other words, \( a | b \). Hence, Proposition 2.2.3 (b) yields \( |a| \leq |b| \) (since \( b \neq 0 \)). But \( |a| \geq a \) (since \( |x| \geq x \) for each \( x \in \mathbb{R} \)), so that \( a \leq |a| \leq |b| \). Also, \( |a| \geq -a \) (since \( |x| \geq -x \) for each \( x \in \mathbb{R} \)) and thus \( -a \leq |a| \leq |b| \), so that \( a \geq -|b| \). Combining this with \( a \leq |b| \), we obtain \(-|b| \leq a \leq |b| \) and thus \( a \in \{-|b|, -|b| + 1, \ldots, |b|\} \) (since \( a \) is an integer).

From Example 2.2.2 (c), we know that 0 | \( b \) only when \( b = 0 \). Thus, we don’t have 0 | \( b \) (since \( b \neq 0 \)).

If we had \( a = 0 \), then we would have \( 0 = a \cdot b \), which would contradict the fact that we don’t have 0 | \( b \). Thus, we cannot have \( a = 0 \). Hence, \( a \neq 0 \). Combining \( a \in \{-|b|, -|b| + 1, \ldots, |b|\} \) with \( a \neq 0 \), we obtain \( a \in \{-|b|, -|b| + 1, \ldots, |b|\} \setminus \{0\} \).

We have proven this for each divisor \( a \) of \( b \). Thus, we conclude that all divisors of \( b \) belong to the set \( \{-|b|, -|b| + 1, \ldots, |b|\} \setminus \{0\} \). This proves Proposition 2.9.2 (c).

Thanks to Proposition 2.9.2, we have a method to find all divisors of an integer \( b \): If \( b = 0 \), then Proposition 2.9.2 (b) directly yields the result; otherwise, Proposition 2.9.2 (c) shows that there is only a finite set of numbers we have to check. When \( b \) is large, this is slow, but to some extent that is because the problem is computationally hard (or at least suspected to be hard).

---

\(^{13}\)Reflection through the origin" is just a poetic way to say “negative”; i.e., the reflection of a number \( a \) through the origin is \(-a\).
2.9.2. Common divisors

It is somewhat more interesting to consider the common divisors of two or more integers:

**Definition 2.9.3.** Let \( b_1, b_2, \ldots, b_k \) be integers. Then, the common divisors of \( b_1, b_2, \ldots, b_k \) are defined to be the integers \( a \) that satisfy

\[
(a \mid b_i \text{ for all } i \in \{1, 2, \ldots, k\})
\]

(in other words, that divide all of the integers \( b_1, b_2, \ldots, b_k \)). We let \( \text{Div}(b_1, b_2, \ldots, b_k) \) denote the set of these common divisors.

Note that the concept of common divisors encompasses the concept of divisors: The common divisors of a single integer \( b \) are merely the divisors of \( b \). Thus, \( \text{Div}(b) \) is the set of all divisors of \( b \) whenever \( b \in \mathbb{Z} \). (Of course, speaking of “common divisors” of just one integer is like speaking of a conspiracy of just one person. But the definition fits, and we algebraists don’t exclude cases just because they are ridiculous.)

(Also, the common divisors of an empty list of integers are all the integers, because the requirement (18) is vacuously true for \( k = 0 \). In other words, \( \text{Div}() = \mathbb{Z} \).)

Here are some more interesting examples of common divisors:

**Example 2.9.4.** (a) The common divisors of 6 and 8 are \(-2, -1, 1, 2\). (In order to see this, just observe that the divisors of 6 are \(-6, -3, -2, -1, 1, 2, 3, 6\), whereas the divisors of 8 are \(-8, -4, -2, -1, 1, 2, 4, 8\); now you can find the common divisors of 6 and 8 by taking the numbers common to these two lists.) Thus,

\[
\text{Div}(6, 8) = \{-2, -1, 1, 2\}.
\]

(b) The common divisors of 6 and 14 are \(-2, -1, 1, 2\) again. (In order to see this, just observe that the divisors of 6 are \(-6, -3, -2, -1, 1, 2, 3, 6\), whereas the divisors of 14 are \(-14, -7, -2, -1, 1, 2, 7, 14\).)

(c) The common divisors of 6, 10 and 15 are \(-1\) and 1. (In order to see this, note that:

- The divisors of 6 are \(-6, -3, -2, -1, 1, 2, 3, 6\).
- The divisors of 10 are \(-10, -5, -2, -1, 1, 2, 5, 10\).
- The divisors of 15 are \(-15, -5, -3, -1, 1, 3, 5, 15\).

The only numbers common to these three lists are \(-1\) and 1.) However:

- The common divisors of 6 and 10 are \(-2, -1, 1, 2\).
- The common divisors of 6 and 15 are \(-3, -1, 1, 3\).
• The common divisors of 10 and 15 are −5, −1, 1, 5.

This illustrates the fact that three numbers can have pairwise nontrivial common divisors (where “nontrivial” means “distinct from 1 and −1”), but the only common divisors of all three of them may still be just 1 and −1.

**Proposition 2.9.5.** Let \( b_1, b_2, \ldots, b_k \) be finitely many integers that are not all 0. Then, the set \( \text{Div}(b_1, b_2, \ldots, b_k) \) has a largest element, and this largest element is a positive integer.

**Proof of Proposition 2.9.5.** The integer 1 satisfies \((1 \mid b_i \text{ for all } i \in \{1, 2, \ldots, k\})\). Thus, 1 is a common divisor of \( b_1, b_2, \ldots, b_k \) (by the definition of “common divisor”). In other words, \( 1 \in \text{Div}(b_1, b_2, \ldots, b_k) \) (by the definition of \( \text{Div}(b_1, b_2, \ldots, b_k) \)). Hence, the set \( \text{Div}(b_1, b_2, \ldots, b_k) \) is nonempty.

Moreover, it is easy to see that the set \( \text{Div}(b_1, b_2, \ldots, b_k) \) is finite.  

[Proof: We have assumed that \( b_1, b_2, \ldots, b_k \) are not all 0. In other words, there exists a \( j \in \{1, 2, \ldots, k\} \) such that \( b_j \) is nonzero. Consider such a \( j \).

Let \( d \in \text{Div}(b_1, b_2, \ldots, b_k) \). Thus, \( d \) is a common divisor of \( b_1, b_2, \ldots, b_k \) (by the definition of \( \text{Div}(b_1, b_2, \ldots, b_k) \)). In other words, \( d \mid b_i \) for all \( i \in \{1, 2, \ldots, k\} \) (by the definition of “common divisor”). Applying this to \( i = j \), we obtain \( d \mid b_j \). Hence, \( d \) is a divisor of \( b_j \).

But Proposition 2.9.2 (applied to \( b = b_j \)) shows that all divisors of \( b_j \) belong to the set \( \{-|b_j|, -|b_j| + 1, \ldots, |b_j|\} \setminus \{0\} \). Hence, \( d \) must belong to this set (since \( d \) is a divisor of \( b_j \)). In other words, \( d \in \{-|b_j|, -|b_j| + 1, \ldots, |b_j|\} \setminus \{0\} \).

Now, forget that we fixed \( d \). We thus have shown that \( d \in \{-|b_j|, -|b_j| + 1, \ldots, |b_j|\} \setminus \{0\} \) for each \( d \in \text{Div}(b_1, b_2, \ldots, b_k) \). In other words,

\[
\text{Div}(b_1, b_2, \ldots, b_k) \subseteq \{-|b_j|, -|b_j| + 1, \ldots, |b_j|\} \setminus \{0\}.
\]

Thus, the set \( \text{Div}(b_1, b_2, \ldots, b_k) \) is finite (since the set \( \{-|b_j|, -|b_j| + 1, \ldots, |b_j|\} \setminus \{0\} \) is finite).

Now we know that the set \( \text{Div}(b_1, b_2, \ldots, b_k) \) is a nonempty finite set of integers. Thus, this set \( \text{Div}(b_1, b_2, \ldots, b_k) \) has a largest element (since every nonempty finite set of integers has a largest element). It remains to prove that this largest element is a positive integer.

Let \( g \) be this largest element. Thus, we must prove that \( g \) is a positive integer. Clearly, \( g \) is an integer (since all elements of \( \text{Div}(b_1, b_2, \ldots, b_k) \) are integers); it thus remains to show that \( g \) is positive.

The element \( g \) is the largest element of the set \( \text{Div}(b_1, b_2, \ldots, b_k) \), and thus is \( \geq \) to every element of this set. In other words, \( g \leq x \) for each \( x \in \text{Div}(b_1, b_2, \ldots, b_k) \). Applying this to \( x = 1 \), we obtain \( g \leq 1 \) (since \( 1 \in \text{Div}(b_1, b_2, \ldots, b_k) \)). Hence, \( g \) is positive. This completes the proof of Proposition 2.9.5.  

The following exercise shows that the set \( \text{Div}(b_1, b_2, \ldots, b_k) \) depends only on the set \( \{b_1, b_2, \ldots, b_k\} \), but not on the numbers \( b_1, b_2, \ldots, b_k \) themselves. Thus, for example, any integers \( a, b \) and \( c \) satisfy \( \text{Div}(a, b, c) = \text{Div}(c, a, b) \) (since \( \{a, b, c, a\} = \{c, a, b\} \)) and \( \text{Div}(a, a, b, a) = \text{Div}(a, b, a) \) (since \( \{a, a, b, a\} = \{a, b\} \)).
Exercise 2.9.1. Let \( b_1, b_2, \ldots, b_k \) be finitely many integers. Let \( c_1, c_2, \ldots, c_\ell \) be finitely many integers. Prove that if
\[
\{ b_1, b_2, \ldots, b_k \} = \{ c_1, c_2, \ldots, c_\ell \},
\]
then
\[
\text{Div} (b_1, b_2, \ldots, b_k) = \text{Div} (c_1, c_2, \ldots, c_\ell).
\]

2.9.3. Greatest common divisors

Proposition 2.9.5 allows us to make a crucial definition:

Definition 2.9.6. Let \( b_1, b_2, \ldots, b_k \) be finitely many integers. The greatest common divisor of \( b_1, b_2, \ldots, b_k \) is defined as follows:

- If \( b_1, b_2, \ldots, b_k \) are not all 0, then it is defined as the largest element of the set \( \text{Div} (b_1, b_2, \ldots, b_k) \). This largest element is well-defined (by Proposition 2.9.5), and is a positive integer (by Proposition 2.9.5 again).

- If \( b_1, b_2, \ldots, b_k \) are all 0, then it is defined to be 0. (This is a slight abuse of the word "greatest common divisor", because 0 is not actually the greatest among the common divisors of \( b_1, b_2, \ldots, b_k \) in this case. In fact, when \( b_1, b_2, \ldots, b_k \) are all 0, every integer is a common divisor of \( b_1, b_2, \ldots, b_k \), so that there is no greatest among these common divisors, because there is no "greatest integer". Nevertheless, defining the greatest common divisor of \( b_1, b_2, \ldots, b_k \) to be 0 in this case will prove to be a good decision, as it will greatly reduce the number of exceptions in our results.)

Thus, in either case, this greatest common divisor is a nonnegative integer. We denote it by \( \gcd (b_1, b_2, \ldots, b_k) \). (Some authors also call it \((b_1, b_2, \ldots, b_k)\), which is rather dangerous as the same notation stands for a \( k \)-tuple. We shall avoid this notation at all cost, but you should be aware of it when reading number-theoretical literature.)

We shall also use the word "\( \gcd \)" as shorthand for "greatest common divisor".

The greatest common divisors you will most commonly see are those of two integers. Indeed, any other \( \gcd \) can be rewritten in terms of these: for example,
\[
\gcd (a, b, c, d, e) = \gcd (a, \gcd (b, \gcd (c, \gcd (d, e))))
\]
for all \( a, b, c, d, e \in \mathbb{Z} \). This is, in fact, a consequence of Proposition 2.9.20(d) (which we will prove later), applied several times.

First, let us observe several properties of greatest common divisors:
Proposition 2.9.7. (a) We have \( \gcd(a, 0) = \gcd(a) = |a| \) for all \( a \in \mathbb{Z} \).
(b) We have \( \gcd(a, b) = \gcd(b, a) \) for all \( a, b \in \mathbb{Z} \).
(c) We have \( \gcd(a, au + b) = \gcd(a, b) \) for all \( a, b, u \in \mathbb{Z} \).
(d) If \( a, b, c \in \mathbb{Z} \) satisfy \( b \equiv c \mod a \), then \( \gcd(a, b) = \gcd(a, c) \).
(e) If \( a, b \in \mathbb{Z} \) are such that \( a \) is positive, then \( \gcd(a, b) = \gcd(a, b\%a) \).
(f) We have \( \gcd(a, b) \mid a \) and \( \gcd(a, b) \mid b \) for all \( a, b \in \mathbb{Z} \).
(g) We have \( \gcd(-a, b) = \gcd(a, b) \) for all \( a, b \in \mathbb{Z} \).
(h) We have \( \gcd(a, -b) = \gcd(a, b) \) for all \( a, b \in \mathbb{Z} \).
(i) If \( a, b \in \mathbb{Z} \) satisfy \( a \mid b \), then \( \gcd(a, b) = |a| \).
(j) The greatest common divisor of the empty list of integers is \( \gcd() = 0 \).

Proposition 2.9.7 is not difficult and we could start proving it right away. However, such a proof would require some annoying case distinctions due to the special treatment that the “\( b_1, b_2, \ldots, b_k \) are all 0” case required in Definition 2.9.6. Fortunately, we can circumnavigate these annoyances by stating a simple rule for how the \( \gcd \) of \( k \) integers \( b_1, b_2, \ldots, b_k \) can be computed from their set of common divisors (including the case when \( b_1, b_2, \ldots, b_k \) are all 0):

**Lemma 2.9.8.** Let \( b_1, b_2, \ldots, b_k \) be finitely many integers. Then,

\[
\gcd(b_1, b_2, \ldots, b_k) = \begin{cases} 
\max(\text{Div}(b_1, b_2, \ldots, b_k)), & \text{if } 0 \notin \text{Div}(b_1, b_2, \ldots, b_k); \\
0, & \text{if } 0 \in \text{Div}(b_1, b_2, \ldots, b_k). 
\end{cases}
\]

(Here, \( \max S \) denotes the largest element of a set \( S \) of integers, whenever this largest element exists.)

**Proof of Lemma 2.9.8** We are in one of the following two cases:

**Case 1:** The integers \( b_1, b_2, \ldots, b_k \) are not all 0.

**Case 2:** The integers \( b_1, b_2, \ldots, b_k \) are all 0.

Let us consider Case 1 first. In this case, the integers \( b_1, b_2, \ldots, b_k \) are not all 0. Hence, \( \gcd(b_1, b_2, \ldots, b_k) \) is defined as the largest element of the set \( \text{Div}(b_1, b_2, \ldots, b_k) \) (by Definition 2.9.6). In other words,

\[
\gcd(b_1, b_2, \ldots, b_k) = \max(\text{Div}(b_1, b_2, \ldots, b_k)). \tag{19}
\]

On the other hand, \( 0 \notin \text{Div}(b_1, b_2, \ldots, b_k) \). Hence,

\[
\begin{align*}
\max(\text{Div}(b_1, b_2, \ldots, b_k)), & \quad \text{if } 0 \notin \text{Div}(b_1, b_2, \ldots, b_k); \\
0, & \quad \text{if } 0 \in \text{Div}(b_1, b_2, \ldots, b_k)
\end{align*}
\]

\(\text{proof.}\) Assume the contrary. Thus, \( 0 \in \text{Div}(b_1, b_2, \ldots, b_k) \). In other words, \( 0 \) is a common divisor of \( b_1, b_2, \ldots, b_k \) (by the definition of \( \text{Div}(b_1, b_2, \ldots, b_k) \)). In other words, \( 0 \mid b_i \) for all \( i \in \{1, 2, \ldots, k\} \) (by the definition of “common divisor”). Thus, for all \( i \in \{1, 2, \ldots, k\} \), we have \( b_i = 0 \) (since \( 0 \mid b_i \), so that \( b_i = 0c \) for some integer \( c \); but this yields \( b_i = 0c = 0 \)). In other words, \( b_1, b_2, \ldots, b_k \) are all 0. But this contradicts the fact that \( b_1, b_2, \ldots, b_k \) are not all 0. This contradiction shows that our assumption was false; qed.
Comparing this with (19), we obtain
\[ \gcd(b_1, b_2, \ldots, b_k) = \begin{cases} 
\max(\Div(b_1, b_2, \ldots, b_k)), & \text{if } 0 \not\in \Div(b_1, b_2, \ldots, b_k); \\
0, & \text{if } 0 \in \Div(b_1, b_2, \ldots, b_k). 
\end{cases} \]

Hence, Lemma 2.9.8 is proven in Case 1.

Let us now consider Case 2. In this case, the integers \(b_1, b_2, \ldots, b_k\) are all 0. Hence, \(\gcd(b_1, b_2, \ldots, b_k)\) is defined as 0 (by Definition 2.9.6). In other words,
\[ \gcd(b_1, b_2, \ldots, b_k) = 0. \quad (20) \]

On the other hand, \(0 \in \Div(b_1, b_2, \ldots, b_k)\) \(^{15}\). Hence,
\[ \begin{cases} 
\max(\Div(b_1, b_2, \ldots, b_k)), & \text{if } 0 \not\in \Div(b_1, b_2, \ldots, b_k); \\
0, & \text{if } 0 \in \Div(b_1, b_2, \ldots, b_k) 
\end{cases} = 0. \]

Comparing this with (20), we obtain
\[ \gcd(b_1, b_2, \ldots, b_k) = \begin{cases} 
\max(\Div(b_1, b_2, \ldots, b_k)), & \text{if } 0 \not\in \Div(b_1, b_2, \ldots, b_k); \\
0, & \text{if } 0 \in \Div(b_1, b_2, \ldots, b_k). 
\end{cases} \]

Hence, Lemma 2.9.8 is proven in Case 2.

We have now proven Lemma 2.9.8 in both Cases 1 and 2. Thus, Lemma 2.9.8 always holds.

A corollary of Lemma 2.9.8 is the following:

**Lemma 2.9.9.** Let \(b_1, b_2, \ldots, b_k\) be finitely many integers. Let \(c_1, c_2, \ldots, c_\ell\) be finitely many integers. If
\[ \Div(b_1, b_2, \ldots, b_k) = \Div(c_1, c_2, \ldots, c_\ell), \]
then
\[ \gcd(b_1, b_2, \ldots, b_k) = \gcd(c_1, c_2, \ldots, c_\ell). \]

**Proof of Lemma 2.9.9** Assume that \(\Div(b_1, b_2, \ldots, b_k) = \Div(c_1, c_2, \ldots, c_\ell)\). Lemma 2.9.8 yields
\[ \gcd(b_1, b_2, \ldots, b_k) = \begin{cases} 
\max(\Div(b_1, b_2, \ldots, b_k)), & \text{if } 0 \not\in \Div(b_1, b_2, \ldots, b_k); \\
0, & \text{if } 0 \in \Div(b_1, b_2, \ldots, b_k) 
\end{cases} = \begin{cases} 
\max(\Div(c_1, c_2, \ldots, c_\ell)), & \text{if } 0 \not\in \Div(c_1, c_2, \ldots, c_\ell); \\
0, & \text{if } 0 \in \Div(c_1, c_2, \ldots, c_\ell) 
\end{cases} \]

\(^{15}\)Proof. The integers \(b_1, b_2, \ldots, b_k\) are all 0. In other words, \(b_i = 0\) for all \(i \in \{1, 2, \ldots, k\}\). Hence, \(0 \mid b_i\) for all \(i \in \{1, 2, \ldots, k\}\) (since each \(i \in \{1, 2, \ldots, k\}\) satisfies \(b_i = 0 = 0 \cdot 0\)). In other words, 0 is a common divisor of \(b_1, b_2, \ldots, b_k\) (by the definition of “common divisor”). In other words, \(0 \in \Div(b_1, b_2, \ldots, b_k)\) (by the definition of \(\Div(b_1, b_2, \ldots, b_k)\)).
(since \(\text{Div}(b_1, b_2, \ldots, b_k) = \text{Div}(c_1, c_2, \ldots, c_\ell)\)). But Lemma \(^{2.9.8}\) (applied to \(c_1, c_2, \ldots, c_\ell\) instead of \(b_1, b_2, \ldots, b_k\)) yields

\[
\gcd(c_1, c_2, \ldots, c_\ell) = \begin{cases} 
\max(\text{Div}(c_1, c_2, \ldots, c_\ell)), & \text{if } 0 \notin \text{Div}(c_1, c_2, \ldots, c_\ell); \\
0, & \text{if } 0 \in \text{Div}(c_1, c_2, \ldots, c_\ell).
\end{cases}
\]

Comparing these two equalities, we obtain \(\gcd(b_1, b_2, \ldots, b_k) = \gcd(c_1, c_2, \ldots, c_\ell)\). This proves Lemma \(^{2.9.9}\). \(\square\)

**Proof of Proposition \(^{2.9.7}\) (a)** Here is a sketch of the proof: The number 0 is a “joker” when it comes to common divisors: For example, if \(a \in \mathbb{Z}\), then the common divisors of \(a\) and 0 are the same as the divisors of \(a\), because every integer divides 0. Thus, if \(a \in \mathbb{Z}\) is nonzero, then the greatest common divisor of \(a\) and 0 is the greatest divisor of \(a\), which is \(|a|\) (an easy consequence of Proposition \(^{2.9.2}\) (b)).

For the sake of completeness, let us give a detailed proof of Proposition \(^{2.9.7}\) (a):

Let \(a \in \mathbb{Z}\). Definition \(^{2.9.6}\) (specifically, its case when \(b_1, b_2, \ldots, b_k\) are all 0) shows that \(\gcd(0, 0) = 0\) and \(\gcd(0) = 0\). Combining this with \(|0| = 0\), we obtain \(\gcd(0, 0) = \gcd(0) = |0|\). In other words, Proposition \(^{2.9.7}\) (a) holds if \(a = 0\). Thus, for the rest of this proof, we WLOG assume that \(a \neq 0\). Hence, the two integers \(a, 0\) are not all zero. Thus, \(\gcd(a, 0)\) is defined to be the largest element of the set \(\text{Div}(a, 0)\) (by Definition \(^{2.9.6}\)). Likewise, \(\gcd(a)\) is the largest element of the set \(\text{Div}(a)\).

We shall now prove that \(\text{Div}(a, 0) = \text{Div}(a)\). Indeed, for any integer \(x\), we have the following chain of equivalences:

\[
(x \in \text{Div}(a, 0)) \\
\iff (x \text{ is a common divisor of } a \text{ and } 0) \quad \text{(by the definition of } \text{Div}(a, 0)) \\
\iff (x | a \text{ and } x | 0) \quad \text{(by the definition of a “common divisor”)} \\
\iff (x | a) \quad \text{(since } x | 0 \text{ always holds (since } 0 = x \cdot 0)) \\
\iff (x \text{ is a common divisor of } a) \quad \text{(by the definition of a “common divisor”)} \\
\iff (x \in \text{Div}(a)) \quad \text{(by the definition of } \text{Div}(a)).
\]

In other words, an integer belongs to \(\text{Div}(a, 0)\) if and only if it belongs to \(\text{Div}(a)\). Thus, \(\text{Div}(a, 0) = \text{Div}(a)\) (since both \(\text{Div}(a, 0)\) and \(\text{Div}(a)\) are sets of integers). Thus, Lemma \(^{2.9.9}\) (applied to \((a, 0)\) and \((a)\) instead of \((b_1, b_2, \ldots, b_k)\) and \((c_1, c_2, \ldots, c_\ell)\)) yields \(\gcd(a, 0) = \gcd(a)\).

For any integer \(x\), we have the following chain of equivalences:

\[
(x \in \text{Div}(a)) \\
\iff (x \text{ is a common divisor of } a) \quad \text{(by the definition of } \text{Div}(a)) \\
\iff (x | a) \quad \text{(by the definition of a “common divisor”)} \\
\iff (x \text{ is a divisor of } a).
\]

Thus, \(\text{Div}(a)\) is the set of all divisors of \(a\).

**Exercise \(^{2.2.1}\) (b)** yields \(|a| | a\). In other words, \(|a|\) is a divisor of \(a\).

Moreover, \(a\) is nonzero (since \(a \neq 0\)). Hence, Proposition \(^{2.9.2}\) (b) (applied to \(b = a\)) shows that all divisors of \(a\) belong to the set \(\{-|a|, -|a| + 1, \ldots, |a|\}\). Hence, they belong to the set \(\{-|a|, -|a| + 1, \ldots, |a|\}\), and thus are \(|a|\).
Recall that $|a|$ is a divisor of $a$. Since we also know that all divisors of $a$ are $\leq |a|$, we can thus conclude that $|a|$ is the largest divisor of $a$. In other words, $|a|$ is the largest element of the set $\text{Div}(a)$ (since $\text{Div}(a)$ is the set of all divisors of $a$). In other words, $|a|$ is gcd$(a)$ (since gcd$(a)$ is the largest element of the set $\text{Div}(a)$). Thus, gcd$(a) = |a|$. Combining this with gcd$(a,0) = \text{gcd}(a)$, this yields gcd$(a,0) = \text{gcd}(a) = |a|$. Thus, Proposition 2.9.7(a) is finally proven.

(b) For any integer $x$, we have the following chain of equivalences:

\[(x \in \text{Div}(a,b)) \iff (x \text{ is a common divisor of } a \text{ and } b) \quad \text{(by the definition of } \text{Div}(a,b)) \]
\[
\iff (x | a \text{ and } x | b) \quad \text{(by the definition of a "common divisor")}
\]
\[
\iff (x | b \text{ and } x | a) \quad \text{(by the definition of a "common divisor")}
\]
\[
\iff (x \text{ is a common divisor of } b \text{ and } a) 
\quad \text{(by the definition of a "common divisor")}
\]
\[
\iff (x \in \text{Div}(b,a)) \quad \text{(by the definition of } \text{Div}(b,a)).
\]

In other words, an integer belongs to $\text{Div}(a,b)$ if and only if it belongs to $\text{Div}(b,a)$. Thus, $\text{Div}(a,b) = \text{Div}(b,a)$ (since both $\text{Div}(a,b)$ and $\text{Div}(b,a)$ are sets of integers). Thus, Lemma 2.9.9 (applied to $(a,b)$ and $(b,a)$ instead of $(b_1,b_2,\ldots,b_k)$ and $(c_1,c_2,\ldots,c_l)$) yields gcd$(a,b) = \text{gcd}(b,a)$. This proves Proposition 2.9.7(b).

Let us prove part (d) now, and then derive part (c) from it.

(d) Let $a,b,c \in \mathbb{Z}$ satisfy $b \equiv c \text{ mod } a$. We must prove that gcd$(a,b) = \text{gcd}(a,c)$. To do so, we shall first prove that $\text{Div}(a,b) = \text{Div}(a,c)$.

From $b \equiv c \text{ mod } a$, we obtain $c \equiv b \text{ mod } a$ (by Proposition 2.3.4(c)). Hence, our situation is symmetric with respect to $b$ and $c$.

We shall now show that $\text{Div}(a,b) \subseteq \text{Div}(a,c)$. Indeed, let $x \in \text{Div}(a,b)$. Then, $x \mid a$ and $x \mid b$ (by the definition of $\text{Div}(a,b)$). In other words, $x \mid a$ and $x \mid b$ (by the definition of a "common divisor"). From $x \mid b$, we obtain $b \equiv 0 \text{ mod } x$. But from $x \mid a$ and $c \equiv b \text{ mod } a$, we obtain $c \equiv b \text{ mod } x$ (by Proposition 2.3.4(e), applied to $a$, $x$, $c$ and $b$ instead of $n$, $m$, $c$ and $b$). Thus, $c \equiv b \equiv 0 \text{ mod } x$, so that $x \mid c$. Combining $x \mid a$ and $x \mid c$, we see that $x$ is a common divisor of $a$ and $c$ (by the definition of a "common divisor"). In other words, $x \in \text{Div}(a,c)$ (by the definition of $\text{Div}(a,c)$).

Now, forget that we fixed $x$. We thus have proven that $x \in \text{Div}(a,c)$ for each $x \in \text{Div}(a,b)$. In other words, $\text{Div}(a,b) \subseteq \text{Div}(a,c)$.

The same argument (but with the roles of $b$ and $c$ swapped) shows that $\text{Div}(a,c) \subseteq \text{Div}(a,b)$ (since our situation is symmetric with respect to $b$ and $c$). Combining this with $\text{Div}(a,b) \subseteq \text{Div}(a,c)$, we obtain $\text{Div}(a,b) = \text{Div}(a,c)$. Thus, Lemma 2.9.9 (applied to $(a,b)$ and $(a,c)$ instead of $(b_1,b_2,\ldots,b_k)$ and $(c_1,c_2,\ldots,c_l)$) yields gcd$(a,b) = \text{gcd}(a,c)$. This proves Proposition 2.9.7(d).

(c) Let $a,b,u \in \mathbb{Z}$. Then, $ua + b \equiv b \text{ mod } a$ (since $(ua + b) - b = ua$ is clearly divisible by $a$). Thus, Proposition 2.9.7(d) (applied to $ua + b$ and $b$ instead of $b$ and $c$) yields gcd$(a,ua + b) = \text{gcd}(a,b)$. This proves Proposition 2.9.7(c).
(e) Let \( a, b \in \mathbb{Z} \) be such that \( a \) is positive. Then, \( b \equiv b \mod a \) (by Corollary 2.6.9 (a), applied to \( a \) and \( b \) instead of \( n \) and \( u \)), thus \( b \equiv b \mod a \). Hence, 
\[
\text{gcd}(a, b) = \text{gcd}(b, a) \quad \text{(by Proposition 2.9.7 (d), applied to } c = b \mod a).
\]
This proves Proposition 2.9.7 (e).

(f) Let \( a, b \in \mathbb{Z} \). We must prove that \( \text{gcd}(a, b) \mid a \) and \( \text{gcd}(a, b) \mid b \).

If the two integers \( a, b \) are all 0, then this is obvious.\(^{16}\) Hence, for the rest of this proof, WLOG assume that \( a, b \) are not all 0. Thus, \( \text{gcd}(a, b) \) is defined to be the largest element of the set \( \text{Div}(a, b) \) (by Definition 2.9.6). Hence, \( \text{gcd}(a, b) \) is an element of this set \( \text{Div}(a, b) \). In other words, \( \text{gcd}(a, b) \) is a common divisor of \( a \) and \( b \) (by the definition of \( \text{Div}(a, b) \)). In other words, \( \text{gcd}(a, b) \mid a \) and \( \text{gcd}(a, b) \mid b \).

This proves Proposition 2.9.7 (f).

(g) Let \( a, b \in \mathbb{Z} \). We must prove that \( \text{gcd}(-a, b) = \text{gcd}(a, b) \). Again, we shall achieve this via showing that \( \text{Div}(-a, b) = \text{Div}(a, b) \).

First, we will show that \( \text{Div}(a, b) \subseteq \text{Div}(-a, b) \). Indeed, let \( x \in \text{Div}(a, b) \). Then, \( x \) is a common divisor of \( a \) and \( b \) (by the definition of \( \text{Div}(a, b) \)). In other words, \( x \mid a \) and \( x \mid b \) (by the definition of a "common divisor"). We have \( a \mid -a \) (since \( -a = a \cdot (-1) \)). Thus, \( x \mid a \mid -a \). Combining \( x \mid -a \) and \( x \mid b \), we see that \( x \) is a common divisor of \( -a \) and \( b \) (by the definition of a "common divisor"). In other words, \( x \in \text{Div}(-a, b) \).

Now, forget that we fixed \( x \). We thus have proven that \( x \in \text{Div}(-a, b) \) for each \( x \in \text{Div}(a, b) \). In other words, \( \text{Div}(a, b) \subseteq \text{Div}(-a, b) \).

The same argument (but applied to \( -a \) instead of \( a \)) shows that \( \text{Div}(-a, b) \subseteq \text{Div}(-(-a), b) \). Since \( -(-a) = a \), this rewrites as \( \text{Div}(-a, b) \subseteq \text{Div}(a, b) \). Combining this with \( \text{Div}(a, b) \subseteq \text{Div}(-a, b) \), we obtain \( \text{Div}(-a, b) = \text{Div}(a, b) \). Thus, \( \text{gcd}(-a, b) = \text{gcd}(a, b) \). This proves Proposition 2.9.7 (g).

(h) We can prove this similarly to how we just proved Proposition 2.9.7 (g), but it is easier to derive it from what was already shown.

Let \( a, b \in \mathbb{Z} \). Proposition 2.9.7 (b) (applied to \( -b \) instead of \( b \)) yields
\[
\text{gcd}(a, -b) = \text{gcd}(-b, a) = \text{gcd}(b, a) \quad \text{(by Proposition 2.9.7 (g), applied to } b \text{ and } a \text{ instead of } a \text{ and } b).
\]
This proves Proposition 2.9.7 (h).

(i) Let \( a, b \in \mathbb{Z} \) satisfy \( a \mid b \). From \( a \mid b \), we obtain \( b \equiv 0 \mod a \). Hence, Proposition 2.9.7 (d) (applied to \( c = 0 \)) yields \( \text{gcd}(a, b) = \text{gcd}(a, 0) = |a| \) (by Proposition 2.9.7 (a)). This proves Proposition 2.9.7 (i).

(j) The empty list of integers \( () \) has the property that all its entries are 0 (indeed, this is vacuously true because it has no entries at all). Thus, its greatest common divisor is defined to be 0 (by the "If \( b_1, b_2, \ldots, b_k \) are not all 0" case of Definition 2.9.6). In other words, \( \text{gcd}() = 0 \). This proves Proposition 2.9.7 (j). \( \Box \)

\(^{16}\)Proof. Assume that \( a, b \) are all 0. Then, \( a = 0 = \text{gcd}(a, b) \cdot 0 \), so that \( \text{gcd}(a, b) \mid a \); similarly, \( \text{gcd}(a, b) \mid b \). Hence, we have proven that \( \text{gcd}(a, b) \mid a \) and \( \text{gcd}(a, b) \mid b \) if the integers \( a, b \) are all 0.
Remark 2.9.10. Proposition 2.9.7(c) says that if we add a multiple of $a$ to $b$, then $\gcd(a,b)$ does not change. Similarly, if we add a multiple of $b$ to $a$, then $\gcd(a,b)$ does not change (i.e., we have $\gcd(vb + a, b) = \gcd(a, b)$ for all $a,b, v \in \mathbb{Z}$).

However, if we simultaneously add a multiple of $a$ to $b$ and a multiple of $b$ to $a$, then $\gcd(a,b)$ may well change: i.e., we may have $\gcd(vb + a, ua + b) \neq \gcd(a, b)$ for all $a,b,u,v \in \mathbb{Z}$. Examples are easy to find (just take $v = 1$ and $u = 1$).

Proposition 2.9.7 gives a quick way to compute $\gcd(a,b)$ for two nonnegative integers $a$ and $b$, by repeatedly applying division with remainder. For example, let us compute $\gcd(210, 45)$ as follows:

$$\gcd(210, 45) = \gcd(45, 210) \quad \text{(by Proposition 2.9.7(b))}$$

$$= \gcd(45, 210 \mod 45) \quad \text{(by Proposition 2.9.7(e))}$$

$$= \gcd(45, 30) = \gcd(30, 45) \quad \text{(by Proposition 2.9.7(b))}$$

$$= \gcd(30, 45 \mod 30) \quad \text{(by Proposition 2.9.7(e))}$$

$$= \gcd(30, 15) = \gcd(15, 30) \quad \text{(by Proposition 2.9.7(b))}$$

$$= \gcd(15, 30 \mod 15) \quad \text{(by Proposition 2.9.7(e))}$$

$$= \gcd(15, 0) = 15 \quad \text{(by Proposition 2.9.7(a))}$$

This method of computing $\gcd(a,b)$ is called the Euclidean algorithm, and is usually much faster than the divisors of $a$ or the divisors of $b$ can be found!

The following exercise shows that the number $\gcd(b_1,b_2,\ldots,b_k)$ depends only on the set $\{b_1,b_2,\ldots,b_k\}$, but not on the numbers $b_1,b_2,\ldots,b_k$ themselves. Thus, for example, any integers $a$, $b$ and $c$ satisfy $\gcd(a,b,c,a) = \gcd(c,a,b)$ (since $\{a,b,c,a\} = \{c,a,b\}$) and $\gcd(a,a,b,a) = \gcd(a,b,a)$ (since $\{a,a,b,a\} = \{a,b,b\}$).

Exercise 2.9.2. Let $b_1,b_2,\ldots,b_k$ be finitely many integers. Let $c_1,c_2,\ldots,c_\ell$ be finitely many integers. Prove that if

$$\{b_1,b_2,\ldots,b_k\} = \{c_1,c_2,\ldots,c_\ell\},$$

then

$$\gcd(b_1,b_2,\ldots,b_k) = \gcd(c_1,c_2,\ldots,c_\ell).$$
2.9.4. Bezout’s theorem

The following fact about gcds is one of the most important facts in number theory:

**Theorem 2.9.11.** Let $a$ and $b$ be two integers. Then, there exist integers $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that

$$\gcd(a, b) = xa + yb.$$ 

Theorem 2.9.11 is often stated as follows: “If $a$ and $b$ are two integers, then $\gcd(a, b)$ is a $\mathbb{Z}$-linear combination of $a$ and $b$”. The notion “$\mathbb{Z}$-linear combination of $a$ and $b$” simply means “a number of the form $xa + yb$ with $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$” (this is exactly the notion of a “linear combination” in linear algebra, except that now the scalars must come from $\mathbb{Z}$), so this is just a restatement of Theorem 2.9.11.

Theorem 2.9.11 is known as Bezout’s theorem (or Bezout’s identity)\(^\text{17}\). We shall prove it in several steps. The first step is to show it when $a$ and $b$ are nonnegative:

**Lemma 2.9.12.** Let $a \in \mathbb{N}$ and $b \in \mathbb{N}$. Then, there exist integers $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that

$$\gcd(a, b) = xa + yb.$$ 

**Proof of Lemma 2.9.12.** The following proof uses a strategy similar to the Euclidean algorithm (making one of $a$ and $b$ smaller repeatedly until one of $a$ and $b$ becomes 0), and can in fact be viewed as a “protocol” of the algorithm\(^\text{18}\).

We use strong induction on $a + b$. Thus, we fix an $n \in \mathbb{N}$, and assume (as induction hypothesis) that Lemma 2.9.12 holds whenever $a + b < n$. We must now prove that Lemma 2.9.12 holds whenever $a + b = n$.

We have assumed that Lemma 2.9.12 holds whenever $a + b < n$. In other words, the following statement holds:

**Statement 1:** Let $a \in \mathbb{N}$ and $b \in \mathbb{N}$ be such that $a + b < n$. Then, there exist integers $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $\gcd(a, b) = xa + yb$.

Now, we must prove that Lemma 2.9.12 holds whenever $a + b = n$. Let us first prove this in the case when $b \geq a$:

**Statement 2:** Let $a \in \mathbb{N}$ and $b \in \mathbb{N}$ be such that $a + b = n$ and $b \geq a$. Then, there exist integers $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $\gcd(a, b) = xa + yb$.

---

\(^{17}\)or Bezout’s theorem for integers if you want to be more precise (as there are similar theorems for other objects)

\(^{18}\)or, rather, of a more primitive version of the Euclidean algorithm, in which we apply not the full power of Proposition 2.9.7\(^{(e)}\) but only the identity $\gcd(a, b) = \gcd(a, b - a)$
[Proof of Statement 2: We are in one of the following two cases:

Case 1: We have \( a = 0 \).

Case 2: We have \( a \neq 0 \).

Let us first consider Case 1. In this case, we have \( a = 0 \). Now, Proposition 2.9.7 (a) (applied to \( b \) instead of \( a \)) yields \( \gcd(b, 0) = \gcd(b) = |b| \in \{b, -b\} \). In other words, \( \gcd(b, 0) = ub \) for some \( u \in \{1, -1\} \). Consider this \( u \). Now, Proposition 2.9.7 (b) yields

\[
\gcd(a, b) = \gcd\left( b, \frac{a}{u} \right) = \gcd(b, 0) = ub = 0a + ub.
\]

Hence, there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, b) = xa + yb \) (namely, \( x = 0 \) and \( y = u \)). Thus, Statement 2 is proven in Case 1.

Let us next consider Case 2. In this case, we have \( a \neq 0 \) (since \( a \in \mathbb{N} \)), so that \( a + b > b \). Hence, \( b < a + b = n \).

From \( b \geq a \), we obtain \( b - a \in \mathbb{N} \). Moreover, \( a \in \mathbb{N} \) and \( b - a \in \mathbb{N} \) satisfy \( a + (b - a) = b < n \). Therefore, we can apply Statement 1 to \( b - a \) instead of \( b \). Thus we obtain that there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, b - a) = xa + y(b - a) \). Fix two such integers \( x \) and \( y \), and denote them by \( x_0 \) and \( y_0 \). Thus, \( x_0 \) and \( y_0 \) are two integers such that \( \gcd(a, b - a) = x_0a + y_0(b - a) \).

Also, Proposition 2.9.7 (c) (applied to \( u = -1 \)) yields \( \gcd(a, (-1)a + b) = \gcd(a, b) \). Hence,

\[
\gcd(a, b) = \gcd\left( a, (-1)a + b \right) = \gcd(a, b - a) = x_0a + y_0(b - a) = x_0a + y_0b - y_0a = (x_0 - y_0)a + y_0b.
\]

Hence, there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, b) = xa + yb \) (namely, \( x = x_0 - y_0 \) and \( y = y_0 \)). Thus, Statement 2 is proven in Case 2.

We have now proven Statement 2 in both Cases 1 and 2. Hence, Statement 2 is always proven.

Now, we can prove that Lemma 2.9.12 holds whenever \( a + b = n \):

Statement 3: Let \( a \in \mathbb{N} \) and \( b \in \mathbb{N} \) be such that \( a + b = n \). Then, there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, b) = xa + yb \).

[Proof of Statement 3: We are in one of the following two cases:

Case 1: We have \( b \geq a \).

Case 2: We have \( b < a \).

Let us first consider Case 1. In this case, we have \( b \geq a \). Hence, Statement 2 shows that there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, b) = xa + yb \). Thus, Statement 3 is proven in Case 1.
Let us next consider Case 2. In this case, we have \( b < a \). Hence, \( a > b \), so that \( a \geq b \). This shows that we can apply Statement 2 to \( b \) and \( a \) instead of \( a \) and \( b \). Thus we obtain that there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(b, a) = xb + ya \). Fix two such integers \( x \) and \( y \), and denote them by \( x_0 \) and \( y_0 \). Thus, \( x_0 \) and \( y_0 \) are two integers such that \( \gcd(b, a) = x_0b + y_0a \). Now, Proposition 2.9.7(b) yields \( \gcd(a, b) = \gcd(b, a) = x_0b + y_0a = y_0a + x_0b \). Hence, there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, b) = xa + yb \) (namely, \( x = y_0 \) and \( y = x_0 \)). Thus, Statement 2 is proven in Case 2.

We have now proven Statement 3 in both Cases 1 and 2. Hence, Statement 3 is always proven.

By proving Statement 3, we have shown that Lemma 2.9.12 holds whenever \( a + b = n \). This completes the induction step. Thus, Lemma 2.9.12 is proven by strong induction.

Next, we shall prove Theorem 2.9.11 when \( a \in \mathbb{N} \) but \( b \) may be negative:

Lemma 2.9.13. Let \( a \in \mathbb{N} \) and \( b \in \mathbb{Z} \). Then, there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that

\[
\gcd(a, b) = xa + yb.
\]

Proof of Lemma 2.9.13. We are in one of the following two cases:

Case 1: We have \( b \geq 0 \).

Case 2: We have \( b < 0 \).

Let us first consider Case 1. In this case, we have \( b \geq 0 \). Thus, \( b \in \mathbb{N} \) (since \( b \in \mathbb{Z} \)). Therefore, Lemma 2.9.12 shows that there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, b) = xa + yb \). Thus, Lemma 2.9.13 is proven in Case 1.

Let us now consider Case 2. In this case, we have \( b < 0 \). Hence, \( -b > 0 \), so that \( -b \in \mathbb{N} \) (since \( -b \in \mathbb{Z} \)). Therefore, Lemma 2.9.12 (applied to \( -b \) instead of \( b \)) shows that there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, -b) = xa + y(-b) \). Fix such integers, and denote them by \( x_0 \) and \( y_0 \). Thus, \( x_0 \in \mathbb{Z} \) and \( y_0 \in \mathbb{Z} \) are integers such that \( \gcd(a, -b) = x_0a + y_0(-b) \).

Now, Proposition 2.9.7(h) yields \( \gcd(a, -b) = \gcd(a, b) \). Hence,

\[
\gcd(a, b) = \gcd(a, -b) = x_0a + y_0(-b) = x_0a + (-y_0)b.
\]

Hence, there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, b) = xa + yb \) (namely, \( x = x_0 \) and \( y = -y_0 \)). Thus, Lemma 2.9.13 is proven in Case 2.

We have now proven Lemma 2.9.13 in both Cases 1 and 2. Hence, Lemma 2.9.13 is proven.

Now, we can prove the whole Theorem 2.9.11

Proof of Theorem 2.9.11. Theorem 2.9.11 can be derived from Lemma 2.9.13 in the same way as Lemma 2.9.13 was derived from Lemma 2.9.12 (except that this time, we have to distinguish between the cases \( a \geq 0 \) and \( a < 0 \), and we have to use
Proposition 2.9.7 (g) instead of Proposition 2.9.7 (h)). Again, let us give the detailed argument for the sake of completeness:

We are in one of the following two cases:

Case 1: We have $a \geq 0$.
Case 2: We have $a < 0$.

Let us first consider Case 1. In this case, we have $a \geq 0$. Thus, $a \in \mathbb{N}$ (since $a \in \mathbb{Z}$). Therefore, Lemma 2.9.13 shows that there exist integers $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $\gcd(a, b) = xa + yb$. Thus, Theorem 2.9.11 is proven in Case 1.

Let us now consider Case 2. In this case, we have $a < 0$. Hence, $-a > 0$, so that $-a \in \mathbb{N}$ (since $-a \in \mathbb{Z}$). Therefore, Lemma 2.9.13 (applied to $-a$ instead of $a$) shows that there exist integers $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $\gcd(-a, b) = x(-a) + yb$. Fix such integers, and denote them by $x_0$ and $y_0$. Thus, $x_0 \in \mathbb{Z}$ and $y_0 \in \mathbb{Z}$ are integers such that $\gcd(-a, b) = x_0(-a) + y_0b$.

Now, Proposition 2.9.7 (g) yields $\gcd(-a, b) = \gcd(a, b)$. Hence, $\gcd(a, b) = \gcd(-a, b) = x_0(-a) + y_0b = (-x_0)a + y_0b$.

Hence, there exist integers $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $\gcd(a, b) = xa + yb$ (namely, $x = -x_0$ and $y = y_0$). Thus, Theorem 2.9.11 is proven in Case 2.

We have now proven Theorem 2.9.11 in both Cases 1 and 2. Hence, Theorem 2.9.11 is proven.

Exercise 2.9.3. Let $u$ be an integer.

(a) Prove that $u^b - 1 \equiv u^a - 1 \mod u^{b-a} - 1$ for any $a \in \mathbb{N}$ and $b \in \mathbb{N}$ satisfying $b \geq a$.

(b) Prove that $\gcd(u^a - 1, u^b - 1) = \left| u^{\gcd(a, b)} - 1 \right|$ for all $a \in \mathbb{N}$ and $b \in \mathbb{N}$.

2.9.5. First applications of Bezout’s theorem

An important corollary of Theorem 2.9.11 is the following fact:

Theorem 2.9.14. Let $a, b \in \mathbb{Z}$. Then:

(a) For each $m \in \mathbb{Z}$, we have the following logical equivalence:

$$ (m \mid a \text{ and } m \mid b) \iff (m \mid \gcd(a, b)). $$

(b) The common divisors of $a$ and $b$ are precisely the divisors of $\gcd(a, b)$.

(c) We have $\text{Div}(a, b) = \text{Div}(\gcd(a, b))$.

The three parts of this theorem are saying the same thing from slightly different perspectives; the importance of the theorem nevertheless justifies this repetition. To prove the theorem, we first show the following:

Lemma 2.9.15. Let $m, a, b \in \mathbb{Z}$ be such that $m \mid a$ and $m \mid b$. Then, $m \mid \gcd(a, b)$. 


Proof of Lemma 2.9.15. Theorem 2.9.11 shows that there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that
\[
\text{gcd}(a, b) = xa + yb. \tag{22}
\]
Consider these \( x \) and \( y \). Now, \( m \mid a \mid xa \), so that \( xa \equiv 0 \mod m \). Also, \( m \mid b \mid yb \), thus \( yb \equiv 0 \mod m \). Adding the congruences \( xa \equiv 0 \mod m \) and \( yb \equiv 0 \mod m \) together, we find \( xa + yb \equiv 0 + 0 = 0 \mod m \); in other words, \( m \mid xa + yb \). In view of (22), this rewrites as \( m \mid \text{gcd}(a, b) \). This proves Lemma 2.9.15.

Proof of Theorem 2.9.14. (a) Let \( m \in \mathbb{Z} \). In order to prove (21), we need to prove the \( \implies \) and \( \iff \) directions of the equivalence (21). But this is easy: The \( \implies \) direction is just the statement of Lemma 2.9.15, whereas the \( \iff \) direction is trivial (to wit: if \( m \mid \text{gcd}(a, b) \), then \( m \mid \text{gcd}(a, b) \mid a \) (by Proposition 2.9.7(e)) and \( m \mid \text{gcd}(a, b) \mid b \) (by Proposition 2.9.7(e)), and thus \( m \mid a \) and \( m \mid b \)). Hence, the equivalence (21) is proven. This proves Theorem 2.9.14(a).

(b) The common divisors of \( a \) and \( b \) are precisely the integers \( m \) that satisfy \( (m \mid a \text{ and } m \mid b) \) (by the definition of “common divisor”). In view of the equivalence (21), this rewrites as follows: The common divisors of \( a \) and \( b \) are precisely the integers \( m \) that satisfy \( m \mid \text{gcd}(a, b) \). In other words, the common divisors of \( a \) and \( b \) are precisely the divisors of \( \text{gcd}(a, b) \). This proves Theorem 2.9.14(b).

(c) Recall that each \( c \in \mathbb{Z} \) satisfies
\[
\text{Div}(c) = \{ \text{the common divisors of } c \} \quad (\text{by the definition of } \text{Div}(c))
\]
\[
= \{ \text{the integers } x \text{ such that } x \mid c \}
\]
\[
= \{ \text{the integers } x \text{ such that } x \mid c \quad (\text{by the definition of “common divisors”})
\]
\[
= \{ \text{the divisors of } c \}.
\]
Applying this to \( c = \text{gcd}(a, b) \), we obtain
\[
\text{Div}(\text{gcd}(a, b)) = \{ \text{the divisors of } \text{gcd}(a, b) \}. \tag{23}
\]
The definition of \( \text{Div}(a, b) \) yields
\[
\text{Div}(a, b) = \{ \text{the common divisors of } a \text{ and } b \}
\]
\[
= \{ \text{the divisors of } \text{gcd}(a, b) \} \quad (\text{by Theorem 2.9.14(b)})
\]
\[
= \text{Div}(\text{gcd}(a, b)) \quad (\text{by (23)}).
\]
This proves Theorem 2.9.14(c).

The following corollary of Theorem 2.9.11 lets us “combine” two divisibilities \( a \mid c \) and \( b \mid c \). In fact, Proposition 2.2.4(e) would already allow us to “combine” them to form \( ab \mid cc = c^2 \); but we can also “combine” them to \( ab \mid \text{gcd}(a, b) \cdot c \) using the following fact:
Theorem 2.9.16. Let $a, b, c \in \mathbb{Z}$ satisfy $a \mid c$ and $b \mid c$. Then, $ab \mid \gcd(a, b) \cdot c$.

Example 2.9.17. Let $a = 6$ and $b = 10$ and $c = 30$. Then, $a = 6 \mid 30 = c$ and $b = 10 \mid 30 = c$. Thus, Theorem 2.9.11 yields $ab \mid \gcd(a, b) \cdot c$. And indeed, this is true, since $ab = 6 \cdot 10 \mid 2 \cdot 30 = \gcd(a, b) \cdot c$ (because $\gcd(a, b) = \gcd(6, 10) = 2$). Note that this latter divisibility is actually an equality: we have $6 \cdot 10 = 2 \cdot 30$. Note also that we do not obtain $ab \mid c$ (and indeed, this does not hold).

Proof of Theorem 2.9.16. Theorem 2.9.11 yields that there exist integers $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $\gcd(a, b) = xa + yb$. Consider these $x$ and $y$.

We have $a \mid c$. In other words, there exists an integer $u$ such that $c = au$. Consider this $u$.

We have $b \mid c$. In other words, there exists an integer $v$ such that $c = bv$. Consider this $v$.

Now,

$$\gcd(a, b) \cdot c = (xa + yb) \cdot c = xa \cdot c + yb \cdot c = xabv + ybau = ab(xv + yu).$$

Thus, there exists an integer $d$ such that $\gcd(a, b) \cdot c = abd$ (namely, $d = xv + yu$). In other words, $ab \mid \gcd(a, b) \cdot c$. This proves Theorem 2.9.16.

Here is another corollary of Theorem 2.9.11 whose usefulness will become clearer later on:

Theorem 2.9.18. Let $a, b, c \in \mathbb{Z}$ satisfy $a \mid bc$. Then, $a \mid \gcd(a, b) \cdot c$.

At this point, you should see that Theorem 2.9.16 allows “strengthening” divisibilities: You give it a “weak” divisibility $a \mid bc$, and obtain a “stronger” divisibility $a \mid \gcd(a, b) \cdot c$ from it (stronger because $\gcd(a, b)$ is usually smaller than $b$).

Proof of Theorem 2.9.18. Theorem 2.9.11 yields that there exist integers $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$ such that $\gcd(a, b) = xa + yb$. Consider these $x$ and $y$.

We have $a \mid bc \mid ybc$; in other words, $ybc \equiv 0 \mod a$. Also, $a \mid axc$, so that $axc \equiv 0 \mod a$. Adding the two congruences $axc \equiv 0 \mod a$ and $ybc \equiv 0 \mod a$ together, we obtain $axc + ybc \equiv 0 + 0 = 0 \mod a$. In view of $axc + ybc = (xa + yb) \cdot c \equiv \gcd(a, b) \cdot c$, this rewrites as $\gcd(a, b) \cdot c \equiv 0 \mod a$. In other words, $a \mid \gcd(a, b) \cdot c$. This proves Theorem 2.9.18.

Corollary 2.9.19. Let $s, a, b \in \mathbb{Z}$. Then,

$$\gcd(sa, sb) = |s| \gcd(a, b).$$
Proof of Corollary 2.9.19. We shall prove that the two integers \( \gcd(sa, sb) \) and \( s \gcd(a, b) \) mutually divide each other (i.e., they satisfy \( \gcd(sa, sb) \mid s \gcd(a, b) \) and \( s \gcd(a, b) \mid \gcd(sa, sb) \)). Then, Exercise 2.2.2 will let us conclude that \( \gcd(sa, sb) = |s| \gcd(a, b) \). This will then rewrite as \( \gcd(sa, sb) = |s| \gcd(a, b) \), and we will be done. (This trick is actually a common strategy for proving equalities between \( \gcd \)s.)

For the sake of brevity, let us set \( g = \gcd(sa, sb) \) and \( h = s \gcd(a, b) \). So our first goal is to prove that \( g \mid h \) and \( h \mid g \).

Proof of \( g \mid h \): Theorem 2.9.11 yields that there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, b) = xa + yb \). Consider these \( x \) and \( y \).

Proposition 2.9.7 (applied to \( sa \) and \( sb \) instead of \( a \) and \( b \)) yields that \( \gcd(sa, sb) \mid sa \) and \( \gcd(sa, sb) \mid sb \). From \( g = \gcd(sa, sb) \mid sa \), we obtain \( g \mid sa \mid xsa \), thus \( xsa \equiv 0 \mod g \). Similarly, \( ysb \equiv 0 \mod g \). Adding these two congruences together,

\[
h = s \gcd(a, b) = s(xa + yb) = xsa + ysb \equiv 0 \mod g.
\]

In other words, \( g \mid h \). Thus, we have proven \( g \mid h \).

Proof of \( h \mid g \): Proposition 2.9.7 (applied to \( s \gcd(a, b) \), \( s, a \) instead of \( a_1, a_2, b_1, b_2 \)) yields \( s \gcd(a, b) \mid sa \). Similarly, \( s \gcd(a, b) \mid sb \). Hence, Lemma 2.9.15 (applied to \( s \gcd(a, b), sa \) and \( sb \) instead of \( m, a \) and \( b \)) yields \( s \gcd(a, b) \mid \gcd(sa, sb) \). In view of \( g = \gcd(sa, sb) \) and \( h = s \gcd(a, b) \), this rewrites as \( h \mid g \). So we have proven \( h \mid g \).

Now, Exercise 2.2.2 (applied to \( g \) and \( h \) instead of \( a \) and \( b \)) yields \( |g| = |h| \).

But recall that a \( \gcd \) of any finitely many integers is nonnegative (by Definition 2.9.6). Hence, in particular, \( \gcd(a, b) \) and \( \gcd(sa, sb) \) are nonnegative. From \( g = \gcd(sa, sb) \), we obtain

\[
|g| = |\gcd(sa, sb)| = \gcd(sa, sb)
\]

(since \( \gcd(sa, sb) \) is nonnegative). Also, from \( h = s \gcd(a, b) \), we obtain

\[
|h| = |s \gcd(a, b)| = |s| \cdot |\gcd(a, b)| = gcd(a, b) = gcd(a, b) (\text{by } 3)
\]

\[
= |s| \gcd(a, b) \text{ (since } gcd(a, b) \text{ is nonnegative)}
\]

\[
= |s| \gcd(a, b) .
\]

Hence,

\[
\gcd(sa, sb) = |g| = |h| = |s| \gcd(a, b) .
\]

This proves Corollary 2.9.19. \( \square \)
Exercise 2.9.4. Let \( a_1, a_2, b_1, b_2 \in \mathbb{Z} \) satisfy \( a_1 \mid b_1 \) and \( a_2 \mid b_2 \). Prove that
\[
\gcd(a_1, a_2) \mid \gcd(b_1, b_2).
\]

Exercise 2.9.5. Let \( a, b \in \mathbb{Z} \).
(a) Prove that \( \gcd(a, |b|) = \gcd(a, b) \).
(b) Prove that \( \gcd(|a|, b) = \gcd(a, b) \).
(c) Prove that \( \gcd(|a|, |b|) = \gcd(a, b) \).

2.9.6. gcds of multiple numbers

The following theorem generalizes some of the previous facts to gcds of multiple integers:

Theorem 2.9.20. Let \( b_1, b_2, \ldots, b_k \) be integers.

(a) For each \( m \in \mathbb{Z} \), we have the following logical equivalence:
\[
(m \mid b_i \text{ for all } i \in \{1, 2, \ldots, k\}) \iff (m \mid \gcd(b_1, b_2, \ldots, b_k)).
\]

(b) The common divisors of \( b_1, b_2, \ldots, b_k \) are precisely the divisors of \( \gcd(b_1, b_2, \ldots, b_k) \).
(c) We have \( \text{Div}(b_1, b_2, \ldots, b_k) = \text{Div}(\gcd(b_1, b_2, \ldots, b_k)) \).
(d) If \( k > 0 \), then
\[
\gcd(b_1, b_2, \ldots, b_k) = \gcd(\gcd(b_1, b_2, \ldots, b_{k-1}), b_k).
\]

Proof of Theorem 2.9.20

Forget that we fixed \( b_1, b_2, \ldots, b_k \). Rather than prove the four parts of Theorem 2.9.20 separately, we shall prove them together as a package.

We shall proceed by induction on \( k \):

Induction base: Theorem 2.9.20 holds for \( k = 0 \).

[Proof: This is a straightforward exercise in dealing with empty sets, 0-tuples and vacuous truths. For the sake of completeness, here is the full argument:

Assume that \( k = 0 \). We must prove that Theorem 2.9.20 holds.

Let \( b_1, b_2, \ldots, b_k \) be integers. Of course, these are 0 integers, since \( k = 0 \).
We don’t have \( k > 0 \) (since \( k = 0 \)). Hence, Theorem 2.9.20 (d) is vacuously true.
All of \( b_1, b_2, \ldots, b_k \) are 0 (indeed, this is vacuously true). Thus, \( \gcd(b_1, b_2, \ldots, b_k) = 0 \) (by Definition 2.9.6).

For each \( m \in \mathbb{Z} \), we have the logical equivalence
\[
(m \mid b_i \text{ for all } i \in \{1, 2, \ldots, k\})
\iff (m \mid 0) \quad \text{(since there exists no } i \in \{1, 2, \ldots, k\})
\iff (m \mid 0) \quad \text{(since } m \mid 0 \text{ is always true)}
\iff (m \mid \gcd(b_1, b_2, \ldots, b_k)) \quad \text{(since } 0 = \gcd(b_1, b_2, \ldots, b_k)\text{).}
This proves Theorem 2.9.20 (a) (in the case $k = 0$, that is). Parts (b) and (c) of Theorem 2.9.20 are restatements of Theorem 2.9.20 (a) and can be derived from it in the same way as we derived parts (b) and (c) of Theorem 2.9.14 from Theorem 2.9.14 (a).

Thus, all four parts of Theorem 2.9.20 are proven for $k = 0$. This completes the induction base.

**Induction step:** Let $\ell$ be a positive integer. Assume that Theorem 2.9.20 holds for $k = \ell - 1$. We must prove that Theorem 2.9.20 holds for $k = \ell$.

We have assumed that Theorem 2.9.20 holds for $k = \ell - 1$. In other words, the following statement holds:

**Statement 1:** Let $b_1, b_2, \ldots, b_{\ell - 1}$ be integers.

(a) For each $m \in \mathbb{Z}$, we have the following logical equivalence:

$$ (m \mid b_i \text{ for all } i \in \{1, 2, \ldots, \ell - 1\}) \iff (m \mid \gcd (b_1, b_2, \ldots, b_{\ell - 1})). $$

(b) The common divisors of $b_1, b_2, \ldots, b_{\ell - 1}$ are precisely the divisors of $\gcd (b_1, b_2, \ldots, b_{\ell - 1})$.

(c) We have $\text{Div} (b_1, b_2, \ldots, b_{\ell - 1}) = \text{Div} (\gcd (b_1, b_2, \ldots, b_{\ell - 1}))$.

(d) If $\ell - 1 > 0$, then

$$ \gcd (b_1, b_2, \ldots, b_{\ell - 1}) = \gcd (\gcd (b_1, b_2, \ldots, b_{\ell - 2}), b_{\ell - 1}). $$

Recall that we must prove that Theorem 2.9.20 holds for $k = \ell$. In other words, we must prove the following statement:

**Statement 2:** Let $b_1, b_2, \ldots, b_\ell$ be integers.

(a) For each $m \in \mathbb{Z}$, we have the following logical equivalence:

$$ (m \mid b_i \text{ for all } i \in \{1, 2, \ldots, \ell\}) \iff (m \mid \gcd (b_1, b_2, \ldots, b_\ell)). $$

(b) The common divisors of $b_1, b_2, \ldots, b_\ell$ are precisely the divisors of $\gcd (b_1, b_2, \ldots, b_\ell)$.

(c) We have $\text{Div} (b_1, b_2, \ldots, b_\ell) = \text{Div} (\gcd (b_1, b_2, \ldots, b_\ell))$.

(d) If $\ell > 0$, then

$$ \gcd (b_1, b_2, \ldots, b_\ell) = \gcd (\gcd (b_1, b_2, \ldots, b_{\ell - 1}), b_\ell). $$

**Proof of Statement 2:** (d) Let us begin with part (d). Assume that $\ell > 0$ (though we already know that this is true).

Let $g = \gcd (b_1, b_2, \ldots, b_\ell)$ and $h = \gcd (\gcd (b_1, b_2, \ldots, b_{\ell - 1}), b_\ell)$. 


If the integers \( b_1, b_2, \ldots, b_\ell \) are all 0, then Statement 2 (d) holds. Hence, for the rest of this proof, we WLOG assume that the integers \( b_1, b_2, \ldots, b_\ell \) are not all 0. Therefore, \( \gcd (b_1, b_2, \ldots, b_\ell) \) is the largest element of the set \( \text{Div} (b_1, b_2, \ldots, b_\ell) \) (by Definition 2.9.6). In other words, \( g \) is the largest element of the set \( \text{Div} (b_1, b_2, \ldots, b_\ell) \) (since \( g = \gcd (b_1, b_2, \ldots, b_\ell) \)).

Furthermore, the two integers \( \gcd (b_1, b_2, \ldots, b_{\ell-1}) \) and \( b_\ell \) are not all 0. Hence, \( \gcd (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) \) is the largest element of the set \( \text{Div} (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) \) (by Definition 2.9.6). In other words, \( h \) is the largest element of the set \( \text{Div} (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) \) (since \( h = \gcd (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) \)).

We intend to show that \( g = h \). For that, it suffices to prove \( g \leq h \) and \( h \leq g \).

**Proof of \( g \leq h \):** Recall that \( g \) is the largest element of the set \( \text{Div} (b_1, b_2, \ldots, b_\ell) \). Therefore, \( g \in \text{Div} (b_1, b_2, \ldots, b_\ell) \). In other words, \( g \) is a common divisor of \( b_1, b_2, \ldots, b_\ell \). Hence, \( g \mid b_i \) for each \( i \in \{1, 2, \ldots, \ell \} \). Thus, in particular, \( g \mid b_i \) for each \( i \in \{1, 2, \ldots, \ell - 1\} \). But Statement 1 (a) (applied to \( m = g \)) shows that we have the equivalence

\[
(g \mid b_i \text{ for all } i \in \{1, 2, \ldots, \ell - 1\}) \iff (g \mid \gcd (b_1, b_2, \ldots, b_{\ell-1})).
\]

Hence, we have \( g \mid \gcd (b_1, b_2, \ldots, b_{\ell-1}) \) (since we know that \( g \mid b_i \) for all \( i \in \{1, 2, \ldots, \ell - 1\} \)). Combining this with \( g \mid b_\ell \), we conclude that \( g \) is a common divisor of \( \gcd (b_1, b_2, \ldots, b_{\ell-1}) \) and \( b_\ell \). In other words, \( g \in \text{Div} (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) \).

Therefore, \( g \leq h \) (since \( h \) is the largest element of the set \( \text{Div} (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) \)).

**Proof of \( h \leq g \):** We have

\[
h = \gcd (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) \mid \gcd (b_1, b_2, \ldots, b_{\ell-1})
\]

(by Proposition 2.9.7 (f), applied to \( a = \gcd (b_1, b_2, \ldots, b_{\ell-1}) \) and \( b = b_\ell \)). Also,

\[
h = \gcd (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) \mid b_\ell
\]

(by Proposition 2.9.7 (f), applied to \( a = \gcd (b_1, b_2, \ldots, b_{\ell-1}) \) and \( b = b_\ell \)).

But Statement 1 (a) (applied to \( m = h \)) shows that we have the equivalence

\[
(h \mid b_i \text{ for all } i \in \{1, 2, \ldots, \ell - 1\}) \iff (h \mid \gcd (b_1, b_2, \ldots, b_{\ell-1})).
\]

**Proof.** Assume that \( b_1, b_2, \ldots, b_\ell \) are all 0. Then, \( \gcd (b_1, b_2, \ldots, b_\ell) = 0 \) (by Definition 2.9.6). Moreover, \( b_1, b_2, \ldots, b_{\ell-1} \) are all 0 (since \( b_1, b_2, \ldots, b_\ell \) are all 0), and thus \( \gcd (b_1, b_2, \ldots, b_{\ell-1}) = 0 \). Finally, \( b_\ell = 0 \) (since \( b_1, b_2, \ldots, b_\ell \) are all 0). Comparing \( \gcd (b_1, b_2, \ldots, b_\ell) = 0 \)

with \( \gcd (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) = \gcd (0, 0) = 0 \), we obtain \( \gcd (b_1, b_2, \ldots, b_\ell) = \gcd (\gcd (b_1, b_2, \ldots, b_{\ell-1}), b_\ell) \). In other words, Statement 2 (d) holds.

**Proof.** Assume the contrary. Thus, both \( \gcd (b_1, b_2, \ldots, b_{\ell-1}) \) and \( b_\ell \) are 0. Thus, in particular, \( b_\ell = 0 \). If the \( \ell - 1 \) integers \( b_1, b_2, \ldots, b_{\ell-1} \) were all 0, then the \( \ell \) integers \( b_1, b_2, \ldots, b_\ell \) would be all 0 (since \( b_\ell = 0 \)), which would contradict the fact that the integers \( b_1, b_2, \ldots, b_\ell \) are not all 0. Hence, the \( \ell - 1 \) integers \( b_1, b_2, \ldots, b_{\ell-1} \) are not all 0. Thus, \( \gcd (b_1, b_2, \ldots, b_{\ell-1}) \) is a positive integer (by Definition 2.9.6). Thus, \( \gcd (b_1, b_2, \ldots, b_{\ell-1}) > 0 \), which contradicts the fact that \( \gcd (b_1, b_2, \ldots, b_{\ell-1}) \) is 0. This contradiction shows that our assumption was false, qed.
Thus, we have \((h \mid b_i \text{ for all } i \in \{1, 2, \ldots, \ell - 1\})\) (since we have \(h \mid \gcd(b_1, b_2, \ldots, b_{\ell - 1})\)).

This divisibility \(h \mid b_i\) holds not only for all \(i \in \{1, 2, \ldots, \ell - 1\}\), but also for \(i = \ell\) (because \(h \mid b_\ell\)). Thus, we conclude that \(h \mid b_i\) for all \(i \in \{1, 2, \ldots, \ell\}\). In other words, \(h\) is a common divisor of \(b_1, b_2, \ldots, b_\ell\). In other words, \(h \in \Div(b_1, b_2, \ldots, b_\ell)\). Thus, \(h \leq g\) (since \(g\) is the largest element of the set \(\Div(b_1, b_2, \ldots, b_\ell)\)).

Combining \(h \leq g\) with \(g \leq h\), we obtain \(g = h\). In other words,

\[
gcd(b_1, b_2, \ldots, b_\ell) = \gcd(\gcd(b_1, b_2, \ldots, b_{\ell - 1}), b_\ell)
\]

(since \(g = \gcd(b_1, b_2, \ldots, b_\ell)\) and \(h = \gcd(\gcd(b_1, b_2, \ldots, b_{\ell - 1}), b_\ell)\)). Hence, Statement 2 (d) is proven.

(a) Let \(m \in \mathbb{Z}\). Then, we have the equivalence

\[
(m \mid b_i \text{ for all } i \in \{1, 2, \ldots, \ell\})
\]

\[
\iff (m \mid b_i \text{ for all } i \in \{1, 2, \ldots, \ell - 1\}) \text{ and } m \mid b_\ell
\]

\[
\iff (m \mid \gcd(b_1, b_2, \ldots, b_{\ell - 1}) \text{ and } m \mid b_\ell)
\]

\[
\iff (m \mid \gcd(b_1, b_2, \ldots, b_{\ell - 1}) \text{ and } m \mid b_\ell)
\]

\[
\iff (m \mid \gcd(b_1, b_2, \ldots, b_{\ell - 1}, b_\ell)
\]

(by \Theorem 2.9.14 (a), applied to \(a = \gcd(b_1, b_2, \ldots, b_{\ell - 1})\) and \(b = b_\ell\))

Thus, Statement 2 (a) follows.

Statement 2 (b) is a restatement of Statement 2 (a) (in the same way that \Theorem 2.9.14 (b) is a restatement of \Theorem 2.9.14 (a)).

Statement 2 (c) is a restatement of Statement 2 (b) (in the same way that \Theorem 2.9.14 (c) is a restatement of \Theorem 2.9.14 (b)).

We are thus done proving Statement 2.

In other words, we have proven that Theorem \textbf{2.9.20} holds for \(k = \ell\). This completes the induction step. Thus, Theorem \textbf{2.9.20} is proven by induction. \(\square\)

Theorem \textbf{2.9.20} (d) is the reason why most properties of gcds of multiple numbers can be derived from corresponding properties of gcds of two numbers. For example, we can easily prove the following analogue of Corollary \textbf{2.9.19} for gcds of three numbers:
Exercise 2.9.6. Let $s, a, b, c \in \mathbb{Z}$. Prove that $\text{gcd} (sa, sb, sc) = |s| \text{gcd} (a, b, c)$.

More generally, Corollary 2.9.19 can be generalized to any number of integers:

Exercise 2.9.7. Let $s \in \mathbb{Z}$, and let $a_1, a_2, \ldots, a_k$ be integers. Prove that $\text{gcd} (sa_1, sa_2, \ldots, sa_k) = |s| \text{gcd} (a_1, a_2, \ldots, a_k)$.

Bezout’s theorem (Theorem 2.9.11) also holds for any finite number of integers:

Theorem 2.9.21. Let $b_1, b_2, \ldots, b_k$ be integers. Then, there exist integers $x_1, x_2, \ldots, x_k$ such that

$$\text{gcd} (b_1, b_2, \ldots, b_k) = x_1 b_1 + x_2 b_2 + \cdots + x_k b_k.$$ 

Once again, we can restate Theorem 2.9.21 by using the concept of a $\mathbb{Z}$-linear combination. Let us define this concept finally:

Definition 2.9.22. Let $b_1, b_2, \ldots, b_k$ be numbers. A $\mathbb{Z}$-linear combination of $b_1, b_2, \ldots, b_k$ shall mean a number of the form $x_1 b_1 + x_2 b_2 + \cdots + x_k b_k$, where $x_1, x_2, \ldots, x_k$ are integers.

Thus, Theorem 2.9.21 can be restated as follows:

Theorem 2.9.23. Let $b_1, b_2, \ldots, b_k$ be integers. Then, $\text{gcd} (b_1, b_2, \ldots, b_k)$ is a $\mathbb{Z}$-linear combination of $b_1, b_2, \ldots, b_k$.

Proof of Theorem 2.9.23. We shall prove this by induction on $k$:

Induction base: Recall that the empty list $()$ satisfies $\text{gcd} () = 0$ (by Definition 2.9.6, since all entries of the empty list are 0). But 0 is a $\mathbb{Z}$-linear combination of an empty list of numbers, because $0 = ()$ (empty sum). Combining these facts, we conclude that $\text{gcd} ()$ is a $\mathbb{Z}$-linear combination of an empty list of numbers. But this is precisely the claim of Theorem 2.9.23 for $k = 0$. Thus, Theorem 2.9.23 holds for $k = 0$. This completes the induction base.

Induction step: Let $\ell$ be a positive integer. Assume that Theorem 2.9.23 holds for $k = \ell - 1$. We must prove that Theorem 2.9.23 holds for $k = \ell$.

We have assumed that Theorem 2.9.23 holds for $k = \ell - 1$. In other words, the following statement holds:

Statement 1: Let $b_1, b_2, \ldots, b_{\ell - 1}$ be integers. Then, $\text{gcd} (b_1, b_2, \ldots, b_{\ell - 1})$ is a $\mathbb{Z}$-linear combination of $b_1, b_2, \ldots, b_{\ell - 1}$.

Our goal is to prove that Theorem 2.9.23 holds for $k = \ell$. In other words, we must prove the following statement:

Statement 2: Let $b_1, b_2, \ldots, b_{\ell}$ be integers. Then, $\text{gcd} (b_1, b_2, \ldots, b_{\ell})$ is a $\mathbb{Z}$-linear combination of $b_1, b_2, \ldots, b_{\ell}$.
Proof of Statement 2: Statement 1 shows that \( \gcd(b_1, b_2, \ldots, b_{\ell-1}) \) is a \( \mathbb{Z} \)-linear combination of \( b_1, b_2, \ldots, b_{\ell-1} \). In other words, there exist \( \ell - 1 \) integers \( y_1, y_2, \ldots, y_{\ell-1} \) such that

\[
\gcd(b_1, b_2, \ldots, b_{\ell-1}) = y_1b_1 + y_2b_2 + \cdots + y_{\ell-1}b_{\ell-1}.
\]

Consider these \( y_1, y_2, \ldots, y_{\ell-1} \).

Furthermore, Theorem 2.9.11 (applied to \( a = \gcd(b_1, b_2, \ldots, b_{\ell-1}) \) and \( b = b_{\ell} \)) yields that there exist two integers \( x \) and \( y \) such that

\[
\gcd(\gcd(b_1, b_2, \ldots, b_{\ell-1}), b_{\ell}) = x \gcd(b_1, b_2, \ldots, b_{\ell-1}) + yb_{\ell}.
\]

Consider these \( x \) and \( y \).

Now, \( \ell > 0 \); thus, Theorem 2.9.20 (d) (applied to \( k = \ell \)) yields

\[
\gcd(b_1, b_2, \ldots, b_{\ell}) = \gcd(\gcd(b_1, b_2, \ldots, b_{\ell-1}), b_{\ell})
\]

\[
= x \gcd(b_1, b_2, \ldots, b_{\ell-1}) + yb_{\ell}
\]

\[
= y_1b_1 + y_2b_2 + \cdots + y_{\ell-1}b_{\ell-1} + yb_{\ell}
\]

\[
= xy_1b_1 + xy_2b_2 + \cdots + xy_{\ell-1}b_{\ell-1} + yb_{\ell}.
\]

This is clearly a \( \mathbb{Z} \)-linear combination of \( b_1, b_2, \ldots, b_{\ell} \). Thus, \( \gcd(b_1, b_2, \ldots, b_{\ell}) \) is a \( \mathbb{Z} \)-linear combination of \( b_1, b_2, \ldots, b_{\ell} \). So Statement 2 is proven.

In other words, we have proven that Theorem 2.9.23 holds for \( k = \ell \). This completes the induction step. Thus, Theorem 2.9.23 is proven by induction.

Proof of Theorem 2.9.21 We have just proven Theorem 2.9.23, which is a restatement of Theorem 2.9.21. Thus, Theorem 2.9.21 is also proven.

For future reference, let us restate Theorem 2.9.20 (a) as follows:

**Corollary 2.9.24.** Let \( b_1, b_2, \ldots, b_k \) be integers. For each \( m \in \mathbb{Z} \), we have the following logical equivalence:

\[(m \mid b_1 \text{ and } m \mid b_2 \text{ and } \cdots \text{ and } m \mid b_k) \iff (m \mid \gcd(b_1, b_2, \ldots, b_k)).\]

Proof of Corollary 2.9.24 Let \( m \in \mathbb{Z} \). Then, we have the following chain of equivalences:

\[(m \mid b_1 \text{ and } m \mid b_2 \text{ and } \cdots \text{ and } m \mid b_k) \iff (m \mid b_i \text{ for all } i \in \{1, 2, \ldots, k\}) \iff (m \mid \gcd(b_1, b_2, \ldots, b_k)) \quad (\text{by Theorem 2.9.20 (a)}).

This proves Corollary 2.9.24.
**Theorem 2.9.25.** Let \(b_1, b_2, \ldots, b_k\) be integers, and let \(c_1, c_2, \ldots, c_\ell\) be integers. Then,
\[
\gcd (b_1, b_2, \ldots, b_k, c_1, c_2, \ldots, c_\ell) = \gcd (\gcd (b_1, b_2, \ldots, b_k), \gcd (c_1, c_2, \ldots, c_\ell)).
\]

Our proof of this theorem will rely on a simple trick, which we state as a lemma:

**Lemma 2.9.26.** Let \(a\) and \(b\) be two integers.

(a) If each \(m \in \mathbb{Z}\) satisfies the implication \((m \mid a) \implies (m \mid b)\), then \(a \mid b\).

(b) If each \(m \in \mathbb{Z}\) satisfies the equivalence \((m \mid a) \iff (m \mid b)\), then \(|a| = |b|\).

Lemma 2.9.26 (b) says that the divisors of an integer \(a\) uniquely determine \(|a|\) (that is, they uniquely determine \(a\) up to sign). Thus, when you want to prove that two integers have the same absolute values, it suffices to prove that they have the same divisors. If you know that your two integers are nonnegative, then you can prove this way that they are equal (since their absolute values are just themselves). This is exactly how we will prove that the left and right hand sides in Theorem 2.9.25 are equal.

**Proof of Lemma 2.9.26.**

(a) Assume that each \(m \in \mathbb{Z}\) satisfies the implication \((m \mid a) \implies (m \mid b)\). Then, applying this to \(m = a\), we obtain the implication \((a \mid a) \implies (a \mid b)\). Since \(a \mid a\) holds, we thus obtain \(a \mid b\). This proves Lemma 2.9.26 (a).

(b) Assume that each \(m \in \mathbb{Z}\) satisfies the equivalence \((m \mid a) \iff (m \mid b)\). Thus, each \(m \in \mathbb{Z}\) satisfies the implication \((m \mid a) \implies (m \mid b)\) (since this implication is part of the equivalence we just assumed). Thus, Lemma 2.9.26 (a) yields \(a \mid b\).

Recall again that each \(m \in \mathbb{Z}\) satisfies the equivalence \((m \mid a) \iff (m \mid b)\). Thus, each \(m \in \mathbb{Z}\) satisfies the implication \((m \mid b) \implies (m \mid a)\) (since this implication is also part of the equivalence). Hence, Lemma 2.9.26 (a) (applied to \(b\) and \(a\) instead of \(a\) and \(b\)) yields \(b \mid a\).

Hence, Exercise 2.2.2 yields \(|a| = |b|\). This proves Lemma 2.9.26 (b).

Lemma 2.9.26 is a simple case of what is known in category theory as the Yoneda lemma.

**Proof of Theorem 2.9.25.** Let \(m \in \mathbb{Z}\). Corollary 2.9.24 (applied to \(k + \ell\) and \((b_1, b_2, \ldots, b_k, c_1, c_2, \ldots, c_\ell)\) instead of \(k\) and \((b_1, b_2, \ldots, b_k)\)) shows that we have the following equivalence:
\[
(m \mid b_1 \text{ and } m \mid b_2 \text{ and } \cdots \text{ and } m \mid b_k \text{ and } m \mid c_1 \text{ and } m \mid c_2 \text{ and } \cdots \text{ and } m \mid c_\ell) \iff (m \mid \gcd (b_1, b_2, \ldots, b_k, c_1, c_2, \ldots, c_\ell)).
\]
This proves Theorem 2.9.25.

We thus have the following chain of equivalences:

\[
(m \mid \gcd(b_1, b_2, \ldots, b_k, c_1, c_2, \ldots, c_{\ell})) \\
\iff (m \mid b_1 \text{ and } m \mid b_2 \text{ and } \ldots \text{ and } m \mid b_k \text{ and } m \mid c_1 \text{ and } m \mid c_2 \text{ and } \ldots \text{ and } m \mid c_{\ell})
\]

\[
\iff \left( (m \mid b_i \text{ for all } i \in \{1, 2, \ldots, k\}) \text{ and } (m \mid c_i \text{ for all } i \in \{1, 2, \ldots, \ell\}) \right)
\]

(by Theorem 2.9.20(a))

\[
\iff (m \mid \gcd(b_1, b_2, \ldots, b_k)) \text{ and } (m \mid \gcd(c_1, c_2, \ldots, c_{\ell}))
\]

(by Theorem 2.9.20(a), applied to \(c\) and \(b\) instead of \(k\) and \((b_1, b_2, \ldots, b_k)\))

Now, forget that we fixed \(m\). We thus have shown that each \(m \in \mathbb{Z}\) satisfies the equivalence

\[
(m \mid \gcd(b_1, b_2, \ldots, b_k, c_1, c_2, \ldots, c_{\ell})) \\
\iff (m \mid \gcd(\gcd(b_1, b_2, \ldots, b_k), \gcd(c_1, c_2, \ldots, c_{\ell}))).
\]

Hence, Lemma 2.9.26(b) (applied to \(a = \gcd(b_1, b_2, \ldots, b_k, c_1, c_2, \ldots, c_{\ell})\) and \(b = \gcd(\gcd(b_1, b_2, \ldots, b_k), \gcd(c_1, c_2, \ldots, c_{\ell}))\)) yields

\[
|\gcd(b_1, b_2, \ldots, b_k, c_1, c_2, \ldots, c_{\ell})| \\
= |\gcd(\gcd(b_1, b_2, \ldots, b_k), \gcd(c_1, c_2, \ldots, c_{\ell}))|.
\]

But a gcd of integers is always nonnegative (by Definition 2.9.6); thus, the absolute value of a gcd is always this gcd itself. Therefore, we can remove the absolute value signs on both sides of (24). We thus obtain

\[
\gcd(b_1, b_2, b_k, c_1, c_2, \ldots, c_{\ell}) = \gcd(\gcd(b_1, b_2, \ldots, b_k), \gcd(c_1, c_2, \ldots, c_{\ell})).
\]

This proves Theorem 2.9.25.

2.9.7. On converses of Bezout’s theorem

Some words of warning are in order. Theorem 2.9.11 says that if \(a\) and \(b\) are two integers, then \(\gcd(a, b)\) is a \(\mathbb{Z}\)-linear combination of \(a\) and \(b\). Note the indefinite article “a” here: There are (usually) many \(\mathbb{Z}\)-linear combinations of \(a\) and \(b\), but only one gcd. It is definitely not true that every \(\mathbb{Z}\)-linear combination of \(a\) and \(b\) must be \(\gcd(a, b)\). However, all these \(\mathbb{Z}\)-linear combinations are multiples of the gcd, as the following (simple) proposition says:
Proposition 2.9.27. Let $a$ and $b$ be two integers. Then, any integers $x$ and $y$ satisfy $\gcd(a, b) \mid xa + yb$.

Proof of Proposition 2.9.27 Let $x$ and $y$ be integers. Let $g = \gcd(a, b)$. Thus, $g = \gcd(a, b) \mid a$ (by Proposition 2.9.7 (f)). Hence, $g \mid a \mid xa$ (since $xa = ax$). In other words, $xa \equiv 0 \pmod{g}$. Similarly, $yb \equiv 0 \pmod{g}$. Adding these two congruences together, we obtain $xa + yb \equiv 0 + 0 = 0 \pmod{g}$. In other words, $g \mid xa + yb$. In other words, $\gcd(a, b) \mid xa + yb$ (since $g = \gcd(a, b)$). This proves Proposition 2.9.27. \qed

A similar proposition holds for $\mathbb{Z}$-linear combinations of any number of integers $b_1, b_2, \ldots, b_k$.

2019-02-06 lecture

2.10. Coprime integers

2.10.1. Definition

The concept of a gcd leads to one of the most important notions of number theory:

Definition 2.10.1. Let $a$ and $b$ be two integers. We say that $a$ is coprime to $b$ if and only if $\gcd(a, b) = 1$.

Instead of “coprime”, some authors say “relatively prime” or even “prime” (but the latter language risks confusion with a more standard notion of “prime” that we will see later on.)

Example 2.10.2. (a) The number 2 is coprime to 3, since $\gcd(2, 3) = 1$.

(b) The number 6 is not coprime to 15, since $\gcd(6, 15) = 3 \neq 1$.

(c) Let $a$ be an integer. We claim (as a generalization of part (a)) that the number $a$ is coprime to $a+1$. To prove this, we note that

$$\gcd\left(a, a+1\right) = \gcd(a, 1a+1) = \gcd(a, 1)$$

(by Proposition 2.9.7 (c), applied to $u = 1$ and $b = 1$)

$$\mid 1$$

(by Proposition 2.9.7 (e), applied to $b = 1$),

and thus $\gcd(a, a+1) = 1$ (by Exercise 2.2.5 since $\gcd(a, a+1)$ is a nonnegative integer), which means that $a$ is coprime to $a+1$.

(d) Let $a$ be an integer. When is $a$ coprime to $a+2$? If we try to compute $\gcd(a, a+2)$, we find

$$\gcd\left(a, a+2\right) = \gcd(a, 1a+2) = \gcd(a, 2)$$

(by Proposition 2.9.7 (c), applied to $u = 1$ and $b = 2$).
It remains to find $\gcd(a, 2)$. Proposition 2.9.7 (e) (applied to $b = 2$) yields $\gcd(a, 2) \mid a$ and $\gcd(a, 2) \mid 2$. Since $\gcd(a, 2)$ is a nonnegative integer and is a divisor of 2 (because $\gcd(a, 2) \mid 2$), we see that $\gcd(a, 2)$ must be either 1 or 2 (since the only nonnegative divisors of 2 are 1 and 2). If $a$ is even, then 2 is a common divisor of $a$ and 2, and thus must be the greatest common divisor of $a$ and 2 (because a common divisor of $a$ and 2 cannot be greater than 2); in other words, we have $\gcd(a, 2) = 2$ in this case. On the other hand, if $a$ is odd, then 2 is not a common divisor of $a$ and 2 (since 2 does not divide $a$), and thus cannot be the greatest common divisor of $a$ and 2; hence, in this case, we have $\gcd(a, 2) \neq 2$ and thus $\gcd(a, 2) = 1$. Summarizing, we conclude that

$$\gcd(a, 2) = \begin{cases} 
2, & \text{if } a \text{ is even;} \\
1, & \text{if } a \text{ is odd.}
\end{cases}$$

Now, recall that $\gcd(a, a+2) = \gcd(a, 2) = \begin{cases} 
2, & \text{if } a \text{ is even;} \\
1, & \text{if } a \text{ is odd.}
\end{cases}$ Hence, $a$ is coprime to $a+2$ if and only if $a$ is odd.

Following the book [GrKnPa94], we introduce a slightly quaint notation:

**Definition 2.10.3.** Let $a$ and $b$ be two integers. We write “$a \perp b$” to signify that $a$ is coprime to $b$.

Note that the “$\perp$” relation is symmetric:

**Proposition 2.10.4.** Let $a$ and $b$ be two integers. Then, $a \perp b$ if and only if $b \perp a$.

**Proof of Proposition 2.10.4** We have the following chain of equivalences:

$$(a \perp b) \iff (a \text{ is coprime to } b) \quad \text{(by the definition of “$\perp$”)}
\iff (\gcd(a, b) = 1) \quad \text{(by the definition of “coprime”)}
\iff (\gcd(b, a) = 1) \quad \text{ (since Proposition 2.9.7 (b) yields $\gcd(a, b) = \gcd(b, a)$)}
\iff (b \text{ is coprime to } a) \quad \text{ (by the definition of “coprime”)}
\iff (b \perp a) \quad \text{ (by the definition of “$\perp$”).}

This proves Proposition 2.10.4.

**Definition 2.10.5.** Let $a$ and $b$ be two integers. Proposition 2.10.4 shows that $a$ is coprime to $b$ if and only if $b$ is coprime to $a$. Hence, we shall sometimes use a more symmetric terminology for this situation: We shall say that “$a$ and $b$ are coprime” to mean that $a$ is coprime to $b$ (or, equivalently, that $b$ is coprime to $a$).
Exercise 2.10.1. Let $a \in \mathbb{Z}$. Prove the following:

(a) We have $1 \perp a$.
(b) We have $0 \perp a$ if and only if $|a| = 1$.

2.10.2. Properties of coprime integers

We can now state multiple theorems about coprime numbers. The first one states that we can “cancel” a factor $b$ from a divisibility $a | bc$ as long as this factor is coprime to $a$:

Theorem 2.10.6. Let $a, b, c \in \mathbb{Z}$ satisfy $a | bc$ and $a \perp b$. Then, $a | c$.

Proof of Theorem 2.10.6. We have $a \perp b$; in other words, $a$ is coprime to $b$ (by Definition 2.10.3). In other words, $\gcd(a, b) = 1$ (by the definition of “coprime”). Now, Theorem 2.9.18 yields $a | \gcd(a, b) \cdot c = c$. This proves Theorem 2.10.6. □

I like to think of Theorem 2.10.7 as a way of removing “unsolicited guests” from divisibilities. Indeed, it says that we can remove the factor $b$ from $a | bc$ if we know that $b$ is “unrelated” (i.e., coprime) to $a$.

The next theorem lets us “combine” two divisibilities $a | c$ and $b | c$ to $ab | c$ as long as $a$ and $b$ are coprime:

Theorem 2.10.7. Let $a, b, c \in \mathbb{Z}$ satisfy $a | c$ and $b | c$ and $a \perp b$. Then, $ab | c$.

Proof of Theorem 2.10.7. We have $a \perp b$; in other words, $a$ is coprime to $b$ (by Definition 2.10.3). In other words, $\gcd(a, b) = 1$ (by the definition of “coprime”). Now, Theorem 2.9.16 yields $ab | \gcd(a, b) \cdot c = c$. This proves Theorem 2.10.7. □

Theorem 2.10.7 can be restated as follows: If $a$ and $b$ are two coprime divisors of an integer $c$, then $ab$ is also a divisor of $c$. This is often helpful when proving divisibilities where the left hand side (i.e., the number in front of the “$|$” sign) can be split into a product of two mutually coprime factors. Similar reasoning works with several coprime factors (see Exercise 2.10.3 below).

The next theorem (still part of the fallout of Bezout’s theorem) is important, but we will not truly appreciate it until later:

Theorem 2.10.8. Let $a, n \in \mathbb{Z}$.

(a) There exists a $b \in \mathbb{Z}$ such that $ab \equiv \gcd(a, n) \mod n$.
(b) If $a \perp n$, then there exists an $a' \in \mathbb{Z}$ such that $aa' \equiv 1 \mod n$.
(c) If there exists an $a' \in \mathbb{Z}$ such that $aa' \equiv 1 \mod n$, then $a \perp n$.

If $a, n \in \mathbb{Z}$, then an integer $a' \in \mathbb{Z}$ satisfying $aa' \equiv 1 \mod n$ is called a modular inverse of $a$ modulo $n$. The word “modular inverse” is chosen in analogy to the
Theorem 2.10.8

Proof of Theorem 2.10.8. (a) Theorem 2.9.11 (applied to \( b = n \)) yields that there exist integers \( x \in \mathbb{Z} \) and \( y \in \mathbb{Z} \) such that \( \gcd(a, n) = xa + yn \). Consider these \( x \) and \( y \). We have \( ax = xa \equiv xa + yn \mod n \) (since \( xa - (xa + yn) = -yn = n (-y) \) is clearly divisible by \( n \)). Thus, \( ax \equiv xa + yn = \gcd(a, n) \mod n \). Therefore, there exists a \( a' \in \mathbb{Z} \) such that \( aa' \equiv (a, n) \mod n \) (namely, \( a' = x \)). This proves Theorem 2.10.8 (a).

(b) Assume that \( a \perp n \). In other words, \( a \) is coprime to \( n \) (by Definition 2.10.3). In other words, \( \gcd(a, n) = 1 \) (by the definition of “coprime”). Now, Theorem 2.10.8 (a) yields that there exists a \( a' \in \mathbb{Z} \) such that \( aa' \equiv \gcd(a, n) \mod n \). In view of \( \gcd(a, n) = 1 \), this rewrites as follows: There exists an \( a' \in \mathbb{Z} \) such that \( aa' \equiv 1 \mod n \). This proves Theorem 2.10.8 (b).

(c) Assume that there exists an \( a' \in \mathbb{Z} \) such that \( aa' \equiv 1 \mod n \). Consider this \( a' \).

Proposition 2.9.7 (f) yields \( \gcd(a, n) \mid a \) and \( \gcd(a, n) \mid n \). Set \( g = \gcd(a, n) \).

Then, \( g \) is a nonnegative integer.

Now, \( g = \gcd(a, n) \mid a \mid aa' \), so that \( aa' \equiv 0 \mod g \). But also \( g = \gcd(a, n) \mid n \). Hence, from \( aa' \equiv 1 \mod n \), we obtain \( aa' \equiv 1 \mod g \) (by Proposition 2.3.4 (e), applied to \( g, aa' \) and \( 1 \) instead of \( m, a \) and \( b \)). Hence, \( 1 \equiv aa' \equiv 0 \mod g \). Equivalently, \( g \mid 1 - 0 = 1 \). Hence, \( g = 1 \) (by Exercise 2.2.5 since \( g \) is a nonnegative integer). Thus, \( \gcd(a, n) = g = 1 \). In other words, \( a \) is coprime to \( n \). In other words, \( a \perp n \). This proves Theorem 2.10.8 (c).

Theorem 2.10.9. Let \( a, b, c \in \mathbb{Z} \) such that \( a \perp c \) and \( b \perp c \). Then, \( ab \perp c \).

Proof of Theorem 2.10.9. Theorem 2.10.8 (b) (applied to \( n = c \)) yields that there exists an \( a' \in \mathbb{Z} \) such that \( aa' \equiv 1 \mod c \). Consider this \( a' \).

Theorem 2.10.8 (b) (applied to \( b \) and \( c \) instead of \( a \) and \( n \)) yields that there exists a \( b' \in \mathbb{Z} \) such that \( bb' \equiv 1 \mod c \). Consider this \( b' \).

Multiplying the two congruences \( aa' \equiv 1 \mod c \) and \( bb' \equiv 1 \mod c \), we obtain \((aa')(bb') \equiv 1 \cdot 1 = 1 \mod c \).

Now, define the integers \( r = ab \) and \( s = ab' \). Then, \( r = ab = ab' = (ab)(a'b') \equiv (aa')(bb') \equiv 1 \mod c \). Hence, there exists an \( r' \in \mathbb{Z} \) such that \( rr' \equiv 1 \mod c \) (namely, \( r' = s \)). Thus, Theorem 2.10.8 (c) (applied to \( r \) and \( c \) instead of \( a \) and \( n \)) yields that \( r \perp c \). In view of \( r = ab \), this rewrites as \( ab \perp c \). This proves Theorem 2.10.9.

Exercise 2.10.2. Let \( c \in \mathbb{Z} \). Let \( a_1, a_2, \ldots, a_k \) be integers such that each \( i \in \{1, 2, \ldots, k\} \) satisfies \( a_i \perp c \). Prove that \( a_1a_2 \cdots a_k \perp c \).
We can similarly generalize Theorem 2.10.7 to show that the product of several mutually coprime divisors of an integer $c$ must again be a divisor of $c$:

**Exercise 2.10.3.** Let $c \in \mathbb{Z}$. Let $b_1, b_2, \ldots, b_k$ be integers that are mutually coprime (i.e., they satisfy $b_i \perp b_j$ for all $i \neq j$). Assume that $b_i \mid c$ for each $i \in \{1, 2, \ldots, k\}$. Prove that $b_1 b_2 \cdots b_k \mid c$.

**Exercise 2.10.4.** Let $a, b \in \mathbb{Z}$ be such that $a \perp b$. Let $n, m \in \mathbb{N}$. Prove that $a^n \perp b^m$.

The above results have one important application to congruences. Recall that if $a, b, c$ are integers satisfying $ab = ac$, then we can “cancel” $a$ from the equality $ab = ac$ to obtain $b = c$ as long as $a$ is nonzero. Something similar is true for congruences modulo $n$, but the condition “$a$ is nonzero” has to be replaced by “$a$ is coprime to $n$”:

**Lemma 2.10.10.** Let $a, b, c, n$ be integers such that $a \perp n$ and $ab \equiv ac \mod n$. Then, $b \equiv c \mod n$.

Lemma 2.10.10 says that we can cancel an integer $a$ from a congruence $ab \equiv ac \mod n$ as long as $a$ is coprime to $n$. Let us give two proofs of this lemma, to illustrate the uses of some of the previous results:

**First proof of Lemma 2.10.10.** We have $ab \equiv ac \mod n$. In other words, $n \mid ab - ac = a(b - c)$. But Proposition 2.10.4 (applied to $n$ instead of $b$) shows that $a \perp n$ if and only if $n \perp a$. Thus, we have $n \perp a$ (since $a \perp n$).

Thus, we know that $n \mid a(b - c)$ and $n \perp a$. Hence, Theorem 2.10.6 (applied to $n$, $a$ and $b - c$ instead of $a$, $b$ and $c$) yields $n \mid b - c$. In other words, $b \equiv c \mod n$. This proves Lemma 2.10.10.

**Second proof of Lemma 2.10.10.** Theorem 2.10.8(b) yields that there exists an $a' \in \mathbb{Z}$ such that $aa' \equiv 1 \mod n$ (since $a \perp n$). Consider this $a'$. Now, let us multiply the (trivial) congruence $a' \equiv a' \mod n$ with the congruence $ab \equiv ac \mod n$. We thus find

$$a'ab \equiv a'a \equiv 1 \mod n$$

Hence,

$$c \equiv a' \equiv a \equiv 1b \equiv b \equiv b \mod n.$$ 

In other words, $b \equiv c \mod n$. This proves Lemma 2.10.10.

For future use, let us restate Exercise 2.10.2 in a form that uses “unordered” finite products $\prod_{i \in I} b_i$ instead of $a_1 a_2 \cdots a_k$: 

---
Exercise 2.10.5. Let \( c \in \mathbb{Z} \). Let \( I \) be a finite set. For each \( i \in I \), let \( b_i \) be an integer such that \( b_i \perp c \). Prove that \( \prod_{i \in I} b_i \perp c \).

Exercise 2.10.6. Let \( a, b, c \) be three integers such that \( a \equiv b \mod c \). Prove that if \( a \perp c \), then \( b \perp c \).

Exercise 2.10.7. Let \( a, b \in \mathbb{Z} \). Prove that \( b - a \perp b \) holds if and only if \( a \perp b \).

2.10.3. An application to sums of powers

Let us show an application of Theorem 2.10.7. First, we shall prove a simple lemma:

Lemma 2.10.11. Let \( d \in \mathbb{N} \). Let \( x \) and \( y \) be integers.

(a) We have \( x - y \mid x^d - y^d \).

(b) We have \( x + y \mid x^d + y^d \) if \( d \) is odd.

Proof of Lemma 2.10.11. (a) Here are two ways of proving this:

First proof of Lemma 2.10.11 (a): We have \( x \equiv y \mod x - y \) (since \( x - y \mid x - y \)). Thus, Exercise 2.3.4 (applied to \( n = x - y \), \( a = x \), \( b = y \) and \( k = d \)) yields \( x^d \equiv y^d \mod x - y \). In other words, \( x - y \mid x^d - y^d \). This proves Lemma 2.10.11 (a).

Second proof of Lemma 2.10.11 (a): Recall that

\[
(a - b) \left( a^{k-1} + a^{k-2} b + a^{k-3} b^2 + \cdots + a b^{k-2} + b^{k-1} \right) = a^k - b^k \tag{25}
\]

for every \( a, b \in \mathbb{Q} \) and \( k \in \mathbb{N} \). (This is a well-known identity, and it appears (with \( k \) renamed as \( n \)) as the first half of Exercise 1 on homework set #0.) Applying this identity to \( a = x, b = y \) and \( k = d \), we obtain

\[
(x - y) \left( x^{d-1} + x^{d-2} y + x^{d-3} y^2 + \cdots + x y^{d-2} + y^{d-1} \right) = x^d - y^d.
\]

Thus, \( x - y \mid x^d - y^d \) (since \( x^{d-1} + x^{d-2} y + x^{d-3} y^2 + \cdots + x y^{d-2} + y^{d-1} \) is an integer). This proves Lemma 2.10.11 (a).

(b) Assume that \( d \) is odd. Thus, \( (-1)^d = -1 \). Now, Lemma 2.10.11 (a) (applied to \(-y\) instead of \( y \)) yields \( x - (-y) \mid x^d - (-y)^d \). Since \( x - (-y) = x + y \) and \( x^d - (-y)^d = x^d - (-1)^d y^d = x^d - (-1) y^d = x^d + y^d \), this rewrites as \( x + y \mid x^d + y^d \). This proves Lemma 2.10.11 (b).

Next, let us recall a basic fact from combinatorics (the “Little Gauss” sum):

Proposition 2.10.12. Let \( n \in \mathbb{N} \). Then,

\[
1 + 2 + \cdots + n = \frac{n(n + 1)}{2}.
\]
Proof of Proposition 2.10.12. Here is one of several equally valid arguments:

$$2 \cdot (1 + 2 + \cdots + n) = (1 + 2 + \cdots + n) + \underbrace{(1 + 2 + \cdots + n)}_{= n + (n-1) + \cdots + 1}$$

(here, we have reversed the order of the addends)

$$= \left( \sum_{k=1}^{n} k \right) + \left( \sum_{k=1}^{n} (n+1-k) \right)$$

$$= \sum_{k=1}^{n} (n+1) = n(n+1).$$

Thus, $$1 + 2 + \cdots + n = \frac{n(n+1)}{2},$$ so that Proposition 2.10.12 is proven. \qed

Proposition 2.10.12 tells us what the sum $$1 + 2 + \cdots + n$$ of the first $$n$$ positive integers is. One might also ask what the sum $$1^2 + 2^2 + \cdots + n^2$$ of their squares is, and similarly for higher powers. While this is tangential to our course, let us collect some formulas for this:

**Proposition 2.10.13.** Let $$n \in \mathbb{N}$$. Then:

(a) We have $$1 + 2 + \cdots + n = \frac{1}{2}n(n+1).$$

(b) We have $$1^2 + 2^2 + \cdots + n^2 = \frac{1}{3}n(n+1)(2n+1).$$

(c) We have $$1^3 + 2^3 + \cdots + n^3 = \frac{1}{4}n^2(n+1)^2.$$

(d) We have $$1^4 + 2^4 + \cdots + n^4 = \frac{1}{5}n(2n+1)(n+1)(3n+3n^2-1).$$

(e) We have $$1^5 + 2^5 + \cdots + n^5 = \frac{1}{6}n^2(n+1)^2(2n+2n^2-1).$$

Each part of Proposition 2.10.13 can be straightforwardly proven by induction on $$n$$; we don’t need ingenious arguments like the one we gave above for Proposition 2.10.12 (and in fact, such arguments cannot always be found).

You probably see a pattern in Proposition 2.10.13. It appears that for each positive integer $$d$$, there exists some polynomial $$p_d(x)$$ of degree $$d+1$$ with rational coefficients such that each $$n \in \mathbb{N}$$ satisfies $$1^d + 2^d + \cdots + n^d = p_d(n)$$. This is indeed the case. Indeed, this is proven (e.g.) in [Galvin17, Proposition 23.2] and in [Grinbe17, Theorem 3.7]. The polynomial $$p_d(x)$$ is uniquely determined for each $$d$$, and can be explicitly computed via the formula

$$p_d(x) = \sum_{k=1}^{d} k! \binom{d}{k} \left( \frac{x+1}{k+1} \right).$$
where \( \binom{x+1}{k+1} = \frac{(x+1)x(x-1)\cdots(x-k+1)}{(k+1)!} \) and where \( \binom{d}{k} \) is a Stirling number of the 2nd kind. Without going into the details of what Stirling numbers of the 2nd kind are, let me say that \( k! \binom{d}{k} \) is the number of surjective maps from \( \{1, 2, \ldots, d\} \) to \( \{1, 2, \ldots, k\} \). For example,

\[
p_2(x) = \sum_{k=1}^{2} k! \binom{2}{k} \frac{x+1}{k+1} = 1! \binom{2}{1} \left( \frac{x+1}{2} \right) + 2! \binom{2}{2} \left( \frac{x+1}{3} \right)
\]

\[
= \left( \frac{x+1}{2} \right) + 2 \left( \frac{x+1}{3} \right) = \frac{(x+1)x}{2} + 2 \cdot \frac{(x+1)x(x-1)}{6}
\]

and thus

\[
1^2 + 2^2 + \cdots + n^2 = p_2(n) = \frac{1}{6} n (n+1) (2n+1) \quad \text{for each } n \in \mathbb{N}.
\]

This recovers the claim of Proposition 2.10.13 (b). The combinatorial proof presented in [Galvin17 Proposition 23.2] is highly recommended reading for anyone interested in this kind of formulas.

Let us note that the polynomials \( p_d(x) \) do not have integer coefficients, but nevertheless all their values \( p_d(n) \) for \( n \in \mathbb{N} \) are integers.

Let us now show the power of Theorem 2.10.7 on the following exercise:

**Exercise 2.10.8.** Let \( n \in \mathbb{N} \). Let \( d \) be an odd positive integer. Prove that

\[
1 + 2 + \cdots + n \mid 1^d + 2^d + \cdots + n^d.
\]

**[Hint: Use Proposition 2.10.12 to reduce the claim to proving that \( n(n+1) \mid 2 \left( 1^d + 2^d + \cdots + n^d \right) \). But Theorem 2.10.7 shows that in order to prove this, it suffices to prove \( n \mid 2 \left( 1^d + 2^d + \cdots + n^d \right) \) and \( n+1 \mid 2 \left( 1^d + 2^d + \cdots + n^d \right) \), because \( n \perp n+1 \).]**

2.10.4. More properties of gcds and coprimality

The following is a random collection of further exercises on gcds.

**Exercise 2.10.9.** Let \( a, b, x, y \) be integers such that \( xa + yb = 1 \). Prove that \( a \perp b \).

**Exercise 2.10.10.** Let \( u, v, x, y \in \mathbb{Z} \). Prove that \( \text{gcd}(u,v) \cdot \text{gcd}(x,y) = \text{gcd}(ux, uy, vx, vy) \).
Exercise 2.10.11. Let \( a, b, c \in \mathbb{Z} \).

(a) Prove that \( \gcd(a, b) \cdot \gcd(a, c) = \gcd(\gcd(a, b), \gcd(a, c)) \), where \( g = \gcd(a, b, c) \).

(b) Prove that \( \gcd(a, b) \cdot \gcd(a, c) = \gcd(a, bc) \) if \( b \perp c \).

Exercise 2.10.12. Let \( a \) and \( b \) be two integers that are not both zero. Let \( g = \gcd(a, b) \). Prove that \( \frac{a}{g} \) and \( \frac{b}{g} \) are integers satisfying \( \frac{a}{g} \perp \frac{b}{g} \).

Exercise 2.10.13. Let \( a \) and \( b \) be two integers. Let \( k \in \mathbb{N} \). Prove that \( \gcd(a^k, b^k) = (\gcd(a, b))^k \).

The next exercise is simply claiming the well-known fact that any rational number can be written as a reduced fraction:

Exercise 2.10.14. Let \( r \in \mathbb{Q} \). Prove that there exist two coprime integers \( a \) and \( b \) satisfying \( r = \frac{a}{b} \).

As an application of some of the preceding results, we can prove that certain numbers are irrational:

Exercise 2.10.15. Prove the following:

(a) If a positive integer \( u \) is not a perfect square\(^{21}\), then \( \sqrt{u} \) is irrational.

(b) If \( u \) and \( v \) are two positive integers, then \( \sqrt{u} + \sqrt{v} \) is irrational unless both \( u \) and \( v \) are perfect squares.

Exercise 2.10.15 invites a rather natural generalization: If \( u_1, u_2, \ldots, u_k \) are several positive integers that are not all perfect squares, then must \( \sqrt{u_1} + \sqrt{u_2} + \cdots + \sqrt{u_k} \) always be irrational? It turns out that the answer is “yes”, but this is not as easy to prove anymore as the two cases \( k = 1 \) and \( k = 2 \) that we handled in Exercise 2.10.15. Proofs of the general version can be found in [Boreic08] (actually, a stronger statement is proven there, although it takes some work to derive ours from it).

Let us generalize Exercise 2.10.10 a bit:

Exercise 2.10.16. Let \( x, y \in \mathbb{Z} \), and let \( a_1, a_2, \ldots, a_k \) be finitely many integers. Prove that

\[
\gcd(a_1, a_2, \ldots, a_k) \cdot \gcd(x, y) = \gcd(a_1 x, a_2 x, \ldots, a_k x, a_1 y, a_2 y, \ldots, a_k y).
\]

We can extend this exercise further to several integers instead of \( x \) and \( y \), but this extension would be notationally awkward, so we only state it for the case of three integers:

\(^{21}\)A perfect square means the square of an integer.
Exercise 2.10.17. Let \( x, y, z \in \mathbb{Z} \), and let \( a_1, a_2, \ldots, a_k \) be finitely many integers. Prove that

\[
gcd(a_1, a_2, \ldots, a_k) \cdot gcd(x, y, z) = gcd(a_1x, a_2x, \ldots, a_kx, a_1y, a_2y, \ldots, a_ky, a_1z, a_2z, \ldots, a_kz).
\]

We leave it to the reader to state and solve an exercise generalizing Exercise 2.10.16 and Exercise 2.10.17.

Exercise 2.10.18. Let \( a, b, c \in \mathbb{Z} \). Prove that

\[
gcd(b, c) \cdot gcd(c, a) \cdot gcd(a, b) = gcd(a, b, c) \cdot gcd(bc, ca, ab).
\]

2019-02-08 lecture

2.11. Lowest common multiples

Common multiples are, in a sense, a “mirror version” of common divisors. Here is their definition:

**Definition 2.11.1.** Let \( b_1, b_2, \ldots, b_k \) be integers. Then, the common multiples of \( b_1, b_2, \ldots, b_k \) are defined to be the integers \( a \) that satisfy

\[
(b_i \mid a \text{ for all } i \in \{1, 2, \ldots, k\}).
\]

(In other words, a common multiple of \( b_1, b_2, \ldots, b_k \) is an integer that is a multiple of each of \( b_1, b_2, \ldots, b_k \).) We let \( \text{Mul}(b_1, b_2, \ldots, b_k) \) denote the set of these common multiples.

**Example 2.11.2.** The common multiples of \( 4, 6 \) are \( \ldots, -36, -24, -12, 0, 12, 24, 36, \ldots \), that is, all multiples of 12.

The common multiples of \( 1, 2, 3 \) are all multiples of 6.

Note that the common multiples of a single integer \( b \) are simply the multiples of \( b \). (Also, the common multiples of an empty list of integers are all the integers; in other words, \( \text{Mul}() = \mathbb{Z} \).)

Note that the definition of common multiples of \( b_1, b_2, \ldots, b_k \) (Definition 2.11.1) is the same as the definition of common divisors of \( b_1, b_2, \ldots, b_k \) except that the divisibility has been flipped (i.e., it says “\( b_i \mid a \)” instead of “\( a \mid b_i \)”)

This is why common multiples are a “mirror version” of common divisors. This analogy is not perfect – in particular, (for example) two nonzero integers have infinitely many common multiples but only finitely many common divisors. We shall now introduce lowest common multiples, which correspond to greatest common divisors in this analogy. However, we have to prove a simple proposition first:
**Proposition 2.11.3.** Let \( b_1, b_2, \ldots, b_k \) be finitely many nonzero integers. Then, the set \( \text{Mul}(b_1, b_2, \ldots, b_k) \) has a smallest positive element.

Proposition 2.11.3 is similar to Proposition 2.9.5 (and will play a similar role), but note the differences: It requires all of \( b_1, b_2, \ldots, b_k \) to be nonzero (unlike Proposition 2.9.5, which needed only one of them to be nonzero), and it does not claim finiteness of any set.

**Proof of Proposition 2.11.3** We claim that 
\[
|b_1 b_2 \cdots b_k| \in \text{Mul}(b_1, b_2, \ldots, b_k).
\] (26)

**Proof of (26):** Let \( i \in \{1,2,\ldots,k\} \). Then, the product \( b_1 b_2 \cdots b_k \) can be written as 
\[
b_1 b_2 \cdots b_k = b_i \cdot (b_1 b_2 \cdots b_{i-1} b_{i+1} b_{i+2} \cdots b_k),
\]
and thus is divisible by \( b_i \). In other words, \( b_i \mid b_1 b_2 \cdots b_k \). But Exercise 2.2.1 (a) (applied to \( a = b_1 b_2 \cdots b_k \)) yields \( b_1 b_2 \cdots b_k \mid |b_1 b_2 \cdots b_k| \). Altogether, \( b_i \mid b_1 b_2 \cdots b_k \mid |b_1 b_2 \cdots b_k| \).

Now forget that we fixed \( i \). We thus have proven \( b_i \mid |b_1 b_2 \cdots b_k| \) for all \( i \in \{1,2,\ldots,k\} \). In other words, \( |b_1 b_2 \cdots b_k| \) is a common multiple of \( b_1, b_2, \ldots, b_k \) (by the definition of a “common multiple”). In other words, \( |b_1 b_2 \cdots b_k| \in \text{Mul}(b_1, b_2, \ldots, b_k) \). This proves (26).]

We know that \( b_1, b_2, \ldots, b_k \) are nonzero integers. Hence, their product \( b_1 b_2 \cdots b_k \) is a nonzero integer as well. Thus, its absolute value \( |b_1 b_2 \cdots b_k| \) is a positive integer. Hence, \( |b_1 b_2 \cdots b_k| \) is a positive element of \( \text{Mul}(b_1, b_2, \ldots, b_k) \) (since (26) shows that it is an element of \( \text{Mul}(b_1, b_2, \ldots, b_k) \)). Thus, the set \( \text{Mul}(b_1, b_2, \ldots, b_k) \) has a positive element. Therefore, this set \( \text{Mul}(b_1, b_2, \ldots, b_k) \) has a smallest positive element as well. This proves Proposition 2.11.3.

**Definition 2.11.4.** Let \( b_1, b_2, \ldots, b_k \) be finitely many integers. The lowest common multiple of \( b_1, b_2, \ldots, b_k \) is defined as follows:

- If \( b_1, b_2, \ldots, b_k \) are all nonzero, then it is defined as the smallest positive element of the set \( \text{Mul}(b_1, b_2, \ldots, b_k) \). This smallest positive element is well-defined (by Proposition 2.11.3), and it is a positive integer (obviously).

- If \( b_1, b_2, \ldots, b_k \) are not all nonzero (i.e., at least one of \( b_1, b_2, \ldots, b_k \) is zero), then it is defined to be 0.

Thus, in either case, this lowest common multiple is a nonnegative integer. We denote it by \( \text{lcm}(b_1, b_2, \ldots, b_k) \). (Some authors also call it \( |b_1, b_2, \ldots, b_k| \).

We shall also use the word “lcm” as shorthand for “lowest common multiple”.

---

\( \text{Proposition 2.11.3} \) is similar to \( \text{Proposition 2.9.5} \) (and will play a similar role), but note the differences: It requires all of \( b_1, b_2, \ldots, b_k \) to be nonzero (unlike Proposition 2.9.5, which needed only one of them to be nonzero), and it does not claim finiteness of any set.

**Proof of Proposition 2.11.3** We claim that 
\[
|b_1 b_2 \cdots b_k| \in \text{Mul}(b_1, b_2, \ldots, b_k).
\] (26)

**Proof of (26):** Let \( i \in \{1,2,\ldots,k\} \). Then, the product \( b_1 b_2 \cdots b_k \) can be written as 
\[
b_1 b_2 \cdots b_k = b_i \cdot (b_1 b_2 \cdots b_{i-1} b_{i+1} b_{i+2} \cdots b_k),
\]
and thus is divisible by \( b_i \). In other words, \( b_i \mid b_1 b_2 \cdots b_k \). But Exercise 2.2.1 (a) (applied to \( a = b_1 b_2 \cdots b_k \)) yields \( b_1 b_2 \cdots b_k \mid |b_1 b_2 \cdots b_k| \). Altogether, \( b_i \mid b_1 b_2 \cdots b_k \mid |b_1 b_2 \cdots b_k| \).

Now forget that we fixed \( i \). We thus have proven \( b_i \mid |b_1 b_2 \cdots b_k| \) for all \( i \in \{1,2,\ldots,k\} \). In other words, \( |b_1 b_2 \cdots b_k| \) is a common multiple of \( b_1, b_2, \ldots, b_k \) (by the definition of a “common multiple”). In other words, \( |b_1 b_2 \cdots b_k| \in \text{Mul}(b_1, b_2, \ldots, b_k) \). This proves (26).

We know that \( b_1, b_2, \ldots, b_k \) are nonzero integers. Hence, their product \( b_1 b_2 \cdots b_k \) is a nonzero integer as well. Thus, its absolute value \( |b_1 b_2 \cdots b_k| \) is a positive integer. Hence, \( |b_1 b_2 \cdots b_k| \) is a positive element of \( \text{Mul}(b_1, b_2, \ldots, b_k) \) (since (26) shows that it is an element of \( \text{Mul}(b_1, b_2, \ldots, b_k) \)). Thus, the set \( \text{Mul}(b_1, b_2, \ldots, b_k) \) has a positive element. Therefore, this set \( \text{Mul}(b_1, b_2, \ldots, b_k) \) has a smallest positive element as well. This proves Proposition 2.11.3.

**Definition 2.11.4.** Let \( b_1, b_2, \ldots, b_k \) be finitely many integers. The lowest common multiple of \( b_1, b_2, \ldots, b_k \) is defined as follows:

- If \( b_1, b_2, \ldots, b_k \) are all nonzero, then it is defined as the smallest positive element of the set \( \text{Mul}(b_1, b_2, \ldots, b_k) \). This smallest positive element is well-defined (by Proposition 2.11.3), and it is a positive integer (obviously).

- If \( b_1, b_2, \ldots, b_k \) are not all nonzero (i.e., at least one of \( b_1, b_2, \ldots, b_k \) is zero), then it is defined to be 0.

Thus, in either case, this lowest common multiple is a nonnegative integer. We denote it by \( \text{lcm}(b_1, b_2, \ldots, b_k) \). (Some authors also call it \( |b_1, b_2, \ldots, b_k| \).

We shall also use the word “lcm” as shorthand for “lowest common multiple”.

---

\( \text{Proposition 2.11.3} \) is similar to \( \text{Proposition 2.9.5} \) (and will play a similar role), but note the differences: It requires all of \( b_1, b_2, \ldots, b_k \) to be nonzero (unlike Proposition 2.9.5, which needed only one of them to be nonzero), and it does not claim finiteness of any set.
Some authors say “least common multiple” instead of “lowest common multiple”.

We are slightly abusing the word “lowest common multiple”, of course; it would be more precise to say “lowest positive common multiple”, and even this would only hold for the case when \( b_1, b_2, \ldots, b_k \) are all nonzero. Taken literally, a “lowest common multiple” of 2 and 3 would not exist, since 2 and 3 have infinitely many negative common multiples.

Note that the lcm of a single number is the absolute value of this number: i.e., we have \( \text{lcm}(a) = |a| \) for each \( a \in \mathbb{Z} \). (This is easy to prove.) Also, the lcm of an empty list of numbers is 1: that is, \( \text{lcm}() = 1 \).

We observe a trivial property of lcms, which (for the sake of brevity) we only state for two integers \( a \) and \( b \) despite it holding for any number of integers (with the same proof):

**Proposition 2.11.5.** Let \( a, b \in \mathbb{Z} \).

(a) We have \( 0 \in \text{Mul}(a, b) \).

(b) We have \( \text{lcm}(a, b) \in \text{Mul}(a, b) \).

(c) We have \( a \mid \text{lcm}(a, b) \) and \( b \mid \text{lcm}(a, b) \).

**Proof of Proposition 2.11.5** (a) The integer 0 clearly satisfies \( (a \mid 0 \text{ and } b \mid 0) \). In other words, 0 is a common multiple of \( a \) and \( b \) (by the definition of a “common multiple”). In other words, \( 0 \in \text{Mul}(a, b) \) (by the definition of \( \text{Mul}(a, b) \)). This proves Proposition 2.11.5 (a).

(b) If the two integers \( a \) and \( b \) are not all nonzero, then Proposition 2.11.5 (b) holds\(^{23}\). Hence, for the rest of this proof, we WLOG assume that the two integers \( a \) and \( b \) are all nonzero. Thus, Definition 2.11.4 yields that \( \text{lcm}(a, b) \) is the smallest positive element of the set \( \text{Mul}(a, b) \). Hence, \( \text{lcm}(a, b) \in \text{Mul}(a, b) \). This proves Proposition 2.11.5 (b).

(c) Proposition 2.11.5 (b) yields \( \text{lcm}(a, b) \in \text{Mul}(a, b) \). In other words, \( \text{lcm}(a, b) \) is a common multiple of \( a \) and \( b \) (by the definition of \( \text{Mul}(a, b) \)). In other words, we have \( (a \mid \text{lcm}(a, b) \text{ and } b \mid \text{lcm}(a, b)) \) (by the definition of “common multiple”). This proves Proposition 2.11.5 (c). \( \square \)

The following theorem yields a good way of computing lcms of two numbers (since we already know how to compute gcds via the Euclidean algorithm):

**Theorem 2.11.6.** Let \( a, b \in \mathbb{Z} \). Then, \( \text{gcd}(a, b) \cdot \text{lcm}(a, b) = |ab| \).

**Proof of Theorem 2.11.6** If at least one of the two numbers \( a \) and \( b \) is 0, then Theorem 2.11.6 holds\(^{24}\). Hence, for the rest of this proof, we WLOG assume that none of the

\(^{23}\)Proof. Assume that the two integers \( a \) and \( b \) are not all nonzero. Hence, Definition 2.11.4 shows that \( \text{lcm}(a, b) = 0 \in \text{Mul}(a, b) \) (by Proposition 2.11.5 (a)). Thus, Proposition 2.11.5 (b) holds.

\(^{24}\)Proof. Assume that at least one of the two numbers \( a \) and \( b \) is 0. Thus, the product \( ab \) is 0. Hence, \( ab = 0 \), so that \( |ab| = 0 \).

On the other hand, the two numbers \( a, b \) are not all nonzero (since at least one of the two numbers \( a \) and \( b \) is 0). Hence, Definition 2.11.4 shows that \( \text{lcm}(a, b) = 0 \). Comparing \( \text{gcd}(a, b) \cdot \)
two numbers $a$ and $b$ is 0. In other words, $a$ and $b$ are nonzero. Thus, Definition 2.11.4 yields that lcm $(a, b)$ is the smallest positive element of the set Mul $(a, b)$. Also, gcd $(a, b)$ is a positive integer (since $a$ and $b$ are nonzero) and thus nonzero. Hence, we can define $c \in \mathbb{Q}$ by $c = \frac{ab}{\gcd(a, b)}$. Consider this $c$. From $c = \frac{ab}{\gcd(a, b)}$, we obtain $ab = \gcd(a, b) \cdot c$.

Let $d = |c|$. The number $c = \frac{ab}{\gcd(a, b)}$ is nonzero (since $a$ and $b$ are nonzero). Hence, its absolute value $|c|$ is positive. In other words, $d$ is positive (since $d = |c|$). From $ab = \gcd(a, b) \cdot c$, we obtain

$$|ab| = \left| \gcd(a, b) \cdot c \right| = \left| \gcd(a, b) \right| \cdot \left| c \right| = d$$

(by (3), applied to $\gcd(a, b)$ and $c$ instead of $x$ and $y$)

$$= \gcd(a, b) \cdot d.$$ (27)

Solving this for $d$, we find $d = \frac{|ab|}{\gcd(a, b)}$ (since $\gcd(a, b)$ is nonzero).

We have $\gcd(a, b) \mid b$ (by Proposition 2.9.7 (f)). Thus, $\frac{b}{\gcd(a, b)}$ is an integer. Now, $c = \frac{ab}{\gcd(a, b)} = \frac{a}{\gcd(a, b)} \cdot \frac{b}{\gcd(a, b)}$ is the product of two integers (since $a$ and $\frac{b}{\gcd(a, b)}$ are integers). Therefore, $c$ itself is an integer. Thus, $d$ is an integer as well (since $d = |c|$). Moreover, $c = \frac{ab}{\gcd(a, b)}$ shows that $b \mid c$ (since $\frac{b}{\gcd(a, b)}$ is an integer). But Exercise 2.2.1 (a) (applied to $c$ instead of $a$) yields $c \mid |c|$ (this means “$c$ divides $|c|$”). In other words, $c \mid d$ (since $d = |c|$). Hence, $a \mid c \mid d$.

So we have proven that $a \mid d$. Similarly, $b \mid d$. Thus, we know that $(a \mid d$ and $b \mid d)$. In other words, $d$ is a common multiple of $a$ and $b$ (by the definition of a “common multiple”). In other words, $d \in \text{Mul}(a, b)$ (by the definition of Mul $(a, b)$). Thus, $d$ is a positive element of the set Mul $(a, b)$ (since $d \in \text{Mul}(a, b)$).

We shall now show that $d$ is the smallest positive element of this set. Indeed, let $x$ be any positive element of Mul $(a, b)$. We are going to prove that $x \geq d$.

In fact, $x \in \text{Mul}(a, b)$. In other words, $x$ is a common multiple of $a$ and $b$. In other words, we have $(a \mid x$ and $b \mid x)$. Hence, Theorem 2.9.16 (applied to $x$ instead of $c$) yields $ab \mid \gcd(a, b) \cdot x$. Both numbers $\gcd(a, b) \cdot x$ are positive; hence, their product $\gcd(a, b) \cdot x$ is positive as well, and thus we have $\gcd(a, b) \cdot x \neq 0$. Hence, Proposition 2.2.3 (b) (applied to $ab$ and $\gcd(a, b) \cdot x$ instead of $a$ and $b$)

\[ \text{lcm}(a, b) = 0 \] with \[ |ab| = 0 \], we obtain \[ \gcd(a, b) \cdot \text{lcm}(a, b) = |ab| \]. In other words, Theorem 2.11.6 holds.
yields $|ab| \leq |\gcd(a, b) \cdot x| = \gcd(a, b) \cdot x$ (since $\gcd(a, b) \cdot x$ is positive). Thus,
\[
\gcd(a, b) \cdot x \geq |ab| = \gcd(a, b) \cdot d \quad \text{(by (27)).}
\]
We can divide this inequality by $\gcd(a, b)$ (since $\gcd(a, b)$ is positive), and thus obtain $x \geq d$.

Now, forget that we fixed $x$. We thus have proven that each positive element $x$ of the set $\Mul(a, b)$ satisfies $x \geq d$. Hence, $d$ is the smallest positive element of the set $\Mul(a, b)$ (since we already know that $d$ is a positive element of the set $\Mul(a, b)$). In other words, $d = \lcm(a, b)$ (since $\lcm(a, b)$ is the smallest positive element of the set $\Mul(a, b)$). In other words, $d = \lcm(a, b)$. Hence, (27) becomes $|ab| = \gcd(a, b) \cdot \frac{d}{\lcm(a, b)} = \gcd(a, b) \cdot \lcm(a, b)$. This proves Theorem 2.11.6. \qed

Next, we state an analogue of Theorem 2.9.14 (with all divisibilities flipped):

**Theorem 2.11.7.** Let $a, b \in \mathbb{Z}$. Then:

(a) For each $m \in \mathbb{Z}$, we have the following logical equivalence:

$$(a \mid m \text{ and } b \mid m) \iff (\lcm(a, b) \mid m).$$

(28)

(b) The common multiples of $a$ and $b$ are precisely the multiples of $\lcm(a, b)$.

(c) We have $\Mul(a, b) = \Mul(\lcm(a, b))$.

Again, the three parts of this theorem are saying the same thing from slightly different perspectives. Our proof of Theorem 2.11.7 will rely on the following lemma:

**Lemma 2.11.8.** Let $m, a, b \in \mathbb{Z}$ be such that $a \mid m$ and $b \mid m$. Then, $\lcm(a, b) \mid m$.

Lemma 2.11.8 is similar to Lemma 2.9.15 but its proof is not:

**Proof of Lemma 2.11.8.** If at least one of the two numbers $a$ and $b$ is 0, then Lemma 2.11.8 holds.\(^{25}\) Hence, for the rest of this proof, we WLOG assume that none of the two numbers $a$ and $b$ is 0. In other words, $a$ and $b$ are nonzero. Thus, Definition 2.11.4 yields that $\lcm(a, b)$ is the smallest positive element of the set $\Mul(a, b)$. Set $n = \lcm(a, b)$. Thus, $n$ is the smallest positive element of the set $\Mul(a, b)$ (since $\lcm(a, b)$ is the smallest positive element of the set $\Mul(a, b)$). Therefore, $n$ is a positive integer and belongs to $\Mul(a, b)$.

Now, $n$ is a common multiple of $a$ and $b$ (since $n$ belongs to $\Mul(a, b)$). In other words, we have $(a \mid n$ and $b \mid n)$.

\(^{25}\)Proof. Assume that at least one of the two numbers $a$ and $b$ is 0. In other words, $a = 0$ or $b = 0$. Let us WLOG assume that $a = 0$ (since the proof in the case $b = 0$ is analogous). We have $a \mid m$, thus $0 = a \mid m$.

On the other hand, the two numbers $a, b$ are not all nonzero (since at least one of the two numbers $a$ and $b$ is 0). Hence, Definition 2.11.4 shows that $\lcm(a, b) = 0 = a \mid m$. In other words, Lemma 2.11.8 holds.
Our goal is to prove that \( \text{lcm}(a, b) \mid m \). In other words, our goal is to prove that \( n \mid m \) (since \( n = \text{lcm}(a, b) \)). Assume the contrary. Thus, we don’t have \( n \mid m \). Hence, we don’t have \( m \mid n = 0 \) (because Corollary 2.6.9 (b) (applied to \( u = m \)) shows that we have \( n \mid m \) if and only if \( m \mid n = 0 \)). In other words, we have \( m \mid n \neq 0 \).

Corollary 2.6.9 (a) (applied to \( u = m \)) yields that \( m \mid n \in \{0, 1, \ldots, n - 1\} \) and \( m \mid n \equiv m \mod n \). Combining \( m \mid n \in \{0, 1, \ldots, n - 1\} \) with \( m \mid n \neq 0 \), we obtain \( m \mid n \in \{0, 1, \ldots, n - 1\} \setminus \{0\} = \{1, 2, \ldots, n - 1\} \). Hence, \( m \mid n \) is a positive integer and satisfies \( m \mid n \leq n - 1 < n \).

From \( m \mid n \equiv m \mod n \) and \( a \mid n \), we obtain \( m \mid n \equiv m \mod a \) (by Proposition 2.3.4 (e), applied to \( a, m \mid n \) and \( m \mid n \)). But \( m \equiv 0 \mod a \) (since \( a \mid m \)). Thus, \( m \mid n \equiv m \equiv 0 \mod a \). In other words, \( a \mid m \mid n \). Similarly, \( b \mid m \mid n \).

So we have proven that \( (a \mid m \mid n \text{ and } b \mid m \mid n) \). In other words, \( m \mid n \) is a common multiple of \( a \) and \( b \). In other words, \( m \mid n \in \text{Mul}(a, b) \). Therefore, \( m \mid n \) is a positive element of \( \text{Mul}(a, b) \) (since \( m \mid n \) is positive). Thus, \( m \mid n \geq n \) (since \( n \) is the smallest positive element of \( \text{Mul}(a, b) \)). This contradicts the fact that \( m \mid n < n \). This contradiction shows that our assumption was false. Hence, Lemma 2.11.8 is proven.

Proof of Theorem 2.11.7. (a) Let \( m \in \mathbb{Z} \). In order to prove (28), we need to prove the “\( \Rightarrow \)” and “\( \Leftarrow \)” directions of the equivalence (28). But this is easy: The “\( \Rightarrow \)” direction is just the statement of Lemma 2.11.8 whereas the “\( \Leftarrow \)” direction is trivial (to wit: if \( \text{lcm}(a, b) \mid m \), then

\[
\begin{align*}
\frac{a}{\text{lcm}(a, b)} & \mid m \\
\frac{b}{\text{lcm}(a, b)} & \mid m
\end{align*}
\]

and thus \( (a \mid m \text{ and } b \mid m) \)). Hence, the equivalence (28) is proven. This proves Theorem 2.11.7 (a).

(b) Theorem 2.11.7 (b) can be derived from Theorem 2.11.7 (a) in the same way as Theorem 2.9.14 (b) was derived from Theorem 2.9.14 (a) (after the necessary changes are made – such as flipping all divisibility relations and replacing “divisor” by “multiple”).

(c) Theorem 2.11.7 (c) can be derived from Theorem 2.11.7 (b) in the same way as Theorem 2.9.14 (c) was derived from Theorem 2.9.14 (b) (after the necessary changes are made – such as flipping all divisibility relations and replacing “divisor” by “multiple”).

Our next claim is an analogue of Theorem 2.9.20.
Theorem 2.11.9. Let $b_1, b_2, \ldots, b_k$ be integers.

(a) For each $m \in \mathbb{Z}$, we have the following logical equivalence:

\[(b_i \mid m \text{ for all } i \in \{1,2,\ldots,k\}) \iff (\text{lcm} (b_1, b_2, \ldots, b_k) \mid m)\].

(b) The common multiples of $b_1, b_2, \ldots, b_k$ are precisely the multiples of $\text{lcm} (b_1, b_2, \ldots, b_k)$.
(c) We have $\text{Mul} (b_1, b_2, \ldots, b_k) = \text{Mul} (\text{lcm} (b_1, b_2, \ldots, b_k))$.
(d) If $k > 0$, then

\[\text{lcm} (b_1, b_2, \ldots, b_k) = \text{lcm} (\text{lcm} (b_1, b_2, \ldots, b_{k-1}), b_k)\].

Proof of Theorem 2.11.9 (sketched). It is not hard to transform our above proof of Theorem 2.9.20 into a proof of Theorem 2.11.9. To do so, we need (of course) to flip the divisibility relations and replace “divisor” by “multiple” and “gcd” by “lcm”. (Some more changes need to be made as well – for example, the induction base needs to be handled differently, and the WLOG assumption that “the integers $b_1, b_2, \ldots, b_t$ are not all 0” needs to be replaced by a WLOG assumption that “the integers $b_1, b_2, \ldots, b_t$ are all nonzero”. Also, “largest element” needs to be replaced by “smallest positive element”. But these are fairly straightforward changes; the main thrust of the argument remains unchanged.)

Exercise 2.11.1. Let $a, b \in \mathbb{Z}$.

(a) Prove that $\text{lcm} (a, b) = \text{lcm} (b, a)$.
(b) Prove that $\text{lcm} (-a, b) = \text{lcm} (a, b)$.
(c) Prove that $\text{lcm} (a, -b) = \text{lcm} (a, b)$.
(d) If $a \mid b$, then $\text{lcm} (a, b) = |b|$.
(e) Let $s \in \mathbb{Z}$. Prove that $\text{lcm} (sa, sb) = |s| \text{lcm} (a, b)$.

Exercise 2.11.2. Let $a, b, c$ be three integers.

(a) Prove that $\text{gcd} (a, b, c) \cdot \text{lcm} (bc, ca, ab) = |abc|$.
(b) Prove that $\text{lcm} (a, b, c) \cdot \text{gcd} (bc, ca, ab) = |abc|$.

2.12. The Chinese remainder theorem (elementary form)

Theorem 2.12.1. Let $m$ and $n$ be two coprime integers. Let $a, b \in \mathbb{Z}$.

(a) There exists an integer $x \in \mathbb{Z}$ such that

\[(x \equiv a \pmod{m} \text{ and } x \equiv b \pmod{n})\].

(b) If $x_1$ and $x_2$ are two such integers $x$, then $x_1 \equiv x_2 \pmod{mn}$.

Theorem 2.12.1 is known as the Chinese remainder theorem. More precisely, there is a sizeable cloud of results that share this name; Theorem 2.12.1 is one of the
most elementary and basic of these results. A more general result is Theorem 2.12.4 further below. However, the strongest and most general “Chinese remainder theorems” rely on concept from abstract algebra such as rings and ideals; it will take us a while to get to them.

Theorem 2.12.1 has gotten its name from the fact that a first glimpse of it appears in “Master Sun’s Mathematical Manual” from the 3rd century AD; it took centuries until it become a theorem with proof and precise statement.

The claim of Theorem 2.12.1 (b) is often restated as “This integer \( x \) (i.e., the integer \( x \) satisfying \( x \equiv a \mod m \) and \( x \equiv b \mod n \)) is unique modulo \( mn \).” The “modulo \( mn \)” here signifies that what we are not claiming literal uniqueness (which would mean that if \( x_1 \) and \( x_2 \) are two such integers \( x \), then \( x_1 = x_2 \)), but merely claiming a weaker form (namely, that if \( x_1 \) and \( x_2 \) are two such integers \( x \), then \( x_1 \equiv x_2 \mod mn \).

Example 2.12.2. Theorem 2.12.1 (a) (applied to \( m = 5 \), \( n = 6 \) and \( a = 3 \) and \( b = 2 \)) shows that there exists an integer \( x \in \mathbb{Z} \) such that

\[
(x \equiv 3 \mod 5 \text{ and } x \equiv 2 \mod 6).
\]

We will soon find such an integer, after we have proved Theorem 2.12.1.

Proof of Theorem 2.12.1. The integers \( m \) and \( n \) are coprime. In other words, \( m \perp n \) (by Proposition 2.10.4).

(a) Theorem 2.10.8 (b) (applied to \( m \) instead of \( a \)) shows that there exists a \( m' \in \mathbb{Z} \) such that \( mm' \equiv 1 \mod n \).

Similarly, there exists an \( n' \in \mathbb{Z} \) such that \( nn' \equiv 1 \mod m \) (since \( m \) and \( n \) play symmetric roles in Theorem 2.12.1).

Now, set \( x_0 = nn'a + mm'b \). Then,

\[
x_0 = \underbrace{nn'}_{\equiv 1 \mod m} a + \underbrace{mm'}_{\equiv 0 \mod m} b \equiv 1a + 0 = a \mod m
\]

(here, we have used the Principle of substitutivity for congruences, which we described in Section 2.5) and similarly \( x_0 \equiv b \mod n \). Thus, there exists an integer \( x \in \mathbb{Z} \) such that \( (x \equiv a \mod m \text{ and } x \equiv b \mod n) \) (namely, \( x = x_0 \)). This proves Theorem 2.12.1 (a).

(b) Let \( x_1 \) and \( x_2 \) be two such integers \( x \). We want to prove that \( x_1 \equiv x_2 \mod mn \).

We know that \( x_1 \) is an integer \( x \) such that \( (x \equiv a \mod m \text{ and } x \equiv b \mod n) \). Thus, \( x_1 \equiv a \mod m \text{ and } x_1 \equiv b \mod n \).

In particular, \( x_1 \equiv a \mod m \), and similarly \( x_2 \equiv a \mod m \). Thus, \( x_1 \equiv a \equiv x_2 \mod m \), so that \( m \mid x_1 - x_2 \). Similarly, \( n \mid x_1 - x_2 \). Since \( m \perp n \), we thus obtain \( mn \mid x_1 - x_2 \) (by Theorem 2.10.7, applied to \( m, n \) and \( x_1 - x_2 \) instead of \( a, b \) and \( c \)). In other words, \( x_1 \equiv x_2 \mod mn \). This proves Theorem 2.12.1. \( \square \)
Example 2.12.3. Assume that we want to find an $x \in \mathbb{Z}$ such that
\[(x \equiv 3 \text{ mod } 5 \text{ and } x \equiv 2 \text{ mod } 6).\]

To compute such an $x$, let us follow the proof of Theorem 2.12.1 (a) above.

We need a modular inverse $5'$ of 5 modulo 6. Such an inverse is 5, since $5 \cdot 5 \equiv 1 \text{ mod } 6$. (In this particular case, finding this modular inverse was easy, because all we had to do is to test the 6 numbers 0, 1, 2, 3, 4, 5; it is clear that a modular inverse of $a$ modulo $m$, if it exists, can be found within the set \{0, 1, \ldots, m - 1\}.

In general, there is a quick way to find a modular inverse of an integer $a$ modulo an integer $m$ using the “Extended Euclidean algorithm”)

We need a modular inverse 6 of 6 modulo 5. Such an inverse is 1, since $6 \cdot 1 \equiv 1 \text{ mod } 5$.

Now, the proof of Theorem 2.12.1 (a) tells us that $x_0 = 6 \cdot 6' \cdot 3 + 5 \cdot 5' \cdot 2$ is an integer $x \in \mathbb{Z}$ such that $(x \equiv 3 \text{ mod } 5 \text{ and } x \equiv 2 \text{ mod } 6)$. This $x_0$ is
\[6 \cdot 6' \cdot 3 + 5 \cdot 5' \cdot 2 = 6 \cdot 1 \cdot 3 + 5 \cdot 5 \cdot 2 = 68.\]

So we have found an $x \in \mathbb{Z}$ such that $(x \equiv 3 \text{ mod } 5 \text{ and } x \equiv 2 \text{ mod } 6)$, namely $x = 68$. (We can easily check this: $68 \equiv 3 \text{ mod } 5$ since $68 - 3 = 5 \cdot 13$; and $68 \equiv 2 \text{ mod } 6$ since $68 - 2 = 6 \cdot 11$.)

There is also a version of Theorem 2.12.1 for multiple integers:

Theorem 2.12.4. Let $m_1, m_2, \ldots, m_k$ be $k$ mutually coprime integers. Let $a_1, a_2, \ldots, a_k \in \mathbb{Z}$.

(a) There exists an integer $x$ such that
\[(x \equiv a_i \text{ mod } m_i \text{ for all } i \in \{1, 2, \ldots, k\}).\]

(b) If $x_1$ and $x_2$ are two such integers $x$, then $x_1 \equiv x_2 \text{ mod } m_1 m_2 \cdots m_k$.

Again, Theorem 2.12.4 (b) is often stated in the form “This integer $x$ is unique modulo $m_1 m_2 \cdots m_k$”.

Clearly, Theorem 2.12.1 is the particular case of Theorem 2.12.4 obtained for $k = 2$.

Proof of Theorem 2.12.4 Forget that we fixed $k$ and $m_1, m_2, \ldots, m_k$ and $a_1, a_2, \ldots, a_k$.

(a) We shall prove Theorem 2.12.4 (a) by induction on $k$.

Induction base: Let us check that Theorem 2.12.4 (a) holds for $k = 0$. Indeed, if $k = 0$, then Theorem 2.12.4 (a) states the following:

Claim 0: Let $m_1, m_2, \ldots, m_0$ be 0 mutually coprime integers. Let $a_1, a_2, \ldots, a_0 \in \mathbb{Z}$. There exists an integer $x$ such that
\[(x \equiv a_i \text{ mod } m_i \text{ for all } i \in \{1, 2, \ldots, 0\}).\]
But Claim 0 is true, because \([30]\) is vacuously true for any integer \(x\) (so we can take, for example, \(x = 0\)). In other words, Theorem \(2.12.4\) holds for \(k = 0\); thus, the induction base is complete.

Needless to say, Claim 0 is not an interesting statement, but it is a perfectly valid induction base! (But you are free to check the case \(k = 1\) by hand – its proof is almost as easy as that for \(k = 0\).

**Induction step:** Let \(\ell\) be a positive integer. Assume that Theorem \(2.12.4\) holds for \(k = \ell - 1\). We must now prove that Theorem \(2.12.4\) holds for \(k = \ell\).

We have assumed that Theorem \(2.12.4\) holds for \(k = \ell - 1\). In other words, the following claim holds:

**Claim 1:** Let \(m_1, m_2, \ldots, m_{\ell-1}\) be \(\ell - 1\) mutually coprime integers. Let 
\(a_1, a_2, \ldots, a_{\ell-1} \in \mathbb{Z}\). There exists an integer \(x\) such that 
\((x \equiv a_i \mod m_i \text{ for all } i \in \{1, 2, \ldots, \ell - 1\})\). \((31)\)

We must prove that Theorem \(2.12.4\) holds for \(k = \ell\). In other words, we must prove the following claim:

**Claim 2:** Let \(m_1, m_2, \ldots, m_{\ell}\) be \(\ell\) mutually coprime integers. Let 
\(a_1, a_2, \ldots, a_{\ell} \in \mathbb{Z}\). There exists an integer \(x\) such that 
\((x \equiv a_i \mod m_i \text{ for all } i \in \{1, 2, \ldots, \ell\})\). \((32)\)

**Proof of Claim 2:** The main idea of this proof is to combine Claim 1 (applied to \(m_1, m_2, \ldots, m_{\ell-1}\)) with Theorem \(2.12.1\) (applied to the coprime integers \(m_1 m_2 \cdots m_{\ell-1}\) and \(m_{\ell}\)). In details:

The \(\ell\) integers \(m_1, m_2, \ldots, m_{\ell}\) are mutually coprime. Thus, the \(\ell - 1\) integers 
\(m_1, m_2, \ldots, m_{\ell-1}\) are mutually coprime. Hence, Claim 1 shows that there exists an 
integer \(x\) such that

\((x \equiv a_i \mod m_i \text{ for all } i \in \{1, 2, \ldots, \ell - 1\})\).

Consider this \(x\), and denote it by \(u\). Thus, \(u\) is an integer such that 
\((u \equiv a_i \mod m_i \text{ for all } i \in \{1, 2, \ldots, \ell - 1\})\). \((33)\)

Define an integer \(m = m_1 m_2 \cdots m_{\ell-1}\).

The integers \(m\) and \(m_{\ell}\) are coprime. \(27\) Hence, Theorem \(2.12.1\) (applied to
\(n = m_{\ell}, a = u\) and \(b = a_{\ell}\)) yields that there exists an integer \(x \in \mathbb{Z}\) such that 
\((x \equiv u \mod m \text{ and } x \equiv a_{\ell} \mod m_{\ell})\).

---

\(26\) Since there exists no \(i \in \{1, 2, \ldots, 0\}\)

\(27\) Recall that the \(\ell\) integers \(m_1, m_2, \ldots, m_{\ell}\) are mutually coprime. In other words, \(m_i \perp m_j\)
for any \(i, j \in \{1, 2, \ldots, \ell\}\) satisfying \(i \neq j\). Applying this to \(j = \ell\), we conclude that \(m_i \perp m_{\ell}\)
for any \(i \in \{1, 2, \ldots, \ell\}\) satisfying \(i \neq \ell\). In other words, \(m_i \perp m_{\ell}\) for any \(i \in \{1, 2, \ldots, \ell - 1\}\) (since
the numbers \(i \in \{1, 2, \ldots, \ell\}\) satisfying \(i \neq \ell\) are precisely the numbers \(i \in \{1, 2, \ldots, \ell - 1\}\)).
In other words, each \(i \in \{1, 2, \ldots, \ell - 1\}\) satisfies \(m_i \perp m_{\ell}\). Hence, Exercise \(2.10.2\) (applied to
\(c = m_{\ell}, k = \ell - 1\) and \(a_i = m_i\)) yields that \(m_1 m_2 \cdots m_{\ell-1} \perp m_{\ell}\). This rewrites as \(m \perp m_{\ell}\) (since
\(m = m_1 m_2 \cdots m_{\ell-1}\)). In other words, the integers \(m\) and \(m_{\ell}\) are coprime.
Consider this \( x \), and denote it by \( v \). Thus, \( v \) is an integer such that

\[
(v \equiv u \mod m \text{ and } v \equiv a_i \mod m_i).
\]

Now, let \( i \in \{1, 2, \ldots, \ell - 1\} \). Then,

\[
m = m_1m_2 \cdots m_{\ell-1} = m_i \cdot (m_1m_2 \cdots m_{i-1}m_{i+1}m_{i+2} \cdots m_{\ell-1});
\]

thus, \( m_i \mid m \) (since \( m_1m_2 \cdots m_{i-1}m_{i+1}m_{i+2} \cdots m_{\ell-1} \) is an integer). But as we just have shown, we have \( v \equiv u \mod m \). Hence, Proposition 2.3.4(e) (applied to \( v, u, m \) and \( m_i \) instead of \( a, b, n \) and \( m \)) yields \( v \equiv u \mod m_i \) (since \( m_i \mid m \)). Hence,

\[
v \equiv u \equiv a_i \mod m_i \quad \text{(by (33))}.
\]

Now, forget that we fixed \( i \). We thus have proven the congruence \( v \equiv a_i \mod m_i \) for each \( i \in \{1, 2, \ldots, \ell - 1\} \). But this congruence also holds for \( i = \ell \) (since \( v \equiv a_\ell \mod m_\ell \)). Hence, this congruence holds for all \( i \in \{1, 2, \ldots, \ell\} \). In other words, we have

\[
v \equiv a_i \mod m_i \text{ for all } i \in \{1, 2, \ldots, \ell\}.
\]

Thus, there exists an integer \( x \) such that

\[
(x \equiv a_i \mod m_i \text{ for all } i \in \{1, 2, \ldots, \ell\})
\]

(namely, \( x = v \)). This proves Claim 2.]

We have now proven Claim 2. In other words, Theorem 2.12.4(a) is true for \( k = \ell \). Thus, the induction step is complete, so we have proven Theorem 2.12.4(a) by induction.

(b) Let \( k \) and \( m_1, m_2, \ldots, m_k \) and \( a_1, a_2, \ldots, a_k \) be as in Theorem 2.12.4. Let \( x_1 \) and \( x_2 \) be two integers \( x \) such that (29). We must prove that \( x_1 \equiv x_2 \mod m_1m_2 \cdots m_k \).

We know that \( x_1 \) is an integer \( x \) such that (29). In other words, \( x_1 \) is an integer and has the property that

\[
(x_1 \equiv a_i \mod m_i \text{ for all } i \in \{1, 2, \ldots, k\}). \quad (34)
\]

Now, let \( i \in \{1, 2, \ldots, k\} \). Then, (34) yields \( x_1 \equiv a_i \mod m_i \). Similarly, \( x_2 \equiv a_i \mod m_i \). Hence, \( x_1 \equiv a_i \equiv x_2 \mod m_i \). In other words, \( m_i \mid x_1 - x_2 \).

Now, forget that we fixed \( i \). We thus have shown that \( m_i \mid x_1 - x_2 \) for each \( i \in \{1, 2, \ldots, k\} \). Hence, Exercise 2.10.3(a) (applied to \( c = x_1 - x_2 \) and \( b_i = m_i \)) shows that \( m_1m_2 \cdots m_k \mid x_1 - x_2 \) (since \( m_1, m_2, \ldots, m_k \) are mutually coprime). In other words, \( x_1 \equiv x_2 \mod m_1m_2 \cdots m_k \). This proves Theorem 2.12.4(b). \( \square \)

2.13. Primes

2.13.1. Definition and the Sieve of Eratosthenes
**Definition 2.13.1.** Let $p$ be an integer greater than 1. We say that $p$ is prime if the only positive divisors of $p$ are 1 and $p$. A prime integer is often just called a prime.

Note that we required $p$ to be greater than 1 here. Thus, 1 does not count as prime even though its only positive divisor is 1 itself.

**Example 2.13.2.** (a) The only positive divisors of 7 are 1 and 7. Thus, 7 is a prime.
(b) The positive divisors of 14 are 1, 2, 7 and 14. These are more than just 1 and 14. Thus, 14 is not a prime.
(c) None of the numbers 4, 6, 8, 10, 12, 14, 16, … (that is, the multiples of 2 that are larger than 2) is a prime. Indeed, if $p$ is any of the numbers, then $p$ has a positive divisor other than 1 and $p$ (namely, 2), and therefore does not meet the definition of “prime”.
(d) None of the numbers 6, 9, 12, 15, 18, … (that is, the multiples of 3 that are larger than 3) is a prime. Indeed, if $p$ is any of the numbers, then $p$ has a positive divisor other than 1 and $p$ (namely, 3), and therefore does not meet the definition of “prime”.

Parts (c) and (d) of Example 2.13.2 suggest a method for finding all primes up to a given integer:

**Example 2.13.3.** Let us say we want to find all primes that are $\leq 30$.

**Step 1:** All such primes must lie in $\{2, 3, \ldots, 30\}$ (since a prime is always an integer greater than 1); thus, let us first write down all elements of $\{2, 3, \ldots, 30\}$:

\[
\begin{array}{cccccccc}
2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 \\
\end{array}
\]

(We are using a table just in order to fit these elements on a page.)

We now plan to remove non-prime numbers from this table until only primes are left.

**Step 2:** First, let us remove all multiples of 2 that are larger than 2 from our table, because none of them is a prime (see Example 2.13.2 (c)). We thus are left with

\[
\begin{array}{cccc}
2 & 3 & 5 & 7 \\
11 & 13 & 15 & 17 & 19 \\
21 & 23 & 25 & 27 & 29
\end{array}
\]

**Step 3:** Next, let us remove all multiples of 3 that are larger than 3 from our table, because none of them is a prime (see Example 2.13.2 (d)). We thus are left with

\[
\begin{array}{ccc}
2 & 3 & 5 & 7 \\
11 & 13 & 17 & 19 \\
23 & 25 & 29
\end{array}
\]
Step 4: Next, let us remove all multiples of 4 that are larger than 4 from our table, because none of them is a prime (for similar reasons). It turns out that this does not change the table at all, because all such multiples have already been removed in Step 2. This is not a coincidence: Since 4 itself has been removed, we know that 4 was a multiple of some number \(d < 4\) (in this case, \(d = 2\)) whose multiples have been removed; therefore, all multiples of 4 are also multiples of \(d\) and thus have been removed along with 4.

Step 5: Next, let us remove all multiples of 5 that are larger than 5 from our table, because none of them is a prime (for similar reasons). We thus are left with

\[
\begin{array}{cccc}
2 & 3 & 5 & 7 \\
11 & 13 & 17 & 19 \\
23 & 29
\end{array}
\]

Step 6: Next, let us remove all multiples of 6 that are larger than 6 from our table, because none of them is a prime. Just as Step 4, this does not change the table, since all such multiples have already been removed in Step 2.

Step 7: Next, let us remove all multiples of 7 that are larger than 7 from our table, because none of them is a prime. Again, this does not change the table, since all such multiples have already been removed.

Proceed likewise until Step 30, at which point the table has become

\[
\begin{array}{cccc}
2 & 3 & 5 & 7 \\
11 & 13 & 17 & 19 \\
23 & 29
\end{array}
\]

(You are reading it right: None of the steps from Step 6 to Step 30 causes any changes to the table, since all multiples that these steps attempt to remove have already been removed beforehand.)

The resulting table has the following property: If \(p\) is an element of this table, then \(p\) cannot be a multiple of any \(d \in \{2, 3, \ldots, p - 1\}\) (because if it was such a multiple, then it would have been removed from the table in Step \(d\) or earlier). In other words, if \(p\) is an element of this table, then \(p\) cannot have any divisor \(d \in \{2, 3, \ldots, p - 1\}\). In other words, if \(p\) is an element of this table, then the only positive divisors of \(p\) are 1 and \(p\). In other words, if \(p\) is an element of this table, then \(p\) is prime. Conversely, any prime \(\leq 30\) is in our table, since the only numbers we have removed from the table were guaranteed to be non-prime. Thus, the table now contains all the primes \(\leq 30\) and only them. So we conclude that the primes \(\leq 30\) are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29.

This method of finding primes is known as the sieve of Eratosthenes. We could have made it more efficient using the following two tricks:

- If a number \(d \in \{2, 3, \ldots, 30\}\) has been removed from the table before Step \(d\), then we know immediately that Step \(d\) will not change the table (because
all multiples of $d$ have already been removed before this step). Thus, we
do not need to make this step.

- If $d \in \{2, 3, \ldots, 30\}$ satisfies $d^2 > 30$, then Step $d$ will not change the table
  (because if $m \in \{2, 3, \ldots, 30\}$ is a multiple of $d$ that is larger than $d$, then $m$
is also a multiple of the integer $m/d$ as well (since $d | m$ and thus $m/d \in \mathbb{Z}$
and of course $m/d | m$), and therefore $m$ has already been removed in Step
$m/d$ (which has already happened before Step $d$ (because $d^2 > 30 \geq m$ and
therefore $d > m/d$)). Thus, we only need to take the Steps
$d$ with $d^2 \leq 30$.

Together, these tricks tell us that the only steps we need to take are the Steps
2, 3 and 5.

2019-02-11 lecture

2.13.2. Basic properties of primes

**Proposition 2.13.4.** Let $p$ be a prime. Then, each $i \in \{1, 2, \ldots, p - 1\}$ is coprime
to $p$.

**Proof of Proposition 2.13.4.** Let $i \in \{1, 2, \ldots, p - 1\}$. We must prove that $i$ is coprime
to $p$.

From $i \in \{1, 2, \ldots, p - 1\}$, we obtain $1 \leq i \leq p - 1$ and thus $i \geq 1 > 0$, so that
$i \neq 0$. Hence, $i$ and $p$ are not all zero. Also, $|i| = i$ (since $i > 0$).

Also, $\text{gcd}(i, p)$ is a positive integer (since $i$ and $p$ are not all zero). Thus,
$|\text{gcd}(i, p)| = \text{gcd}(i, p)$.

Proposition 2.9.7 (f) (applied to $a = i$ and $b = p$) shows that $\text{gcd}(i, p) | i$ and
$\text{gcd}(i, p) | p$. From $\text{gcd}(i, p) | i$ and $i \neq 0$, we obtain $|\text{gcd}(i, p)| \leq |i|$ (by Exercise
2.2.3 (b), applied to $a = \text{gcd}(i, p)$ and $b = i$). In view of $|\text{gcd}(i, p)| = \text{gcd}(i, p)$
and $|i| = i$, this rewrites as $\text{gcd}(i, p) \leq i$. Hence, $\text{gcd}(i, p) \leq i \leq p - 1 < p$ and
therefore $\text{gcd}(i, p) \neq p$.

We know that $p$ is prime. In other words, the only positive divisors of $p$ are 1
and $p$ (by the definition of “prime”).

The integer $\text{gcd}(i, p)$ is a positive divisor of $p$ (since $\text{gcd}(i, p)$ is positive and
satisfies $\text{gcd}(i, p) | p$), and thus must be either 1 or $p$ (since the only positive
divisors of $p$ are 1 and $p$). Since we know that $\text{gcd}(i, p) \neq p$, we thus conclude that
$\text{gcd}(i, p) = 1$. In other words, $i$ is coprime to $p$ (by the definition of “coprime”).
This proves Proposition 2.13.4.

Note that this proposition characterizes primes: If $p > 1$ is an integer such that
each $i \in \{1, 2, \ldots, p - 1\}$ is coprime to $p$, then $p$ is prime. (The proof of this is left
as an easy exercise.)
Proposition 2.13.5. Let $p$ be a prime. Let $a \in \mathbb{Z}$. Then, either $p \mid a$ or $p \perp a$.

Proof of Proposition 2.13.5. Assume the contrary. Thus, neither $p \mid a$ nor $p \perp a$.

We know that $p$ is prime. In other words, $p$ is an integer greater than 1 such that the only positive divisors of $p$ are 1 and $p$ (by the definition of “prime”).

In particular, $p$ is greater than 1. Hence, $p > 1 > 0$, so that $p \neq 0$. Hence, $a$ and $p$ are not all zero. Thus, gcd $(a, p)$ is a positive integer.

Proposition 2.9.7 (applied to $b = p$) shows that gcd $(a, p) \mid a$ and gcd $(a, p) \mid p$. If we had gcd $(a, p) = p$, then we would obtain $p = \gcd (a, p) \mid a$, which would contradict the fact that we do not have $p \mid a$. Hence, we cannot have gcd $(a, p) = p$.

In other words, we have gcd $(a, p) \neq p$.

The integer gcd $(a, p)$ is a positive divisor of $p$ (since gcd $(a, p)$ is positive and satisfies gcd $(a, p) \mid p$), and thus must be either 1 or $p$ (since the only positive divisors of $p$ are 1 and $p$). Since we know that gcd $(a, p) \neq p$, we thus conclude that gcd $(a, p) = 1$. But Proposition 2.9.7 (applied to $b = p$) yields gcd $(a, p) = \gcd (p, a)$. Thus, gcd $(p, a) = \gcd (a, p) = 1$. In other words, $p$ is coprime to $a$ (by the definition of “coprime”). In other words, $p \perp a$. This contradicts the fact that we don’t have $p \perp a$.

This contradiction shows that our assumption was false. Hence, Proposition 2.13.5 is proven.

We note that a converse of Proposition 2.13.5 holds as well: If $p > 1$ is an integer such that each $a \in \mathbb{Z}$ satisfies either $p \mid a$ or $p \perp a$, then $p$ is a prime. This is easy to prove and left to the reader.

Exercise 2.13.1. Let $p$ and $q$ be two distinct primes. Prove that $p \perp q$.

Theorem 2.13.6. Let $p$ be a prime. Let $a, b \in \mathbb{Z}$ such that $p \mid ab$. Then, $p \mid a$ or $p \mid b$.

Proof of Theorem 2.13.6. Assume the contrary. Thus, neither $p \mid a$ nor $p \mid b$.

Proposition 2.13.5 yields that either $p \mid a$ or $p \perp a$. Hence, $p \perp a$ (since $p \mid a$ does not hold). But $p \mid ab$. Hence, Theorem 2.10.6 (applied to $p, a$ and $b$ instead of $a, b$ and $c$) yields $p \mid b$. This contradicts the fact that we don’t have $p \mid b$.

This contradiction shows that our assumption was false. Hence, Theorem 2.13.6 is proven.

Again, Theorem 2.13.6 has a converse:

Exercise 2.13.2. Let $p > 1$ be an integer. Assume that for every $a, b \in \mathbb{Z}$ satisfying $p \mid ab$, we must have $p \mid a$ or $p \mid b$. Prove that $p$ is prime.

There is also a version of Theorem 2.13.6 for products of multiple integers:
Proposition 2.13.7. Let \( p \) be a prime. Let \( a_1, a_2, \ldots, a_k \) be integers such that \( p \mid a_1 a_2 \cdots a_k \). Then, \( p \mid a_i \) for some \( i \in \{1, 2, \ldots, k\} \).

We could prove Proposition 2.13.7 by induction on \( k \). But here is a more direct argument:

Proof of Proposition 2.13.7. Assume the contrary. Thus, there exists no \( i \in \{1, 2, \ldots, k\} \) such that \( p \mid a_i \). In other words, for each \( i \in \{1, 2, \ldots, k\} \), we have

\[
\text{not } p \mid a_i.
\]  

(35)

Now, let \( i \in \{1, 2, \ldots, k\} \). Then, we don’t have \( p \mid a_i \) (by (35)). But Proposition 2.13.5 (applied to \( a = a_i \)) shows that either \( p \mid a_i \) or \( p \perp a_i \). Hence, we have \( p \perp a_i \) (since we don’t have \( p \mid a_i \)). In other words, \( a_i \perp p \) (by Proposition 2.10.4).

Now, forget that we fixed \( i \). We thus have proven that each \( i \in \{1, 2, \ldots, k\} \) satisfies \( a_i \perp p \). Hence, Exercise 2.10.2 (applied to \( c = p \)) yields \( a_1 a_2 \cdots a_k \perp p \). In other words, \( \gcd(a_1 a_2 \cdots a_k, p) = 1 \). Hence, Proposition 2.9.7 (b) yields \( \gcd(p, a_1 a_2 \cdots a_k) = \gcd(a_1 a_2 \cdots a_k, p) = 1 \).

But \( p \) is prime; thus, \( p > 1 \). Hence, \( p \) is positive. Recall that \( p \mid a_1 a_2 \cdots a_k \); thus, Proposition 2.9.7 (i) (applied to \( a = p \) and \( b = a_1 a_2 \cdots a_k \)) yields \( \gcd(p, a_1 a_2 \cdots a_k) = \gcd(p, a_1 a_2 \cdots a_k) = |p| = p \) (since \( p \) is positive). Comparing this with \( \gcd(p, a_1 a_2 \cdots a_k) = 1 \), we obtain \( p = 1 \). This contradicts \( p > 1 \). This contradiction shows that our assumption was wrong. This proves Proposition 2.13.7.

Exercise 2.13.3. Let \( p \) be a prime. Let \( k \) be a positive integer. Let \( a \in \mathbb{Z} \). Prove that \( a \perp p^k \) holds if and only if \( p \nmid a \).

2.13.3. Prime factorization I

The next simple proposition says that every integer \( n > 1 \) is divisible by at least one prime:

Proposition 2.13.8. Let \( n > 1 \) be an integer. Then, there exists at least one prime \( p \) such that \( p \mid n \).

Proof of Proposition 2.13.8. Clearly, \( n \) is a divisor of \( n \) such that \( n > 1 \). Thus, there exists a divisor \( q \) of \( n \) such that \( q > 1 \) (namely, \( q = n \)). Let \( d \) be the smallest such divisor.\(^{28}\) Thus, \( d \) is a divisor of \( n \) and satisfies \( d > 1 \). The integer \( d \) is positive (since \( d > 1 > 0 \)) and satisfies \( d \mid n \) (since \( d \) is a divisor of \( n \)).

We claim that \( d \) is a prime.

[Proof: Let \( e \) be any positive divisor of \( d \). Assume (for the sake of contradiction) that \( e \notin \{1, d\} \). Thus, \( e \neq 1 \) and \( e \neq d \). Now, \( e \) is a divisor of \( d \); thus, \( e \mid d \mid n \). In other words, \( e \) is a divisor of \( n \). Also, \( e > 1 \) (because \( e \) is positive and \( e \neq 1 \)). Hence, \( e \) is a divisor \( q \) of \( n \) such that \( q > 1 \).

\(^{28}\)This exists, because the set of possible candidates is nonempty (by the previous sentence) and finite.
But \( d \) was defined as the **smallest** divisor \( q \) of \( n \) such that \( q > 1 \). Hence, any such divisor is \( \geq d \). In other words, any divisor \( q \) of \( n \) such that \( q > 1 \) must satisfy \( q \geq d \). Applying this to \( q = e \), we conclude that \( e \geq d \) (since \( e \) is a divisor \( q \) of \( n \) such that \( q > 1 \)). Combined with \( e \neq d \), this yields \( e > d \).

But \( e \mid d \) and \( d \neq 0 \) (since \( d > 1 > 0 \)). Hence, \( |e| \leq |d| \) (by Exercise 2.2.3(b), applied to \( a = e \) and \( b = d \)). Since \( e \) is positive, we have \( |e| = e \), so that \( e = |e| \leq |d| = d \) (since \( d \) is positive). This contradicts \( e > d \). This contradiction shows that our assumption (that \( e \notin \{1,d\} \)) was false. Thus, we have proven that \( e \in \{1,d\} \). In other words, \( e \) is either 1 or \( d \).

Now, forget that we fixed \( e \). We thus have proven that if \( e \) is any positive divisor of \( d \), then \( e \in \{1,d\} \). In other words, any positive divisor of \( d \) is either 1 or \( d \). Thus, the only positive divisors of \( d \) are 1 and \( d \) (since 1 and \( d \) clearly are positive divisors of \( d \)). In other words, \( d \) is prime (by the definition of “prime”).]

So we know that \( d \mid n \), and that \( d \) is prime. Hence, there exists at least one prime \( p \) such that \( p \mid n \) (namely, \( p = d \)). This proves Proposition 2.13.8.

---

**Definition 2.13.9.** Let \( n \) be an integer. A **prime factor** of \( n \) means a prime \( p \) such that \( p \mid n \). Some say “prime divisor” instead of “prime factor”.

Thus, Proposition 2.13.8 says that each integer \( n > 1 \) has at least one prime divisor.

**Proposition 2.13.10.** Let \( n \) be a positive integer. Then, \( n \) can be written as a product of finitely many primes.

**Example 2.13.11.** (a) The integer 60 can be written as a product of four primes: namely, \( 60 = 2 \cdot 2 \cdot 3 \cdot 5 \).

(b) The integer 1 is the product of 0 many primes (because a product of 0 many primes is the empty product, which is defined to be 1).

**Proof of Proposition 2.13.10.** We shall prove Proposition 2.13.10 by strong induction on \( n \). Thus, we fix a positive integer \( N \), and we assume (as the induction hypothesis) that Proposition 2.13.10 holds whenever \( n < N \). We must now prove that Proposition 2.13.10 holds for \( n = N \). In other words, we must prove that \( N \) can be written as a product of finitely many primes.

If \( N = 1 \), then this is obvious (because 1 is a product of 0 many primes\(^\text{29}\)). Thus, for the rest of this proof, we WLOG assume that \( N \neq 1 \). Hence, \( N > 1 \) (since \( N \) is a positive integer). Therefore, Proposition 2.13.8 (applied to \( n = N \)) shows that there exists at least one prime \( p \) such that \( p \mid N \). Consider this \( p \).

We have \( p \mid N \). In other words, there exists an integer \( c \) such that \( N = pc \). Consider this \( c \). We have \( p > 1 \) (since \( p \) is prime); thus, \( p \) is positive. Hence, \( p \neq 0 \). Thus, solving the equality \( N = pc \) for \( c \), we find \( c = N/p < N/1 \) (since \( N > 1 \)).

---

\(^{29}\)See Example 2.13.11 (b).
positive), so that \( c < N/1 = N \). But our induction hypothesis says that Proposition 2.13.10 holds whenever \( n < N \). Hence, we can apply Proposition 2.13.10 to \( n = c \) (since \( c < N \)). We thus conclude that \( c \) can be written as a product of finitely many primes. In other words, there exist primes \( q_1, q_2, \ldots, q_k \) such that \( c = q_1 q_2 \cdots q_k \). Consider these \( q_1, q_2, \ldots, q_k \).

But

\[
N = p \underbrace{c}_{=q_1 q_2 \cdots q_k} = pq_1 q_2 \cdots q_k.
\]

Hence, \( N \) can be written as a product of finitely many primes (namely, of the primes \( p, q_1, q_2, \ldots, q_k \)). In other words, Proposition 2.13.10 holds for \( n = N \). This completes the induction step. Hence, Proposition 2.13.10 is proven by strong induction. \( \square \)

Proposition 2.13.10 shows that every positive integer \( n \) can be represented as a product of finitely many primes. Such a representation – or, more precisely, the list of the primes it contains – will be called the prime factorization of \( n \). Rigorously speaking, this means that we make the following definition:

**Definition 2.13.12.** Let \( n \) be a positive integer. A prime factorization of \( n \) means a tuple \((p_1, p_2, \ldots, p_k)\) of primes such that \( n = p_1 p_2 \cdots p_k \).

Keep in mind that “tuple” always means “ordered tuple” unless we say otherwise.

**Example 2.13.13.** (a) The prime factorizations of 12 are

\[
(2, 2, 3), \quad (2, 3, 2), \quad (3, 2, 2).
\]

Indeed, these three 3-tuples are prime factorizations of 12 because \( 12 = 2 \cdot 2 \cdot 3 = 2 \cdot 3 \cdot 2 = 3 \cdot 2 \cdot 2 \). It is not hard to check that they are the only prime factorizations of 12.

(b) If \( p \) is a prime, then the only prime factorization of \( p \) is the 1-tuple \((p)\).

(c) If \( p \) is a prime and \( i \in \mathbb{N} \), then the only prime factorization of \( p^i \) is the \( i \)-tuple \((p, p, \ldots, p)\). This is not quite obvious at this point (though it is not hard to derive from Proposition 2.13.7).

(d) The only prime factorization of 1 is the 0-tuple ( ).

This example suggests that all prime factorizations of a given positive integer \( n \) are equal to each other up to the order of their entries (i.e., are permutations of each other). This is indeed true, and we are going to prove this soon (in Theorem 2.13.31 below).
2.13.4. Permutations

First of all: what is a “permutation”, and what exactly does “equal to each other up to the order of their entries” mean?

Informally speaking, a permutation of a tuple \((a_1, a_2, \ldots, a_k)\) is a tuple obtained from \((a_1, a_2, \ldots, a_k)\) by rearranging its entries (without inserting new entries, or removing or duplicating existing entries). To be rigorous, we need to encode this rearrangement via a bijective map \(\sigma: \{1, 2, \ldots, k\} \rightarrow \{1, 2, \ldots, k\}\) which will tell us which entry of our original tuple will go to which position in the rearranged tuple. Such bijective maps, too, are called permutations – but permutations of sets, not of tuples. So let us first define permutations of a set, and then use this to define permutations of a tuple:

**Definition 2.13.14.** Let \(A\) be a set. A permutation of \(A\) means a bijective map \(A \rightarrow A\).

**Example 2.13.15.** (a) The map \(\{1, 2, 3, 4\} \rightarrow \{1, 2, 3, 4\}\) that sends 1, 2, 3, 4 to 3, 1, 4, 2 (respectively) is a permutation of \(\{1, 2, 3, 4\}\).

(b) The map \(\{1, 2, 3\} \rightarrow \{1, 2, 3\}\) that sends 1, 2, 3 to 2, 3, 1 (respectively) is a permutation of \(\{1, 2, 3\}\).

(c) For each set \(A\), the identity map \(\text{id}: A \rightarrow A\) is a permutation of \(A\).

Thus, we have defined permutations of a set. We shall later study such permutations in more detail, at least for finite sets \(A\).

Now we can define permutations of a tuple:

**Definition 2.13.16.** Let \((p_1, p_2, \ldots, p_k)\) be a \(k\)-tuple. A permutation of \((p_1, p_2, \ldots, p_k)\) means a \(k\)-tuple of the form \((p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(k)})\) where \(\sigma\) is a permutation of the set \(\{1, 2, \ldots, k\}\). A permutation of \((p_1, p_2, \ldots, p_k)\) is also known as a rearrangement of \((p_1, p_2, \ldots, p_k)\).

**Example 2.13.17.** (a) The 4-tuple \((1, 3, 1, 2)\) is a permutation of the 4-tuple \((3, 2, 1, 1)\). In fact, if we denote the 4-tuple \((3, 2, 1, 1)\) by \((p_1, p_2, p_3, p_4)\), then there exists a permutation \(\sigma\) of the set \(\{1, 2, 3, 4\}\) such that \((1, 3, 1, 2) = \left(p_{\sigma(1)}, p_{\sigma(2)}, p_{\sigma(3)}, p_{\sigma(4)}\right)\). (Actually, there exist two such permutations \(\sigma\): One of them sends 1, 2, 3, 4 to 3, 1, 4, 2, while the other sends 1, 2, 3, 4 to 4, 1, 3, 2.)

(b) Any \(k\)-tuple is a permutation of itself. Indeed, if \((p_1, p_2, \ldots, p_k)\) is any \(k\)-tuple, then \((p_1, p_2, \ldots, p_k) = \left(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(k)}\right)\) if we let \(\sigma\) be the identity map \(\text{id}: \{1, 2, \ldots, k\} \rightarrow \{1, 2, \ldots, k\}\).

The following fact is easy and fundamental:

---

\(^{30}\)Recall: a prime factorization is a tuple.
Proposition 2.13.18. Let \((p_1, p_2, \ldots, p_k)\) be a \(k\)-tuple. If \((q_1, q_2, \ldots, q_k)\) is a permutation of \((p_1, p_2, \ldots, p_k)\), then \((p_1, p_2, \ldots, p_k)\) is a permutation of \((q_1, q_2, \ldots, q_k)\).

Proof of Proposition 2.13.18. If you don’t insist on formalization, this is obvious: Any rearrangement of the entries of a \(k\)-tuple can be undone by another rearrangement (which places the entries back in their old positions). Thus, \((p_1, p_2, \ldots, p_k)\) can be obtained from \((q_1, q_2, \ldots, q_k)\) by rearranging the entries.

Here is a formal proof:

Assume that \((q_1, q_2, \ldots, q_k)\) is a permutation of \((p_1, p_2, \ldots, p_k)\). In other words, the \(k\)-tuple \((q_1, q_2, \ldots, q_k)\) has the form \(p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(k)}\) for some permutation \(\sigma\) of the set \(\{1, 2, \ldots, k\}\) (by Definition 2.13.16). Consider this \(\sigma\), and denote it by \(\tau\). Thus, \(\tau\) is a permutation of the set \(\{1, 2, \ldots, k\}\) and has the property that \((q_1, q_2, \ldots, q_k) = (p_{\tau(1)}, p_{\tau(2)}, \ldots, p_{\tau(k)})\).

Now, \(\tau\) is a permutation of the set \(\{1, 2, \ldots, k\}\). In other words, \(\tau\) is a bijective map \(\{1, 2, \ldots, k\} \rightarrow \{1, 2, \ldots, k\}\) (by Definition 2.13.14). So the map \(\tau\) is bijective, hence invertible. Thus, its inverse \(\tau^{-1}\) is well-defined and is also invertible \(^{31}\), hence bijective. So we know that \(\tau^{-1}\) is a bijective map \(\{1, 2, \ldots, k\} \rightarrow \{1, 2, \ldots, k\}\). In other words, \(\tau^{-1}\) is a permutation of the set \(\{1, 2, \ldots, k\}\) (by Definition 2.13.14).

We have \((q_1, q_2, \ldots, q_k) = (p_{\tau(1)}, p_{\tau(2)}, \ldots, p_{\tau(k)})\). In other words,

\[
q_i = p_{\tau(i)} \quad \text{for each } i \in \{1, 2, \ldots, k\}.
\]

Hence, for each \(j \in \{1, 2, \ldots, k\}\), we have

\[
q_{\tau^{-1}(j)} = p_{\tau(\tau^{-1}(j))} \quad \text{(by }\text{36}, \text{ applied to } i = \tau^{-1}(j)\text{)}
= p_j \quad \text{(since } \tau \left( \tau^{-1}(j) \right) = j\text{)}.
\]

In other words, \((q_{\tau^{-1}(1)}, q_{\tau^{-1}(2)}, \ldots, q_{\tau^{-1}(k)}) = (p_1, p_2, \ldots, p_k)\). Hence, the \(k\)-tuple \((p_1, p_2, \ldots, p_k)\) has the form \((q_{\sigma(1)}, q_{\sigma(2)}, \ldots, q_{\sigma(k)})\) for some permutation \(\sigma\) of the set \(\{1, 2, \ldots, k\}\) (namely, \(\sigma = \tau^{-1}\)). In other words, the \(k\)-tuple \((p_1, p_2, \ldots, p_k)\) is a permutation of the \(k\)-tuple \((q_1, q_2, \ldots, q_k)\) (by Definition 2.13.16). This proves Proposition 2.13.18.

Now, we can say what we mean when we say that two tuples differ only in the order of their entries:

Definition 2.13.19. We say that two tuples differ only in the order of their entries if they are permutations of each other.

The next lemma that we shall use is a basic fact from elementary combinatorics:

\(^{31}\)And its inverse is \((\tau^{-1})^{-1} = \tau\).
Lemma 2.13.20. Let $P$ be a set. Let $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_\ell)$ be two tuples of elements of $P$. Assume that for each $p \in P$, we have

$$\text{(the number of times } p \text{ appears in } (a_1, a_2, \ldots, a_k)) = \text{(the number of times } p \text{ appears in } (b_1, b_2, \ldots, b_\ell)).$$

(37)

Then, the two tuples $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_\ell)$ differ only in the order of their entries (i.e., are permutations of each other). (In other words, we have $k = \ell$, and there exists a permutation $\sigma$ of the set $\{1, 2, \ldots, \ell\}$ such that $(a_1, a_2, \ldots, a_k) = (b_{\sigma(1)}, b_{\sigma(2)}, \ldots, b_{\sigma(\ell)})$.)

Lemma 2.13.20 is an intuitively obvious fact: It says that if two tuples (of any objects – e.g., numbers) have the property that any object occurs as often in the first tuple as it does in the second tuple, then the two tuples differ only in the order of their entries. From the formal point of view, though, it is a statement that needs proof. Let us merely sketch how such a proof can be obtained, without going into the details:

Proof of Lemma 2.13.20 (sketched). We can WLOG assume that the set $P$ is finite (since otherwise, we can replace $P$ by the finite subset $\{a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_1\}$, without breaking the assumption that $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_\ell)$ are two tuples of elements of $P$). Assume this (at least if you don’t want to use the Axiom of Choice$^{32}$).

For each $p \in P$, define two sets

$$A_p = \{i \in \{1, 2, \ldots, k\} \mid a_i = p\};$$

$$B_p = \{j \in \{1, 2, \ldots, \ell\} \mid b_j = p\}.$$

The equation (37) then says that $|A_p| = |B_p|$ for each $p \in P$. Hence, for each $p \in P$, there exists a bijection $\phi_p : A_p \rightarrow B_p$ (because if two sets have the same size, then there exists a bijection between them). Pick such a bijection $\phi_p$ for each $p \in P$. (This does not require the Axiom of Choice, since $P$ is finite.)

Now, define a map $\sigma : \{1, 2, \ldots, k\} \rightarrow \{1, 2, \ldots, \ell\}$ as follows: For each $i \in \{1, 2, \ldots, k\}$, set $\sigma(i) = \phi_p(i)$, where $p = a_i$. Thus, for each $p \in P$, the map $\sigma$ sends each $i \in A_p$ to an element of $B_p$ (because if $i \in A_p$, then $a_i = p$, and thus the definition of $\sigma$ yields $\sigma(i) = \phi_p(i) \in B_p$).

It is not hard to see that this map $\sigma$ is a bijection. (Its inverse map sends each $j \in \{1, 2, \ldots, \ell\}$ to $\phi_p^{-1}(j)$, where $p = b_j$.) Thus, we have found a bijection from $\{1, 2, \ldots, k\}$ to $\{1, 2, \ldots, \ell\}$. This shows that the sets $\{1, 2, \ldots, k\}$ and $\{1, 2, \ldots, \ell\}$ have the same size; in other words, $k = \ell$. Thus, the bijection $\sigma$ is actually a bijection from $\{1, 2, \ldots, \ell\}$ to $\{1, 2, \ldots, \ell\}$. In other words, $\sigma$ is a permutation of the set $\{1, 2, \ldots, \ell\}$.

Finally, it is easy to see that $(a_1, a_2, \ldots, a_k) = (b_{\sigma(1)}, b_{\sigma(2)}, \ldots, b_{\sigma(\ell)})$. (Indeed, let $i \in \{1, 2, \ldots, k\}$, and set $p = a_i$; then, the definition of $\sigma$ yields $\sigma(i) = \phi_p(i) \in B_p$ and therefore $b_{\sigma(i)} = a_i$. Since this holds for each $i$, we thus conclude that $(b_{\sigma(1)}, b_{\sigma(2)}, \ldots, b_{\sigma(\ell)}) = (a_1, a_2, \ldots, a_k)$.)

$^{32}$I don’t.
(a_1, a_2, \ldots, a_k). Thus, (a_1, a_2, \ldots, a_k) = (b_{v(1)}, b_{v(2)}, \ldots, b_{v(k)}) = (b_{v(1)}, b_{v(2)}, \ldots, b_{v(\ell)}) \ (\text{since } k = \ell).) Thus, we have found a permutation \( \sigma \) of the set \( \{1, 2, \ldots, \ell\} \) such that \( (a_1, a_2, \ldots, a_k) = (b_{v(1)}, b_{v(2)}, \ldots, b_{v(\ell)}) \). In other words, the two tuples \((a_1, a_2, \ldots, a_k)\) and \((b_1, b_2, \ldots, b_\ell)\) are permutations of each other. This proves Lemma 2.13.20. 

Lemma 2.13.20 has a converse that is much simpler:

**Lemma 2.13.21.** Let \( P \) be a set. Let \((a_1, a_2, \ldots, a_k)\) and \((b_1, b_2, \ldots, b_\ell)\) be two tuples of elements of \( P \). Assume that these two tuples \((a_1, a_2, \ldots, a_k)\) and \((b_1, b_2, \ldots, b_\ell)\) differ only in the order of their entries (i.e., are permutations of each other). Then, for each \( p \in P \), we have

\[
\begin{align*}
\text{(the number of times } p \text{ appears in } (a_1, a_2, \ldots, a_k)) &= \text{(the number of times } p \text{ appears in } (b_1, b_2, \ldots, b_\ell)).
\end{align*}
\]

We leave the proof of this lemma to the reader.

**2.13.5. \( p \)-valuations**

Now, let us come back to number theory. We first claim that a nonzero integer \( n \) can only be divisible by finitely many powers of a given prime \( p \). More precisely:

**Lemma 2.13.22.** Let \( p \) be a prime. Let \( n \) be a nonzero integer. Then, there exists a largest \( m \in \mathbb{N} \) such that \( p^m \mid n \).

The proof of this lemma will rely on a simple inequality, which we leave as an exercise:

**Exercise 2.13.4.** Let \( p \) be an integer such that \( p > 1 \). Prove that \( p^k > k \) for each \( k \in \mathbb{N} \).

*Proof of Lemma 2.13.22* We know that \( p \) is a prime. Thus, \( p \) is an integer and \( p > 1 \) (by the definition of a “prime”). This is all we shall need from our assumption that \( p \) is prime.

Let \( W \) be the set of all \( m \in \mathbb{N} \) satisfying \( p^m \mid n \). Then, \( W \) is a set of integers. Moreover, \( 0 \) is an \( m \in \mathbb{N} \) satisfying \( p^m \mid n \) (since \( p^0 = 1 \mid n \)); in other words, \( 0 \in W \) (by the definition of \( W \)). Hence, the set \( W \) is nonempty.

Let \( u = |n| \). Thus, \( u \in \mathbb{N} \).

Exercise 2.13.4 yields that \( p^k > k \) for each \( k \in \mathbb{N} \). Thus, each \( g \in W \) satisfies \( g \in \{0, 1, \ldots, u-1\} \). In other words, \( W \subseteq \{0, 1, \ldots, u-1\} \). Hence, the set \( W \)

33\text{Proof:} Let \( g \in W \). Thus, \( g \) is an \( m \in \mathbb{N} \) satisfying \( p^m \mid n \) (by the definition of \( W \)). In other words, \( g \in \mathbb{N} \) and \( p^g \mid n \). Also, \( n \neq 0 \) (since \( n \) is nonzero). Hence, Proposition 2.2.3(b) (applied to \( a = p^g \) and \( b = n \)) yields \( |p^g| \leq |n| = u \). But \( p \) is positive (since \( p > 1 \)); thus, \( p^g \) is positive. Hence, \( |p^g| = p^g \). Thus, \( p^g = |p^g| \leq u \). But recall that \( p^k > k \) for each \( k \in \mathbb{N} \). Applying this to \( k = g \), we find \( p^g > g \). Hence, \( g < p^g \leq u \), so that \( g \in \{0, 1, \ldots, u-1\} \) (since \( g \in \mathbb{N} \)). QED.
is finite (since the set \(\{0, 1, \ldots, u - 1\}\) is finite). Thus, \(W\) is a finite nonempty set of integers. Therefore, the set \(W\) has a largest element. In view of how \(W\) was defined, this can be restated as follows: There exists a largest \(m \in \mathbb{N}\) such that \(p^m | n\). This proves Lemma 2.13.22.

**Definition 2.13.23.** Let \(p\) be a prime.

(a) Let \(n\) be a nonzero integer. Then, \(v_p(n)\) shall denote the largest \(m \in \mathbb{N}\) such that \(p^m | n\). This is well-defined (by Lemma 2.13.22). This nonnegative integer \(v_p(n)\) will be called the \(p\)-valuation (or the \(p\)-adic valuation) of \(n\).

(b) We extend this definition of \(v_p(n)\) to the case of \(n = 0\) as follows: Set \(v_p(0) = \infty\), where \(\infty\) is a new symbol. This symbol \(\infty\) is supposed to model "positive infinity"; in particular, we take it to satisfy the following rules:

- We have \(k + \infty = \infty + k = \infty\) for all integers \(k\).
- We have \(\infty + \infty = \infty\).
- Each integer \(k\) satisfies \(k < \infty\) and \(\infty > k\) (and thus \(k \leq \infty\) and \(\infty \geq k\)).
- No integer \(k\) satisfies \(k \geq \infty\) or \(\infty \leq k\) (or \(k > \infty\) or \(\infty < k\)).
- If \(S\) is a nonempty set of integers, then \(\min (S \cup \{\infty\}) = \min S\) (provided that \(\min S\) exists).
- If \(S\) is any set of integers, then \(\max (S \cup \{\infty\}) = \infty\).

(Note, however, that \(\infty\) is not supposed to be a "first class citizen" of the number system. In particular, \(\infty - \infty\) is not defined. More generally, \(k - \infty\) is never defined, whatever \(k\) is. Indeed, any definition of \(k - \infty\) would break some of the familiar rules of arithmetic. The only operations that we shall subject \(\infty\) to are addition, minimum and maximum.)

Note that the rules for the symbol \(\infty\) yield that

\[
k + \infty = \infty + k = \max \{k, \infty\} = \infty
\]

and

\[
\min \{k, \infty\} = k
\]

for each \(k \in \mathbb{Z} \cup \{\infty\}\). It is not hard to see that basic properties of inequalities (such as "if \(a \leq b\) and \(b \leq c\), then \(a \leq c\")) and of addition (such as "\((a + b) + c = a + (b + c)\") and of the interplay between inequalities and addition (such as "if \(a \leq b\), then \(a + c \leq b + c\") are still valid in \(\mathbb{Z} \cup \{\infty\}\) (that is, they still hold if we plug \(\infty\) for one or more of the variables). However, of course, we cannot "cancel" \(\infty\) from equalities (i.e., we cannot cancel \(\infty\) from \(a + \infty = b + \infty\) to obtain \(a = b\)) or inequalities.
Example 2.13.24. (a) We have \( v_5(50) = 2 \). Indeed, 2 is the largest \( m \in \mathbb{N} \) such that \( 5^m \mid 50 \) (because \( 5^2 = 25 \mid 50 \) but \( 5^3 = 125 \nmid 50 \)).

(b) We have \( v_5(51) = 0 \). Indeed, 0 is the largest \( m \in \mathbb{N} \) such that \( 5^m \mid 51 \) (because \( 5^0 = 1 \mid 51 \) but \( 5^1 = 5 \nmid 51 \)).

c) We have \( v_5(55) = 1 \). Indeed, 1 is the largest \( m \in \mathbb{N} \) such that \( 5^m \mid 55 \) (because \( 5^1 = 5 \mid 55 \) but \( 5^2 = 25 \nmid 55 \)).

d) We have \( v_5(0) = \infty \) (by Definition 2.13.23(b)).

Definition 2.13.23(a) can be restated in the following more intuitive way: Given a prime \( p \) and a nonzero integer \( n \), we let \( v_p(n) \) be the number of times we can divide \( n \) by \( p \) without leaving \( \mathbb{Z} \). Definition 2.13.23(b) is consistent with this picture, because we can clearly divide 0 by \( p \) infinitely often without leaving \( \mathbb{Z} \).

From this point of view, the following lemma should be obvious:

Lemma 2.13.25. Let \( p \) be a prime. Let \( i \in \mathbb{N} \). Let \( n \in \mathbb{Z} \). Then, \( p^i \mid n \) if and only if \( v_p(n) \geq i \).

Proof of Lemma 2.13.25 First, let us notice that \( p^i \mid 0 \). Also, Definition 2.13.23(b) yields \( v_p(0) = \infty \geq i \) (according to our rules for the symbol \( \infty \)). Hence, both statements \( (p^i \mid 0) \) and \( (v_p(0) \geq i) \) hold. Thus, \( p^i \mid 0 \) if and only if \( v_p(0) \geq i \). In other words, Lemma 2.13.25 holds if \( n = 0 \). Thus, for the rest of this proof, we WLOG assume that \( n \neq 0 \). Hence, \( n \) is nonzero. Thus, \( v_p(n) \) is the largest \( m \in \mathbb{N} \) such that \( p^m \mid n \) (by Definition 2.13.23(a)). Hence, \( v_p(n) \) itself is an \( m \in \mathbb{N} \) such that \( p^m \mid n \). In other words, \( v_p(n) \in \mathbb{N} \) and \( p^{v_p(n)} \mid n \).

We must prove that \( p^i \mid n \) if and only if \( v_p(n) \geq i \). Let us prove the “\( \implies \)” and “\( \impliedby \)” directions of this “if and only if” statement separately:

\( \implies \): Assume that \( p^i \mid n \). We must prove that \( v_p(n) \geq i \).

The integer \( i \) is an \( m \in \mathbb{N} \) such that \( p^m \mid n \) (since \( p^i \mid n \)). But \( v_p(n) \) is the largest such \( m \) (by Definition 2.13.23(a)). Hence, \( v_p(n) \geq i \). This proves the “\( \implies \)” direction of Lemma 2.13.25.

\( \impliedby \): Assume that \( v_p(n) \geq i \). We must prove that \( p^i \mid n \).

We have \( v_p(n) \geq i \), thus \( i \leq v_p(n) \). Hence, Exercise 2.2.4 (applied to \( p \), \( i \) and \( v_p(n) \)) instead of \( n \), \( a \) and \( b \) yields \( p^i \mid p^{v_p(n)} \). Thus, \( p^i \mid p^{v_p(n)} \mid n \).

Hence, we have proven \( p^i \mid n \). This proves the “\( \impliedby \)” direction of Lemma 2.13.25.

Corollary 2.13.26. Let \( p \) be a prime. Let \( n \in \mathbb{Z} \). Then, \( v_p(n) = 0 \) if and only if \( p \nmid n \).

Proof of Corollary 2.13.26 \( \implies \): Assume that \( v_p(n) = 0 \). We must prove that \( p \nmid n \).

We don’t have \( v_p(n) \geq 1 \) (since \( v_p(n) = 0 < 1 \)). But Lemma 2.13.25 (applied to \( i = 1 \)) shows that \( p^1 \mid n \) if and only if \( v_p(n) \geq 1 \). Hence, we don’t have \( p^1 \mid n \) (since we don’t have \( v_p(n) \geq 1 \)). In other words, we have \( p^1 \nmid n \). In other words, \( p \nmid n \) (since \( p = p^1 \)). This proves the “\( \implies \)” direction of Corollary 2.13.26.
\[ \quad \quad \leftarrow: \text{Assume that } p \nmid n. \text{ We must prove that } v_p(n) = 0. \]

We don’t have \( p \mid n \) (since \( p \nmid n \)). In other words, we don’t have \( p^1 \mid n \) (since \( p^1 = p \)). But Lemma \ref{lem:prime_factorization} (applied to \( i = 1 \)) shows that \( p^1 \mid n \) if and only if \( v_p(n) \geq 1 \). Hence, we don’t have \( v_p(n) \geq 1 \) (since we don’t have \( p^1 \mid n \)). In other words, \( v_p(n) < 1 \).

If we had \( n = 0 \), then we would have \( p \mid 0 = n \), which would contradict \( p \nmid n \). Hence, we don’t have \( n = 0 \). Thus, \( p \) is nonzero. Hence, Definition \ref{def:valuation} \( (a) \) shows that \( v_p(n) \in \mathbb{N} \). In light of this, we can conclude \( v_p(n) = 0 \) from \( v_p(n) < 1 \). This proves the “\( \leftarrow \)” direction of Corollary \ref{cor:valuation_properties} \( \square \)

Here is another property of \( p \)-valuations that is useful in their study:

**Lemma 2.13.27.** Let \( p \) be a prime. Let \( n \in \mathbb{Z} \) be nonzero. Then:

\( (a) \) There exists a nonzero integer \( u \) such that \( u \perp p \) and \( n = up^{v_p(n)} \).

\( (b) \) If \( i \in \mathbb{N} \) and \( w \in \mathbb{Z} \) are such that \( w \perp p \) and \( n = wp^i \), then \( v_p(n) = i \).

Before we prove this formally, let us show the idea behind this lemma. Recall that, given a prime \( p \) and a nonzero integer \( n \), the number \( v_p(n) \) counts how often we can divide \( n \) by \( p \) without leaving \( \mathbb{Z} \). What happens after we have divided \( n \) by \( p \) this many times? We get a number \( u \) that is still an integer, but is no longer divisible by \( p \), and thus must be coprime to \( p \) (by Proposition \ref{prop:coprime}). This is what Lemma \ref{lem:valuation_prop} \( (a) \) says. Lemma \ref{lem:valuation_prop} \( (b) \) is a converse statement: It says that if we divide \( n \) by \( p \) some number of times (say, \( i \) times) and obtain an integer coprime to \( p \), then \( i \) must be \( v_p(n) \).

**Proof of Lemma 2.13.27**  Definition \ref{def:valuation} \( (a) \) shows that \( v_p(n) \) is the largest \( m \in \mathbb{N} \) such that \( p^m \mid n \). Hence, \( v_p(n) \) itself is an \( m \in \mathbb{N} \) such that \( p^m \mid n \). In other words, \( v_p(n) \in \mathbb{N} \) and \( p^{v_p(n)} \mid n \).

Thus, in particular, \( p^{v_p(n)} \mid n \). In other words, there exists an integer \( c \) such that \( n = p^{v_p(n)}c \). Consider this \( c \). We have \( n = p^{v_p(n)}c = cp^{v_p(n)} \).

Assume (for the sake of contradiction) that \( p \mid c \). Thus, there exists an integer \( d \) such that \( c = pd \). Consider this \( d \). Now,

\[
\begin{align*}
n &= p^{v_p(n)}c \\
&= p^{v_p(n)}pd \\
&= p^{v_p(n)+1}d.
\end{align*}
\]

Hence, \( p^{v_p(n)+1} \mid n \) (since \( d \) is an integer). In other words, \( v_p(n) + 1 \) is an \( m \in \mathbb{N} \) such that \( p^m \mid n \). But we know that \( v_p(n) \) is the \textbf{largest} such \( m \) (by Definition \ref{def:valuation} \( (a) \)). Hence, we conclude that \( v_p(n) \geq v_p(n) + 1 \). But this is clearly absurd. This contradiction shows that our assumption (that \( p \mid c \)) was wrong. Hence, we do not have \( p \mid c \).

But Proposition \ref{prop:coprime} (applied to \( a = c \)) shows that either \( p \mid c \) or \( p \perp c \). Hence, \( p \perp c \) (since we do not have \( p \mid c \)). In other words, \( c \perp p \) (because of Proposition \ref{prop:coprime}).
If we had \( c = 0 \), then we would have \( n = p^{v_p(n)} \frac{c}{0} = 0 \), which would contradict the fact that \( n \) is nonzero. Hence, we cannot have \( c = 0 \). Thus, \( c \) is nonzero.

Now, we know that \( c \) is a nonzero integer satisfying \( c \perp p \) and \( n = c p^{v_p(n)} \). Hence, there exists a nonzero integer \( u \) such that \( u \perp p \) and \( n = up^{v_p(n)} \) (namely, \( u = c \)). This proves Lemma 2.13.27 (a).

(b) Let \( i \in \mathbb{N} \) and \( w \in \mathbb{Z} \) be such that \( w \perp p \) and \( n = wp^i \). We must prove that \( v_p(n) = i \).

From \( w \perp p \), we obtain \( p \perp w \) (by Proposition 2.10.4). In other words, \( \gcd(p, w) = 1 \).

We have \( n = wp^i = p^i w \) and thus \( p^i \mid n \) (since \( w \) is an integer). But Lemma 2.13.25 yields that \( p^i \mid n \) if and only if \( v_p(n) \geq i \). Hence, we have \( v_p(n) \geq i \) (since we have \( p^i \mid n \)).

Now, we shall prove that \( v_p(n) \leq i \). Indeed, assume the contrary. Thus, \( v_p(n) > i \), so that \( v_p(n) \geq i + 1 \) (since \( v_p(n) \) and \( i \) are integers). But Lemma 2.13.25 (applied to \( i + 1 \) instead of \( i \)) shows that \( p^{i+1} \mid n \) if and only if \( v_p(n) \geq i + 1 \). Thus, we have \( p^{i+1} \mid n \) (since we have \( v_p(n) \geq i + 1 \)). In other words, \( p^{i+1} \mid p^i w \) (since \( p^{i+1} = p^i \) and \( n = wp^i \)). But \( p \) is a prime; thus, \( p > 1 > 0 \) and therefore \( p \neq 0 \). Hence, \( p^i \neq 0 \). Thus, Exercise 2.2.3 (applied to \( p, w \) and \( p^i \) instead of \( a, b \) and \( c \)) shows that \( p \mid w \) holds if and only if \( pp^i \mid wp^i \). Hence, \( p \mid w \) holds (since \( pp^i \mid wp^i \) holds). Thus, Proposition 2.9.7 (i) (applied to \( p \) and \( w \) instead of \( a \) and \( b \)) yields \( \gcd(p, w) = |p| = p \) (since \( p > 0 \)). Comparing this with \( \gcd(p, w) = 1 \), we find \( p = 1 \). This contradicts \( p > 1 \).

This contradiction shows that our assumption was false. Hence, \( v_p(n) \leq i \) is proven. Combining this with \( v_p(n) \geq i \), we obtain \( v_p(n) = i \). This proves Lemma 2.13.27 (b).

The next property of \( p \)-adic valuations is crucial, as it reveals how they can be computed and bounded:

**Theorem 2.13.28.** Let \( p \) be a prime.

(a) We have \( v_p(ab) = v_p(a) + v_p(b) \) for any two integers \( a \) and \( b \).

(b) We have \( v_p(a + b) \geq \min\{v_p(a), v_p(b)\} \) for any two integers \( a \) and \( b \).

(c) We have \( v_p(1) = 0 \).

(d) We have \( v_p(q) = \begin{cases} 1, & \text{if } q = p; \\ 0, & \text{if } q \neq p \end{cases} \) for any prime \( q \).

Note that Theorem 2.13.28 (a) gives a formula for \( v_p(ab) \) in terms of \( v_p(a) \) and \( v_p(b) \), but there is no such formula for \( v_p(a + b) \) (since \( v_p(a) \) and \( v_p(b) \) do not uniquely determine \( v_p(a + b) \)). Thus, Theorem 2.13.28 (b) only gives a bound.

**Proof of Theorem 2.13.28** (a) Let \( a \) and \( b \) be two integers. We must prove that \( v_p(ab) = v_p(a) + v_p(b) \).
If $a = 0$, then this is true\(^{34}\). Thus, for the rest of the proof of Theorem 2.13.28 (a), we WLOG assume that $a \neq 0$. For similar reasons, we WLOG assume that $b \neq 0$.

The integer $a$ is nonzero (since $a \neq 0$). Thus, Lemma 2.13.27 (applied to $n = a$) shows that there exists a nonzero integer $u$ such that $u \perp p$ and $a = up^{v_p(a)}$.

Thus, $a \neq 0$ and $b \neq 0$.

The number $b$ is nonzero (since $b \neq 0$). Thus, Lemma 2.13.27 (applied to $n = b$) shows that there exists a nonzero integer $u$ such that $u \perp p$ and $b = up^{v_p(b)}$.

Consider this $u$, and denote it by $y$. Thus, $y$ is a nonzero integer such that $y \perp p$ and $b = yp^{v_p(b)}$.

We have $x \perp p$ and $y \perp p$. Thus, Theorem 2.10.9 (applied to $x, y$ and $p$ instead of $a, b$ and $c$) shows that $xy \perp p$.

The integer $ab$ is nonzero (since $a \neq 0$ and $b \neq 0$). Furthermore, multiplying the equalities $a = xp^{v_p(a)}$ and $b = yp^{v_p(b)}$, we obtain

$$ab = \left(xp^{v_p(a)}\right) \left(yp^{v_p(b)}\right) = (xy) \left(p^{v_p(a)} \cdot p^{v_p(b)}\right) = (xy) p^{v_p(a) + v_p(b)}.$$  

Thus, Lemma 2.13.27 (applied to $n = ab$, $i = v_p(a) + v_p(b)$ and $w = xy$) shows that $v_p(ab) = v_p(a) + v_p(b)$ (since $v_p(a) + v_p(b) \in \mathbb{N}$ and $xy \in \mathbb{Z}$ and $xy \perp p$).

This proves Theorem 2.13.28 (a).

(b) Let $a$ and $b$ be two integers. We must prove that $v_p(a + b) \geq \min \{v_p(a), v_p(b)\}$.

If $a = 0$, then this is true\(^{35}\). Thus, for the rest of the proof of Theorem 2.13.28 (b), we WLOG assume that $a \neq 0$. For similar reasons, we WLOG assume that $b \neq 0$.

The integer $a$ is nonzero (since $a \neq 0$). Thus, $v_p(a) \in \mathbb{N}$ (by Definition 2.13.23 (a)). Similarly, $v_p(b) \in \mathbb{N}$.

Let $m = \min \{v_p(a), v_p(b)\}$. Thus, $m \in \mathbb{N}$ (since $v_p(a) \in \mathbb{N}$ and $v_p(b) \in \mathbb{N}$).

We have $m = \min \{v_p(a), v_p(b)\} \leq v_p(a)$; in other words, $v_p(a) \geq m$. But Lemma 2.13.25 (applied to $n = a$ and $i = m$) shows that $p^m | a$ if and only if $v_p(a) \geq m$. Hence, we have $p^m | a$ (since $v_p(a) \geq m$). In other words, $a \equiv 0 \mod p^m$. Similarly, $b \equiv 0 \mod p^m$. Adding these two congruences together, we obtain $a + b \equiv 0 + 0 = 0 \mod p^m$. In other words, $p^m | a + b$.

But Lemma 2.13.25 (applied to $n = a + b$ and $i = m$) shows that $p^m | a + b$ if and only if $v_p(a + b) \geq m$. Hence, we have $v_p(a + b) \geq m$ (since $p^m | a + b$). Thus,

\[^{34}\text{Proof. Assume that } a = 0. \text{ Then } \sum_{i=0}^{a} b = 0 \text{ and thus } v_p(ab) = v_p(0) = \infty \text{ (by Definition 2.13.23 (b)). Also, from } a = 0, \text{ we obtain } v_p(a) = v_p(0) = \infty. \text{ Hence, } v_p(a) + v_p(b) = \infty + v_p(b) = \infty \text{ (since } \infty + k = \infty \text{ for each } k \in \mathbb{Z} \cup \{\infty\}). \text{ Comparing this with } v_p(ab) = \infty, \text{ we obtain } v_p(ab) = v_p(a) + v_p(b). \text{ This is exactly what we wanted to prove.}\]

\[^{35}\text{Proof. Assume that } a = 0. \text{ Then, } v_p\left(\sum_{i=0}^{a} b\right) = v_p(b) \geq \min \{v_p(a), v_p(b)\} \text{ (since any element of a set is } \geq \text{ to the minimum of this set). \text{ This is exactly what we wanted to prove.}\]
\(v_p(a + b) \geq m = \min \{v_p(a), v_p(b)\}\). This proves Theorem 2.13.28 (b).

\textbf{(c)} Exercise 2.10.1 (a) (applied to \(a = p\)) yields \(1 \perp p\). Also, \(1 = 1 \cdot p^0\). Thus, Lemma 2.13.27 (b) (applied to \(n = 1, i = 0\) and \(w = 1\)) yields \(v_p(1) = 0\). This proves Theorem 2.13.28 (c).

\textbf{(d)} Let \(q\) be a prime. We must prove that \(v_p(q) = \begin{cases} 1, & \text{if } q = p; \\ 0, & \text{if } q \neq p. \end{cases}\)

We are in one of the following two cases:

\textit{Case 1:} We have \(q = p\).
\textit{Case 2:} We have \(q \neq p\).

Let us first consider Case 1. In this case, we have \(q = p\). But Exercise 2.10.1 (a) (applied to \(a = p\)) yields \(1 \perp p\). Also, \(p = 1 \cdot p^1\). Thus, Lemma 2.13.27 (b) (applied to \(n = p, i = 1\) and \(w = 1\)) yields \(v_p(p) = 1\). From \(q = p\), we obtain \(v_p(q) = v_p(p) = 1\). Comparing this with \(\begin{cases} 1, & \text{if } q = p; \\ 0, & \text{if } q \neq p. \end{cases}\) (since \(q = p\)), we obtain \(v_p(q) = \begin{cases} 1, & \text{if } q = p; \\ 0, & \text{if } q \neq p. \end{cases}\). Hence, Theorem 2.13.28 (d) is proven in Case 1.

Let us now consider Case 2. In this case, we have \(q \neq p\). Thus, the primes \(q\) and \(p\) are distinct. Hence, Exercise 2.13.1 (applied to \(q\) and \(p\) instead of \(p\) and \(q\)) yields \(q \perp p\). Also, \(q = q \cdot p^0\) (since \(p^0 = 1\)). Thus, Lemma 2.13.27 (b) (applied to \(n = q, i = 0\) and \(w = q\)) yields \(v_p(q) = 0\). Comparing this with \(\begin{cases} 1, & \text{if } q = p; \\ 0, & \text{if } q \neq p. \end{cases}\) (since \(q \neq p\)), we obtain \(v_p(q) = \begin{cases} 1, & \text{if } q = p; \\ 0, & \text{if } q \neq p. \end{cases}\). Hence, Theorem 2.13.28 (d) is proven in Case 2.

We have now proven Theorem 2.13.28 (d) in each of the two Cases 1 and 2. Thus, Theorem 2.13.28 (d) is always proven.

\textbf{Corollary 2.13.29.} Let \(p\) be a prime. Let \(a_1, a_2, \ldots, a_k\) be \(k\) integers. Then, \(v_p(a_1 a_2 \cdots a_k) = v_p(a_1) + v_p(a_2) + \cdots + v_p(a_k)\).

\textbf{Proof of Corollary 2.13.29} This follows straightforwardly by induction on \(k\), using Theorem 2.13.28 (a) (as well as Theorem 2.13.28 (c) for the induction base). We leave the details to the reader, who has seen this sort of proof several times already.

\textbf{Exercise 2.13.5.} Let \(p\) be a prime. Let \(n \in \mathbb{Z}\). Prove that \(v_p(|n|) = v_p(n)\).
Exercise 2.13.6. Let $p$ be a prime. Let $a \in \mathbb{Z}$ and $k \in \mathbb{N}$. Prove that $v_p (a^k) = kv_p (a)$.

Exercise 2.13.7. Let $p_1, p_2, \ldots, p_u$ be finitely many distinct primes. Let $a_1, a_2, \ldots, a_u$ be nonnegative integers.

(a) Prove that $v_{p_i} (p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u}) = a_i$ for each $i \in \{1, 2, \ldots, u\}$.

(b) Prove that $v_p (p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u}) = 0$ for each prime $p$ satisfying $p \not\in \{p_1, p_2, \ldots, p_u\}$.

2.13.6. Prime factorization II

Proposition 2.13.30. Let $n$ be a positive integer. Let $(a_1, a_2, \ldots, a_k)$ be a prime factorization of $n$. Let $p$ be a prime. Then,

$(\text{the number of times } p \text{ appears in the tuple } (a_1, a_2, \ldots, a_k))$

$= (\text{the number of } i \in \{1, 2, \ldots, k\} \text{ such that } a_i = p)$

$= v_p (n)$.

Proof of Proposition 2.13.30. We have assumed that $(a_1, a_2, \ldots, a_k)$ is a prime factorization of $n$. Thus, $a_1, a_2, \ldots, a_k$ are primes satisfying $n = a_1 a_2 \cdots a_k$. Hence, for each $i \in \{1, 2, \ldots, k\}$, the integer $a_i$ is prime and thus satisfies

$$v_p (a_i) = \begin{cases} 1, & \text{if } a_i = p; \\ 0, & \text{if } a_i \neq p \end{cases}$$

(by Theorem 2.13.28 (d), applied to $q = a_i$).
From $n = a_1a_2\cdots a_k$, we obtain

$$v_p(n) = v_p(a_1a_2\cdots a_k) = v_p(a_1) + v_p(a_2) + \cdots + v_p(a_k) \quad \text{(by Corollary 2.13.29)}$$

$$= \sum_{i=1}^{k} v_p(a_i) = \sum_{i\in\{1,2,\ldots,k\}} \begin{cases} 1, & \text{if } a_i = p; \\ 0, & \text{if } a_i \neq p \end{cases} \quad \text{(by Corollary 2.13.29)}$$

$$= \sum_{i\in\{1,2,\ldots,k\}; a_i=p} 1 + \sum_{i\in\{1,2,\ldots,k\}; a_i\neq p} 0 = \sum_{i\in\{1,2,\ldots,k\}; a_i=p} 1$$

$$= (\text{the number of } i \in \{1,2,\ldots,k\} \text{ such that } a_i = p) \cdot 1$$

$$= (\text{the number of } i \in \{1,2,\ldots,k\} \text{ such that } a_i = p)$$

This proves Proposition 2.13.30 $\square$

We are finally ready to prove the so-called **Fundamental Theorem of Arithmetic**:

**Theorem 2.13.31.** Let $n$ be a positive integer.

(a) There exists a prime factorization of $n$.

(b) Any two such factorizations differ only in the order of their entries (i.e., are permutations of each other).

**Proof of Theorem 2.13.31** (a) Proposition 2.13.10 shows that $n$ can be written as a product of finitely many primes. In other words, there exist finitely many primes $p_1, p_2, \ldots, p_k$ such that $n = p_1p_2\cdots p_k$. Consider these primes. Thus, $(p_1, p_2, \ldots, p_k)$ is a prime factorization of $n$ (by the definition of “prime factorization”). Hence, there exists a prime factorization of $n$. This proves Theorem 2.13.31(a).

(b) Let $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_\ell)$ be two prime factorizations of $n$. We must prove that $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_\ell)$ differ only in the order of their entries (i.e., are permutations of each other).

Let $P$ be the set of all primes. Note that $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_\ell)$ are prime factorizations of $n$. Hence, $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_\ell)$ are tuples of primes, i.e., tuples of elements of $P$. 
Let $p \in P$. Thus, $p$ is a prime (by the definition of $P$). Hence, Proposition 2.13.30 shows that

\[
(\text{the number of times } p \text{ appears in the tuple } (a_1, a_2, \ldots, a_k))
= (\text{the number of } i \in \{1, 2, \ldots, k\} \text{ such that } a_i = p)
= v_p(n).
\]

Similarly,

\[
(\text{the number of times } p \text{ appears in the tuple } (b_1, b_2, \ldots, b_\ell))
= (\text{the number of } i \in \{1, 2, \ldots, \ell\} \text{ such that } b_i = p)
= v_p(n).
\]

Comparing these two equalities, we conclude that

\[
(\text{the number of times } p \text{ appears in } (a_1, a_2, \ldots, a_k))
= (\text{the number of times } p \text{ appears in } (b_1, b_2, \ldots, b_\ell)).
\] (39)

Now, forget that we fixed $p$. We thus have proven (39) for each $p \in P$. Hence, Lemma 2.13.20 shows that the tuples $(a_1, a_2, \ldots, a_k)$ and $(b_1, b_2, \ldots, b_\ell)$ differ only in the order of their entries (i.e., are permutations of each other). This completes our proof of Theorem 2.13.31 (b).\[\square\]

2.13.7. The canonical factorization

You have seen finite products such as

\[
\prod_{i \in \{1,2,3,4,5\}} i = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 5! = 120
\]

and

\[
\prod_{i \in \{3,5,7\}} (i^2 + 1) = (3^2 + 1) \cdot (5^2 + 1) \cdot (7^2 + 1) = 13000.
\]

Sometimes, infinite products (i.e., products ranging over infinite sets) also make sense. Many examples of well-defined infinite products arise from analysis and have to do with convergence. Here, we are doing algebra and thus shall only consider a very elementary, non-analytic meaning of convergence. Namely, we will consider infinite products that have only finitely many factors different from 1. For example, the product $2 \cdot 7 \cdot 4 \cdot 1 \cdot 1 \cdot 1 \cdots$ is of such form. It is easy to give infinitely many 1’s.

\[\text{[36] Here and in the following, } n! \text{ denotes the product } 1 \cdot 2 \cdots n \text{ whenever } n \in \mathbb{N}. \text{ Thus, in particular,}
\]

\[
0! = (\text{empty product}) = 1, \quad 1! = 1, \quad 2! = 1 \cdot 2 = 2,
3! = 1 \cdot 2 \cdot 3 = 6, \quad 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24, \quad 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120.
\]
a meaning to such products: Just throw away all the 1’s (since multiplying by 1 does not change a number) and take the product of the remaining (finitely many) numbers. So, for example, our product $2 \cdot 7 \cdot 4 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot \cdots$ should evaluate to infinitely many 1’s.

$2 \cdot 7 \cdot 4 = 56$.

This is indeed a meaningful and useful definition. For example, the set of all prime numbers is infinite (by Theorem 2.13.43 below), but nevertheless, for each nonzero integer $n$, the product $\prod_{p \text{ prime}} p^{v_p(n)}$ (where the “$\prod$” symbol means a product ranging over all primes $p$) is well-defined due to having only finitely many factors different from 1:

**Lemma 2.13.32.** Let $n$ be a nonzero integer.

(a) We have $v_p(n) = 0$ for every prime $p > |n|$. (Note that “for every prime $p > |n|$” is shorthand for “for every prime $p$ satisfying $p > |n|$.”)

(b) The product $\prod_{p \text{ prime}} p^{v_p(n)}$ has only finitely many factors different from 1.

(Here and in the following, the “$\prod$” symbol means a product ranging over all primes $p$.)

**Proof of Lemma 2.13.32.** (a) Let $p$ be a prime such that $p > |n|$. We must prove that $v_p(n) = 0$.

We have $p > 1$ (since $p$ is prime); thus, $p > 1 > 0$ and therefore $|p| = p > |n|$.

We have $n \neq 0$ (since $n$ is nonzero). Thus, if we had $p \mid n$, then we would have $|p| \leq |n|$ (by Proposition 2.2.3 (b), applied to $a = p$ and $b = n$), which would contradict $|p| > |n|$. Thus, we cannot have $p \mid n$. In other words, we have $p \nmid n$. But Corollary 2.13.26 yields that $v_p(n) = 0$ if and only if $p \nmid n$. Hence, $v_p(n) = 0$ (since $p \nmid n$). This proves Lemma 2.13.32 (a).

(b) For every prime $p > |n|$, we have $v_p(n) = 0$ (by Lemma 2.13.32 (a)) and thus $p^{v_p(n)} = p^0 = 1$. Thus, all but finitely many primes $p$ satisfy $p^{v_p(n)} = 1$ (since all but finitely many primes $p$ satisfy $p > |n|$). Therefore, all but finitely many factors of the product $\prod_{p \text{ prime}} p^{v_p(n)}$ are 1. In other words, the product $\prod_{p \text{ prime}} p^{v_p(n)}$ has only finitely many factors different from 1. This proves Lemma 2.13.32 (b). □

**Corollary 2.13.33.** Let $n$ be a positive integer. Then,

$$n = \prod_{p \text{ prime}} p^{v_p(n)}.$$  

Here, the infinite product $\prod_{p \text{ prime}} p^{v_p(n)}$ is well-defined (according to Lemma 2.13.32 (b)).

This expression $n = \prod_{p \text{ prime}} p^{v_p(n)}$ is called the **canonical factorization** of $n$. 
Proof of Corollary 2.13.33. Theorem 2.13.31 (a) shows that there exists a prime factorization of $n$. Consider such a factorization, and denote it by $(a_1, a_2, \ldots, a_k)$. Thus, $(a_1, a_2, \ldots, a_k)$ is a prime factorization of $n$; in other words, $a_1, a_2, \ldots, a_k$ are primes satisfying $n = a_1 a_2 \cdots a_k$. For each prime $p$, we have

\[(\text{the number of } i \in \{1, 2, \ldots, k\} \text{ such that } a_i = p) = v_p(n) \quad (40)\]

(by Proposition 2.13.30). Now,

\[
n = a_1 a_2 \cdots a_k = \prod_{i \in \{1, 2, \ldots, k\}} a_i
\]

\[
= \prod_{\text{prime } p} \prod_{i \in \{1, 2, \ldots, k\}; a_i = p} a_i
\]

\[
= \prod_{\text{prime } p} \sum_{i \in \{1, 2, \ldots, k\}; a_i = p} 1
\]

\[
= \prod_{\text{prime } p} v_p(n)
\]

This proves Corollary 2.13.33. \qed

The next exercise says that a nonnegative integer $n$ is uniquely determined by the family $(v_p(n))_{p \text{ prime}}$ of its $p$-valuations for all primes $p$:

Exercise 2.13.8. Let $n$ and $m$ be two nonnegative integers. Assume that

\[v_p(n) = v_p(m) \quad \text{for every prime } p. \quad (41)\]

Prove that $n = m$.

Corollary 2.13.34. Let $n$ be a nonzero integer. Then,

\[|n| = \prod_{\text{prime } p} p^{v_p(n)}.\]

Here, the infinite product $\prod_{\text{prime } p} p^{v_p(n)}$ is well-defined (according to Lemma 2.13.32 (b)).
Proof of Corollary 2.13.34 The integer \(|n|\) is positive (since \(n\) is nonzero). Hence, Corollary 2.13.33 (applied to \(|n|\) instead of \(n\)) yields

\[
|n| = \prod_{p \text{ prime}} \frac{p^{v_p(|n|)}}{p^{v_p(n)}} = \prod_{p \text{ prime}} p^{v_p(n)}.
\]

(since Exercise 2.13.3 yields \(v_p(|n|) = v_p(n)\))

This proves Corollary 2.13.34.

We can furthermore use \(p\)-adic valuations to check divisibility of integers:

**Proposition 2.13.35.** Let \(n\) and \(m\) be integers. Then, \(n \mid m\) if and only if each prime \(p\) satisfies \(v_p(n) \leq v_p(m)\).

**Proof of Proposition 2.13.35.** If \(m = 0\), then Proposition 2.13.35 is true.\(^{37}\) Hence, for the rest of this proof, we WLOG assume that \(m \neq 0\). Therefore, \(m\) is nonzero. Hence, \(v_p(m) \in \mathbb{N}\) (by Definition 2.13.23 (a)), so that \(v_p(m) < \infty\).

If \(n = 0\), then Proposition 2.13.35 is true.\(^{38}\) Hence, for the rest of this proof, we WLOG assume that \(n \neq 0\). Therefore, \(n\) is nonzero.

The statement of Proposition 2.13.35 does not change if we replace \(n\) and \(m\) by \(|n|\) and \(|m|\), respectively.\(^{39}\) Hence, we can WLOG assume that \(n\) and \(m\) are nonnegative. Assume this. Then, \(n \geq 0\), so that \(n > 0\) (since \(n\) is nonzero). Hence, \(n\) is a positive integer. Thus, Corollary 2.13.33 yields

\[
n = \prod_{p \text{ prime}} p^{v_p(n)}.
\]

Proof. Assume that \(m = 0\). Thus, each prime \(p\) satisfies \(v_p(m) \geq v_p(n)\), so that \(v_p(n) \leq v_p(m)\). Also, \(n | 0 = m\). Thus, the statements “\(n \mid m\)” and “each prime \(p\) satisfies \(v_p(n) \leq v_p(m)\)” are both true. Hence, \(n \mid m\) if and only if each prime \(p\) satisfies \(v_p(n) \leq v_p(m)\). In other words, Proposition 2.13.35 is true. Qed.

Proof. Assume that \(n = 0\). Thus, each prime \(p\) satisfies \(v_p(n) = v_p(0) = \infty\) (by Definition 2.13.23 (b)) and thus \(v_p(m) < \infty = v_p(n)\) (since \(2\) is a prime). Hence, we don’t have \(v_2(n) \leq v_2(m)\). Thus, the statement “each prime \(p\) satisfies \(v_p(n) \leq v_p(m)\)” is false (since \(p = 2\) is a counterexample).

If we had \(n \mid m\), then there would be an integer \(c\) such that \(m = nc\). This would then lead to \(m = \prod_{n \neq 0} p^{v_p(n)}\) \(c = 0\), which would contradict \(m \neq 0\). Hence, we cannot have \(n \mid m\). Thus, the statements “\(n \mid m\)” and “each prime \(p\) satisfies \(v_p(n) \leq v_p(m)\)” are both false. Hence, \(n \mid m\) if and only if each prime \(p\) satisfies \(v_p(n) \leq v_p(m)\). In other words, Proposition 2.13.35 is true. Qed.

Indeed, the statement “\(n \mid m\)” does not change (since Proposition 2.2.3 (a) yields that we have \(n \mid m\) if and only if \(|n| \mid |m|\), and the statement “each prime \(p\) satisfies \(v_p(n) \leq v_p(m)\)” does not change either (because Exercise 2.13.3 shows that \(v_p(|n|) = v_p(n)\) and \(v_p(|m|) = v_p(m)\)).
Similarly,
\[ m = \prod_{p \text{ prime}} p^{v_p(m)}. \] (43)

Our goal is to prove that \( n \mid m \) if and only if each prime \( p \) satisfies \( v_p(n) \leq v_p(m) \). We shall now prove the “\( \iff \)" and “\( \implies \)" directions of this “if and only if” statement separately.

\( \iff \): Assume that each prime \( p \) satisfies \( v_p(n) \leq v_p(m) \). We must prove that \( n \mid m \).

The product \( \prod_{p \text{ prime}} p^{v_p(m) - v_p(n)} \) is well-defined.

We have assumed that each prime \( p \) satisfies \( v_p(n) \leq v_p(m) \). In other words, each prime \( p \) satisfies \( v_p(m) - v_p(n) \geq 0 \) and therefore \( p^{v_p(m) - v_p(n)} \in \mathbb{Z} \). Hence, the product \( \prod_{p \text{ prime}} p^{v_p(m) - v_p(n)} \) is a product of integers, and thus itself an integer.

Let us denote this product by \( c \). Thus,
\[ c = \prod_{p \text{ prime}} p^{v_p(m) - v_p(n)}. \] (44)

Thus, \( c \) is an integer (since we have just shown that \( \prod_{p \text{ prime}} p^{v_p(m) - v_p(n)} \) is an integer).

Multiplying the equalities (42) and (44), we obtain
\[ nc = \left( \prod_{p \text{ prime}} p^{v_p(n)} \right) \left( \prod_{p \text{ prime}} p^{v_p(m) - v_p(n)} \right) = \prod_{p \text{ prime}} \left( p^{v_p(n)} p^{v_p(m) - v_p(n)} \right) \]
\[ = \prod_{p \text{ prime}} p^{v_p(m)} = m \quad \text{(by (43))}. \]

In other words, \( m = nc \). Hence, \( n \mid m \). This completes the proof of the “\( \iff \)" direction of Proposition 2.13.35.

\( \implies \): Assume that \( n \mid m \). We must prove that each prime \( p \) satisfies \( v_p(n) \leq v_p(m) \).

\[ \text{Proof.} \] Let \( p \) be a prime such that \( p > |m| \). Then, \( v_p(m) = 0 \) (by Lemma 2.13.32(a), applied to \( m \) instead of \( n \)), so that \( v_p(m) - v_p(n) \leq v_p(m) = 0 \). On the other hand, \( v_p(n) \leq v_p(m) \) (since we assumed that each prime \( p \) satisfies \( v_p(n) \leq v_p(m) \)); thus, \( v_p(m) - v_p(n) \geq 0 \). Combining this with \( v_p(m) - v_p(n) \leq 0 \), we obtain \( v_p(m) - v_p(n) = 0 \). Hence, \( p^{v_p(m) - v_p(n)} = p^0 = 1 \).

Now, forget that we fixed \( p \). We thus have proven that every prime \( p > |m| \) satisfies \( p^{v_p(m) - v_p(n)} = 1 \). Hence, all but finitely many primes \( p \) satisfy \( p^{v_p(m) - v_p(n)} = 1 \) (since all but finitely many primes \( p \) satisfy \( p > |m| \)). In other words, the product \( \prod_{p \text{ prime}} p^{v_p(m) - v_p(n)} \) has only finitely many factors different from 1. Hence, this product is well-defined.
So let $p$ be a prime. Recall that $n \mid m$. In other words, there exists some integer $b$ such that $m = nb$. Consider this $b$. Now,

$$v_p \left( \frac{m}{nb} \right) = v_p(nb) = v_p(n) + v_p(b) \geq 0$$

(by Theorem 2.13.28(a), applied to $a = n$)

$$\geq v_p(n),$$

so that $v_p(n) \leq v_p(m)$. Now, forget that we fixed $p$. We thus have proven that each prime $p$ satisfies $v_p(n) \leq v_p(m)$. This completes the proof of the “$\Rightarrow$” direction of Proposition 2.13.35.

Let us extract one of the steps of our above proof into a separate lemma, since we shall use the same reasoning later on:

**Lemma 2.13.36.** For each prime $p$, let $a_p$ and $b_p$ be nonnegative integers such that

$$a_p \leq b_p. \quad (45)$$

Assume that all but finitely many primes $p$ satisfy $b_p = 0$. Then, the products $\prod_{p \text{ prime}} p^{a_p}$ and $\prod_{p \text{ prime}} p^{b_p}$ are both well-defined and satisfy $\prod_{p \text{ prime}} p^{a_p} \mid \prod_{p \text{ prime}} p^{b_p}$.

**Proof of Lemma 2.13.36** This is going to be really boring: The well-definedness part is all about bookkeeping finiteness information, whereas the $\prod_{p \text{ prime}} p^{a_p} \mid \prod_{p \text{ prime}} p^{b_p}$ claim is proven just as we proved the “$\Leftarrow$” direction of Proposition 2.13.35. For the sake of completeness, let us nevertheless give the complete proof:

All but finitely many primes $p$ satisfy $b_p = 0$. In other words, there exists some finite set $S$ of primes such that every prime $p \notin S$ satisfies

$$b_p = 0. \quad (46)$$

Consider this $S$. Clearly, all but finitely many primes $p$ satisfy $p \notin S$ (since $S$ is finite).

Now, every prime $p \notin S$ satisfies

$$a_p = 0 \quad (47)$$

Hence, all but finitely many primes $p$ satisfy $a_p = 0$ (since all but finitely many primes $p$ satisfy $p \notin S$). Thus, all but finitely many primes $p$ satisfy $p^{a_p} = p^0 = 1$. In other words, only finitely many primes $p$ satisfy $p^{a_p} \neq 1$. In other words, only finitely many factors of the product $\prod_{p \text{ prime}} p^{a_p}$ are different from 1. Hence, this product $\prod_{p \text{ prime}} p^{a_p}$ is well-defined.

Also, all but finitely many primes $p$ satisfy $b_p = 0$. Therefore, all but finitely many primes $p$ satisfy $p^{b_p} = p^0 = 1$. In other words, only finitely many primes $p$ satisfy $p^{b_p} \neq 1$.

\[\text{Proof. Let } p \notin S \text{ be a prime. Then, } (45) \text{ yields } a_p \leq b_p = 0 \text{ (by } (46)). \text{ Thus, } a_p = 0 \text{ (since } a_p \text{ is a nonnegative integer), qed.}\]
In other words, only finitely many factors of the product $\prod_{p \text{ prime}} p^{b_p}$ are different from 1. Hence, this product $\prod_{p \text{ prime}} p^{b_p}$ is well-defined.

The product $\prod_{p \text{ prime}} p^{b_p - a_p}$ is well-defined. Denote this product by $c$.

For each prime $p$, we have $b_p - a_p \geq 0$ (by (15)) and thus $b_p - a_p \in \mathbb{N}$. Hence, for each prime $p$, the number $p^{b_p - a_p}$ is an integer. Therefore, $\prod_{p \text{ prime}} p^{b_p - a_p}$ is a product of integers, and thus itself an integer. In other words, $c$ is an integer (since $c = \prod_{p \text{ prime}} p^{b_p - a_p}$).

But from $c = \prod_{p \text{ prime}} p^{b_p - a_p}$, we obtain

$$\left( \prod_{p \text{ prime}} p^{b_p} \right) c = \left( \prod_{p \text{ prime}} p^{a_p} \right) \left( \prod_{p \text{ prime}} p^{b_p - a_p} \right) = \prod_{p \text{ prime}} \left( p^{a_p} p^{b_p - a_p} \right) = \prod_{p \text{ prime}} p^{b_p}.$$

Thus, $\prod_{p \text{ prime}} p^{a_p} \mid \prod_{p \text{ prime}} p^{b_p}$ (since $c$ is an integer). This completes the proof of Lemma 2.13.36.

**Corollary 2.13.37.** For each prime $p$, let $b_p$ be a nonnegative integer. Assume that all but finitely many primes $p$ satisfy $b_p = 0$. Let $n = \prod_{p \text{ prime}} p^{b_p}$. Then,

$$v_q(n) = b_q \quad \text{for each prime } q.$$

**Proof of Corollary 2.13.37.** The product $\prod_{p \text{ prime}} p^{b_p}$ is well-defined. (This can be shown just as in the proof of Lemma 2.13.36.) Now, choose a list $(a_1, a_2, \ldots, a_k)$ of primes that contains each prime $p$ exactly $b_p$ times. (Such a list clearly exists: For example, we can pick

$$\left( \begin{array}{cccc} 2,2,\ldots,2,3,3,\ldots,3,5,5,\ldots,5 \end{array} \right).$$

This is indeed a finite list, since all but finitely many primes $p$ satisfy $b_p = 0$.) Now, the list $(a_1, a_2, \ldots, a_k)$ contains each prime $p$ exactly $b_p$ times (and no other entries). Hence, the product $a_1 a_2 \cdots a_k$ of the entries of this list contains each prime

$$p^{b_p} - a_p = 0 - 0 = 0 \quad \text{and therefore } p^{b_p - a_p} = p^0 = 1.$$

Thus, all but finitely many primes $p$ satisfy $p^{b_p - a_p} = 1$ (since all but finitely many primes $p$ satisfy $p \notin S$). In other words, only finitely many primes $p$ satisfy $p^{b_p - a_p} \neq 1$. In other words, only finitely many factors of the product $\prod_{p \text{ prime}} p^{b_p - a_p}$ are different from 1. Hence, this product $\prod_{p \text{ prime}} p^{b_p - a_p}$ is well-defined.
\( p \) exactly \( b_p \) times as a factor (and no other factors). Thus, this product equals 
\[
\prod_{\text{prime}} p^{b_p}.\]
In other words, \( a_1 a_2 \cdots a_k = \prod_{\text{prime}} p^{b_p} \). Hence,
\[
n = \prod_{\text{prime}} p^{b_p} = a_1 a_2 \cdots a_k.
\]
Thus, \((a_1, a_2, \ldots, a_k)\) is a prime factorization of \( n \) (since \((a_1, a_2, \ldots, a_k)\) is a tuple of primes).

Let \( q \) be a prime. Proposition 2.13.30 (applied to \( p = q \)) yields
\[
\text{(the number of times } q \text{ appears in the tuple } (a_1, a_2, \ldots, a_k)) = \text{(the number of } i \in \{1, 2, \ldots, k\} \text{ such that } a_i = q) = v_{q}(n).
\]
Thus,
\[
v_{q}(n) = \text{(the number of times } q \text{ appears in the tuple } (a_1, a_2, \ldots, a_k)) = b_q \]
(since the list \((a_1, a_2, \ldots, a_k)\) contains each prime \( p \) exactly \( b_p \) times, and thus contains the prime \( q \) exactly \( b_q \) times). This proves Corollary 2.13.37.

**Exercise 2.13.9.** Let \( n \) be a nonzero integer. Let \( a \) and \( b \) be two integers. Assume that
\[
a \equiv b \mod{p^{v_p(n)}} \quad \text{for every prime } p. \tag{48}
\]
Prove that \( a \equiv b \mod{n} \).

**2019-02-13 lecture**

Canonical factorizations can also be used to describe gcds and lcms:

**Proposition 2.13.38.** Let \( n \) and \( m \) be two nonzero integers. Then,
\[
\gcd(n, m) = \prod_{\text{prime}} p^{\min\{v_p(n), v_p(m)\}} \tag{49}
\]
and
\[
\text{lcm}(n, m) = \prod_{\text{prime}} p^{\max\{v_p(n), v_p(m)\}}. \tag{50}
\]

**Proof of Proposition 2.13.38** If \( p \) is any prime, then \( v_p(n) \) and \( v_p(m) \) are nonnegative integers (since \( n \) and \( m \) are nonzero), and thus so are \( \min\{v_p(n), v_p(m)\} \) and \( \max\{v_p(n), v_p(m)\} \).
It is easy to see that the infinite products
\[
\prod_{p \text{ prime}} p^{\min\{v_p(n),v_p(m)\}} \quad \text{ and } \quad \prod_{p \text{ prime}} p^{\max\{v_p(n),v_p(m)\}}
\]
are well-defined. \footnote{Proof. Let \( p \) be a prime such that \( p > \max\{|m|,|n|\} \). Thus, \( p > \max\{|m|,|n|\} \geq |m| \) and therefore \( v_p(m) = 0 \) (by Lemma 2.13.32(a), applied to \( m \) instead of \( n \)). Similarly, \( v_p(n) = 0 \).

Hence, \( \max\{v_p(n),v_p(m)\} = \max\{0,0\} = 0 \) and therefore \( p^{\max\{v_p(n),v_p(m)\}} = p^0 = 1 \).

Now, forget that we fixed \( p \). We thus have proven that every prime \( p > \max\{|m|,|n|\} \)
satisfies \( p^{\max\{v_p(n),v_p(m)\}} = 1 \). Hence, all but finitely many primes \( p \) satisfy \( p^{\max\{v_p(n),v_p(m)\}} = 1 \)
(since all but finitely many primes \( p \) satisfy \( p > \max\{|m|,|n|\} \)). In other words, the product
\[
\prod_{p \text{ prime}} p^{\max\{v_p(n),v_p(m)\}}
\]
has only finitely many factors different from 1. Hence, this product is well-defined. Similarly, we can show that the product
\[
\prod_{p \text{ prime}} p^{\min\{v_p(n),v_p(m)\}}
\]
is well-defined.

Define two nonnegative integers
\[
g = \prod_{p \text{ prime}} p^{\min\{v_p(n),v_p(m)\}} \quad \text{ and } \quad h = \gcd(n,m).
\]

Note that \( h = \gcd(n,m) \) is a positive integer (since \( n \) and \( m \) are nonzero) and thus nonzero. Thus, \( v_p(h) \) is a nonnegative integer for each prime \( p \).

Corollary 2.13.34 yields \( |n| = \prod_{p \text{ prime}} p^{v_p(n)} \). But each prime \( p \) satisfies
\[
\min\{v_p(n),v_p(m)\} \leq v_p(n) \quad \text{(since \( \min\{v_p(n),v_p(m)\} \leq \min\{v_p(n),v_p(m)\} \)}
\]
Hence, Lemma 2.13.36 (applied to \( a_p = \min\{v_p(n),v_p(m)\} \) and \( b_p = v_p(n) \)) yields
\[
\prod_{p \text{ prime}} p^{\min\{v_p(n),v_p(m)\}} | \prod_{p \text{ prime}} p^{v_p(n)}. \quad \text{This rewrites as } \quad g \mid |n| \quad \text{(since } \gcd(n,m) = h \).
\]

On the other hand, Proposition 2.13.35 (applied to \( h \) and \( n \) instead of \( n \) and \( m \)) shows that \( h \mid n \) if and only if each prime \( p \) satisfies \( v_p(h) \leq v_p(n) \). Thus, each prime \( p \) satisfies \( v_p(h) \leq v_p(n) \).

Now, fix any prime \( p \). Then, \( v_p(h) \leq v_p(n) \) (as we have just seen) and \( v_p(h) \leq v_p(m) \) (similarly). Combining these two inequalities, we obtain
\[
v_p(h) \leq \min\{v_p(n),v_p(m)\}
\]
(since \( \min\{v_p(n),v_p(m)\} \) must be one of the two numbers \( v_p(n) \) and \( v_p(m) \), but we have just seen that \( v_p(h) \) is \( \leq \) to each of these two numbers).

Now, forget that we fixed \( p \). We thus have show that each prime \( p \) satisfies \( v_p(h) \leq \min\{v_p(n),v_p(m)\} \). Hence, Lemma 2.13.36 (applied to \( a_p = v_p(h) \) and
\( b_p = \min \{ v_p(n), v_p(m) \} \) yields \( \prod_{p \text{ prime}} p^{v_p(h)} \mid \prod_{p \text{ prime}} p^{\min \{ v_p(n), v_p(m) \}} \). But \( h \) is positive; hence, Corollary 2.13.33 (applied to \( h \) instead of \( n \)) yields

\[
h = \prod_{p \text{ prime}} p^{v_p(h)} \mid \prod_{p \text{ prime}} p^{\min \{ v_p(n), v_p(m) \}} = g.
\]

Thus, we know that \( g \mid h \) and \( h \mid g \). Hence, Exercise 2.2.2 (applied to \( a = g \) and \( b = h \)) yields \( |g| = |h| \). But \( g \) is nonnegative; thus, \( |g| = g \). Hence, \( g = |g| = |h| = h \) (since \( h \) is positive). In view of (51), this rewrites as \( \prod_{p \text{ prime}} p^{\min \{ v_p(n), v_p(m) \}} = \gcd(n, m) \). This proves (49).

The proof of (50) is entirely analogous to the proof of (49) we just gave: We merely need to flip all divisibilities and inequalities and replace “\( \min \)” by “\( \max \)” everywhere, and use Lemma 2.11.8 instead of Lemma 2.9.15.

\[\text{Example 2.13.39.}\] For this example, set \( n = 3^2 \cdot 5 \cdot 7^8 \) and \( m = 2 \cdot 3^3 \cdot 7^2 \). Let us compute \( \gcd(n, m) \) and \( \operatorname{lcm}(n, m) \) using Proposition 2.13.38.

From \( n = 3^2 \cdot 5 \cdot 7^8 \), we obtain (using Corollary 2.13.37) that

\[
v_3(n) = 2, \quad v_5(n) = 1, \quad v_7(n) = 8, \quad \text{and} \quad v_p(n) = 0 \quad \text{for each prime} \quad p \not\in \{3, 5, 7\}.
\]

Similarly, from \( m = 2 \cdot 3^3 \cdot 7^2 \), we obtain

\[
v_2(m) = 1, \quad v_3(m) = 3, \quad v_7(m) = 2, \quad \text{and} \quad v_p(m) = 0 \quad \text{for each prime} \quad p \not\in \{2, 3, 7\}.
\]

Now, (49) yields

\[
\gcd(n, m) = \prod_{p \text{ prime}} p^{\min \{ v_p(n), v_p(m) \}} = 2^0 \cdot 3^2 \cdot 5^0 \cdot 7^2 = 3^2 \cdot 7^2.
\]
Likewise, (50) yields
\[
\text{lcm}(n,m) = \prod_{p \text{ prime}} p^{\max\{v_p(n),v_p(m)\}} = 2^{\max\{v_2(n),v_2(m)\}} \cdot 3^{\max\{v_3(n),v_3(m)\}} \cdot 5^{\max\{v_5(n),v_5(m)\}} \cdot 7^{\max\{v_7(n),v_7(m)\}}. 
\]

Proposition 2.13.38 can be generalized to the case of \(k\) integers \(b_1, b_2, \ldots, b_k\) instead of two integers \(n, m\):

**Proposition 2.13.40.** Let \(b_1, b_2, \ldots, b_k\) be finitely many nonzero integers, with \(k > 0\). Then,
\[
gcd(b_1, b_2, \ldots, b_k) = \prod_{p \text{ prime}} p^{\min\{v_p(b_1),v_p(b_2),\ldots,v_p(b_k)\}} \tag{52}
\]
an
\[
lcm(b_1, b_2, \ldots, b_k) = \prod_{p \text{ prime}} p^{\max\{v_p(b_1),v_p(b_2),\ldots,v_p(b_k)\}}. \tag{53}
\]

**Proof of Proposition 2.13.40.** The proof of Proposition 2.13.40 is analogous to the proof of Proposition 2.13.38, with two minor exceptions:

- Instead of applying Lemma 2.9.15 (in the proof of (52)), we now have to apply the analogous claim for \(k\) integers.\footnote{Namely: Let \(m \in \mathbb{Z}\) and let \(b_1, b_2, \ldots, b_k\) be integers such that \((m \mid b_i) \text{ for all } i \in \{1, 2, \ldots, k\}\). Then, \(m \mid \gcd(b_1, b_2, \ldots, b_k)\).} The latter claim follows from Theorem 2.9.20 (a).

- Instead of applying Lemma 2.11.8 (in the proof of (53)), we now have to apply the analogous claim for \(k\) integers\footnote{Namely: Let \(m \in \mathbb{Z}\) and let \(b_1, b_2, \ldots, b_k\) be integers such that \((b_i \mid m) \text{ for all } i \in \{1, 2, \ldots, k\}\). Then, \(\text{lcm}(b_1, b_2, \ldots, b_k) \mid m\).}. The latter claim follows from Theorem 2.11.9 (a). Alternatively, instead of applying this claim, we can argue as follows: Setting \(g = \prod_{p \text{ prime}} p^{\max\{v_p(b_1),v_p(b_2),\ldots,v_p(b_k)\}}\) and \(h = \text{lcm}(b_1, b_2, \ldots, b_k)\), we see that \((b_i \mid g \text{ for all } i \in \{1, 2, \ldots, k\})\) (by an argument analogous to the one we used to show \((g \mid n \text{ and } g \mid m)\) in the original proof of (49)).
is a common multiple of \(b_1, b_2, \ldots, b_k\). In other words, \(g \in \text{Mul} (b_1, b_2, \ldots, b_k)\).

Hence, \(g\) is a positive element of \(\text{Mul} (b_1, b_2, \ldots, b_k)\) (since \(g\) is positive). Hence, \(g \geq \text{lcm} (b_1, b_2, \ldots, b_k)\) (since \(\text{lcm} (b_1, b_2, \ldots, b_k)\) is the smallest positive element of \(\text{Mul} (b_1, b_2, \ldots, b_k)\)). In other words, \(g \geq h\) (since \(h = \text{lcm} (b_1, b_2, \ldots, b_k)\)).

On the other hand, we prove \(g \mid h\) (similarly to how we proved \(h \mid g\) in the original proof of \((49)\)). Thus, Proposition 2.2.3 \((b)\) (applied to \(g\) and \(h\) instead of \(a\) and \(b\)) yields \(|g| \leq |h|\) (since \(h \neq 0\)). Since \(g\) is positive, we have \(|g| = g\) and thus \(g = |g| \leq |h| = h\) (since \(h\) is positive). Combining this with \(g \geq h\), we obtain \(g = h\). As before, this completes the proof of \((53)\).

\[
\square
\]

We can use Propositions 2.13.38 and 2.13.40 to reprove certain facts about \(\text{lcm}\)s and \(\gcd\)s. For example, let us prove Theorem 2.11.6 and solve Exercise 2.11.2.

**Second proof of Theorem 2.11.6 (sketched).** WLOG assume that \(a\) and \(b\) are nonzero (since otherwise, the claim of Theorem 2.11.6 easily reduces to \(0 = 0\)). Then, \(ab\) is nonzero as well.

Hence, Corollary 2.13.34 (applied to \(n = ab\)) yields

\[
|ab| = \prod_{p \text{ prime}} p^{v_p(ab)}.
\]

Now, Proposition 2.13.38 yields

\[
\gcd (a, b) = \prod_{p \text{ prime}} p^{\min \{v_p(a), v_p(b)\}} \quad \text{and} \quad \text{lcm} (a, b) = \prod_{p \text{ prime}} p^{\max \{v_p(a), v_p(b)\}}.
\]

Multiplying these two equalities, we get

\[
\gcd (a, b) \cdot \text{lcm} (a, b) = \left( \prod_{p \text{ prime}} p^{\min \{v_p(a), v_p(b)\}} \right) \cdot \left( \prod_{p \text{ prime}} p^{\max \{v_p(a), v_p(b)\}} \right)
\]

\[
= \prod_{p \text{ prime}} \left( p^{\min \{v_p(a), v_p(b)\} + \max \{v_p(a), v_p(b)\}} \right)
\]

\[
= \prod_{p \text{ prime}} p^{v_p(a) + v_p(b)}
\]

(since \(\min \{u, v\} + \max \{u, v\} = u + v\) for any reals \(u, v\))

\[
= \prod_{p \text{ prime}} p^{v_p(ab)} = \prod_{p \text{ prime}} p^{v_p(ab)} = |ab|.
\]

Hence, Theorem 2.11.6 is proven again.

See Section 6.56 for a second solution to Exercise 2.11.2 using Propositions 2.13.38 and 2.13.40 (and a slightly more general result that can be proven in the same way). \(\square\)
Exercise 2.13.10. Let \( n \) and \( m \) be two integers. Let \( p \) be a prime.
(a) Prove that \( v_p(\gcd(n,m)) = \min\{v_p(n), v_p(m)\} \).
(b) Prove that \( v_p(\lcm(n,m)) = \max\{v_p(n), v_p(m)\} \).

Exercise 2.13.11. Let \( a, b, c \) be three integers.
(a) Prove that \( \gcd(a, \lcm(b,c)) = \lcm(\gcd(a,b), \gcd(a,c)) \).
(b) Prove that \( \lcm(a, \gcd(b,c)) = \gcd(\lcm(a,b), \lcm(a,c)) \).

The two parts of Exercise 2.13.11 can be regarded as “distributivity laws”, but for the binary operations \( \gcd \) and \( \lcm \) (or \( \lcm \) and \( \gcd \), respectively) instead of + and ·.

2.13.8. Coprimality through prime factors

Proposition 2.13.41. Let \( n \) and \( m \) be two integers. Then, \( n \perp m \) if and only if there exists no prime \( p \) that divides both \( n \) and \( m \).

Proof of Proposition 2.13.41. \( \implies \): Assume that \( n \perp m \). We must prove that there exists no prime \( p \) that divides both \( n \) and \( m \).

Let \( p \) be a prime that divides both \( n \) and \( m \). Thus, \( p \mid n \) and \( p \mid m \). Hence, \( p \mid \gcd(n,m) \) (by Lemma 2.9.15, applied to \( p, n \) and \( m \) instead of \( n, a \) and \( b \)). But \( \gcd(n,m) = 1 \) (since \( n \perp m \)). Hence, \( p \mid \gcd(n,m) = 1 \). Hence, Proposition 2.13.4 (applied to \( a = p \) and \( b = 1 \)) yields \( |p| \leq |1| = 1 \).

But \( p \) is a prime; thus, \( p > 1 > 0 \), so that \( |p| = p \) and thus \( p = |p| \leq 1 \). This contradicts \( p > 1 \).

Now, forget that we fixed \( p \). We thus have obtained a contradiction for each prime \( p \) that divides both \( n \) and \( m \). Hence, there exists no prime \( p \) that divides both \( n \) and \( m \). This proves the “\( \implies \)” direction of Proposition 2.13.41.

\( \Longleftarrow \): Assume that there exists no prime \( p \) that divides both \( n \) and \( m \). We must prove that \( n \perp m \).

Assume the contrary. Thus, we don’t have \( n \perp m \). In other words, we don’t have \( \gcd(n,m) = 1 \). In other words, \( \gcd(n,m) \neq 1 \). Hence, there exists at least one prime \( p \) such that \( p \mid \gcd(n,m) \). \(^{46}\) Consider this \( p \).

We have \( p \mid \gcd(n,m) \mid n \) and \( p \mid \gcd(n,m) \mid m \). Thus, the prime \( p \) divides both \( n \) and \( m \). This contradicts the assumption that there exists no prime \( p \) that divides both \( n \) and \( m \).

This contradiction shows that our assumption was false. Hence, \( n \perp m \) is proven. This proves the “\( \Longleftarrow \)” direction of Proposition 2.13.41. □

\(^{46}\) Proof. This is obvious if \( \gcd(n,m) = 0 \) (because in that case, we can take \( p = 2 \), or any other prime). Thus, for the rest of this proof, we WLOG assume that \( \gcd(n,m) \neq 0 \).

Thus, \( \gcd(n,m) > 1 \) (since \( \gcd(n,m) \) is a nonnegative integer satisfying \( \gcd(n,m) \neq 0 \) and \( \gcd(n,m) \neq 1 \)). Hence, Proposition 2.13.8 (applied to \( \gcd(n,m) \) instead of \( n \)) yields that there exists at least one prime \( p \) such that \( p \mid \gcd(n,m) \). Qed.
Corollary 2.13.42. Let $n$ and $m$ be two nonzero integers. Then:

(a) The infinite sum $\sum_{p \text{ prime}} v_p(n) v_p(m)$ is well-defined (i.e., all but finitely many primes $p$ satisfy $v_p(n) v_p(m) = 0$).

(b) We have $n \perp m$ if and only if

$$\sum_{p \text{ prime}} v_p(n) v_p(m) = 0.$$

Proof of Corollary 2.13.42 (sketched). (a) For every prime $p > |n|$, we have $v_p(n) = 0$ (by Lemma 2.13.32 (a)) and thus $v_p(n) v_p(m) = 0$. Now, Corollary 2.13.42 (a) follows easily.

(b) A sum of nonnegative reals is 0 if and only if each of its addends is 0. Thus, the sum $\sum_{p \text{ prime}} v_p(n) v_p(m)$ is 0 if and only if we have $v_p(n) v_p(m) = 0$ for all primes $p$ (because all the addends $v_p(n) v_p(m)$ of our sum are nonnegative reals). Hence, we have the following chain of equivalences:

$$\left( \sum_{p \text{ prime}} v_p(n) v_p(m) = 0 \right)$$

$$\iff (v_p(n) v_p(m) = 0 \text{ for all primes } p)$$

$$\iff ((v_p(n) = 0 \text{ or } v_p(m) = 0) \text{ for all primes } p)$$

$$\iff ((p \mid n \text{ or } p \mid m) \text{ for all primes } p)$$

since Corollary 2.13.26 yields the equivalences $(v_p(n) = 0) \iff (p \nmid n)$ and $(v_p(m) = 0) \iff (p \nmid m)$ for each prime $p$

$$\iff (\text{there exists no prime } p \text{ such that } (p \mid n \text{ and } p \mid m))$$

$$\iff (\text{there exists no prime } p \text{ that divides both } n \text{ and } m)$$

$$\iff (n \perp m) \quad \text{(by Proposition 2.13.41)}.$$

This proves Corollary 2.13.42 (b). \qed

Corollary 2.13.42 (b) is the reason for the notation “$\perp$” that we are using for coprimality. In fact, when $n$ is a positive integer, we can regard the $p$-valuations $v_p(n)$ as the “coordinates” of $n$ in an (infinite-dimensional) Cartesian coordinate system. Then, the sum $\sum_{p \text{ prime}} v_p(n) v_p(m)$ in Corollary 2.13.42 is something like a “dot product” between $n$ and $m$. Thus, Corollary 2.13.42 (b) shows that two integers $n$ and $m$ are coprime if and only if their “dot product” is 0. But for vectors in a Euclidean space, the dot product is 0 if and only if the vectors are orthogonal. Thus, coprime integers are like orthogonal vectors. Of course, this analogy should
be taken with a grain of salt; in particular, our “dot product” is far from being bilinear.\footnote{Or, rather, it is bilinear with respect to multiplication: If we denote \( \sum_{p \text{ prime}} v_p(n) v_p(m) \) by \( \langle n, m \rangle \), then we have \( \langle n_1 n_2, m \rangle = \langle n_1, m \rangle + \langle n_2, m \rangle \) and \( \langle n, m_1 m_2 \rangle = \langle n, m_1 \rangle + \langle n, m_2 \rangle \) for arbitrary integers \( n_1, n_2, n, m, m_1, m_2 \).}

### 2.13.9. There are infinitely many primes

#### Theorem 2.13.43

There are infinitely many primes.

**Proof of Theorem 2.13.43**

The following proof is a classic, appearing in Euclid’s *Elements*:

Let \((p_1, p_2, \ldots, p_k)\) be any finite list of primes. We shall find a new prime distinct from each of \(p_1, p_2, \ldots, p_k\).

Indeed, \(p_1, p_2, \ldots, p_k\) are primes, and thus are integers > 1 (by the definition of a “prime”). Hence, they are positive integers; thus, their product \(p_1 p_2 \cdots p_k\) is a positive integer as well. Thus, \(p_1 p_2 \cdots p_k > 0\).

Now, let \(n = p_1 p_2 \cdots p_k + 1\). Then, \(n = p_1 p_2 \cdots p_k + 1 > 1\). Hence, Proposition 2.13.8 shows that there exists at least one prime \(p\) such that \(p \mid n\). Consider this \(p\).

We claim that \(p\) is distinct from each of \(p_1, p_2, \ldots, p_k\).

*Proof: Assume the contrary. Thus, \(p = p_i\) for some \(i \in \{1, 2, \ldots, k\}\). Consider this \(i\).

We have \(p_1 p_2 \cdots p_k = p_i \cdot (p_1 p_2 \cdots p_{i-1} p_{i+1} p_{i+2} \cdots p_k)\). Thus, \(p_i \mid p_1 p_2 \cdots p_k\) (since \(p_1 p_2 \cdots p_{i-1} p_{i+1} p_{i+2} \cdots p_k\) is an integer). Hence, \(p = p_i \mid p_1 p_2 \cdots p_k\). In other words, \(p_1 p_2 \cdots p_k \equiv 0 \mod p\).

\[
\begin{align*}
n &= p_1 p_2 \cdots p_k + 1 \\ &\equiv 0 \mod p
\end{align*}
\]

Hence, \(1 \equiv n \mod p\). But \(p \mid n\) and thus \(n \equiv 0 \mod p\). Hence, \(1 \equiv n \equiv 0 \mod p\); in other words, \(p \mid 1 - 0 = 1\). Thus, Proposition 2.2.3(b) (applied to \(a = p\) and \(b = 1\)) yields \(|p| \leq |1| = 1\). But \(p\) is a prime; thus, \(p > 1\), so that \(|p| = p > 1\). This contradicts \(|p| \leq 1\). This contradiction shows that our assumption was wrong.

"qed."

Thus, we have proven that \(p\) is distinct from each of \(p_1, p_2, \ldots, p_k\). Hence, there exists a prime distinct from each of \(p_1, p_2, \ldots, p_k\) (namely, \(p\)).

Now, forget that we fixed \(p_1, p_2, \ldots, p_k\). We thus have proven that if \((p_1, p_2, \ldots, p_k)\) is any finite list of primes, then there exists a prime distinct from each of \(p_1, p_2, \ldots, p_k\). In other words, given any finite list of primes, there exists at least one prime that is not in this list. In other words, no finite list of primes can cover all the primes. In other words, there are infinitely many primes. This proves Theorem 2.13.43. \(\square\)
Note that our proof of Theorem 2.13.43 is constructive: It gives an algorithm to construct arbitrarily many distinct primes. This algorithm is not very efficient, since $p_1 p_2 \cdots p_k + 1$ can be very large even if $p_1, p_2, \ldots, p_k$ are fairly small. In practice, the sieve of Eratosthenes is much better for generating primes. Much faster algorithms are known.

Exercise 2.13.12. Let $p$ be a prime. Let $a \in \mathbb{Z}$ be such that $a^2 \equiv 1 \mod p$. Prove that $a \equiv 1 \mod p$ or $a \equiv -1 \mod p$.

2.14. Euler’s totient function ($\phi$-function)

2.14.1. Definition and some formulas

Recall that $\mathbb{P}$ stands for the set of all positive integers.

Definition 2.14.1. We define a function $\phi : \mathbb{P} \to \mathbb{N}$ as follows: For each $n \in \mathbb{P}$, we let $\phi(n)$ be the number of all $i \in \{1, 2, \ldots, n\}$ that are coprime to $n$. In other words,

$$\phi(n) = |\{i \in \{1, 2, \ldots, n\} \mid i \perp n\}|. \quad (54)$$

This function $\phi$ is called Euler’s totient function or just $\phi$-function.

Example 2.14.2. (a) We have $\phi(12) = 4$, since the number of all $i \in \{1, 2, \ldots, 12\}$ that are coprime to 12 is 4 (indeed, these $i$ are 1, 5, 7 and 11).

(b) We have $\phi(13) = 12$, since the number of all $i \in \{1, 2, \ldots, 13\}$ that are coprime to 13 is 12 (indeed, these $i$ are 1, 2, \ldots, 12).

(c) We have $\phi(14) = 6$, since the number of all $i \in \{1, 2, \ldots, 14\}$ that are coprime to 14 is 6 (indeed, these $i$ are 1, 3, 5, 9, 11, 13).

(d) We have $\phi(1) = 1$, since the number of all $i \in \{1, 2, \ldots, 1\}$ that are coprime to 1 is 1 (indeed, the only such $i$ is 1).

The $\phi$-function $\phi$ is denoted by $\varphi$ by some authors.

Proposition 2.14.3. Let $p$ be a prime. Then, $\phi(p) = p - 1$.

Proof of Proposition 2.14.3. Here is the idea: The definition of $\phi$ shows that $\phi(p)$ is the number of all $i \in \{1, 2, \ldots, p\}$ that are coprime to $p$. But we know exactly what these $i$ are: They are just the first $p - 1$ positive integers 1, 2, \ldots, $p - 1$. (In fact, Proposition 2.13.4 shows that each of the integers $1, 2, \ldots, p - 1$ is coprime to $p$, whereas $\gcd(p, p) = p > 1$ shows that $p$ is not coprime to $p$.) Thus, $\phi(p)$ is the number of these $p - 1$ integers; in other words, $\phi(p) = p - 1$.

For one last time, here is the proof in detail:

We have $p > 1$ (since $p$ is a prime), thus $p \neq 1$. Also, $p \mid p$; hence, Proposition 2.9.7 (i) (applied to $a = p$ and $b = p$) yields $\gcd(p, p) = |p| = p$ (since $p > 1 > 0$).

Now, we claim that

$$\{i \in \{1, 2, \ldots, p\} \mid i \perp p\} \subseteq \{1, 2, \ldots, p - 1\}. \quad (55)$$
**Proof of (55):** Let \( i \in \{1, 2, \ldots, p\} \) be such that \( i \perp p \). From \( i \perp p \), we obtain \( \gcd(i, p) = 1 \). If we had \( i = p \), then we would have \( \gcd(i, p) = \gcd(p, p) = p \neq 1 \), which would contradict \( \gcd(i, p) = 1 \). Thus, we cannot have \( i = p \). Hence, we have \( i \neq p \). Combining this with \( i \in \{1, 2, \ldots, p\} \), we obtain \( i \in \{1, 2, \ldots, p\} \setminus \{p\} = \{1, 2, \ldots, p-1\} \). Hence, we have \( i \neq p \). Combining this with \( i \in \{1, 2, \ldots, p\} \), we obtain \( i \in \{1, 2, \ldots, p\} \setminus \{p\} = \{1, 2, \ldots, p-1\} \).

Now, forget that we fixed \( i \). We thus have proven that every \( i \in \{1, 2, \ldots, p\} \) satisfying \( i \perp p \) must belong to \( \{1, 2, \ldots, p-1\} \). In other words, \( \{i \in \{1, 2, \ldots, p\} \mid i \perp p\} \subseteq \{1, 2, \ldots, p-1\} \). This proves (55).

Conversely, we have \( \{1, 2, \ldots, p-1\} \subseteq \{i \in \{1, 2, \ldots, p\} \mid i \perp p\} \).

**Proof of (56):** Let \( j \in \{1, 2, \ldots, p-1\} \). Thus, \( j \) is coprime to \( p \) (by Proposition 2.13.4, applied to \( i = j \)). In other words, \( j \perp p \). Also, \( j \in \{1, 2, \ldots, p-1\} \subseteq \{1, 2, \ldots, p\} \). Hence, \( j \) is an \( i \in \{1, 2, \ldots, p\} \) satisfying \( i \perp p \). In other words, \( j \in \{i \in \{1, 2, \ldots, p\} \mid i \perp p\} \) for each \( j \in \{1, 2, \ldots, p-1\} \). In other words, \( \{1, 2, \ldots, p-1\} \subseteq \{i \in \{1, 2, \ldots, p\} \mid i \perp p\} \). This proves (56).

Combining (55) with (56), we obtain \( \{i \in \{1, 2, \ldots, p\} \mid i \perp p\} = \{1, 2, \ldots, p-1\} \).

Now, (54) (applied to \( n = p \)) yields
\[
\phi(p) = \left| \left\{ i \in \{1, 2, \ldots, p\} \mid i \perp p \right\} \right| = \left| \{1, 2, \ldots, p-1\} \right| = p-1.
\]

This proves Proposition 2.14.3.

Proposition 2.14.3 can be generalized as follows:

**Exercise 2.14.1.** Let \( p \) be a prime. Let \( k \) be a positive integer. Prove that \( \phi(p^k) = (p-1)p^{k-1} \).

**Theorem 2.14.4.** Let \( m \) and \( n \) be two coprime positive integers. Then, \( \phi(mn) = \phi(m) \cdot \phi(n) \).

We will prove Theorem 2.14.4 later (in Section 2.16.3).

**Theorem 2.14.5.** Let \( n \) be a positive integer. Then,
\[
\phi(n) = \prod_{\substack{p \text{ prime; } \newline p | n}} (p-1)p^{v_p(n)-1} = n \cdot \prod_{\substack{p \text{ prime; } \newline p | n}} \left( 1 - \frac{1}{p} \right).
\]

Theorem 2.14.5 will be proven in Section 2.16.3.
Exercise 2.14.2. Let \( n \) be a positive integer.
(a) Prove that
\[
n - \phi(n) = |\{i \in \{1, 2, \ldots, n\} \mid \text{we don’t have } i \perp n\}|.
\]
(b) We have \( n - \phi(n) \ge 0 \).
(c) Let \( d \) be a positive divisor of \( n \). Prove that \( d - \phi(d) \le n - \phi(n) \).
(d) Let \( d \) be a positive divisor of \( n \) such that \( d \ne n \). Prove that \( d - \phi(d) < n - \phi(n) \).

2.14.2. The totient sum theorem

Theorem 2.14.6. Let \( n \) be a positive integer. Then,
\[
\sum_{d \mid n} \phi(d) = n.
\]
Here and in the following, the symbol “\( \sum \)” stands for “sum over all positive divisors \( d \) of \( n \)”.

For example, for \( n = 12 \), Theorem 2.14.6 states that
\[
\phi(1) + \phi(2) + \phi(3) + \phi(4) + \phi(6) + \phi(12) = 12.
\]

Before we prove Theorem 2.14.6 let us motivate an argument via a classical puzzle:

Exercise 2.14.3. You have a corridor with 1000 lamps, which are initially all off. Each lamp has a lightswitch controlling its state.

Every night, a ghost glides through the corridor (always in the same direction) and flips some of the switches:

On the 1st night, the ghost flips every switch.
On the 2nd night, the ghost flips switches 2, 4, 6, 8, 10, \ldots.
On the 3rd night, the ghost flips switches 3, 6, 9, 12, 15, \ldots.

etc.

(That is: For each \( k \in \{1, 2, \ldots, 1000\} \), the ghost spends the \( k \)-th night flipping switches \( k, 2k, 3k, \ldots \).)

Which lamps will be on after 1000 nights?

In more rigorous terms, Exercise 2.14.3 is simply asking which of the numbers 1, 2, \ldots, 1000 have an odd number of positive divisors. (Indeed, the situation after 1000 nights looks as follows: For each \( n \in \{1, 2, \ldots, 1000\} \), the \( n \)-th switch has been flipped exactly once for each positive divisor of \( n \); thus, the \( n \)-th lamp is on if and only if \( n \) has an odd number of positive divisors.)
Experiments reveal that among the first 10 positive integers, only three have an odd number of positive divisors: namely, 1, 4 and 9. (For example, 9 has the 3 positive divisors 1, 3 and 9.) This suggests the following:

**Proposition 2.14.7.** A positive integer \( n \) has an odd number of positive divisors if and only if \( n \) is a perfect square.

**Proof of Proposition 2.14.7.** Fix a positive integer \( n \). If \( d \) is a positive divisor of \( n \), then \( n/d \) is a positive divisor of \( n \) as well. This allows us to define a map

\[
F : \{ \text{positive divisors of } n \} \to \{ \text{positive divisors of } n \},
\]

\[
d \mapsto n/d.
\]

This map \( F \) has the property that \( F \circ F = \text{id} \), because each \( d \in \{ \text{positive divisors of } n \} \) satisfies

\[
(F \circ F)(d) = F \left( \left( \underbrace{F(d)}_{=n/d} \right) \right) = F \left( \frac{n}{d} \right) \quad \text{(by the definition of } F \text{)}
\]

\[
= \frac{n}{(n/d)} \quad \text{(by the definition of } F \text{)}
\]

\[
= d = \text{id}(d).
\]

Hence, the map \( F \) is inverse to itself. Thus, the map \( F \) is invertible, i.e., is a bijection.

For the rest of this proof, the word “divisor” shall mean “positive divisor of \( n \)”.

Thus, \( F \) is a map from \( \{ \text{divisors} \} \) to \( \{ \text{divisors} \} \).

The rough idea from here on is the following. The map \( F \) “pairs up” each divisor \( d \) with the divisor \( F(d) = n/d \). Thus, the divisors are “grouped into pairs”, except for those that satisfy \( d = n/d \) (because these would have to be paired up with themselves). When \( n \) is not a perfect square, there are no such “exceptional” divisors, since \( d = n/d \) means \( n = d^2 \). When \( n \) is a perfect square, there is exactly one such “exceptional” divisor, namely \( \sqrt{n} \). So the number of divisors is even if \( n \) is not a perfect square, and odd otherwise (because clearly, all the pairs have no effect on the parity of the total number of divisors, and thus can be forgotten). In other words, \( n \) has an odd number of positive divisors if and only if \( n \) is a perfect square.

There are several ways to make this argument rigorous; here is the easiest (though perhaps the least instructive one): A divisor \( d \) shall be called

\(^{48} \text{Proof. Let } d \text{ be a positive divisor of } n. \text{ Thus, } d \text{ is a positive integer satisfying } d \mid n. \text{ But Proposition 2.2.3(c) (applied to } a = d \text{ and } b = n) \text{ yields that } d \mid n \text{ if and only if } \frac{n}{d} \in \mathbb{Z} \text{ (since } d \neq 0). \text{ Hence, } \frac{n}{d} \in \mathbb{Z} \text{ (since } d \mid n). \text{ In other words, } n/d \in \mathbb{Z}. \text{ Moreover, } n/d \text{ is positive (since } n \text{ and } d \text{ are positive).}

\text{So } n/d \text{ is a positive integer (since } n/d \in \mathbb{Z}) \text{ and is a divisor of } n \text{ (since } n = (n/d) \cdot d). \text{ Hence, } n/d \text{ is a positive divisor of } n. \text{ Qed.}

\(^{49} \text{We shall give a more rigorous proof shortly.} \)
• small if \(d < n/d\);
• medium if \(d = n/d\);
• large if \(d > n/d\).

It is easy to see that if \(d\) is a small divisor, then \(F(d)\) is a large divisor\(^{50}\). Hence, the map

\[
F^+: \{\text{small divisors}\} \rightarrow \{\text{large divisors}\},
\]

\[
d \mapsto F(d)
\]

is well-defined. Similarly, the map

\[
F^-: \{\text{large divisors}\} \rightarrow \{\text{small divisors}\},
\]

\[
d \mapsto F(d)
\]

is well-defined. These two maps \(F^+\) and \(F^-\) are both restrictions of the map \(F\), and thus are mutually inverse (since the map \(F\) is inverse to itself). Hence, the map \(F^+\) is invertible, i.e., is a bijection. Thus, we have found a bijection from \(\{\text{small divisors}\}\) to \(\{\text{large divisors}\}\) (namely, \(F^+\)). Therefore,

\[
|\{\text{small divisors}\}| = |\{\text{large divisors}\}|.
\]

\[(57)\]

On the other hand, let us take a look at medium divisors. If \(d\) is a medium divisor, then \(d = n/d\) (by the definition of “medium”), so that \(d^2 = n\) and thus \(n\) must be a perfect square. Thus, if \(n\) is not a perfect square, then there are no medium divisors. In other words, if \(n\) is not a perfect square, then

\[
|\{\text{medium divisors}\}| = 0.
\]

\[(58)\]

But if \(n\) is a perfect square, then \(n\) has exactly one medium divisor\(^{51}\). In other words, if \(n\) is a perfect square, then

\[
|\{\text{medium divisors}\}| = 1.
\]

\[(59)\]

But each divisor is either small or medium or large, and there are no overlaps between these three classes (i.e., a divisor cannot be small and medium at the same time, or small

\(^{50}\text{Proof.}\) Let \(d\) be a small divisor. Thus, \(d < n/d\). Hence, \(n/d > d = n/(n/d)\). In view of \(F(d) = n/d\), this rewrites as \(F(d) > n/(F(d))\). In other words, \(F(d)\) is a large divisor (by the definition of “large divisor”). Qed.

\(^{51}\text{Proof.}\) Assume that \(n\) is a perfect square. Thus, \(n = w^2\) for some \(w \in \mathbb{Z}\). Consider this \(w\). Clearly, \(w \neq 0\) (since \(ww = w^2 = n \neq 0\)), so that \(|w| > 0\).

Let \(u = |w|\). Thus, \(u \in \mathbb{Z}\) (since \(w \in \mathbb{Z}\)). Hence, \(u\) is a positive integer (since \(u = |w| > 0\)).

Moreover, from \(u = |w|\), we obtain \(u^2 = |w|^2 = w^2 = n\). Hence, \(u = n/u\).

This positive integer \(u\) satisfies \(uu = u^2 = n\) and thus \(u \mid n\). Hence, \(u\) is a positive divisor of \(n\) (that is, a divisor, as we call it). This divisor \(u\) is medium, since it satisfies \(u = n/u\).

Moreover, if \(d\) is any medium divisor, then \(d = n/d\) (by the definition of “medium”), thus \(d^2 = n = u^2\), thus \(\sqrt{d^2} = \sqrt{u^2} = |u| = u\) (since \(u\) is positive), thus \(u = \sqrt{d^2} = |d| = d\) (since \(d\) is positive) and therefore \(d = u\). In other words, any medium divisor must equal \(u\). This shows that \(u\) is the only medium divisor (since we already know that \(u\) is a medium divisor). Hence, \(n\) has exactly one medium divisor.
and large, or medium and large). Thus, in order to count the number of all divisors, we can add the number of small divisors, the number of medium divisors and the number of large divisors. In other words:

\[ |\{\text{divisors}\}| = |\{\text{small divisors}\}| + |\{\text{medium divisors}\}| + |\{\text{large divisors}\}| \]

(by \((57)\))

\[ = 2 \cdot |\{\text{large divisors}\}| + |\{\text{medium divisors}\}| \equiv 0 \mod 2 \]

Hence, if \(n\) is not a perfect square, then

\[ |\{\text{divisors}\}| \equiv |\{\text{medium divisors}\}| = 0 \mod 2 \]

(by \((58)\)). In other words, if \(n\) is not a perfect square, then the number of divisors is even.

On the other hand, if \(n\) is a perfect square, then

\[ |\{\text{divisors}\}| \equiv |\{\text{medium divisors}\}| = 1 \mod 2 \]

(by \((59)\)). In other words, if \(n\) is a perfect square, then the number of divisors is odd.

Combining the results of the previous two paragraphs, we conclude that the number of divisors is odd if \(n\) is a perfect square, and is even otherwise. In other words, \(n\) has an odd number of positive divisors if and only if \(n\) is a perfect square. This proves Proposition 2.14.7.

Having proven Proposition 2.14.7 we now can answer Exercise 2.14.3. The 31 lamps \(1^2, 2^2, \ldots, 31^2\) (and no others) will be on after the 1000 nights. (Indeed, these 31 lamps correspond to the 31 perfect squares in the set \(\{1, 2, \ldots, 1000\}\).)

The bijection \(F\) from the proof of Proposition 2.14.7 will serve us well in our proof of Theorem 2.14.6. Beside that, we need the following lemma:

**Lemma 2.14.8.** Let \(n\) be a positive integer. Let \(d\) be a positive divisor of \(n\). Then,

\[ \{\text{the number of } i \in \{1, 2, \ldots, n\} \text{ such that } \gcd(i, n) = d\} = \phi\left(\frac{n}{d}\right) \]

**Proof of Lemma 2.14.8.** We have \(d \mid n\) (since \(d\) is a divisor of \(n\)) and \(d \neq 0\) (since \(d\) is positive). Thus, Proposition 2.2.3 (c) (applied to \(d\) and \(n\) instead of \(a\) and \(b\)) yields that \(d \mid n\) if and only if \(\frac{n}{d} \in \mathbb{Z}\). Thus, \(\frac{n}{d} \in \mathbb{Z}\) (since \(d \mid n\)). In other words, \(n/d \in \mathbb{Z}\). Thus, \(n/d\) is an integer. This integer \(n/d\) is positive (since \(n\) and \(d\) are positive). Hence, \(\phi\left(\frac{n}{d}\right)\) is well-defined.

Define two sets \(I\) and \(J\) by

\[ I = \{i \in \{1, 2, \ldots, n\} \mid \gcd(i, n) = d\} \] (60)
and
\[ J = \{ i \in \{1,2,\ldots,n/d\} \mid i \perp n/d \}. \]  
(61)

But (54) (applied to \( n/d \) instead of \( n \)) yields
\[ \phi(n/d) = |\{ i \in \{1,2,\ldots,n/d\} \mid i \perp n/d \}| = |J| \]  
(62)
(since \( \{ i \in \{1,2,\ldots,n/d\} \mid i \perp n/d \} = J \)).

We shall next construct a bijection from \( I \) to \( J \) (which will show that \( |I| = |J| \)). For each \( a \in I \), we have \( a/d \in J \). Hence, we can define a map
\[ f : I \to J, \]
\[ a \mapsto a/d. \]

\footnote{Proof. Let \( a \in I \). Thus, \( a \in I = \{ i \in \{1,2,\ldots,n\} \mid \gcd(i,n) = d \} \). In other words, \( a \) is an element of \( \{1,2,\ldots,n\} \) satisfying \( \gcd(a,n) = d \).

Thus, \( d = \gcd(a,n) \mid a \). But Proposition \( \ref{prop:divisibility} \) (applied to \( d \) and \( a \) instead of \( a \) and \( b \)) yields that \( d \mid a \) if and only if \( a/d \in \mathbb{Z} \). Thus, \( a/d \in \mathbb{Z} \) (since \( d \mid a \)). In other words, \( a/d \in \mathbb{Z} \).

Also, \( a \in \{1,2,\ldots,n\} \), so that \( 0 < a \leq n \). We can divide this chain of inequalities by \( d \) (since \( d \) is positive), and thus obtain \( 0 < a/d \leq n/d \). Hence, \( a/d \in \{1,2,\ldots,n/d\} \) (since \( a/d \in \mathbb{Z} \)).

Furthermore, Corollary \( \ref{cor:gcd} \) (applied to \( d, a/d \) and \( n/d \) instead of \( s, a \) and \( b \)) yields
\[ \gcd(d(a/d), d(n/d)) = \frac{|d|}{d} \gcd(a/d, n/d) = d \gcd(a/d, n/d). \]
(since \( d \) is positive)

Solving this for \( \gcd(a/d, n/d) \), we obtain
\[ \gcd(a/d, n/d) = \frac{1}{d} \gcd\left(\frac{d(a/d)}{d}, \frac{d(n/d)}{d}\right) = \frac{1}{d} \gcd(a,n) = \frac{1}{d} = 1. \]
In other words, \( a/d \perp n/d \).

So we know that \( a/d \) is an element of \( \{1,2,\ldots,n/d\} \) satisfying \( a/d \perp n/d \). In other words, \( a/d \in \{ i \in \{1,2,\ldots,n/d\} \mid i \perp n/d \} = J \). Qed.}
For each \( b \in J \), we have \( bd \in I \). Thus, we can define a map
\[
g : J \to I, \\
b \mapsto bd.
\]

The two maps \( f \) and \( g \) are mutually inverse (since the map \( f \) divides its input by \( d \), while the map \( g \) multiplies its input by \( d \)). Hence, \( f \) is invertible, i.e., is a bijection. Thus, there exists a bijection from \( I \) to \( J \) (namely, \( f \)). Hence, \(|I| = |J| = \phi(n/d)\) (by (62)). Thus,
\[
\phi(n/d) = |I| = |\{i \in \{1, 2, \ldots, n\} \mid \gcd(i, n) = d\}| \quad (\text{by (60)})
\]
This proves Lemma 2.14.8.

**Proof of Theorem 2.14.6.** Consider the map \( F \) we defined in the proof of Proposition 2.14.7. This map \( F \) is a bijection (as we have seen back in that proof). In other words, the map
\[
\{\text{positive divisors of } n\} \to \{\text{positive divisors of } n\}, \\
d \mapsto n/d
\]
is a bijection (since this is precisely the map \( F \)). Thus, we can substitute \( n/d \) for \( d \) in the sum \( \sum_{d|n} \phi(d) \) (and, more generally, in any sum that ranges over all positive divisors \( d \) of \( n \)). We thus obtain
\[
\sum_{d|n} \phi(d) = \sum_{d|n} \phi(n/d). \quad (63)
\]

**Proof.** Let \( b \in J \). Thus, \( b \in J = \{i \in \{1, 2, \ldots, n/d\} \mid i \perp n/d\} \). In other words, \( b \) is an element of \( \{1, 2, \ldots, n/d\} \) satisfying \( b \perp n/d \).

From \( b \in \{1, 2, \ldots, n/d\} \subseteq \mathbb{Z} \) and \( d \in \mathbb{Z} \), we obtain \( bd \in \mathbb{Z} \). From \( b \in \{1, 2, \ldots, n/d\} \), we obtain \( 0 < b \leq n/d \). We can multiply this chain of inequalities by \( d \) (since \( d \) is positive), and thus obtain \( 0 < bd \leq n \). Thus, \( bd \in \{1, 2, \ldots, n\} \) (since \( bd \in \mathbb{Z} \)). Corollary 2.9.19 (applied to \( d, b \) and \( n/d \) instead of \( s, a \) and \( b \)) yields
\[
\gcd (db, d(n/d)) = \underbrace{|d|}_{=d} \frac{\gcd (b, n/d)}{=d} = d.
\]
(since \( d \) is positive) (since \( b \perp n/d \))

Thus, \( d = \gcd \left( \frac{db}{=bd}, \frac{d(n/d)}{=n} \right) = \gcd (bd, n) \), so that \( \gcd (bd, n) = d \).

So we know that \( bd \) is an element of \( \{1, 2, \ldots, n\} \) satisfying \( \gcd (bd, n) = d \). In other words, \( bd \in \{i \in \{1, 2, \ldots, n\} \mid \gcd (i, n) = d\} = \mathbb{I} \), qed.
But
\[ n = |\{1, 2, \ldots, n\}| = (\text{the number of } i \in \{1, 2, \ldots, n\}) \]
\[ = \sum_{d|n} (\text{the number of } i \in \{1, 2, \ldots, n\} \text{ such that } \gcd(i, n) = d) \]
\[ = \sum_{d|n} \phi\left(\frac{n}{d}\right) \]
(by Lemma 2.14.8)
\[ = \sum_{d|n} \phi(d) \]
(because if \( i \in \{1, 2, \ldots, n\} \), then \( \gcd(i, n) \) is a positive divisor of \( n \))
\[ = \sum_{d|n} \phi\left(\frac{n}{d}\right) = \sum_{d|n} \phi\left(\frac{n}{d}\right) \]
(by (63)).

This proves Theorem 2.14.6. \[ \square \]

**Exercise 2.14.4.** Let \( n \in \mathbb{N} \) satisfy \( n > 2 \). Prove that \( \phi(n) \) is even.

**Exercise 2.14.5.** Let \( n \in \mathbb{N} \) satisfy \( n > 1 \). Prove that
\[ \sum_{i \in \{1, 2, \ldots, n\}; i \perp n} i = n \phi(n) / 2. \]

---

### 2.15. Fermat, Euler, Wilson

#### 2.15.1. Fermat and Euler: statements

The following theorem is known as *Fermat’s Little Theorem* (often abbreviated as “FLT”):

**Theorem 2.15.1.** Let \( p \) be a prime. Let \( a \in \mathbb{Z} \).

(a) If \( p \nmid a \), then \( a^{p-1} \equiv 1 \mod p \).

(b) We always have \( a^p \equiv a \mod p \).

The word “little” in the name of Theorem 2.15.1 is meant to distinguish the theorem from “Fermat’s Last Theorem”, a much more difficult result only proven in the 1990s. (Unfortunately, the latter result is also abbreviated as “FLT”.)

We will prove Theorem 2.15.1 soon, by showing a more general result (Theorem 2.15.3). But before we do so, let us convince ourselves that the parts (a) and (b) of Theorem 2.15.1 are equivalent:

**Remark 2.15.2.** Theorem 2.15.1 (b) follows from Theorem 2.15.1 (a), because (using the notations of Theorem 2.15.1):

- If \( p \nmid a \), then Theorem 2.15.1 (a) yields \( a^{p-1} \equiv 1 \mod p \), thus \( a^p = a \equiv a \mod p \).
If \( p \mid a \), then both \( a^p \) and \( a \) are \( \equiv 0 \mod p \) (because \( p \mid a \) entails \( a \equiv 0 \mod p \) since \( p > 0 \)), and therefore \( a^{p^\ell} \equiv 0 \mod p \).

Conversely, Theorem 2.15.1 (a) follows from Theorem 2.15.1 (b) by the following argument: Let \( p \) and \( a \) be as in Theorem 2.15.1. Assume that \( p \nmid a \). Then, \( p \perp a \) (by Proposition 2.13.5), so that \( a \perp p \). Thus, we can “cancel” \( a \) from any congruence modulo \( p \) (by Lemma 2.10.10). Doing this to the congruence \( a^p \equiv a \mod p \) (which follows from Theorem 2.15.1 (b)), we obtain \( a^{p-1} \equiv 1 \mod p \).

The next result is known as Euler’s theorem:

**Theorem 2.15.3.** Let \( n \) be a positive integer. Let \( a \in \mathbb{Z} \) be coprime to \( n \).

Then, \( a^\phi(n) \equiv 1 \mod n \).

Theorem 2.15.3 yields Theorem 2.15.1 (a), since \( \phi(p) = p - 1 \) when \( p \) is prime. Since we also know that Theorem 2.15.1 (b) follows from Theorem 2.15.1 (a), we see that a proof of Theorem 2.15.3 will immediately yield the whole Theorem 2.15.1. Before we give said proof, let us show an example of how Theorem 2.15.3 can be used:

**Exercise 2.15.1.** What is the last digit of \( 3^{45} \)?

**Notational remark:** An expression of the form “\( a^{bc} \)” always means \( a^{(bc)} \), not \( (a^b)^c \). (Actually, there is no need for an extra notation for \( (a^b)^c \), because \( (a^b)^c = a^{bc} \).)

**Solution to Exercise 2.15.1 (sketched).** The last digit of a positive integer \( n \) is \( n \mod 10 \) (that is, the remainder of \( n \) upon division by 10). So we need to work modulo 10.

Since 3 is coprime to 10, we can apply Theorem 2.15.3 to \( n = 10 \) and \( a = 3 \). We thus get \( 3^{\phi(10)} \equiv 1 \mod 10 \). Since \( \phi(10) = 4 \), this rewrites as \( 3^4 \equiv 1 \mod 10 \). Now, \( 4^{5} = 4 \cdot 4^4 \), so that

\[
3^{45} = 3^{4 \cdot 4^4} = \left(3^4 \equiv 1 \mod 10 \right)^{4^4} \equiv 1^{4^4} = 1 \mod 10.
\]

So the last digit of \( 3^{45} \) is 1.

Theorem 2.15.3 is also the reason why certain rational numbers (such as \( \frac{2}{7} = 0.285714 \ldots \)) have purely periodic decimal expansions, while others (such as

\[\text{See below for details of this argument.}\]

\[\text{The bar ( ) over the “285714” means that we are repeating 285714 over and over. So } 0.285714 = \]

\[0.285714285714285714285714 \ldots \]
\[
\frac{1}{12} = 0.08\overline{3} = 0.083333\ldots \text{ or } \frac{1}{2} = 0.5\overline{0} = 0.50000\ldots \]
have their periods start only after some initial nonrepeating block. We refer [ConradE, §4] to the details of this. 

2.15.2. Proving Euler and Fermat

Our proof of Theorem 2.15.3 will rely on the following lemma:

**Lemma 2.15.4.** Let \( n \) be a positive integer. Then,

\[
\phi(n) = |\{i \in \{0, 1, \ldots, n-1\} \mid i \perp n\}|.
\]

**First proof of Lemma 2.15.4 (sketched).** If \( n = 1 \), then this lemma can easily be proven by hand. Thus, WLOG assume that \( n \neq 1 \). Hence, \( n > 1 \) (since \( n \) is a positive integer). Thus, neither 0 nor \( n \) is coprime to \( n \) (since \( \gcd(0, n) = n \neq 1 \) and \( \gcd(n, n) = n > 1 \)). Hence, 

\[
\{i \in \{0, 1, \ldots, n-1\} \mid i \perp n\} = \{i \in \{1, 2, \ldots, n\} \mid i \perp n\}
\]

(because these sets could only differ in the elements 0 and \( n \), but none of these two elements belongs to any of these two sets), and therefore

\[
|\{i \in \{0, 1, \ldots, n-1\} \mid i \perp n\}| = |\{i \in \{1, 2, \ldots, n\} \mid i \perp n\}| = \phi(n)
\]

(by (54)). This proves Lemma 2.15.4. \( \square \)

**Second proof of Lemma 2.15.4.** We have \( n \mid n \) and thus \( n \equiv 0 \mod n \). Hence, Proposition 2.9.7 (d) (applied to \( a = n \), \( b = n \) and \( c = 0 \)) yields that \( \gcd(n, n) = \gcd(0, n) = \gcd(0, n) \) (by Proposition 2.9.7 (b)). Hence, \( \gcd(0, n) = 1 \) holds if and only if \( \gcd(n, n) = 1 \). In other words, the number 0 is coprime to \( n \) if and only if \( n \) is coprime to \( n \). Hence, if we remove \( n \) from the set \( \{1, 2, \ldots, n\} \) and add 0 instead (so that our set becomes \( \{0, 1, \ldots, n-1\} \)), then the number of elements coprime to \( n \) in that set does not change. In other words,

\[
\begin{align*}
(\text{the number of all } i \in \{0, 1, \ldots, n-1\} \text{ that are coprime to } n) \\
= (\text{the number of all } i \in \{1, 2, \ldots, n\} \text{ that are coprime to } n) \\
= |\{i \in \{1, 2, \ldots, n\} \mid i \perp n\}| = \phi(n) \quad \text{(by (54))}.
\end{align*}
\]

In brief, the rule is as follows: Any fraction \( \frac{a}{b} \) with \( a, b \in \mathbb{Z} \) (and \( b \neq 0 \)) has such a decimal representation with a period. (A period means a part that gets repeated over and over.) A fraction \( \frac{a}{b} \) is called purely periodic if its period (in decimal notation) begins straight after the decimal point. So \( \frac{2}{7} \) is purely periodic but \( \frac{1}{12} \) and \( \frac{1}{2} \) are not. Now, the answer is that a fraction \( \frac{a}{b} \) (with \( a \perp b \)) is purely periodic if and only if \( b \perp 10 \) (in other words, \( 2 \nmid b \) and \( 5 \nmid b \)). This can be proven using Theorem 2.15.3.

\[56\]

since neither 0 nor \( n \) is coprime to \( n \)
In other words,

\[ \phi(n) = (\text{the number of all } i \in \{0,1,\ldots,n-1\} \text{ that are coprime to } n) = |\{i \in \{0,1,\ldots,n-1\} \mid i \perp n\}|. \]

This proves Lemma 2.15.4 again. \(\square\)

**Proof of Theorem 2.15.3** Let

\[ C = \{i \in \{0,1,\ldots,n-1\} \mid i \perp n\}. \]

Then, Lemma 2.15.4 says that \(\phi(n) = |C|\).

Now, set

\[ z = \prod_{i \in C} i. \] (64)

Exercise 2.10.5 (applied to \(I = C\), \(c = n\) and \(b_i = i\)) yields \(\prod_{i \in C} i \perp n\) (since each \(i \in C\) satisfies \(i \perp n\)). In other words, \(z \perp n\) (since \(z = \prod_{i \in C} i\)).

We have \((ai) \% n \in C\) for each \(i \in C\).

[Proof: Let \(i \in C\). Corollary 2.6.9 (applied to \(u = ai\)) yields that \((ai) \% n \in \{0,1,\ldots,n-1\}\) and \((ai) \% n \equiv ai \mod n\). Thus, \(ai \equiv (ai) \% n \mod n\).

From \(a \perp n\) and \(i \perp n\), we obtain \(ai \perp n\) (by Theorem 2.10.9, applied to \(i\) and \(n\) instead of \(b\) and \(c\)). Hence, Exercise 2.10.6 (applied to \(ai\), \((ai) \% n\) and \(n\) instead of \(a\), \(b\) and \(c\)) yields \((ai) \% n \perp n\) (since \(ai \equiv (ai) \% n \mod n\)). Combining this with \((ai) \% n \in \{0,1,\ldots,n-1\}\), we obtain \((ai) \% n \in C\) (by the definition of \(C\)), qed.]

Thus, we can define a map

\[ f : C \to C, \]

\[ i \mapsto (ai) \% n. \]

The map \(f\) is injective.

[Proof: Let \(i\) and \(j\) be two elements of \(C\) such that \(f(i) = f(j)\). We must prove that \(i = j\).

We have \(f(i) = f(j)\). In view of \(f(i) = (ai) \% n\) (by the definition of \(f\)) and \(f(j) = (aj) \% n\), this rewrites as \((ai) \% n = (aj) \% n\). But Exercise 2.6.1 (applied to \(u = ai\) and \(v = aj\)) shows that \(ai \equiv aj \mod n\) if and only if \((ai) \% n = (aj) \% n\).

Hence, we have \(ai \equiv aj \mod n\) (since \((ai) \% n = (aj) \% n\)). By Lemma 2.10.10, we can “cancel” \(a\) from this congruence (since \(a \perp n\)), and obtain \(i \equiv j \mod n\). But both \(i\) and \(j\) belong to \(C\) and thus belong to \(\{0,1,\ldots,n-1\}\) (by the definition of \(C\)). Hence, from \(i \equiv j \mod n\), we can easily obtain that \(i = j\). \(^{58}\)

Now, forget that we fixed \(i\) and \(j\). We thus have proven that if \(i\) and \(j\) and two elements of \(C\) such that \(f(i) = f(j)\), then \(i = j\). In other words, \(f\) is injective.]

\(^{58}\)Proof. Corollary 2.6.9 (c) (applied to \(u = j\) and \(c = i\)) yields \(i = j \% n\) (since \(i \equiv j \mod n\) and \(i \in \{0,1,\ldots,n-1\}\)). But Corollary 2.6.9 (c) (applied to \(u = j\) and \(c = j\)) yields \(j = j \% n\) (since \(j \equiv j \mod n\) and \(j \in \{0,1,\ldots,n-1\}\)). Hence, \(i = j \% n = j\).
The map \( f \) is surjective.

[Proof: Let \( i \in C \). We shall prove that \( i \in f(C) \).

Indeed, \( i \in C \). By the definition of \( C \), this means that \( i \in \{0, 1, \ldots, n - 1\} \) and \( i \perp n \).

But Proposition 2.10.8 (b) shows that there exists an \( a' \in \mathbb{Z} \) such that \( aa' \equiv 1 \mod n \) (since \( a \perp n \)). Consider this \( a' \), and denote it by \( u \). Thus, \( u \) is an element of \( \mathbb{Z} \) and satisfies \( au \equiv 1 \mod n \). From \( ua = au \equiv 1 \mod n \), we conclude that there exists an \( u' \in \mathbb{Z} \) such that \( uu' \equiv 1 \mod n \) (namely, \( u' = a \)). Hence, Theorem 2.10.8 (c) (applied to \( u \) and \( u' \) instead of \( a \) and \( a' \)) shows that \( u \perp n \).

Now, Corollary 2.6.9 (a) (applied to \( ui \) instead of \( u \)) shows that \((ui) \mod n \in \{0, 1, \ldots, n - 1\}\) and \((ui) \mod n \equiv ui \mod n \). Set \( j = (ui) \mod n \). Thus, \( j = (ui) \mod n \) and \( j \equiv (ui) \mod n \). Multiplying the congruences \( a \equiv a \mod n \) and \( j \equiv ui \mod n \), we obtain

\[
aj \equiv au \mod n = 1i = i \mod n.
\]

In other words, \( i \equiv aj \mod n \). Therefore, Corollary 2.6.9 (c) (applied to \( aj \) and \( i \) instead of \( u \) and \( c \)) shows that \((aj) \mod n \in \{0, 1, \ldots, n - 1\}\) and \((aj) \mod n \equiv (aj) \mod n \).

Combining \( u \perp n \) with \( i \perp n \), we obtain \( ui \perp n \) (by Theorem 2.10.9 applied to \( u, i \) and \( n \) instead of \( a, b \) and \( c \)). Also, \( ui \equiv j \mod n \) (since \( i \equiv ui \mod n \)). Hence, Exercise 2.10.6 (applied to \( ui, j \) and \( n \) instead of \( a, b \) and \( c \)) yields \( j \perp n \). From \( j \in \{0, 1, \ldots, n - 1\} \) and \( j \perp n \), we obtain \( j \in C \) (by the definition of \( C \)). Thus, \( f(j) \) is well-defined. The definition of \( f \) yields \( f(j) = (aj) \mod n = i \) (since \( i = (aj) \mod n \)).

Hence, \( i = f\left(\begin{array}{c} j \\ \in C \end{array}\right) \in f(C) \).

Now, forget that we fixed \( i \). We thus have proven that \( i \in f(C) \) for each \( i \in C \). In other words, \( C \subseteq f(C) \). In other words, the map \( f \) is surjective.]

Now we know that the map \( f \) is injective and surjective. Hence, this map \( f \) is bijective. In other words, \( f \) is a bijection from \( C \) to \( C \). Thus, we can substitute \( f(s) \) for \( i \) in the product \( \prod_{i \in C} i \). So we obtain

\[
\prod_{i \in C} i = \prod_{s \in C} f(s) \quad \text{(65)}
\]

But for each \( s \in C \), we have

\[
f(s) = (as) \mod n \quad \text{(by the definition of \( f \))}
\]

\[
\equiv as \mod n \quad \text{(by Corollary 2.6.9 (a), applied to \( u = as \))}.
\]

Hence, \( (9) \) (applied to \( S = C, a_s = f(s) \) and \( b_s = as \)) yields

\[
\prod_{s \in C} f(s) \equiv \prod_{s \in C} (as) = a_{\mid C \mid} \prod_{s \in C} s = a_{\mid C \mid} z = a_{\phi(n)} z \mod n
\]

= \prod_{i \in C} i \quad \text{(by (64))}.
(since $|C| = \phi(n)$). Now, (64) becomes
\[
z = \prod_{i \in C} i = \prod_{s \in C} f(s) \quad \text{(by (65))}
\]
\[
\equiv a^{\phi(n)}z = za^{\phi(n)} \mod n.
\]
Thus, $z \cdot 1 = z \equiv za^{\phi(n)} \mod n$. Lemma 2.10.10 lets us “cancel” $z$ from this congruence (since $z \perp n$). We thus obtain $1 \equiv a^{\phi(n)} \mod n$. This proves Theorem 2.15.3.

Proof of Theorem 2.15.1. As we have explained above, Theorem 2.15.1 follows from Theorem 2.15.3.

Here is the argument in more detail:

(a) Assume that $p \nmid a$. Proposition 2.14.3 yields $\phi(p) = p-1$. But Proposition 2.13.5 yields that either $p \mid a$ or $p \perp a$. Hence, $p \perp a$ (since $p \nmid a$). In other words, $a$ is coprime to $p$. Hence, Theorem 2.15.3 (applied to $n = p$) yields $a^{\phi(p)} \equiv 1 \mod p$.

This rewrites as $a^{p-1} \equiv 1 \mod p$ (since $\phi(p) = p-1$). This proves Theorem 2.15.1 (a).

(b) We are in one of the following two cases:

Case 1: We have $p \nmid a$.

Case 2: We have $p \mid a$.

Let us first consider Case 1. In this case, we have $p \nmid a$. Hence, Theorem 2.15.1 (a) yields $a^{p-1} \equiv 1 \mod p$. Multiplying this congruence with the congruence $a \equiv a \mod p$, we obtain $a^{p-1}a \equiv a \equiv a \mod p$. In view of $a^{p-1}a = a^p$, this rewrites as $a^p \equiv a \mod p$. Hence, Theorem 2.15.1 (b) is proven in Case 1.

Let us now consider Case 2. In this case, we have $p \mid a$. In other words, $a \equiv 0 \mod p$. Taking this congruence to the $p$-th power, we obtain $a^p \equiv 0^p = 0 \mod p$ (since $p > 0$). Thus, $a^p \equiv 0 \equiv a \mod p$ (since $a \equiv 0 \mod p$). Hence, Theorem 2.15.1 (b) is proven in Case 2.

We have now proven Theorem 2.15.1 (b) in both Cases 1 and 2. Hence, Theorem 2.15.1 (b) always holds.

The next exercise shows an amusing (and useful) corollary of Fermat’s Little Theorem: a situation in which congruent exponents lead to congruent powers (albeit under rather specific conditions, and with the congruent powers being congruent modulo a different number than the exponents):

Exercise 2.15.2. Let $p$ be a prime. Let $a \in \mathbb{Z}$ be such that $p \nmid a$. Let $u, v \in \mathbb{N}$ satisfy $u \equiv v \mod p-1$. Then, $a^u \equiv a^v \mod p$.

2.15.3. The Pigeonhole Principles

In our above proof of Theorem 2.15.3, we have proven that the map $f : C \to C$ (that we constructed) is injective and surjective. It turns out that this was, to some extent, wasteful: It would have been enough to prove one of the two properties only (i.e., injectivity or surjectivity). The reason for this are the following two basic facts about finite sets:
Theorem 2.15.5 (Pigeonhole Principle for Injections). Let $A$ and $B$ be two finite sets such that $|A| \geq |B|$. Let $f : A \to B$ be an injective map. Then, $f$ is bijective.

Theorem 2.15.6 (Pigeonhole Principle for Surjections). Let $A$ and $B$ be two finite sets such that $|A| \leq |B|$. Let $f : A \to B$ be an surjective map. Then, $f$ is bijective.

Theorem 2.15.5 is called the Pigeonhole Principle for Injections, due to the following interpretation: If $a$ pigeons sit in $b$ pigeonholes with $a \geq b$ (that is, there are at least as many pigeons as there are pigeonholes), and if no two pigeons are sharing the same hole, then every hole must have at least one pigeon in it. (This corresponds to the statement of Theorem 2.15.5 if you let $A$ be the set of pigeons, $B$ be the set of holes, and $f$ be the map that sends each pigeon to the hole it is sitting in. The injectivity of $f$ is then precisely the statement that no two pigeons are sharing the same hole.)

Likewise, Theorem 2.15.6 is called the Pigeonhole Principle for Surjections, due to the following interpretation: If $a$ pigeons sit in $b$ pigeonholes with $a \leq b$ (that is, there are at most as many pigeons as there are pigeonholes), and if each hole contains at least one pigeon, then no two pigeons are sharing the same hole.

Theorem 2.15.5 and Theorem 2.15.6 are both basic facts of set theory; how to prove them depends on how you define the size of a finite set in the first place. See [Grinbe15, solution to Exercise 1.1] for one way of proving them (more precisely, Theorem 2.15.5 is the “$\Rightarrow$” direction of [Grinbe15, Lemma 1.5], while Theorem 2.15.6 is the “$\Leftarrow$” direction of [Grinbe15, Lemma 1.4]).

Now, Theorem 2.15.5 can be used to simplify our above proof of Theorem 2.15.3. Indeed, in the latter proof, once we have shown that $f$ is injective, we can immediately apply Theorem 2.15.5 (to $A = C$ and $B = C$) in order to conclude that $f$ is bijective (since $C$ is a finite set and satisfies $|C| \geq |C|$). The proof of surjectivity of $f$ is thus unnecessary. Alternatively, we could have omitted the proof of injectivity of $f$, and instead used the surjectivity of $f$ to apply Theorem 2.15.6 (to $A = C$ and $B = C$) in order to conclude that $f$ is bijective (since $C$ is a finite set and satisfies $|C| \leq |C|$). Either way, we would have obtained a shorter proof.

2.15.4. Wilson

The next theorem is known as Wilson’s theorem:

Theorem 2.15.7. Let $p$ be a prime. Then, $(p - 1)! \equiv -1 \mod p$.

We shall prove Theorem 2.15.7 using modular inverses modulo $p$. The main idea is that we can “pair up” each factor in the product $(p - 1)! = 1 \cdot 2 \cdot \ldots \cdot (p - 1)$ with its modular inverse modulo $p$, where of course we take the unique modular inverse that belongs to the set $\{1, 2, \ldots, p - 1\}$. This relies on the following lemma:
Lemma 2.15.8. Let $p$ be a prime. Set $A = \{1, 2, \ldots, p - 1\}$.

(a) If $a_1$ and $a_2$ are two elements of $A$ satisfying $a_1 \equiv a_2 \mod p$, then $a_1 = a_2$.
(b) For each $a \in A$, there exists a unique $a' \in A$ satisfying $aa' \equiv 1 \mod p$.
(c) Define a map $J : A \to A$ as follows: For each $a \in A$, we let $J(a)$ denote the unique $a' \in A$ satisfying $aa' \equiv 1 \mod p$. (This unique $a'$ indeed exists, by Lemma 2.10.10 (applied to $a\equiv 2 mod p$).)

Then, this map $J$ is a bijection satisfying $J \circ J = id$.

Proof of Lemma 2.15.8
(a) Let $a_1$ and $a_2$ be two elements of $A$ satisfying $a_1 \equiv a_2 \mod p$. We must prove that $a_1 = a_2$.

We have $a_1 \equiv a_2 \mod p$. Hence, Corollary 2.6.9 (c) (applied to $p$, $a_2$ and $a_1$ instead of $n$, $u$ and $c$) yields $a_1 = a_2 \mod p$ (since $a_1 \in A = \{1, 2, \ldots, p - 1\} \subseteq \{0, 1, \ldots, p - 1\}$).

Also, $a_2 \equiv a_2 \mod p$. Thus, Corollary 2.6.9 (c) (applied to $p$, $a_2$ and $a_2$ instead of $n$, $u$ and $c$) yields $a_2 = a_2 \mod p$ (since $a_2 \in A = \{1, 2, \ldots, p - 1\} \subseteq \{0, 1, \ldots, p - 1\}$).

Comparing this with $a_1 \equiv a_2 \mod p$, we obtain $a_1 = a_2$. This proves Lemma 2.15.8 (a).

(b) Let $a \in A$. Thus, $a \in A = \{1, 2, \ldots, p - 1\}$. Hence, Proposition 2.13.4 (applied to $i = 0$) shows that $a$ is coprime to $p$. In other words, $a \perp p$. Hence, Theorem 2.10.8 (a) shows that there exists a $b \in \mathbb{Z}$ such that $ab \equiv \gcd(a, p) \mod p$. Consider this $b$.

We have $ab \equiv \gcd(a, p) \equiv 1 \mod p$ (since $a \perp p$). Let $c = b \mod p$. Corollary 2.6.9 (a) (applied to $n = p$ and $u = b$) yields $b \mod p \in \{0, 1, \ldots, p - 1\}$ and $b \mod p \equiv b \mod p$.

Now, $c = b \mod p \in \{0, 1, \ldots, p - 1\}$ and $a \equiv c \equiv ab \equiv 1 \mod p$.

Assume (for the sake of contradiction) that $c = 0$. Thus, $a \equiv c \equiv 0$ and thus $0 = ac \equiv 1 \mod p$. Hence, $1 \equiv 0 \mod p$. In other words, $p \mid 1 - 0 = 1$. Hence, Exercise 2.2.5 (applied to $g = p$) yields $p = 1$. But $p > 1$ (since $p$ is prime). This contradicts $p = 1$. This contradiction shows that our assumption (that $c = 0$) is false.

Hence, $c \neq 0$. Combining this with $c \in \{0, 1, \ldots, p - 1\}$, we obtain $c \in \{0, 1, \ldots, p - 1\} \setminus \{0\} = \{1, 2, \ldots, p - 1\} = A$. Recall that $ac \equiv 1 \mod p$.

Thus, there exists at least one $a' \in A$ satisfying $aa' \equiv 1 \mod p$ (namely, $a' = c$). It remains to prove that there is only one such $a'$.

Indeed, let $a_1'$ and $a_2'$ be two elements of $A$ satisfying $aa' \equiv 1 \mod p$. We shall prove that $a_1' = a_2'$.

We know that $a_1'$ is an element of $A$ satisfying $aa' \equiv 1 \mod p$. In other words, $a_1'$ is an element of $A$ and satisfies $aa_1' \equiv 1 \mod p$. Similarly, $a_2'$ is an element of $A$ and satisfies $aa_2' \equiv 1 \mod p$. Hence, $1 \equiv aa_2' \mod p$, so that $aa_1' \equiv 1 \equiv aa_2' \mod p$. Thus, Lemma 2.10.10 (applied to $a_1'$, $a_2'$ and $p$ instead of $b$, $c$ and $n$) yields $a_1' \equiv a_2' \mod p$ (since $a \perp p$). Hence, Lemma 2.15.8 (a) (applied to $a_1 = a_1'$ and $a_2 = a_2'$) yields $a_1' = a_2'$.

Now, forget that we fixed $a_1'$ and $a_2'$. We thus have shown that if $a_1'$ and $a_2'$ are two elements of $A$ satisfying $aa' \equiv 1 \mod p$, then $a_1' = a_2'$. In other words, there exists at most one $a' \in A$ satisfying $aa' \equiv 1 \mod p$. Thus, there exists a unique such
a' (because we have already shown that there exists at least one such a'). In other words, there exists a unique a' ∈ A satisfying aa' ≡ 1 mod p. This proves Lemma 2.15.8.(b).

(c) Let a ∈ A. Then, J(a) is the unique a' ∈ A satisfying aa' ≡ 1 mod p (by the definition of J). Hence, J(a) is an a' ∈ A satisfying aa' ≡ 1 mod p. In other words, J(a) is an element of A and satisfies

\[ aJ(a) ≡ 1 \mod p. \]  \hspace{1cm} (66)

Now, forget that we fixed a. We thus have proven (66) for each a ∈ A.

Now, let a ∈ A be arbitrary. Then, J(a) ∈ A (since J is a map from A to A). Thus, (66) (applied to J(a)) instead of a yields J(a)J(J(a)) ≡ 1 mod p. Also, from J(a) ∈ A, we obtain J(J(a)) ∈ A (since J is a map from A to A). On the other hand, (66) yields aJ(a) ≡ 1 mod p. Thus, 1 ≡ aJ(a) mod p. Now,

\[ J(a)J(J(a)) ≡ 1 ≡ aJ(a) = J(a)a \mod p. \]

But J(a) ∈ A = \{1, 2, \ldots, p − 1\}. Hence, Proposition 2.13.4 (applied to i = J(a)) shows that J(a) is coprime to p. In other words, J(a) ⊥ p. Hence, Lemma 2.10.10 (applied to J(a), J(J(a)), a and p instead of a, b, c and n) yields J(J(a)) ≡ a mod p (since J(a)J(J(a)) ≡ J(a)a \mod p). Therefore, Lemma 2.15.8(a) (applied to a_1 = J(J(a)) and a_2 = a) yields J(J(a)) = a. Thus, (J ◦ J)(a) = J(J(a)) = a = id(a).

Now, forget that we fixed a. We thus have proven that (J ◦ J)(a) = id(a) for each a ∈ A. In other words, J ◦ J = id. Hence, the maps J and J are mutually inverse. Thus, the map J is invertible, i.e., is a bijection. Thus, Lemma 2.15.8(c) is proven.

\[ \square \]

Remark 2.15.9. Let S be a set. An involution on S means a map f : S → S satisfying f ◦ f = id. Thus, Lemma 2.15.8(c) says that the map J : A → A defined in this lemma is an involution on A.

We are now ready to prove Theorem 2.15.7.

First proof of Theorem 2.15.7. We have (2 − 1)! = 1! = 1 ≡ −1 \mod 2 (since 1 − (−1) = 2 is divisible by 2). In other words, Theorem 2.15.7 holds when p = 2. Hence, for the rest of this proof, we WLOG assume that we don’t have p = 2. Hence, p ≠ 2. Thus, 1 ≠ p − 1.

But p is a prime; thus, p > 1, so that p ≥ 2 (since p is an integer). Combining this with p ≠ 2, we obtain p > 2, so that p ≥ 3 (since p is an integer).

Define the set A and the map J : A → A as in Lemma 2.15.8. Hence, Lemma 2.15.8(c) shows that this map J is a bijection satisfying J ◦ J = id. The equality J ◦ J = id shows that the map J is inverse to itself. For each a ∈ A, we have

\[ aJ(a) ≡ 1 \mod p. \]  \hspace{1cm} (67)

(This congruence is proven in the same way as it was proven in our above proof of Lemma 2.15.8(c).)
Now, the rest of our proof shall follow the following plan (using the same “pairing” idea that we have seen in our proof of Proposition 2.14.7 and in the solution to Exercise 2.14.4): We will use the map \( J \) to establish a pairing between the factors of the product \( 1 \cdot 2 \cdot \cdots \cdot (p - 1) \) (pairing up each factor \( a \) with the factor \( J(a) \)), which will pair up almost all of them – more precisely, all of them except for the very first and very last factors (since these two factors would have to pair up with themselves). For example, if \( p = 11 \), then we have the following table of values of \( J \):

<table>
<thead>
<tr>
<th>( a )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J(a) )</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

(Indeed, for example, \( J(2) = 6 \), since 6 is the unique \( a' \in A \) satisfying \( 2 \cdot a' \equiv 1 \mod 11 \), and thus we pair up the factors of the product \( 1 \cdot 2 \cdot \cdots \cdot (p - 1) \) as follows:

\[
1 \cdot 2 \cdot \cdots \cdot (p - 1) = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10
\]

\[
= 1 \cdot (2 \cdot 6) \cdot (3 \cdot 4) \cdot (5 \cdot 9) \cdot (7 \cdot 8) \cdot 10.
\]

By the definition of the map \( J \), each pair has the form \((a, J(a))\) for some \( a \in A \), and thus the product of any two different factors paired up with each other is \( \equiv 1 \mod p \) (by (67)). For example, if \( p = 11 \), then we have

\[
1 \cdot 2 \cdot \cdots \cdot (p - 1) \equiv 1 \mod 11 \quad 1 \cdot 10 \mod 11.
\]

Thus, any two different factors paired up with each other “neutralize” each other when being multiplied (as long as we are computing modulo \( p \)). Hence, the product of all the \( p - 1 \) factors will reduce (when working modulo \( p \)) to the product of the two factors that have not been paired up, which will be \( 1 \cdot (p - 1) = p - 1 \equiv -1 \mod p \).

Here are the details of this argument:

An element \( a \) of \( A \) will be called

- **small** if \( a < J(a) \);
- **medium** if \( a = J(a) \);
- **large** if \( a > J(a) \).

Now, we claim that the medium elements of \( A \) are precisely 1 and \( p - 1 \).

\[\text{59 The reason why it is precisely these two factors that will not be paired up is not completely trivial. It follows from Exercise 2.13.12.}\]
Proof: We have $1 \leq p - 1$ (since $p \geq 2$). Thus, $1 \in \{1,2,\ldots,p-1\} = A$ and $p-1 \in \{1,2,\ldots,p-1\} = A$. The element 1 of $A$ is medium. The element $p-1$ of $A$ is medium. Hence, the two numbers 1 and $p-1$ are medium elements of $A$. It remains to prove that these two numbers are the only medium elements of $A$.

Indeed, let $a$ be a medium element of $A$. We shall show that $a = 1$ or $a = p - 1$.

Indeed, assume the contrary. Thus, neither $a = 1$ nor $a = p - 1$ holds.

If we had $a \equiv 1 \mod p$, then Lemma 2.15.8 (a) (applied to $a_1 = a$ and $a_2 = 1$) would yield $a = 1$, which would contradict the fact that $a = 1$ does not hold. Thus, we do not have $a \equiv 1 \mod p$.

If we had $a \equiv p-1 \mod p$, then Lemma 2.15.8 (a) (applied to $a_1 = a$ and $a_2 = p-1$) would yield $a = p - 1$, which would contradict the fact that $a = p - 1$ does not hold. Thus, we do not have $a \equiv p-1 \mod p$.

We have assumed that $a$ is medium. In other words, $a = J(a)$. But 67 yields $aJ(a) \equiv 1 \mod p$. Thus, $a^2 = a \cdot \frac{a}{a} = aJ(a) \equiv 1 \mod p$. Hence, Exercise 2.13.12 shows that $a \equiv 1$ or $a \equiv -1 \mod p$. Hence, we must have $a \equiv -1 \mod p$ (since we do not have $a \equiv 1 \mod p$). Thus, $a \equiv -1 \equiv p-1 \mod p$ (since $p-1 \equiv -1 \mod p$). This contradicts the fact that we do not have $a \equiv p-1 \mod p$.

This contradiction shows that our assumption was false. Hence, $a = 1$ or $a = p - 1$.

Now, forget that we fixed $a$. We thus have proven that every medium element $a$ of $A$ satisfies $a = 1$ or $a = p - 1$. In other words, every medium element of $A$ is either 1 or $p-1$. Since we know that 1 and $p-1$ actually are medium elements of $A$, we thus conclude that the medium elements of $A$ are precisely 1 and $p-1$.

So we have shown that the medium elements of $A$ are precisely 1 and $p - 1$. Since these two elements are distinct (because $p - 1 \neq 1$), we thus obtain

$$\prod_{a \in A; \text{a is medium}} a = 1 \cdot (p-1) = p - 1 \equiv -1 \mod p. \tag{68}$$

Proof. We have $1 \in A$. Hence, $J(1) \in A$ (since $J$ is a map from $A$ to $A$). Furthermore, 67 (applied to $a = 1$) yields $1J(1) \equiv 1 \mod p$. Thus, $1 \equiv 1J(1) = J(1) \mod p$. Thus, Lemma 2.15.8 (a) (applied to $a_1 = 1$ and $a_2 = J(1)$) yields $1 = J(1)$. In other words, the element 1 of $A$ is medium.

Proof. We have $p - 1 \in A$. Hence, $J(p-1) \in A$ (since $J$ is a map from $A$ to $A$). Furthermore, 67 (applied to $a = p - 1$) yields $(p-1)J(p-1) \equiv 1 \mod p$. Multiplying this congruence with the obvious congruence $p-1 \equiv p - 1 \mod p$, we obtain

$$(p-1)(p-1)J(p-1) \equiv (p-1)1 = p - 1 \mod p.$$

Hence,

$$p - 1 \equiv (p-1)(p-1)J(p-1) = \left(\frac{p-1}{\equiv -1 \mod p}\right)^2 J(p-1) \equiv (-1)^2 J(p-1) \equiv 1 = J(p-1) \mod p.$$

Thus, Lemma 2.15.8 (a) (applied to $a_1 = p - 1$ and $a_2 = J(p-1)$) yields $p - 1 = J(p-1)$. In other words, the element $p - 1$ of $A$ is medium.
It is easy to see that if \( a \) is a small element of \( A \), then \( J(a) \) is a large element of \( A \). Hence, the map

\[
J^+: \{ \text{small elements of } A \} \to \{ \text{large elements of } A \},
\]

\[
a \mapsto J(a)
\]
is well-defined. Similarly, the map

\[
J^-: \{ \text{large elements of } A \} \to \{ \text{small elements of } A \},
\]

\[
a \mapsto J(a)
\]
is well-defined. These two maps \( J^+ \) and \( J^- \) are both restrictions of the map \( J \), and thus are mutually inverse (since the map \( J \) is inverse to itself). Hence, the map \( J^+ \) is invertible, i.e., is a bijection. In other words, the map

\[
\{ \text{small elements of } A \} \to \{ \text{large elements of } A \},
\]

\[
a \mapsto J(a)
\]
is a bijection (since this map is just the map \( J^+ \)). Thus, we can substitute \( J(b) \) for \( a \) in the product \( \prod_{a \in A; \text{a is large}} a \). We thus obtain

\[
\prod_{a \in A; \text{a is large}} a = \prod_{b \in A; \text{b is small}} J(b) = \prod_{a \in A; \text{a is small}} J(a)
\]  

(69)

(here, we have renamed the index \( b \) as \( a \) in the product). Now, the definition of \( (p-1)! \)

---

\textit{Proof.} Let \( a \) be a small element of \( A \). Thus, \( a < J(a) \). Note that \( J(a) \in A \) (since \( J \) is a map from \( A \) to \( A \)). But \( J \circ J = \text{id} \), so that \( (J \circ J)(a) = \text{id} a = a < J(a) \). In view of \( (J \circ J)(a) = J(J(a)) \), this rewrites as \( J(J(a)) < J(a) \). In other words, \( J(a) > J(J(a)) \). In other words, the element \( J(a) \) of \( A \) is large (by the definition of “large”). Qed.
yields

\[(p - 1)! = 1 \cdot 2 \cdot \ldots \cdot (p - 1) = \prod_{a \in A} a\]

\[
= \left( \prod_{\substack{a \in A; \ a \text{ is small}}} a \right) \cdot \left( \prod_{\substack{a \in A; \ a \text{ is medium}}} a \right) \cdot \left( \prod_{\substack{a \in A; \ a \text{ is large}}} a \right)
\]

\[
\equiv -1 \mod p \quad \text{(by (68))}
\]

\[
= \prod_{a \in A; \ a \text{ is small}} J(a)
\]

\[
= \prod_{a \in A; \ a \text{ is small}} (aJ(a)) \mod p. \quad \text{(70)}
\]

But it is clear that

\[
\prod_{a \in A; \ a \text{ is small}} (aJ(a)) \equiv 1 \mod p \quad \text{(by (67))}
\]

Hence, (70) rewrites as

\[
(p - 1)! \equiv - \prod_{a \in A; \ a \text{ is small}} (aJ(a)) \equiv -1 \mod p. \quad \text{(71)}
\]

\[63\text{Proof. Here is this argument in more detail:}\]

Every \(a \in \{ \text{small elements of } A \} \) satisfies \(aJ(a) \equiv 1 \mod p \) (by (67)). Renaming the index \(a\) as \(s\) in this statement, we obtain the following: Every \(s \in \{ \text{small elements of } A \} \) satisfies \(sJ(s) \equiv 1 \mod p \). Hence, (9) (applied to \(n = p, S = \{ \text{small elements of } A \}, a_s = sJ(s)\) and \(b_s = 1\)) yields

\[
\prod_{s \in \{ \text{small elements of } A \}} (sJ(s)) \equiv 1 \mod p.
\]

In view of

\[
\prod_{s \in \{ \text{small elements of } A \}} (sJ(s)) = \prod_{a \in \{ \text{small elements of } A \}} (aJ(a)) \quad \text{(here, we have renamed the index } s \text{ as } a \text{ in the product)}
\]

this rewrites as \(\prod_{a \in A; \ a \text{ is small}} (aJ(a)) \equiv 1 \mod p.\)
This proves Theorem \textcolor{red}{2.15.7}

Later, in Section \textcolor{red}{3.5}, we shall give a different version of this proof.

Theorem \textcolor{red}{2.15.7} has a converse:

\textbf{Exercise 2.15.3.} If an integer \( p > 1 \) satisfies \((p - 1)! \equiv -1 \mod p\), then prove that \( p \) is a prime.

(This is actually easier to prove than Theorem \textcolor{red}{2.15.7} itself.)

\textbf{Exercise 2.15.4.} Let \( p \) be a prime. Prove that
\[
(p - 1)! \equiv p - 1 \mod 1 + 2 + \cdots + (p - 1).
\]

\textbf{Exercise 2.15.5.} Let \( p \) be an odd prime. Write \( p \) in the form \( p = 2k + 1 \) for some \( k \in \mathbb{N} \). Prove that \( k!^2 \equiv -(-1)^k \mod p \).

[\textbf{Hint:} Each \( j \in \mathbb{Z} \) satisfies \( j \cdot (p - j) \equiv -j^2 \mod p \).]

\section*{2.16. The Chinese Remainder Theorem as a bijection}

\subsection*{2.16.1. The bijection \( K_{m,n} \)}

Here comes another of the many facts known as the “Chinese Remainder Theorem”:

\textbf{Theorem 2.16.1.} Let \( m \) and \( n \) be two coprime positive integers. Then, the map
\[
K_{m,n} : \{0, 1, \ldots, mn - 1\} \rightarrow \{0, 1, \ldots, m - 1\} \times \{0, 1, \ldots, n - 1\},
\]
\[
a \mapsto (a \% m, a \% n)
\]
is well-defined and is a bijection.

\textbf{Example 2.16.2. (a)} Theorem \textcolor{red}{2.16.1} (applied to \( m = 3 \) and \( n = 2 \)) says that the map
\[
K_{3,2} : \{0, 1, 2, 3, 4, 5\} \rightarrow \{0, 1, 2\} \times \{0, 1\},
\]
\[
a \mapsto (a \% 3, a \% 2)
\]
is a bijection. This map sends
\[
0, 1, 2, 3, 4, 5 \text{ to } (0, 0), (1, 1), (2, 0), (0, 1), (1, 0), (2, 1),
\]
respectively (since \( 0 \% 3 = 0 \) and \( 0 \% 2 = 0 \) and \( 1 \% 3 = 1 \) and \( 1 \% 2 = 1 \) and \( 2 \% 3 = 2 \) and \( 2 \% 2 = 0 \) and so on). This list of values shows that this map is bijective (since
it takes on every possible value in \( \{0,1,2\} \times \{0,1\} \) exactly once. Theorem \textit{2.16.1} says that this holds for arbitrary coprime \( m \) and \( n \).

(b) Let us see how Theorem \textit{2.16.1} fails when \( m \) and \( n \) are not coprime. For example, take \( m = 6 \) and \( n = 4 \). Then, the map

\[
K_{6,4} : \{0,1,\ldots,23\} \to \{0,1,2,3,4,5\} \times \{0,1,2,3\}, \\
a \mapsto (a \% 6, a \% 4)
\]

is not a bijection. Indeed, it is neither injective (for example, it sends both 0 and 12 to the same pair \((0,0)\)) nor surjective (for example, it never takes the value \((1,2)\)).

\textbf{Proof of Theorem \textit{2.16.1}} For every \( a \in \{0,1,\ldots, mn - 1\} \), we have \((a \% m, a \% n) \in \{0,1,\ldots, m - 1\} \times \{0,1,\ldots, n - 1\} \). Hence, the map \( K_{m,n} \) is well-defined. It remains to prove that this map \( K_{m,n} \) is a bijection. To that aim, we shall prove that \( K_{m,n} \) is injective and surjective.

[Proof that the map \( K_{m,n} \) is injective: Let \( a, b \in \{0,1,\ldots, mn - 1\} \) be such that \( K_{m,n} (a) = K_{m,n} (b) \). We want to prove \( a = b \).

The definition of \( K_{m,n} \) yields \( K_{m,n} (a) = (a \% mn, a \% mn) \) and \( K_{m,n} (b) = (b \% mn, b \% mn) \). Hence, the equality \( K_{m,n} (a) = K_{m,n} (b) \) (which is true by assumption) rewrites as \((a \% m, a \% n) = (b \% m, b \% n)\). In other words, \( a \% m = b \% m \) and \( a \% n = b \% n \).

Now, Exercise \textit{2.6.1} (applied to \( u = a \) and \( v = b \)) yields that \( a \equiv b \mod n \) if and only if \( a \% n = b \% n \). Hence, \( a \equiv b \mod n \) (since \( a \% n = b \% n \)). In other words, \( n \mid a - b \).

Now, we have \( m \perp n \) (since \( m \) and \( n \) are coprime) and \( m \mid a - b \) and \( n \mid a - b \). Hence, Theorem \textit{2.10.7} (applied to \( m \), \( n \) and \( a - b \) instead of \( a \), \( b \) and \( c \)) yields \( mn \mid a - b \). In other words, \( a \equiv b \mod mn \). Hence, Corollary \textit{2.6.9} (c) (applied to \( mn \), \( b \) and \( a \) instead of \( n \), \( u \) and \( c \)) yields \( a = b \% (mn) \) (since \( a \in \{0,1,\ldots, mn - 1\} \)).

On the other hand, \( b \equiv b \mod mn \). Hence, Corollary \textit{2.6.9} (c) (applied to \( mn \), \( b \) and \( b \) instead of \( n \), \( u \) and \( c \)) yields \( b = b \% (mn) \) (since \( b \in \{0,1,\ldots, mn - 1\} \)). Comparing this with \( a = b \% (mn) \), we obtain \( a = b \).

Now, forget that we fixed \( a \) and \( b \). We thus have shown that if \( a, b \in \{0,1,\ldots, mn - 1\} \) are such that \( K_{m,n} (a) = K_{m,n} (b) \), then \( a = b \). In other words, the map \( K_{m,n} \) is injective.]

[Proof that the map \( K_{m,n} \) is surjective: Fix \((a,b) \in \{0,1,\ldots, mn - 1\} \times \{0,1,\ldots, n - 1\} \). We want to find a \( c \in \{0,1,\ldots, mn - 1\} \) such that \( K_{m,n} (c) = (a,b) \).

We have \((a,b) \in \{0,1,\ldots, mn - 1\} \times \{0,1,\ldots, n - 1\} \). In other words, \( a \in \{0,1,\ldots, mn - 1\} \) and \( b \in \{0,1,\ldots, n - 1\} \). Theorem \textit{2.12.1} (a) shows that there exists an integer

\textit{Proof.} Let \( a \in \{0,1,\ldots, mn - 1\} \). We must prove that \((a \% m, a \% n) \in \{0,1,\ldots, m - 1\} \times \{0,1,\ldots, n - 1\} \).

Corollary \textit{2.6.9} (a) (applied to \( u = a \)) yields that \( a \% n \in \{0,1,\ldots, n - 1\} \) and \( a \% m \equiv a \mod m \). Thus, \( a \% n \in \{0,1,\ldots, n - 1\} \). The same argument (applied to \( m \) instead of \( n \)) yields \( a \% m \in \{0,1,\ldots, m - 1\} \). Combining this with \( a \% n \in \{0,1,\ldots, n - 1\} \), we obtain \((a \% m, a \% n) \in \{0,1,\ldots, m - 1\} \times \{0,1,\ldots, n - 1\} \). Qed.
x ∈ ℤ such that

\[ x \equiv a \mod m \text{ and } x \equiv b \mod n. \]

Consider such an x. We have x ≡ a mod m, thus a ≡ x mod m. From x ≡ b mod n, we obtain b ≡ x mod n.

Let y = x % (mn). Then, Corollary \[2.6.9\] (a) (applied to \(mn\) and x instead of n and u) yields \(x\% (mn) \in \{0, 1, \ldots, mn - 1\}\) and \(x\% (mn) \equiv x \mod mn\). Hence, \(x \equiv x\% (mn) = y \mod mn\) (since \(y = x\% (mn)\)).

Since \(m \mid mn\), we thus obtain \(x \equiv y \mod m\) (by Proposition \[2.3.4\] (e), applied to \(mn\), x, y and m instead of n, a, b and m). Thus, \(a \equiv x \equiv y \mod m\). Hence, Corollary \[2.6.9\] (c) (applied to m, y and a instead of n, u and c) yields \(a = y\%m\) (since \(a \in \{0, 1, \ldots, m - 1\}\)).

Also, from \(x \equiv y \mod mn\) and \(n \mid mn\), we obtain \(x \equiv y \mod n\) (by Proposition \[2.3.4\] (e), applied to \(mn\), x, y and n instead of n, a, b and m). Thus, \(b \equiv x \equiv y \mod n\). Hence, Corollary \[2.6.9\] (c) (applied to y and b instead of u and c) yields \(b = y\%n\) (since \(b \in \{0, 1, \ldots, n - 1\}\)).

From \(a = y\%m\) and \(b = y\%n\), we obtain \((a, b) = (y\%m, y\%n)\).

We have \(y \in x\% (mn) \in \{0, 1, \ldots, mn - 1\}\); thus, the definition of the map \(K_{m,n}\) yields \(K_{m,n}(y) = (y\%m, y\%n) = (a, b)\) (since \((a, b) = (y\%m, y\%n)\)). Thus, there exists a \(c \in \{0, 1, \ldots, mn - 1\}\) such that \(K_{m,n}(c) = (a, b)\) (namely, \(c = y\)).

Now, forget that we fixed \((a, b)\). We thus have shown that for any \((a, b) \in \{0, 1, \ldots, m - 1\} \times \{0, 1, \ldots, n - 1\}\), there exists a \(c \in \{0, 1, \ldots, mn - 1\}\) such that \(K_{m,n}(c) = (a, b)\). In other words, the map \(K_{m,n}\) is surjective.

We have now proven that the map \(K_{m,n}\) is both injective and surjective. Hence, this map \(K_{m,n}\) is bijective, i.e., is a bijection. This proves Theorem \[2.16.1\].

[Remark: We could have saved ourselves some of the work done in this proof by invoking the Pigeonhole Principle. Indeed, our goal was to show that the map \(K_{m,n}\) is bijective. By the Pigeonhole Principle for Injections (Theorem \[2.15.5\]), it suffices to prove that it is injective, because \(\{0, 1, \ldots, mn - 1\}\) and \(\{0, 1, \ldots, m - 1\} \times \{0, 1, \ldots, n - 1\}\) are two finite sets of the same size. Alternatively, by the Pigeonhole Principle for Surjections (Theorem \[2.15.6\]), it would instead suffice to prove that the map \(K_{m,n}\) is surjective].

### 2.16.2. Coprime remainders

For the rest of this section, we shall use the following notation:

#### Definition 2.16.3. Let \(n\) be a positive integer. Then, let \(C_n\) be the subset \(\{i \in \{0, 1, \ldots, n - 1\} \mid i \perp n\}\) of \(\{0, 1, \ldots, n - 1\}\).

Now, we claim the following:

#### Proposition 2.16.4. Let \(m\) and \(n\) be two coprime positive integers. Consider the map \(K_{m,n}\) defined in Theorem \[2.16.1\]. Then,

\[ K_{m,n}(C_{mn}) = C_m \times C_n. \]
(Here, $K_{m,n}(C_{mn})$ denotes the image of the subset $C_{mn}$ of $\{0,1,\ldots,mn-1\}$ under the map $K_{m,n}$; that is, $K_{m,n}(C_{mn}) = \{K_{m,n}(x) \mid x \in C_{mn}\}$.)

**Example 2.16.5.** Theorem 2.16.1 (applied to $m = 3$ and $n = 5$) says that the map

$$K_{3,5} : \{0,1,\ldots,14\} \to \{0,1,2\} \times \{0,1,2,3,4\},$$

$$a \mapsto (a\%3, a\%5)$$

is a bijection. Proposition 2.16.4 (applied to $m = 3$ and $n = 5$) says that this map satisfies $K_{3,5}(C_{15}) = C_3 \times C_5$. In view of

- $C_{15} = \{i \in \{0,1,\ldots,14\} \mid i \perp 15\} = \{1,2,4,7,8,11,13,14\}$,
- $C_3 = \{i \in \{0,1,2\} \mid i \perp 3\} = \{1,2\}$, and
- $C_5 = \{i \in \{0,1,2,3,4\} \mid i \perp 5\} = \{1,2,3,4\}$,

this rewrites as

$$K_{3,5}(\{1,2,4,7,8,11,13,14\}) = \{1,2\} \times \{1,2,3,4\}.$$

And indeed, this can easily be checked: The map $K_{3,5}$ sends

- $1, 2, 4, 7, 8, 11, 13, 14,$ to
- $(1,1), (2,2), (1,4), (1,2), (2,3), (2,1), (1,3), (2,4)$,

respectively, which entails

$$K_{3,5}(\{1,2,4,7,8,11,13,14\}) = \{1,2\} \times \{1,2,3,4\}.$$

**Proof of Proposition 2.16.4** Theorem 2.16.1 yields that the map $K_{m,n}$ is well-defined and is a bijection.

The definition of $C_n$ yields

$$C_n = \{i \in \{0,1,\ldots,n-1\} \mid i \perp n\} \subseteq \{0,1,\ldots,n-1\}.$$  

The definition of $C_m$ yields

$$C_m = \{i \in \{0,1,\ldots,m-1\} \mid i \perp m\} \subseteq \{0,1,\ldots,m-1\}.$$  

The definition of $C_{mn}$ yields

$$C_{mn} = \{i \in \{0,1,\ldots,mn-1\} \mid i \perp mn\} \subseteq \{0,1,\ldots,mn-1\}.$$  

Hence, $K_{m,n}(C_{mn})$ is well-defined. Also, from $C_m \subseteq \{0,1,\ldots,m-1\}$ and $C_n \subseteq \{0,1,\ldots,n-1\}$, we obtain

$$C_m \times C_n \subseteq \{0,1,\ldots,m-1\} \times \{0,1,\ldots,n-1\}.$$
Now, we claim that
\[ K_{m,n} (C_{mn}) \subseteq C_m \times C_n. \tag{72} \]

[Proof of (72): Let \( z \in K_{m,n} (C_{mn}) \). Thus, \( z = K_{m,n} (x) \) for some \( x \in C_{mn} \). Consider this \( x \).

We have \( x \in C_{mn} = \{ i \in \{0,1,\ldots,mn-1\} \mid i \perp mn \} \). In other words, \( x \) is an \( i \in \{0,1,\ldots,mn-1\} \) satisfying \( i \perp mn \). In other words, \( x \) is an element of \( \{0,1,\ldots,mn-1\} \) and satisfies \( x \perp mn \). In other words, \( x \perp mn \) (since \( mn = mn \)).

We have \( x \mid x \) and \( m \mid mn \). Hence, Exercise 2.9.4 (applied to \( a_1 = x \), \( a_2 = m \), \( b_1 = x \) and \( b_2 = mn \)) yields \( \gcd (x,m) \mid \gcd (x,mn) = 1 \) (since \( x \perp mn \)). Since \( \gcd (x,m) \) is a nonnegative integer \(^{65}\) this entails \( \gcd (x,m) = 1 \) (by Exercise 2.2.5, applied to \( g = \gcd (x,m) \)). In other words, \( x \perp m \). But Corollary 2.6.9 (a) (applied to \( m \) and \( x \) instead of \( n \) and \( u \)) yields \( x \% m \in \{0,1,\ldots,m-1\} \) and \( x \% m \equiv x \mod m \). From \( x \% m \equiv x \mod m \), we obtain \( x \equiv x \% m \mod m \). Hence, Exercise 2.10.6 (applied to \( x \), \( x \% m \) and \( m \) instead of \( a \), \( b \) and \( c \)) yields \( x \% m \perp m \) (since \( x \perp m \)). Hence, \( x \% m \) is an \( i \in \{0,1,\ldots,m-1\} \) satisfying \( i \perp m \) (since \( x \% m \in \{0,1,\ldots,m-1\} \)). In other words, \( x \% m \in \{ i \in \{0,1,\ldots,m-1\} \mid i \perp m \} \). In other words, \( x \% m \in C_m \) (since \( C_m = \{ i \in \{0,1,\ldots,m-1\} \mid i \perp m \} \)).

The same argument (with the roles of \( m \) and \( n \) swapped) yields \( x \% n \in C_n \) (since \( x \perp mn \)). Now,

\[ z = K_{m,n} (x) = (x \% m, x \% n) \quad (\text{by the definition of } K_{m,n}) \]
\[ \in C_m \times C_n \quad (\text{since } x \% m \in C_m \text{ and } x \% n \in C_n). \]

Now, forget that we fixed \( z \). We thus have proven that \( z \in C_m \times C_n \) for each \( z \in K_{m,n} (C_{mn}) \). In other words, \( K_{m,n} (C_{mn}) \subseteq C_m \times C_n \). This proves (72).

Next, we claim that
\[ C_m \times C_n \subseteq K_{m,n} (C_{mn}). \tag{73} \]

[Proof of (73): Let \( y \in C_m \times C_n \). We shall prove that \( y \in K_{m,n} (C_{mn}) \).

We have \( y \in C_m \times C_n \subseteq \{0,1,\ldots,m-1\} \times \{0,1,\ldots,n-1\} = K_{m,n} (\{0,1,\ldots,mn-1\}) \) (since the map \( K_{m,n} \) is a bijection). In other words, there exists some \( x \in \{0,1,\ldots,mn-1\} \) such that \( y = K_{m,n} (x) \). Consider this \( x \). The definition of \( K_{m,n} \) yields \( K_{m,n} (x) = (x \% m, x \% n) \). Hence,

\[ (x \% m, x \% n) = K_{m,n} (x) = y \in C_m \times C_n. \]

In other words, \( x \% m \in C_m \) and \( x \% n \in C_n \).

We have \( x \% m \in C_m = \{ i \in \{0,1,\ldots,m-1\} \mid i \perp m \} \). In other words, \( x \% m \) is an \( i \in \{0,1,\ldots,m-1\} \) satisfying \( i \perp m \). In other words, \( x \% m \) is an element of \( \{0,1,\ldots,m-1\} \) and satisfies \( x \% m \perp m \).

But Corollary 2.6.9 (a) (applied to \( m \) and \( x \) instead of \( n \) and \( u \)) yields \( x \% m \in \{0,1,\ldots,m-1\} \) and \( x \% m \equiv x \mod m \). Hence, Exercise 2.10.6 (applied to \( x \% m \), \( x \) and \( m \) instead of \( a \), \( b \) and \( c \)) yields \( x \perp m \) (since \( x \% m \perp m \)). This yields \( m \perp x \)

\(^{65}\text{because any gcd is a nonnegative integer}\)
(by Proposition 2.10.4). The same argument (applied to \(n\) instead of \(m\)) yields \(n \perp x\) (since \(x \% n \in C_n\)). Hence, Theorem 2.10.9 (applied to \(m, n\) and \(x\) instead of \(a, b\) and \(c\)) yields \(mn \perp x\). This yields \(x \perp mn\) (by Proposition 2.10.4). Thus, \(x\) is an \(i \in \{0, 1, \ldots, mn - 1\}\) satisfying \(i \perp mn\) (since \(x \in \{0, 1, \ldots, mn - 1\}\)). In other words, \(x \in \{i \in \{0, 1, \ldots, mn - 1\} \mid i \perp mn\}\). In other words, \(x \in C_{mn}\) (since \(C_{mn} = \{i \in \{0, 1, \ldots, mn - 1\} \mid i \perp mn\}\)). Hence, \(y = K_{m,n} \begin{pmatrix} x \\ \in C_{mn} \end{pmatrix} \in K_{m,n}(C_{mn})\).

Now, forget that we fixed \(y\). We thus have shown that \(y \in K_{m,n}(C_{mn})\) for each \(y \in C_m \times C_n\). In other words, \(C_m \times C_n \subseteq K_{m,n}(C_{mn})\). This proves (73).

Combining (72) with (73), we obtain \(K_{m,n}(C_{mn}) = C_m \times C_n\). This proves Proposition 2.16.4.

2.16.3. Proving the formula for \(\phi\)

We now can prove Theorem 2.14.4:

First proof of Theorem 2.14.4 The definition of \(C_n\) yields

\[
C_n = \{i \in \{0, 1, \ldots, n - 1\} \mid i \perp n\}.
\]

Lemma 2.15.4 yields

\[
\phi(n) = \left| \{i \in \{0, 1, \ldots, n - 1\} \mid i \perp n\} \right| = |C_n|.
\]

The same argument (applied to \(mn\) instead of \(n\)) yields \(\phi(mn) = |C_{mn}|\).

We have shown that \(\phi(n) = |C_n|\), so that \(|C_n| = \phi(n)\). The same argument (applied to \(m\) instead of \(n\)) yields \(|C_m| = \phi(m)\).

It is well-known that any two finite sets \(A\) and \(B\) satisfy \(|A \times B| = |A| \cdot |B|\) \(^{66}\). Applying this to \(A = C_m\) and \(B = C_n\), we obtain

\[
|C_m \times C_n| = |C_m| \cdot |C_n| = \phi(m) \cdot \phi(n).
\]

Note that \(C_{mn}\) is a subset of \(\{0, 1, \ldots, mn - 1\}\) (since the definition of \(C_{mn}\) yields \(C_{mn} = \{i \in \{0, 1, \ldots, mn - 1\} \mid i \perp mn\} \subseteq \{0, 1, \ldots, mn - 1\}\)).

Now, consider the map \(K_{m,n}\) defined in Theorem 2.16.1. Then, Theorem 2.16.1 shows that this map \(K_{m,n}\) is a bijection. Thus, in particular, \(K_{m,n}\) is injective. Hence,

\(^{66}\)This is the so-called **product rule** in its simplest form (see, e.g., [Loehr11] 1.5) or [LeLeMe18] §15.2.1).
$|K_{m,n}(S)| = |S|$ for each subset $S$ of $\{0,1,\ldots,mn-1\}$. Applying this to $S = C_{mn}$, we obtain $|K_{m,n}(C_{mn})| = |C_{mn}|$. Thus,

$$|C_{mn}| = \left| \frac{K_{m,n}(C_{mn})}{C_{mn}} \right| = |C_m \times C_n| = \phi(m) \cdot \phi(n).$$

Hence, $\phi(mn) = |C_{mn}| = \phi(m) \cdot \phi(n)$. This proves Theorem 2.14.4.

We now take aim at proving Theorem 2.14.5. First, let us extend Theorem 2.14.4 to products of $k$ mutually coprime integers:

**Exercise 2.16.1.** Let $n_1, n_2, \ldots, n_k$ be mutually coprime positive integers. Prove that $\phi(n_1n_2 \cdots n_k) = \phi(n_1) \cdot \phi(n_2) \cdots \phi(n_k)$.

**Exercise 2.16.2.** Let $I$ be a finite set. For each $i \in I$, let $n_i$ be an integer. Assume that every two distinct elements $i$ and $j$ of $I$ satisfy $n_i \perp n_j$. (75) Prove that

$$\phi\left(\prod_{i \in I} n_i\right) = \prod_{i \in I} \phi(n_i).$$

We are finally ready to prove Theorem 2.14.5.

**Proof of Theorem 2.14.5** If $p$ is a prime satisfying $p \nmid n$, then $v_p(n) = 0$ (by Corollary 2.13.26) and therefore

$$p^{v_p(n)} = p^0 = 1.$$ (76)

If $p$ is a prime satisfying $p \mid n$, then $v_p(n)$ is a positive integer and therefore we have

$$\phi\left(p^{v_p(n)}\right) = (p-1) p^{v_p(n)-1}$$ (77)

(by Exercise 2.14.1, applied to $k = v_p(n)$).

---

67 This follows from the following general principle: If $f : X \to Y$ is an injective map between two finite sets $X$ and $Y$, then $|f(S)| = |S|$ for each subset $S$ of $X$.

68 Proof. Let $p$ be a prime satisfying $p \mid n$. If we had $v_p(n) = 0$, then we would have $p \nmid n$ (by Corollary 2.13.26), and this would contradict $p \mid n$. Hence, we cannot have $v_p(n) = 0$. Thus, $v_p(n) \neq 0$. But $v_p(n) \in \mathbb{N}$ (since $n$ is nonzero). Thus, from $v_p(n) \neq 0$, we conclude that $v_p(n)$ is a positive integer. Qed.
Corollary 2.13.33 yields

\[ n = \prod_{p \text{ prime}} p^{v_p(n)} = \left( \prod_{p \text{ prime; } p \nmid n} p^{v_p(n)} \right) \cdot \left( \prod_{p \text{ prime; } p|n} p^{v_p(n)} \right) \]

(since each prime \( p \) satisfies either \( p \mid n \) or \( p \nmid n \), but not both at the same time)

\[ = \left( \prod_{p \text{ prime; } p|n} p^{v_p(n)} \right) \cdot \left( \prod_{p \text{ prime; } p \nmid n} 1 \right) = \prod_{p \text{ prime; } p|n} p^{v_p(n)}. \quad (78) \]

Let \( P \) be the set of all primes \( p \) satisfying \( p \mid n \). This set \( P \) is finite\(^{69}\). For each \( i \in P \), the number \( i^{v_i(n)} \) is an integer (since \( v_i(n) \in \mathbb{N} \) (since \( n \) is nonzero)). Moreover, every two distinct elements \( i \) and \( j \) of \( P \) satisfy \( i^{v_i(n)} \perp j^{v_j(n)} \). Hence, Exercise 2.16.2 (applied to \( I = P \) and \( n_i = i^{v_i(n)} \)) yields

\[ \phi \left( \prod_{i \in P} i^{v_i(n)} \right) = \prod_{i \in P} \phi \left( i^{v_i(n)} \right). \]

Renaming the index \( i \) as \( p \) in both products, we can rewrite this equality as

\[ \phi \left( \prod_{p \in P} p^{v_p(n)} \right) = \prod_{p \in P} \phi \left( p^{v_p(n)} \right). \quad (79) \]

But the product signs “\( \prod \)” in this equality can be replaced by “\( \prod \)” without changing their meaning (since \( P \) is the set of all primes \( p \) satisfying \( p \mid n \)). Hence,

\[^{69}\text{Proof.} \text{ We shall show that } P \subseteq \{1, 2, \ldots, n\}. \]

Indeed, let \( p \in P \). Thus, \( p \) is a prime satisfying \( p \mid n \) (by the definition of \( P \)). Hence, \( p \) is positive (since \( p \) is prime); thus, \( p \neq 0 \). Thus, from \( p \mid n \), we obtain \( |p| \leq |n| \) (by Proposition 2.2.3(b), applied to \( a = p \) and \( b = n \)). In view of \( |p| = p \) (since \( p \) is positive) and \( |n| = n \) (since \( n \) is positive), this rewrites as \( p \leq n \). Hence, \( p \in \{1, 2, \ldots, n\} \) (since \( p \) is a positive integer).

Now, forget that we fixed \( p \). We thus have shown that \( p \in \{1, 2, \ldots, n\} \) for each \( p \in P \). In other words, \( P \subseteq \{1, 2, \ldots, n\} \). Hence, the set \( P \) is finite (since the set \( \{1, 2, \ldots, n\} \) is finite).

\[^{70}\text{Proof.} \text{ Let } i \text{ and } j \text{ be two distinct elements of } P. \text{ All elements of } P \text{ are primes (by the definition of } P); \text{ thus, } i \text{ and } j \text{ are primes (since } i \text{ and } j \text{ are elements of } P). \text{ Also, } i \neq j \text{ (since } i \text{ and } j \text{ are distinct). Hence, Exercise 2.13.1 (applied to } p = i \text{ and } q = j \text{) yields } i \perp j. \text{ But } v_i(n) \in \mathbb{N} \text{ (since } n \text{ is nonzero) and } v_j(n) \in \mathbb{N} \text{ (for the same reason). Hence, Exercise 2.10.4 (applied to } i, j, v_i(n) \text{ and } v_j(n) \text{ instead of } a, b, n \text{ and } m \text{) yields } i^{v_i(n)} \perp j^{v_j(n)}. \text{ Qed.} \]
the equality (79) rewrites as

\[
\phi \left( \prod_{\substack{p \text{ prime}; \\ p|n}} p^{v_p(n)} \right) = \prod_{\substack{p \text{ prime}; \\ p|n}} \phi \left( p^{v_p(n)} \right). \tag{80}
\]

Now, applying the map \( \phi \) to both sides of the equality (78), we find

\[
\phi(n) = \phi \left( \prod_{\substack{p \text{ prime}; \\ p|n}} p^{v_p(n)} \right) = \prod_{\substack{p \text{ prime}; \\ p|n}} \phi \left( p^{v_p(n)} \right) = \prod_{\substack{p \text{ prime}; \\ p|n}} \left( p^{v_p(n)} \left( 1 - \frac{1}{p} \right) \right) \quad \text{(by (80))}
\]

\[
= \left( \prod_{\substack{p \text{ prime}; \\ p|n}} p^{v_p(n)} \right) \cdot \prod_{\substack{p \text{ prime}; \\ p|n}} \left( 1 - \frac{1}{p} \right) = n \cdot \prod_{\substack{p \text{ prime}; \\ p|n}} \left( 1 - \frac{1}{p} \right). \quad \text{(by (78))}
\]

This proves Theorem 2.14.5.

Theorem 2.15.3 generalizes Theorem 2.15.1 (a). Likewise, the following exercise generalizes Theorem 2.15.1 (b):

**Exercise 2.16.3.** Let \( a \) be an integer, and let \( n \) be a positive integer. Prove that

\[
a^n \equiv a^{n - \phi(n)} \mod n.
\]

[Hint: Use Exercises 2.13.9 and 2.14.2 and Theorems 2.15.3 and 2.14.4]

2019-02-18 lecture

2.17. Binomial coefficients

2.17.1. Definitions and basics

Next, we shall introduce and briefly study binomial coefficients. While binomial coefficients belong more to (enumerative) combinatorics than to algebra, they are used significantly in algebra, so we have to derive some of their properties.
Here is the definition of binomial coefficients (at least the one I am going to follow in these notes):

**Definition 2.17.1.** Let \( n \in \mathbb{Q} \) and \( k \in \mathbb{N} \). Then, we define the *binomial coefficient* \( \binom{n}{k} \) as follows:

(a) If \( k \in \mathbb{N} \), then we set
\[
\binom{n}{k} = \frac{n(n-1)(n-2) \cdots (n-k+1)}{k!} = \frac{\prod_{i=0}^{k-1} (n-i)}{k!}.
\]

(b) If \( k \notin \mathbb{N} \), then we set \( \binom{n}{k} = 0 \).

This definition is exactly the definition of \( \binom{n}{k} \) that we used in homework set #0. It is also almost exactly the definition given in [GrKnPa94, (5.1)] (except that we are allowing \( k \) to be non-integer, while the authors of [GrKnPa94] do not). Definition 2.17.1(a) is also identical with the definition of binomial coefficients in [Grinbe15]. Our choice to require \( n \in \mathbb{Q} \) is more or less arbitrary – we could have as well made the same definition for \( n \in \mathbb{R} \) or \( n \in \mathbb{C} \) (but I am not aware of this generality being of much use).

Generally, when you read literature on binomial coefficients, be aware that some authors use somewhat different definitions of \( \binom{n}{k} \). All known definitions give the same results when \( n \) and \( k \) are nonnegative integers, but in the other cases there may be discrepancies.

Here are some examples of binomial coefficients:

**Example 2.17.2.** (a) Definition 2.17.1(a) yields
\[
\binom{n}{2} = \frac{n(n-1)}{2!} = \frac{n(n-1)}{2}
\]
for all \( n \in \mathbb{Q} \). Thus, for example,
\[
\binom{5}{2} = \frac{5 \cdot 4}{2} = 10.
\]

(b) Definition 2.17.1(a) yields
\[
\binom{n}{3} = \frac{n(n-1)(n-2)}{3!} = \frac{n(n-1)(n-2)}{6}
\]
for
all \( n \in \mathbb{Q} \). Thus, for example,
\[
\binom{5}{3} = \frac{5 \cdot 4 \cdot 3}{6} = \frac{60}{6} = 10;
\]
\[
\binom{1}{3} = \frac{1 \cdot 0 \cdot (-1)}{6} = \frac{0}{6} = 0;
\]
\[
\binom{-2}{3} = \frac{(-2) \cdot (-3) \cdot (-4)}{6} = \frac{-24}{6} = -4;
\]
\[
\binom{1/2}{3} = \frac{(1/2) \cdot (-1/2) \cdot (-3/2)}{6} = \frac{3/8}{6} = \frac{1}{16}.
\]

(c) Definition 2.17.1 (a) yields \( \binom{n}{1} = \frac{n}{1!} = \frac{n}{1} = n \) for all \( n \in \mathbb{Q} \).

(d) Definition 2.17.1 (b) yields \( \binom{4}{1/2} = 0 \) (since \( 1/2 \notin \mathbb{N} \)).

The binomial coefficients \( \binom{n}{k} \) for \( n \in \mathbb{N} \) and \( k \in \{0, 1, \ldots, n\} \) are particularly important. They are usually tabulated in a triangle-shaped table known as Pascal’s triangle, which starts as follows:

\[
\begin{array}{cccccccc}
1 \\
1 & 1 \\
1 & 2 & 1 \\
1 & 3 & 3 & 1 \\
1 & 4 & 6 & 4 & 1 \\
1 & 5 & 10 & 10 & 5 & 1 \\
1 & 6 & 15 & 20 & 15 & 6 & 1
\end{array}
\]

In this table, the binomial coefficient \( \binom{n}{k} \) appears as the \( k \)-th entry (from the left) of the \( n \)-th row (but we count the rows from 0; that is, the topmost row, consisting just of a single “1”, is actually the 0-th row). We advise the reader to peruse the Wikipedia article for the history and the multiple illustrious properties of Pascal’s triangle.

The expression \( \binom{n}{k} \) is pronounced as “\( n \) choose \( k \)”. The reason for the word “choose” will become clearer once we have seen Theorem 2.17.10 further below.

Some of these properties are so fundamental that we are going to list them right now:

**Theorem 2.17.3.** Let \( n \in \mathbb{N} \) and \( k \in \mathbb{N} \) be such that \( n \geq k \). Then,
\[
\binom{n}{k} = \frac{n!}{k! \,(n-k)!}.
\]
Proof of Theorem 2.17.3: This was Exercise 3 (a) on homework set #0.

Several authors use the formula \( \binom{n}{k} = \frac{n!}{k!(n-k)!} \) as a definition of the binomial coefficients. However, this definition has the massive disadvantage of being less general than Definition 2.17.1 (since it only covers the case when \( n, k \in \mathbb{N} \) and \( n \geq k \)). To us, this formula is not a definition, but a result that can be proven.

**Theorem 2.17.4.** Let \( n \in \mathbb{N} \) and \( k \in \mathbb{Q} \) be such that \( k > n \). Then,

\[
\binom{n}{k} = 0.
\]

Proof of Theorem 2.17.4: This was Exercise 3 (b) on homework set #0.

**Theorem 2.17.5.** Let \( n \in \mathbb{Q} \). Then,

\[
\binom{n}{0} = 1.
\]

Proof of Theorem 2.17.5: Definition 2.17.1 (applied to \( k = 0 \)) yields

\[
\binom{n}{0} = \frac{\prod_{i=0}^{0-1} (n-i)}{0!} = \frac{1}{1}
\]

(since \( \prod_{i=0}^{0-1} (n-i) = (\text{empty product}) = 1 \) and \( 0! = 1 \)). Thus, \( \binom{n}{0} = \frac{1}{1} = 1 \). This proves Theorem 2.17.5.

**Theorem 2.17.6.** Let \( n \in \mathbb{N} \) and \( k \in \mathbb{Q} \). Then,

\[
\binom{n}{k} = \binom{n}{n-k}.
\]

Theorem 2.17.6 is known as the *symmetry of binomial coefficients*. Note that it fails if \( n \notin \mathbb{N} \); thus, be careful when applying it!

Proof of Theorem 2.17.6: This was Exercise 3 (c) on homework set #0.

**Theorem 2.17.7.** Let \( n \in \mathbb{Q} \) and \( k \in \mathbb{Q} \). Then,

\[
\binom{-n}{k} = (-1)^k \binom{k+n-1}{k}.
\]
Theorem 2.17.7 is one of the versions of the upper negation formula.

Proof of Theorem 2.17.7 This was Exercise 3 (d) on homework set #0.

Theorem 2.17.8. Any \( n \in \mathbb{Q} \) and \( k \in \mathbb{Q} \) satisfy

\[
\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.
\]

Theorem 2.17.8 is known as the recurrence of the binomial coefficients, and is the reason why each entry of Pascal’s triangle is the sum of the two entries above it.

Proof of Theorem 2.17.8 This was Exercise 3 (e) on homework set #0.

Theorem 2.17.9. Any \( n \in \mathbb{Q} \) and \( k \in \mathbb{Q} \) satisfy

\[
k \binom{n}{k} = n \binom{n-1}{k-1}.
\]

Proof of Theorem 2.17.9 This was Exercise 3 (f) on homework set #0.

2.17.2. Combinatorial interpretation

The next property of binomial coefficients is one of the major motivations for defining them:

Theorem 2.17.10. Let \( n \in \mathbb{N} \) and \( k \in \mathbb{Q} \). Let \( N \) be an \( n \)-element set. Then, \( \binom{n}{k} \) is the number of \( k \)-element subsets of \( N \).

We shall refer to Theorem 2.17.10 as the Combinatorial interpretation of binomial coefficients. Theorem 2.17.10 can be restated as “\( \binom{n}{k} \) is the number of ways to choose \( k \) elements (with no repetitions and with no regard for the order) from a given \( n \)-element set (when \( n \in \mathbb{N} \))”. This is the reason why \( \binom{n}{k} \) is called “\( n \) choose \( k \)”. Note, however, that Theorem 2.17.10 does not directly help us compute \( \binom{n}{k} \) when \( n \notin \mathbb{N} \).

\[71\] Of course, this does not apply to the “1” at the apex of Pascal’s triangle (unless we extend the triangle further to the top by a \((-1)-st\) row).
Proof of Theorem 2.17.10. What follows is an outline of the proof. For a detailed proof, see [Grinbe15, Exercise 3.4], where I thoroughly prove Theorem 2.17.10 in the case \( k \in \mathbb{N} \). (The remaining case \( k \notin \mathbb{N} \) is obvious, because in that case the theorem simply says \( 0 = 0 \).)

We proceed by induction on \( n \):

**Induction base:** Let \( n, k \) and \( N \) be as in Theorem 2.17.10 and let us assume that \( n = 0 \). From \( n = 0 \), we obtain \( \binom{n}{k} = \binom{0}{k} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0 \end{cases} \) (this is easy to derive from Definition 2.17.1\(^{72}\)). On the other hand, the set \( N \) is empty (since \( |N| = n = 0 \)). Thus, its only subset is \( \emptyset \), which is a 0-element subset. Hence, \( N \) has exactly one 0-element subset, and no subsets of any other size. Hence, the number of \( k \)-element subsets of \( N \) is \( \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0 \end{cases} \). Comparing this with \( \binom{n}{k} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0 \end{cases} \), we conclude that \( \binom{n}{k} \) is the number of \( k \)-element subsets of \( N \). Thus, we have proven Theorem 2.17.10 under the assumption that \( n = 0 \). This completes the induction base.

**Induction step:** Let \( m \) be a positive integer. Assume (as the induction hypothesis) that Theorem 2.17.10 holds for \( n = m - 1 \). We must now prove that Theorem 2.17.10 holds for \( n = m \).

Let \( k \in \mathbb{Q} \). Let \( N \) be an \( m \)-element set. Thus, \( |N| = m > 0 \). Hence, the set \( N \) is nonempty; in other words, there exists some \( a \in N \). Pick such an \( a \). (It does not matter which one we choose, but we need to leave it fixed from now on.) Clearly, \( |N \setminus \{a\}| = m - 1 \) (since \( |N| = m \)). In other words, \( N \setminus \{a\} \) is an \((m - 1)\)-element set.

Now, the \( k \)-element subsets of \( N \) can be classified into two types:

- We say that a \( k \)-element subset is **type-1** if it doesn’t contain \( a \).
- We say that a \( k \)-element subset is **type-2** if it does contain \( a \).

(We shall use the adjectives “type-1” and “type-2” for \( k \)-element subsets of \( N \) only. Thus, whenever we say “type-1 subset” in the following, we will always mean “type-1 \( k \)-element subset of \( \mathbb{N} \)”, and similarly for “type-2 subset”.)

\(^{72}\)To wit:

- If \( k = 0 \), then \( \binom{0}{k} = \binom{0}{0} = 1 \) (by Theorem 2.17.5).
- If \( k > 0 \), then Theorem 2.17.4 (applied to 0 instead of \( n \)) yields \( \binom{0}{k} = 0 \).
- If \( k < 0 \), then \( k \notin \mathbb{N} \) and thus \( \binom{0}{k} = 0 \) (by Definition 2.17.1 (b)).

Thus, in all three cases \( k = 0, k > 0 \) and \( k < 0 \), we conclude that \( \binom{0}{k} = \begin{cases} 1, & \text{if } k = 0; \\ 0, & \text{if } k \neq 0 \end{cases} \).
Clearly, any $k$-element subset of $N$ is either type-1 or type-2 (but never both at the same time).

The type-1 subsets are precisely the $k$-element subsets of the $(m - 1)$-element set $N \setminus \{a\}$. By our induction hypothesis, we know that Theorem 2.17.10 holds for $n = m - 1$. Hence, we can apply Theorem 2.17.10 to $m - 1$ and $N \setminus \{a\}$ instead of $n$ and $N$. We thus conclude that \( \binom{m - 1}{k} \) is the number of $k$-element subsets of $N \setminus \{a\}$. In other words, \( \binom{m - 1}{k} \) is the number of type-1 subsets (since the type-1 subsets are precisely the $k$-element subsets of $N \setminus \{a\}$). In other words, \( \binom{m - 1}{k} = \text{(the number of type-1 subsets)} \). (81)

Now, let us count the type-2 subsets\(^\text{73}\) This is a bit harder, since they are not subsets of $N \setminus \{a\}$ anymore. However, they are in 1-to-1 correspondence (aka bijection) with some such subsets. Namely, there is a bijection

\[
\{(k - 1)\text{-element subsets of } N \setminus \{a\}\} \to \{\text{type-2 subsets}\},
\]

\[
S \mapsto S \cup \{a\}.
\]

(The inverse of this bijection sends each type-2 subset $T$ to $T \setminus \{a\}$. You can easily show that these two maps are actually well-defined and mutually inverse, so that they really are bijections.) This bijection shows that

\[|\{\text{type-2 subsets}\}| = |\{(k - 1)\text{-element subsets of } N \setminus \{a\}\}|. \tag{82}\]

But recall that Theorem 2.17.10 holds for $n = m - 1$. Hence, we can apply Theorem 2.17.10 to $m - 1$, $k - 1$ and $N \setminus \{a\}$ instead of $n$, $k$ and $N$. We thus conclude that

\[
\binom{m - 1}{k - 1}
\]

is the number of $(k - 1)$-element subsets of $N \setminus \{a\}$. In other words, \( \binom{m - 1}{k - 1} = |\{(k - 1)\text{-element subsets of } N \setminus \{a\}\}|. \)

Comparing this equality with (82), we obtain

\[
\binom{m - 1}{k - 1} = |\{\text{type-2 subsets}\}|
\]

\[= \text{(the number of type-2 subsets)} \]. \tag{83}\]

Now, recall that any $k$-element subset of $N$ is either type-1 or type-2 (but never both at the same time). Hence, we can count all $k$-element subsets of $N$ by first

\(^{73}\)Keep in mind that “type-2 subset” means “type-2 $k$-element subset of $N$.”
counting the type-1 subsets, then counting the type-2 subsets, and then adding these two results. We thus find

$$74$$

(by Theorem 2.17.8 (applied to $n = m$) yields $\binom{m}{k} = \binom{m-1}{k} + \binom{m-1}{k-1}$). In other words, $\binom{m}{k}$ is the number of $k$-element subsets of $N$.

Now, forget that we fixed $N$ and $k$. We thus have shown that if $k \in \mathbb{Q}$ and if $N$ is an $m$-element set, then $\binom{m}{k}$ is the number of $k$-element subsets of $N$. In other words, Theorem 2.17.10 holds for $n = m$. This completes the induction step. Hence, Theorem 2.17.10 is proven.

Corollary 2.17.11. Let $n \in \mathbb{N}$ and $k \in \mathbb{Q}$. Then, $\binom{n}{k}$ is a nonnegative integer.

Proof of Corollary 2.17.11. Let $N = \{1, 2, \ldots, n\}$; thus, $N$ is an $n$-element set. Hence, Theorem 2.17.10 shows that $\binom{n}{k}$ is the number of $k$-element subsets of $N$. But the latter number is clearly a nonnegative integer (since it counts something). Thus, $\binom{n}{k}$ is a nonnegative integer. This proves Corollary 2.17.11.

Proposition 2.17.12. Let $n \in \mathbb{Z}$ and $k \in \mathbb{Q}$. Then, $\binom{n}{k}$ is an integer.

Proof of Proposition 2.17.12. If $n \geq 0$, then this follows from Corollary 2.17.11 (because $n \geq 0$ implies $n \in \mathbb{N}$, and thus we can apply Corollary 2.17.11). Thus, for the rest of this proof, we WLOG assume that $n < 0$. Hence, $n \leq -1$ (since $n$ is an integer), so that $n + 1 \leq 0$ and thus $-1 - (n + 1) \geq 0$. Therefore, $-1 - (n + 1) \in \mathbb{N}$ (since $-1 - (n + 1)$ is an integer).

If $k \notin \mathbb{N}$, then $\binom{n}{k}$ is a integer (since Definition 2.17.1 (b) yields $\binom{n}{k} = 0$ in this case). Thus, for the rest of this proof, we WLOG assume that $k \in \mathbb{N}$.

The combinatorial principle we are using in the following computation is the so-called sum rule in its simplest form (see, e.g., [Loehr11, 1.1] or [LeLeMe18, §15.2.3]).
Thus, \( k + (-n) - 1 = \binom{k}{\in \mathbb{N}} + (-n+1) \in \mathbb{N} \). Hence, Corollary 2.17.11 (applied to \( k + (-n) - 1 \) instead of \( n \)) yields that \( \binom{k + (-n) - 1}{k} \) is a nonnegative integer. Thus, \( \binom{n}{k} \in \mathbb{Z} \). In other words, \( \binom{n}{k} \) is an integer. Thus, Proposition 2.17.12 is proven. 

Exercise 2.17.1. Let \( k \in \mathbb{N} \). Prove that the product of any \( k \) consecutive integers is divisible by \( k! \).

Exercise 2.17.2. In this exercise, we shall use the Iverson bracket notation: If \( A \) is any statement, then \( [A] \) stands for the integer \( 1 \), if \( A \) is true; \( 0 \), if \( A \) is false (which is also known as the truth value of \( A \)). For instance, \( [1 + 1 = 2] = 1 \) and \( [1 + 1 = 1] = 0 \).

(a) Prove that \( n/!k = \sum_{i=1}^{n} [k \mid i] \) for any \( n \in \mathbb{N} \) and any positive integer \( k \).

(b) Prove that \( v_p(n) = \sum_{i \geq 1} [p^i \mid n] \) for any prime \( p \) and any nonzero integer \( n \).

There, the sum \( \sum_{i \geq 1} [p^i \mid n] \) is a sum over all positive integers; but it is well-defined, since it has only finitely many nonzero addends.

(c) Prove that \( v_p(n!) = \sum_{i \geq 1} n/!p^i \) for any prime \( p \) and any \( n \in \mathbb{N} \). (Here, the expression “\( \sum_{i \geq 1} n/!p^i \)” should be understood as \( \sum_{i \geq 1} (n/!p^i) \). Again, this sum \( \sum_{i \geq 1} (n/!p^i) \) is well-defined, since it has only finitely many nonzero addends.)

(d) Use part (c) to prove Corollary 2.17.11 again.

The claim of Exercise 2.17.2 (c) is usually rewritten in the form \( v_p(n!) = \sum_{i \geq 1} \left\lfloor \frac{n}{p^i} \right\rfloor \) (which is equivalent, because of Proposition 2.8.3; in this form, it is known as Legendre’s formula or as de Polignac’s formula (see, e.g., [Grinbe16, Theorem 1.3.3]). It is often a helpful tool in proving divisibility properties of factorials and binomial coefficients. One application, for example, is to quickly compute how many zeroes the decimal expansion of \( n! \) ends with. (Note that Exercise 2.17.2 (b) can be rewritten as \( v_p(n) = \sum_{i \geq 1 \mid p^i \mid n \}} 1 \); in this form it appears in [Grinbe16, Lemma 1.3.4].)
2.17.3. Binomial formula and Vandermonde convolution

One of the staples of enumerative combinatorics are identities that involve binomial coefficients. Hundreds of such identities have been found (see, e.g., Henry W. Gould’s website for a list of some of them; see also [GrKnPa94, Chapter 5] and [Grinbe15, Chapter 3] for introductions). At this point, let us only show two of the most important ones (not counting the ones we have already shown above). Probably the most famous one is the binomial formula:

**Theorem 2.17.13.** Let \( x, y \) be any numbers (e.g., rational or real or complex numbers). Let \( n \in \mathbb{N} \). Then,

\[
(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.
\]

Theorem 2.17.13 is known as the binomial formula or the binomial theorem. It generalizes the well-known and beloved identities

\[
(x + y)^2 = x^2 + 2xy + y^2;
\]
\[
(x + y)^3 = x^3 + 3x^2y + 3xy^2 + y^3;
\]
\[
(x + y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4
\]
(as well as \((x + y)^1 = x^1 + y^1\) and \((x + y)^0 = 1\), of course).

**Proof of Theorem 2.17.13 (sketched).** This can be proven by a straightforward induction on \( n \) (using Theorem 2.17.8 in the induction step). See [Grinbe15, Exercise 3.6] for details of this proof. Alternatively, see [Galvin17, Identity 11.4] for combinatorial proofs (which rely on Theorem 2.17.10).

The next identity we want to show is the Vandermonde convolution identity:

**Theorem 2.17.14.** Let \( x, y \in \mathbb{Q} \) and \( n \in \mathbb{N} \). Then,

\[
\binom{x+y}{n} = \sum_{k=0}^{n} \binom{x}{k} \binom{y}{n-k}.
\]

For example, for \( n = 2 \), Theorem 2.17.14 says that

\[
\binom{x+y}{2} = \binom{x}{0} \binom{y}{2} + \binom{x}{1} \binom{y}{1} + \binom{x}{2} \binom{y}{0} = \binom{y}{2} + xy + \binom{x}{2}.
\]

The proof of Theorem 2.17.14 that we are soon going to sketch is similar to the one given in [Grinbe15, §3.3.3] (but, unlike the latter proof, we will use polynomials...
in 1 variable only). It will not be a complete proof, since it will rely on some properties of polynomials, and not only have we not proven these properties – we have actually not rigorously defined polynomials yet! (We will do so later, after we have introduced rings.) See [Grinbe15, §3.3.2] for another (more boring and tedious, but conceptually simpler) proof of Theorem \ref{2.17.14}.

Our proof of Theorem \ref{2.17.14} proceeds via several intermediate steps. The first one is to prove Theorem \ref{2.17.14} in the particular case when $x, y \in \mathbb{N}$:

**Lemma 2.17.15.** Let $a, b \in \mathbb{N}$ and $n \in \mathbb{N}$. Then,

\[
\binom{a + b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k}.
\]

(We have renamed the variables $x$ and $y$ from Theorem \ref{2.17.14} as $a$ and $b$ here, since we will soon use the letter “$x$” for something completely different.)

*Proof of Lemma 2.17.15 (sketched).* Let

\[ C = \{1, 2, \ldots, a\} \cup \{-1, -2, \ldots, -b\}. \]

Thus, $C$ is an $(a + b)$-element set, containing only positive and negative integers. How many $n$-element subsets does $C$ have?

- On the one hand: The set $C$ is an $(a + b)$-element set. Hence, Theorem \ref{2.17.10} (applied to $a + b$, $n$ and $C$ instead of $n$, $k$ and $N$) shows that the number of $n$-element subsets of $C$ is $\binom{a + b}{n}$.

- On the other hand: Let us classify the $n$-element subsets of $C$ according to how many positive elements they have. We claim the following:

  **Claim 1:** For each $k \in \{0, 1, \ldots, n\}$, the number of $n$-element subsets of $C$ having *exactly* $k$ positive elements is $\binom{a}{k} \binom{b}{n-k}$.

  *Proof of Claim 1:* Let $k \in \{0, 1, \ldots, n\}$. In order to choose an $n$-element subset of $C$ having exactly $k$ positive elements, we need to choose

  - its $k$ positive elements from the set of all positive elements of $C$ (that is, from the set $\{1, 2, \ldots, a\}$), and
  - its remaining $n - k$ (negative) elements from the set of all negative elements of $C$ (that is, from the set $\{-1, -2, \ldots, -b\}$).

  In other words, we need to choose

  - a $k$-element subset of the set $\{1, 2, \ldots, a\}$, and
  - an $(n - k)$-element subset of the set $\{-1, -2, \ldots, -b\}$. 


Theorem 2.17.10 (applied to $a, k$ and $\{1, 2, \ldots, a\}$ instead of $n, k$ and $N$) shows that the number of $k$-element subsets of the set $\{1, 2, \ldots, a\}$ is $\binom{a}{k}$ (since $\{1, 2, \ldots, a\}$ is an $a$-element set). Similarly, the number of $(n - k)$-element subsets of the set $\{-1, -2, \ldots, -b\}$ is $\binom{b}{n - k}$. Since we need to choose one of the former subsets and one of the latter subsets (and our choices are independent – i.e., any of the former subsets can be combined with any of the latter), we thus conclude that the total number of options we have is $\binom{a}{k} \binom{b}{n - k}$ \footnote{In other words, the number of $n$-element subsets of $C$ having exactly $k$ positive elements is $\binom{a}{k} \binom{b}{n - k}$. This proves Claim 1.] Now, the total number of $n$-element subsets of $C$ is\footnote{The combinatorial principle we are using here is the so-called \textit{product rule} (see, e.g., [Loehr11, 1.8] or [LeLeMe18, §15.2.1]).} 

\[
\sum_{k=0}^{n} \left( \text{the number of } n \text{-element subsets of } C \text{ having exactly } k \text{ positive elements} \right) = \binom{a}{k} \binom{b}{n - k} 
\]

(by Claim 1)

\[
= \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n - k}. 
\]

Now, we have computed the number of $n$-element subsets of $C$ in two ways. The first way yielded the result $\binom{a + b}{n}$, while the second way yielded $\sum_{k=0}^{n} \binom{a}{k} \binom{b}{n - k}$. But these two results clearly have to be equal. In other words, we have

\[
\binom{a + b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n - k}. 
\]

Thus, Lemma 2.17.15 holds.

(This was an example of a proof by \textit{double counting}, also known as a \textit{combinatorial proof}. See [LeLeMe18, §15.10] for some more examples of such proofs, and see most textbooks on combinatorics for more.) \qed
This shows that Theorem 2.17.14 holds for all \( x \in \mathbb{N} \) and \( y \in \mathbb{N} \). In order to extend its reach to arbitrary rational \( a \) and \( b \), we shall use the “polynomial identity trick”. First, let us briefly explain what polynomials are, without giving a formal definition.

Informally, a polynomial (in 1 variable \( x \), with rational coefficients) is an “expression” of the form \( a_k x^k + a_{k-1} x^{k-1} + \cdots + a_0 \), where \( a_k, a_{k-1}, \ldots, a_0 \) are (fixed) rational numbers and where \( x \) is a (so far meaningless) symbol (called indeterminate or variable). For example, \( 4x^3 + 2x^2 - \frac{1}{3}x + \frac{2}{7} \) is a polynomial, and so is \( 0x^3 + x^2 - 0x + \frac{1}{3} \). We can omit terms of the form “0\( x^i \)” when writing down a polynomial and treat the result as being the same polynomial; thus, \( 0x^3 + x^2 - 0x + \frac{1}{3} \) can also be written as \( x^2 - 0x + \frac{1}{3} \) and as \( x^2 + \frac{1}{3} \). Likewise, we can treat the “+” signs as signifying addition and behaving like it, so, e.g., commutativity holds:

\[
2x^3 + 5x = 5x + 2x^3
\]

are the same polynomial (but \( 2x + 5x^3 \) is different). We also pretend that distributivity holds, so “like terms” can be combined: e.g., we have

\[
4x^3 + 9x^3 = (4 + 9) x^3 = 13x^3 \quad \text{or} \quad 4x^3 - 12x^3 = (4 - 12) x^3 = -8x^3.
\]

Thus, we can add two polynomials: e.g.,

\[
\left(3x^2 - 1x + \frac{1}{2}\right) + (6x - 7) = 3x^2 + \left(-1 + 6\right)x + \left(\frac{1}{2} - 7\right) = 3x^2 + 5x + \left(-\frac{13}{2}\right).
\]

By pretending that the \( x^i \) (with \( i \in \mathbb{N} \)) are actual powers of the symbol \( x \), and that multiplication obeys the associativity law (so that \((\lambda x^i)x^j = \lambda(x^i x^j) = \lambda x^{i+j} \) for rational \( \lambda \) and \( i, j \in \mathbb{N} \)), we can multiply polynomials as well (first use distributivity to expand the product):

\[
(3x - 5) \left(x^2 + 3x + 2\right) = 3x \left(x^2 + 3x + 2\right) - 5 \left(x^2 + 3x + 2\right)
\]

\[
= 3x^3 + 9x^2 + 6x - 5x^2 - 15x - 10
\]

\[
= 3x^3 + 4x^2 - 9x - 10.
\]

Most importantly, it is possible to substitute a number into a polynomial: If \( u \in \mathbb{Q} \) and if \( P = a_k x^k + a_{k-1} x^{k-1} + \cdots + a_0 \) is a polynomial, then we define \( P(u) \) (called the evaluation of \( P \) at \( u \), or the result of substituting \( u \) for \( x \) in \( P \)) to be the number \( a_k u^k + a_{k-1} u^{k-1} + \cdots + a_0 \). More generally, if the polynomial \( P \) is given in any of its forms (e.g., as a product of other polynomials), then we can compute \( P(u) \) by replacing each \( x \) appearing in this form by an \( u \). For example, if \( P = (2x + 1) (3x + 1) - (4x + 1) (5x + 1) \), then \( P(u) = (2u + 1) (3u + 1) - (4u + 1) (5u + 1) \); thus, we do not need to expand \( P \) before substituting \( u \) into it.

Even more generally, \( u \) does not have to be a rational number in order to be substituted in a polynomial \( P \) – it can be (roughly speaking!) anything that can
be taken to the $i$-th power for $i \in \mathbb{N}$ and that can be added and multiplied by a rational number. For example, $u$ can be a real number or a square matrix or another polynomial. (We will later learn the precise meaning of “anything” here.)

We have been vague in our definition of polynomials, since making it rigorous would take us a fair way afield. But we will eventually (in April?) define polynomials rigorously and prove that all of the above claims (e.g., about associativity and distributivity) actually hold. For now, we need a basic property of polynomials:

**Proposition 2.17.16.** Let $P$ and $Q$ be two polynomials in 1 variable $x$ with rational coefficients. Assume that infinitely many $u \in \mathbb{Q}$ satisfy $P(u) = Q(u)$. Then, $P = Q$ (as polynomials).

We will prove Proposition 2.17.16 later.

Note that polynomials are not functions – despite the fact that we can substitute numbers into them and obtain other numbers. However, in many regards, they behave like functions. For what we are going to do in this section, the difference does not matter; we can treat polynomials as functions here.

With Lemma 2.17.15, we have proven Theorem 2.17.14 in the case when $x$ and $y$ belong to $\mathbb{N}$. Our goal, however, is to prove it for arbitrary $x, y \in \mathbb{Q}$. Let us first go to the intermediate level of generality – allowing $x$ to be arbitrary, but still requiring $y \in \mathbb{N}$. Thus, we want to prove the following lemma:

**Lemma 2.17.17.** Let $a \in \mathbb{Q}$, $b \in \mathbb{N}$ and $n \in \mathbb{N}$. Then,

\[
\binom{x + b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k}.
\]

**Proof of Lemma 2.17.15 (sketched).** Let us define a polynomial $P$ in 1 variable $x$ with rational coefficients as follows:

\[
P = \binom{x + b}{n}.
\] (84)

The “binomial coefficient” $\binom{x + b}{n}$ here is to be understood by extending Definition 2.17.1 (a) in the obvious fashion to the case when $n$ is a polynomial (in our case, $x + b$) rather than a rational number. Thus,

\[
\binom{x + b}{n} = \frac{(x + b)(x + b - 1)(x + b - 2) \cdots (x + b - n + 1)}{n!}.
\]

Let us also define a polynomial $Q$ in 1 variable $x$ with rational coefficients as follows:

\[
Q = \sum_{k=0}^{n} \binom{x}{k} \binom{b}{n-k}.
\] (85)
(Again, the “binomial coefficients” \( \binom{x}{k} \) are defined via our extension of Definition 2.17.1(a), and can be explicitly written as \( \binom{x}{k} = \frac{x(x-1)(x-2)\cdots(x-k+1)}{k!} \).

Meanwhile, the \( \binom{b}{n-k} \) are just constant integers.)

Now, for each \( u \in \mathbb{N} \), we have

\[
P(u) = \binom{u+b}{n} = \sum_{k=0}^{n} \binom{u}{k} \binom{b}{n-k} \quad \text{(by substituting \( u \) for \( x \) in the equality (84))}
\]

(by Lemma 2.17.15 applied to \( u \) instead of \( a \)) and

\[
Q(u) = \sum_{k=0}^{n} \binom{u}{k} \binom{b}{n-k} \quad \text{(by substituting \( u \) for \( x \) in the equality (85)).}
\]

Comparing these two equalities, we obtain \( P(u) = Q(u) \) for all \( u \in \mathbb{N} \). Hence, infinitely many \( u \in \mathbb{Q} \) satisfy \( P(u) = Q(u) \) (since infinitely many \( u \in \mathbb{Q} \) satisfy \( u \in \mathbb{N} \)). Thus, Proposition 2.17.16 yields \( P = Q \). In view of (84) and (85), this rewrites as

\[
\binom{x+b}{n} = \sum_{k=0}^{n} \binom{x}{k} \binom{b}{n-k}. \quad (86)
\]

Now, substituting \( a \) for \( x \) in this equality of polynomials, we obtain

\[
\binom{a+b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k}. \]

This proves Lemma 2.17.17. \( \square \)

Let us summarize the main idea of this proof: We replaced the rational number \( a \) by the indeterminate \( x \), thus transforming the identity we were proving into an equality between two polynomials (namely, \( P = Q \)). But in order to prove an equality between polynomials, it suffices to prove that it holds at infinitely many numbers (by Proposition 2.17.16); thus, in particular, it suffices to check it at all non-negative integers. But this is precisely what we did in Lemma 2.17.15 above. This kind of argument (with its use of Proposition 2.17.16) is known as the “polynomial identity trick”.

Now, let us extend the reach of Lemma 2.17.17 further, allowing both \( a \) and \( b \) to be arbitrary (and thus obtaining the whole Theorem 2.17.14):

\[\textbf{Lemma 2.17.18.} \text{ Let } a, b \in \mathbb{Q} \text{ and } n \in \mathbb{N}. \text{ Then,} \]

\[
\binom{a+b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k}. \]
Proof of Lemma 2.17.18 (sketched). Deriving Lemma 2.17.18 from Lemma 2.17.17 is very similar to deriving Lemma 2.17.17 from Lemma 2.17.15. The main difference is that we replace $b$ (rather than $a$) by the indeterminate $x$ now.

Here are the details: Let us define a polynomial $P$ in 1 variable $x$ with rational coefficients as follows:

$$P = \binom{a + x}{n}.$$  \hspace{1cm} (87)

Let us also define a polynomial $Q$ in 1 variable $x$ with rational coefficients as follows:

$$Q = \sum_{k=0}^{n} \binom{a}{k} \binom{x}{n-k}.$$  \hspace{1cm} (88)

Now, for each $u \in \mathbb{N}$, we have

$$P(u) = \binom{a + u}{n} \quad \text{(by substituting $u$ for $x$ in the equality (87))}$$

$$= \sum_{k=0}^{n} \binom{a}{k} \binom{u}{n-k}$$  \hspace{1cm} (by Lemma 2.17.17 applied to $u$ instead of $b$)

(by substituting $u$ for $x$ in the equality (88)). Comparing these two equalities, we obtain $P(u) = Q(u)$ for all $u \in \mathbb{N}$. Hence, infinitely many $u \in \mathbb{Q}$ satisfy $P(u) = Q(u)$ (since infinitely many $u \in \mathbb{Q}$ satisfy $u \in \mathbb{N}$). Thus, Proposition 2.17.16 yields $P = Q$. In view of (87) and (88), this rewrites as

$$\binom{a + x}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{x}{n-k}.$$  \hspace{1cm} \bigstar

Now, substituting $b$ for $x$ in this equality of polynomials, we obtain

$$\binom{a + b}{n} = \sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k}.$$  \hspace{1cm} \bigstar

This proves Lemma 2.17.18.

Proof of Theorem 2.17.14 (sketched). Theorem 2.17.14 is just Lemma 2.17.18 with $a$ and $b$ renamed as $x$ and $y$.

2.17.4. Some divisibilities and congruences

So far we have been proving identities between binomial coefficients. Let us now step to divisibilities and congruences.

Proposition 2.17.12 shows that binomial coefficients $\binom{n}{k}$ are integers whenever $n$ is an integer. This allows us to study divisibilities and congruences between binomial coefficients (and you have seen a few of them on homework set #1). One of the most important such divisibilities is the following fact:
Theorem 2.17.19. Let $p$ be a prime. Let $k \in \{1, 2, \ldots, p-1\}$. Then, $p \mid \binom{p}{k}$.

First proof of Theorem 2.17.19\footnote{by an application of Proposition 2.17.12} Applying Theorem 2.17.9 to $n = p$, we obtain

$$\frac{p}{k} = \binom{p-1}{k-1}.$$ 

Thus, $p \mid \frac{p}{k}$ (since $\binom{p-1}{k-1}$ is an integer\footnote{by an application of Proposition 2.17.12}). But Proposition 2.13.4 (applied to $i = k$) yields that $k$ is coprime to $p$. In other words, $k \perp p$, and thus $p \perp k$. Hence, Theorem 2.10.6 (applied to $a = p$, $b = k$ and $c = \binom{p}{k}$) yields $p \mid \binom{p}{k}$ (since $p \mid \frac{p}{k}$). This proves Theorem 2.17.19. \qed

We shall see a second, combinatorial proof of Theorem 2.17.19 further below; it will rely on the concept of group actions.

Let us state two congruences for binomial coefficients, which we will show later using tools from abstract algebra:

Theorem 2.17.20 (Lucas’s congruence). Let $p$ be a prime. Let $a, b \in \mathbb{Z}$. Let $c, d \in \{0, 1, \ldots, p-1\}$. Then,

$$\binom{pa+c}{pb+d} \equiv \binom{a}{b} \binom{c}{d} \mod p.$$ 

Theorem 2.17.21 (Babbage’s congruence). Let $p$ be a prime. Let $a, b \in \mathbb{Z}$. Then,

$$\binom{pa}{pb} \equiv \binom{a}{b} \mod p^2.$$ 

For the impatient: Elementary proofs of Theorem 2.17.20 and Theorem 2.17.21 can be found in [Grinbe17].

Remark 2.17.22. Lucas’s congruence has the following consequence: Let $p$ be a prime. Let $a, b \in \mathbb{N}$. Write $a$ and $b$ in base $p$ as follows:

$$a = a_k p^k + a_{k-1} p^{k-1} + \cdots + a_0 p^0 \quad \text{and} \quad b = b_k p^k + b_{k-1} p^{k-1} + \cdots + b_0 p^0.$$
with $k \in \mathbb{N}$ and $a_k, a_{k-1}, \ldots, a_0, b_k, b_{k-1}, \ldots, b_0 \in \{0, 1, \ldots, p-1\}$. (Note that we allow “leading zeroes” – i.e., any of $a_k$ and $b_k$ can be 0.) Then,

$${a \choose b} \equiv (a_k \bmod p) (a_{k-1} \bmod p) \cdots (a_0 \bmod p) \mod p.$$  

(This can be easily proven by induction on $k$, using Theorem 2.17.20 in the induction step.) This allows for quick computation of remainders of $${a \choose b}$$ modulo prime numbers, and also explains (when applied to $p = 2$) why we can obtain (an approximation of) Sierpinski’s triangle from Pascal’s triangle by coloring all even numbers white and all odd numbers black.

See [Mestro14] and [Granvi05] for overviews of more complicated divisibilities and congruences for binomial coefficients.

2019-02-20 lecture

2.17.5. Integer-valued polynomials

Now that we have introduced polynomials (albeit informally and on somewhat shaky foundations) and binomial coefficients (albeit briefly), it would be a shame to leave unmentioned a subject that connects the two particularly closely: the integer-valued polynomials. We are going to state a few basic facts, but we will not prove them.

If $f = a_kx^k + a_{k-1}x^{k-1} + \cdots + a_0$ is a polynomial (in 1 variable $x$, with rational coefficients), then the rational numbers $a_k, a_{k-1}, \ldots, a_0$ are called the coefficients of $f$. The coefficients of a polynomial $f$ are uniquely determined by $f$ (except for the fact that we can always add terms of the form $0x^\ell$ and thus obtain extra coefficients that are equal to 0). (This fact is not obvious, given our “definition” of polynomials above. We will later define polynomials more formally as sequences of coefficients; then this will become clear.)

If $f = a_kx^k + a_{k-1}x^{k-1} + \cdots + a_0$ is a polynomial (in 1 variable $x$, with rational coefficients) such that $a_k \neq 0$ (each polynomial that is not just 0 can be uniquely written in such a form), then the integer $k$ is called the degree of $f$.

**Definition 2.17.23.** A polynomial $P$ with rational coefficients is said to be integer-valued if $(P(n) \in \mathbb{Z}$ for all $n \in \mathbb{Z}$).

Of course, a polynomial with integer coefficients is always integer-valued. But there are other integer-valued polynomials, too:

---

78For example, why cannot we start with (say) $6x^2 + 5x + 4$, then rewrite it as $(2x + 1)(3x + 1) + 3$, then do some other transformations (using commutativity, associativity and other laws), and finally end up with a polynomial that has different coefficients (say, $3x^2 + 9x + 4$) ? We cannot, but it is not easy to prove with what we have.
Example 2.17.24. (a) The polynomial \( \binom{x}{2} = \frac{x(x-1)}{2} = \frac{1}{2} x^2 - \frac{1}{2} x \) is integer-valued (since \( \binom{n}{2} \in \mathbb{Z} \) for each \( n \in \mathbb{Z} \)), but its coefficients are \( \frac{1}{2}, -\frac{1}{2}, 0 \).

(b) More generally: If \( k \in \mathbb{N} \) is arbitrary, then the polynomial \( \binom{x}{k} = x(x-1)(x-2) \cdots (x-k+1) / k! \) is integer-valued (since \( \binom{n}{k} \in \mathbb{Z} \) for each \( n \in \mathbb{Z} \)).

(c) If \( p \) is any prime, then the polynomial \( \frac{x^p - x}{p} \) is integer-valued (since Theorem 2.15.1 (b) yields \( a^p \equiv a \mod p \) for each \( a \in \mathbb{Z} \), which means that \( \frac{a^p - a}{p} \in \mathbb{Z} \) for each \( a \in \mathbb{Z} \)). Its coefficients are not integers.

This suggests the following question: How can we describe the integer-valued polynomials? The following result of Pólya \([Polya19]\) gives an answer:

Theorem 2.17.25. Let \( k \in \mathbb{N} \).

(a) Any polynomial \( P \) (in 1 variable \( x \), with rational coefficients) of degree \( k \) can be uniquely written in the form

\[
P(x) = a_k \binom{x}{k} + a_{k-1} \binom{x}{k-1} + \cdots + a_0 \binom{x}{0}
\]

with rational \( a_k, a_{k-1}, \ldots, a_0 \).

(b) The polynomial \( P \) is integer-valued if and only if these \( a_k, a_{k-1}, \ldots, a_0 \) are integers.

For example, the integer-valued polynomial \( \frac{x^3 - x}{3} \) can be written as

\[
\frac{x^3 - x}{3} = a_3 \binom{x}{3} + a_2 \binom{x}{2} + a_1 \binom{x}{1} + a_0 \binom{x}{0}
\]

for

\[
a_3 = 2, \quad a_2 = 2, \quad a_1 = 0, \quad a_0 = 0.
\]

These \( a_3, a_2, a_1, a_0 \) are integers – exactly as Theorem 2.17.25 (b) says.

I sketched a proof of Theorem 2.17.25 (b) in a talk in 2013 (\[http://www.cip.ifi.lmu.de/~grinberg/storrs2013.pdf\])\footnote{In this talk, I refer to integer-valued polynomials as “integral-valued polynomials”.} See also \[daSilv12\] for a self-contained proof.
2.18. Counting divisors

Now that we have seen some combinatorial reasoning (e.g., in the proof of Theorem 2.17.14), let us solve a rather natural counting problem: Let us count the divisors of a nonzero integer \( n \).

**Proposition 2.18.1.** Let \( n \in \mathbb{Z} \) be nonzero. Then:

(a) The product \( \prod_{p \text{ prime}} (v_p(n) + 1) \) is well-defined, since all but finitely many of its factors are 1.

(b) We have

\[
\text{(the number of positive divisors of } n) = \prod_{p \text{ prime}} (v_p(n) + 1).
\]

(c) We have

\[
\text{(the number of divisors of } n) = 2 \prod_{p \text{ prime}} (v_p(n) + 1).
\]

**Example 2.18.2.** If \( n = 12 \), then

\[
\text{(the number of positive divisors of } n) = 6
\]

(since the positive divisors of \( n = 12 \) are 1, 2, 3, 4, 6, 12) and

\[
\prod_{p \text{ prime}} (v_p(n) + 1) = \left( \frac{v_2(n) + 1}{2} \right) \left( \frac{v_3(n) + 1}{1} \right) \prod_{\substack{p \text{ prime;} \; p \notin \{2,3\} \atop = 0}} (v_p(n) + 1) = (2 + 1) (1 + 1) \prod_{\substack{p \text{ prime;} \; p \notin \{2,3\} \atop = 1}} 1 = (2 + 1) (1 + 1) = 6.
\]

This confirms Proposition 2.18.1 (b) for \( n = 12 \). In order to confirm Proposition 2.18.1 (c) for \( n = 12 \) as well, we observe that (the number of divisors of \( n \)) = 12 (since the divisors of \( n = 12 \) are \(-12, -6, -4, -3, -2, -1, 1, 2, 3, 4, 6, 12\)).

The function

\[
\{1, 2, 3, \ldots \} \rightarrow \mathbb{N},
\]

\[ n \mapsto \text{(the number of positive divisors of } n) \]

is known as the divisor function and is commonly denoted by \( \tau \). So Proposition 2.18.1 (b) gives a formula for \( \tau(n) \). See [Grinbe16, Theorem 2.1.7 (proof sketched in §2.7)] for a different proof of this formula.
Our proof of Proposition 2.18.1 will rely on the following lemma, which classifies all divisors of a positive integer in terms of its prime factorization:

**Lemma 2.18.3.** Let \( p_1, p_2, \ldots, p_u \) be finitely many distinct primes. For each \( i \in \{1, 2, \ldots, u\} \), let \( a_i \) be a nonnegative integer. Let \( n = p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u} \).

Define a set \( T \) by

\[
T = \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\} = \{(b_1, b_2, \ldots, b_u) \mid b_i \in \{0, 1, \ldots, a_i\} \text{ for each } i \in \{1, 2, \ldots, u\}\}.
\]

Then, the map

\[
\Λ : T \to \{\text{positive divisors of } n\},
\]

\[
(b_1, b_2, \ldots, b_u) \mapsto p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u}
\]

is well-defined and bijective.

**Example 2.18.4.** For this example, let \( u = 2, p_1 = 2, p_2 = 3, a_1 = 2 \) and \( a_2 = 1 \). Define the integer \( n \) and the set \( T \) as in Lemma 2.18.3 then,

\[
n = p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u} = 2^2 \cdot 3^1 = 12
\]

and

\[
T = \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\} = \{0, 1, 2\} \times \{0, 1\} = \{(0, 0), (0, 1), (1, 0), (1, 1), (2, 0), (2, 1)\}.
\]

Now, Lemma 2.18.3 says that the map

\[
\Λ : T \to \{\text{positive divisors of } n\},
\]

\[
(b_1, b_2, \ldots, b_u) \mapsto p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u}
\]

is well-defined and bijective. Here is a table of values of this map \( \Λ \):

<table>
<thead>
<tr>
<th>(\mathbf{b})</th>
<th>(0, 0)</th>
<th>(0, 1)</th>
<th>(1, 0)</th>
<th>(1, 1)</th>
<th>(2, 0)</th>
<th>(2, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Lambda(\mathbf{b}))</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

**Proof of Lemma 2.18.3** The numbers \( p_1, p_2, \ldots, p_u \) are primes, and thus positive integers. Hence, the product \( p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u} \) is a positive integer as well (since \( a_1, a_2, \ldots, a_u \) are nonnegative integers). In other words, \( n \) is a positive integer (since \( n = \)


Each $i \in \{1, 2, \ldots, u\}$ satisfies
\[
v_{p_i} \left( \frac{n}{p_1^{a_{i_1}} \cdots p_u^{a_{i_u}}} \right) = v_{p_i} \left( p_1^{a_{i_1}} p_2^{a_{i_2}} \cdots p_u^{a_{i_u}} \right) = a_i \quad \text{(89)}
\]
(by Exercise 2.13.7 (a)). Furthermore, if $p$ is a prime satisfying $p \not\in \{p_1, p_2, \ldots, p_u\}$, then
\[
v_{p_1} \left( \frac{n}{p_1^{a_{i_1}} \cdots p_u^{a_{i_u}}} \right) = v_{p_1} \left( p_1^{a_{i_1}} p_2^{a_{i_2}} \cdots p_u^{a_{i_u}} \right) = 0 \quad \text{(90)}
\]
(by Exercise 2.13.7 (b)).

For each $(b_1, b_2, \ldots, b_u) \in T$, we have
\[
p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u} \in \{ \text{positive divisors of } n \} \quad \text{(91)}
\]

[Proof of (91)]: Let $(b_1, b_2, \ldots, b_u) \in T$. We must prove (91).

We have \((b_1, b_2, \ldots, b_u) \in T = \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\}\). In other words, \(b_i \in \{0, 1, \ldots, a_i\}\) for each \(i \in \{1, 2, \ldots, u\}\). Hence, for each \(i \in \{1, 2, \ldots, u\}\), we have
\[
a_i - b_i \in \{0, 1, \ldots, a_i\} \quad \text{(since } b_i \in \{0, 1, \ldots, a_i\}\text{)}
\]
and thus \(p_i^{a_i - b_i}\) is an integer. Hence, \(\prod_{i=1}^{u} p_i^{a_i - b_i}\) is a product of integers, and thus is an integer as well. Now,
\[
n = \prod_{i=1}^{u} p_i^{a_i} = \prod_{i=1}^{u} \left( p_i^{b_i} p_i^{a_i - b_i} \right) = \left( \prod_{i=1}^{u} p_i^{b_i} \right) \left( \prod_{i=1}^{u} p_i^{a_i - b_i} \right)
\]
Thus, \(p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u} | n\) (since \(\prod_{i=1}^{u} p_i^{a_i - b_i}\) is an integer). In other words, \(p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u}\) is a divisor of \(n\). Hence, \(p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u}\) is a positive divisor of \(n\) (since \(p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u}\) is clearly positive). In other words, \(p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u} \in \{ \text{positive divisors of } n \}\). This proves (91).]

The equality (91) shows that the map \(\Lambda\) in Lemma 2.18.3 is well-defined. It remains to prove that it is bijective.
We shall achieve this by constructing an inverse to $\Lambda$.
Indeed, for each $d \in \{\text{positive divisors of } n\}$, we have

$$(v_{p_1}(d), v_{p_2}(d), \ldots, v_{p_a}(d)) \in T. \quad (92)$$

[Proof of (92): Let $d \in \{\text{positive divisors of } n\}$. Thus, $d$ is a positive divisor of $n$. In other words, $d$ is a positive integer satisfying $d \mid n$.

Fix $i \in \{1, 2, \ldots, u\}$. We shall show that $v_{p_i}(d) \in \{0, 1, \ldots, a_i\}$.

The integer $d$ is positive and thus nonzero. Hence, $v_{p_i}(d) \in \mathbb{N}$. But Proposition $2.13.35$ (applied to $d$ and $n$ instead of $n$ and $m$) shows that $d \mid n$ if and only if each prime $p$ satisfies $v_p(d) \leq v_p(n)$. Thus, each prime $p$ satisfies $v_p(d) \leq v_p(n)$ (since $d \mid n$). Applying this to $p = p_i$, we obtain $v_{p_i}(d) \leq v_{p_i}(n) = a_i$ (by $89$). Hence, $v_{p_i}(d) \in \{0, 1, \ldots, a_i\}$ (since $v_{p_i}(d) \in \mathbb{N}$).

Now, forget that we fixed $i$. We thus have shown that $v_{p_i}(d) \in \{0, 1, \ldots, a_i\}$ for each $i \in \{1, 2, \ldots, u\}$. In other words,

$$(v_{p_1}(d), v_{p_2}(d), \ldots, v_{p_a}(d)) \in \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\}. \quad (92)$$

This rewrites as $$(v_{p_1}(d), v_{p_2}(d), \ldots, v_{p_a}(d)) \in T$$ (since $T = \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\}$). Thus, $92$ is proven.]

We now define a map

$$V : \{\text{positive divisors of } n\} \to T,$$

$$d \mapsto (v_{p_1}(d), v_{p_2}(d), \ldots, v_{p_a}(d)).$$

This map is well-defined, because of $92$.

Now, we claim that $\Lambda \circ V = \text{id}$. \quad (92)

[Proof: Let $d \in \{\text{positive divisors of } n\}$. We shall show that $(\Lambda \circ V)(d) = \text{id}(d)$.

Indeed, $d$ is a positive divisor of $n$ (since $d \in \{\text{positive divisors of } n\}$). Hence, $d$ is a positive integer and satisfies $d \mid n$. But Proposition $2.13.35$ (applied to $d$ and $n$ instead of $n$ and $m$) shows that $d \mid n$ if and only if each prime $p$ satisfies $v_p(d) \leq v_p(n)$. Thus, each prime $p$ satisfies $v_p(d) \leq v_p(n)$ (since $d \mid n$). Hence, if $p$ is a prime satisfying $p \notin \{p_1, p_2, \ldots, p_a\}$, then we have

$$v_p(d) \leq v_p(n) = 0 \quad \text{(by } 90\text{)}$$

and therefore

$$v_p(d) = 0 \quad \text{(since } v_p(d) \in \mathbb{N} \cup \{\infty\} \text{ and } v_p(d) \leq 0)$$

and therefore

$$p^{v_p(d)} = p^0 = 1. \quad (93)$$

The elements $p_1, p_2, \ldots, p_u$ are distinct. Thus, the map $\{1, 2, \ldots, u\} \to \{p_1, p_2, \ldots, p_u\}, \ i \mapsto p_i$ is a bijection $80$.

---

80Indeed, this map is injective, since the elements $p_1, p_2, \ldots, p_u$ are distinct; and it is surjective, since its image is clearly $\{p_1, p_2, \ldots, p_u\}$.
But $d$ is a positive integer. Thus, Corollary 2.13.33 (applied to $d$ instead of $n$) yields

$$d = \prod_{p \text{ prime}} p^v_p(d) = \left( \prod_{p \in \{p_1, p_2, \ldots, p_u\}} p^v_p(d) \right) \left( \prod_{p \not\in \{p_1, p_2, \ldots, p_u\}} p^v_p(d) \right)$$

since each prime $p$ satisfies either $p \in \{p_1, p_2, \ldots, p_u\}$ or $p \not\in \{p_1, p_2, \ldots, p_u\}$ (but not both simultaneously)

$$= \left( \prod_{p \in \{p_1, p_2, \ldots, p_u\}} p^v_p(d) \right) \left( \prod_{p \not\in \{p_1, p_2, \ldots, p_u\}} 1 \right) = \prod_{p \in \{p_1, p_2, \ldots, p_u\}} p^v_p(d)$$

here, we have substituted $p_i$ for $p$ in the product,

(since the map $\{1, 2, \ldots, u\} \to \{p_1, p_2, \ldots, p_u\}$, $i \mapsto p_i$ is a bijection).

Comparing this with

$$((\Lambda \circ V)(d)) = \Lambda \left( \frac{V(d)}{(v_{p_1}(d), v_{p_2}(d), \ldots, v_{p_u}(d))} \right) = \Lambda \left( (v_{p_1}(d), v_{p_2}(d), \ldots, v_{p_u}(d)) \right)$$

(by the definition of $V$)

$$= p_1^{v_{p_1}(d)} p_2^{v_{p_2}(d)} \cdots p_u^{v_{p_u}(d)} = \prod_{i=1}^{u} p_i^{v_{p_i}(d)},$$

we obtain $(\Lambda \circ V)(d) = d = \text{id}(d)$.

Now, forget that we fixed $d$. We thus have shown that $(\Lambda \circ V)(d) = \text{id}(d)$ for each $d \in \{\text{positive divisors of } n\}$. In other words, $\Lambda \circ V = \text{id}$.

Next, we claim that $V \circ \Lambda = \text{id}$.

[Proof: Let $b \in T$. We shall show that $(V \circ \Lambda)(b) = \text{id}(b)$.

Indeed, we have $b \in T = \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\}$. Thus, $b$ is a $u$-tuple of nonnegative integers. Hence, write $b$ in the form $b = (b_1, b_2, \ldots, b_u)$ for some $u$ nonnegative integers $b_1, b_2, \ldots, b_u$. Then, the definition of $\Lambda$ yields $\Lambda(b) = p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u}$. Hence, for each $i \in \{1, 2, \ldots, u\}$, we have

$$v_{p_i}\left( \Lambda(b) \right) = v_{p_i}\left( p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u} \right) = b_i.$$
(by Exercise 2.13.7(a), applied to \(b_i\) instead of \(a_i\)). In other words,
\[
(v_{p_1}(\Lambda(b)), v_{p_2}(\Lambda(b)), \ldots, v_{p_u}(\Lambda(b))) = (b_1, b_2, \ldots, b_u).
\]

Now,
\[
(V \circ \Lambda)(b) = V(\Lambda(b)) = (v_{p_1}(\Lambda(b)), v_{p_2}(\Lambda(b)), \ldots, v_{p_u}(\Lambda(b))) = (b_1, b_2, \ldots, b_u) = b = \text{id}(b).
\]

Now, forget that we fixed \(b\). We have thus proven that \((V \circ \Lambda)(b) = \text{id}(b)\) for each \(b \in T\). In other words, \(V \circ \Lambda = \text{id}\).

We have now proven the equalities \(\Lambda \circ V = \text{id}\) and \(V \circ \Lambda = \text{id}\). These equalities show that the maps \(\Lambda\) and \(V\) are mutually inverse. Hence, the map \(\Lambda\) is invertible, i.e., bijective. This completes the proof of Lemma 2.18.3.

**Proof of Proposition 2.18.1** The integer \(|n|\) is positive (since \(n\) is nonzero) and thus nonzero. We observe that
\[
\{\text{positive divisors of } |n|\} = \{\text{positive divisors of } n\} \quad (94)
\]

Hence,
\[
\text{(the number of positive divisors of } |n|) = \text{(the number of positive divisors of } n) \quad (96)
\]

The same argument (but with the word “positive” removed) yields
\[
\text{(the number of divisors of } |n|) = \text{(the number of divisors of } n) \quad (97)
\]

**Proof of (94):** Let \(d \in \{\text{positive divisors of } |n|\}\). Thus, \(d\) is a positive divisor of \(|n|\). In other words, \(d\) is a positive integer and satisfies \(d \mid |n|\). But \(|n| \mid n\) (by Exercise 2.2.1(b), applied to \(a = n\)). Hence, Proposition 2.2.4(b) (applied to \(a = d, b = |n|\) and \(c = n\)) shows that \(d \mid n\). Thus, \(d\) is a positive integer and satisfies \(d \mid n\). In other words, \(d\) is a positive divisor of \(n\). In other words, \(d \in \{\text{positive divisors of } n\}\).

Now, forget that we fixed \(d\). We thus have proven that \(d \in \{\text{positive divisors of } n\}\) for each \(d \in \{\text{positive divisors of } |n|\}\). In other words,
\[
\{\text{positive divisors of } |n|\} \subseteq \{\text{positive divisors of } n\} \quad (95)
\]

Let \(e \in \{\text{positive divisors of } n\}\). Thus, \(e\) is a positive divisor of \(n\). In other words, \(e\) is a positive integer and satisfies \(e \mid n\). But \(n \mid |n|\) (by Exercise 2.2.1(a), applied to \(a = n\)). Hence, Proposition 2.2.4(b) (applied to \(a = e, b = n\) and \(c = |n|\)) shows that \(e \mid |n|\). Thus, \(e\) is a positive integer and satisfies \(e \mid |n|\). In other words, \(e\) is a positive divisor of \(|n|\). In other words, \(e \in \{\text{positive divisors of } |n|\}\).

Now, forget that we fixed \(e\). We thus have proven that \(e \in \{\text{positive divisors of } |n|\}\) for each \(e \in \{\text{positive divisors of } n\}\). In other words,
\[
\{\text{positive divisors of } n\} \subseteq \{\text{positive divisors of } |n|\} \quad (94)
\]

Combining this with (95), we obtain \(\{\text{positive divisors of } |n|\} = \{\text{positive divisors of } n\}\). Thus, (94) is proven.
Finally, Exercise 2.13.5 yields that

\[ \nu_p(\lfloor n \rfloor) = \nu_p(n) \quad \text{for each prime } p. \]  

(98)

The claim of Proposition 2.18.1 does not change if we replace \( n \) by \( \lfloor n \rfloor \) (because of (96), (97) and (98)). Thus, we can WLOG assume that \( n \geq 0 \) (since otherwise, we can just replace \( n \) by \( \lfloor n \rfloor \)). Assume this. Combining \( n \neq 0 \) (since \( n \) is nonzero) with \( n \geq 0 \), we find \( n > 0 \). Hence, \( n \) is a positive integer.

\( \textbf{(a)} \) For every prime \( p > \lfloor n \rfloor \), we have \( \nu_p(n) = 0 \) (by Lemma 2.13.32 \( \textbf{(a)} \)) and thus \( \nu_p(n) + 1 = 1 \). Thus, all but finitely many primes \( p \) satisfy \( \nu_p(n) + 1 = 1 \) (since all but finitely many primes \( p \) satisfy \( p > \lfloor n \rfloor \)). Therefore, all but finitely many factors of the product \( \prod_{p \text{ prime}} (\nu_p(n) + 1) \) are 1. In other words, the product \( \prod_{p \text{ prime}} (\nu_p(n) + 1) \) has only finitely many factors different from 1. Hence, this product is well-defined. This proves Proposition 2.18.1 \( \textbf{(a)} \).

\( \textbf{(b)} \) For every prime \( p > \lfloor n \rfloor \), we have \( \nu_p(n) = 0 \) (by Lemma 2.13.32 \( \textbf{(a)} \)). Thus, all but finitely many primes \( p \) satisfy \( \nu_p(n) = 0 \) (since all but finitely many primes \( p \) satisfy \( p > \lfloor n \rfloor \)). In other words, the set of all primes \( p \) satisfying \( \nu_p(n) \neq 0 \) is finite. Let \( P \) be this set. Thus, \( P \) is finite.

Let \( \{p_1, p_2, \ldots, p_u\} \) be a list of elements of \( P \), with no repetitions.\(^{82}\) Thus,

\[ \{p_1, p_2, \ldots, p_u\} = P. \]

The elements \( p_1, p_2, \ldots, p_u \) are distinct (since \( \{p_1, p_2, \ldots, p_u\} \) is a list with no repetitions). Thus, the map \( \{1, 2, \ldots, u\} \rightarrow \{p_1, p_2, \ldots, p_u\}, i \mapsto p_i \) is a bijection.\(^{83}\) Moreover, the elements \( p_1, p_2, \ldots, p_u \) belong to \( \{p_1, p_2, \ldots, p_u\} = P \), and thus are primes (since \( P \) is a set of primes).

If \( p \) is a prime such that \( p \notin \{p_1, p_2, \ldots, p_u\} \), then

\[ \nu_p(n) = 0. \]  

(99)

\[ \text{[Proof of (99)]} \] Recall that \( P \) is the set of all primes \( p \) satisfying \( \nu_p(n) \neq 0 \) (by the definition of \( P \)). Hence, every prime \( p \) satisfying \( \nu_p(n) \neq 0 \) must belong to \( P \). Thus, if \( p \) is a prime that does not belong to \( P \), then \( p \) cannot satisfy \( \nu_p(n) \neq 0 \). In other words, if \( p \) is a prime that does not belong to \( P \), then \( p \) must satisfy \( \nu_p(n) = 0 \). In other words, if \( p \) is a prime such that \( p \notin P \), then \( \nu_p(n) = 0 \). Since \( \{p_1, p_2, \ldots, p_u\} = P \), this rewrites as follows: If \( p \) is a prime such that \( p \notin \{p_1, p_2, \ldots, p_u\} \), then \( \nu_p(n) = 0 \). This proves (99).

If \( p \) is a prime such that \( p \notin \{p_1, p_2, \ldots, p_u\} \), then

\[ p^{\nu_p(n)} = p^0 \quad \text{(since (99) yields } \nu_p(n) = 0) \]

\[ = 1 \]  

(100)

\(^{82}\)Such a list exists, since \( P \) is finite.

\(^{83}\)Indeed, this map is injective, since the elements \( p_1, p_2, \ldots, p_u \) are distinct; and it is surjective, since its image is clearly \( \{p_1, p_2, \ldots, p_u\} \).
and
\[ v_p(n) + 1 = 1. \]  
(by (99))

For each \( i \in \{1, 2, \ldots, u\} \), define a nonnegative integer \( a_i \) by
\[ a_i = v_{p_i}(n). \]  
(102)

This is well-defined, since \( p_i \) is a prime (because \( p_1, p_2, \ldots, p_u \) are primes) and since \( n \) is nonzero.

Define a set \( T \) as in Lemma 2.18.3.

Recall that \( n \) is a positive integer. Thus, Corollary 2.13.33 yields

\[
n = \prod_{p \text{ prime}; p \in \{p_1, p_2, \ldots, p_u\}} p^{v_p(n)} = \left( \prod_{p \text{ prime}; p \in \{p_1, p_2, \ldots, p_u\}} p^{v_p(n)} \right) \left( \prod_{p \text{ prime}; p \not\in \{p_1, p_2, \ldots, p_u\}} p^{v_p(n)} \right)
\]

(100)

\[
= \left( \prod_{p \text{ prime}; p \in \{p_1, p_2, \ldots, p_u\}} p^{v_p(n)} \right) \left( \prod_{p \text{ prime}; p \not\in \{p_1, p_2, \ldots, p_u\}} (v_p(n) = 1) \right) = \prod_{p \in \{p_1, p_2, \ldots, p_u\}} p^{v_p(n)}
\]

(101)

(here, we have substituted \( p_i \) for \( p \) in the product, since each \( p_i \) is a prime in \( \{p_1, p_2, \ldots, p_u\} \), \( i \rightarrow p_i \) is a bijection)

\[
= \prod_{p \in \{p_1, p_2, \ldots, p_u\}} p^{v_p(n)} = \prod_{i=1}^u p_i^{v_{p_i}(n)} = p_i^{a_i} (\text{since (102) yields } v_{p_i}(n) = a_i)
\]

(102)

Hence, Lemma 2.18.3 shows that the map
\[
\Lambda : T \rightarrow \{ \text{positive divisors of } n \}, \quad (b_1, b_2, \ldots, b_u) \mapsto p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u}
\]
is well-defined and bijective.
Thus, there is a bijective map from $T$ to $\{\text{positive divisors of } n\}$ (namely, $\Lambda$). Hence,

$$|\{\text{positive divisors of } n\}|$$

$$= |T| = |\{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\}|$$  

(since $T = \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\}$)

$$= |\{0, 1, \ldots, a_1\}| \cdot |\{0, 1, \ldots, a_2\}| \cdots |\{0, 1, \ldots, a_u\}|$$  

(since the product rule $|A_1 \times A_2 \times \cdots \times A_u| = |A_1| \cdot |A_2| \cdots |A_u|$ holds whenever $A_1, A_2, \ldots, A_u$ are any $u$ finite sets)

$$= \prod_{i=1}^{u} |\{0, 1, \ldots, a_i\}| = \prod_{i=1}^{u} (a_i + 1).$$

Comparing this with

$$\prod_{p \text{ prime}} (v_p(n) + 1)$$

$$= \left( \prod_{p \in \{p_1, p_2, \ldots, p_u\}} (v_p(n) + 1) \right) \left( \prod_{p \text{ prime}; p \not\in \{p_1, p_2, \ldots, p_u\}} (v_p(n) + 1) \right)$$  

(by (101))

(since each prime $p$ satisfies either $p \in \{p_1, p_2, \ldots, p_u\}$ or $p \not\in \{p_1, p_2, \ldots, p_u\}$ (but not both simultaneously))

$$= \left( \prod_{p \text{ prime}; p \in \{p_1, p_2, \ldots, p_u\}} (v_p(n) + 1) \right) \left( \prod_{p \text{ prime}; p \not\in \{p_1, p_2, \ldots, p_u\}} 1 \right) = \prod_{p \text{ prime}; p \in \{p_1, p_2, \ldots, p_u\}} (v_p(n) + 1)$$

(since each $p \in \{p_1, p_2, \ldots, p_u\}$ is a prime)

$$= \prod_{p \in \{p_1, p_2, \ldots, p_u\}} (v_p(n) + 1) = \prod_{i=1}^{u} \left( \frac{v_{p_i}(n)}{a_i} + 1 \right)$$  

(by (102))

(here, we have substituted $p_i$ for $p$ in the product, since the map $\{1, 2, \ldots, u\} \to \{p_1, p_2, \ldots, p_u\}$, $i \mapsto p_i$ is a bijection)

$$= \prod_{i=1}^{u} (a_i + 1),$$

we obtain

$$|\{\text{positive divisors of } n\}| = \prod_{p \text{ prime}} (v_p(n) + 1).$$
Hence,
\[
\text{(the number of positive divisors of } n) = |\{\text{positive divisors of } n\}| = \prod_{p \text{ prime}} (v_p (n) + 1).
\]

This proves Proposition 2.18.1 (b).

(c) Every divisor of } n \text{ is either positive or negative}^{84} \text{(but clearly cannot be both at the same time). Hence,}
\[
\text{(the number of divisors of } n) = (\text{the number of positive divisors of } n) + (\text{the number of negative divisors of } n).
\]

If } d \text{ is a positive divisor of } n, \text{ then } -d \text{ is a negative divisor of } n. ^{85} \text{ Hence, we can define a map}
\[
A : \{\text{positive divisors of } n\} \rightarrow \{\text{negative divisors of } n\},
\]
\[d \mapsto -d.
\]
For similar reasons, we can define a map
\[
B : \{\text{negative divisors of } n\} \rightarrow \{\text{positive divisors of } n\},
\]
\[d \mapsto -d.
\]
Consider these two maps } A \text{ and } B. \text{ Clearly, } A \circ B = \text{id (since each negative divisor}
\[
d \text{ of } n \text{ satisfies } (A \circ B) (d) = A \left( B (d) \right) = A (-d) = - (-d) = d = \text{id (d)} \text{ and}
\]
\[
B \circ A = \text{id (similarly). Thus, these maps } A \text{ and } B \text{ are mutually inverse. Hence, the}
\text{map } A \text{ is invertible, i.e., a bijection.}
\text{Hence, there is a bijection between } \{\text{positive divisors of } n\} \text{ and } \{\text{negative divisors of } n\}
\text{(namely, } A). \text{ Thus,}
\[
|\{\text{negative divisors of } n\}| = |\{\text{positive divisors of } n\}|,
\]

\[84\text{Proof. Let } d \text{ be a divisor of } n. \text{ We must prove that } d \text{ is either positive or negative.}
\text{We have } d \mid n \text{ (since } d \text{ is a divisor of } n). \text{ Thus, there is an integer } e \text{ such that } n = de. \text{ Consider this } e. \text{ If we had } d = 0, \text{ then we would have } n = \underbrace{d}_{=0} e = 0 \text{, which would contradict the fact that}
\text{ } n \text{ is nonzero. Hence, we cannot have } d = 0. \text{ In other words, we have } d \neq 0. \text{ Thus, } d \text{ is a nonzero integer. Hence, } d \text{ is either positive or negative. Qed.}
\]

\[85\text{Proof. Let } d \text{ be a positive divisor of } n. \text{ We must prove that } -d \text{ is a negative divisor of } n. \text{ Clearly,}
\text{ } -d \text{ is negative (since } d \text{ is positive).}
\text{We have assumed that } d \text{ is a positive divisor of } n. \text{ In other words, } d \text{ is a positive integer and}
satisfies } d \mid n. \text{ But } d = (-d) (-1), \text{ thus, } -d \mid d \text{ (since } -1 \text{ is an integer). Hence, } -d \mid d \mid n. \text{ Hence,}
\text{ } -d \text{ is a divisor of } n. \text{ Thus, } -d \text{ is a negative divisor of } n \text{ (since } -d \text{ is negative). Qed.}
so that
\[
\text{(the number of negative divisors of } n) \\
= |\{\text{negative divisors of } n\}| = |\{\text{positive divisors of } n\}| \\
= \text{(the number of positive divisors of } n).
\]

Therefore,
\[
\text{(the number of divisors of } n) \\
= \text{(the number of positive divisors of } n) + \underbrace{\text{(the number of negative divisors of } n)}_{\text{=(the number of positive divisors of } n)} \\
= 2 \cdot \underbrace{\text{(the number of positive divisors of } n)}_{\text{=} 2 \prod_{p \text{ prime}} (v_p(n) + 1)} + \underbrace{\text{(the number of positive divisors of } n)}_{\text{=} 2 \prod_{p \text{ prime}} (v_p(n) + 1)}. \\
\]

\[
= \prod_{p \text{ prime}} (v_p(n) + 1).
\]

(by Proposition 2.18.1(b))

Hence, Proposition 2.18.1(c) follows. \(\square\)

**Remark 2.18.5.** Proposition 2.18.1 can be used to re-prove Proposition 2.14.7. We leave the details of this argument to the reader.

The method by which we proved Proposition 2.18.1 can be used (with a minor modification) to not just count the positive divisors of a positive integer \(n\), but also (for example) to compute their sum or the sum of their squares. This relies on the following basic property of \(\sum\) and \(\prod\) signs:

**Lemma 2.18.6.** Let \(n \in \mathbb{N}\). For every \(i \in \{1, 2, \ldots, n\}\), let \(Z_i\) be a finite set. For every \(i \in \{1, 2, \ldots, n\}\) and every \(k \in Z_i\), let \(p_{i,k}\) be a number. Then,
\[
\prod_{i=1}^{n} \sum_{k \in Z_i} p_{i,k} = \sum_{(k_1, k_2, \ldots, k_n) \in Z_1 \times Z_2 \times \cdots \times Z_n} \prod_{i=1}^{n} p_{i,k_i}.
\]

(Note that if \(n = 0\), then the Cartesian product \(Z_1 \times Z_2 \times \cdots \times Z_n\) has no factors; it is what is called an *empty Cartesian product*. It is understood to be a 1-element set, and its single element is the 0-tuple (\(\) (also known as the empty list).)

Lemma 2.18.6 is essentially a version of the distributivity law (or the FOIL method) for expanding a product of several sums, each of which has several factors. For example, if we take \(n = 3\) and \(Z_i = \{1, 2\}\) for each \(i \in \{1, 2, 3\}\), then Lemma 2.18.6 says that
\[
(p_{1,1} + p_{1,2}) (p_{2,1} + p_{2,2}) (p_{3,1} + p_{3,2}) \\
= p_{1,1}p_{2,1}p_{3,1} + p_{1,1}p_{2,1}p_{3,2} + p_{1,1}p_{2,2}p_{3,1} + p_{1,1}p_{2,2}p_{3,2} \\
+ p_{1,2}p_{2,1}p_{3,1} + p_{1,2}p_{2,1}p_{3,2} + p_{1,2}p_{2,2}p_{3,1} + p_{1,2}p_{2,2}p_{3,2}
\]
(which is precisely what you get if you expand the product 
\((p_{1,1} + p_{1,2}) (p_{2,1} + p_{2,2}) (p_{3,1} + p_{3,2})\) using the distributivity law). For another example, if we take \(n = 2\) and \(Z_i = \{1, 2, 3\}\) for each \(i \in \{1, 2\}\), then Lemma 2.18.6 says that
\[
(p_{1,1} + p_{1,2} + p_{1,3}) (p_{2,1} + p_{2,2} + p_{2,3}) = p_{1,1}p_{2,1} + p_{1,1}p_{2,2} + p_{1,1}p_{2,3} \\
+ p_{1,2}p_{2,1} + p_{1,2}p_{2,2} + p_{1,2}p_{2,3} \\
+ p_{1,3}p_{2,1} + p_{1,3}p_{2,2} + p_{1,3}p_{2,3}
\]
(which is, again, simply the result of expanding the left hand side). In the general case, the idea behind Lemma 2.18.6 is that if you expand the product
\[
\prod_{i=1}^{n} \sum_{k=1}^{m_i} p_{i,k}
\]
\[
= \prod_{i=1}^{n} (p_{i,1} + p_{i,2} + \cdots + p_{i,m_i})
\]
\[
= (p_{1,1} + p_{1,2} + \cdots + p_{1,m_1}) (p_{2,1} + p_{2,2} + \cdots + p_{2,m_2}) \cdots (p_{n,1} + p_{n,2} + \cdots + p_{n,m_n})
\]
then you get a sum of \(m_1m_2 \cdots m_n\) terms, each of which has the form
\[
p_{1,k_1}p_{2,k_2} \cdots p_{n,k_n} = \prod_{i=1}^{n} p_{i,k_i}
\]
for some \((k_1, k_2, \ldots, k_n) \in \{1, 2, \ldots, m_1\} \times \{1, 2, \ldots, m_2\} \times \cdots \times \{1, 2, \ldots, m_n\}\). See [Grinbe15] proof of Lemma 7.160 for a rigorous proof of Lemma 2.18.6 (which uses induction and the distributivity law).

Now, we can state a formula for the sum of all positive divisors of a positive integer \(n\), and more generally for the sum of the \(k\)-th powers of these positive divisors, where \(k\) is a fixed integer:

**Exercise 2.18.1.** Let \(n\) be a positive integer. Let \(k \in \mathbb{Z}\). Prove that:

(a) The product \(\prod_{p \text{ prime}} \left( p^{0k} + p^{1k} + \cdots + p^{v_p(n)k} \right)\) is well-defined, since all but finitely many of its factors are 1.

(b) We have
\[
\sum_{d|n} d^k = \prod_{p \text{ prime}} \left( p^{0k} + p^{1k} + \cdots + p^{v_p(n)k} \right).
\]

(Recall that the summation sign “\(\sum\)” means a sum over all positive divisors \(d\) of \(n\).)

---

66We are here assuming (for the sake of simplicity) that each set \(Z_i\) is \(\{1, 2, \ldots, m_i\}\) for some \(m_i \in \mathbb{N}\).

This does not weaken the reach of Lemma 2.18.6 since each finite set \(Z_i\) can be relabelled as \(\{1, 2, \ldots, m_i\}\) for \(m_i = |Z_i|\).
Example 2.18.7. If \( n = 6 \), then the positive divisors of \( n \) are 1, 2, 3, 6. Thus, in this case, the claim of Exercise 2.18.1 \((b)\) becomes

\[
1^k + 2^k + 3^k + 6^k = \prod_{p \text{ prime}} \left( p^{0k} + p^{1k} + \ldots + p^{v_p(6)k} \right).
\]

This equality can easily be verified, since the right hand side is

\[
\prod_{p \text{ prime}} \left( p^{0k} + p^{1k} + \ldots + p^{v_p(6)k} \right)
= \left(2^{0k} + 2^{1k} + \ldots + 2^{v_2(6)k} \right) \cdot \left(3^{0k} + 3^{1k} + \ldots + 3^{v_3(6)k} \right)
\]

\[
\cdot \prod_{p \text{ prime}; p\not\in\{2,3\}} \left( p^{0k} + p^{1k} + \ldots + p^{v_p(6)k} \right)
\]

\[
= \left(2^{0k} + 2^{1k} \right) \cdot \left(3^{0k} + 3^{1k} \right) \cdot \prod_{p \text{ prime}; p\not\in\{2,3\}} p^{0k}
\]

\[
= \left(1 + 2^{k} \right) \cdot \left(1 + 3^{k} \right) = \frac{1}{1^{k}} + 2^{k} + 3^{k} + 2^{k} \cdot 3^{k} = 1^k + 2^k + 3^k + 6^k.
\]

Note that Proposition 2.18.1 \((b)\) is the particular case of Exercise 2.18.1 \((b)\) obtained when setting \( k = 0 \) (because each integer \( z \) satisfies \( z^0 = 1 \), and thus \( \sum_{d|n} d^0 \) is the number of positive divisors of \( n \)).

3. Equivalence relations and residue classes

3.1. Relations

Loosely speaking, a relation on a set \( S \) is a property that two elements \( a \) and \( b \) of \( S \) (or, more formally, a pair \((a, b) \in S \times S\) of two elements of \( S\)) can either have or not have. For example, equality (denoted \( = \)) is a relation, since two elements \( a \) and \( b \) of \( S \) are either equal (i.e., satisfy \( a = b \)) or not equal. Likewise, the divisibility relation (denoted \( | \)) is a relation on \( \mathbb{Z} \), since two elements \( a \) and \( b \) of \( \mathbb{Z} \) either satisfy \( a | b \) or do not.

A formal definition of relations proceeds as follows:
Definition 3.1.1. Fix a set $S$. A binary relation on $S$ is a subset of $S \times S$ (that is, a set of pairs of elements of $S$).

If $R$ is a binary relation (on $S$), and if $a, b \in S$, then we write $aRb$ for $(a, b) \in R$.

The word “relation” shall always mean “binary relation” unless we say otherwise.

So a relation on a set $S$ is, formally speaking, a subset of $S \times S$—but in practice, we think of it as a property that holds for some pairs $(a, b) \in S \times S$ (namely, for the ones that belong to this subset) and does not hold for some others (namely, for the ones that do not belong to this subset). In order to define a relation $R$ on a given set $S$, it suffices to tell which pairs $(a, b) \in S \times S$ satisfy $aRb$ (because then, $R$ will simply be the set of all these pairs $(a, b)$). Let us define several relations on the set $\mathbb{Z}$ by this strategy:

Example 3.1.2. Let $S = \mathbb{Z}$.

(a) The relation $=$ is a binary relation on $S$. As a subset of $S \times S$, this relation is

$$\{(a, b) \in S \times S \mid a = b\} = \{(c, c) \mid c \in S\} = \{\ldots, (-1, -1), (0, 0), (1, 1), \ldots\}.$$  

(b) The relation $<$ is a binary relation on $S$. As a subset of $S \times S$, this relation is

$$\{(a, b) \in S \times S \mid a < b\}.$$  

(c) The relation $\leq$ is a binary relation on $S$. As a subset of $S \times S$, this relation is

$$\{(a, b) \in S \times S \mid a \leq b\}.$$  

(d) The relation $\neq$ is also a binary relation on $S$.

(e) Fix $n \in \mathbb{Z}$. Define a relation $\equiv$ on $S = \mathbb{Z}$ by

$$\left(\frac{a}{n} b\right) \iff (a \equiv b \mod n).$$  

As a subset of $S \times S = \mathbb{Z} \times \mathbb{Z}$, this relation $\equiv$ is

$$\{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid a \equiv b \mod n\} = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid \text{there exists an integer } d \text{ such that } b = a + nd\}$$

(by Exercise 2.3.7)

$$= \{(a, a + nd) \mid a, d \in \mathbb{Z}\}.$$  

Note that the relation $\equiv_0$ is exactly the relation $=$ (by Example 2.3.2(e)).

---

\[\text{footnote:} \text{Here, the word "some" can mean "none" or "all" or anything inbetween.}\]
(f) Define a binary relation \( N \) on \( S \) by
\[
(a N b) \iff \text{(false)}
\]
(that is, \( a N b \) never holds, no matter what \( a \) and \( b \) are). As a subset of \( S \times S \), this relation \( N \) is just the empty subset of \( S \times S \).

(g) On the other extreme: Define a binary relation \( A \) on \( S \) by
\[
(a A b) \iff \text{(true)}
\]
(that is, \( a A b \) holds for all \( a \) and \( b \)). As a subset of \( S \times S \), this relation \( A \) is the whole set \( S \times S \). Note that the relation \( A \) is exactly the relation \( \equiv \) (by Example 2.3.2 (d)).

(h) The relation \( \mid \) (divisibility) is also a relation on \( S = \mathbb{Z} \).

(i) The relation \( \perp \) (coprimality) is also a relation on \( S = \mathbb{Z} \).

(j) We have defined several relations on the set \( S = \mathbb{Z} \) now. The relations \( =, \neq, N \) and \( A \) (or, rather, relations analogous to them) can be defined on any set.

### 3.2. Equivalence relations

Relations occur frequently in mathematics, and there is a bunch of properties that a relation can have or not have. (See the Wikipedia article on binary relations for a long list of such properties.) We shall need only the following three:

**Definition 3.2.1.** Let \( R \) be a binary relation on a set \( S \).

(a) We say that \( R \) is **reflexive** if every \( a \in S \) satisfies \( aRa \).

(b) We say that \( R \) is **symmetric** if every \( a, b \in S \) satisfying \( aRb \) satisfy \( bRa \).

(c) We say that \( R \) is **transitive** if every \( a, b, c \in S \) satisfying \( aRb \) and \( bRc \) satisfy \( a Rc \).

(Here are mnemonics for the three words we just defined:

- “Reflexive” should make you think of \( R \) as a mirror through which \( a \) can see itself (that is, satisfy \( aRa \)).

- “Symmetric” means that the roles of \( a \) and \( b \) in \( aRb \) are interchangeable – a symmetry.

- “Transitive” means that you can “transit” an element \( b \) on your way from \( a \) to \( c \) (that is, if you treat \( aRb \) as the existence of a “path” from \( a \) to \( b \), and \( bRc \) as the existence of a “path” from \( b \) to \( c \), then you can combine a “path” from \( a \) to \( b \) with a “path” from \( b \) to \( c \) to get a “path” from \( a \) to \( c \)).
Let us see some examples of these properties of relations.

**Example 3.2.2.** Let $S$ be the set $\mathbb{Z}$. Consider the relations on $\mathbb{Z}$ defined in Example 3.1.2.

(a) The relation $=$ is reflexive, symmetric and transitive.
(b) The relation $<$ is transitive, but neither reflexive nor symmetric.
(c) The relation $\leq$ is transitive and reflexive, but not symmetric.
(d) The relation $\neq$ is symmetric, but neither reflexive nor transitive.
(e) For each $n \in \mathbb{Z}$, the relation $\equiv_n$ is reflexive, symmetric and transitive.
(f) The relation $\equiv_n$ is symmetric and transitive, but not reflexive.
(g) The relation $\equiv_{\equiv_{\equiv}}$ is reflexive, symmetric and transitive.
(h) The divisibility relation $|$ is reflexive and transitive, but not symmetric.
(i) The coprimality relation $\perp$ is symmetric, but neither reflexive nor transitive.

**Proof of Example 3.2.2.** (a) Indeed:

- The relation $=$ is reflexive, because every $a \in S$ satisfies $a = a$.
- The relation $=$ is symmetric, because every $a, b \in S$ satisfying $a = b$ satisfy $b = a$.
- The relation $=$ is transitive, because every $a, b, c \in S$ satisfying $a = b$ and $b = c$ satisfy $a = c$.

(b) Indeed:

- The relation $<$ is transitive (because every $a, b, c \in S$ satisfying $a < b$ and $b < c$ satisfy $a < c$).
- Not every $a \in S$ satisfies $a < a$ (in fact, no $a \in S$ satisfies $a < a$); thus, $<$ is not reflexive.
- Similarly, $<$ is not symmetric, since $a < b$ does not imply $b < a$ (quite the opposite).

(c) Indeed:

- The relation $\leq$ is transitive (because every $a, b, c \in S$ satisfying $a \leq b$ and $b \leq c$ satisfy $a \leq c$).
- The relation $\leq$ is reflexive (since every $a \in S$ satisfies $a \leq a$).
- The relation $\leq$ is not symmetric (since $a \leq b$ does not imply $b \leq a$; for example, $1 \leq 2$ holds but $2 \leq 1$ does not).

See further below for the proofs of the claims made in this example.
(d) Indeed:

- The relation $\neq$ is symmetric (because every $a, b \in S$ satisfying $a \neq b$ satisfy $b \neq a$).
- The relation $\neq$ is not reflexive (since we don’t have $2 \neq 2$).
- The relation $\neq$ is not transitive (since $2 \neq 3$ and $3 \neq 2$ do not lead to $2 \neq 2$).

(e) Let $n \in \mathbb{Z}$.

- Proposition 2.3.4 (b) shows that every $a, b, c \in \mathbb{Z}$ satisfying $a \equiv b \mod n$ and $b \equiv c \mod n$ satisfy $a \equiv c \mod n$. In other words, every $a, b, c \in S$ satisfying $a \equiv b$ and $b \equiv c$ satisfy $a \equiv c$ (since the definition of $\equiv$ shows that the three statements

$$(a \equiv b) \mod n, \quad (b \equiv c) \mod n, \quad (a \equiv c) \mod n,$$

are equivalent to

$$(a \equiv b \mod n), \quad (b \equiv c \mod n), \quad (a \equiv c \mod n),$$

respectively). But this means precisely that the relation $\equiv$ is transitive.
- Similarly, the relation $\equiv$ is reflexive (by Proposition 2.3.4 (a)).
- Similarly, the relation $\equiv$ is symmetric (by Proposition 2.3.4 (c)).

(f) This may appear strange, but is a completely straightforward consequence of the concept of “vacuous truth”:

- Every $a, b \in S$ satisfying $a \nmid b$ satisfy $b \nmid a$ (because there are no such $a, b$ to begin with – since $a \nmid b$ never holds). Thus, $\nmid$ is symmetric.
- Similarly, $\nmid$ is transitive.
- But $\nmid$ is not reflexive, since (for example) $1 \nmid 1$ does not hold.

(g) All of this is trivial, because $a \nmid b$ holds for all $a, b \in S$.

(h) The divisibility relation $|$ is reflexive (by Proposition 2.2.4 (a)) and transitive (by Proposition 2.2.4 (b)), but not symmetric (since $1 \mid 2$ does not lead to $2 \mid 1$).

(i) The coprimality relation is symmetric (by Proposition 2.10.4), but neither reflexive (since we don’t have $2 \perp 2$) nor transitive (since $2 \perp 3$ and $3 \perp 2$ do not lead to $2 \perp 2$).
**Definition 3.2.3.** An *equivalence relation* on a set $S$ means a relation on $S$ that is reflexive, symmetric and transitive.

**Example 3.2.4.** Let $S$ be any set. The relation $=$ on the set $S$ is an equivalence relation, because it is reflexive, symmetric and transitive.

**Example 3.2.5.** Let $n \in \mathbb{Z}$. The relation relation $\equiv_n$ on $\mathbb{Z}$ (defined in Example 3.1.2(e)) is an equivalence relation, because (as we saw in Example 3.2.2(e)) it is reflexive, symmetric and transitive.

**Example 3.2.6.** Here are some examples from elementary plane geometry: Congruence (e.g., of triangles) is an equivalence relation. Similarity is also an equivalence relation. The same holds for direct similarity (i.e., orientation-preserving similarity). The same holds for parallelism of lines.

**Example 3.2.7.** Let $S$ and $T$ be two sets, and let $f : S \to T$ be a map. Define a relation $\equiv_f$ on $S$ by

$$(a \equiv_f b) \iff (f(a) = f(b)).$$

This relation $\equiv_f$ is an equivalence relation.

**Proof of Example 3.2.7** Indeed:

- The relation $\equiv_f$ is reflexive, because every $a \in S$ satisfies $a \equiv_f a$ (since $f(a) = f(a)$).

- The relation $\equiv_f$ is symmetric, because every $a, b \in S$ satisfying $a \equiv_f b$ satisfy $b \equiv_f a$. (Indeed, $a \equiv_f b$ means $f(a) = f(b)$, which entails $f(b) = f(a)$, which in turn rewrites as $b \equiv_f a$.)

- The relation $\equiv_f$ is transitive, because every $a, b, c \in S$ satisfying $a \equiv_f b$ and $b \equiv_f c$ satisfy $a \equiv_f c$. (Indeed, the assumptions $a \equiv_f b$ and $b \equiv_f c$ rewrite as $f(a) = f(b)$ and $f(b) = f(c)$; therefore, $f(a) = f(b) = f(c)$, which rewrites as $a \equiv_f c$.)

Thus, $\equiv_f$ is an equivalence relation. \hfill $\Box$

We will soon learn that every equivalence relation on a set $S$ is actually of the form $\equiv_f$ for some set $T$ and some map $f : S \to T$. (Namely, this is proven in Exercise 3.3.3 below.)
Example 3.2.8. Let $S$ be the set of all points on the landmass of the Earth, and let $\sim$ be the relation on $S$ defined by

$$(a \sim b) \iff \text{(there is a land route from $a$ to $b$)}.$$ 

This $\sim$ is an equivalence relation (with the caveat that $S$ is not a mathematical object and thus not really well-defined).

Example 3.2.9. Let

$$S = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\}) = \{(a_1, a_2) \mid a_1 \in \mathbb{Z} \text{ and } a_2 \in \mathbb{Z} \setminus \{0\}\}.$$ 

This is the set of all pairs whose first entry is an integer and whose second entry is a nonzero integer. We define a relation $\sim_*$ on $S$ by

$$((a_1, a_2) \sim_* (b_1, b_2)) \iff (a_1 b_2 = a_2 b_1).$$ 

This relation $\sim_*$ is an equivalence relation.

Proof of Example 3.2.9. Indeed:

- The relation $\sim_*$ is reflexive.

  [Proof: Let $a \in S$. Thus, $a \in S = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$; in other words, we can write $a$ as $a = (a_1, a_2)$ for some $a_1 \in \mathbb{Z}$ and $a_2 \in \mathbb{Z} \setminus \{0\}$. Consider these $a_1$ and $a_2$.

  Clearly, $a_1 a_2 = a_2 a_1$. In other words, $(a_1, a_2) \sim_* (a_1, a_2)$ (because the definition of the relation $\sim$ yields that $(a_1, a_2) \sim (a_1, a_2)$ means $a_1 a_2 = a_2 a_1$). In other words, $a \sim_* a$ (since $a = (a_1, a_2)$).

  Now, forget that we fixed $a$. We thus have shown that every $a \in S$ satisfies $a \sim_* a$. In other words, the relation $\sim_*$ is reflexive.]

- The relation $\sim_*$ is symmetric.

  [Proof: Let $a, b \in S$ be such that $a \sim_* b$. We shall prove that $b \sim_* a$.

  We have $a \in S = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$; in other words, we can write $a$ as $a = (a_1, a_2)$ for some $a_1 \in \mathbb{Z}$ and $a_2 \in \mathbb{Z} \setminus \{0\}$. Similarly, we can write $b$ as $b = (b_1, b_2)$ for some $b_1 \in \mathbb{Z}$ and $b_2 \in \mathbb{Z} \setminus \{0\}$. Consider these $a_1, a_2, b_1$ and $b_2$.

  We have assumed that $a \sim_* b$. In other words, $(a_1, a_2) \sim_*(b_1, b_2)$ (since $a = (a_1, a_2)$ and $b = (b_1, b_2)$). In other words, $a_1 b_2 = b_2 a_1$ (because this is what $(a_1, a_2) \sim_*(b_1, b_2)$ means, by the definition of the relation $\sim_*$). Thus, $b_2 a_1 = a_1 b_2 = a_2 b_1 = b_1 a_2$; in other words, $b_1 a_2 = b_2 a_1$. In other words, $(b_1, b_2) \sim_*(a_1, a_2)$ (by the definition of the relation $\sim_*$). In other words, $b \sim_* a$ (since $a = (a_1, a_2)$ and $b = (b_1, b_2)$).]
Now, forget that we fixed $a$ and $b$. We thus have shown that every $a, b \in S$ satisfying $a \sim b$ satisfy $b \sim a$. In other words, the relation $\sim$ is symmetric.]

- The relation $\sim$ is transitive.

[Proof: Let $a, b, c \in S$ be such that $a \sim b$ and $b \sim c$. We shall prove that $a \sim c$.

We have $a \in S = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$; in other words, we can write $a$ as $a = (a_1, a_2)$ for some $a_1 \in \mathbb{Z}$ and $a_2 \in \mathbb{Z} \setminus \{0\}$. Similarly, we can write $b$ as $b = (b_1, b_2)$ for some $b_1 \in \mathbb{Z}$ and $b_2 \in \mathbb{Z} \setminus \{0\}$. Similarly, we can write $c$ as $c = (c_1, c_2)$ for some $c_1 \in \mathbb{Z}$ and $c_2 \in \mathbb{Z} \setminus \{0\}$. Consider these $a_1, a_2, b_1, b_2, c_1$ and $c_2$. Note that $b_2 \in \mathbb{Z} \setminus \{0\}$, so that $b_2 \neq 0$.

We have assumed that $a \sim b$. In other words, $(a_1, a_2) \sim (b_1, b_2)$ (since $a = (a_1, a_2)$ and $b = (b_1, b_2)$). In other words, $a_1 b_2 = a_2 b_1$ (by the definition of the relation $\sim$). Similarly (by exploiting the assumption $b \sim c$ instead of $a \sim b$), we can obtain $b_1 c_2 = b_2 c_1$. Hence,

$\begin{align*}
(a_1 b_2) c_2 &= a_2 b_1 c_2 = a_2 b_2 c_1,
\end{align*}$

We can cancel $b_2$ from this equality (since $b_2 \neq 0$), and thus obtain $a_1 c_2 = a_2 c_1$. In other words, $(a_1, a_2) \sim (c_1, c_2)$ (by the definition of the relation $\sim$). In other words, $a \sim c$ (since $a = (a_1, a_2)$ and $c = (c_1, c_2)$).

Now, forget that we fixed $a, b, c$. We thus have shown that every $a, b, c \in S$ satisfying $a \sim b$ and $b \sim c$ satisfy $a \sim c$. In other words, the relation $\sim$ is transitive.]

We have now proven that the relation $\sim$ is reflexive, symmetric and transitive. In other words, $\sim$ is an equivalence relation (by the definition of “equivalence relation”). This proves Example 3.2.9. □

The relation $\sim$ from Example 3.2.9 may appear familiar to you. In fact, its definition can be restated as follows:

$$(a_1, a_2) \sim (b_1, b_2) \iff \left( \frac{a_1}{a_2} = \frac{b_1}{b_2} \right),$$

and this makes the claims of Example 3.2.9 a lot more obvious. However, this is (in a sense) circular reasoning: The statement “$\frac{a_1}{a_2} = \frac{b_1}{b_2}$” only makes sense if the rational numbers have been defined, but the definition of rational numbers (at least the usual definition, given in [Swanso18, §3.6] and in many other places) already relies on the claims of Example 3.2.9 (Namely, the rational numbers are defined as the equivalence classes of the relation $\sim$; this is explained in Example 3.3.6 below.) Thus, our above proof of Example 3.2.9 was not a waste of time, but

\[\text{since } \frac{a_1}{a_2} \text{ and } \frac{b_1}{b_2} \text{ are (in general) not integers but rational numbers,}\]
rather an important prerequisite for the construction of rational numbers (one of the cornerstones of mathematics).

If you are familiar with basic linear algebra, you may notice that the relation \(\sim\) from Example 3.2.9 can also be regarded as linear dependence. Namely, two pairs \((a_1, a_2)\) and \((b_1, b_2)\) in \(\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})\) satisfy \((a_1, a_2) \sim (b_1, b_2)\) if and only if the vectors \((a_1, a_2)\) and \((b_1, b_2)\) in \(\mathbb{Q}^2\) are linearly dependent.

One simple property of symmetric relations will come useful:

**Proposition 3.2.10.** Let \(\sim\) be a symmetric relation on a set \(S\). Let \(a, b \in S\). Then, \(a \sim b\) if and only if \(b \sim a\).

**Proof of Proposition 3.2.10.** The relation \(\sim\) is symmetric. Thus, if \(a\) and \(b\) satisfy \(a \sim b\), then they also satisfy \(b \sim a\) (by the definition of “symmetric”). In other words, we have the logical implication \((a \sim b) \implies (b \sim a)\). But the same argument (with the roles of \(a\) and \(b\) interchanged) yields the implication \((b \sim a) \implies (a \sim b)\). Combining these two implications, we obtain the equivalence \((a \sim b) \iff (b \sim a)\). This proves Proposition 3.2.10. \(\square\)

### 3.3. Equivalence classes

#### 3.3.1. Definition of equivalence classes

We can now state one of the most important definitions in mathematics:

**Definition 3.3.1.** Let \(\sim\) be an equivalence relation on a set \(S\).

(a) For each \(a \in S\), we define a subset \([a]_\sim\) of \(S\) by

\[
[a]_\sim = \{ b \in S \mid b \sim a \}.
\]

This subset \([a]_\sim\) is called the **equivalence class** of \(a\), or the \(\sim\)-equivalence class of \(a\).

(b) The **equivalence classes** of \(\sim\) are defined to be the sets \([a]_\sim\) for \(a \in S\). They are also known as the \(\sim\)-equivalence classes.

**Example 3.3.2.** Consider the relation \(\equiv_3\) on \(\mathbb{Z}\) (defined in Example 3.1.2 (e)). We have

\[
[5]_{\equiv_3} = \left\{ b \in \mathbb{Z} \mid b \equiv 5 \pmod{3} \right\} = \left\{ b \in \mathbb{Z} \mid b \equiv 5 \pmod{3} \right\} = \{ \ldots, -4, -1, 2, 5, 8, 11, 14, \ldots \}
\]

\[\text{Note, however, that linear dependence is no longer an equivalence relation if we allow the vector (0,0) in our set } S, \text{ because then, it is no longer transitive (for example, (1,1) and (0,0) are linearly dependent, and (0,0) and (1,2) are linearly dependent, but (1,1) and (1,2) are not).}\]
and

\[ [3]_3 = \left\{ b \in \mathbb{Z} \mid b \equiv 3 \pmod{3} \right\} = \left\{ b \in \mathbb{Z} \mid b \equiv 3 \text{ mod } 3 \right\} = \{\ldots, -6, -3, 0, 3, 6, 9, 12, \ldots \} \]

and

\[ [2]_3 = \left\{ b \in \mathbb{Z} \mid b \equiv 2 \pmod{3} \right\} = \left\{ b \in \mathbb{Z} \mid b \equiv 2 \text{ mod } 3 \right\} = \{\ldots, -4, -1, 2, 5, 8, 11, 14, \ldots \} \]

Note that \([5]_3 = [2]_3\), as you can easily see.

### 3.3.2. Basic properties

**Proposition 3.3.3.** Let \(\sim\) be an equivalence relation on a set \(S\). Let \(a \in S\). Then,

\[ [a]_\sim = \{b \in S \mid a \sim b\}. \]

**Proof of Proposition 3.3.3.** The relation \(\sim\) is symmetric (since it is an equivalence relation). Thus, for any \(b \in S\), we have \((a \sim b\) if and only if \(b \sim a\) (by Proposition 3.2.10). Hence, \([b \in S \mid a \sim b] = [b \in S \mid b \sim a]\). Comparing this with (103), we obtain \([a]_\sim = \{b \in S \mid a \sim b\}\). This proves Proposition 3.3.3.

Proposition 3.3.3 shows that we can replace the condition “\(b \sim a\)” by “\(a \sim b\)” in Definition 3.3.1 (a) without changing the meaning of the definition. (Some authors, such as Swanson in [Swanson18, Definition 2.3.6], do exactly that.)

**Proposition 3.3.4.** Let \(\sim\) be an equivalence relation on a set \(S\). Let \(a \in S\). Then, \(a \in [a]_\sim\).

**Proof of Proposition 3.3.4.** The relation \(\sim\) is reflexive (since it is an equivalence relation). Thus, \(a \sim a\). In other words, \(a\) is a \(b \in S\) satisfying \(b \sim a\). In other words, \(a \in \{b \in S \mid b \sim a\}\). But \([a]_\sim = \{b \in S \mid b \sim a\}\) (by the definition of \([a]_\sim\)). Hence, \(a \in \{b \in S \mid b \sim a\} = [a]_\sim\). This proves Proposition 3.3.4.

Proposition 3.3.4 shows that all equivalence classes of an equivalence relation are nonempty sets (because each equivalence class \([a]_\sim\) contains at least the element \(a\)).

**Theorem 3.3.5.** Let \(\sim\) be an equivalence relation on a set \(S\). Let \(x, y \in S\).

(a) If \(x \sim y\), then \([x]_\sim = [y]_\sim\).

(b) If not \(x \sim y\), then the sets \([x]_\sim\) and \([y]_\sim\) are disjoint.

(c) We have \(x \sim y\) if and only if \(x \in [y]_\sim\).

(d) We have \(x \sim y\) if and only if \(y \in [x]_\sim\).

(e) We have \(x \sim y\) if and only if \([x]_\sim = [y]_\sim\).
Proof of Theorem 3.3.5. The relation ∼ is transitive (since it is an equivalence relation) and symmetric (for the same reason).

The definition of \([x]_\sim\) yields \([x]_\sim = \{b \in S \mid b \sim x\}\). Similarly, \([y]_\sim = \{b \in S \mid b \sim y\}\).

(a) Assume that \(x \sim y\). Thus, \(y \sim x\) (since the relation ∼ is symmetric).

Let \(c \in [x]_\sim\). Thus, \(c \in [x]_\sim = \{b \in S \mid b \sim x\}\). Thus, \(c \sim x\). From \(c \sim x\) and \(x \sim y\), we obtain \(c \sim y\) (since the relation ∼ is transitive). Hence, \(c \in \{b \in S \mid b \sim y\}\). In other words, \(c \in [y]_\sim\) (since \([y]_\sim = \{b \in S \mid b \sim y\}\)).

Forget that we fixed \(c\). We thus have proven that \(c \in [y]_\sim\) for each \(c \in [x]_\sim\). Thus, \([x]_\sim \subseteq [y]_\sim\). The same argument (with \(x\) and \(y\) switched) yields \([y]_\sim \subseteq [x]_\sim\) (since \(y \sim x\)). Combining \([x]_\sim \subseteq [y]_\sim\) with \([y]_\sim \subseteq [x]_\sim\), we obtain \([x]_\sim = [y]_\sim\). This proves Theorem 3.3.5 (a).

(b) Assume that we don’t have \(x \sim y\). Let \(c \in [x]_\sim \cap [y]_\sim\). We aim for a contradiction.

We have \(c \in [x]_\sim \cap [y]_\sim \subseteq [x]_\sim = \{b \in S \mid b \sim x\}\), so that \(c \sim x\). Likewise, \(c \sim y\). From \(c \sim x\), we obtain \(x \sim c\) (since the relation ∼ is symmetric). Combining this with \(c \sim y\), we obtain \(x \sim y\) (since ∼ is transitive). This contradicts our assumption that we don’t have \(x \sim y\).

Now, forget that we fixed \(c\). So we have found a contradiction for each \(c \in [x]_\sim \cap [y]_\sim\). Thus, there is no such \(c\). In other words, \([x]_\sim \cap [y]_\sim = \emptyset\). In other words, the sets \([x]_\sim\) and \([y]_\sim\) are disjoint. This proves Theorem 3.3.5 (b).

(c) Recall that \([y]_\sim = \{b \in S \mid b \sim y\}\). Thus, we have \(x \in [y]_\sim\) if and only if \(x \sim y\). In other words, we have \(x \sim y\) if and only if \(x \in [y]_\sim\). This proves Theorem 3.3.5 (c).

(d) Theorem 3.3.5 (e) (applied to \(y\) and \(x\) instead of \(x\) and \(y\)) shows that we have \(y \sim x\) if and only if \(y \in [x]_\sim\). In other words, we have the logical equivalence \((y \sim x) \iff (y \in [x]_\sim)\).

Proposition 3.2.10 (applied to \(a = x\) and \(b = y\)) shows that we have \(x \sim y\) if and only if \(y \sim x\). Thus, we have the following chain of logical equivalences:

\[
(x \sim y) \iff (y \sim x) \iff (y \in [x]_\sim).
\]

In other words, we have \(x \sim y\) if and only if \(y \in [x]_\sim\). This proves Theorem 3.3.5 (d).

(e) \(\implies\): Assume that \(x \sim y\). Then, Theorem 3.3.5 (a) yields \([x]_\sim = [y]_\sim\). Thus, the “\(\implies\)” direction of Theorem 3.3.5 (e) is proven.

\(\iff\): Assume that \([x]_\sim = [y]_\sim\). Then, Proposition 3.3.4 (applied to \(a = x\)) yields \(x \in [x]_\sim = [y]_\sim = \{b \in S \mid b \sim y\}\). In other words, \(x \sim y\). This proves the “\(\iff\)” direction of Theorem 3.3.5 (e).

Theorem 3.3.5 yields an important property of equivalence classes:

Exercise 3.3.1. Let ∼ be an equivalence relation on a set S. Prove that any two equivalence classes of ∼ are either identical or disjoint.

In the following, we will try to use Greek letters for equivalence classes and Roman letters for their representatives (as we did in the solution to Exercise 3.3.1 above).
3.3.3. More examples

Example 3.3.6. Consider the relation $\sim$ on $S = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ defined in Example 3.2.9. Its equivalence classes are the rational numbers. Indeed, the equivalence class $[(a_1, a_2)]_{\sim}$ of a pair $(a_1, a_2) \in S$ is commonly denoted by $\frac{a_1}{a_2}$ (or by $a_1 / a_2$).

This is how rational numbers are defined!

Equivalence classes appear in real life too, at least in the modern world. When you say that the sun rises approximately at 7 AM in February\footnote{in Minneapolis} what do “7 AM” and “February” mean? Clearly, “February” is not a specific month in history, since each year has its own February. Rather, it stands for an equivalence class of months, with respect to the relation of “being an integer number of years apart”. Similarly, “7 AM” means an equivalence class of moments with respect to the relation of “being an integer number of days apart”. Likewise, “the horse” in “the horse has a lifespan of 25 years” refers not to a specific horse, but to the whole species, which is an equivalence class of creatures with respect to a certain relation\footnote{According to Darwin, the relation is “being able to procreate” – although this is not per se an equivalence relation, so some tweaks need to be made ("reflexive-and-transitive closure") to turn it into one.}. Finally, the equivalence classes of the relation $\sim$ in Example 3.2.8 are commonly referred to as “continents”\footnote{at least if one considers Eurasia to be a single continent} or “islands”. Equivalence classes provide a way to refer to multiple objects (usually similar in some way) as if they were one.

3.3.4. The “is a permutation of” relation on tuples

Let us give a few more mathematical examples for equivalences and equivalence classes:

Definition 3.3.7. Let $A$ be a set, and let $k \in \mathbb{N}$. As we know, $A^k$ denotes the set of all $k$-tuples of elements of $A$.

The relation $\sim_{\text{perm}}$ on $A^k$ is defined as follows:

$$\left( p \sim_{\text{perm}} q \right) \iff (p \text{ is a permutation of } q).$$

(We are using Definition 2.13.16 here.) For example, $(3, 8, 8, 2) \sim_{\text{perm}} (8, 3, 2, 8)$.
Exercise 3.3.2. Prove that the relation $\sim \text{perm}$ is an equivalence relation.

Definition 3.3.8. Let $A$ be a set, and let $k \in \mathbb{N}$. The relation $\sim \text{perm}$ on $A^k$ is an equivalence relation (by Exercise 3.3.2). Its equivalence classes are called the unordered $k$-tuples of elements of $A$. For example, for $k = 2$ and $A = \mathbb{Z}$, the two 2-tuples $(6, 8)$ and $(8, 6)$ are permutations of each other, so $(6, 8) \sim \text{perm} (8, 6)$ and thus $[(6, 8)] \sim \text{perm} = [(8, 6)] \sim \text{perm}$.

3.3.5. The “is a cyclic rotation of” relation on tuples

Another example of an equivalence relation is the following:

Definition 3.3.9. Again, let $A$ be a set and $k \in \mathbb{N}$. If $a = (a_1, a_2, \ldots, a_k) \in A^k$, then a cyclic rotation of $a$ means a $k$-tuple of the form

$$(a_{i+1}, a_{i+2}, \ldots, a_k, a_1, a_2, \ldots, a_i) \in A^k$$

for some $i \in \{0, 1, \ldots, k\}$.

For example, the cyclic rotations of the 3-tuple $(1, 4, 5)$ are $(1, 4, 5)$, $(4, 5, 1)$ and $(5, 1, 4)$.

(Here is an equivalent description of cyclic rotations: Let $C$ be the map $A^k \rightarrow A^k$ that sends each $k$-tuple $(a_1, a_2, \ldots, a_k)$ to $(a_2, a_3, \ldots, a_k, a_1)$. Then, it is easy to see that a cyclic rotation of $a$ is the same as a $k$-tuple of the form $C^i (a)$ for some $i \in \{0, 1, \ldots, k\}$. But it is also easy to see that $C^k = \text{id}$. Thus, the $C^i (a)$ for $i \in \{0, 1, \ldots, k\}$ are exactly the $C^i (a)$ for $i \in \mathbb{N}$.)

The relation $\sim \text{cyc}$ on $A^k$ is defined as follows:

$$(p \sim \text{cyc} q) \iff (p \text{ is a cyclic rotation of } q) \iff \left(p = C^i (q) \text{ for some } i \in \mathbb{N}\right).$$

This relation $\sim \text{cyc}$ is an equivalence relation. Its equivalence classes are called necklaces of length $k$ over $A$.

We shall not prove the statements claimed in this definition, since they are particular cases of more general results that will be proven below (about groups acting on sets).
For example, the necklaces of length 3 over the set $A = \{1, 2\}$ are

\[
[(1, 1, 1)]_{\sim_{\text{cyc}}} = \{(1, 1, 1)\},
\]
\[
[(1, 1, 2)]_{\sim_{\text{cyc}}} = \{(1, 1, 2), (1, 2, 1), (2, 1, 1)\},
\]
\[
[(1, 2, 2)]_{\sim_{\text{cyc}}} = \{(1, 2, 2), (2, 2, 1), (2, 1, 2)\},
\]
\[
[(2, 2, 2)]_{\sim_{\text{cyc}}} = \{(2, 2, 2)\}.
\]

This may suggest that a necklace $[(a_1, a_2, \ldots, a_k)]_{\sim_{\text{cyc}}}$ is uniquely determined by how often each element appears in the tuple $(a_1, a_2, \ldots, a_k)$. But this is not true in general; for example, if $A = \{1, 2, 3\}$, then

\[
[(1, 2, 3)]_{\sim_{\text{cyc}}} = \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\}
\]
\[
[(1, 3, 2)]_{\sim_{\text{cyc}}} = \{(1, 3, 2), (3, 2, 1), (2, 1, 3)\}
\]

are two different necklaces of length 3 over the set $A = \{1, 2, 3\}$.

How many necklaces of length $k$ over a $q$-element set $A$ exist? It turns out that there is a nice formula for this, involving Euler’s totient function $\phi$:

**Theorem 3.3.10.** Let $k$ be a positive integer. Let $A$ be a $q$-element set (where $q \in \mathbb{N}$). Then, the number of necklaces of length $k$ over the set $A$ is

\[
\frac{1}{k} \sum_{d \mid k} \phi(d) q^{k/d}.
\]

Note that it is not (a priori) clear that $\frac{1}{k} \sum_{d \mid k} \phi(d) q^{k/d}$ is an integer! Actually, this holds even when $q$ is a negative integer, even though there exist no $q$-element sets in that case. Thus, $\frac{1}{k} \sum_{d \mid k} \phi(d) q^{k/d}$ is another integer-valued polynomial for each positive integer $k$.

We will prove Theorem 3.3.10 using the concept of group actions further below.

### 3.3.6. Definition of the quotient set and the projection map

**Definition 3.3.11.** Let $S$ be a set, and let $\sim$ be an equivalence relation on $S$.

(a) The set of equivalence classes of $\sim$ is denoted by $S/\sim$. It is called the **quotient** (or **quotient set**) of $S$ by $\sim$.

(b) The map

\[
S \rightarrow S/\sim, \quad s \mapsto [s]_{\sim}
\]
(which sends each element \( s \in S \) to its equivalence class) is called the canonical projection (onto the quotient), and we will denote it by \( \pi_\sim \).

(c) An element of an equivalence class of \( \sim \) is also called a representative of this class.

**Exercise 3.3.3.** Let \( S \) be a set.

Recall that if \( T \) is a further set, and if \( f : S \to T \) is a map, then an equivalence relation \( \equiv_f \) is defined on the set \( S \). (See Example 3.2.7 for its definition.)

Now, let \( \sim \) be any equivalence relation on \( S \). Prove that \( \sim \) has the form \( \equiv_f \) for a properly chosen set \( T \) and a properly chosen \( f : S \to T \).

More precisely, prove that \( \sim \) equals \( \equiv_f \), where \( T \) is the quotient set \( S/\sim \) and \( f : S \to T \) is the canonical projection \( \pi_\sim : S \to S/\sim \).

**Hint:** To prove that two relations \( R_1 \) and \( R_2 \) on \( S \) are equal, you need to check that every pair \((a, b)\) of elements of \( S \) satisfies the equivalence \((aR_1b) \iff (aR_2b)\).

### 3.4. \( \mathbb{Z}/n \) ("integers modulo \( n \")

We now come to one of the most important example of equivalence classes: the residue classes of integers modulo a given positive integer \( n \).

**Convention 3.4.1.** For the whole Section 3.4, we fix an integer \( n \).

#### 3.4.1. Definition of \( \mathbb{Z}/n \)

**Definition 3.4.2.** (a) Define a relation \( \equiv_n \) on the set \( \mathbb{Z} \) by

\[
(a \equiv_n b) \iff (a \equiv b \text{ mod } n).
\]

(This is precisely the relation \( \equiv \) from Example 3.1.2 (e).)

Recall that \( \equiv_n \) is an equivalence relation (by Example 3.2.5).

(b) A **residue class modulo \( n \)** means an equivalence class of the relation \( \equiv_n \). For example,

\[
\begin{align*}
[0]_5 &= \{ \ldots, -15, -10, -5, 0, 5, 10, 15, 20, \ldots \}, \\
[1]_5 &= \{ \ldots, -14, -9, -4, 1, 6, 11, 16, 21, \ldots \}, \\
[2]_5 &= \{ \ldots, -13, -8, -3, 2, 7, 12, 17, 22, \ldots \}, \\
[3]_5 &= \{ \ldots, -12, -7, -2, 3, 8, 13, 18, 23, \ldots \}, \\
[4]_5 &= \{ \ldots, -11, -6, -1, 4, 9, 14, 19, 24, \ldots \}.
\end{align*}
\]
are all the residue classes modulo 5. As you see, these classes are in 1-to-1 correspondence with the 5 possible remainders 0, 1, 2, 3, 4 modulo 5. This generalizes (see Theorem 3.4.4 below). First, let us introduce a few notations:

**Definition 3.4.3.** (a) If $i$ is an integer, then we denote the residue class $[i]_n$ by $[i]_n$. (Some authors denote this residue class by $i_n$ or $i \mod n$. Be careful with the notation $i \mod n$, since other authors use it for the integer $i \% n$ when $n$ is positive.)

(b) The set $\mathbb{Z}/\equiv_n$ of residue classes modulo $n$ is called $\mathbb{Z}/n$. (Some authors call it $\mathbb{Z}/(n)$ or $\mathbb{Z}/n\mathbb{Z}$ or $\mathbb{Z}_n$. Be careful with the notation $\mathbb{Z}_n$, since it has a different meaning, too.)

### 3.4.2. What $\mathbb{Z}/n$ looks like

Let us now state and rigorously prove what we have just observed on the example of $n = 5$:

**Theorem 3.4.4.** Assume that the integer $n$ is positive.

The set $\mathbb{Z}/n$ has exactly $n$ elements, namely $[0]_n, [1]_n, \ldots, [n-1]_n$. (In particular, these elements $[0]_n, [1]_n, \ldots, [n-1]_n$ are distinct.)

Before we prove this, let us make a simple observation:

**Proposition 3.4.5.** (a) Each element of $\mathbb{Z}/n$ can be written in the form $[s]_n$ for some integer $s$.

(b) Let $a$ and $b$ be integers. Then, we have $[a]_n = [b]_n$ if and only if $a \equiv b \mod n$.

**Proof of Proposition 3.4.5.** (a) If $\sigma \in \mathbb{Z}/n$, then $\sigma$ is a residue class modulo $n$ (by the definition of $\mathbb{Z}/n$), and thus is an equivalence class of the relation $\equiv_n$ (by the definition of a residue class). Hence, this $\sigma$ can be written in the form $[s]_n$ for some integer $s$. In other words, this $\sigma$ can be written in the form $[s]_n$ for some integer $s$ (since we have defined $[s]_n$ to be a shorthand for $[s]_n$). In other words, each element of $\mathbb{Z}/n$ can be written in the form $[s]_n$ for some integer $s$. This proves Proposition 3.4.5 (a).

(b) Theorem 3.3.5 (e) (applied to $\mathbb{Z}/n$, $\equiv_n$, $a$ and $b$ instead of $S$, $\sim$, $x$ and $y$) shows that we have $a \equiv b \mod n$ if and only if $[a]_n = [b]_n$. Thus, we have the logical equivalence

$$
(a \equiv b) \iff ([a]_n = [b]_n).
$$

(104)

Definition 3.4.3 (a) shows that $[a]_n = [a]_n$ and $[b]_n = [b]_n$. Hence, we have the
following chain of logical equivalences:

\[(a_n = b_n) \iff (a \equiv_n n = b \equiv_n n) \iff (a \equiv b) \quad \text{(by (104))}\]

(by the definition of the relation \(\equiv_n\)). In other words, we have \([a]_n = [b]_n\) if and only if \(a \equiv b \mod n\). This proves Proposition 3.4.5(b). 

Proof of Theorem 3.4.4. We have a map

\[\pi_n : \mathbb{Z} \to \mathbb{Z}/n,\]

\[s \mapsto [s]_n.\]

(This is simply the map \(\pi\) defined in Definition 3.3.11(b), applied to the case when \(S = \mathbb{Z}\) and when \(\sim \) is the equivalence relation \(\equiv_n\).)

We restrict this map \(\equiv_n\) to the set \(\{0, 1, \ldots, n - 1\}\); we thus obtain a map

\[P : \{0, 1, \ldots, n - 1\} \to \mathbb{Z}/n,\]

\[s \mapsto [s]_n.\]

Our goal is to prove that this map \(P\) is bijective.

In general, there are two ways in which one usually proves that a map is bijective: One way is to prove that it is surjective and injective; the other way is by constructing an inverse to this map. Both ways can be used here; let us follow the second way, since it demonstrates an important point about equivalence classes.

So we want to construct an inverse to the map \(P\). To do so, we try to define a map

\[R : \mathbb{Z}/n \to \{0, 1, \ldots, n - 1\},\]

\[[s]_n \mapsto s \% n\]

(that is, a map \(R : \mathbb{Z}/n \to \{0, 1, \ldots, n - 1\}\) that sends each residue class \([s]_n\) to the remainder \(s \% n\)). Can we do this? Would this map \(R\) be actually well-defined?

First of all, our definition of \(R\) does indeed specify a value of \(R(\sigma)\) for each \(\sigma \in \mathbb{Z}/n\). This is because each element of \(\mathbb{Z}/n\) can be written in the form \([s]_n\) for some integer \(s\) (because of Proposition 3.4.5(a)), and therefore our definition tells us where this element should go under \(R\).

Furthermore, if \(s\) is an integer, then \(s \% n \in \{0, 1, \ldots, n - 1\}\) (by Corollary 2.6.9(a), applied to \(u = s\)). Hence, our definition of \(R\) does not require the map \(R\) to take values lying outside of its target.

\[\text{This is one way in which maps can fail to be well-defined. For example, the map}\]

\[\mathbb{N} \to \mathbb{N}, \quad i \mapsto i - 1\]

\[\text{is not well-defined for this reason (because } i - 1 \notin \mathbb{N} \text{ for } i = 0).\]
However, there is one more thing that could go wrong with our definition of $R$: One element $\sigma$ of $\mathbb{Z}/n$ can be written as $[s]_n$ for several different integers $s$. For instance, $[2]_5 = [7]_5 = [12]_5 = [17]_5 = \cdots$. If the remainders $s \% n$ of these integers $s$ were different, then the map $R$ would have to send the class $\sigma$ to several different numbers, and this is not something a map can do. To see an example where this does go wrong, let us try to define a map

$$R_{\text{wrong}} : \mathbb{Z}/n \rightarrow \{0, 1, \ldots, n-1\},$$

$$[s]_n \mapsto s \% (n + 1).$$

So this definition of $R_{\text{wrong}}$ is identical to our definition of $R$ above, except that we are sending $[s]_n$ to $s \% (n + 1)$ rather than to $s \% n$. However, $R_{\text{wrong}}$ does not actually exist. In fact, if this ostensible map $R_{\text{wrong}}$ would exist, then it would have to send $[0]_n$ to $0 \% (n + 1) = 0$ and send $[-n]_n$ to $(-n) \% (n + 1) = 1$ however, $[0]_n$ and $[-n]_n$ are the same residue class (since $0 \equiv -n \mod n$), whereas 0 and 1 are not the same number, and thus this map $R_{\text{wrong}}$ would send the same class to two different numbers. Thus, the map $R_{\text{wrong}}$ does not exist.

We shall now check that our above definition of $R$ does not suffer from this problem. In other words, we shall check that in the definition of

$$R : \mathbb{Z}/n \rightarrow \{0, 1, \ldots, n-1\},$$

$$[s]_n \mapsto s \% n,$$

any two possible integers $s$ leading to the same class $[s]_n$ also lead to the same remainder $s \% n$. In other words, we shall prove the following claim:

**Claim 1:** If $s_1$ and $s_2$ are two integers such that $[s_1]_n = [s_2]_n$, then $s_1 \% n = s_2 \% n$.

**Proof of Claim 1:** Let $s_1$ and $s_2$ be two integers such that $[s_1]_n = [s_2]_n$.

Proposition 3.4.5 (b) (applied to $s_1$ and $s_2$ instead of $a$ and $b$) shows that we have $[s_1]_n = [s_2]_n$ if and only if $s_1 \equiv s_2 \mod n$. Thus, we have $s_1 \equiv s_2 \mod n$ (since $[s_1]_n = [s_2]_n$). But Exercise 2.6.1 (applied to $u = s_1$ and $v = s_2$) shows that $s_1 \equiv s_2 \mod n$ if and only if $s_1 \% n = s_2 \% n$. Hence, we have $s_1 \% n = s_2 \% n$. This proves Claim 1.

Claim 1 shows that if $s$ is an integer, then $s \% n$ depends only on the residue class $[s]_n$, but not on the actual integer $s$. Thus, if we have a residue class $\sigma \in \mathbb{Z}/n$, then we can write $\sigma$ in the form $\sigma = [s]_n$ for some integer $s$ (since every residue class in $\mathbb{Z}/n$ can be written in this form), and then the integer $s \% n$ will depend only on the class $\sigma$ and not on the specific choice of this integer $s$. Hence, the map $R$ is well-defined.

Now we have two maps

$$P : \{0, 1, \ldots, n-1\} \rightarrow \mathbb{Z}/n,$$

$$s \mapsto [s]_n.$$

The equality $(-n) \% (n + 1) = 1$ follows from writing $-n$ in the form $-n = (-1) \cdot (n + 1) + 1$. 

---

\footnote{The equality $(-n) \% (n + 1) = 1$ follows from writing $-n$ in the form $-n = (-1) \cdot (n + 1) + 1$.}
and
\[ R : \mathbb{Z}/n \to \{0,1,\ldots,n-1\}, \quad [s]_n \mapsto s \% n. \]

We claim that they are mutually inverse. Indeed:

- We have \( P \circ R = \text{id} \).

  \[ \text{Proof: Let } \sigma \in \mathbb{Z}/n. \text{ We shall prove that } (P \circ R)(\sigma) = \text{id}(\sigma). \]

  Proposition 3.4.5 (a) says that each element of \( \mathbb{Z}/n \) can be written in the form \([s]_n\) for some integer \( s \). Hence, \( \sigma \) can be written in this form. In other words, \( \sigma = [s]_n \) for some integer \( s \). Consider this \( s \). The definition of \( R \) yields \( R([s]_n) = s \% n \). Corollary 2.6.9 (a) (applied to \( u = s \)) yields \( s \% n \equiv s \mod n \).

  Now, from \( \sigma = [s]_n \), we obtain
  \[
  (P \circ R)(\sigma) = (P \circ R)([s]_n) = P \left( R \left( [s]_n \right) \right) = P \left( s \% n \right)
  = [s \% n]_n \quad (\text{by the definition of } P)
  = [s]_n \quad (\text{since } s \% n \equiv s \mod n)
  = \sigma = \text{id}(\sigma).
  \]

  Now, forget that we fixed \( \sigma \). We thus have proven that \( (P \circ R)(\sigma) = \text{id}(\sigma) \) for each \( \sigma \in \mathbb{Z}/n \). In other words, \( P \circ R = \text{id} \].

- We have \( R \circ P = \text{id} \).

  \[ \text{Proof: Let } s \in \{0,1,\ldots,n-1\}. \text{ Thus, Corollary 2.6.9 (c) (applied to } u = s \text{ and } c = s) \text{ yields } s = s \% n \text{ (since } s \equiv s \mod n). \text{ But the definition of } P \text{ yields } P(s) = [s]_n. \text{ Hence,}
  \]
  \[
  (R \circ P)(s) = R \left( P(s) \right) = R([s]_n) = s \% n \quad (\text{by the definition of } R)
  = s = \text{id}(s).
  \]

  Now, forget that we fixed \( s \). We thus have proven that \( (R \circ P)(s) = \text{id}(s) \) for each \( s \in \{0,1,\ldots,n-1\} \). In other words, \( R \circ P = \text{id} \].

Combining \( P \circ R = \text{id} \) and \( R \circ P = \text{id} \), we conclude that the maps \( P \) and \( R \) are mutually inverse. Thus, the map \( P \) is invertible, i.e., bijective. Hence, \( P \) is surjective and injective. Since \( P \) is injective, we see that \( P \) must send the distinct elements \( 0,1,\ldots,n-1 \) of its domain to distinct elements. In other words, the \( n \) elements \( P(0), P(1), \ldots, P(n-1) \) of \( \mathbb{Z}/n \) must be distinct.
But recall that \( P(s) = \left[ s \right]_n \) for each \( s \in \{0, 1, \ldots, n - 1\} \) (by the definition of \( P \)). Thus, the \( n \) elements \( P(0), P(1), \ldots, P(n-1) \) can be rewritten as \( [0]_n, [1]_n, \ldots, [n-1]_n \). Hence, the \( n \) elements \( [0]_n, [1]_n, \ldots, [n-1]_n \) are distinct (since the \( n \) elements \( P(0), P(1), \ldots, P(n-1) \) are distinct).

Moreover, \( P \) is surjective. Thus,

\[
\mathbb{Z}/n = P(\{0, 1, \ldots, n-1\}) = \{P(0), P(1), \ldots, P(n-1)\} = \{[0]_n, [1]_n, \ldots, [n-1]_n\}
\]

(since \( P(s) = \left[ s \right]_n \) for each \( s \in \{0, 1, \ldots, n - 1\} \)). In other words, the elements of \( \mathbb{Z}/n \) are exactly the \( n \) elements \( [0]_n, [1]_n, \ldots, [n-1]_n \). These \( n \) elements are distinct (as we have previously shown). Hence, the set \( \mathbb{Z}/n \) has exactly \( n \) elements, namely \( [0]_n, [1]_n, \ldots, [n-1]_n \). This proves Theorem 3.4.4.

Let us summarize some of the facts we have shown in the above proof as a separate proposition:

**Proposition 3.4.6.** Let \( n \) be a positive integer.

(a) The two maps

\[
P : \{0, 1, \ldots, n-1\} \to \mathbb{Z}/n,
\]

\[
s \mapsto \left[ s \right]_n
\]

and

\[
R : \mathbb{Z}/n \to \{0, 1, \ldots, n-1\},
\]

\[
\left[ s \right]_n \mapsto s \% n
\]

are well-defined and mutually inverse, and thus are bijections.

(b) Let \( \alpha \in \mathbb{Z}/n \). Then, there exists a unique \( a \in \{0, 1, \ldots, n-1\} \) satisfying \( \alpha = \left[ a \right]_n \).

**Proof of Proposition 3.4.6 (a)** During the proof of Theorem 3.4.4 above, we have shown that the maps \( P \) and \( R \) are well-defined and mutually inverse. Hence, these maps \( P \) and \( R \) are invertible, i.e., are bijective. In other words, the maps \( P \) and \( R \) are bijections. Thus, Proposition 3.4.6 (a) is proven.

(b) Consider the maps \( P \) and \( R \) from Proposition 3.4.6 (a). Then, Proposition 3.4.6 (a) shows that these two maps \( P \) and \( R \) are well-defined and mutually inverse, and thus are bijections.

We have \( P \circ R = \text{id} \) (since \( P \) and \( R \) are mutually inverse). Hence, \( (P \circ R)(\alpha) = \text{id}(\alpha) = \alpha \). Hence, \( \alpha = (P \circ R)(\alpha) = P(R(\alpha)) = \left[ R(\alpha) \right]_n \) (by the definition of \( P \)). Thus, there exists at least one \( a \in \{0, 1, \ldots, n-1\} \) satisfying \( \alpha = \left[ a \right]_n \) (namely, \( a = R(\alpha) \)). (Indeed, \( R(\alpha) \in \{0, 1, \ldots, n-1\} \) follows from the fact that \( R \) is a map from \( \mathbb{Z}/n \) to \( \{0, 1, \ldots, n-1\} \).)
On the other hand, let \( a \in \{0,1,\ldots,n-1\} \) be such that \( \alpha = [a]_n \). We shall prove that \( a = R(\alpha) \). Indeed, the definition of \( P \) yields \( P(a) = [a]_n = a \); hence, \( a = P^{-1}(a) \) (since the map \( P \) is invertible). But \( P^{-1} = R \) (since \( P \) and \( R \) are mutually inverse). Thus, \( a = R^{-1}(a) = R(\alpha) \).

Now, forget that we fixed \( a \). We thus have shown that every \( a \in \{0,1,\ldots,n-1\} \) satisfying \( \alpha = [a]_n \) must satisfy \( a = R(\alpha) \). In other words, every \( a \in \{0,1,\ldots,n-1\} \) satisfying \( a = [a]_n \) must be equal to \( R(\alpha) \). Hence, there exists at most one such \( a \).

Now, we conclude that there exists a unique \( \alpha \in \{0,1,\ldots,n-1\} \) satisfying \( \alpha = [a]_n \); because we have shown that there exists at least one such \( a \), and we have shown that there exists at most one such \( a \). This proves Proposition 3.4.6(b).

Proposition 3.4.6(b) can be restated as follows: Each residue class \( \alpha \in \mathbb{Z}/n \) has a unique representative in the set \( \{0,1,\ldots,n-1\} \).

### 3.4.3. Making choices that don’t matter: The universal property of quotient sets

In the above proof of Theorem 3.4.4, we have witnessed an important issue in dealing with quotient sets: If you want to define a map \( f \) going out of a quotient set \( S/\sim \) then the easiest way to do so is often to specify \( f([s]_\sim) \) for each \( s \in S \); but in order to ensure that this definition is well-defined (i.e., that our map \( f \) actually exists), we need to verify that the value of \( f([s]_\sim) \) we are specifying depends only on the equivalence class \([s]_\sim\) but not on the representative \( s \). In other words, we need to verify that if \( s_1 \) and \( s_2 \) are two elements of \( S \) such that \([s_1]_\sim = [s_2]_\sim\), then our definition of \( f \) assigns the same value to \( f([s_1]_\sim) \) as it does to \( f([s_2]_\sim) \). This verification (which we did in our above proof by proving Claim 1) is often quite easy, but it is necessary.

Let us restate this strategy for defining maps out of a quotient set more rigorously:

**Remark 3.4.7.** Let \( S \) and \( T \) be two sets, and let \( \sim \) be an equivalence relation on \( S \). Assume that we want to define a map

\[
f : S/\sim \to T, \quad [s]_\sim \mapsto F(s),
\]

where \( F(s) \) is some element of \( T \) for each \( s \in S \). (That is, we want to define a map \( f : S \to T \) such that every \( s \in S \) satisfies \( f([s]_\sim) = F(s) \).

In order to ensure that this \( f \) is well-defined, we need to verify that if \( s_1 \) and \( s_2 \) are two elements of \( S \) such that \([s_1]_\sim = [s_2]_\sim\), then \( F(s_1) = F(s_2) \). If this verification has been done, the map \( f \) is well-defined.

---

96In our case, the quotient set was \( \mathbb{Z}/n \equiv \mathbb{Z}/n \) (also known as \( \mathbb{Z}/n \)), and the map we wanted to define was \( R \).
Further examples of maps out of quotient sets defined in this way can be found in [ConradW].

Let us illustrate this method of defining maps on a few more examples:

**Example 3.4.8.** Let $A$ be a set, and let $k \in \mathbb{N}$. Fix some $c \in A$. We can then define a map

$$\text{mult}_c : A^k \to \mathbb{N},$$

$$(a_1, a_2, \ldots, a_k) \mapsto \text{(the number of } i \in \{1, 2, \ldots, k\} \text{ such that } a_i = c).$$

This map $\text{mult}_c$ simply sends each $k$-tuple to the number of times that $c$ appears in this $k$-tuple. For example, $\text{mult}_5((1, 5, 2, 4, 7, 5, 5, 6)) = 3$, since 5 appears exactly 3 times in the 8-tuple $(1, 5, 2, 4, 7, 5, 5, 6)$ (assuming that $k = 8$ and $A = \mathbb{Z}$). It is clear that this map $\text{mult}_c$ is well-defined. (The number $\text{mult}_c(a)$ for a $k$-tuple $a$ is called the multiplicity of $c$ in $a$. Therefore the notation “$\text{mult}_c$”.)

Now, it stands to reason that the same can be done with unordered $k$-tuples: After all, the number of times that $c$ appears in a $k$-tuple should not depend on the order of the entries of the tuple. To formalize this, however, we need to deal with quotient sets. Indeed, recall that the “unordered $k$-tuples of elements of $A$” were defined (in Definition 3.3.8) as equivalence classes of the relation $\sim_{\text{perm}}$ on the set $A^k$. So $A^k/\sim_{\text{perm}}$ is the set of all unordered $k$-tuples of elements of $A$. The map that counts how often $c$ appears in an unordered $k$-tuple should thus have the form

$$\text{mult}'_c : A^k/\sim_{\text{perm}} \to \mathbb{N},$$

$$[(a_1, a_2, \ldots, a_k)]_{\sim_{\text{perm}}} \mapsto \text{(the number of } i \in \{1, 2, \ldots, k\} \text{ such that } a_i = c).$$

Or, to put it more compactly (making use of the map $\text{mult}_c$ for ordered $k$-tuples defined above), it should have the form

$$\text{mult}'_c : A^k/\sim_{\text{perm}} \to \mathbb{N},$$

$$[a]_{\sim_{\text{perm}}} \mapsto \text{mult}_c a.$$  

The question is: Why is this map $\text{mult}'_c$ well-defined?

Remark 3.4.7 (applied to $A^k$, $\mathbb{N}$ and $\sim_{\text{perm}}$ instead of $S$, $T$ and $\sim$) shows that in order to ensure that this map $\text{mult}'_c$ is well-defined, we need to verify that if $a_1$ and $a_2$ are two elements of $A^k$ (that is, two ordered $k$-tuples) such that $[a_1]_{\sim_{\text{perm}}} = [a_2]_{\sim_{\text{perm}}}$, then $\text{mult}_c(a_1) = \text{mult}_c(a_2)$. Let us do this: Let $a_1$ and $a_2$ be two elements of $A^k$ (that is, two ordered $k$-tuples) such that $[a_1]_{\sim_{\text{perm}}} = [a_2]_{\sim_{\text{perm}}}$. Now, $[a_1]_{\sim_{\text{perm}}} = [a_2]_{\sim_{\text{perm}}}$ entails $a_1 \sim_{\text{perm}} a_2$ (indeed, Theorem 3.3.5(e) shows that we have $a_1 \sim_{\text{perm}} a_2$ if and only if $[a_1]_{\sim_{\text{perm}}} = [a_2]_{\sim_{\text{perm}}}$. In other words, $a_1$ is a permutation of $a_2$ (by the definition of $\sim_{\text{perm}}$). In other words, the

---

97 When reading [ConradW, Example 1.1], keep in mind that rational numbers are defined as equivalence classes of elements of $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$, as we have seen in Example 3.3.6. Thus, $Q$ is actually a quotient set: namely, $Q = S/\sim$ using the notations of Example 3.3.6.
tuples $a_1$ and $a_2$ differ only in the order of their entries. Hence, Lemma 2.13.21 (applied to $A$, $a_1$, $a_2$ and $c$ instead of $P$, $(a_1,a_2,\ldots,a_k)$, $(b_1,b_2,\ldots,b_l)$ and $p$) yields that

$$(\text{the number of times } c \text{ appears in } a_1) = (\text{the number of times } c \text{ appears in } a_2).$$

This rewrites as $\text{mult}_c(a_1) = \text{mult}_c(a_2)$ (since $\text{mult}_c(a_1) = \text{mult}_c(a_2)$). This is what we needed to prove. Thus, we have shown that $\text{mult}_c'$ is well-defined.

On the other hand, if we tried to define a map $\text{first} : A^k/\sim_{\text{perm}} \to \mathbb{N}$,

$$[a]_{\sim_{\text{perm}}} \mapsto (\text{the first entry of } a),$$

(assuming that $k > 0$, so that an ordered $k$-tuple does indeed have a first entry), then we would run into troubles, because it is not true that if $a_1$ and $a_2$ are two elements of $A^k$ such that $[a_1]_{\sim_{\text{perm}}} = [a_2]_{\sim_{\text{perm}}}$, then $(\text{the first entry of } a_1) = (\text{the first entry of } a_2)$. And this is no surprise: There is no such thing as “the first entry” of an unordered $k$-tuple. The first entry of a $k$-tuple is sensitive to reordering of its entries.

We can restate this method of defining maps as a rigorous theorem:

**Theorem 3.4.9.** Let $S$ and $T$ be two sets, and let $\sim$ be an equivalence relation on $S$. For each $s \in S$, let $F(s)$ be an element of $T$. (In other words, let $F$ be a map from $S$ to $T$.) Assume that the following assumption holds:

**Assumption 1:** If $s_1$ and $s_2$ are two elements of $S$ satisfying $s_1 \sim s_2$, then $F(s_1) = F(s_2)$.

Then, there exists a unique map $f : S/\sim \to T$ such that every $s \in S$ satisfies $f([s]_{\sim}) = F(s)$.

Theorem 3.4.9 says that (under the assumption that Assumption 1 holds) we can define a map

$$f : S/\sim \to T,$$

$$[s]_{\sim} \mapsto F(s).$$

For example, the map $R$ defined in our proof of Theorem 3.4.4 was defined in this way (with $\mathbb{Z}/\mathbb{Z}$, $\equiv$ and $s \% n$ playing the roles of $S$, $T$, $\sim$ and $F(s)$), and our proof of Claim 1 was essentially us verifying that Assumption 1 of Theorem 3.4.9 is satisfied.

For the sake of completeness, let us give a formal proof for Theorem 3.4.9 as well:

**Proof of Theorem 3.4.9.** We need to prove the following two statements:

**Statement 1:** There exists at least one map $f : S/\sim \to T$ such that every $s \in S$ satisfies $f([s]_{\sim}) = F(s)$. 

Statement 2: There exists at most one map \( f : S / \sim \rightarrow T \) such that every \( s \in S \) satisfies \( f ([s]_\sim) = F (s) \).

[Proof of Statement 1: We define a map \( \varphi \) as follows:

Let \( \sigma \in S / \sim \). Thus, \( \sigma \) is an equivalence class of \( \sim \) (by the definition of \( S / \sim \)). In other words, \( \sigma = [s]_\sim \) for some element \( s \in S \). In other words, there exists some element \( s \in S \) such that \( \sigma = [s]_\sim \). If \( s_1 \) and \( s_2 \) are two such elements \( s \), then \( F (s_1) = F (s_2) \). \(^{98}\) Thus, the element \( F (s) \in T \) obtained from such an element \( s \) does not depend on the choice of \( s \) (as long as \( \sigma \) is fixed). Hence, we can define \( \varphi (\sigma) \) by setting

\[
\varphi (\sigma) = F (s),
\]

where \( s \) is any element of \( S \) satisfying \( \sigma = [s]_\sim \).

Define \( \varphi (\sigma) \) this way. Thus, we have defined an element \( \varphi (\sigma) \) of \( T \) for each \( \sigma \in S / \sim \). Hence, we have defined a map \( \varphi : S / \sim \rightarrow T \). Moreover, this map has the property that every \( s \in S \) satisfies \( \varphi ([s]_\sim) = F (s) \). (Indeed, this follows from (105) (applied to \( \sigma = [s]_\sim \)), since obviously \( [s]_\sim = [s]_\sim \).)

Hence, there exists at least one map \( f : S / \sim \rightarrow T \) such that every \( s \in S \) satisfies \( f ([s]_\sim) = F (s) \) (namely, the map \( \varphi \)). This proves Statement 1.]

[Proof of Statement 2: Let \( f_1 \) and \( f_2 \) be two maps \( f : S / \sim \rightarrow T \) such that every \( s \in S \) satisfies \( f ([s]_\sim) = F (s) \). We shall show that \( f_1 = f_2 \).

We know that \( f_1 \) is a map \( f : S / \sim \rightarrow T \) such that every \( s \in S \) satisfies \( f ([s]_\sim) = F (s) \). In other words, \( f_1 \) is a map from \( S / \sim \) to \( T \) and has the property that

\[
every \ s \in S \ satisfies \ f_1 ([s]_\sim) = F (s). \quad (106)
\]

Likewise, \( f_2 \) is a map from \( S / \sim \) to \( T \) and has the property that

\[
every \ s \in S \ satisfies \ f_2 ([s]_\sim) = F (s). \quad (107)
\]

Now, let \( \sigma \in S / \sim \) be arbitrary. Thus, \( \sigma \) is an equivalence class of \( \sim \) (by the definition of \( S / \sim \)). In other words, \( \sigma = [s]_\sim \) for some element \( s \in S \). Consider this \( s \). Then, from \( \sigma = [s]_\sim \), we obtain \( f_1 (\sigma) = f_1 ([s]_\sim) = F (s) \) (by (106)). Similarly, \( f_2 (\sigma) = F (s) \). Comparing these two equalities, we find \( f_1 (\sigma) = f_2 (\sigma) \).

Forget that we fixed \( \sigma \). We thus have proven that \( f_1 (\sigma) = f_2 (\sigma) \) for each \( \sigma \in S / \sim \). In other words, \( f_1 = f_2 \).

Forget that we fixed \( f_1 \) and \( f_2 \). We thus have proven that if \( f_1 \) and \( f_2 \) are two maps \( f : S / \sim \rightarrow T \) such that every \( s \in S \) satisfies \( f ([s]_\sim) = F (s) \), then \( f_1 = f_2 \). In other words, there exists at most one such map \( f \). This proves Statement 2.]

Now, we conclude that there exists a unique map \( f : S / \sim \rightarrow T \) such that every \( s \in S \) satisfies \( f ([s]_\sim) = F (s) \) (because Statement 1 shows that there exists at least one such map, while Statement 2 shows that there exists at most one such map). This proves Theorem 3.4.9. \( \square \)

Theorem 3.4.9 is known as the universal property of the quotient set.

\(^{98}\) Proof. Let \( s_1 \) and \( s_2 \) be two such elements \( s \). Then, \( \sigma = [s_1]_\sim \) (since \( s_1 \) is an element \( s \in S \) such that \( \sigma = [s]_\sim \)) and \( \sigma = [s_2]_\sim \) (for similar reasons). Hence, \( [s_1]_\sim = [s_2]_\sim \). But Theorem 3.3.5 (e) (applied to \( x = s_1 \) and \( y = s_2 \)) yields that we have \( s_1 \sim s_2 \) if and only if \( [s_1]_\sim = [s_2]_\sim \). Hence, we have \( s_1 \sim s_2 \) (since \( [s_1]_\sim = [s_2]_\sim \)). Thus, Assumption 1 shows that \( F (s_1) = F (s_2) \), qed.
3.4.4. Projecting from $\mathbb{Z}/n$ to $\mathbb{Z}/d$

As another example of a map from a quotient set, let us define certain maps from $\mathbb{Z}/n$ to $\mathbb{Z}/d$ that exist whenever two integers $n$ and $d$ satisfy $d \mid n$:

**Proposition 3.4.10.** Let $n$ be an integer. Let $d$ be a divisor of $n$. Then, there is a map

$$\pi_{n,d} : \mathbb{Z}/n \to \mathbb{Z}/d,$$

$$[s]_n \mapsto [s]_d.$$

**Example 3.4.11. (a)** For example, for $n = 6$ and $d = 2$, Proposition 3.4.10 says that there is a map

$$\pi_{6,2} : \mathbb{Z}/6 \to \mathbb{Z}/2,$$

$$[s]_6 \mapsto [s]_2.$$

This map sends the residue classes

$$[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6$$

to $[0]_2, [1]_2, [2]_2, [3]_2, [4]_2, [5]_2$, respectively.

In other words, it sends the residue classes

$$[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6$$

to $[0]_2, [1]_2, [0]_2, [1]_2, [0]_2, [1]_2$, respectively

(since $[2]_2 = [0]_2$ and $[3]_2 = [1]_2$ and $[4]_2 = [0]_2$ and $[5]_2 = [1]_2$). More generally, for arbitrary positive integers $n$ and $d$ satisfying $d \mid n$, the map $\pi_{n,d}$ sends the $n$ residue classes $[0]_n, [1]_n, \ldots, [n-1]_n$ to $[0]_d, [1]_d, \ldots, [d-1]_d, [0]_d, [1]_d, \ldots, [d-1]_d$ (that is, $[0]_d, [1]_d, \ldots, [d-1]_d$ in this order, repeated $\frac{n}{d}$ many times), respectively.

**(b)** For a non-example, set $n = 3$ and $d = 2$. Then, Proposition 3.4.10 does not apply, since 2 is not a divisor of 3. And for good reason: There is no map

$$\pi_{3,2} : \mathbb{Z}/3 \to \mathbb{Z}/2,$$

$$[s]_3 \mapsto [s]_2.$$  

Indeed, this map would have to send $[0]_3$ and $[3]_3$ to $[0]_2$ and $[3]_2$, respectively; but this means sending two equal inputs to different outputs (since $[0]_3 = [3]_3$ but $[0]_2 \neq [3]_2$), which is impossible. More generally, if a positive integer $d$ is not a divisor of a positive integer $n$, then there is no map

$$\pi_{n,d} : \mathbb{Z}/n \to \mathbb{Z}/d,$$

$$[s]_n \mapsto [s]_d.$$
Proof of Proposition 3.4.10. We must prove that, for an integer \( s \in \mathbb{Z} \), the class \([s]_d \in \mathbb{Z}/d\) depends only on the residue class \([s]_n\), but not on the integer \( s \) itself. In other words, we need to prove the following claim:

Claim 1: If \( s_1 \) and \( s_2 \) are two integers such that \([s_1]_n = [s_2]_n\), then \([s_1]_d = [s_2]_d\).

[Proof of Claim 1: Let \( s_1 \) and \( s_2 \) be two integers such that \([s_1]_n = [s_2]_n\).

Proposition 3.4.5 (b) (applied to \( a = s_1 \) and \( b = s_2 \)) shows that we have \([s_1]_n = [s_2]_n\) if and only if \( s_1 \equiv s_2 \mod n \). Thus, we have \( s_1 \equiv s_2 \mod n \) (since \([s_1]_n = [s_2]_n\)). Hence, Proposition 2.3.4 (e) (applied to \( s_1 \), \( s_2 \) and \( d \) instead of \( a \), \( b \) and \( m \)) yields \( s_1 \equiv s_2 \mod d \) (since \( d \mid n \)).

But Proposition 3.4.5 (b) (applied to \( d \), \( s_1 \) and \( s_2 \) instead of \( n \), \( a \) and \( b \)) shows that we have \([s_1]_d = [s_2]_d\) if and only if \( s_1 \equiv s_2 \mod d \). Thus, we have \([s_1]_d = [s_2]_d\) (since \( s_1 \equiv s_2 \mod d \)). This proves Claim 1.]

Having proven Claim 1, we can now conclude that the map

\[
\pi_{n,d} : \mathbb{Z}/n \to \mathbb{Z}/d,
\]

\([s]_n \mapsto [s]_d\)

is well-defined. (This can be regarded as a consequence of applying Theorem 3.4.9 to \( \mathbb{Z}/n \), \( \mathbb{Z}/d \), \( \equiv \) and \([s]_d\) instead of \( S \), \( T \), \( \sim \) and \( F(s) \). The Claim 1 that we proved above guarantees that Assumption 1 of Theorem 3.4.9 is satisfied.) Hence, Proposition 3.4.10 is proven.

The next exercise is unrelated to \( \mathbb{Z}/n \), but has been placed in this section because it relies on the same sort of “well-definedness” argument that we have seen in our proofs above:

**Exercise 3.4.1.** Fix a prime \( p \). For each nonzero rational number \( r \), define an integer \( w_p(r) \) (called the extended \( p \)-adic valuation of \( r \)) as follows: We write \( r = a/b \) for two nonzero integers \( a \) and \( b \), and we set \( w_p(r) = v_p(a) - v_p(b) \). (It also makes sense to set \( w_p(0) = \infty \), but we shall not concern ourselves with this border case in this exercise.)

(a) Prove that this is well-defined – i.e., that \( w_p(r) \) does not depend on the precise choice of \( a \) and \( b \) satisfying \( r = a/b \).

(b) Prove that \( w_p(n) = v_p(n) \) for each nonzero integer \( n \).

(c) Prove that \( w_p(ab) = w_p(a) + w_p(b) \) for any two nonzero rational numbers \( a \) and \( b \).

(d) Prove that \( w_p(a + b) \geq \min\{w_p(a), w_p(b)\} \) for any two nonzero rational numbers \( a \) and \( b \) if \( a + b \neq 0 \).

### 3.4.5. Addition, subtraction and multiplication in \( \mathbb{Z}/n \)

Let us recall the concept of a binary operation (defined in Definition 1.6.1). We shall now define several such operations on the set \( \mathbb{Z}/n \)\textsuperscript{99}. We will check afterwards that these operations are indeed well-defined.
**Definition 3.4.12.** (a) We define a binary operation $+$ on $\mathbb{Z}/n$ (called *addition*) by setting

$$[a]_n + [b]_n = [a + b]_n$$

for any integers $a$ and $b$.

(In other words, we define a binary operation $+$ on $\mathbb{Z}/n$ as follows: For any $\alpha, \beta \in \mathbb{Z}/n$, we let $\alpha + \beta = [a + b]_n$, where $a$ and $b$ are two integers satisfying $\alpha = [a]_n$ and $\beta = [b]_n$.)

(b) We define a binary operation $-$ on $\mathbb{Z}/n$ (called *subtraction*) by setting

$$[a]_n - [b]_n = [a - b]_n$$

for any integers $a$ and $b$.

(c) We define a binary operation $\cdot$ on $\mathbb{Z}/n$ (called *multiplication*) by setting

$$[a]_n \cdot [b]_n = [a \cdot b]_n$$

for any integers $a$ and $b$. We also write $[a]_n [b]_n$ for $[a]_n \cdot [b]_n$.

**Theorem 3.4.13.** Everything defined in Definition 3.4.12 is well-defined.

*Proof of Theorem 3.4.13.* (a) Let us first prove that the binary operation $+$ in Definition 3.4.12 (a) is well-defined.

Indeed, here we are in the same situation in which we were when defining the map $R$ in the proof of Theorem 3.4.4. We are trying to define a map (in the current case, the binary operation $+$, which should be a map from $(\mathbb{Z}/n) \times (\mathbb{Z}/n)$ to $\mathbb{Z}/n$) by specifying how it acts on inputs of the form $[a]_n$, but our definition refers to the integer $a$. (Actually, it is a little bit more complicated: We have two inputs $[a]_n$ and $[b]_n$ and thus two integers $a$ and $b$. But the problem we are facing is the same.) We want to prove that this map is well-defined. This requires checking that the output (that is, $[a + b]_n$) depends only on the two classes $[a]_n$ and $[b]_n$, but not on the integers $a$ and $b$.

So we have to prove the following:

**Claim 1:** Let $a_1$ and $a_2$ be two integers such that $[a_1]_n = [a_2]_n$. Let $b_1$ and $b_2$ be two integers such that $[b_1]_n = [b_2]_n$. Then,

$$[a_1 + b_1]_n = [a_2 + b_2]_n.$$  

*Proof of Claim 1:* Proposition 3.4.5 (b) (applied to $a_1$ and $a_2$ instead of $a$ and $b$) shows that we have $[a_1]_n = [a_2]_n$ if and only if $a_1 \equiv a_2 \text{mod } n$. Thus, we have $a_1 \equiv a_2 \text{mod } n$ (since $[a_1]_n = [a_2]_n$). Similarly, $b_1 \equiv b_2 \text{mod } n$ (since $[b_1]_n = [b_2]_n$). Adding these two congruences together, we obtain $a_1 + b_1 \equiv a_2 + b_2 \text{mod } n$.

But Proposition 3.4.5 (b) (applied to $a_1 + b_1$ and $a_2 + b_2$ instead of $a$ and $b$) shows that we have $[a_1 + b_1]_n = [a_2 + b_2]_n$ if and only if $a_1 + b_1 \equiv a_2 + b_2 \text{mod } n$. Thus, we have $[a_1 + b_1]_n = [a_2 + b_2]_n$ (since $a_1 + b_1 \equiv a_2 + b_2 \text{mod } n$). This proves Claim 1.)

Claim 1 shows that in Definition 3.4.12 (a), the residue class $[a + b]_n$ depends only on the two classes $[a]_n$ and $[b]_n$, but not on the integers $a$ and $b$. Thus, the binary operation $+$ in Definition 3.4.12 (a) is indeed well-defined.
(b) The binary operation $-$ in Definition 3.4.12 (b) is well-defined. This can be proven in the same way as we just proved that the binary operation $+$ in Definition 3.4.12 (a) is well-defined; the only difference is that we now have to subtract the congruences $a_1 \equiv a_2 \mod n$ and $b_1 \equiv b_2 \mod n$ instead of adding them together.

(c) The binary operation $\cdot$ in Definition 3.4.12 (c) is well-defined. This can be proven in the same way as we just proved that the binary operation $+$ in Definition 3.4.12 (a) is well-defined; the only difference is that we now have to multiply the congruences $a_1 \equiv a_2 \mod n$ and $b_1 \equiv b_2 \mod n$ instead of adding them together.

Thus, we have proven that all three operations $+, -$ and $\cdot$ in Definition 3.4.12 are well-defined. This proves Theorem 3.4.13.

Recall that $\mathbb{Z}/n$ is a finite set (of size $n$) whenever $n$ is a positive integer. Hence, for each given positive integer $n$, we can tabulate all the values of the operations $+, -$ and $\cdot$; the resulting tables are called addition tables, subtraction tables and multiplication tables (like in high school, except that we are working with residue classes now).

**Example 3.4.14.** (a) If $n = 3$, then the addition, subtraction and multiplication tables for $\mathbb{Z}/n = \mathbb{Z}/3$ are

\[
\begin{array}{ccc}
+ & [0]_3 & [1]_3 & [2]_3 \\
[0]_3 & [0]_3 & [1]_3 & [2]_3 \\
[1]_3 & [1]_3 & [2]_3 & [0]_3 \\
[2]_3 & [2]_3 & [0]_3 & [1]_3 \\
\end{array}
\quad
\begin{array}{ccc}
- & [0]_3 & [1]_3 & [2]_3 \\
[0]_3 & [0]_3 & [2]_3 & [1]_3 \\
[1]_3 & [1]_3 & [0]_3 & [2]_3 \\
[2]_3 & [2]_3 & [1]_3 & [0]_3 \\
\end{array}
\quad
\begin{array}{ccc}
\cdot & [0]_3 & [1]_3 & [2]_3 \\
[0]_3 & [0]_3 & [1]_3 & [2]_3 \\
[1]_3 & [1]_3 & [2]_3 & [0]_3 \\
[2]_3 & [2]_3 & [1]_3 & [0]_3 \\
\end{array}
\]

(Here, the entry in the row corresponding to $\alpha$ and the column corresponding to $\beta$ is $\alpha + \beta$, $\alpha - \beta$ and $\alpha \cdot \beta$, respectively.)

(b) If $n = 2$, then the addition, subtraction and multiplication tables for $\mathbb{Z}/n = \mathbb{Z}/2$ are

\[
\begin{array}{cc}
+ & [0]_2 & [1]_2 \\
[0]_2 & [0]_2 & [1]_2 \\
[1]_2 & [1]_2 & [0]_2 \\
\end{array}
\quad
\begin{array}{cc}
- & [0]_2 & [1]_2 \\
[0]_2 & [0]_2 & [1]_2 \\
[1]_2 & [1]_2 & [0]_2 \\
\end{array}
\quad
\begin{array}{cc}
\cdot & [0]_2 & [1]_2 \\
[0]_2 & [0]_2 & [1]_2 \\
[1]_2 & [1]_2 & [0]_2 \\
\end{array}
\]

(In particular, the addition table is the same as the multiplication table, because any $\alpha, \beta \in \mathbb{Z}/2$ satisfy $\alpha + \beta = \alpha - \beta$. This follows from Exercise 2.3.1)

**Remark 3.4.15.** We **cannot** define a division operation on $\mathbb{Z}/n$ by setting

\[
[a]_n / [b]_n := [a/b]_n
\]

for any integers $a$ and $b$.

Indeed, leaving aside the issues that $b$ could be 0 or $a/b$ could be non-integer, this would still not be well-defined, because the class $[a/b]_n$ depends not just on
For the outputs of our binary operations $+, -, \cdot$ on $\mathbb{Z}/n$, we shall use the same terminology as with integers:

**Definition 3.4.16.**

(a) If $\alpha$ and $\beta$ are two elements of $\mathbb{Z}/n$, then we shall refer to $\alpha + \beta$ as the sum of $\alpha$ and $\beta$.

(b) If $\alpha$ and $\beta$ are two elements of $\mathbb{Z}/n$, then we shall refer to $\alpha - \beta$ as the difference of $\alpha$ and $\beta$.

(c) If $\alpha$ and $\beta$ are two elements of $\mathbb{Z}/n$, then we shall refer to $\alpha \cdot \beta$ (also known as $\alpha \beta$) as the product of $\alpha$ and $\beta$.

(d) If $\alpha$ is an element of $\mathbb{Z}/n$, then the difference $[0]_n - \alpha$ shall be denoted by $-\alpha$.

Caution: While the remainder $i \% n$ and the residue class $[i]_n$ encode the same information about an integer $i$ (for a fixed positive integer $n$), they are not the same thing! For example, any two integers $u$ and $v$ satisfy $[u]_n + [v]_n = [u + v]_n$ but don’t always satisfy $u \% n + v \% n = (u + v) \% n$ \footnote{Thus, it is important to distinguish between $i \% n$ and $[i]_n$.}

**Remark 3.4.17.** We can view the residue classes modulo 24 (that is, the elements of $\mathbb{Z}/24$) as the hours of the day. For example, the time “2 AM” can be viewed as the residue class $[2]_{24}$, whereas the time “3 PM” can be viewed as the residue class $[15]_{24}$. From this point of view, addition of residue classes is a rather familiar operation: For example, the statement that “10 hours from 3 PM is 1 AM” is saying $[15]_{24} + [10]_{24} = [1]_{24}$.

### 3.4.6. Scaling by $r \in \mathbb{Z}$

Let us define another operation – not binary this time – on $\mathbb{Z}/n$:

\footnote{Here is a specific example:}

$$[2]_5 + [3]_5 = [2 + 3]_5 = [5]_5 = [0]_5,$$

but

$$2 \% 5 + 3 \% 5 = 2 + 3 = 5 \neq 0 \% 5;$$

Exercise 2.6.3 (a) addresses how $u \% n + v \% n$ differs from $(u + v) \% n$. 


Definition 3.4.18. Fix \( r \in \mathbb{Z} \). For any \( \alpha \in \mathbb{Z}/n \), we define a residue class \( r\alpha \in \mathbb{Z}/n \) by setting

\[
(r[a]_n = [ra]_n \quad \text{for any } a \in \mathbb{Z}).
\]

(In other words, for any \( \alpha \in \mathbb{Z}/n \), we let \( r\alpha = [ra]_n \), where \( a \) is an integer satisfying \( \alpha = [a]_n \).) This is well-defined, because of Proposition 3.4.19 (a) below.

We also write \( r \cdot [a]_n \) for \( r[a]_n \).

Proposition 3.4.19. Fix \( r \in \mathbb{Z} \).

(a) For any \( \alpha \in \mathbb{Z}/n \), the residue class \( r\alpha \in \mathbb{Z}/n \) in Definition 3.4.18 is well-defined.

(b) For any \( \alpha \in \mathbb{Z}/n \), we have \( r\alpha = [r]_n \cdot \alpha \).

Proof of Proposition 3.4.19. (a) We are again in the same situation in which we were when defining the map \( R \) in the proof of Theorem 3.4.4: We are trying to define a map (in this case, the map \( \mathbb{Z}/n \to \mathbb{Z}/n, \alpha \mapsto r\alpha \)) by specifying how it acts on inputs of the form \([a]_n\), but our definition refers to the integer \( a \). We want to prove that this map is well-defined. This requires checking that the output (that is, \([ra]_n\)) depends only on the class \([a]_n\), but not on the integer \( a \). So we have to prove the following:

Claim 1: Let \( a_1 \) and \( a_2 \) be two integers such that \([a_1]_n = [a_2]_n\). Then, \([ra_1]_n = [ra_2]_n\).

[Proof of Claim 1: Proposition 3.4.5 (b) (applied to \( a_1 \) and \( a_2 \) instead of \( a \) and \( b \)) shows that we have \([a_1]_n = [a_2]_n\) if and only if \( a_1 \equiv a_2 \mod n \). Thus, we have \( a_1 \equiv a_2 \mod n \) (since \([a_1]_n = [a_2]_n\)). On the other hand, we have the (obvious) congruence \( r \equiv r \mod n \). Multiplying this congruence by the congruence \( a_1 \equiv a_2 \mod n \), we obtain \( ra_1 \equiv ra_2 \mod n \). But Proposition 3.4.5 (b) (applied to \( ra_1 \) and \( ra_2 \) instead of \( a \) and \( b \)) shows that we have \([ra_1]_n = [ra_2]_n\) if and only if \( ra_1 \equiv ra_2 \mod n \). Thus, we have \([ra_1]_n = [ra_2]_n\) (since \( ra_1 \equiv ra_2 \mod n \)). This proves Claim 1.]

Claim 1 shows that in Definition 3.4.18, the residue class \([ra]_n\) depends only on the class \([a]_n\), but not on the integer \( a \). Thus, the residue class \( r\alpha \) is indeed well-defined for each \( \alpha \in \mathbb{Z}/n \). This proves Proposition 3.4.19 (a).

(b) Let \( \alpha \in \mathbb{Z}/n \). Proposition 3.4.5 (a) shows that each element of \( \mathbb{Z}/n \) can be written in the form \([s]_n\) for some integer \( s \). Thus, \( \alpha \in \mathbb{Z}/n \) can be written in this form. In other words, \( \alpha = [a]_n \) for some integer \( a \). Consider this \( a \). Comparing

\[
r \cdot \alpha = r[a]_n = [ra]_n \quad \text{(by Definition 3.4.18)}
\]
with
\[ [r]_n \cdot [\alpha]_n = [a]_n = [r \cdot a]_n \] (by Definition 3.4.12 (c))
\[ = [ra]_n, \]
we obtain \( r \cdot \alpha = [r]_n \cdot \alpha \). This proves Proposition 3.4.19 (b). \( \square \)

For a fixed \( r \in \mathbb{Z} \), we shall refer to the map
\[ \mathbb{Z}/n \to \mathbb{Z}/n, \quad \alpha \mapsto r \cdot \alpha \]
as scaling by \( r \). This map is actually the same as multiplication by the residue class \([r]_n\) (by Proposition 3.4.19 (b)). So why did we define it “from scratch” rather than piggybacking on the already established definition of multiplication in \( \mathbb{Z}/n \) (Definition 3.4.12 (c))? The reason is that scaling operations appear much more frequently in algebra than multiplication operations. (For example, every vector space has a scaling operation, but usually there is no way of multiplying two vectors.) Thus, it is useful to have seen a scaling operation constructed independently.

3.4.7. \( k \)-th powers for \( k \in \mathbb{N} \)

Similarly to Definition 3.4.18, we can define what it means to take the \( k \)-th power of a residue class in \( \mathbb{Z}/n \), when \( k \) is a nonnegative integer.

**Definition 3.4.20.** Fix \( k \in \mathbb{N} \).
For any \( \alpha \in \mathbb{Z}/n \), we define a residue class \( \alpha^k \in \mathbb{Z}/n \) by setting
\[ \left( ([a]_n)^k = [a^k]_n \right) \text{ for any } a \in \mathbb{Z}. \]
(In other words, for any \( \alpha \in \mathbb{Z}/n \), we let \( \alpha^k = [a^k]_n \), where \( a \) is an integer satisfying \( \alpha = [a]_n \).) This is well-defined, because of Proposition 3.4.21 below.

If \( \alpha \in \mathbb{Z}/n \), then we shall refer to \( \alpha^k \) as the \( k \)-th power of \( \alpha \).

**Proposition 3.4.21.** Fix \( k \in \mathbb{N} \). For any \( \alpha \in \mathbb{Z}/n \), the residue class \( \alpha^k \in \mathbb{Z}/n \) in Definition 3.4.20 is well-defined.

**Proof of Proposition 3.4.21** This proof is analogous to the above proof of Proposition 3.4.19 (a); but instead of multiplying the two congruences \( r \equiv r \mod n \) and \( a_1 \equiv a_2 \mod n \), we now need to take the \( k \)-th power of the congruence \( a_1 \equiv a_2 \mod n \). (Exercise 2.3.4 allows us to do that.) \( \square \)

2019-02-25 lecture

3.4.8. Rules and properties for the operations
Math 4281 notes page 215

[]

We shall now study some properties of the many “arithmetical” operations we have defined on \( \mathbb{Z} \).

These properties should all look familiar, as they mirror the classical properties of the arithmetic operations on integers, rational numbers and real numbers (with the caveat that scaling by \( r \) and 1). For example, the expression \( \alpha \cdot \beta + \gamma \cdot \delta \) means \( (\alpha \cdot \beta) + (\gamma \cdot \delta) \) and not \( \alpha \cdot (\beta + \gamma) \cdot \delta \). Likewise, the expression \( \alpha \beta^k + r\gamma \) (with \( r \in \mathbb{Z} \)) should be understood as \( "(\alpha (\beta^k)) + (r\gamma)" \) and not in any other way.

We shall now study some properties of the many “arithmetical” operations we have defined on \( \mathbb{Z}/n \).

**Theorem 3.4.23.** The following rules for addition, subtraction, multiplication and scaling in \( \mathbb{Z}/n \) hold:

(a) We have \( \alpha + \beta = \beta + \alpha \) for any \( \alpha, \beta \in \mathbb{Z}/n \).

(b) We have \( \alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma \) for any \( \alpha, \beta, \gamma \in \mathbb{Z}/n \).

(c) We have \( \alpha + [0]_n = [0]_n + \alpha = \alpha \) for any \( \alpha \in \mathbb{Z}/n \).

(d) We have \( \alpha \cdot [1]_n = [1]_n \cdot \alpha = \alpha \) for any \( \alpha \in \mathbb{Z}/n \).

(e) We have \( \alpha \cdot \beta = \beta \cdot \alpha \) for any \( \alpha, \beta \in \mathbb{Z}/n \).

(f) We have \( \alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma \) for any \( \alpha, \beta, \gamma \in \mathbb{Z}/n \).

(g) We have \( \alpha \cdot (\beta + \gamma) = \alpha \beta + \alpha \gamma \) and \( (\alpha + \beta) \cdot \gamma = \alpha \gamma + \beta \gamma \) for any \( \alpha, \beta, \gamma \in \mathbb{Z}/n \).

(h) We have \( \alpha \cdot [0]_n = [0]_n \cdot \alpha = [0]_n \) for any \( \alpha \in \mathbb{Z}/n \).

(i) If \( \alpha, \beta, \gamma \in \mathbb{Z}/n \), then we have the equivalence \( (\alpha - \beta = \gamma) \iff (\alpha = \beta + \gamma) \).

(j) We have \( r (\alpha + \beta) = r\alpha + r\beta \) for any \( r \in \mathbb{Z} \) and \( \alpha, \beta \in \mathbb{Z}/n \).

(k) We have \( (r + s) \alpha = ra + sa \) for any \( r, s \in \mathbb{Z} \) and \( \alpha \in \mathbb{Z}/n \).

(l) We have \( r (sa) = (rs) \alpha \) for any \( r, s \in \mathbb{Z} \) and \( \alpha \in \mathbb{Z}/n \).

(m) We have \( r (\alpha \beta) = (ra) \beta = (r\beta) \alpha \) for any \( r \in \mathbb{Z} \) and \( \alpha, \beta \in \mathbb{Z}/n \).

(n) We have \( -(r\alpha) = (-r) \alpha = r(-\alpha) \) for any \( r \in \mathbb{Z} \) and \( \alpha \in \mathbb{Z}/n \).

(o) We have \( 1\alpha = \alpha \) for any \( \alpha \in \mathbb{Z}/n \).

(p) We have \( -(1) \alpha = -\alpha \) for any \( \alpha \in \mathbb{Z}/n \).

(q) We have \( -(\alpha + \beta) = (-\alpha) + (-\beta) \) for any \( \alpha, \beta \in \mathbb{Z}/n \).

(r) We have \( -[0]_n = [0]_n \).

(s) We have \( -(-\alpha) = \alpha \) for any \( \alpha \in \mathbb{Z}/n \).

(t) We have \( -(\alpha \beta) = (-\alpha) \beta = \alpha (-\beta) \) for any \( \alpha, \beta \in \mathbb{Z}/n \).

(u) We have \( \alpha - \beta - \gamma = \alpha - (\beta + \gamma) \) for any \( \alpha, \beta, \gamma \in \mathbb{Z}/n \). (Here and in the following, “\( \alpha - \beta - \gamma \)” should be read as “\( (\alpha - \beta) - \gamma \)”.)

These properties should all look familiar, as they mirror the classical properties of the arithmetic operations on integers, rational numbers and real numbers (with the caveat that the residue classes \([0]_n\) and \([1]_n\) take on the roles of the numbers 0 and 1). For example, Theorem 3.4.23 (g) corresponds to the laws of distributivity for numbers. Parts (a), (b), (c), (i), (j), (k), (l) and (o) of Theorem 3.4.23 furthermore are reminiscent of the axioms for a vector space (with the caveat that scaling by \( r \) is only defined for integers \( r \) here, so \( \mathbb{Z}/n \) is not precisely a vector space).
Proof of Theorem 3.4.23 Let us first prove Theorem 3.4.23 (f):

(f) Let \(\alpha, \beta, \gamma \in \mathbb{Z}/n\). Proposition 3.4.5 (a) shows that each element of \(\mathbb{Z}/n\) can be written in the form \([s]_n\) for some integer \(s\). Thus, in particular, the element \(\alpha\) can be written in this form. In other words, there exists an integer \(a\) such that \(\alpha = [a]_n\). Similarly, there exists an integer \(b\) such that \(\beta = [b]_n\). Similarly, there exists an integer \(c\) such that \(\gamma = [c]_n\). Consider these integers \(a, b, c\).

Now,

\[
\alpha \cdot \left(\beta \cdot \gamma\right) = [a]_n \cdot \left([b]_n \cdot [c]_n\right) = [a]_n \cdot [b \cdot c]_n
\]

(by Definition 3.4.12 (c))

\[
\begin{align*}
\left(\alpha \cdot \beta\right) \cdot \gamma &= \left([a]_n \cdot [b]_n\right) \cdot [c]_n = [a \cdot b]_n \cdot [c]_n \\
&= \left([a \cdot b \cdot c]_n\right)
\end{align*}
\]

(by Definition 3.4.12 (c)).

But it is well-known that multiplication of integers is associative. Thus, \(a \cdot (b \cdot c) = (a \cdot b) \cdot c\). Hence, \([a \cdot (b \cdot c)]_n = [(a \cdot b) \cdot c]_n\). In other words, the right hand sides of the equalities (108) and (109) are equal. Hence, the left hand sides of these equalities must also be equal. In other words, \(\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma\). This proves Theorem 3.4.23 (f).

The idea of the above proof of Theorem 3.4.23 (f) was simple: We fixed a representative for each residue class involved (namely, we fixed representatives \(a, b, c\) for the residue classes \(\alpha, \beta, \gamma\), and rewrote each of the two sides of the alleged equality (which, in our case, was \(\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma\) in terms of these representatives (obtaining \([a \cdot (b \cdot c)]_n\) for the left hand side, and \([(a \cdot b) \cdot c]_n\) for the right hand side). Thus, the equality that we had to prove followed from the analogous equality for integers (in our case, \(a \cdot (b \cdot c) = (a \cdot b) \cdot c\), which was well-known. In short, we have realized that the equality \(\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma\) for \(\alpha, \beta, \gamma \in \mathbb{Z}/n\) is “inherited from \(\mathbb{Z}\)” (in the sense that it follows straightforwardly from the corresponding property \(a \cdot (b \cdot c) = (a \cdot b) \cdot c\) of integers \(a, b, c \in \mathbb{Z}\)). This strategy proves all the other parts of Theorem 3.4.23 in the same way, except for part (i).\(^{102}\) Part (i) does not lend itself to such a proof, since it claims not an equality but rather an equivalence between two equalities. So let us prove part (i) separately:

\(^{102}\)The reason why this works is that the operations \(+, -, \cdot\) on \(\mathbb{Z}/n\) as well as scaling by integers are defined by picking a representative of each residue class and doing the analogous operation with the representatives (and then taking the residue class again).
(i) Let \( \alpha, \beta, \gamma \in \mathbb{Z}/n \). Proposition 3.4.5 (a) shows that each element of \( \mathbb{Z}/n \) can be written in the form \([s]_n\) for some integer \( s \). Thus, in particular, the element \( \alpha \) can be written in this form. In other words, there exists an integer \( a \) such that \( \alpha = [a]_n \). Similarly, there exists an integer \( b \) such that \( \beta = [b]_n \). Similarly, there exists an integer \( c \) such that \( \gamma = [c]_n \). Consider these integers \( a, b, c \).

We have \( \alpha - \beta = [a]_n - [b]_n = [a - b]_n \) (by Definition 3.4.12 (b)) and \( \beta + \gamma = [b]_n + [c]_n = [b + c]_n \) (by Definition 3.4.12 (a)). Now, we have the following chain of logical equivalences:

\[
(\alpha - \beta = \gamma) \iff ([a - b]_n = [c]_n) \quad \text{(since } \alpha - \beta = [a - b]_n \text{ and } \gamma = [c]_n \text{)} \iff (a - b \equiv c \mod n) \quad (110)
\]

(since Proposition 3.4.5 (b) (applied to \( a - b \) and \( c \) instead of \( a \) and \( b \)) shows that we have \([a - b]_n = [c]_n \) if and only if \( a - b \equiv c \mod n \)). Also, we have the following chain of logical equivalences:

\[
(\alpha = \beta + \gamma) \iff ([a]_n = [b + c]_n) \quad \text{(since } \alpha = [a]_n \text{ and } \beta + \gamma = [b + c]_n \text{)} \iff (a \equiv b + c \mod n) \quad (111)
\]

(since Proposition 3.4.5 (b) (applied to \( b + c \) instead of \( b \)) shows that we have \([a]_n = [b + c]_n \) if and only if \( a \equiv b + c \mod n \)). Finally, Exercise 2.3.8 shows that we have \( a - b \equiv c \mod n \) if and only if \( a \equiv b + c \mod n \); thus, we have the equivalence \((a - b \equiv c \mod n) \iff (a \equiv b + c \mod n)\).

Now, we have the following chain of logical equivalences:

\[
(\alpha - \beta = \gamma) \iff (a - b \equiv c \mod n) \quad \text{(by (110))} \iff (a \equiv b + c \mod n) \iff (\alpha = \beta + \gamma) \quad \text{(by (111))}.
\]

This proves Theorem 3.4.23 (i). \( \square \)

Recall the concept of a finite sum of integers (i.e., a sum of the form \( \sum_{i \in I} a_i \), where \( I \) is a finite set and \( a_i \) is an integer for each \( i \in I \)), and the analogous concept of a finite product of integers (i.e., a product of the form \( \prod_{i \in I} a_i \)). These concepts are defined recursively\(^{103}\) and satisfy various rules\(^{104}\). See [Grinbe15 §1.4] for a comprehensive list of these rules and [Grinbe15 §2.14] for their proofs.

---

\(^{103}\)See [Grinbe15 §1.4.1 and §1.4.3] for their definitions, and [Grinbe15 §2.14] for a proof that these are well-defined.

\(^{104}\)such as \( \sum_{i \in I} (a_i + b_i) = \sum_{i \in I} a_i + \sum_{i \in I} b_i \) (where \( a_i \) and \( b_i \) are two integers for each \( i \in I \)) or \( \sum_{i \in I} a_i = \sum_{i \in I \setminus J} a_i \) (where \( J \) is a subset of \( I \))
Definition 3.4.24. In the same vein, we define the concept of a finite sum of residue classes in \( \mathbb{Z}/n \) (i.e., a sum of the form \( \sum_{i \in I} \alpha_i \), where \( I \) is a finite set and \( \alpha_i \in \mathbb{Z}/n \) for each \( i \in I \)), and the analogous concept of a finite product of residue classes in \( \mathbb{Z}/n \) (i.e., a product of the form \( \prod_{i \in I} \alpha_i \), where \( I \) is a finite set and \( \alpha_i \in \mathbb{Z}/n \) for each \( i \in I \)).

More precisely, the concept of a finite sum \( \sum_{i \in I} \alpha_i \) (with \( I \) being a finite set, and with \( \alpha_i \in \mathbb{Z}/n \) for each \( i \in I \)) is defined recursively as follows:

- If the set \( I \) is empty (that is, \(|I| = 0\)), then \( \sum_{i \in I} \alpha_i \) is defined to be \( [0]_n \in \mathbb{Z}/n \) (and called an empty sum).
- Otherwise, we pick an arbitrary element \( t \in I \), and set
  \[
  \sum_{i \in I} \alpha_i = \alpha_t + \sum_{i \in I \setminus \{t\}} \alpha_i.
  \]

(The sum \( \sum_{i \in I \setminus \{t\}} \alpha_i \) on the right hand side is a sum over a smaller set than \( I \), whence we can assume it to already be defined in this recursive definition.)

This definition is well-defined (i.e., the choice of element \( t \) does not influence the final value of the sum), by Proposition 3.4.25 (a) below.

The concept of a finite product \( \prod_{i \in I} \alpha_i \) is defined similarly, except that we use multiplication instead of addition (and we define the empty product to be \( [1]_n \) instead of \( [0]_n \)).

We will use the usual shorthands for special kinds of finite sums and products. For example, if \( I \) is an interval \( \{p, p+1, \ldots, q\} \) of integers (and if \( \alpha_i \in \mathbb{Z}/n \) for each \( i \in I \)), then the sum \( \sum_{i \in I} \alpha_i \) will also be denoted by \( \sum_{p}^{q} \alpha_i \) or \( \alpha_p + \alpha_{p+1} + \cdots + \alpha_q \).

Likewise for products. Thus, for example, \( \alpha_1 + \alpha_2 + \cdots + \alpha_k \) and \( \alpha_1 \alpha_2 \cdots \alpha_k \) are well-defined whenever \( \alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{Z}/n \).

Proposition 3.4.25. (a) Definition 3.4.24 is well-defined.
(b) Finite sums (\( \sum_{i \in I} \alpha_i \)) and finite products (\( \prod_{i \in I} \alpha_i \)) of elements \( \alpha_i \in \mathbb{Z}/n \) satisfy the same rules that finite sums and finite products of integers satisfy.
(c) If \( a_1, a_2, \ldots, a_k \) are \( k \) integers, then
\[
[a_1]_n + [a_2]_n + \cdots + [a_k]_n = [a_1 + a_2 + \cdots + a_k]_n \quad \text{and} \quad [a_1]_n \cdot [a_2]_n \cdots [a_k]_n = [a_1 a_2 \cdots a_k]_n.
\]

Proof of Proposition 3.4.25 (a) In [Grinbe15 Theorem 2.118 (a)], it is proven that fi-
finite sums of integers are well-defined. The same argument (but relying on Theorem 3.4.23 instead of the usual rules of commutativity, associativity etc. for integers) shows that finite sums of elements $\alpha_i \in \mathbb{Z}/n$ are well-defined. The analogous fact for products is proven in the same way, except that we need to replace $[0]_n$ by $[1]_n$ and properties of addition by corresponding properties of multiplication.

(b) The proofs of the properties of finite sums and finite products of elements of $\mathbb{Z}/n$ are identical to the analogous proofs for integers, but (again) rely on Theorem 3.4.23 instead of the usual rules of commutativity, associativity etc. for integers.

(c) This can be proven by a straightforward induction on $k$. □

Also, the standard rules for exponents apply to residue classes:

**Theorem 3.4.26.** (a) We have $\alpha^0 = [1]_n$ for any $\alpha \in \mathbb{Z}/n$.
(b) We have $\alpha^1 = \alpha$ for any $\alpha \in \mathbb{Z}/n$.
(c) We have $\alpha^k = \underbrace{\alpha \cdot \alpha \cdots \alpha}_k$ for any $\alpha \in \mathbb{Z}/n$ and $k \in \mathbb{N}$.
(d) We have $\alpha^{u+v} = \alpha^u \alpha^v$ for any $\alpha \in \mathbb{Z}/n$ and any $u, v \in \mathbb{N}$.
(e) We have $(\alpha \beta)^k = \alpha^k \beta^k$ for any $\alpha, \beta \in \mathbb{Z}/n$ and $k \in \mathbb{N}$.
(f) We have $(\alpha^u)^v = \alpha^{uv}$ for any $\alpha \in \mathbb{Z}/n$ and any $u, v \in \mathbb{N}$.

**Proof of Theorem 3.4.26.** Each part of Theorem 3.4.26 follows from the analogous property of integers, in the same way as we derived Theorem 3.4.23 (f) from the associativity of multiplication for integers. (Note that Proposition 3.4.25 (c) has to be used in proving Theorem 3.4.26 (c).) □

Also, the binomial formula holds for residue classes:

**Theorem 3.4.27.** Let $\alpha, \beta \in \mathbb{Z}/n$ and $m \in \mathbb{N}$. Then,

$$\left(\alpha + \beta\right)^m = \sum_{k=0}^{m} \binom{m}{k} \alpha^k \beta^{m-k}.$$ 

**Proof.** This follows from Theorem 2.17.13, in the same way as we derived Theorem 3.4.23 (f) from the associativity of multiplication for integers. □

### 3.5. Modular inverses revisited

**Convention 3.5.1.** For the whole Section 3.5, we fix a positive integer $n$.

In this section, we will see how modular inverses become actual inverses when we consider residue classes instead of numbers.

Recall that if $a$ is an integer, then an *inverse of $a$ in $\mathbb{Z}$* means an integer $a' \in \mathbb{Z}$ satisfying $aa' = 1$. The only two integers that have an inverse in $\mathbb{Z}$ are 1 and $-1$. The integer 1 has only one inverse (namely, itself). The integer $-1$ has only one inverse (namely, itself). Thus, “inverse in $\mathbb{Z}$” is not a very interesting notion.

Let us now define an analogous notion for $\mathbb{Z}/n$:
Definition 3.5.2. Let $\alpha \in \mathbb{Z}/n$. An inverse of $\alpha$ means an $\alpha' \in \mathbb{Z}/n$ such that $\alpha \cdot \alpha' = [1]_n$.

For example, $[2]_5$ is an inverse of $[3]_5$ for $n = 5$, since $[3]_5 \cdot [2]_5 = [3 \cdot 2]_5 = [6]_5 = [1]_5$.

It turns out that inverses of residue classes $\alpha \in \mathbb{Z}/n$ exist much more frequently than inverses of integers in $\mathbb{Z}$:

Proposition 3.5.3. Let $a \in \mathbb{Z}$.

(a) If $[a]_n \in \mathbb{Z}/n$ has an inverse, then $a \perp n$.

(b) If $a \perp n$, then $[a]_n \in \mathbb{Z}/n$ has a unique inverse.

As we will see in the proof of this proposition, the inverse of a residue class $[a]_n$ is simply the residue class $[a']_n$ of a modular inverse $a'$ of $a$ modulo $n$; thus, the existence part of Proposition 3.5.3 (i.e., the claim that $[a]_n$ has an inverse) is just Theorem 2.10.8 (b) in disguise. However, before we start proving Proposition 3.5.3, let us state the uniqueness part (i.e., the claim that the inverse of $[a]_n$ is unique) as a separate fact:

Proposition 3.5.4. Let $\alpha \in \mathbb{Z}/n$. Then, $\alpha$ has at most one inverse.

Proof of Proposition 3.5.4 Let $\beta$ and $\gamma$ be two inverses of $\alpha$. We shall show that $\beta = \gamma$.

We have $\alpha \cdot \beta = [1]_n$ (since $\beta$ is an inverse of $\alpha$) and $\alpha \cdot \gamma = [1]_n$ (since $\gamma$ is an inverse of $\alpha$). Theorem 3.4.23 (e) (applied to $\gamma$ and $\alpha$ instead of $\alpha$ and $\beta$) yields $\gamma \cdot \alpha = \alpha \cdot \gamma = [1]_n$.

Theorem 3.4.23 (d) (applied to $\gamma$ instead of $\alpha$) yields $\gamma \cdot [1]_n = [1]_n \cdot \gamma = \gamma$. Also, Theorem 3.4.23 (d) (applied to $\beta$ instead of $\alpha$) yields $\beta \cdot [1]_n = [1]_n \cdot \beta = \beta$.

Theorem 3.4.23 (f) (applied to $\gamma$, $\alpha$ and $\beta$ instead of $\alpha$, $\beta$ and $\gamma$) shows that $\gamma \cdot (\alpha \cdot \beta) = (\gamma \cdot \alpha) \cdot \beta$. Now, comparing

$$\gamma \cdot \left( \begin{array}{c} \alpha \cdot \beta \\ = [1]_n \end{array} \right) = \gamma \cdot [1]_n = \gamma$$

with

$$\gamma \cdot (\alpha \cdot \beta) = \left( \begin{array}{c} \gamma \cdot \alpha \\ = [1]_n \end{array} \right) \cdot \beta = [1]_n \cdot \beta = \beta,$$

we obtain $\beta = \gamma$.

Now, forget that we fixed $\beta$ and $\gamma$. We thus have proven that if $\beta$ and $\gamma$ are two inverses of $\alpha$, then $\beta = \gamma$. In other words, any two inverses of $\alpha$ must be equal. In other words, $\alpha$ has at most one inverse. This proves Proposition 3.5.4. \qed
Note that in the above proof of Proposition 3.5.4, we have never had to pick a representative of the residue class \(a\) (nor of any other class). This is because this proof is actually an instance of a much more general argument. And indeed, you might recall that a very similar argument is used to prove the classical facts that

- a map has at most one inverse;
- a matrix has at most one inverse.

To be more precise, the proofs of these two facts differ slightly from our proof of Proposition 3.5.4 because the definitions of an inverse of a map and of an inverse of a matrix differ from Definition 3.5.2. Indeed, in Definition 3.5.2, we have only required the inverse \(a'\) of \(a \in \mathbb{Z}/n\) to satisfy the single equation \(a \cdot a' = [1]_n\), whereas an inverse \(g\) of a map \(f\) is required to satisfy the two equations \(f \circ g = \text{id}\) and \(g \circ f = \text{id}\) (and likewise, an inverse \(B\) of a matrix \(A\) is required to satisfy the two equations \(AB = I\) and \(BA = I\) for the appropriate identity matrices \(I\)).

But this difference is not substantial: The multiplication of residue classes in \(\mathbb{Z}/n\) is commutative (by Theorem 3.4.23 (e)) (unlike the composition of maps or the multiplication of matrices); thus, the single equation \(a \cdot a' = [1]_n\) automatically implies \(a' \cdot a = [1]_n\). Hence, we could have as well required \(a'\) to satisfy both equations \(a \cdot a' = [1]_n\) and \(a' \cdot a = [1]_n\) in Definition 3.5.2 and nothing would change.

Let us now prove Proposition 3.5.3.

**Proof of Proposition 3.5.3**

(a) Assume that \([a]_n \in \mathbb{Z}/n\) has an inverse. Let \(b\) be this inverse. Write the residue class \(\beta \in \mathbb{Z}/n\) in the form \([b]_n\) for some integer \(b\). Now, \(\beta\) is an inverse of \([a]_n\). In other words, \([a]_n \cdot \beta = [1]_n\) (by the definition of “inverse”). But \([a]_n \cdot \beta = [a]_n \cdot [b]_n = [a \cdot b]_n\) (by the definition of multiplication on \(\mathbb{Z}/n\)).

Comparing this with \([a]_n \cdot \beta = [1]_n\) we obtain \([a \cdot b]_n = [1]_n\). By Proposition 3.4.5 (applied to \(a \cdot b\) and 1 instead of \(a\) and \(b\)), this yields \(a \cdot b \equiv 1 \mod n\). Hence, there exists an \(a' \in \mathbb{Z}\) such that \(aa' \equiv 1 \mod n\) (namely, \(a' = b\)). Thus, Theorem 2.10.8 (c) yields \(a \perp n\). This proves Proposition 3.5.3(a).

(b) Assume that \(a \perp n\). Hence, Theorem 2.10.8 (b) yields that there exists an \(a' \in \mathbb{Z}\) such that \(aa' \equiv 1 \mod n\). Consider this \(a'\). From \(aa' \equiv 1 \mod n\), we obtain \([aa']_n = [1]_n\) (by Proposition 3.4.5 (b), applied to \(aa'\) and 1 instead of \(a\) and \(b\)). But the definition of multiplication on \(\mathbb{Z}/n\) yields \([a]_n \cdot [a']_n = [a \cdot a']_n = [aa']_n = [1]_n\).

In other words, \([a']_n\) is an inverse of \([a]_n\). Hence, \([a]_n\) has at least one inverse (namely, \([a']_n\)).

But Proposition 3.5.4 (applied to \(\alpha = [a]_n\)) shows that \([a]_n\) has at most one inverse.

Thus, we conclude that \([a]_n\) has a unique inverse (since we already know that \([a]_n\) has at least one inverse and has at most one inverse). This proves Proposition 3.5.3. \(\square\)
**Corollary 3.5.5.** Let $U_n$ be the set of all residue classes $\alpha \in \mathbb{Z}/n$ that have an inverse. Then:

(a) For an integer $a$, we have the logical equivalence $([a]_n \in U_n) \iff (a \perp n)$.

(b) We have $|U_n| = \phi(n)$.

**Proof of Corollary 3.5.5.** (a) Let $a$ be an integer. Proposition 3.5.3 (a) yields the logical implication

$$([a]_n \text{ has an inverse}) \implies (a \perp n).$$

(112)

But Proposition 3.5.3 (b) yields the logical implication

$$(a \perp n) \implies ([a]_n \text{ has a unique inverse}) \implies ([a]_n \text{ has an inverse}).$$

(113)

Combining the two implications (112) and (113), we obtain the equivalence

$$([a]_n \text{ has an inverse}) \iff (a \perp n).$$

(114)

But $U_n$ was defined as the set of all residue classes $\alpha \in \mathbb{Z}/n$ that have an inverse. Hence, we have the following chain of equivalences:

$$([a]_n \in U_n) \iff ([a]_n \text{ has an inverse}) \iff (a \perp n)$$

(by (114)). This proves Corollary 3.5.5 (a).

(b) Theorem 3.4.4 says that the set $\mathbb{Z}/n$ has exactly $n$ elements, namely $[0]_n, [1]_n, \ldots, [n-1]_n$ (and in particular, these $n$ elements are all distinct). Thus, the map

$$P : \{0, 1, \ldots, n-1\} \to \mathbb{Z}/n,$$

$$s \mapsto [s]_n$$

is bijective.\(^{105}\) Consider this map $P$. For each integer $a$, we have the logical equivalence

$$
\begin{pmatrix}
P(a) \\
= [a]_n \\
\in U_n
\end{pmatrix}
\iff ([a]_n \in U_n) \iff (a \perp n)
$$

(115)

(by Corollary 3.5.5 (a)).

But the map $P$ is bijective. Hence, we can substitute $P(a)$ for $\alpha$ when counting

\(^{105}\)We also have explicitly proven this fact during our proof of Theorem 3.4.4.
the number of \( \alpha \in U_n \). We thus obtain

\[
\begin{align*}
& \text{(the number of } \alpha \in U_n) \\
& = (\text{the number of } a \in \{0, 1, \ldots, n - 1\} \text{ such that } P(a) \in U_n) \\
& = (\text{the number of } a \in \{0, 1, \ldots, n - 1\} \text{ such that } a \perp n) \\
& \quad \left(\text{because for each } a \in \{0, 1, \ldots, n - 1\}, \text{we have the logical equivalence } (P(a) \in U_n) \iff (a \perp n) \text{ (by (115))}\right)
\end{align*}
\]

\[
\begin{align*}
& = \left| \{a \in \{0, 1, \ldots, n - 1\} \mid a \perp n\} \right| \\
& = \left| \{i \in \{0, 1, \ldots, n - 1\} \mid i \perp n\} \right| \quad \text{(here, we have renamed the index } a \text{ as } i) \\
& = \phi(n) \\
& \quad \text{(by Lemma 2.15.4). This proves Corollary 3.5.5(b).}
\end{align*}
\]

**Definition 3.5.6.** Let \( \alpha \in \mathbb{Z}/n \) be a residue class that has an inverse. Then, Proposition 3.5.4 shows that \( \alpha \) has a unique inverse. This inverse can thus be called “the inverse” of \( \alpha \); it will be denoted by \( \alpha^{-1} \).

For example, \(([3]_5)^{-1} = [2]_5\) for \( n = 5 \), since \([2]_5\) is an inverse (and thus the inverse) of \([3]_5\).

Let us state a couple properties of inverses in \( \mathbb{Z}/n \):

**Exercise 3.5.1.** (a) Let \( \alpha \in \mathbb{Z}/n \) be a residue class that has an inverse. Prove that its inverse \( \alpha^{-1} \) has an inverse as well, and this inverse is \((\alpha^{-1})^{-1} = \alpha\).

(b) Let \( \alpha, \beta \in \mathbb{Z}/n \) be two residue classes that have inverses. Prove that their product \( \alpha \beta \) has an inverse as well, and this inverse is \((\alpha \beta)^{-1} = \alpha^{-1} \beta^{-1}\).

The concept of inverses in \( \mathbb{Z}/n \) lets us prove Theorem 2.15.7 (Wilson’s theorem) again – or, rather, restate our previous proof of Theorem 2.15.7 in more natural terms:

**Second proof of Theorem 2.15.7 (sketched).** Theorem 3.4.4 (applied to \( n = p \)) shows that the set \( \mathbb{Z}/p \) has exactly \( p \) elements, namely \([0]_p, [1]_p, \ldots, [p - 1]_p\). In particular, these elements \([0]_p, [1]_p, \ldots, [p - 1]_p\) are distinct.

If \( p = 2 \), then the claim of Theorem 2.15.7 is easy to check (as we have done in our First proof above). Thus, we WLOG assume that \( p \neq 2 \) for the rest of this proof. Thus, \( p - 1 \neq 1 \); in other words, the numbers 1 and \( p - 1 \) are distinct. But \( p \) is a prime; thus, \( p > 1 \), so that the elements 1 and \( p - 1 \) belong to the set \( \{0, 1, \ldots, p - 1\} \). Thus, the two residue classes \([1]_p\) and \([p - 1]_p\) are distinct.\footnote{Proof. Recall that the elements \([0]_p, [1]_p, \ldots, [p - 1]_p\) are distinct. In other words, if \( i \) and \( j \) are two distinct elements of \( \{0, 1, \ldots, p - 1\} \), then \([i]_p \neq [j]_p\). We can apply this to \( i = 1 \) and \( j = p - 1 \), since 1 and \( p - 1 \) are distinct. Thus, we obtain \([1]_p \neq [p - 1]_p\). In other words, the two residue classes \([1]_p\) and \([p - 1]_p\) are distinct.}
Recall that
\[(p - 1)! = 1 \cdot 2 \cdots (p - 1).\]
Thus,
\[
[(p - 1)!]_p = [1 \cdot 2 \cdots (p - 1)]_p = [1]_p \cdot [2]_p \cdots [p - 1]_p
\]
(by Proposition 3.4.25(c)).

Let \(U_p\) be the set of all residue classes \(\alpha \in \mathbb{Z}/p\) that have an inverse. Then, \([0]_p \notin U_p\). On the other hand, the \(p - 1\) residue classes \([1]_p, [2]_p, \ldots, [p - 1]_p\) all belong to \(U_p\). Combining these two sentences, we conclude that the \(p - 1\) residue classes \([1]_p, [2]_p, \ldots, [p - 1]_p\) are precisely the elements of \(U_p\) (since the set \(\mathbb{Z}/p\) has exactly \(p\) elements, namely \([0]_p, [1]_p, \ldots, [p - 1]_p\)). Thus, these \(p - 1\) residue classes have inverses (because belonging to \(U_p\) means having an inverse), and their inverses in turn have inverses (by Exercise 3.5.1(a)) and thus belong to \(U_p\) (because belonging to \(U_p\) means having an inverse). Thus, the map
\[
J : U_p \to U_p,
\]
\[
\alpha \mapsto \alpha^{-1}
\]
(sending each of the \(p - 1\) residue classes \([1]_p, [2]_p, \ldots, [p - 1]_p\) to its inverse) is well-defined. Moreover, each \(\alpha \in U_p\) satisfies \((\alpha^{-1})^{-1} = \alpha\); in other words, each \(\alpha \in U_p\) satisfies \(J(J(\alpha)) = \alpha\). In other words, \(J \circ J = \text{id}\). Hence, the map \(J\) is invertible, i.e., bijective.

(Note that this map \(J\) is similar to the map \(J\) constructed back in our first proof of Theorem 2.15.7 above, but unlike the latter, it acts on residue classes, not on actual numbers.)

Note that
\[
[1]_p \cdot [2]_p \cdots [p - 1]_p = \prod_{\alpha \in U_p} \alpha,
\]
since the \(p - 1\) residue classes \([1]_p, [2]_p, \ldots, [p - 1]_p\) are precisely the elements of \(U_p\) (and are distinct).

Now, we shall complete the proof using the same “pairing” that we used in our first proof of Theorem 2.15.7, except that we will now be pairing up residue classes rather than numbers. Namely, we will use the map \(J\) to establish a pairing between

\[\text{Proof. We have } p > 1; \text{ thus, we don’t have } |p| = 1. \text{ But Exercise 2.10.1(b) (applied to } a = p) \text{ shows that we have } 0 \perp p \text{ if and only if } |p| = 1. \text{ Thus, we don’t have } 0 \perp p \text{ (since we don’t have } |p| = 1).\]

Corollary 3.5.5(a) (applied to \(n = p\) and \(a = 0\)) shows that we have the logical equivalence \((0]_p \in U_p) \iff (0 \perp p)\). Since we don’t have \(0 \perp p\), we thus conclude that we don’t have \([0]_p \in U_p\). In other words, we have \([0]_p \notin U_p\).

\[\text{Proof. We must show that } [i]_p \in U_p \text{ for each } i \in \{1, 2, \ldots, p - 1\}.\]

So let \(i \in \{1, 2, \ldots, p - 1\}\). Then, Proposition 2.13.3 shows that \(i\) is coprime to \(p\). In other words, \(i \perp p\).

But Corollary 3.5.5(a) (applied to \(n = p\) and \(a = i\)) yields the equivalence \(([i]_p \in U_p) \iff (i \perp p)\). Hence, we have \([i]_p \in U_p\) (since \(i \perp p\). Qed.
the factors of the product \([1]_p \cdot [2]_p \cdots \cdot [p-1]_p = \prod_{\alpha \in U_p} \alpha\) (pairing up each factor \(\alpha\) with the factor \(J(\alpha) = \alpha^{-1}\)), which will pair up almost all of them – more precisely, all of them except for the very first and very last factors (since these two factors would have to pair up with themselves\(^{109}\)). For example, if \(p = 11\), then we have the following table of values of \(J:\)

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>([1]_{11})</th>
<th>([2]_{11})</th>
<th>([3]_{11})</th>
<th>([4]_{11})</th>
<th>([5]_{11})</th>
<th>([6]_{11})</th>
<th>([7]_{11})</th>
<th>([8]_{11})</th>
<th>([9]_{11})</th>
<th>([10]_{11})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(J(\alpha))</td>
<td>([1]_{11})</td>
<td>([6]_{11})</td>
<td>([4]_{11})</td>
<td>([3]_{11})</td>
<td>([9]_{11})</td>
<td>([2]_{11})</td>
<td>([8]_{11})</td>
<td>([7]_{11})</td>
<td>([5]_{11})</td>
<td>([10]_{11})</td>
</tr>
</tbody>
</table>

(since, for example, \(J([2]_{11}) = ([2]_{11})^{-1} = [6]_{11}\), and thus we pair up the factors of the product \([1]_p \cdot [2]_p \cdots \cdot [p-1]_p\) as follows:

\[
\begin{align*}
[1]_p \cdot [2]_p \cdots \cdot [p-1]_p &= [1]_{11} \cdot [2]_{11} \cdot [3]_{11} \cdot [4]_{11} \cdot [5]_{11} \cdot [6]_{11} \cdot [7]_{11} \cdot [8]_{11} \cdot [9]_{11} \cdot [10]_{11} \\
&= [1]_{11} \cdot ([2]_{11} \cdot [6]_{11}) \cdot ([3]_{11} \cdot [4]_{11}) \\
&\quad \cdot ([5]_{11} \cdot [9]_{11}) \cdot ([7]_{11} \cdot [8]_{11}) \cdot [10]_{11}.
\end{align*}
\] (116)

By the definition of the map \(J\), each pair has the form \((\alpha, J(\alpha)) = (\alpha, \alpha^{-1})\) for some \(\alpha \in U_p\), and thus the product of any two different factors paired up with each other

\(^{109}\) The reason why it is precisely these two factors that will not be paired up is the following:

Clearly, the factors \([a]_p\) that cannot be paired up are exactly the factors \([a]_p\) that satisfy \(J([a]_p) = [a]_p\) – i.e., the ones that are their own inverses. So we must prove that a residue class \([a]_p\) with \(a \in \{1, 2, \ldots, p - 1\}\) is its own inverse if and only if \(a\) is either 1 or \(p - 1\). But this follows from the following chain of equivalences:

\[
\begin{align*}
& (\text{the residue class } [a]_p \text{ is its own inverse}) \\
& \iff ([a]_p \cdot [a]_p = [1]_p) \iff ([a^2]_p = [1]_p) \quad \text{(since } [a]_p \cdot [a]_p = [a \cdot a]_p = [a^2]_p) \\
& \iff (a^2 \equiv 1 \pmod{p}) \quad \text{(by Proposition 3.4.5(b))} \\
& \iff (a \equiv 1 \pmod{p} \text{ or } a \equiv -1 \pmod{p}) \\
& \quad \left(\text{indeed, Exercise 2.13.12 yields the implication } (a^2 \equiv 1 \pmod{p}) \implies (a \equiv 1 \pmod{p} \text{ or } a \equiv -1 \pmod{p}); \right. \\
& \left. \text{but the converse implication is easy to check} \right) \\
& \iff ([a]_p = [1]_p \text{ or } [a]_p = [-1]_p) \quad \text{(by Proposition 3.4.5(b))} \\
& \iff ([a]_p = [1]_p \text{ or } [a]_p = [p-1]_p) \quad \text{(since } [-1]_p = [p-1]_p \text{ (because } -1 \equiv p-1 \pmod{p}) \right) \\
& \iff (a = 1 \text{ or } a = p-1) \\
& (\text{since the elements } [0]_p, [1]_p, \ldots, [p-1]_p \text{ are distinct}).
\]
is $[1]_p$ (since $a a^{-1} = [1]_p$). For example, if $p = 11$, then we have

$$\begin{align*}
[1]_p \cdot [2]_p \cdot \cdots \cdot [p-1]_p &= [1]_{11} \cdot ([2]_{11} \cdot [6]_{11}) \cdot ([3]_{11} \cdot [4]_{11}) \cdot ([5]_{11} \cdot [9]_{11}) \cdot ([7]_{11} \cdot [8]_{11}) \cdot [10]_{11} \\
&= [1]_{11} \cdot [10]_{11}.
\end{align*}$$

Thus, any two different factors paired up with each other “neutralize” each other when being multiplied. Hence, the product of all the $p - 1$ factors will reduce to the product of the two factors that have not been paired up, which will be $[1]_p \cdot [p-1]_p = [p-1]_p$. Since this product was $[(p-1)!]_p$, we thus obtain

$$[(p-1)!]_p = [p-1]_p. \quad (117)$$

In other words, $(p-1)! \equiv p - 1 \equiv -1 \mod p$. Hence, Theorem 2.15.7 is proven again.

Once again, if you like your proofs rigorous and formal, you may be wondering how this “pairing up” argument can be formalized. Here is one way to do so: We proceed similarly to how we formalized our first proof of Theorem 2.15.7 above, but with a minor complication. We want to call an element $\alpha$ of $U_p$:

- small if $\alpha < J(\alpha)$;
- medium if $\alpha = J(\alpha)$;
- large if $\alpha > J(\alpha)$.

However, in order for this definition to make sense, we need to define two relations $<$ and $>$ on the set $\mathbb{Z}/p$; otherwise, it is not clear what “$\alpha < J(\alpha)$” and “$\alpha > J(\alpha)$” should mean. Fortunately, this is easy: For example, we can

- consider the bijection $R : \mathbb{Z}/p \to \{0,1,\ldots,p-1\}$ defined in Proposition 3.4.6 (a) (applied to $n = p$);
- define the binary relation $<$ on the set $\mathbb{Z}/p$ by setting
  $$(\alpha < \beta) \iff (R(\alpha) < R(\beta)) \quad \text{for any } \alpha, \beta \in \mathbb{Z}/p;$$
- define the binary relation $>$ on the set $\mathbb{Z}/p$ by setting
  $$(\alpha > \beta) \iff (R(\alpha) > R(\beta)) \quad \text{for any } \alpha, \beta \in \mathbb{Z}/p.$$

The two relations we have just defined have the property that each $\alpha, \beta \in \mathbb{Z}/p$ satisfy either $\alpha < \beta$ or $\alpha = \beta$ or $\alpha > \beta$ but never two or more of these three statements simultaneously (indeed, this follows easily from the fact that $R$ is a bijection). Thus, each element of $U_p$ is either small or medium or large (and there is no overlap between these three classes of elements). Hence, the argument that we used to prove (71) in our first proof of Theorem
2.15.7 can be adapted in order to prove \([ (p - 1)! ]_p = [-1]_p \), except that we have to use residue classes in \( U_p \) instead of elements of \( A \). (We could have just as well used any other bijection from \( \mathbb{Z}/p \) to \( \{1, \ldots, p-1\} \) instead of \( R \) here.) Of course, \([ (p - 1)! ]_p = [-1]_p \) immediately yields \((p - 1)! \equiv -1 \pmod{p}\) (by an application of Proposition 3.4.5 (b)), and thus the proof of Theorem 2.15.7 is complete.

\[ \]

3.6. The Chinese Remainder Theorem as a bijection between residue classes

**Definition 3.6.1.** Let \( n \) be a positive integer. Let \( d \) be a positive divisor of \( n \). Then, define the map

\[ \pi_{n,d} : \mathbb{Z}/n \to \mathbb{Z}/d, \]

\[ [s]_n \mapsto [s]_d. \]

(This is well-defined, according to Proposition 3.4.10.)

See Example 3.4.11 (a) for what this map looks like.

We can now state another version of the “Chinese Remainder Theorem”, which claims the existence of a certain bijection. We have already seen such a version (Theorem 2.16.1), but that one claimed a bijection between two sets of remainders, whereas the following version claims a bijection between two sets of residue classes. Other than that, the two versions are rather similar.

**Theorem 3.6.2.** Let \( m \) and \( n \) be two coprime positive integers. Then, the map

\[ S_{m,n} : \mathbb{Z}/(mn) \to (\mathbb{Z}/m) \times (\mathbb{Z}/n), \]

\[ \alpha \mapsto (\pi_{mn,m}(\alpha), \pi_{mn,n}(\alpha)) \]

is well-defined and is a bijection. It sends each \([s]_{mn}\) (with \( s \in \mathbb{Z} \)) to the pair \(([s]_m, [s]_n)\).

**Example 3.6.3.** (a) Theorem 3.6.2 (applied to \( m = 3 \) and \( n = 2 \)) says that the map

\[ S_{3,2} : \mathbb{Z}/6 \to (\mathbb{Z}/3) \times (\mathbb{Z}/2), \]

\[ \alpha \mapsto (\pi_{6,3}(\alpha), \pi_{6,2}(\alpha)) \]

is a bijection. This map sends

\[
\begin{align*}
[0]_6, & \quad [1]_6, \quad [2]_6, \quad [3]_6, \quad [4]_6, \quad [5]_6 \\
([0]_3, [0]_2), & \quad ([1]_3, [1]_2), \quad ([2]_3, [2]_2), \quad ([3]_3, [3]_2), \quad ([4]_3, [4]_2), \quad ([5]_3, [5]_2),
\end{align*}
\]

\[ \]

\[ \text{viz. by splitting up the product into a product over small elements, a product over medium elements, and a product over large elements} \]

\[ \]

\[ \]
The maps

\[
\begin{align*}
([0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6) & \to ([0]_3, [0]_2), ([1]_3, [1]_2), ([2]_3, [0]_2), ([0]_3, [1]_2), ([1]_3, [0]_2), ([2]_3, [1]_2),
\end{align*}
\]

respectively (since \([2]_2 = [0]_2\) and \([3]_3 = [0]_3\) and \([3]_2 = [1]_2\) and so on). This list of values shows that this map is bijective (since it takes on every possible value in \((\mathbb{Z}/3) \times (\mathbb{Z}/2)\) exactly once). Theorem 3.6.2 says that this holds for arbitrary coprime \(m\) and \(n\).

(b) Let us see how Theorem 3.6.2 fails when \(m\) and \(n\) are not coprime. For example, take \(m = 6\) and \(n = 4\). Then, the map

\[
S_{6,4} : \mathbb{Z}/24 \to (\mathbb{Z}/6) \times (\mathbb{Z}/4),
\]

\[
\alpha \mapsto (\pi_{24,6}(\alpha), \pi_{24,4}(\alpha))
\]

is not a bijection. Indeed, it is neither injective (for example, it sends both \([0]_{24}\) and \([12]_{24}\) to the same pair \(([0]_6, [0]_4)\)) nor surjective (for example, it never takes the value \(([1]_6, [2]_4)\)).

The following proof of Theorem 3.6.2 has the same structure as our proof of Theorem 2.16.1 above, but is shorter since residue classes are easier to deal with than remainders.

Proof of Theorem 3.6.2. The maps \(\pi_{mn,m}\) and \(\pi_{mn,n}\) are well-defined, since \(m\) and \(n\) are positive divisors of \(mn\). Thus, the map

\[
S_{m,n} : \mathbb{Z}/(mn) \to (\mathbb{Z}/m) \times (\mathbb{Z}/n),
\]

\[
\alpha \mapsto (\pi_{mn,m}(\alpha), \pi_{mn,n}(\alpha))
\]

is well-defined. Consider this map \(S_{m,n}\). Clearly, for each \(s \in \mathbb{Z}\), we have

\[
S_{m,n}([s]_{mn}) = \begin{cases} 
\pi_{mn,m}([s]_{mn}) , & \pi_{mn,n}([s]_{mn}) \\
= [s]_m , & = [s]_n \\
\end{cases} \\
\text{(by the definition of } \pi_{mn,m}) \quad \text{(by the definition of } \pi_{mn,n})
\]

(by the definition of \(S_{m,n}\))

\[
= ([s]_m, [s]_n).
\]

(118)

In other words, the map \(S_{m,n}\) sends each \([s]_{mn}\) (with \(s \in \mathbb{Z}\)) to the pair \(([s]_m, [s]_n)\).

It thus remains to prove that \(S_{m,n}\) is a bijection. To that aim, we shall prove that \(S_{m,n}\) is injective and surjective.

[Proof that the map \(S_{m,n}\) is injective: Let \(\alpha, \beta \in \mathbb{Z}/(mn)\) be such that \(S_{m,n}(\alpha) = S_{m,n}(\beta)\). We want to prove \(\alpha = \beta\).]
Write the residue classes \( \alpha \) and \( \beta \) in the forms \( \alpha = [a]_{mn} \) and \( \beta = [b]_{mn} \) for two integers \( a \) and \( b \). (This is possible, because of Proposition 3.4.5(a).) From \( \alpha = [a]_{mn} \), we obtain \( S_{m,n} (\alpha) = S_{m,n} ([a]_{mn}) = ([a]_m, [a]_n) \) (by (118), applied to \( s = a \)). Similarly, \( S_{m,n} (\beta) = ([b]_m, [b]_n) \). Thus, the equality \( S_{m,n} (\alpha) = S_{m,n} (\beta) \) (which we have assumed to hold) rewrites as \( ([a]_m, [a]_n) = ([b]_m, [b]_n) \). In other words, \( [a]_m = [b]_m \) and \( [a]_n = [b]_n \).

Now, we have \( [a]_m = [b]_m \); equivalently, \( a \equiv b \mod m \) (by Proposition 3.4.5(b)); in other words, \( m \mid a - b \). Similarly, \( n \mid a - b \).

Now, we have \( m \perp n \) (since \( m \) and \( n \) are coprime) and \( m \mid a - b \) and \( n \mid a - b \). Hence, Theorem 2.10.7 (applied to \( m \) and \( n \) and \( a - b \) instead of \( a, b \) and \( c \)) yields \( mn \mid a - b \). In other words, \( a \equiv b \mod mn \). In other words, \( [a]_{mn} = [b]_{mn} \) (by Proposition 3.4.5(b)). In other words, \( \alpha = \beta \) (since \( \alpha = [a]_{mn} \) and \( \beta = [b]_{mn} \)).

Now, forget that we fixed \( \alpha \) and \( \beta \). We thus have shown that if \( a, b \in \mathbb{Z}/(mn) \) are such that \( S_{m,n} (\alpha) = S_{m,n} (\beta) \), then \( \alpha = \beta \). In other words, the map \( S_{m,n} \) is injective.

[Proof that the map \( S_{m,n} \) is surjective: Fix \( (\alpha, \beta) \in (\mathbb{Z}/m) \times (\mathbb{Z}/n) \). We want to find a \( \gamma \in \mathbb{Z}/(mn) \) such that \( S_{m,n} (\gamma) = (\alpha, \beta) \).

We have \( \alpha \in \mathbb{Z}/m \). Thus, we can write the residue class \( \alpha \) as \( \alpha = [a]_m \) for some integer \( a \) (because of Proposition 3.4.5(a)). Similarly, we can write the residue class \( \beta \) as \( \beta = [b]_n \) for some integer \( b \). Consider these two integers \( a \) and \( b \). Theorem 2.12.1(a) shows that there exists an integer \( x \in \mathbb{Z} \) such that

\[
(x \equiv a \mod m \text{ and } x \equiv b \mod n).
\]

Consider such an \( x \). We have \( [x]_m = [a]_m \) (since \( x \equiv a \mod m \)) and \( [x]_n = [b]_n \) (since \( x \equiv b \mod n \)). Now, (118) (applied to \( s = x \)) yields

\[
S_{m,n} ([x]_{mn}) = \left( \frac{[x]_m}{[a]_m = \alpha}, \frac{[x]_n}{[b]_n = \beta} \right) = (\alpha, \beta).
\]

Thus, there exists a \( \gamma \in \mathbb{Z}/(mn) \) such that \( S_{m,n} (\gamma) = (\alpha, \beta) \) (namely, \( \gamma = [x]_{mn} \)).

Now, forget that we fixed \( (\alpha, \beta) \). We thus have shown that for any \( (\alpha, \beta) \in (\mathbb{Z}/m) \times (\mathbb{Z}/n) \), there exists a \( \gamma \in \mathbb{Z}/(mn) \) such that \( S_{m,n} (\gamma) = (\alpha, \beta) \). In other words, the map \( S_{m,n} \) is surjective.

We have now proven that the map \( S_{m,n} \) is both injective and surjective. Hence, this map \( S_{m,n} \) is bijective, i.e., is a bijection. This completes the proof of Theorem 3.6.2.

[Remark: As in the proof of Theorem 2.16.1, we could have saved ourselves some of the work by invoking the Pigeonhole Principle. Indeed, our goal was to show that the map \( S_{m,n} \) is bijective. By the Pigeonhole Principle, it suffices to prove that it is injective or that it is surjective, since \( \mathbb{Z}/(mn) \) and \( (\mathbb{Z}/m) \times (\mathbb{Z}/n) \) are finite sets of the same size. But such a proof would be harder to generalize to certain settings that we might later want to generalize Theorem 3.6.2 to.]
We have already proven Theorem 2.14.4 using Theorem 2.16.1. Let us now reprove it using Theorem 3.6.2 instead (by a rather similar argument, but using residue classes instead of remainders):

Second proof of Theorem 2.14.4: For every positive integer $g$, we let $U_g$ be the set of all residue classes $\alpha \in \mathbb{Z}/g$ that have an inverse. Then, $U_n$ is exactly the set that was called $U_n$ in Corollary 3.5.5. Hence, Corollary 3.5.5 (b) yields $\phi(n) = |U_n|$. Similarly, $\phi(m) = |U_m|$ and $\phi(mn) = |U_{mn}|$.

Theorem 3.6.2 says that the map

$$S_{m,n}: \mathbb{Z}/(mn) \to (\mathbb{Z}/m) \times (\mathbb{Z}/n),$$

$$\alpha \mapsto (\pi_{mn,m}(\alpha), \pi_{mn,n}(\alpha))$$

is well-defined and is a bijection. Consider this map $S_{m,n}$. This map $S_{m,n}$ is a bijection, i.e., is injective and surjective. Moreover, the definition of $S_{m,n}$ yields

$$S_{m,n}([1]_{mn}) = \begin{pmatrix} \pi_{mn,m}([1]_{mn}) \\ \pi_{mn,n}([1]_{mn}) \end{pmatrix} = ([1]_m, [1]_n).$$

Let us first prove a trivial fact:

Claim 1: Let $\alpha, \beta \in \mathbb{Z}/(mn)$. Then, $\pi_{mn,m}(\alpha \beta) = \pi_{mn,m}(\alpha) \cdot \pi_{mn,m}(\beta)$.

[Proof of Claim 1: Write the residue classes $\alpha$ and $\beta$ as $\alpha = [a]_{mn}$ and $\beta = [b]_{mn}$ for some integers $a$ and $b$. (This can be done because of Proposition 3.4.5 (a).) Now, $\frac{\alpha}{[a]_{mn}} \cdot \frac{\beta}{[b]_{mn}} = [a]_{mn} \cdot [b]_{mn} = [ab]_{mn}$ (by the definition of multiplication on $\mathbb{Z}/(mn)$). Hence,

$$\pi_{mn,m}\left(\frac{\alpha \beta}{[ab]_{mn}}\right) = \pi_{mn,m}( [ab]_{mn} ) = [ab]_m \quad \text{(by the definition of } \pi_{mn,m}).$$

On the other hand, from $\alpha = [a]_{mn}$, we obtain $\pi_{mn,m}(\alpha) = \pi_{mn,m}( [a]_{mn} ) = [a]_m$ (by the definition of $\pi_{mn,m}$). Similarly, $\pi_{mn,m}(\beta) = [b]_m$. Hence,

$$\pi_{mn,m}(\alpha) \cdot \pi_{mn,m}(\beta) = [a]_m \cdot [b]_m = [ab]_m$$

(by the definition of multiplication on $\mathbb{Z}/m$). Comparing this with $\pi_{mn,m}(\alpha \beta) = [ab]_m$, we obtain $\pi_{mn,m}(\alpha) \cdot \pi_{mn,m}(\beta)$. This proves Claim 1.]

Also, from $U_m \subset \mathbb{Z}/m$ and $U_n \subset \mathbb{Z}/n$, we obtain $U_m \times U_n \subset (\mathbb{Z}/m) \times (\mathbb{Z}/n)$. 


Now, we claim that
\[
S_{m,n} (U_{mn}) \subseteq U_m \times U_n.
\] (119)

[Proof of (119): Let \( \zeta \in S_{m,n} (U_{mn}) \). Thus, \( \zeta = S_{m,n} (\alpha) \) for some \( \alpha \in U_{mn} \). Consider this \( \alpha \).

We have \( \alpha \in U_{mn} \). In other words, \( \alpha \) is a residue class in \( \mathbb{Z}/(mn) \) that has an inverse (since \( U_{mn} \) was defined as the set of all residue classes in \( \mathbb{Z}/(mn) \) that have an inverse). Thus, \( \alpha \) is a residue class in \( \mathbb{Z}/(mn) \) and has an inverse \( \beta \in \mathbb{Z}/(mn) \). Consider this \( \beta \). We know that \( \beta \) is an inverse of \( \alpha \); in other words, \( \alpha \beta = [1]_{mn} \) (by the definition of “inverse”).

Now, Claim 1 yields \( \pi_{mn,m} (\alpha \beta) = \pi_{mn,m} (\alpha) \cdot \pi_{mn,m} (\beta) \), and thus

\[
\pi_{mn,m} (\alpha) \cdot \pi_{mn,m} (\beta) = \pi_{mn,m} \left( \frac{\alpha \beta}{[1]_{mn}} \right) = \pi_{mn,m} ([1]_{mn}) = [1]_m
\]

(by the definition of \( \pi_{mn,m} \)). Thus, \( \pi_{mn,m} (\beta) \) is an inverse of \( \pi_{mn,m} (\alpha) \) in \( \mathbb{Z}/m \) (by the definition of “inverse”). Hence, \( \pi_{mn,m} (\alpha) \) is a residue class in \( \mathbb{Z}/m \) that has an inverse (namely, \( \pi_{mn,m} (\beta) \)). In other words, \( \pi_{mn,m} (\alpha) \in U_m \) (since \( U_m \) was defined as the set of all residue classes in \( \mathbb{Z}/m \) that have an inverse). Similarly, \( \pi_{mn,n} (\alpha) \in U_n \). Now,

\[
\zeta = S_{m,n} (\alpha) = \left( \begin{array}{c}
\pi_{mn,m} (\alpha) \in U_m \\
\pi_{mn,n} (\alpha) \in U_n
\end{array} \right)
\]

(by the definition of \( S_{m,n} \))

\[
\in U_m \times U_n.
\]

Now, forget that we fixed \( \zeta \). We thus have proven that \( \zeta \in U_m \times U_n \) for each \( \zeta \in S_{m,n} (U_{mn}) \). In other words, \( S_{m,n} (U_{mn}) \subseteq U_m \times U_n \). This proves (119).

Next, we claim that
\[
U_m \times U_n \subseteq S_{m,n} (U_{mn}) \ .
\] (120)

[Proof of (120): Let \( \theta \in U_m \times U_n \). We shall prove that \( \theta \in S_{m,n} (U_{mn}) \).

We have \( \theta \in U_m \times U_n \subseteq (\mathbb{Z}/m) \times (\mathbb{Z}/n) = S_{m,n} (\mathbb{Z}/(mn)) \) (since the map \( S_{m,n} \) is a bijection). In other words, there exists some \( \alpha \in \mathbb{Z}/(mn) \) such that \( \theta = S_{m,n} (\alpha) \). Consider this \( \alpha \). The definition of \( S_{m,n} \) yields \( S_{m,n} (\alpha) = (\pi_{mn,m} (\alpha), \pi_{mn,n} (\alpha)) \).

Hence,

\[
(\pi_{mn,m} (\alpha), \pi_{mn,n} (\alpha)) = S_{m,n} (\alpha) = \theta \in U_m \times U_n.
\]

In other words, \( \pi_{mn,m} (\alpha) \in U_m \) and \( \pi_{mn,n} (\alpha) \in U_n \).

We have \( \pi_{mn,m} (\alpha) \in U_m \). In other words, \( \pi_{mn,m} (\alpha) \) is a residue class in \( \mathbb{Z}/m \) that has an inverse (since \( U_m \) was defined as the set of all residue classes in \( \mathbb{Z}/m \) that have an inverse). In other words, \( \pi_{mn,m} (\alpha) \) is a residue class in \( \mathbb{Z}/m \) and has an inverse \( \gamma \in \mathbb{Z}/m \). Likewise, \( \pi_{mn,n} (\alpha) \) is a residue class in \( \mathbb{Z}/n \) and has an inverse \( \delta \in \mathbb{Z}/n \). Consider these \( \gamma \) and \( \delta \).
We have \((\gamma, \delta) \in (\mathbb{Z}/m) \times (\mathbb{Z}/n)\). Since the map \(S_{m,n}\) is surjective, we can thus find a \(\beta \in \mathbb{Z}/(mn)\) such that \(S_{m,n}(\beta) = (\gamma, \delta)\). Consider this \(\beta\). We have

\[(\gamma, \delta) = S_{m,n}(\beta) = (\pi_{mn,m}(\beta), \pi_{mn,n}(\beta))\]

(by the definition of \(S_{m,n}\)). In other words, \(\gamma = \pi_{mn,m}(\beta)\) and \(\delta = \pi_{mn,n}(\beta)\).

Now, we want to prove that \(\beta\) is an inverse of \(\alpha\) (in \(\mathbb{Z}/(mn)\)). Indeed,

\[
\pi_{mn,m}(\alpha \beta) = \pi_{mn,m}(\alpha) \cdot \pi_{mn,m}(\beta) = \gamma
\]

(by Claim 1)

\[
= \pi_{mn,m}(\alpha) \cdot \gamma = [1]_m
\]

(since \(\gamma\) is an inverse of \(\pi_{mn,m}(\alpha)\))

and similarly \(\pi_{mn,n}(\alpha \beta) = [1]_n\). Now, the definition of \(S_{m,n}\) yields

\[
S_{m,n}(\alpha \beta) = \left(\pi_{mn,m}(\alpha \beta), \pi_{mn,n}(\alpha \beta)\right) = ([1]_m, [1]_n).
\]

Comparing this with \(S_{m,n}([1]_{mn}) = ([1]_m, [1]_n)\), we obtain \(S_{m,n}(\alpha \beta) = S_{m,n}([1]_{mn})\).

Since the map \(S_{m,n}\) is injective, we thus conclude that \(\alpha \beta = [1]_{mn}\). In other words, \(\beta\) is an inverse of \(\alpha\) (by the definition of “inverse”). Hence, \(\alpha\) is a residue class in \(\mathbb{Z}/(mn)\) that has an inverse (namely, \(\beta\)). In other words, \(\alpha \in U_{mn}\) (since \(U_{mn}\) was defined as the set of all residue classes in \(\mathbb{Z}/(mn)\) that have an inverse). Now,

\[
\theta = S_{m,n}\left(\underbrace{a}_{\in U_{mn}}\right) \in S_{m,n}(U_{mn}).
\]

Now, forget that we fixed \(\theta\). We thus have shown that \(\theta \in S_{m,n}(U_{mn})\) for each \(\theta \in U_m \times U_n\). In other words, \(U_m \times U_n \subseteq S_{m,n}(U_{mn})\). This proves \([120]\).

Combining \([119]\) with \([120]\), we obtain

\[
S_{m,n}(U_{mn}) = U_m \times U_n. \tag{121}
\]

It is well-known that any two finite sets \(A\) and \(B\) satisfy \(|A \times B| = |A| \cdot |B|\) \([11]\). Applying this to \(A = U_m\) and \(B = U_n\), we obtain

\[
|U_m \times U_n| = |U_m| \cdot |U_n| = \phi(m) \cdot \phi(n).
\]

Note that \(U_{mn}\) is a subset of \(\mathbb{Z}/(mn)\) (by its definition).

\([11]\) This is the so-called product rule in its simplest form (see, e.g., [Loehr11 1.5] or [LeLeMe18 §15.2.1]).
Recall that the map \( S_{m,n} \) is injective. Hence, \( |S_{m,n}(T)| = |T| \) for each subset \( T \) of \( \mathbb{Z}/(mn) \). Applying this to \( T = U_{mn} \), we obtain \( |S_{m,n}(U_{mn})| = |U_{mn}| \). Thus,

\[
|U_{mn}| = \left| \frac{S_{m,n}(U_{mn})}{= U_m \times U_n \ (\text{by } \{121\})} \right| = |U_m \times U_n| = \phi(m) \cdot \phi(n).
\]

Hence, \( \phi(mn) = |U_{mn}| = \phi(m) \cdot \phi(n) \). So Theorem 2.14.4 is proven again. \( \square \)

### 3.7. Substitutivity and chains of congruences revisited

Proposition 3.4.5 (b) can be stated as follows: Given an integer \( n \), two integers \( a \) and \( b \) are congruent to each other modulo \( n \) if and only if their residue classes \( [a]_n \) and \( [b]_n \) are equal. This lets us see congruences modulo \( n \) in a new light (namely, as equalities). In particular, some previous results about congruences now become trivial. For example, we can obtain a very short proof of Proposition 2.4.4 using residue classes:

**Proof of Proposition 2.4.4** We have the chain of congruences \( a_1 \equiv a_2 \equiv \cdots \equiv a_k \mod n \). In other words,

\[
a_i \equiv a_{i+1} \mod n \quad \text{holds for each } i \in \{1,2,\ldots,k-1\}
\]

(by Definition 2.4.3). Thus, for each \( i \in \{1,2,\ldots,k-1\} \), we have \( a_i \equiv a_{i+1} \mod n \) and therefore \( [a_i]_n = [a_{i+1}]_n \) (by Proposition 3.4.5 (b), applied to \( a = a_i \) and \( b = a_{i+1} \)). In other words, we have the chain of equalities \( [a_1]_n = [a_2]_n = \cdots = [a_k]_n \).

From this chain, we immediately obtain \( [a_u]_n = [a_v]_n \) (by Proposition 2.4.2 applied to \( [a_i]_n \) instead of \( a_i \)). Hence, Proposition 3.4.5 (b) (applied to \( a = a_u \) and \( b = a_v \)) shows that \( a_u \equiv a_v \mod n \). This proves Proposition 2.4.4 \( \square \)

We can also prove the Principle of substitutivity of congruences (which we informally stated in Section 2.5 and abbreviated as “PSC”):

**Proof of the PSC (informal).** We have \( x \equiv x' \mod n \). Hence, Proposition 3.4.5 (b) (applied to \( a = x \) and \( b = x' \)) yields \( [x]_n = [x']_n \).

Now, let \( a \) be the expression \( A \), except that each integer appearing in it has been replaced by its residue class modulo \( n \). (For example, if \( A \) is the expression “\( 3 - 2 \cdot 7 + 6 \)”, then \( a \) will be “\( [3]_n - [2]_n \cdot [7]_n + [6]_n \)”.)

Likewise, let \( a' \) be the expression \( A' \), except that each integer appearing in it has been replaced by its residue class modulo \( n \).

The expression \( A' \) differs from \( A \) only in that some appearance of \( x \) in it has been replaced by \( x' \). Thus, the expression \( a' \) differs from \( a \) only in that some appearance

---

\( \text{[12]} \) This follows from the following general principle: If \( f : X \rightarrow Y \) is an injective map between two finite sets \( X \) and \( Y \), then \( |f(T)| = |T| \) for each subset \( T \) of \( X \).
of \([x]_n\) in it has been replaced by \([x']_n\). This replacement does not change the value of the expression, since \([x]_n = [x']_n\). Thus,

\[
\text{the value of } \alpha' = \text{(the value of } \alpha). 
\]

We have defined \(\alpha\) to be the expression \(A\), except that each integer appearing in it has been replaced by its residue class modulo \(n\). Thus, the value of \(\alpha\) is the residue class of the value of \(A\) modulo \(n\). (For example, if \(A\) is “3 – 2 \cdot 7 + 6”, then \(\alpha\) will be “\([3]_n - [2]_n \cdot [7]_n + [6]_n\)”, and thus

\[
\begin{align*}
\text{the value of } \alpha &= [3]_n - [2]_n \cdot [7]_n + [6]_n \\
&= [2]_n \cdot [7]_n \quad \text{(by Definition 3.4.12(c))} \\
&= [3]_n - [2 \cdot 7]_n + [6]_n \quad \text{(by Definition 3.4.12(b))} \\
&= [3 - 2 \cdot 7]_n + [6]_n = [3 - 2 \cdot 7 + 6]_n \quad \text{(by Definition 3.4.12(a))},
\end{align*}
\]

which is precisely the residue class of the value of \(A\) modulo \(n\).) In other words, we have

\[
\text{(the value of } \alpha) = [\text{the value of } A]_n.
\]

Similarly,

\[
\text{(the value of } \alpha') = [\text{the value of } A']_n.
\]

Hence,

\[
[\text{the value of } A']_n = (\text{the value of } \alpha') = (\text{the value of } \alpha) = [\text{the value of } A]_n.
\]

Hence, Proposition 3.4.5(b) (applied to \(a = (\text{the value of } A')\) and \(b = (\text{the value of } A)\)) yields (the value of \(A') \equiv (\text{the value of } A) \mod n\). In other words, the value of the expression \(A'\) is congruent to the value of \(A\) modulo \(n\). This proves the PSC.

3.8. A couple of applications of elementary number theory

In the following short section, we shall see two practical applications of the above number-theoretical studies. The first is a method for encrypting information (the RSA cryptosystem); the second is a trick by which computations with large integers can be split up into more manageable pieces (and distributed across several computers, or parallelized across several cores). We shall be brief, since applications are not a focus of these notes; for further details, see [GalQua17] and the MathOverflow answer [https://mathoverflow.net/a/10022/]. If you are interested in further applications, you may also want to consult the other answers to [https://mathoverflow.net/questions/10014](https://mathoverflow.net/questions/10014) (for a list of uses of the Chinese Remainder Theorem – mostly, but not entirely, inside mathematics), as well as [UspHea39, Appendix to Chapter VII] (for applications of modular arithmetic to calendar computations), and the Wikipedia page on “Universal hashing” (for an application of residue classes modulo primes).
3.8.1. The RSA cryptosystem

Let us present the RSA cryptosystem. This is one of the first modern methods for encrypting data. (The name “RSA” stands for the initials of its three authors: Rivest, Shamir and Adleman.)

This cryptosystem addresses a fairly standard situation: Albert and Julia want to communicate secretly (i.e., Albert wants to send messages to Julia, and Julia to Albert), without having to give each other keys in advance.

Albert wants to send encrypted messages that only Julia can read, and receive encrypted messages that only he can read. The channel of communication may have eavesdroppers. How can they do that?

Setup:

- Albert generates two distinct large and sufficiently random primes $p$ and $q$. (This involves a lot of technicalities like actually finding large primes. See Keith Conrad’s note The Solovay-Strassen test for an algorithm for generating large primes and for a more comprehensive treatment. As to what “large” means, we refer to the Wikipedia article on “key size”.)

- Albert computes the positive integer $m = pq$. This number $m$ (called the modulus) he makes public. (Note that factoring a number into a product of primes is computationally a lot harder than multiplying a bunch of primes. Thus, eavesdroppers will not (likely) be able to reconstruct the primes $p$ and $q$ from their (public) product $m$.)

- Albert computes the positive integer $\ell = (p - 1) (q - 1)$, but keeps this number private.

- Albert randomly picks an $e \in \{2, 3, \ldots, \ell - 1\}$ such that $e \perp \ell$. (Again, we...
omit the details of how to pick such an \( e \) randomly\(^{116}\). This number \( e \) will be called the encryption key, and Albert keeps it private.

- Albert computes a positive modular inverse \( d \) of \( e \) modulo \( \ell \) (that is, a positive integer \( d \) such that \( ed \equiv 1 \mod \ell \)). This number \( d \) exists by Theorem 2.10.8(b); it will be called the decryption key.

- Albert publishes the pair \((e, m)\) as his public key.

- We assume that the messages that Albert and Julia want to send to each other are elements of \( \{0, 1, \ldots, m - 1\} \). This assumption is perfectly reasonable, because these messages originally exist in some digital form (e.g., as bitstrings), and it is easy to translate them from this form into elements of \( \{0, 1, \ldots, m - 1\} \) by some universally agreed rule (e.g., if a bitstring \((a_1, a_2, \ldots, a_k)\) is short enough, then the integer \( a_1 2^{k-1} + a_2 2^{k-2} + \cdots + a_k 2^{k-k} \) will belong to \( \{0, 1, \ldots, m - 1\} \), and thus we can translate this bitstring into this latter integer; otherwise, we break it up into shorter chunks and send those as separate messages).

**Encrypting a message:**

If Julia wants to send a message \( a \in \{0, 1, \ldots, m - 1\} \) to Albert, then she does the following:

- She computes the residue class \( \alpha := [a]_m \in \mathbb{Z}/m \).
- She computes \( \alpha^e \) in \( \mathbb{Z}/m \). (This can be computed quickly using binary exponentiation: If \( \beta \in \mathbb{Z}/m \), then all powers of \( \beta \) can be computed recursively via the formulas \( \beta^{2k} = (\beta^k)^2 \) and \( \beta^{2k+1} = (\beta^k)^2 \beta \).)
- She sends the residue class \( \alpha^e \) (or, more precisely, its unique representative in the set \( \{0, 1, \ldots, m - 1\} \)) to Albert.

**Decrypting a message:**

Albert receives the residue class \( \beta = \alpha^e \) (or, more precisely, a representative thereof, which he can easily turn into the residue class), and recovers the original message \( a \) as follows:

- He sets \( \gamma = \beta^d \). This \( \gamma \) is the same \( \alpha \) that Julia computed, as we shall see below.

\(^{116}\)The rough idea is “pick \( e \in \{2, 3, \ldots, \ell - 1\} \) randomly; check (using the Euclidean algorithm) whether \( e \perp \ell \); if not, then pick another \( e \), and keep repeating this until you hit an \( e \) such that \( e \perp \ell \).” In theory, you could be unlucky and keep picking bad \( e \)’s forever; but in reality, you will soon hit an \( e \) that satisfies \( e \perp \ell \).

\(^{117}\)This unique representative exists by Proposition 3.4.6(b) (and can be computed by picking an arbitrary representative \( b \) first, and then taking its remainder \( b \% m \)).
He recovers the original message \( a \in \{0, 1, \ldots, m - 1\} \) as the unique representative of the residue class \( \gamma = \alpha \) in \( \{0, 1, \ldots, m - 1\} \) (since Julia defined \( \alpha \) as the residue class of \( a \)).

This way, Julia can send a message to Albert that no eavesdropper can read – unless said eavesdropper knows \( d \), or possesses an algorithm hitherto unknown to the world, or has an incredibly fast computer, or Albert’s randomly picked numbers were not random enough\(^{118}\), or one of myriad other practical mistakes has been made. The proper implementation of the RSA cryptosystem, and the real-life considerations needed to prevent “leakage” of sensitive data such as the decryption key \( d \), are a subject in its own right, which we shall not discuss here.

Albert’s method for recovering Julia’s message relies on the following fact (which we shall prove a bit later):

**Lemma 3.8.1.** Let \( p \) and \( q \) be two distinct primes. Let \( N \) be a positive integer such that \( N \equiv 1 \mod (p - 1)(q - 1) \). Then:

- (a) Each \( a \in \mathbb{Z} \) satisfies \( a^N \equiv a \mod pq \).
- (b) Each \( \alpha \in \mathbb{Z}/(pq) \) satisfies \( \alpha^N = \alpha \).

Now, when Albert receives \( \beta = \alpha^e \) from Julia, we have

\[
\beta^d = (\alpha^e)^d = \alpha^{ed}.
\]

But \( d \) was a modular inverse of \( e \) modulo \( \ell \); thus, \( ed \equiv 1 \mod \ell \). Since \( \ell = (p - 1)(q - 1) \), we thus have \( ed \equiv 1 \mod (p - 1)(q - 1) \). Hence, Lemma 3.8.1 (b) (applied to \( N = ed \)) yields \( \alpha^{ed} = \alpha \) (since \( \alpha \in \mathbb{Z}/(pq) \)). Thus,

\[
\beta^d = \alpha^{ed} = \alpha.
\]

Thus, the residue class \( \gamma = \beta^d \) that Albert computes is exactly Julia’s \( \alpha \); hence, Albert correctly recovers the message.

---

\(^{118}\) Computers cannot generate “truly” random numbers (whatever this would even mean!); thus, you have to get by with number generators which try their best at being unpredictable. Lots of creativity has gone into finding ways to come up with numbers that are “as random as possible.” Software alone is, per se, deterministic and thus can at most come up with numbers that “look random” (“pseudorandom number generators”). Nondeterministic input must come from the outside world. This is why certain programs that generate keys ask you to move your mouse around the screen – they are, in fact, using your mouse movements as a source of randomness. Better randomness comes from hardware random number generators, such as Geiger counters or lava lamps.

What happens if your randomly picked prime numbers are not random enough? In the worst case, you never find two distinct primes to begin with. In a more realistic case, your distinct primes will all belong to a small and predictable set, and an eavesdropper can easily find them simply by checking all possibilities. In less obvious cases, different keys you generate for different purposes will occasionally have some primes in common, in which case an easy application of the Chinese Remainder Theorem will allow an eavesdropper to reconstruct them and decrypt your messages. See [https://factorable.net](https://factorable.net) for a study of RSA keys in the wild, which found a lot of common primes.
Proof of Lemma 3.8.1 (sketched). (a) Let \( a \in \mathbb{Z} \). We need to show that \( a^N \equiv a \mod pq \).

In other words, we need to show that \( pq \mid a^N - a \). Since \( p \perp q \), it suffices to prove that \( p \mid a^N - a \) and \( q \mid a^N - a \) (because then, Theorem 2.10.7 will yield \( pq \mid a^N - a \)).

Let us prove that \( p \mid a^N - a \) first. Two cases are possible:

Case 1: We have \( p \mid a \).

Case 2: We have \( p \nmid a \).

Let us first consider Case 1. In this case, we have \( p \mid a \). Thus, \( a \equiv 0 \mod p \). Hence, \( a^N \equiv 0^N = 0 \mod p \) (since \( N \) is positive). Thus, \( \frac{a^N}{p} - a \equiv 0 \mod p \).

\[ 0 \mod p, \text{ so that } p \mid a^N - a. \text{ Thus, we have proven } p \mid a^N - a \text{ in Case 1.} \]

Now, let us consider Case 2. In this case, we have \( p \nmid a \). But we have \( p - 1 \mid (p - 1) (q - 1) \) and \( N \equiv 1 \mod (p - 1) (q - 1) \); hence, Proposition 2.3.4 (e) (applied to \((p - 1) (q - 1), p - 1, N \) and \( 1 \) instead of \( n, m, a \) and \( b \)) yields \( N \equiv 1 \mod p - 1 \). Hence, Exercise 2.15.2 (applied to \( u = N \) and \( v = 1 \)) yields \( a^N \equiv a^1 = a \mod p \). In other words, \( p \mid a^N - a \). Thus, we have proven \( p \mid a^N - a \) in Case 2.

So we have proven \( p \mid a^N - a \) in both Cases. Hence, \( p \mid a^N - a \) always holds.

Similarly, we can prove \( q \mid a^N - a \). This completes our proof of Lemma 3.8.1 (a).

(b) Let \( a \in \mathbb{Z}/(pq) \). Then, we can write \( a \) in the form \( a = [a]_{pq} \) for some \( a \in \mathbb{Z} \) (by Proposition 3.4.3 (a)). Consider this \( a \). From \( a = [a]_{pq} \), we obtain \( a^N = ([a]_{pq})^N = [a^N]_{pq} = [a]_{pq} \) (since Lemma 3.8.1 (a) yields \( a^N \equiv a \mod pq \)). Hence, \( a^N = [a]_{pq} = a \). This proves Lemma 3.8.1 (b).

3.8.2. Computing using the Chinese Remainder Theorem

Next, let us outline a simple yet unexpected application of the Chinese Remainder Theorem.

Assume that you have an expression \( a \) that is made of integers, addition, subtraction and multiplication. For example, say

\[
a = 400 \cdot 405 \cdot 409 \cdot 413 - 401 \cdot 404 \cdot 408 \cdot 414.
\]

(122)

Assume that computing \( a \) directly is too hard, because the intermediate results will be forbiddingly huge numbers, but you know (e.g., from some estimates) that the final result will be a fairly small number. Let’s say (for simplicity) that you know that \( 0 \leq a < 500 \, 000 \).

How can you use this information to compute \( a \) quickly?

One simple trick is to work with residue classes modulo 500 000 instead of working with integer. Thus, instead of computing the number \( a \) directly through the equality (122), we can instead compute its residue class

\[
[a]_{500 \, 000} = [400 \cdot 405 \cdot 409 \cdot 413 - 401 \cdot 404 \cdot 408 \cdot 414]_{500 \, 000} \\
= [400]_{500 \, 000} \cdot [405]_{500 \, 000} \cdot [409]_{500 \, 000} \cdot [413]_{500 \, 000} \\
- [401]_{500 \, 000} \cdot [404]_{500 \, 000} \cdot [408]_{500 \, 000} \cdot [414]_{500 \, 000}
\]
(which is an easier task, because we can always reduce our intermediate results using the fact that every integer \(a\) satisfies \([a]_{500\,000} = [a \mod 500\,000]_{500\,000}\), and then recover \(a\) by observing that \(a\) must be the unique representative of its residue class \([a]_{500\,000}\), that belongs to \(\{0, 1, \ldots, 499\,999\}\) (since \(0 \leq a < 500\,000\)). This is actually how integer arithmetic works in most low-level programming languages; for example, the most popular integer type of the C++ language is “\(\text{int}\)”, which stands not for integers but rather for residue classes modulo \(2^{64}\) (when working on a 64-bit system). (This is where integer overflow comes from.)

Computing \([a]_{500\,000}\) instead of computing \(a\) is already an improvement, but in practice, the “500 000” might actually be a significantly bigger number. Assume, for example, that instead of \(0 \leq a < 500\,000\), you merely know that \(0 \leq a < N\) for some fixed number \(N\) which is small enough that computing in \(\mathbb{Z}/N\) is possible, but large enough that doing the \textbf{whole} computation of \([a]_N\) in \(\mathbb{Z}/N\) is unviable. What can we do then?

One thing we can do is to compute the residue classes \([a]_n\) for several coprime “small” integers \(n\). For example, we can compute \([a]_2\) (by performing the whole computation of \(a\) using residue classes modulo 2 instead of integers) and similarly \([a]_3\) and \([a]_5\) and \([a]_7\) etc.. (We are using prime numbers for \(n\) here, which has certain advantages, but is not strictly necessary; all we need is that the values of \(n\) we are using are coprime\(^{119}\)).

The Chinese Remainder Theorem (in the form of Theorem 3.6.2) shows that if \(m\) and \(n\) are two coprime positive integers, then the map \(S_{m,n}\) from Theorem 3.6.2 (sending each \([s]_m\) to the pair \(([s]_m, [s]_n)\)) is a bijection. In our proof of Theorem 3.6.2 (when proving the surjectivity of \(S_{m,n}\), we gave an explicit way of constructing preimages under this map \(S_{m,n}\) (using Bezout’s theorem, which has a fast algorithm underlying it – the Extended Euclidean algorithm). Thus, we have an explicit way of recovering the residue class \([s]_{mn}\) from the pair \(([s]_m, [s]_n)\) whenever \(s\) is an (unknown) integer (and \(m\) and \(n\) are two coprime positive integers). We shall now refer to this way as the “patching procedure” (since it lets us “patch” two residue classes \([s]_m\) and \([s]_n\) together to a residue class \([s]_{mn}\)).

Now, having computed a bunch of residue classes \([a]_2, [a]_3, [a]_5, [a]_7\) of our unknown integer \(a\) modulo coprime small integers, we can “patch” these classes together:

- From \([a]_2\) and \([a]_3\), we get \([a]_{2,3}\) by the “patching procedure”.
- From \([a]_{2,3}\) and \([a]_5\), we get \([a]_{2,3,5}\) by the “patching procedure”.
- From \([a]_{2,3,5}\) and \([a]_7\), we get \([a]_{2,3,5,7}\) by the “patching procedure”.
- and so on.

We keep “patching” until the product \(2 \cdot 3 \cdot 5 \cdot 7 \cdots\) becomes larger than our \(N\) (which will happen fairly soon, since this product grows super-exponentially with

\(^{119}\)Note that the computations of \([a]_n\) for different values of \(n\) are independent of each other, which comes handy if you have several processors.
the number of “patching” steps). At that point, we have found the residue class 
\([a]_m\) of our unknown integer \(a\) modulo some integer \(m > N\). Since \(0 \leq a < N < m\),
we can thus recover \(a\) itself (as the unique representative of the class \([a]_m\) that lies
in the set \(\{0, 1, \ldots, m - 1\}\)).

This technique has been used a lot (for an example, see [Vogan07, pp. 1031–1033]).

2019-03-01 lecture

3.9. Primitive roots: an introduction

3.9.1. Definition and examples

Let us finally discuss a kind of residue classes that come very useful when they
exist: the \textit{primitive roots} (modulo a positive integer \(n\)). We are not yet able to
ascertain when they exist and when they don’t (this will require some more abstract
algebra); but we can already see some examples of them:

\begin{tcolorbox}
\textbf{Convention 3.9.1.} For the whole Subsection 3.9.1 we fix a positive integer \(n\).
\end{tcolorbox}

\begin{tcolorbox}
\textbf{Definition 3.9.2.} Let \(\alpha \in \mathbb{Z}/n\) be a residue class.
\begin{enumerate}[label=(\alph*)]
\item We say that \(\alpha\) is \textit{invertible} if \(\alpha\) has an inverse.
\item A \textit{power of} \(\alpha\) means a residue class of the form \(\alpha^m\) for some \(m \in \mathbb{N}\).
\item Assume that \(\alpha\) is invertible. Then, \(\alpha\) is said to be a \textit{primitive root modulo} \(n\) if
every invertible residue class \(\beta \in \mathbb{Z}/n\) is a power of \(\alpha\).
\end{enumerate}
\end{tcolorbox}

\begin{tcolorbox}
\textbf{Example 3.9.3.} Let \(n = 9\). The invertible residue classes in \(\mathbb{Z}/9\) are
\([1]_9, [2]_9, [3]_9, [4]_9, [5]_9, [6]_9, [7]_9, [8]_9\).

Clearly, the residue class \([1]_9\) is not a primitive root modulo 9, since all its
powers equal \([1]_9\).

The powers of \([2]_9\) are

\begin{align*}
([2]_9)^0 &= [1]_9, \\
([2]_9)^1 &= [2]_9, \\
([2]_9)^2 &= [4]_9, \\
([2]_9)^3 &= [8]_9, \\
([2]_9)^4 &= [7]_9, \\
([2]_9)^5 &= [5]_9, \\
&\ldots
\end{align*}

Thus, they cover all the six invertible residue classes
\([1]_9, [2]_9, [3]_9, [4]_9, [5]_9, [6]_9, [7]_9, [8]_9\). Hence, \([2]_9\) is a primitive root modulo 9.

It is easy to see that \([5]_9\) also is a primitive root modulo 9, and these two
primitive roots are the only ones.

\end{tcolorbox}
Note that Corollary 3.5.5\(^{(b)}\) shows that there are exactly \(\phi(n)\) invertible residue classes in \(\mathbb{Z}/n\). It is easy to see that any power of an invertible residue class is again invertible.

Euler’s theorem (Theorem 2.15.3) yields that if \(\alpha \in \mathbb{Z}/n\) is an invertible residue class, then \(\alpha^{\phi(n)} = [1]_n\) (because Corollary 3.5.5\(^{(a)}\) shows that \(\alpha\) can be written in the form \(\alpha = [a]_n\) for some integer \(a\) satisfying \(a \perp n\)). Thus, it is easy to see that an invertible residue class \(\alpha \in \mathbb{Z}/n\) has at most \(\phi(n)\) distinct powers. When an invertible residue class \(\alpha \in \mathbb{Z}/n\) has exactly \(\phi(n)\) distinct powers, it is a primitive root (since there are exactly \(\phi(n)\) invertible residue classes in \(\mathbb{Z}/n\)).

**Example 3.9.4.** Let \(n = 8\). The invertible residue classes in \(\mathbb{Z}/8\) are \([1]_8, [3]_8, [5]_8, [7]_8\).

Again, \([1]_8\) is certainly not a primitive root.

The powers of \([3]_8\) are

\[
([3]_8)^0 = [1]_8, \\
([3]_8)^1 = [3]_8, \\
([3]_8)^2 = [9]_8 = [1]_8, \\
\vdots
\]

(so the even powers are \([1]_8\) and the odd powers are \([3]_8\)). So \([3]_8\) is not a primitive root.

The same behavior prevents \([5]_8\) and \([7]_8\) from being primitive roots. Thus, we see that there are no primitive roots modulo 8.

\[\text{Here is a fast way to compute these powers:}\]

\[
([2]_9)^0 = [1]_9, \\
([2]_9)^1 = [2]_9, \\
([2]_9)^2 = \begin{bmatrix} 2^2 \\ =4 \end{bmatrix}_9 = [4]_9, \\
([2]_9)^3 = \begin{bmatrix} 2^3 \\ =8 \end{bmatrix}_9 = [8]_9, \\
([2]_9)^4 = \begin{bmatrix} 2^4 \\ =16 \end{bmatrix}_9 = [16]_9 = [7]_9 \quad (\text{since } 16 \equiv 7 \text{ mod } 9), \\
([2]_9)^5 = [2]_9 \cdot ([2]_9)^4 = [2]_9 \cdot [7]_9 = \begin{bmatrix} 2 \cdot 7 \\ =14 \end{bmatrix}_9 = [14]_9 = [5]_9 \quad (\text{since } 14 \equiv 5 \text{ mod } 9), \\
\vdots
\]
Examples \[3.9.4\] and \[3.9.3\] suggest the following questions: For what \(n\) does a primitive root modulo \(n\) exist, and when it does, how many of them are there? The following theorem – a result proven in 1801 by Gauss – answers both of these questions:

**Theorem 3.9.5.** (a) A primitive root modulo \(n\) exists if and only if \(n\) is

- either 1,
- or a prime \(p\),
- or a power \(p^k\) of an odd prime \(p\) (with \(k\) being a positive integer),
- or 4,
- or \(2p^k\) for an odd prime \(p\) (with \(k\) being a positive integer).

(b) If a primitive root modulo \(n\) exists, then there are precisely \(\phi(\phi(n))\) many of them.

This theorem would be fairly difficult to prove at this point, but will be doable with some abstract algebra (at least in the case \(n = p\)). See [GalQua17, Chapter 4] for a proof.

---

### 4. Complex numbers and Gaussian integers

#### 4.1. Complex numbers

##### 4.1.1. An informal introduction

We now leave (at least for the time being) the study of integers and proceed to consider a much larger “number system”: the complex numbers.

Before we define these numbers rigorously, let me sketch the idea behind their construction. Please suspend your disbelief about the not-quite-kosher reasoning that will follow; we will return to rigorous mathematics in Definition 4.1.1 below.

We know that the number \(-1\) (like any other negative number) has no square root in \(\mathbb{R}\) (because the square of any real number is \(\geq 0\)). But let us audaciously pretend that it does have a square root somewhere else. In other words, let us pretend that there exists a mythical “number” \(i\) such that \(i^2 = -1\). Of course, such a “number” \(i\) will not be a real number, but let us assume (without real justification, for now) that it behaves like a usual number would (to some extent). In particular, let us assume that it can be added, subtracted and multiplied like the numbers that we know and love.

So we have extended the set \(\mathbb{R}\) of real numbers by a new number \(i\). Now, by applying addition, subtraction and multiplication to this new number (and our

\[122\text{Recall: Odd primes are the same as primes } \neq 2.\]
old numbers), we get a bunch of further new numbers – namely, all numbers of the form \(a_0 + a_1i + a_2i^2 + \cdots + a_ki^k\), where \(k \in \mathbb{N}\) and where \(a_0, a_1, \ldots, a_k\) are real numbers. (These can be described as the polynomials in \(i\) with real coefficients.)

However, some of these numbers will be equal; in fact, any number of this form can be reduced to a number of the form \(a + bi\) (with \(a, b \in \mathbb{R}\)), because

\[
\begin{align*}
  i^2 &= -1, \\
  i^3 &= i, \\
  i^4 &= i^2 = -1, \\
  i^5 &= i^3 = -i, \\
  i^6 &= i^4 = -1, \\
  i^7 &= i^5 = i, & \text{etc..}
\end{align*}
\]

For example, the number \(3 + 5i + 9i^2 + 7i^3\) equals \(3 + 5i + 9(-1) + 7(-i) = (3 - 9) + (5 - 7)i = -6 - 2i\).

So all our new numbers have the form \(a + bi\) for two reals \(a\) and \(b\). We call them “complex numbers”. (As we have said, we will give a rigorous definition later.) Since we are assuming that the standard rules of arithmetic still hold for our new numbers, we can easily find formulas for computing the sum, the difference, the product and the quotient of two complex numbers written in the form \(a + bi\):

Namely, for any two complex numbers \(a + bi\) and \(c + di\) (with \(a, b, c, d \in \mathbb{R}\), we have

\[
\begin{align*}
  (a + bi) + (c + di) &= (a + c) + (b + d)i, \quad \text{(123)} \\
  (a + bi) - (c + di) &= (a - c) + (b - d)i; \quad \text{(124)} \\
  (a + bi)(c + di) &= ac + adi + bci + bd, \\
  &= (ac - bd) + (ad + bc)i; \quad \text{(125)} \\
  \frac{a + bi}{c + di} &= \frac{(a + bi)(c - di)}{(c + di)(c - di)} = \frac{ac - adi + bci - bd\overline{i}}{cc - cdi + dci - d\\overline{i}^2} = \frac{ac - adi + bci + bd}{cc - cdi + dci - dd} \quad \text{(if \(c, d\) are not both 0).} \quad \text{(126)}
\end{align*}
\]

(Note that the latter formula is an analogue of the standard procedure for rationalizing denominators that involve square roots:

\[
\frac{a + b\sqrt{2}}{c + d\sqrt{2}} = \frac{(a + b\sqrt{2}) (c - d\sqrt{2})}{(c + d\sqrt{2}) (c - d\sqrt{2})} = \frac{(ac - 2bd) + (bc - ad) \sqrt{2}}{c^2 - 2d^2},
\]

except that the square root that we are trying to exorcise from the denominator is not \(\sqrt{2}\) but \(\sqrt{-1} = i\) now.)

However, not all features of real numbers carry over to complex numbers: Inequalities do not make sense for complex numbers. Indeed, if they would make sense, then we would get a contradiction as follows:

\[\text{Of course, we are assuming that the standard rules -- such as associativity of multiplication -- apply to our "new" numbers.}\]
• If $i \geq 0$, then $i^2 \geq 0$, contradicting $i^2 = -1 < 0$.

• If $i < 0$, then $i^2 = (-i)^2 > 0$ (since $i < 0$ yields $-i > 0$), contradicting $i^2 = -1 < 0$.

Here, we have assumed two things about our relations: First, we have assumed that $i$ is either $\geq 0$ or $< 0$; and second, we have assumed that the square of a non-negative complex number is nonnegative. Sure, we could avoid the contradiction by forfeiting one of these assumptions; but then, the $\geq$ and $<$ relations would not be worth their names any more.

So we appear to be able to extend the four operations $+, -, \cdot, /$ to our weird new numbers, but not the relations $<, \leq, >, \geq$ (at least not in any meaningful way). But how can we be sure that the four operations $+, -, \cdot, /$ don’t already lead to some contradictions?

To answer this question, let us forget our daring postulation of the existence of $i$, and instead give a formal definition of complex numbers:

### 4.1.2. Rigorous definition of the complex numbers

**Definition 4.1.1.** (a) A complex number is defined as a pair $(a, b)$ of two real numbers.

(b) We let $\mathbb{C}$ be the set of all complex numbers.

(c) For each real number $r$, we denote the complex number $(r, 0)$ by $r_\mathbb{C}$.

(d) We let $i$ be the complex number $(0, 1)$. When the notation “$i$” is ambiguous, I will be calling it “$i_\mathbb{C}$” instead. (Some authors call $j$ or $\iota$ or $\sqrt{-1}$.)

(e) We define three binary operations $+, -$ and $\cdot$ on $\mathbb{C}$ by setting

$$
(a, b) + (c, d) = (a + c, b + d),
(a, b) - (c, d) = (a - c, b - d), \quad \text{and}
(a, b) \cdot (c, d) = (ac - bd, ad + bc)
$$

for all $(a, b) \in \mathbb{C}$ and $(c, d) \in \mathbb{C}$.

(f) If $\alpha$ and $\beta$ are two complex numbers, then we write $\alpha \beta$ for $\alpha \cdot \beta$.

(g) If $\alpha$ is a complex number, then the complex number $0_\mathbb{C} - \alpha$ shall be denoted by $-\alpha$.

For example, the definition of the operation $\cdot$ on $\mathbb{C}$ yields

$$
\begin{pmatrix}
\mathbb{C} \\
(0, 1) \\
= (0, 1)
\end{pmatrix}
\begin{pmatrix}
\mathbb{C} \\
(0, 1) \\
= (0, 1)
\end{pmatrix} = (0, 1) (0, 1) = \left(\begin{array}{cc}
0 & -1 \\
1 & 0
\end{array}\right) = \left(\begin{array}{c}
-1 \\
= -1
\end{array}\right) = (0, -1) = (1)_\mathbb{C}.
$$

We will later[124] equate the complex number $(-1)_\mathbb{C}$ with the real number $-1$; thus, this equation will simplify to $ii = -1$. So $i$ “behaves like a square root of $-1$”. But

[124] In Convention 4.1.7
we also have \(( -i)(-i) = (-1)_{\mathbb{C}}\), so \(-i\) fits the same bill. Thus, we didn’t have to postulate the existence of a mythical number \(i\) satisfying \(i^2 = 1\); we simply found such a number in the set \(\mathbb{C}\).

The definitions of the operations \(+, -\) and \(\cdot\) in Definition 4.1.1 are not chosen by accident. We shall later identify each complex number \((a, b)\) with \(a + bi\); then, these definitions will become exactly the equalities (123), (124) and (125) that we derived unrigorously.

We are leaving division of complex numbers undefined so far, because we will later get it more or less for free.

We shall follow the usual “PEMDAS” rules for the order of operations when interpreting expressions involving the operations \(+, -\) and \(\cdot\) on \(\mathbb{C}\). Thus, for example, the expression “\(a + \beta \cdot \gamma\)” shall mean \(a + (\beta \cdot \gamma)\) and not \((a + \beta) \cdot \gamma\).

4.1.3. Rules for \(+, -\) and \(\cdot\)

So we have defined complex numbers as pairs of real numbers, and we have defined three operations on them which we called \(+, -\) and \(\cdot\). But do these operations really deserve these names? Do they still behave as nicely as the corresponding operations on real numbers? Do they, in particular, satisfy the standard rules of arithmetic such as commutativity, associativity and distributivity? The next theorem shows that they indeed do:

Theorem 4.1.2. The following rules for addition, subtraction and multiplication in \(\mathbb{C}\) hold:

(a) We have \(a + \beta = \beta + a\) for any \(a, \beta \in \mathbb{C}\).
(b) We have \(a + (\beta + \gamma) = (a + \beta) + \gamma\) for any \(a, \beta, \gamma \in \mathbb{C}\).
(c) We have \(a + 0_{\mathbb{C}} = 0_{\mathbb{C}} + a = a\) for any \(a \in \mathbb{C}\).
(d) We have \(a \cdot 1_{\mathbb{C}} = 1_{\mathbb{C}} \cdot a = a\) for any \(a \in \mathbb{C}\).
(e) We have \(a \cdot \beta = \beta \cdot a\) for any \(a, \beta \in \mathbb{C}\).
(f) We have \(a \cdot (\beta \cdot \gamma) = (a \cdot \beta) \cdot \gamma\) for any \(a, \beta, \gamma \in \mathbb{C}\).
(g) We have \(a \cdot (\beta + \gamma) = a\beta + a\gamma\) and \((a + \beta) \cdot \gamma = a\gamma + \beta\gamma\) for any \(a, \beta, \gamma \in \mathbb{C}\).
(h) We have \(a \cdot 0_{\mathbb{C}} = 0_{\mathbb{C}} \cdot a = 0_{\mathbb{C}}\) for any \(a \in \mathbb{C}\).
(i) If \(a, \beta, \gamma \in \mathbb{C}\), then we have the equivalence \((a - \beta = \gamma) \iff (a = \beta + \gamma)\).
(j) We have \(- (a + \beta) = (-a) + (-\beta)\) for any \(a, \beta \in \mathbb{C}\).
(k) We have \(-0_{\mathbb{C}} = 0_{\mathbb{C}}\).
(l) We have \(-(-a) = a\) for any \(a \in \mathbb{C}\).
(m) We have \(- (a\beta) = (-a) \beta = a (-\beta)\) for any \(a, \beta \in \mathbb{C}\).
(n) We have \(a - \beta - \gamma = a - (\beta + \gamma)\) for any \(a, \beta, \gamma \in \mathbb{C}\). (Here and in the following, “\(a - \beta - \gamma\)” should be read as “\((a - \beta) - \gamma\)”.)

Proof of Theorem 4.1.2. All parts of this theorem are straightforward. I will only prove the two parts (f) and (i).

(f) Let \(a, \beta, \gamma \in \mathbb{C}\). Thus, \(a\) is a complex number; in other words, \(a\) is a pair of two real numbers (by the definition of complex numbers). Hence, we can write \(a\) in the form \(a = (a, a')\) for two real numbers \(a, a'\). Similarly, we can write \(\beta\) and \(\gamma\) in
the forms $\beta = (b, b')$ and $\gamma = (c, c')$ for four real numbers $b, b', c, c'$. Consider these six real numbers $a, a', b, b', c, c'$. Now, from the equalities $\alpha = (a, a')$, $\beta = (b, b')$ and $\gamma = (c, c')$, we obtain

$$
\alpha \cdot (\beta \cdot \gamma) = (a, a') \cdot \left((b, b') \cdot (c, c')\right)
$$

(by the definition of the operation $\cdot$ on $\mathbb{C}$)

$$
= (a, a') \cdot (bc - b'c', bc' + b'c)
$$

(by the definition of the operation $\cdot$ on $\mathbb{C}$)

$$
= \left(\frac{a (bc - b'c') - a' (bc' + b'c)}{abc - ab'c' - a'b'c' - a'b'c'}\right)
\cdot \left(\frac{a (bc' + b'c) + a' (bc - b'c')}{abc + ab'c' + a'b'c - a'b'c'}\right)
$$

Comparing these two equalities, we see that $\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma$. So Theorem 4.1.2 (f) is proven.

(i) Let $\alpha, \beta, \gamma \in \mathbb{C}$. Thus, $\alpha$ is a complex number; in other words, $\alpha$ is a pair of two real numbers (by the definition of complex numbers). Hence, we can write $\alpha$ in the form $\alpha = (a, a')$ for two real numbers $a, a'$. Similarly, we can write $\beta$ and $\gamma$ in the forms $\beta = (b, b')$ and $\gamma = (c, c')$ for four real numbers $b, b', c, c'$. Consider these six real numbers $a, a', b, b', c, c'$. Now, we have the following chain of logical
equivalences:

$$\begin{pmatrix}
\alpha = (a, a') \\
\beta = (b, b') \\
\gamma = (c, c')
\end{pmatrix}
\iff
\begin{pmatrix}
(a, a') - (b, b') = (c, c') \\
= (a - b, a' - b') \\
(by the definition of the operation \( - \) on \( \mathbb{C} \))
\end{pmatrix}
\iff
\begin{pmatrix}
(a - b = c \quad \text{and} \quad a' - b' = c') \\
= (a = b + c) \quad \iff \quad (a' = b' + c')
\end{pmatrix}
\iff
\begin{pmatrix}
(a, a') = (b + c, b' + c') \\
= (\alpha = \beta + \gamma)
\end{pmatrix}
\iff
\begin{pmatrix}
(a, a') = (b, b') + (c, c') \\
= (b + c, b' + c') \quad (by \ the \ definition \ of \ the \ operation \ + \ on \ \mathbb{C})
\end{pmatrix}
\iff
\begin{pmatrix}
\alpha = \beta + \gamma.
\end{pmatrix}

This proves Theorem 4.1.2 (i).

All the other parts of Theorem 4.1.2 can be proven by direct computations, just as we proved Theorem 4.1.2 (f). \(\Box\)

### 4.1.4. Finite sums and finite products

Recall the concept of a finite sum of real numbers (i.e., a sum of the form \( \sum_{i \in I} a_i \), where \( I \) is a finite set and \( a_i \) is a real number for each \( i \in I \)), and the analogous concept of a finite product of real numbers (i.e., a product of the form \( \prod_{i \in I} a_i \)).

**Definition 4.1.3.** In the same vein, we define the concept of a finite sum of complex numbers (i.e., a sum of the form \( \sum_{i \in I} \alpha_i \), where \( I \) is a finite set and \( \alpha_i \in \mathbb{C} \) for each \( i \in I \)), and the analogous concept of a finite product of complex numbers (i.e., a product of the form \( \prod_{i \in I} \alpha_i \), where \( I \) is a finite set and \( \alpha_i \in \mathbb{C} \) for each \( i \in I \)).

These concepts are well-defined, by Proposition 4.1.4 (a) below.

We will use the usual shorthands for special kinds of finite sums and products. For example, if \( I \) is an interval \( \{p, p + 1, \ldots, q\} \) of integers (and if \( \alpha_i \in \mathbb{C} \) for each \( i \in I \)), then the sum \( \sum_{i \in I} \alpha_i \) will also be denoted by \( \sum_{i=p}^{q} \alpha_i \) or \( \alpha_p + \alpha_{p+1} + \cdots + \alpha_q \).
Likewise for products. Thus, for example, \( \alpha_1 + \alpha_2 + \cdots + \alpha_k \) and \( \alpha_1 \alpha_2 \cdots \alpha_k \) are well-defined whenever \( \alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{C} \).

**Proposition 4.1.4.** (a) Definition 4.1.3 is well-defined. (b) Finite sums (\( \sum_{i \in I} \alpha_i \)) and finite products (\( \prod_{i \in I} \alpha_i \)) of complex numbers \( \alpha_i \in \mathbb{C} \) satisfy the same rules that finite sums and finite products of real numbers satisfy.

**Proof of Proposition 4.1.4.** (a) In [Grinbe15, Theorem 2.118 (a)], it is proven that finite sums of real numbers are well-defined. The same argument (but relying on Theorem 4.1.2 instead of the usual rules of commutativity, associativity etc. for real numbers) shows that finite sums of complex numbers \( \alpha_i \in \mathbb{C} \) are well-defined.

The analogous fact for products is proven in the same way, except that we need to replace 0 by 1 and properties of addition by corresponding properties of multiplication.

(b) The proofs of the properties of finite sums and finite products of elements of \( \mathbb{C} \) are identical to the analogous proofs for real numbers, but (again) rely on Theorem 4.1.2 instead of the usual rules of commutativity, associativity etc. for real numbers.

4.1.5. Embedding \( \mathbb{R} \) into \( \mathbb{C} \)

**Theorem 4.1.5.** For any real numbers \( a \) and \( b \), we have

\[
(a + b)_\mathbb{C} = a_\mathbb{C} + b_\mathbb{C} \quad \text{and} \quad (127)
\]
\[
(a - b)_\mathbb{C} = a_\mathbb{C} - b_\mathbb{C} \quad \text{and} \quad (128)
\]
\[
(ab)_\mathbb{C} = a_\mathbb{C} b_\mathbb{C}. \quad (129)
\]

**Proof of Theorem 4.1.5.** Let \( a \) and \( b \) be two real numbers. Then, the definitions of \( a_\mathbb{C} \), \( b_\mathbb{C} \) and \( (ab)_\mathbb{C} \) yield \( a_\mathbb{C} = (a,0) \) and \( b_\mathbb{C} = (b,0) \) and \( (ab)_\mathbb{C} = (ab,0) \). Now, (129) follows from

\[
\frac{a_\mathbb{C}}{(a,0)} \frac{b_\mathbb{C}}{(b,0)} = (a,0) (b,0) = (a,0) \cdot (b,0) = \left( \frac{ab - 0 \cdot 0}{=ab} a \cdot 0 + 0 \cdot b \right) \quad \text{(by the definition of the operation } \cdot \text{ on } \mathbb{C})
\]

\[
= (ab,0) = (ab)_\mathbb{C}.
\]

Similar straightforward computations prove the equalities (127) and (128). Thus, Theorem 4.1.5 is proven.

**Remark 4.1.6.** If \( a_1, a_2, \ldots, a_k \) are \( k \) reals, then

\[
(a_1)_\mathbb{C} + (a_2)_\mathbb{C} + \cdots + (a_k)_\mathbb{C} = (a_1 + a_2 + \cdots + a_k)_\mathbb{C} \quad \text{and} \quad (127)
\]
\[
(a_1)_\mathbb{C} \cdot (a_2)_\mathbb{C} \cdots \cdots (a_k)_\mathbb{C} = (a_1a_2 \cdots a_k)_\mathbb{C}. \quad (128)
\]
Proof of Remark 4.1.6. This can be proven by a straightforward induction on $k$. □

**Convention 4.1.7.** From now on, for each real number $r$, we shall identify the real number $r$ with the complex number $r_C = (r, 0)$.

Identifying different things is always risky in mathematics; for example, we have seen above why it would be a bad idea to identify residue classes $[a]_n$ of integers modulo a positive integer $n$ with the corresponding remainders $a \% n$ (even though there is a 1-to-1 correspondence between the former and the latter). Nevertheless, the identification made in Convention 4.1.7 is harmless, due to Theorem 4.1.5 and because the map
\[ \mathbb{R} \to \mathbb{C}, \quad r \mapsto r_C \]
is injective (so we are not identifying two different real numbers with one and the same complex numbers).

So we have identified each real number with a complex number. Thus, the complex numbers can be seen as an extension of the real numbers: $\mathbb{R} \subseteq \mathbb{C}$. (Of course, this is not literally true, since formally speaking $r_C$ is a pair while $r$ is a single real number. Nevertheless, we will work as if this was true, and hope that the reader can insert “$\mathbb{C}$” subscripts wherever necessary in order to make our computations literally true.)

When we defined complex numbers as pairs of real numbers in Definition 4.1.1, we were intending that the pair $(a, b)$ would correspond to the complex number $a + bi$ in our previous informal construction of the complex numbers. Convention 4.1.7 makes this actually hold:

**Proposition 4.1.8.** For any $(a, b) \in \mathbb{C}$, we have $(a, b) = a + bi$.

*Proof.* Let $(a, b) \in \mathbb{C}$. Thus, $a$ and $b$ are real numbers. By Convention 4.1.7, we identify these real numbers $a$ and $b$ with the complex numbers $a_C = (a, 0)$ and $b_C = (b, 0)$.

\[ (a, b)_C = (a_C, b_C) = (a, 0) + (b, 0) = (a + b, 0) = (a + bi), \]

To make sure that Convention 4.1.7 cannot spawn such absurdities, we had to prove Theorem 4.1.5.

\[Why does Theorem 4.1.5 matter here? Well, let us assume for a moment that Theorem 4.1.5 was false; specifically, let us assume that there are two real numbers $a$ and $b$ such that $(ab)_C \neq a_C b_C$. Consider these $a$ and $b$. Now, Convention 4.1.7 lets us identify the real numbers $a$, $b$ and $ab$ with the complex numbers $a_C$, $b_C$ and $(ab)_C$. Thus, $ab = (ab)_C \neq a_C b_C = ab$, which is nonsense.\]
\[ b_C = (b, 0), \text{ respectively. Thus, } a = a_C = (a, 0) \text{ and } b = b_C = (b, 0). \text{ Hence,} \]

\[
\begin{align*}
\frac{a}{(a, 0)} + \frac{b}{(b, 0)} \cdot i &= (a, 0) + (b, 0) \cdot (0, 1) \\
&= (a, 0) + \left( b \cdot 0 - 0 \cdot 1, b \cdot 1 + 0 \cdot 0 \right) \\
&= (a, 0) + (0, b) \\
&= (a, b). 
\end{align*}
\]

(by the definition of the operation \( \cdot \) on \( C \))

\[ = (a, b). \]

This proves Proposition 4.1.8.

The next proposition shows that if we multiply a complex number \((b, c)\) with a real number \(a\) (of course, understanding this real number \(a\) as the complex number \(a_C = (a, 0)\)), then the result will simply be \((ab, ac)\) (that is, multiplying a complex number by \(a\) merely multiplies both of its entries by \(a\)):

Proposition 4.1.9. For any \(a \in \mathbb{R}\) and \((b, c) \in \mathbb{C}\), we have \(a \cdot (b, c) = (ab, ac)\). (Here, of course, “\(a \cdot (b, c)\)” means the product \(a_C \cdot (b, c)\).)

Proof. This is straightforward: Let \(a \in \mathbb{R}\) and \((b, c) \in \mathbb{C}\). By Convention 4.1.7, we identify the real number \(a\) with the complex number \(a_C = (a, 0)\). Hence, \(a = a_C = (a, 0)\). Now,

\[
\begin{align*}
\frac{a}{(a, 0)} \cdot (b, c) &= (a, 0) \cdot (b, c) \\
&= \left( \frac{ab - 0c}{ab} \cdot \frac{ac + 0b}{ac} \right) \\
&= (ab, ac). 
\end{align*}
\]

(by the definition of the operation \( \cdot \) on \( C \))

This proves Proposition 4.1.9.

4.1.6. Inverses and division of complex numbers

Definition 4.1.10. A complex number \(a\) is said to be nonzero if and only if it is distinct from the complex number \(0_C = (0, 0)\).

In other words, a complex number \(a\) is nonzero if and only if it is distinct from 0 (since we are identifying the real number 0 with \(0_C\)). Equivalently, a complex number \(\alpha = (a, b)\) is nonzero if and only if \((a, b) \neq (0, 0)\) as pairs (i.e., if and only if at least one of the real numbers \(a\) and \(b\) are nonzero).
We have so far been adding, subtracting and multiplying complex numbers, but never dividing them (except briefly, before we formally defined them). We could define division in the same way as we defined addition, subtraction and multiplication – namely, by an explicit formula for \( \frac{(a, b)}{(c, d)} \) whenever \((c, d)\) is nonzero\(^{126}\). However, it is more instructive to proceed differently, and construct the division from the multiplication that was already defined. After all, if our division is to deserve its name, it should undo multiplication; and this determines it uniquely. We will not define division right away; instead, we start out by defining an inverse of a complex number:

**Definition 4.1.11.** Let \( \alpha \) be a complex number. An inverse of \( \alpha \) means a complex number \( \beta \) such that \( \alpha \beta = 1 \). (Recall that \( 1 = 1_C \) by Convention 4.1.7.)

The complex number 0 has no inverse (because \( 0 \beta = 0 \neq 1 \), no matter what \( \beta \) is). But it turns out that all the other complex numbers have one:

**Theorem 4.1.12.** Let \( \alpha \) be a nonzero complex number. Then, \( \alpha \) has a unique inverse.

**Proof of Theorem 4.1.12** We shall separately prove the existence and the uniqueness of an inverse of \( \alpha \).

**Proof of the existence of the inverse:** Write the complex number \( \alpha \) as \( \alpha = (c, d) \) for two real numbers \( c \) and \( d \). Then, \((c, d) = \alpha \neq 0 \) (since \( \alpha \) is nonzero). Thus, \((c, d) \neq 0 = (0, 0) \). In other words, at least one of the two real numbers \( c \) and \( d \) is nonzero. Hence, at least one of the two real numbers \( c^2 \) and \( d^2 \) is positive\(^{127}\). The other among these two numbers must, of course, be nonnegative\(^{128}\). Hence, \( c^2 + d^2 \) is the sum of a positive real number with a nonnegative real number. Therefore, \( c^2 + d^2 \) itself is positive. Thus, \( c^2 + d^2 \) is a nonzero real number; hence, we can divide by \( c^2 + d^2 \). In particular, we can define a complex number \( \beta \) by

\[
\beta = \left( \frac{c}{c^2 + d^2}, \frac{-d}{c^2 + d^2} \right).
\]

Consider this \( \beta \). Multiplying the equalities \( \alpha = (c, d) \) and \( \beta = \left( \frac{c}{c^2 + d^2}, \frac{-d}{c^2 + d^2} \right) \),

---

\(^{126}\)This formula would be \( \frac{(a, b)}{(c, d)} = \left( \frac{ac + bd}{c^2 + d^2}, \frac{bc - ad}{c^2 + d^2} \right) \).

\(^{127}\)since the square of a nonzero real number is always positive

\(^{128}\)since the square of a real number is always nonnegative
we find
\[
\alpha \beta = (c, d) \left( \frac{c}{c^2 + d^2}, \frac{-d}{c^2 + d^2} \right)
\]
\[
= \left( \frac{c \cdot c - d \cdot d}{c^2 + d^2}, \frac{-d \cdot d + d \cdot c}{c^2 + d^2} \right)
\]
(by the definition of the operation \( \cdot \) on \( \mathbb{C} \))
\[
= (1, 0) = 1_{C}.
\]
Thus, \( \beta \) is an inverse of \( \alpha \) (by the definition of an inverse of \( \alpha \)). Hence, \( \alpha \) has at least one inverse (namely, \( \beta \)).

Proof of the uniqueness of the inverse: We must prove that \( \alpha \) has at most one inverse. This is exactly the statement of Proposition 3.5.4, except that our \( \alpha \) is an element of \( \mathbb{C} \) rather than of \( \mathbb{Z}/n \). But the same argument that we used to prove Proposition 3.5.4 can be applied to \( \alpha \in \mathbb{C} \) instead of \( \alpha \in \mathbb{Z}/n \). Hence, we obtain that \( \alpha \) has at most one inverse.

We have now shown that \( \alpha \) has at least one inverse, and we have shown that \( \alpha \) has at most one inverse. Combining these two results, we conclude that \( \alpha \) has a unique inverse. This proves Theorem 4.1.12.

Defn. 4.1.13. (a) Let \( \beta \) be a nonzero complex number. Theorem 4.1.12 shows that \( \beta \) has a unique inverse. This inverse is called \( \beta^{-1} \), and will be referred to as the inverse of \( \beta \).

(b) Let \( \alpha \) and \( \beta \) be two complex numbers such that \( \beta \neq 0 \). Then, the quotient \( \frac{\alpha}{\beta} \) is defined to be the complex number \( \alpha \cdot \beta^{-1} \). It is sometimes also denoted by \( \frac{\alpha}{\beta} \).

(c) The operation that transforms a pair \( (\alpha, \beta) \) of two complex numbers (with \( \beta \) nonzero) into \( \alpha/\beta \) is called division.

It is easy to see that division undoes multiplication:

Prop. 4.1.14. Let \( \alpha, \beta, \gamma \) be three complex numbers with \( \beta \neq 0 \). Then, we have the equivalence
\[
(\gamma = \frac{\alpha}{\beta}) \iff (\alpha = \beta \gamma).
\]

Proof of Prop. 4.1.14 We have \( \beta \neq 0 \). Thus, \( \beta \) has a well-defined inverse \( \beta^{-1} \). The definition of this inverse yields \( \beta \beta^{-1} = 1 \); now, Theorem 4.1.2 (e) yields \( \beta^{-1} \beta = 1 \).

\textsuperscript{129}Of course, we need to make some obvious modifications, such as replacing every appearance of \( "[1]_n" \) by \( "1" \), and replacing every reference to Theorem 3.4.23 with a reference to Theorem 4.1.2.
\[ \beta \beta^{-1} = 1. \] Also, Theorem 4.1.2(e) yields \( \gamma \beta = \beta \gamma \) and \( \beta^{-1} \alpha = \alpha \beta^{-1} \). The definition of \( \frac{\alpha}{\beta} \) yields \( \frac{\alpha}{\beta} = \alpha \cdot \beta^{-1} = \alpha \beta^{-1} \).

We have to prove the equivalence \( \left( \gamma = \frac{\alpha}{\beta} \right) \iff (\alpha = \beta \gamma) \). Let us prove the “\( \implies \)” and “\( \impliedby \)” directions of this equivalence separately:

\( \implies \): Assume that \( \gamma = \frac{\alpha}{\beta} \). We shall show that \( \alpha = \beta \gamma \).

We have \( \gamma = \frac{\alpha}{\beta} = \alpha \beta^{-1} \). Multiplying both sides of this equality with \( \beta \), we obtain

\[
\gamma \beta = \alpha \beta^{-1} \beta = \alpha \cdot 1 = \alpha.
\]

Hence, \( \alpha = \gamma \beta = \beta \gamma \). This proves the “\( \implies \)” direction of the equivalence \( \left( \gamma = \frac{\alpha}{\beta} \right) \iff (\alpha = \beta \gamma) \).

\( \impliedby \): Assume that \( \alpha = \beta \gamma \). We shall show that \( \gamma = \frac{\alpha}{\beta} \).

We have \( \beta^{-1} \frac{\alpha}{\beta} = \beta^{-1} \beta \gamma = 1 \gamma = \gamma 1 = \gamma \), so that \( \gamma = \beta^{-1} \alpha = \alpha \beta^{-1} = \frac{\alpha}{\beta} \). This proves the “\( \impliedby \)” direction of the equivalence \( \left( \gamma = \frac{\alpha}{\beta} \right) \iff (\alpha = \beta \gamma) \).

Thus, the equivalence \( \left( \gamma = \frac{\alpha}{\beta} \right) \iff (\alpha = \beta \gamma) \) holds (since we have proven both of its directions). That is, we have proven Proposition 4.1.14. \( \square \)

Inverses also have the following properties:

**Proposition 4.1.15.** (a) Let \( \alpha \in \mathbb{C} \) be a complex number that has an inverse (i.e., is nonzero). Then, its inverse \( \alpha^{-1} \) has an inverse as well, and this inverse is \( (\alpha^{-1})^{-1} = \alpha \).

(b) Let \( \alpha, \beta \in \mathbb{C} \) be two complex numbers that have inverses (i.e., are nonzero). Then, their product \( \alpha \beta \) has an inverse as well, and this inverse is \( (\alpha \beta)^{-1} = \alpha^{-1} \beta^{-1} \).

**Proof of Proposition 4.1.15** This proof is completely analogous to the solution to Exercise 3.5.1 (Just replace \( \mathbb{Z}/n \) by \( \mathbb{C} \) ) \( \square \)

**Corollary 4.1.16.** Let \( \alpha, \beta \in \mathbb{C} \) be two nonzero complex numbers. Then, the complex number \( \alpha \beta \) is nonzero as well.

**Proof of Corollary 4.1.16 (sketched).** The complex numbers \( \alpha \) and \( \beta \) are nonzero, and thus have inverses (by Theorem 4.1.12). Hence, Proposition 4.1.15(b) shows that their product \( \alpha \beta \) has an inverse as well. Thus, \( \alpha \beta \neq 0 \) (since 0 has no inverse). This proves Corollary 4.1.16. \( \square \)
4.1.7. Powers of complex numbers

Let us now define powers of complex numbers, where the exponent is a nonnegative integer.

**Definition 4.1.17.** Let \( \alpha \in \mathbb{C} \) and \( n \in \mathbb{N} \). We define a complex number \( \alpha^n \) (called the \( n \)-th power of \( \alpha \)) by setting \( \alpha^n = \underbrace{\alpha \cdot \alpha \cdots \alpha}_{n \text{ times}} \).

Definition 4.1.17 yields
\[
i^2 = ii = (-1)_{\mathbb{C}} = -1.
\]
Moreover, Definition 4.1.17 yields
\[
\alpha^0 = \underbrace{\alpha \cdot \alpha \cdots \alpha}_{0 \text{ times}} = (\text{empty product}) = 1 \quad \text{and} \quad \alpha^1 = \underbrace{\alpha \cdot \alpha \cdots \alpha}_{1 \text{ times}} = \alpha
\]
for each \( \alpha \in \mathbb{C} \).

For another example, Definition 4.1.17 yields
\[
(1 + i)^2 = (1 + i) (1 + i) = 1 + i + i + ii = 1 + i + i + (-1) = i + i = 2i
\]
and
\[
(1 + i)^4 = (1 + i) (1 + i) (1 + i) (1 + i) = 2i \cdot 2i = 4 \cdot ii = 4 (-1) = -4.
\]

We shall use the PEMDAS convention for the order of operations when powers are involved. For example, the expression “\( \alpha \beta^k + \gamma \)” means \( (\alpha (\beta^k)) + \gamma \) rather than (say) \( (\alpha \beta)^k + \gamma \).

Recall that any nonzero complex number \( \alpha \) has an inverse \( \alpha^{-1} \) (by Definition 4.1.13(a)). This allows us to extend our definition of \( \alpha^n \) to negative \( n \) as well:

**Definition 4.1.18.** Let \( \alpha \in \mathbb{C} \) be nonzero. For any negative \( n \in \mathbb{Z} \), we define a complex number \( \alpha^n \) (called the \( n \)-th power of \( \alpha \)) by \( \alpha^n = (\alpha^{-1})^{-n} \). (This is well-defined, since \( (\alpha^{-1})^{-n} \) is already defined by Definition 4.1.17 (because \( n \) is negative and thus \( -n \in \mathbb{N} \)).)

The attentive reader will have noticed that Definition 4.1.18 redefines \( \alpha^{-1} \) when \( \alpha \) is nonzero (indeed, \(-1\) is a negative integer, and thus can be substituted for \( n \) in Definition 4.1.18). Fortunately, this new definition of \( \alpha^{-1} \) does not clash with the original definition (Definition 4.1.13(a)), because if we set \( n = -1 \) in Definition 4.1.18 then we get \( \alpha^{-1} = (\alpha^{-1})^{-1} = \alpha^{-1} \) (where the “\( \alpha^{-1} \)” on the left hand side is
the new meaning defined in Definition 4.1.18, whereas the \(a^{-1}\) on the right hand
side is the old meaning defined in Definition 4.1.13 (a).

If \(a = 0\) and if \(n \in \mathbb{Z}\) is negative, then we leave \(a^n\) undefined.

Powers of complex numbers satisfy the usual rules for exponents:

**Proposition 4.1.19.** (a) We have \(a^{n+1} = a a^n\) for all \(a \in \mathbb{C}\) and \(n \in \mathbb{N}\).

(b) We have \(a^{n+m} = a^n a^m\) for all \(a \in \mathbb{C}\) and \(n, m \in \mathbb{N}\).

(c) We have \((a \beta)^n = a^n \beta^n\) for all \(a, \beta \in \mathbb{C}\) and \(n \in \mathbb{N}\).

(d) We have \((a^n)^m = a^{nm}\) for all \(a \in \mathbb{C}\) and \(n, m \in \mathbb{N}\).

(e) We have \(1^n = 1\) for all \(n \in \mathbb{N}\).

(f) We have \(a^{n+1} = a a^n\) for all nonzero \(a \in \mathbb{C}\) and all \(n \in \mathbb{Z}\).

(g) We have \(a^{-n} = (a^{-1})^n\) for all nonzero \(a \in \mathbb{C}\) and all \(n \in \mathbb{Z}\).

(h) We have \(a^{n+m} = a^n a^m\) for all nonzero \(a \in \mathbb{C}\) and all \(n, m \in \mathbb{Z}\).

(i) We have \((a \beta)^n = a^n \beta^n\) for all nonzero \(a, \beta \in \mathbb{C}\) and all \(n \in \mathbb{Z}\).

(j) We have \(1^n = 1\) for all \(n \in \mathbb{Z}\).

(k) We have \((a^n)^{-1} = a^{-n}\) for all nonzero \(a \in \mathbb{C}\) and all \(n \in \mathbb{Z}\). (In particular, \(a^n\) is nonzero, so that \((a^n)^{-1}\) is well-defined.)

(l) We have \((a^n)^m = a^{nm}\) for all nonzero \(a \in \mathbb{C}\) and all \(n, m \in \mathbb{Z}\). (In particular, \(a^n\) is nonzero, so that \((a^n)^m\) is well-defined for all \(m \in \mathbb{Z}\).)

(m) Complex numbers satisfy the binomial formula: That is, if \(a, \beta \in \mathbb{C}\), then

\[
(a + \beta)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \beta^{n-k}\quad\text{for } n \in \mathbb{N}.
\]

Proposition 4.1.19 can be proven in the same way as the corresponding claims are proven for real (or rational) numbers:

**Exercise 4.1.1.** Prove Proposition 4.1.19

It may be tempting to try to extend Definition 4.1.18 further by defining fractional powers (such as \(a^{1/2}\)). There is a way to do so, but such a definition would be of questionable use and somewhat fragile (in the sense that it would fail to satisfy the rules of exponents). For example, if you wanted to define \((-1)^{1/2}\), then the only reasonable choices would be \(i\) and \(-i\) (since these are the only two complex numbers whose squares are \(-1\)); but with either option, the equality \((a \beta)^{1/2} = a^{1/2} \beta^{1/2}\) would fail if we took \(a = -1\) and \(\beta = -1\). Thus, we prefer to leave powers of the form \(a^n\) for \(n \not\in \mathbb{Z}\) undefined.

### 4.1.8. The Argand diagram

Let us next make a small detour to demonstrate a geometric representation of the complex numbers which, while not strictly necessary for what we intend to do with them, is conducive both to understanding them and to applying them.
Recall that a complex number was defined as a pair of real numbers. On the other hand, a point in the Cartesian plane is also defined as a pair of real numbers (its x-coordinate and its y-coordinate). Thus, it is natural to identify each complex number \((a, b) = a + bi\) with the point \((a, b) \in \mathbb{R}^2\) on the Cartesian plane (i.e., the point with x-coordinate \(a\) and y-coordinate \(b\)). This identification equates each complex number with a unique point in the Cartesian plane, and vice versa:

\[
\text{a + bi} = (a, b)
\]

The picture below shows some of the points (specifically, all the 25 points \((a, b) \in \{-2, -1, 0, 1, 2\}^2\) whose both coordinates are integers between \(-2\) and \(2\)) labeled
with the corresponding complex numbers:

(as well as the unit circle, which passes through the four points labeled 1, i, −1, −i; we will encounter these four points rather often in the following).

This identification of complex numbers with points is called [the Argand diagram or the complex plane](although the latter word has yet another, different meaning). The complex number 0 corresponds to the origin (0, 0) of the plane.

In Definition [4.1.1 (e)], we have introduced three operations on complex numbers; what do they mean geometrically for the corresponding points? The two operations + and − are easiest to understand: They are exactly the usual operations of addition and subtraction for vectors. Thus, if α and β are two complex numbers, then the points labeled by the four complex numbers 0, α, α + β and β
form a parallelogram:

Likewise, the points labeled by the four complex numbers $0, \alpha, \beta$ and $\beta - \alpha$ form a parallelogram. These parallelograms can be degenerate; in particular, the point $-\alpha$ is the reflection of the point $\alpha$ through the origin.\(^\text{130}\)

Multiplication is less evident. The easiest case is multiplying by $i$: If $\alpha$ is a complex number, then the point $i\alpha$ is obtained from the point $\alpha$ by a $90^\circ$ rotation (counterclockwise) around the origin. Thus, the four points $\alpha, i\alpha, -\alpha$ and $-i\alpha$ form

\(^{130}\)We no longer say “the point labeled by $\alpha$”, but simply equate $\alpha$ with that point now.
More generally, if $\beta$ is a complex number, then multiplication by $\beta$ (that is, the map $\mathbb{C} \to \mathbb{C}$, $\alpha \mapsto \alpha \beta$) is a similitude transformation (so it preserves angles and ratios of lengths); more precisely it is a rotation around the origin composed with a homothety from the origin. Combined with the fact that it sends 1 to $\beta$, this uniquely determines it.

This is just the beginning of a rather helpful dictionary between elementary plane geometry and the algebra of complex numbers. See [AndAnd14] for many applications of this point of view, particularly to proving results in plane geometry.

### 4.1.9. Norms and conjugates

Let us now define some further features of complex numbers.

**Definition 4.1.20.** Let $\alpha = (a, b)$ be a complex number.

The *norm* of $\alpha$ is defined to be the real number $a^2 + b^2 \in \mathbb{R}$. This norm is called $N(\alpha)$.

**Proposition 4.1.21.** Let $\alpha$ be a complex number.

(a) We have $N(\alpha) \geq 0$.

(b) We have $N(\alpha) = 0$ if and only if $\alpha = 0$.

**Proof of Proposition 4.1.21.** Write the complex number $\alpha$ in the form $\alpha = (a, b)$ for two real numbers $a$ and $b$. Then, $N(\alpha) = a^2 + b^2$ (by the definition of the norm). But $a^2$ and $b^2$ are squares of real numbers and thus $\geq 0$ (since a square of a real number is always $\geq 0$). Hence, $N(\alpha) = \underbrace{a^2}_{\geq 0} + \underbrace{b^2}_{\geq 0} \geq 0$. This proves Proposition 4.1.21 (a).

(b) We know that $a^2$ and $b^2$ are $\geq 0$. In other words, $a^2$ and $b^2$ are two nonnegative reals. But the sum of two nonnegative reals is $0$ if and only if both of these reals
are 0. Applying this to the two nonnegative reals $a^2$ and $b^2$, we conclude that $a^2 + b^2 = 0$ if and only if both $a^2$ and $b^2$ are 0. In other words, we have the logical equivalence $(a^2 + b^2 = 0) \iff \text{(both } a^2 \text{ and } b^2 \text{ are 0)}$.

Now, we have the following chain of equivalences:

\[
(N(\alpha) = 0) \iff (a^2 + b^2 = 0) \quad \text{(since } N(\alpha) = a^2 + b^2) \iff \text{ both } a^2 \text{ and } b^2 \text{ are 0} \iff (a^2 = 0 \text{ and } b^2 = 0) \iff (a = 0 \text{ and } b = 0) \iff \left(\begin{array}{c} a \\ b \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \end{array}\right) \iff (\alpha = 0_C) \iff (\alpha = 0) .
\]

This proves Proposition 4.1.21 (b). \qed

**Proposition 4.1.22.** Let $a \in \mathbb{R}$. Then, $N(a_C) = a^2$.

**Proof of Proposition 4.1.22.** We have $a_C = (a,0)$ (by the definition of $a_C$). Hence, the definition of the norm yields $N(a_C) = a^2 + 0^2 = a^2$. This proves Proposition 4.1.22. \qed

**Definition 4.1.23.** Let $\alpha = (a,b) \in \mathbb{C}$.

The *conjugate* $\bar{\alpha}$ of $\alpha$ is defined to be the complex number $(a,-b) \in \mathbb{C}$.

From the viewpoint of the Argand diagram, the conjugate $\bar{\alpha}$ of a complex number $\alpha$ is simply the result of reflecting $\alpha$ (or, to be pedantic, the point labeled by $\alpha$) across the x-axis:

Thus, the following is completely self-evident:
**Proposition 4.1.24.** Let \( \alpha \in \mathbb{C} \).

(a) We have \( \alpha = \bar{\alpha} \) if and only if \( \alpha \in \mathbb{R} \). (Keep in mind that we are following Convention 4.1.7, so that the statement “\( \alpha \in \mathbb{R} \)” (for a complex number \( \alpha \)) actually means “\( \alpha = r \mathbb{C} \) for some \( r \in \mathbb{R} \).”)

(b) We always have \( \bar{\alpha} = \alpha \).

Since we don’t want to depend on geometric reasoning, let us nevertheless prove this fact algebraically:

**Proof of Proposition 4.1.24.** Write the complex number \( \alpha \) in the form \( \alpha = (a, b) \) for two real numbers \( a \) and \( b \). Then, \( \bar{\alpha} = (a, -b) \) (by the definition of \( \bar{\alpha} \)). Hence, the definition of \( \bar{\alpha} \) yields \( \bar{\alpha} = \left( a, \frac{-(b)}{b} \right) = (a, b) = \alpha \). This proves Proposition 4.1.24 (b).

(a) \( \iff \): Assume that \( \alpha \in \mathbb{R} \). We must prove that \( \alpha = \bar{\alpha} \).

We have \( \alpha \in \mathbb{R} \). In other words, there exists an \( r \in \mathbb{R} \) such that \( \alpha = r \mathbb{C} \). Consider this \( r \). We have \( \alpha = r \mathbb{C} = (r, 0) \) (by the definition of \( r \mathbb{C} \)). Hence, the definition of \( \bar{\alpha} \) yields \( \bar{\alpha} = \left( r, \frac{-0}{0} \right) = (r, 0) = \alpha \). Thus, \( \alpha = \bar{\alpha} \). This proves the “\( \iff \)” direction of Proposition 4.1.24 (a).

\( \implies \): Assume that \( \alpha = \bar{\alpha} \). We must prove that \( \alpha \in \mathbb{R} \).

We have \( \alpha = \bar{\alpha} \). Thus, \( (a, b) = \alpha = \bar{\alpha} = (a, -b) \). In other words, \( a = a \) and \( b = -b \). From \( b = -b \), we obtain \( 2b = 0 \), thus \( b = 0 \). Hence, \( \alpha = \left( a, \frac{b}{0} \right) = (a, 0) = a \mathbb{C} \) (since \( a \mathbb{C} \) is defined to be \( (a, 0) \)). Thus, there exists an \( r \in \mathbb{R} \) such that \( \alpha = r \mathbb{C} \) (namely, \( r = a \)). In other words, \( \alpha \in \mathbb{R} \). This proves the “\( \implies \)” direction of Proposition 4.1.24 (a). \( \square \)

**Proposition 4.1.25.** Let \( \alpha \in \mathbb{C} \).

(a) We have \( N(\alpha) = \alpha \bar{\alpha} \) (or, more formally: \( (N(\alpha))_{\mathbb{C}} = \alpha \bar{\alpha} \)).

(b) We have \( N(\bar{\alpha}) = N(\alpha) \).

**Proof of Proposition 4.1.25.** Write the complex number \( \alpha \) in the form \( \alpha = (a, b) \) for two real numbers \( a \) and \( b \). Then, \( \bar{\alpha} = (a, -b) \) (by the definition of \( \bar{\alpha} \)) and \( N(\alpha) = a^2 + b^2 \) (by the definition of \( N(\alpha) \)).
(a) Multiplying the equalities $\alpha = (a, b)$ and $\bar{\alpha} = (a, -b)$, we obtain

$$\alpha \bar{\alpha} = (a, b) (a, -b) = \left( \frac{aa - b(-b), a(-b) + ba}{a^2 + b^2} = 0 \right)^{\,\cdot\, 0} = a^2 + b^2 = (N(\alpha), 0) = (N(\alpha))_C$$

(by the definition of the operation $\cdot$ on $\mathbb{C}$)

$$= \left( a^2 + b^2, 0 \right) = (N(\alpha), 0) = (N(\alpha))_C$$

(since $N(\alpha)_C$ is defined to be $(N(\alpha), 0)$). In other words, $(N(\alpha))_C = a\bar{\alpha}$. According to Convention 4.1.7, we are equating the real number $N(\alpha)$ with the complex number $(N(\alpha))_C$; hence, this equality rewrites as $N(\alpha) = a\bar{\alpha}$. This proves Proposition 4.1.25 (a).

(b) Recall that $\bar{\alpha} = (a, -b)$. Thus, the definition of $N(\bar{\alpha})$ yields $N(\bar{\alpha}) = a^2 + (-b)^2 = a^2 + b^2 = N(\alpha)$. This proves Proposition 4.1.25 (b).

2019-03-04 lecture

**Proposition 4.1.26.** Let $\alpha$ and $\beta$ be two complex numbers. Then:

(a) We have $\alpha + \beta = \bar{\alpha} + \bar{\beta}$.

(b) We have $\alpha - \beta = \bar{\alpha} - \bar{\beta}$.

(c) We have $\alpha \cdot \beta = \bar{\alpha} \cdot \bar{\beta}$.

(d) We have $N(\alpha \beta) = N(\alpha) \cdot N(\beta)$.

(e) If $\beta \neq 0$, then $N\left(\frac{\alpha}{\beta}\right) = \frac{N(\alpha)}{N(\beta)}$.

**Proof of Proposition 4.1.26.** Write the complex number $\alpha$ in the form $\alpha = (a, b)$ for two real numbers $a$ and $b$. Then, $\bar{\alpha} = (a, -b)$ (by the definition of $\bar{\alpha}$) and $N(\alpha) = a^2 + b^2$ (by the definition of $N(\alpha)$).

Write the complex number $\beta$ in the form $\beta = (c, d)$ for two real numbers $c$ and $d$. Then, $\bar{\beta} = (c, -d)$ (by the definition of $\bar{\beta}$) and $N(\beta) = c^2 + d^2$ (by the definition of $N(\beta)$).

(c) Multiplying the equalities $\alpha = (a, b)$ and $\beta = (c, d)$, we obtain $\alpha \cdot \beta = (a, b) (c, d) = (ac - bd, ad + bc)$ (by the definition of the operation $\cdot$ on $\mathbb{C}$). Hence, Definition 4.1.23 yields $\bar{\alpha} \cdot \bar{\beta} = (ac - bd, -(ad + bc))$.

On the other hand, multiplying the equalities $\bar{\alpha} = (a, -b)$ and $\bar{\beta} = (c, -d)$ yields

$$\bar{\alpha} \cdot \bar{\beta} = \left( a \underbrace{(-b)}_{=ac-bd} - \underbrace{(-d)}_{=-(ad+bc)} \right) = (ac - bd, - (ad + bc))$$
(by the definition of the operation \( \cdot \) on \( \mathbb{C} \)). Comparing this with \( \alpha \cdot \beta = (ac - bd, -ad + bc) \),
we obtain \( \alpha \cdot \beta = \bar{\alpha} \cdot \bar{\beta} \). This proves Proposition 4.1.26 (c).

Parts (a) and (b) of Proposition 4.1.26 follow by similar (but easier) computations.

(d) Proposition 4.1.25 (a) yields \( N(\alpha) = \alpha \bar{\alpha} \). Similarly, \( N(\beta) = \beta \bar{\beta} \) and \( N(\alpha \beta) = \alpha \beta \bar{\alpha} \bar{\beta} \). Hence,
\[
N(\alpha \beta) = \alpha \beta \bar{\alpha} \bar{\beta} = (\alpha \bar{\alpha} \cdot \bar{\beta}) \cdot (\beta \bar{\beta}) = N(\alpha) \cdot N(\beta).
\]

This proves Proposition 4.1.26 (d).

(e) Assume that \( \beta \neq 0 \). Thus, the quotient \( \frac{\alpha}{\beta} \in \mathbb{C} \) is defined (by Definition 4.1.13 (b)). Proposition 4.1.26 (d) (applied to \( \frac{\alpha}{\beta} \) instead of \( \alpha \)) yields \( N\left(\frac{\alpha}{\beta}\right) = N\left(\alpha \right) \cdot N\left(\beta\right) \cdot N(\beta) \).

In view of \( \frac{\alpha}{\beta} \cdot \beta = \alpha \), this rewrites as
\[
N(\alpha) = N\left(\frac{\alpha}{\beta}\right) \cdot N(\beta).
\]

Also, Proposition 4.1.21 (b) (applied to \( \beta \) instead of \( \alpha \)) yields that we have \( N(\beta) = 0 \)
if and only if \( \beta = 0 \). Hence, \( N(\beta) \neq 0 \) (since \( \beta \neq 0 \)). Thus, we can divide both
sides of the equality (130) by \( N(\beta) \). We thus obtain \( \frac{N(\alpha)}{N(\beta)} = N\left(\frac{\alpha}{\beta}\right) \). Proposition 4.1.26 (e) follows.

The properties of the norm of a complex numbers let us see an old fact in new light: Remember the Brahmagupta–Fibonacci identity (1), which said that
\[
(a^2 + b^2)(c^2 + d^2) = (ad + bc)^2 + (ac - bd)^2
\]
for \( a, b, c, d \in \mathbb{R} \). This identity is equivalent to the identity
\[
N(\alpha) \cdot N(\beta) = N(\alpha \beta)
\]
for the complex numbers \( \alpha = (a, b) = a + bi \) and \( \beta = (c, d) = c + di \). Thus, the identity (1) is just Proposition 4.1.26 (d), restated without the use of complex numbers. This answers the question of how you could have come up with this identity – at least if you know complex numbers. (Brahmagupta must have found it in a different way, since complex numbers were not known to him.)

**Corollary 4.1.27.** Let \( \alpha \in \mathbb{C} \) and \( k \in \mathbb{N} \). Then:

(a) We have \( \alpha^k = \bar{\alpha}^k \).

(b) We have \( N(\alpha^k) = (N(\alpha))^k \).
Proof of Corollary 4.1.27. (a) This follows by induction on $k$, using Proposition 4.1.26 (c) and the fact that $1 = 1$.

(b) This follows by induction on $k$, using Proposition 4.1.26 (d) and the fact that $N(1) = 1.$

Using the norm of a complex number, we can define a notion of absolute value of a complex number:

Definition 4.1.28. Let $\alpha = (a, b)$ be a complex number. The absolute value (or modulus or length) of $\alpha$ is defined to be $\sqrt{N(\alpha)} = \sqrt{a^2 + b^2} \in \mathbb{R}$. (This is well-defined, because Proposition 4.1.21 (a) shows that $N(\alpha) \geq 0$.)

The absolute value of $\alpha$ is denoted by $|\alpha|$. (This notation does not conflict with the classical notation $|a|$ for the absolute value of a real number $a$, because if $a$ is a real number, then Proposition 4.1.22 yields $N(a C) = a^2$ and therefore $\sqrt{N(a C)} = \sqrt{a^2} = |a|$, where “$|a|$” means the classical concept of absolute value of $a$.)

In the Argand diagram, the absolute value $|\alpha|$ of a complex number $\alpha$ is simply the distance of $\alpha$ from the origin. The reason for this is the Pythagorean theorem:

\[ |\alpha| = \sqrt{a^2 + b^2} \]

Good references for the basic properties of complex numbers are [LaNaSc16] and [Swanso18 §3.9–§3.12]. The book [AndAnd14] is a treasure trove of applications and exercises.

4.1.10. Re, Im and the $2 \times 2$-matrix representation

We define some more attributes of a complex number.
Definition 4.1.29. Let $\alpha = (a, b)$ be a complex number (so that $a$ and $b$ are real numbers and $\alpha = a + bi$).

Then, $a$ is called the real part of $\alpha$ and denoted $\text{Re} \, \alpha$ (or $\Re \alpha$).

Also, $b$ is called the imaginary part of $\alpha$ and denoted $\text{Im} \, \alpha$ (or $\Im \alpha$).

The following proposition shows

Proposition 4.1.30. Let $\mathbb{R}^{2 \times 2}$ be the set of all $2 \times 2$-matrices with real entries.

Define a map $\mu : \mathbb{C} \to \mathbb{R}^{2 \times 2}$ by setting

$$
\mu(a, b) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}
$$

for each $(a, b) \in \mathbb{C}$.

(a) We have $\mu(\alpha + \beta) = \mu(\alpha) + \mu(\beta)$ for all $\alpha, \beta \in \mathbb{C}$.

(b) We have $\mu(\alpha - \beta) = \mu(\alpha) - \mu(\beta)$ for all $\alpha, \beta \in \mathbb{C}$.

(c) We have $\mu(\alpha \cdot \beta) = \mu(\alpha) \cdot \mu(\beta)$ for all $\alpha, \beta \in \mathbb{C}$.

(d) The map $\mu$ is injective.

Proof of Proposition 4.1.30 (c) This is a straightforward computation: Let $\alpha, \beta \in \mathbb{C}$.

Write the complex number $\alpha$ in the form $\alpha = (a, b)$ for two real numbers $a$ and $b$.

Write the complex number $\beta$ in the form $\beta = (c, d)$ for two real numbers $c$ and $d$.

Multiplying the equalities $\alpha = (a, b)$ and $\beta = (c, d)$, we obtain

$$
\alpha \cdot \beta = (a, b) \cdot (c, d) = (ac - bd, ad + bc)
$$

(by the definition of $\cdot$). Hence,

$$
\mu(\alpha \cdot \beta) = \mu(ac - bd, ad + bc) = \begin{pmatrix} ac - bd & ad + bc \\ -(ad + bc) & ac - bd \end{pmatrix}
$$

(by the definition of $\mu$). On the other hand, from $\alpha = (a, b)$, we obtain

$$
\mu(\alpha) = \mu(a, b) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}
$$

(by the definition of $\mu$),

and similarly we can find

$$
\mu(\beta) = \begin{pmatrix} c & d \\ -d & c \end{pmatrix}.
$$

Multiplying these two equalities together, we find

$$
\mu(\alpha) \cdot \mu(\beta) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \cdot \begin{pmatrix} c & d \\ -d & c \end{pmatrix} = \begin{pmatrix} ac + b(-d) & ad + bc \\ (-b)c + a(-d) & (-b)d + ac \end{pmatrix}
$$

$$
= \begin{pmatrix} ac - bd & ad + bc \\ -(ad + bc) & ac - bd \end{pmatrix}.
$$
Comparing this with \((131)\), we find \(\mu (\alpha \cdot \beta) = \mu (\alpha) \cdot \mu (\beta)\). This proves Proposition 4.1.30 (c).

Similar (but much simpler) computations prove parts (a) and (b) of Proposition 4.1.30.

(d) We need to show that a complex number \(\alpha\) can always be recovered from its image \(\mu (\alpha)\).

But this is easy: If \(\alpha = (a, b)\) is a complex number, then \(\mu (\alpha) = \mu (a, b) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}\) (by the definition of \(\mu\)), and therefore we can recover \(a\) and \(b\) from \(\mu (\alpha)\) (namely, \(a\) and \(b\) are the two entries of the first row of the matrix \(\mu (\alpha)\)). Hence, we can recover \(\alpha\) from \(\mu (\alpha)\). This shows that the map \(\mu\) is injective; this proves Proposition 4.1.30 (d).

Proposition 4.1.30 really says that (instead of regarding complex numbers as pairs of real numbers) we can regard complex numbers as a specific kind of \(2 \times 2\)-matrices with real entries (by identifying each complex number \(\alpha\) with the matrix \(\mu (\alpha)\)). This viewpoint has the advantage that multiplication of complex numbers becomes a particular case of matrix multiplication. (We could have saved ourselves the trouble of proving the associativity of multiplication for complex numbers if we had taken this viewpoint.)

4.1.11. The fundamental theorem of algebra

Finally, let me mention without proof the so-called Fundamental Theorem of Algebra:

**Theorem 4.1.31.** Let \(p (x)\) be a polynomial of degree \(n\) with complex coefficients. Then, there exist complex numbers \(\alpha_1, \alpha_2, \ldots, \alpha_n\) and \(\beta\) such that

\[
p (x) = \beta (x - \alpha_1) (x - \alpha_2) \cdots (x - \alpha_n).
\]

In other words, any polynomial with complex coefficients can be factored into linear factors. This is in contrast to real numbers, where polynomials can at best be factored into linear and quadratic factors. (For example, the polynomial \(x^2 + 1\) cannot be factored further over the real numbers, but factors as \((x + i) (x - i)\) over the complex numbers.)

The Fundamental Theorem of Algebra is not actually a theorem of algebra. It relies heavily on the concepts of real and complex numbers. So it is actually a theorem of analysis. For a proof, see [LaNaSc16 Theorem 3.2.2].

4.2. Gaussian integers

Inside the set \(\mathbb{C}\) of all complex numbers (an uncountable set) lies a much smaller (countable) set of numbers, which are much closer to integers than even to real numbers. We shall study them partly for their own sake, partly as an instructive
example of what will later call a commutative ring, and partly in order to answer
the questions from Section 1.4 (although complex numbers were never mentioned
in that section).
We shall follow Keith Conrad’s notes [ConradG] for most of this section (but at
the end we will go a bit further in order to answer Question 1.4.2 (b)).

4.2.1. Definitions and basics

We shall now define the Gaussian integers: a middle ground between integers and
complex numbers.

Definition 4.2.1. A Gaussian integer is a complex number \((a, b)\) with \(a, b \in \mathbb{Z}\).

For example, \(3 + 5i = (3, 5)\) and \(3 - 7i = (3, -7)\) are Gaussian integers. So are
\(0 = (0, 0)\), \(1 = (1, 0)\) and \(i = (0, 1)\). Every integer is a Gaussian integer. But
\(\frac{1}{2} + 3i = \left(\frac{1}{2}, 3\right)\) and \(\sqrt{2} + 4i = \left(\sqrt{2}, 4\right)\) are not Gaussian integers.

Remark 4.2.2. In Definition 4.2.1, we have defined Gaussian integers using com-
plex numbers. This can be viewed as somewhat of an overkill, as the notion
of complex numbers depends on the notion of real numbers, which are mostly
useless for Gaussian integers. Thus, one might ask for a different definition of
Gaussian integers – one which relies only on integers and not on real numbers.

Such a definition is easy to make: Just replace every appearance of real num-
ers in Definition 4.1.1 by integers! Thus, define the Gaussian integers as pairs
of two integers; let \(C_Z\) be the set of these pairs; denote the Gaussian integer
\((r, 0)\) by \(r_{C}\) whenever \(r\) is an integer; define the operations \(+, -\) and \(\cdot\) on the
set \(C_Z\) by the same formulas as in Definition 4.1.1 (e); likewise, adapt the rest
of Definition 4.1.1 to integers. Most of what we have done in Section 4.1 can
be straightforwardly adapted to this notion of Gaussian integers (by making
the obvious changes – i.e., mostly, replacing real numbers by integers); the main
exceptions are the following:

- Not every nonzero Gaussian integer has an inverse (in the set of Gaussian
  integers). (In fact, as we will soon see, the only Gaussian integers that
  have inverses are \(1, i, -1, -i\).) Thus, division and negative powers of Gaus-
  sian integers are usually not defined (without leaving the set of Gaussian
  integers).
- The absolute value \(|\alpha|\) of a Gaussian integer \(\alpha\) will usually not be an integer
  (since it is defined as a square root).

This alternative definition of Gaussian integers is equivalent to Definition 4.2.1;
we are using the latter mainly because it is shorter.
Likewise, we could have defined “Gaussian rationals” by adapting Definition
4.1.1 to rational (instead of real) numbers. Unlike the Gaussian integers, these
“Gaussian rationals” do have inverses (when they are nonzero), and thus division and negative powers are well-defined for them.

**Definition 4.2.3.** We let \( \mathbb{Z}[i] \) be the set of all Gaussian integers.

Elementary number theory concerns itself with integers (mostly). Our goal in this section is to replicate as much as we can of this theory in the setting of Gaussian integers, and then see how it can be applied back to answer some questions about the usual integers.

We will try to use Greek letters for Gaussian integers and Roman letters for integers.

**Proposition 4.2.4.** (a) Let \( \alpha \) and \( \beta \) be two Gaussian integers. Then, \( \alpha + \beta, \alpha - \beta \) and \( \alpha \cdot \beta \) are Gaussian integers.

(b) Sums and products of finitely many Gaussian integers are Gaussian integers.

**Proposition 4.2.5.** Let \( \alpha \) be a Gaussian integer. Then, \( \bar{\alpha} \) is a Gaussian integer.

Proof of Proposition 4.2.5. Write the complex number \( \alpha \) as \( \alpha = (a, b) \) with \( a, b \in \mathbb{R} \). Then, \( a, b \in \mathbb{Z} \) (since \( \alpha \) is a Gaussian integer). Hence, \( a, -b \in \mathbb{Z} \). Now, the definition of \( \bar{\alpha} \) yields \( \bar{\alpha} = (a, -b) \). Hence, \( \bar{\alpha} \) is a Gaussian integer (since \( a, -b \in \mathbb{Z} \)). This proves Proposition 4.2.5.

**Proposition 4.2.6.** Let \( \alpha \in \mathbb{Z}[i] \). Then, \( N(\alpha) \in \mathbb{N} \).

Proof of Proposition 4.2.6. Write the complex number \( \alpha \) as \( \alpha = (a, b) \) with \( a, b \in \mathbb{R} \). Then, \( a, b \in \mathbb{Z} \) (since \( \alpha \) is a Gaussian integer). In other words, \( a \) and \( b \) are integers. Hence, \( a^2 \) and \( b^2 \) are nonnegative integers (since the square of an integer is always a nonnegative integer). In other words, \( a^2, b^2 \in \mathbb{N} \). Hence, \( a^2 + b^2 \in \mathbb{N} \). But the definition of \( N(\alpha) \) yields \( N(\alpha) = a^2 + b^2 \in \mathbb{N} \). This proves Proposition 4.2.6.
Definition 4.2.7. (a) A Gaussian integer \( \alpha \in \mathbb{Z}[i] \) is said to be invertible in \( \mathbb{Z}[i] \) if it has an inverse in \( \mathbb{Z}[i] \).

A unit will mean a Gaussian integer that is invertible in \( \mathbb{Z}[i] \).

(b) We define a relation \( \sim \) on \( \mathbb{Z}[i] \) by

\[
(\alpha \sim \beta) \iff (\alpha = \gamma \beta \text{ for some unit } \gamma \in \mathbb{Z}[i]).
\]

This relation will be called unit-equivalence (or equality up to unit).

For comparison, let us consider analogous concepts for integers instead of Gaussian integers. The units of \( \mathbb{Z} \) (that is, the integers that are invertible in \( \mathbb{Z} \)) are 1 and \(-1\). So if we defined a relation \( \sim \) on \( \mathbb{Z} \) in the same way as we defined the relation \( \sim \) on \( \mathbb{Z}[i] \) (but requiring \( \gamma \in \mathbb{Z} \) instead of \( \gamma \in \mathbb{Z}[i] \)), then this relation would just be given by

\[
\left( a \sim_{\mathbb{Z}} b \right) \iff (a = cb \text{ for some } c \in \{1, -1\})
\]

\[
\iff (a = b \text{ or } a = -b) \iff (|a| = |b|). \quad (132)
\]

So the relation \( \sim \) is not very exciting; it is simply “equality up to sign”\(^{131} \)

But the equality \( \sim \) on \( \mathbb{Z}[i] \) cannot be described as simply as this: It is easy to find two Gaussian integers \( \alpha \) and \( \beta \) such that \( |\alpha| = |\beta| \) holds but \( \alpha \sim \beta \) does not (for example, \( \alpha = 16 + 63i \) and \( \beta = 33 + 56i \) both have absolute value 65 but are not unit-equivalent).

Proposition 4.2.8. The relation \( \sim \) on \( \mathbb{Z}[i] \) is an equivalence relation.

Proof of Proposition 4.2.8. Observe the following:

- The relation \( \sim \) is reflexive.
  
  [Proof: Let \( \alpha \in \mathbb{Z}[i] \). Then, \( \alpha = 1\alpha \). But 1 is a unit (since \( 1^{-1} = 1 \in \mathbb{Z}[i] \)). Hence, \( \alpha = \gamma \alpha \) for some unit \( \gamma \in \mathbb{Z}[i] \) (namely, \( \gamma = 1 \)). In other words, \( \alpha \sim \alpha \) (by the definition of the relation \( \sim \)).

  Now, forget that we fixed \( \alpha \). We thus have shown that every \( \alpha \in \mathbb{Z}[i] \) satisfies \( \alpha \sim \alpha \). In other words, the relation \( \sim \) is reflexive.]

- The relation \( \sim \) is symmetric.
  
  [Proof: Let \( \alpha, \beta \in \mathbb{Z}[i] \) be such that \( \alpha \sim \beta \). We shall prove that \( \beta \sim \alpha \).

  We have \( \alpha \sim \beta \). In other words, \( \alpha = \delta \beta \) for some unit \( \delta \in \mathbb{Z}[i] \) (by the definition of the relation \( \sim \)). Consider this \( \delta \). Note that \( \delta \) is a unit; in other words, \( \delta \) is a Gaussian

\(^{131}\)In other words, it is precisely the relation \( \equiv_{\text{abs}} \), where \( \text{abs} : \mathbb{Z} \to \mathbb{N} \) is the map sending each integer \( n \) to its absolute value \( |n| \). (See Example 3.2.7 for how this relation \( \equiv_{\text{abs}} \) is defined.)
integer that has an inverse in \( Z[i] \). Thus, \( \delta^{-1} \) is well-defined (since \( \delta \) has an inverse), and \( \delta^{-1} \in Z[i] \) (since \( \delta \) has an inverse in \( Z[i] \)). Now, \( \delta^{-1} \) is a Gaussian integer (since \( \delta^{-1} \in Z[i] \)) and itself has an inverse in \( Z[i] \) (since its inverse is \( (\delta^{-1})^{-1} = \delta \in Z[i] \)). In other words, \( \delta^{-1} \) is a unit. Furthermore, dividing both sides of the equality \( \alpha = \delta \beta \) by \( \delta \), we find \( \delta^{-1} \alpha = \beta \), so that \( \beta = \delta^{-1} \alpha \). Thus, \( \beta = \gamma \alpha \) for some unit \( \gamma \in Z[i] \) (namely, for \( \gamma = \delta^{-1} \)). In other words, \( \beta \sim \alpha \) (by the definition of the relation \( \sim \)).

Now, forget that we fixed \( \alpha \) and \( \beta \). We thus have shown that every \( \alpha, \beta \in Z[i] \) satisfying \( \alpha \sim \beta \) satisfy \( \beta \sim \alpha \). In other words, the relation \( \sim \) is symmetric.

• The relation \( \sim \) is transitive.

[Proof: Let \( \alpha, \beta, \gamma \in Z[i] \) be such that \( \alpha \sim \beta \) and \( \beta \sim \gamma \). We shall prove that \( \alpha \sim \gamma \).

From \( \alpha \sim \beta \), we conclude that \( \alpha = \delta \beta \) for some unit \( \delta \in Z[i] \) (by the definition of the relation \( \sim \)). From \( \beta \sim \gamma \), we conclude that \( \beta = \epsilon \gamma \) for some unit \( \epsilon \in Z[i] \) (by the definition of the relation \( \sim \)). Consider these two units \( \delta \) and \( \epsilon \). Both \( \delta \) and \( \epsilon \) are units, and thus have inverses in \( Z[i] \) (by the definition of “unit”). In other words, they have inverses, and these inverses \( \delta^{-1} \) and \( \epsilon^{-1} \) belong to \( Z[i] \). Now, Proposition \ref{prop:4.1.15}(b) (applied to \( \delta \) and \( \epsilon \) instead of \( \alpha \) and \( \beta \)) yields that the product \( \delta \epsilon \) has an inverse as well, and this inverse is \( (\delta \epsilon)^{-1} = \delta^{-1} \epsilon^{-1} \). Hence, \((\delta \epsilon)^{-1} = \delta^{-1} \epsilon^{-1} \in Z[i] \) (since both \( \delta^{-1} \) and \( \epsilon^{-1} \) belong to \( Z[i] \)). Thus, \( \delta \epsilon \) is a Gaussian integer (since \( \delta \) and \( \epsilon \) are Gaussian integers) that has an inverse in \( Z[i] \) (since \( (\delta \epsilon)^{-1} \in Z[i] \)). In other words, \( \delta \epsilon \) is a unit (by the definition of a “unit”). This unit \( \delta \epsilon \) satisfies \( \alpha = (\delta \epsilon) \gamma \) (since \( \alpha = \delta \beta = \delta \epsilon \gamma = (\delta \epsilon) \gamma \)). Hence, \( \alpha = \rho \gamma \) for some unit \( \rho \in Z[i] \) (namely, for \( \rho = \delta \epsilon \)). In other words, \( \alpha \sim \gamma \) (by the definition of the relation \( \sim \)).

Now, forget that we fixed \( \alpha, \beta, \gamma \). We thus have shown that every \( \alpha, \beta, \gamma \in Z[i] \) satisfying \( \alpha \sim \beta \) and \( \beta \sim \gamma \) satisfy \( \alpha \sim \gamma \). In other words, the relation \( \sim \) is transitive.]

We have now proven that the relation \( \sim \) is reflexive, symmetric and transitive. In other words, \( \sim \) is an equivalence relation (by the definition of “equivalence relation”). This proves Proposition \ref{prop:4.2.8} \( \square \)

Proposition \ref{prop:4.2.9}. Let \( \alpha \) be a Gaussian integer.

(a) We have \( N(\alpha) = 0 \) if and only if \( \alpha = 0 \).

(b) We have \( N(\alpha) = 1 \) if and only if \( \alpha \) is a unit.

Proof of Proposition \ref{prop:4.2.9} (a) This is a particular case of Proposition \ref{prop:4.1.21}(b).

(b) \( \iff \): Assume that \( N(\alpha) = 1 \). We must prove that \( \alpha \) is a unit.

Proposition \ref{prop:4.1.25} \( \alpha \) yields \( N(\alpha) = a \bar{\alpha} \). Hence, \( a \bar{\alpha} = N(\alpha) = 1 \).

But Proposition \ref{prop:4.2.5} shows that \( \bar{\alpha} \) is a Gaussian integer. In other words, \( \bar{\alpha} \in Z[i] \).

This Gaussian integer \( \bar{\alpha} \in Z[i] \) is an inverse of \( \alpha \) (since \( a \bar{\alpha} = 1 \)). Thus, \( \alpha \) has an inverse in \( Z[i] \) (namely, \( \bar{\alpha} \)). In other words, \( \alpha \) is a unit (by the definition of “unit”).

This proves the “\( \iff \)” direction of Proposition \ref{prop:4.2.9}(b).

\( \Leftarrow \): Assume that \( \alpha \) is a unit. We must prove that \( N(\alpha) = 1 \).

We know that \( \alpha \) is a unit. In other words, \( \alpha \) is invertible in \( Z[i] \). In other words, \( \alpha \) has an inverse \( \alpha^{-1} \in Z[i] \).
This inverse \( a^{-1} \) satisfies \( aa^{-1} = 1 = 1_C = (1, 0) \), so that
\[
N\left(aa^{-1}\right) = N\left((1, 0)\right) = 1^2 + 0^2 \quad \text{(by the definition of } N\left((1, 0)\right)) = 1.
\]

But Proposition 4.1.26 (d) (applied to \( \beta = a^{-1} \)) yields \( N\left(aa^{-1}\right) = N\left(a\right) \cdot N\left(a^{-1}\right) \). Hence, \( N\left(a\right) \cdot N\left(a^{-1}\right) = N\left(aa^{-1}\right) = 1 \). But Proposition 4.2.6 yields \( N\left(a\right) \in \mathbb{N} \). The same argument (applied to \( a^{-1} \) instead of \( a \)) yields \( N\left(a^{-1}\right) = N\left(a\right) \in \mathbb{N} \) (since \( a^{-1} \in \mathbb{Z}[i] \)). Hence, the equality \( N\left(a\right) \cdot N\left(a^{-1}\right) = 1 \) entails that \( N\left(a\right) \mid 1 \). Consequently, \( N\left(a\right) = 1 \) (since \( N\left(a\right) \in \mathbb{N} \)). This proves the "\( \Longleftrightarrow \)" direction of Proposition 4.2.9 (b).

**Proposition 4.2.10.** The units (in \( \mathbb{Z}[i] \)) are \( 1, -1, i, -i \).

**Proof of Proposition 4.2.10.** Each of the four Gaussian integers \( 1, -1, i, -i \) is a unit \(^{132}\).

It remains to prove that there are no other units.

So let \( a \in \mathbb{Z}[i] \) be a unit. We shall prove that \( a \) is either \( 1 \) or \( -1 \) or \( i \) or \( -i \).

Proposition 4.2.9 (b) shows that we have \( N\left(a\right) = 1 \) if and only if \( a \) is a unit. Hence, we have \( N\left(a\right) = 1 \) (since \( a \) is a unit).

Let us write the complex number \( a \) as \( a = (a, b) \). Then, \( a, b \in \mathbb{Z} \) (since \( a \in \mathbb{Z}[i] \)). Furthermore, the definition of \( N\left(a\right) \) yields \( N\left(a\right) = a^2 + b^2 \), so that \( a^2 + b^2 = N\left(a\right) = 1 \). If both integers \( a \) and \( b \) were nonzero, then both their squares \( a^2 \) and \( b^2 \) would be \( \geq 1 \) (because the square of any nonzero integer is \( \geq 1 \)), and thus the sum of these squares would be \( a^2 + b^2 \geq 1 + 1 > 1 \); but this would contradict \( a^2 + b^2 = 1 \). Hence, the two integers \( a \) and \( b \) cannot both be nonzero. In other words, at least one of them is \( 0 \). In other words, we have \( a = 0 \) or \( b = 0 \). Thus, we are in one of the following two cases:

**Case 1:** We have \( a = 0 \).

**Case 2:** We have \( b = 0 \).

(These two cases could theoretically overlap, though it is easy to see that they don’t.)

Let us first consider Case 1. In this case, we have \( a = 0 \). Hence, \( a^2 + b^2 = 0^2 + b^2 = b^2 \), so that \( b^2 = a^2 + b^2 = 1 \). Hence, \( b \) is either 1 or \(-1\). Thus, the complex

\(^{132}\text{Proof.} \) We have \( i(-i) = 1 \). Hence, the Gaussian integer \( i \) has an inverse, namely \( i^{-1} = -i \).

Similarly, the Gaussian integer \(-i\) has an inverse, namely \((-i)^{-1} = i \).

The Gaussian integer \( 1 \) is invertible in \( \mathbb{Z}[i] \) (since its inverse is \( 1^{-1} = 1 \in \mathbb{Z}[i] \)), and thus is a unit.

The Gaussian integer \(-1 \) is invertible in \( \mathbb{Z}[i] \) (since its inverse is \((-1)^{-1} = -1 \in \mathbb{Z}[i] \)), and thus is a unit.

The Gaussian integer \( i \) is invertible in \( \mathbb{Z}[i] \) (since its inverse is \( i^{-1} = -i \in \mathbb{Z}[i] \)), and thus is a unit.

The Gaussian integer \(-i \) is invertible in \( \mathbb{Z}[i] \) (since its inverse is \((-i)^{-1} = i \in \mathbb{Z}[i] \)), and thus is a unit.

Thus, each of the four Gaussian integers \( 1, -1, i, -i \) is a unit.
number \((0, b)\) is either \((0, 1)\) or \((0, -1)\). In other words, the complex number \(\alpha\) is either \(i\) or \(-i\) (since \(\alpha = \begin{pmatrix} a \\ b \end{pmatrix} = (0, b)\) and \(i = (0, 1)\) and \(-i = -(0, 1) = (0, -1)\)). Thus, \(\alpha\) is either 1 or \(-1\) or \(i\) or \(-i\). So we have shown in Case 1 that \(\alpha\) is either 1 or \(-1\) or \(i\) or \(-i\).

Let us next consider Case 2. In this case, we have \(b = 0\). Hence, \(a^2 + b^2 = a^2 + 0^2 = a^2\), so that \(a^2 = a^2 + b^2 = 1\). Hence, \(\alpha\) is either 1 or \(-1\). Thus, the complex number \((a, 0)\) is either \((1, 0)\) or \((-1, 0)\). In other words, the complex number \(\alpha\) is either 1 or \(-1\) (since \(\alpha = \begin{pmatrix} a \\ b \end{pmatrix} = (a, 0)\) and \(1 = (1, 0)\) and \(-1 = (-1, 0)\)).

Thus, \(\alpha\) is either 1 or \(-1\) or \(i\) or \(-i\). So we have shown in Case 2 that \(\alpha\) is either 1 or \(-1\) or \(i\) or \(-i\).

We have now proven in both Cases 1 and 2 that \(\alpha\) is either 1 or \(-1\) or \(i\) or \(-i\). Hence, this always holds.

Now, forget that we fixed \(\alpha\). We thus have shown that if \(\alpha \in \mathbb{Z}[i]\) is a unit, then \(\alpha\) is either 1 or \(-1\) or \(i\) or \(-i\). Thus, \(1, -1, i, -i\) are the only possible units. Since we already know that \(1, -1, i, -i\) are units, we thus conclude that the units are \(1, -1, i, -i\). This proves Proposition 4.2.10.

As a consequence of Proposition 4.2.10, if we are given two Gaussian integers \(\alpha\) and \(\beta\), we can easily check whether \(\alpha \sim \beta\) holds: Just check whether we have \(\alpha = \beta\) or \(\alpha = -1\beta\) or \(\alpha = i\beta\) or \(\alpha = -i\beta\).

**Definition 4.2.11.** The unit-equivalence classes are defined to be the equivalence classes of the relation \(\sim\) on \(\mathbb{Z}[i]\). (We know from Proposition 4.2.8 that this relation \(\sim\) is an equivalence relation.)

---

You have reached the end of the finished part.

TODO: Write on from here.

---

**Proposition 4.2.12.** The unit-equivalence classes are the sets of the form \(\{\alpha, i\alpha, -\alpha, -i\alpha\}\) for some \(\alpha \in \mathbb{C}\).

**Proof.** Easy. \(\square\)

**Proposition 4.2.13.** Let \(\alpha\) be a Gaussian integer. Then, \(\alpha \sim 1\) if and only if \(\alpha\) is a unit.

**Proposition 4.2.14.** Let \(\alpha\) and \(\beta\) be two unit-equivalent Gaussian integers. Then, \(N(\alpha) = N(\beta)\).
Proof. We have $\alpha = \gamma \beta$ for some unit $\gamma \in \mathbb{Z}[i]$ (since $\alpha \sim \beta$). Consider this $\gamma$.

Since $\gamma$ is a unit, we have $N(\gamma) = 1$ (by Proposition 4.2.9 (b)). Thus, $N\left(\begin{array}{c} \alpha \\ = \gamma \beta \end{array}\right) = N(\gamma \beta) = N(\gamma) N(\beta) = N(\beta)$.

4.2.2. Divisibility and congruence

Now, we begin to do proper number theory with Gaussian integers. The next definition is the straightforward analogue of Definition 2.2.1.

**Definition 4.2.15.** Let $\alpha$ and $\beta$ be two Gaussian integers. We say that $\alpha \mid \beta$ (or “$\alpha$ divides $\beta$” or “$\beta$ is divisible by $\alpha$” or “$\beta$ is a multiple of $\alpha$”) if there exists a Gaussian integer $\gamma$ such that $\beta = \alpha \gamma$.

We furthermore say that $\alpha \nmid \beta$ if $\alpha$ does not divide $\beta$.

When making such a definition, we need to be very careful: Potentially, it might create a clash of notations. In fact, every integer is a Gaussian integer. If $a$ and $b$ are integers, then the statement “$a \mid b$” already has a meaning (explained in Definition 2.2.1). Definition 4.2.15 gives this statement a new meaning, because we can consider our integers $a$ and $b$ as Gaussian integers. If these two meanings are not equivalent, then we have laid ourselves a landmine!

Fortunately, these two meanings are equivalent. That is: If $a$ and $b$ are two integers, then the statement “$a \mid b$” interpreted according to Definition 2.2.1 is equivalent to the statement “$a \mid b$” interpreted according to Definition 4.2.15. This follows from the following proposition:

**Proposition 4.2.16.** Let $a \in \mathbb{Z}$ and $\beta = (b, c) \in \mathbb{Z}[i]$. Then, $a \mid \beta$ if and only if $a$ divides both $b$ and $c$. 

Proof. We have the following equivalence:

\[
\begin{align*}
(a &\mid \beta) \\
\iff &\ (\text{there exists a Gaussian integer } \gamma \text{ such that } \beta = a\gamma) \\
\iff &\ (\text{there exists a Gaussian integer } (u,v) \text{ such that } \beta = a(u,v)) \\
\iff &\ (\text{there exist integers } u \text{ and } v \text{ such that } \beta = a(u,v)) \\
\iff &\ \left(\begin{array}{l}
\text{there exist integers } u \text{ and } v \text{ such that } (b,c) = (au,av) \\
\text{(since } \beta = (b,c) \text{ and } a(u,v) = (au,av)) \\
\text{there exist integers } u \text{ and } v \text{ such that } b = au \text{ and } c = av \\
\text{there exist an integer } u \text{ such that } b = au \\
\land \ (\text{there exists an integer } v \text{ such that } c = av) \\
\text{there exists an integer } u \text{ such that } b = au \\
\land \ (\text{there exists an integer } v \text{ such that } c = av) \\
\text{there exist integers } u \text{ and } v \text{ such that } b = au \\
\land \ (\text{there exists an integer } v \text{ such that } c = av) \\
\iff &\ (a \mid b) \\
\iff &\ (a \mid c) \\
\iff &\ (a \mid b) \land (a \mid c) \\
\iff &\ (a \text{ divides } b \text{ and } c).
\end{align*}
\]

The next proposition is a (partial) analogue of Proposition 2.2.3.

**Proposition 4.2.17.** Let \( \alpha \) and \( \beta \) be two Gaussian integers.

(a) If \( \alpha \mid \beta \), then \( N(\alpha) \mid N(\beta) \).

(b) If \( \alpha \mid \beta \) and \( \beta \neq 0 \), then \( N(\alpha) \leq N(\beta) \).

(c) Assume that \( \alpha \neq 0 \). Then, \( \alpha \mid \beta \) if and only if \( \frac{\beta}{\alpha} \in \mathbb{Z}[i] \).

Note that we are using the norms \( N(\alpha) \) and \( N(\beta) \) as analogues of \( |a| \) and \( |b| \) here, since the absolute values \( |\alpha| \) and \( |\beta| \) of Gaussian integers are often irrational and thus it makes no sense to talk of their divisibility.

Note that the converse of Proposition 4.2.17 (a) does not hold. (That is, \( N(\alpha) \mid N(\beta) \) does not yield \( \alpha \mid \beta \).)

**Proof of Proposition 4.2.17** (a) Assume that \( \alpha \mid \beta \). Then, \( \beta = \alpha\gamma \) for some \( \gamma \in \mathbb{Z}[i] \). Thus, \( \text{N}(\beta) = N(\alpha\gamma) = N(\alpha) \cdot N(\gamma) \), so \( N(\alpha) \mid N(\beta) \) since \( N(\gamma) \in \mathbb{N} \).

(b) follows from (a), since the norms are nonnegative integers (and since \( \beta \neq 0 \) implies \( N(\beta) \neq 0 \)).

(c) is proven as in the integer case.

The next proposition is a straightforward analogue of Proposition 2.2.4.

---

- **Math 4281 notes page 274**
- **Proof.**
- **Proposition 4.2.17.**
- **Note:**
- **Proof of Proposition 4.2.17** (a)
- **Note that:**
Proposition 4.2.18. (a) We have $\alpha \mid \alpha$ for every $\alpha \in \mathbb{Z}[i]$. (This is called the reflexivity of divisibility.)

(b) If $\alpha, \beta, \gamma \in \mathbb{Z}[i]$ satisfy $\alpha \mid \beta$ and $\beta \mid \gamma$, then $\alpha \mid \gamma$. (This is called the transitivity of divisibility.)

(c) If $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{Z}[i]$ satisfy $\alpha_1 \mid \beta_1$ and $\alpha_2 \mid \beta_2$, then $\alpha_1 \alpha_2 \mid \beta_1 \beta_2$.

Proof. Analogous to the proof of Proposition 2.2.4.

The next exercise is a Gaussian-integer analogue of Exercise 2.2.2:

Exercise 4.2.1. Let $\alpha$ and $\beta$ be two Gaussian integers such that $\alpha \mid \beta$ and $\beta \mid \alpha$. Prove that $\alpha \sim \beta$.

Solution sketch. WLOG assume that $\alpha \neq 0$. Then, $\beta \neq 0$. Now, from $\alpha \mid \beta$, we obtain $\frac{\beta}{\alpha} \in \mathbb{Z}[i]$. Similarly, $\frac{\alpha}{\beta} \in \mathbb{Z}[i]$. Thus, $\frac{\beta}{\alpha}$ and $\frac{\alpha}{\beta}$ are mutually inverse Gaussian integers. So $\frac{\beta}{\alpha}$ is invertible, i.e., is a unit. But $\beta = \alpha \cdot \frac{\beta}{\alpha}$, so $\beta \sim \alpha$ (since $\frac{\beta}{\alpha}$ is a unit), and thus $\alpha \sim \beta$. This solves Exercise 4.2.1.

Note that the conclusion “$\alpha \sim \beta$” in Exercise 4.2.1 is the proper Gaussian-integer analogue of the conclusion “$|a| = |b|$” in Exercise 2.2.2 (since (132) shows that unit-equivalence on $\mathbb{Z}[i]$ is an analogue of the “have the same absolute value” relation on $\mathbb{Z}$). (We could have stated the weaker conclusion $|\alpha| = |\beta|$ as well, but it would not be half as useful.)

The next exercise is an analogue of Exercise 2.2.3:

Exercise 4.2.2. Let $\alpha, \beta, \gamma$ be three Gaussian integers such that $\gamma \neq 0$. Prove that $\alpha \mid \beta$ holds if and only if $\alpha \gamma \mid \beta \gamma$.

The next exercise is an analogue of Exercise 2.2.4:

Exercise 4.2.3. Let $\nu \in \mathbb{Z}[i]$. Let $a, b \in \mathbb{N}$ be such that $a \leq b$. Prove that $\nu^a \mid \nu^b$.

Needless to say, the $a$ and $b$ in this exercise still have to be nonnegative integers, since Gaussian integers make no sense in exponents.

The next exercise is an analogue of Exercise 2.2.5:

Exercise 4.2.4. Let $\gamma$ be a Gaussian integer such that $\gamma \mid 1$. Prove that $\gamma \sim 1$ (that is, $\gamma$ is a unit, i.e., either 1 or $-1$ or $i$ or $-i$).

Next comes another trivial fact:

Exercise 4.2.5. Let $\alpha$ and $\beta$ be Gaussian integers such that $\alpha \mid \beta$. Prove that $\bar{\alpha} \mid \bar{\beta}$.

2019-03-06 lecture
We have defined congruence for integers in Definition 2.3.1. We can repeat the
same definition for Gaussian integers:

**Definition 4.2.19.** Let \( \nu, \alpha, \beta \in \mathbb{Z}[i] \). We say that \( \alpha \) is congruent to \( \beta \) modulo \( \nu \) if
and only if \( \nu \mid \alpha - \beta \).

We furthermore shall use the notation “\( \alpha \equiv \beta \mod \nu \)” for “\( \alpha \) is congruent to \( \beta \) modulo \( \nu \)”.

Once again, such a definition risks sneaking in ambiguity, but fortunately this
one does not: If \( n, a, b \in \mathbb{Z} \), then the statement “\( a \equiv b \mod n \)” interpreted accord-
ing to Definition 2.3.1 is equivalent to the statement “\( a \equiv b \mod n \)” interpreted
according to Definition 4.2.19 (by treating \( n, a, b \) as Gaussian integers). To do so,
recall that both statements are defined to mean “\( n \mid a - b \)” and the meaning of the
latter statement does not depend on whether we interpret \( n, a, b \) as integers or as
Gaussian integers.

The next proposition is a straightforward analogue of Proposition 2.3.3:

**Proposition 4.2.20.** Let \( \nu \in \mathbb{Z}[i] \) and \( \alpha \in \mathbb{Z}[i] \). Then, \( \alpha \equiv 0 \mod \nu \) if and only if
\( \nu \mid \alpha \).

*Proof.* Analogous to the proof of Proposition 2.3.3. \( \square \)

The next proposition is a straightforward analogue of Proposition 2.3.4:

**Proposition 4.2.21.** Let \( \nu \in \mathbb{Z}[i] \).

(a) We have \( \alpha \equiv \alpha \mod \nu \) for every \( \alpha \in \mathbb{Z}[i] \).

(b) If \( \alpha, \beta, \gamma \in \mathbb{Z}[i] \) satisfy \( \alpha \equiv \beta \mod \nu \) and \( \beta \equiv \gamma \mod \nu \), then \( \alpha \equiv \gamma \mod \nu \).

(c) If \( \alpha, \beta \in \mathbb{Z}[i] \) satisfy \( \alpha \equiv \beta \mod \nu \), then \( \beta \equiv \alpha \mod \nu \).

(d) If \( \alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{Z}[i] \) satisfy \( \alpha_1 \equiv \beta_1 \mod \nu \) and \( \alpha_2 \equiv \beta_2 \mod \nu \), then
\[
\alpha_1 + \alpha_2 \equiv \beta_1 + \beta_2 \mod \nu; \tag{133}
\]
\[
\alpha_1 - \alpha_2 \equiv \beta_1 - \beta_2 \mod \nu; \tag{134}
\]
\[
\alpha_1\alpha_2 \equiv \beta_1\beta_2 \mod \nu. \tag{135}
\]

(e) Let \( \mu \in \mathbb{Z}[i] \) be such that \( \mu \mid \nu \). If \( \alpha, \beta \in \mathbb{Z}[i] \) satisfy \( \alpha \equiv \beta \mod \nu \), then
\( \alpha \equiv \beta \mod \mu \).

*Proof.* Analogous to the proof of Proposition 2.3.4. \( \square \)

### 4.2.3. Division with remainder

Now, let us try to make division with remainder work for Gaussian integers. This is
no longer easy or just a straightforward modification of the corresponding situation
for integers.

Let us recall how division with remainder works for (usual) integers. The relevant fact (Theorem 2.6.1 with \( u \) and \( n \) renamed as \( a \) and \( b \)) is the following:
Theorem 4.2.22. Let $b$ be a positive integer. Let $a \in \mathbb{Z}$. Then, there exists a unique pair $(q, r) \in \mathbb{Z} \times \{0, 1, \ldots, b - 1\}$ such that $a = qb + r$.

With Gaussian integers, it is no longer clear what $\{0, 1, \ldots, b - 1\}$ would be. But when we have Gaussian integers $\alpha$ and $\beta \neq 0$, we may try to find a pair $(\gamma, \rho) \in \mathbb{Z}[i] \times \mathbb{Z}[i]$ such that $\alpha = \gamma \beta + \rho$ and $N(\rho) < N(\beta)$. It turns out that such a pair exists (but is not unique).

To prove this, let us recall a similar statement about integers – which turns out to be not the original Theorem 2.6.1, but rather a modified version thereof:

Theorem 4.2.23. (Minimum-size division theorem) Let $b$ be a positive integer. Let $a \in \mathbb{Z}$. Then, there exist integers $q$ and $r$ such that $a = qb + r$ and $|r| \leq b/2$.

This is the claim of Exercise 2.6.2(a). Note that $(q, r)$ is not always unique.

Now, let us state an analogous fact for Gaussian integers:

Theorem 4.2.24. (Division-with-remainder theorem for Gaussian integers:) Let $\alpha$ and $\beta \neq 0$ be Gaussian integers. There exist Gaussian integers $\gamma$ and $\rho$ such that $\alpha = \gamma \beta + \rho$ and $N(\rho) \leq N(\beta)/2$.

Note that $(\gamma, \rho)$ is not unique. It is best to regard Theorem 4.2.24 as an analogue of Exercise 2.6.2(a), not as an analogue of Theorem 2.6.1; nevertheless, it is the closest we can get to Theorem 2.6.1 in $\mathbb{Z}[i]$, and can often be substituted in places where one would usually want to apply Theorem 2.6.1.

Theorem 4.2.24 can be visualized geometrically (similarly to the visualizations shown in Remark 2.6.8 and Remark 2.6.10). See [ConradG §7] for the details.

Proof of Theorem 4.2.24 (The following proof follows [ConradG, proof of Theorem 3.1].) Let $N = N(\beta)$. Write

$$\frac{\alpha}{\beta} = \frac{\overline{\alpha} \beta}{\overline{\beta} \beta} = \frac{\overline{\alpha} \beta}{N} \quad \text{(since } \beta \overline{\beta} = N(\beta) = N).$$

Note that $\alpha \overline{\beta}$ is a Gaussian integer (since $\alpha$ and $\beta$ are Gaussian integers); thus, we can write it in the form

$$\alpha \overline{\beta} = m + ni \quad \text{for some } m, n \in \mathbb{Z}.$$

Consider these $m, n$. Note that $N = N(\beta) > 0$ since $\beta \neq 0$.

Exercise 2.6.2(a) (applied to $N$ and $m$ instead of $n$ and $u$) shows that there exists a pair $(q_1, r_1) \in \mathbb{Z} \times \mathbb{Z}$ such that

$$m = q_1 N + r_1 \quad \text{and} \quad |r_1| \leq N/2.$$

Consider this pair.
Exercise 2.6.2 (a) (applied to $N$ and $n$ instead of $n$ and $u$) shows that there exists a pair $(q_2, r_2) \in \mathbb{Z} \times \mathbb{Z}$ such that

$$n = q_2 N + r_2 \quad \text{and} \quad |r_2| \leq N/2.$$ 

Consider this pair.

Now,

$$\frac{\alpha}{\beta} = \frac{\alpha \beta}{N} = \frac{(q_1 N + r_1) + (q_2 N + r_2)i}{N},$$

since $\frac{\alpha \beta}{N} = \frac{m}{q_1 N + r_1} + \frac{n}{q_2 N + r_2} i = (q_1 N + r_1) + (q_2 N + r_2)i$)

$$= (q_1 + q_2i) + \frac{r_1 + r_2i}{N}.$$

(136)

Set $\gamma = q_1 + q_2i$ and $\rho = \alpha - \gamma \beta$. Note that $\gamma$ and thus $\rho$ are Gaussian integers.

From $\rho = \alpha - \gamma \beta$, we obtain $\alpha = \gamma \beta + \rho$. Thus it remains to prove $N(\rho) \leq N(\beta)/2$.

The equation (136) becomes $\frac{\alpha}{\beta} = \gamma + \frac{r_1 + r_2i}{N}$ (since $q_1 + q_2i = \gamma$), so that

$$\frac{r_1 + r_2i}{N} = \frac{\alpha}{\beta} - \gamma = \frac{\alpha - \gamma \beta}{\beta} = \frac{\rho}{\beta} \quad \text{(since } \alpha - \gamma \beta = \rho).$$

Hence,

$$\rho = \beta \cdot \frac{r_1 + r_2i}{\beta} = \beta \cdot \frac{r_1 + r_2i}{\beta \beta}$$

and thus

$$N(\rho) = N\left(\frac{r_1 + r_2i}{\beta}\right) = \frac{N(r_1 + r_2i)}{N(\beta))} \quad \text{(by Proposition 4.1.26 (e))}$$

$$= \frac{r_1^2 + r_2^2}{N} \quad \text{(since } N(r_1 + r_2i) = r_1^2 + r_2^2 \text{ and } N(\beta) = N(\beta) = N)$$

$$= \frac{|r_1|^2 + |r_2|^2}{N} \quad \text{(since } r^2 = |r|^2 \text{ for every real } r)$$

$$\leq \frac{(N/2)^2 + (N/2)^2}{N} \quad \text{(since } |r_1| \leq N/2 \text{ and } |r_2| \leq N/2)$$

$$= N/2 = N(\beta) / 2$$

(since $N = N(\beta)$). This proves Theorem 4.2.24.

Note: We cannot define $\alpha/\beta$ or $\alpha\%\beta$ for Gaussian integers $\alpha$ and $\beta$, since there is no uniqueness statement in Theorem 4.2.24.
4.2.4. Common divisors

Next, we define the Gaussian divisors of a Gaussian integer (in analogy to Definition 2.9.1):

**Definition 4.2.25.** Let \( \beta \in \mathbb{Z}[i] \). The **Gaussian divisors** of \( \beta \) are defined as the Gaussian integers that divide \( \beta \).

Note that we are calling them “Gaussian divisors” and not “divisors”, because when \( \beta \) is an actual integer, there are (usually) Gaussian divisors of \( \beta \) that are not divisors of \( \beta \) (in the sense of Definition 2.9.1). For example, \( 1 + i \) is a Gaussian divisor of \( 2 \) (since \( 2 = (1 + i)(1 - i) \)), but the only divisors of \( 2 \) (in the sense of Definition 2.9.1) are \(-2, -1, 1, 2\). This is one of those situations where using the same name for a concept and its Gaussian-integer analogue would lead to ambiguities.

The following is an analogue of Proposition 2.9.2:

**Proposition 4.2.26.** (a) If \( \beta \in \mathbb{Z}[i] \), then \( 1 \) and \( \beta \) are Gaussian divisors of \( \beta \).

(b) The Gaussian divisors of \( 0 \) are all the Gaussian integers.

(c) Let \( \beta \in \mathbb{Z}[i] \) be nonzero. Then, all Gaussian divisors of \( \beta \) belong to the set

\[
\{ x + yi \mid x, y \in \mathbb{Z} \text{ satisfying } |x| \leq |eta| \text{ and } |y| \leq |eta| \}.
\]

**Proof of Proposition 4.2.26.**

Parts (a) and (b) are clear.

(c) Let \( x + yi \) be a Gaussian divisor of \( \beta \). Then, Proposition 4.2.17 (b) yields \( N(x + yi) \leq N(\beta) \), thus \( |x + yi| \leq |eta| \) (since \( |\alpha| = \sqrt{N(\alpha)} \) for every \( \alpha \in \mathbb{C} \)). But \( |x + yi| = \sqrt{x^2 + y^2} \geq \sqrt{x^2} = |x| \), so that \( |x| \leq |x + yi| \leq |eta| \). Similarly, \( |y| \leq |eta| \).

Thus, again, finding all Gaussian divisors of a Gaussian integer \( \beta \) is a problem solvable in finite time. (Indeed, the set in Proposition 4.2.26 (c) is clearly finite.)

The following is a straightforward analogue of Definition 2.9.3:

**Definition 4.2.27.** Let \( \beta_1, \beta_2, \ldots, \beta_k \) be Gaussian integers. Then, the **common Gaussian divisors** of \( \beta_1, \beta_2, \ldots, \beta_k \) are defined to be the Gaussian integers \( \alpha \) that satisfy

\[
(\alpha | \beta_i \text{ for all } i \in \{1, 2, \ldots, k\})
\]

(in other words, that divide all of the Gaussian integers \( \beta_1, \beta_2, \ldots, \beta_k \)). We let \( \text{Div}_{\mathbb{Z}[i]}(\beta_1, \beta_2, \ldots, \beta_k) \) denote the set of these common Gaussian divisors.

The reason why I chose the notation \( \text{Div}_{\mathbb{Z}[i]}(\beta_1, \beta_2, \ldots, \beta_k) \) rather than the simpler notation \( \text{Div}(\beta_1, \beta_2, \ldots, \beta_k) \) is that the latter would be ambiguous. In fact, when \( \beta_1, \beta_2, \ldots, \beta_k \) are integers, the set \( \text{Div}(\beta_1, \beta_2, \ldots, \beta_k) \) of common divisors of
β₁, β₂, . . . , βₖ is not the set Div_{\mathbb{Z}[i]} (β₁, β₂, . . . , βₖ) of common Gaussian divisors of β₁, β₂, . . . , βₖ. (For example, the former set does not contain i, while the latter does.)

We cannot directly define a “greatest common Gaussian divisor of β₁, β₂, . . . , βₖ” to be the greatest element of Div_{\mathbb{Z}[i]} (β₁, β₂, . . . , βₖ), since “greatest” does not make sense for complex numbers. (Even if we wanted “greatest in norm”, it would not a-priori be obvious that there are no ties, i.e., that such a greatest common Gaussian divisor is unique.)

However, it turns out that a “greatest common Gaussian divisor” gcd_{\mathbb{Z}[i]} (β₁, β₂, . . . , βₖ) actually can be defined reasonably (although only up to multiplication by units).

Before we can do so, let us state some basic properties of common Gaussian divisors:

**Proposition 4.2.28.** (a) We have Div_{\mathbb{Z}[i]} (α, 0) = Div_{\mathbb{Z}[i]} (α) for all α ∈ \mathbb{Z}[i].
(b) We have Div_{\mathbb{Z}[i]} (α, β) = Div_{\mathbb{Z}[i]} (β, α) for all α, β ∈ \mathbb{Z}[i].
(c) We have Div_{\mathbb{Z}[i]} (α, ηα + β) = Div_{\mathbb{Z}[i]} (α, β) for all α, β, η ∈ \mathbb{Z}[i].
(d) If α, β, γ ∈ \mathbb{Z}[i] satisfy β ≡ γ mod α, then Div_{\mathbb{Z}[i]} (α, β) = Div_{\mathbb{Z}[i]} (α, γ).
(g) We have Div_{\mathbb{Z}[i]} (ηα, β) = Div_{\mathbb{Z}[i]} (α, β) for all α, β ∈ \mathbb{Z}[i] for every unit η ∈ \mathbb{Z}[i].
(h) We have Div_{\mathbb{Z}[i]} (α, ηβ) = Div_{\mathbb{Z}[i]} (α, β) for all α, β ∈ \mathbb{Z}[i] for every unit η ∈ \mathbb{Z}[i].
(i) If α, β ∈ \mathbb{Z}[i] satisfy α | β, then Div_{\mathbb{Z}[i]} (α, β) = Div_{\mathbb{Z}[i]} (α).
(j) The common Gaussian divisors of the empty list of Gaussian integers are Div_{\mathbb{Z}[i]} () = \mathbb{Z}[i].

**Proof.** Most of these facts are analogues of Proposition 2.9.7 or rather of the corresponding properties of Div (a, b) for two integers a and b that were proven during our proof of Proposition 2.9.7. Their proofs also are straightforward adaptations of the proofs of the latter properties. Let us only sketch the proof of (g), since it may require some extra thinking:

(g) Let η ∈ \mathbb{Z}[i] be a unit.

Claim: The Gaussian divisors of α are exactly the Gaussian divisors of ηα.

(Indeed, any Gaussian divisor of α is clearly a Gaussian divisor of ηα. Conversely, since η⁻¹ is a Gaussian integer, any Gaussian divisor of ηα is a Gaussian divisor of η⁻¹ηα = α.)

The rest is proven just as for integers, except that we don’t make the final step from Div to gcd.

Now, we can compute Div_{\mathbb{Z}[i]} (α, β) for Gaussian integers α and β by a version of the “Euclidean algorithm” that we used to compute gcd (a, b) for integers a and
b. For example, we can compute $\text{Div}_{\mathbb{Z}[i]}(32 + 9i, 4 + 11i)$ as follows:\footnote{This is [ConradG, Example 4.4].}

$\text{Div}_{\mathbb{Z}[i]}(32 + 9i, 4 + 11i)$

\[= \text{Div}_{\mathbb{Z}[i]} \left( 4 + 11i, \frac{32 + 9i}{(2 - 2i)(4 + 11i) + (2 - 5i)} \right) \quad (\text{by Proposition 4.2.28 (b)}) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( 4 + 11i, (2 - 2i)(4 + 11i) + (2 - 5i) \right) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( 4 + 11i, 2 - 5i \right) \quad (\text{by Proposition 4.2.28 (c)}) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( 2 - 5i, \frac{4 + 11i}{(-2 + i)(2 - 5i) + (3 - i)} \right) \quad (\text{by Proposition 4.2.28 (b)}) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( 2 - 5i, (-2 + i)(2 - 5i) + (3 - i) \right) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( 2 - 5i, 3 - i \right) \quad (\text{by Proposition 4.2.28 (c)}) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( -i, \frac{2 - 5i}{(1 - i)(3 - i) - i} \right) \quad (\text{by Proposition 4.2.28 (b)}) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( -i, (1 - i)(3 - i) - i \right) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( 3 - i, (1 - i)(3 - i) - i \right) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( 3 - i, -i \right) \quad (\text{by Proposition 4.2.28 (c)}) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( -i, \frac{3 - i}{(1 + 3i)(-i) + 0} \right) \quad (\text{by Proposition 4.2.28 (b)}) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( -i, (1 + 3i)(-i) + 0 \right) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( -i, 0 \right) \quad (\text{by Proposition 4.2.28 (c)}) \]

\[= \text{Div}_{\mathbb{Z}[i]} \left( -i \right) \quad (\text{by Proposition 4.2.28 (a)}) \]

\[= \{1, i, -1, -i\}. \]

In the same way, for any two Gaussian integers $\alpha$ and $\beta$ we can find a Gaussian integer $\gamma$ such that $\text{Div}_{\mathbb{Z}[i]}(\alpha, \beta) = \text{Div}_{\mathbb{Z}[i]}(\gamma)$. This resulting $\gamma$ will actually be unique up to multiplication by units (i.e., its unit-equivalence class will be unique). Better yet, we have the following:

**Theorem 4.2.29.** (Bezout’s theorem for Gaussian integers:)

Let $\alpha, \beta \in \mathbb{Z}[i]$. Then:
- (a) There exists a $\mathbb{Z}[i]$-linear combination $\gamma$ of $\alpha$ and $\beta$ that is a common Gaussian divisor of $\alpha$ and $\beta$. (Note: A $\mathbb{Z}[i]$-linear combination of $\alpha$ and $\beta$ means a Gaussian integer of the form $\lambda \alpha + \mu \beta$ with $\lambda, \mu \in \mathbb{Z}[i]$.)
(b) Any such γ satisfies Div_{Z[i]}(α, β) = Div_{Z[i]}(γ).

(c) The unit-equivalence class of this γ is uniquely determined.

This theorem is, in a sense, a generalization of Theorem 2.9.11 even though (unlike the latter theorem) it does not rely on an already existing concept of “greatest common divisor” but rather builds the foundation for such a concept. With Theorem 4.2.29 in hand, it makes sense to call γ the “greatest common Gaussian divisor” of α and β, but rigorously speaking this name should be reserved for the unit-equivalence class of γ since γ itself is not unique.

2019-03-08 lecture

Proof of Theorem 4.2.29 (sketched). (a) Rough idea: This is analogous to the proof of Lemma 2.9.12. (But instead of strong induction on a + b, we need to do strong induction on N(α) + N(β). Also, instead of the usual division-with-remainder theorem, you have to use Theorem 4.2.24. Note that the inequality N(ρ) ≤ N(β) / 2 in Theorem 4.2.24 implies N(ρ) < N(β), which is the only inequality you need.)

Actually, here are some more details of the proof. For any α, β ∈ Z[i], we let Lin_{Z[i]}(α, β) be the set of all Z[i]-linear combinations of α and β. (This will be called the Z[i]-linear span of α and β later on, in analogy to spans in classical linear algebra.) Now, the claim of Theorem 4.2.29 (a) can be restated as follows:

\[ \text{Div}_{Z[i]}(α, β) \cap \text{Lin}_{Z[i]}(α, β) \neq \emptyset. \] (138)

We shall prove (138) by strong induction on N(α) + N(β).

So we fix n ∈ N, and assume as the induction hypothesis that (138) holds for all α, β ∈ Z[i] satisfying N(α) + N(β) < n. We must now prove (138) for all α, β ∈ Z[i] satisfying N(α) + N(β) = n.

So let α, β ∈ Z[i] be such that N(α) + N(β) = n. We must prove Div_{Z[i]}(α, β) ∩ Lin_{Z[i]}(α, β) ≠ ∅. We can WLOG assume N(β) ≥ N(α), since otherwise we can swap α with β without changing any of the sets Div_{Z[i]}(α, β) and Lin_{Z[i]}(α, β). Assume this. Furthermore, we WLOG assume that α ≠ 0 (since otherwise, the set Div_{Z[i]}(α, β) ∩ Lin_{Z[i]}(α, β) = Div_{Z[i]}(0, β) ∩ Lin_{Z[i]}(0, β) clearly contains β and thus is ≠ ∅). Hence, N(α) > 0. Now, Theorem 4.2.24 (applied to β and α instead of α and β) yields that there exist Gaussian integers γ and ρ such that ρ = γα + ρ and N(ρ) ≤ N(α) / 2. Consider these γ and ρ. From β = γα + ρ, we obtain ρ = β − γα.

The Gaussian integers β and ρ satisfy β ≡ ρ mod α (since β = γα + ρ). Hence, Proposition 4.2.28 (d) yields Div_{Z[i]}(α, β) = Div_{Z[i]}(α, ρ). Also, it is easy to see that Lin_{Z[i]}(α, β) = Lin_{Z[i]}(α, ρ) (since every λ, μ ∈ Z[i] satisfy

\[ λα + μ \overset{β}{=} = γα + ρ \]

\[ λα + μ \overset{ρ}{=} = β − γα \]
proves Theorem 4.2.29.

(c) Similarly, Div_{\mathcal{Z}[i]} (\alpha, \beta) \nsubseteq \varnothing boils down to proving Div_{\mathcal{Z}[i]} (\alpha, \rho) \nsubseteq \varnothing. But this follows from the induction hypothesis, since

\[ N(\alpha) + \underbrace{N(\rho)}_{\leq N(\alpha)/2} < N(\alpha) + N(\beta) = n. \]

This completes the induction step. Hence, (138) (and thus Theorem 4.2.29 (a)) follows by strong induction.

(b) We shall prove Div_{\mathcal{Z}[i]} (\alpha, \beta) \subseteq Div_{\mathcal{Z}[i]} (\gamma) and Div_{\mathcal{Z}[i]} (\alpha, \beta) \supseteq Div_{\mathcal{Z}[i]} (\gamma) separately:

\(\subseteq\): Since \(\gamma\) is a \(\mathcal{Z}[i]\)-linear combination of \(\alpha\) and \(\beta\), every common Gaussian divisor of \(\alpha\) and \(\beta\) must also divide \(\gamma\). Thus, Div_{\mathcal{Z}[i]} (\alpha, \beta) \subseteq Div_{\mathcal{Z}[i]} (\gamma).

\(\supseteq\): Since \(\gamma\) is a common Gaussian divisor of \(\alpha\) and \(\beta\), every Gaussian divisor of \(\gamma\) must be a common Gaussian divisor of \(\alpha\) and \(\beta\). Thus, Div_{\mathcal{Z}[i]} (\alpha, \beta) \supseteq Div_{\mathcal{Z}[i]} (\gamma).

(c) Let \(\gamma_1\) and \(\gamma_2\) be two such \(\gamma\)'s. We must prove that \(\gamma_1 \sim \gamma_2\).

We have Div_{\mathcal{Z}[i]} (\alpha, \beta) = Div_{\mathcal{Z}[i]} (\gamma_1) and Div_{\mathcal{Z}[i]} (\alpha, \beta) = Div_{\mathcal{Z}[i]} (\gamma_2), so that Div_{\mathcal{Z}[i]} (\gamma_1) = Div_{\mathcal{Z}[i]} (\gamma_2). Now, \(\gamma_1 \in Div_{\mathcal{Z}[i]} (\gamma_1) = Div_{\mathcal{Z}[i]} (\gamma_2), \) thus \(\gamma_1 \mid \gamma_2\).

Similarly, \(\gamma_2 \mid \gamma_1\). Combining these, we obtain \(\gamma_1 \sim \gamma_2\) (by Exercise 4.2.1). This proves Theorem 4.2.29 (c).

Definition 4.2.30. The greatest common Gaussian divisor (or, short, gcd) of two Gaussian integers \(\alpha\) and \(\beta\) is defined to be the \(\gamma\) from Theorem 4.2.29 (a). It is called \(\text{gcd}_{\mathcal{Z}[i]} (\alpha, \beta)\).

So it is a common Gaussian divisor of \(\alpha\) and \(\beta\) and also a \(\mathcal{Z}[i]\)-linear combination of \(\alpha\) and \(\beta\) and satisfies

\[ \text{Div}_{\mathcal{Z}[i]} \left( \text{gcd}_{\mathcal{Z}[i]} (\alpha, \beta) \right) = \text{Div}_{\mathcal{Z}[i]} (\alpha, \beta). \] (139)

However, it is only well-defined up to unit-equivalence. Thus, if you have \(\gamma_1 = \text{gcd}_{\mathcal{Z}[i]} (\alpha, \beta)\) and \(\gamma_2 = \text{gcd}_{\mathcal{Z}[i]} (\alpha, \beta)\), then you cannot conclude that \(\gamma_1 = \gamma_2\) (you can only conclude \(\gamma_1 \sim \gamma_2\)). So, strictly speaking, we should have defined \(\text{gcd}_{\mathcal{Z}[i]} (\alpha, \beta)\) as a unit-equivalence class, not as a concrete Gaussian integer. But we will allow ourselves this abuse of notation. We shall not write equality signs like the one in “\(\gamma_1 = \text{gcd}_{\mathcal{Z}[i]} (\alpha, \beta)\)”, however; we instead prefer to write “\(\gamma_1 \sim \text{gcd}_{\mathcal{Z}[i]} (\alpha, \beta)\)”. Generally, whenever you see \(\text{gcd}_{\mathcal{Z}[i]} (\alpha, \beta)\) in a statement, you should be understanding the statement to hold for every possible choice of \(\text{gcd}_{\mathcal{Z}[i]} (\alpha, \beta)\).

Proposition 4.2.31. Let \(a\) and \(b\) be two integers. Then,

\[ \text{gcd} (a, b) \sim \text{gcd}_{\mathcal{Z}[i]} (a, b). \]

(Of course, the gcd on the left hand side is the gcd of the two integers \(a\) and \(b\) as defined in Definition 2.9.6, whereas the \(\text{gcd}_{\mathcal{Z}[i]}\) on the right hand side is the greatest common Gaussian divisor of the Gaussian integers \(a\) and \(b\).)
Proof. The integer \( \gcd (a, b) \) is a common divisor of \( a \) and \( b \) and also is a \( \mathbb{Z} \)-linear combination of \( a \) and \( b \) (by Bezout). Therefore, it is also a common Gaussian divisor of the Gaussian integers \( a \) and \( b \) and also is a \( \mathbb{Z}[i] \)-linear combination of \( a \) and \( b \). But this yields that it is \( \gcd_{\mathbb{Z}[i]} (a, b) \) (due to the definition of \( \gcd_{\mathbb{Z}[i]} (a, b) \)). \( \square \)

This proposition allows us to write “\( \gcd \)” for both concepts of \( \gcd \) without having to disambiguate the meaning. (We shall not do so, however.)

**Proposition 4.2.32.** Let \( \alpha \) and \( \beta \) be two Gaussian integers, not both equal to 0. Then, the possible values of \( \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \) (that is, strictly speaking, all four elements of the unit-equivalence class \( \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \)) are exactly the elements of \( \text{Div}_{\mathbb{Z}[i]} \) \( (\alpha, \beta) \) having the largest norm.

Proof. First of all, \( \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \) is a common Gaussian divisor of \( \alpha \) and \( \beta \), and thus is \( \neq 0 \) (since \( \alpha \) and \( \beta \) are not both equal to 0). Thus, there are exactly four Gaussian integers unit-equivalent to \( \text{gcd}_{\mathbb{Z}[i]} (\alpha, \beta) \). In other words, there are exactly four possible values of \( \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \). We must show that these values are exactly the elements of \( \text{Div}_{\mathbb{Z}[i]} \) \( (\alpha, \beta) \) having the largest norm.

In other words, we must show the following two claims:

*Claim 1:* We have \( N \left( \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \right) > N(\gamma) \) for each \( \gamma \in \text{Div}_{\mathbb{Z}[i]} (\alpha, \beta) \) that does not satisfy \( \gamma \sim \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \).

*Claim 2:* We have \( N \left( \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \right) = N(\gamma) \) for each \( \gamma \in \text{Div}_{\mathbb{Z}[i]} (\alpha, \beta) \) that does satisfy \( \gamma \sim \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \).

Claim 2 is obvious, since any two unit-equivalent Gaussian integers have the same norm (by Proposition 4.2.14).

*Proof of Claim 1:* Let \( \gamma \in \text{Div}_{\mathbb{Z}[i]} (\alpha, \beta) \) do not satisfy \( \gamma \sim \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \). Now, \( \gamma \in \text{Div}_{\mathbb{Z}[i]} (\alpha, \beta) = \text{Div}_{\mathbb{Z}[i]} \left( \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \right) \) (by (139)). Hence, \( \gamma | \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \).

Let us set \( \delta = \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \). So \( \gamma | \delta \). We have \( \delta \neq 0 \) (because \( \alpha \) and \( \beta \) are not both zero) and thus \( \gamma \neq 0 \) (since \( \gamma | \delta \)). Thus, \( \gamma | \delta \) yields that \( \frac{\delta}{\gamma} \) is a Gaussian integer, which is furthermore nonzero (since \( \delta \neq 0 \)). If this Gaussian integer \( \frac{\delta}{\gamma} \) was a unit, then we would have \( \gamma \sim \delta = \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \), which would contradict the assumption that \( \gamma \) does not satisfy \( \gamma \sim \gcd_{\mathbb{Z}[i]} (\alpha, \beta) \). So \( \frac{\delta}{\gamma} \) is a nonzero Gaussian integer that is not a unit. Hence, \( N \left( \frac{\delta}{\gamma} \right) > 1 \) (because Proposition 4.2.9 yields that every nonzero Gaussian integer that is not a unit must have norm \( > 1 \)). Now,

\[
N(\delta) = N \left( \frac{\delta}{\gamma} \right) \cdot N(\gamma) > N(\gamma).
\]
In other words, \( N\left( \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \right) > N(\gamma) \). This proves Claim 1. \( \square \)

Proposition 4.2.32 shows that \( \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \) is uniquely determined by the set \( \text{Div}_{\mathbb{Z}[i]}(\alpha, \beta) \). (Yes, you have to consider the case \( \alpha = \beta = 0 \) separately in proving this.) Hence, Proposition 4.2.28 yields:

**Proposition 4.2.33.** (a) We have \( \gcd_{\mathbb{Z}[i]}(\alpha, 0) \sim \gcd_{\mathbb{Z}[i]}(\alpha) \) for all \( \alpha \in \mathbb{Z}[i] \).

(b) We have \( \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \sim \gcd_{\mathbb{Z}[i]}(\beta, \alpha) \) for all \( \alpha, \beta \in \mathbb{Z}[i] \).

(c) We have \( \gcd_{\mathbb{Z}[i]}(\alpha, \eta \alpha + \beta) \sim \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \) for all \( \alpha, \beta, \eta \in \mathbb{Z}[i] \).

(d) If \( \alpha, \beta, \gamma \in \mathbb{Z}[i] \) satisfy \( \beta = \gamma \mod \alpha \), then \( \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \sim \gcd_{\mathbb{Z}[i]}(\alpha, \gamma) \).

(e) We have \( \gcd_{\mathbb{Z}[i]}(\eta \alpha, \beta) \sim \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \) for all \( \alpha, \beta \in \mathbb{Z}[i] \) for every unit \( \eta \in \mathbb{Z}[i] \).

(f) We have \( \gcd_{\mathbb{Z}[i]}(\eta \alpha, \beta) \sim \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \) for all \( \alpha, \beta \in \mathbb{Z}[i] \) for every unit \( \eta \in \mathbb{Z}[i] \).

(g) We have \( \gcd_{\mathbb{Z}[i]}(\eta \alpha, \beta) \sim \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \) for all \( \alpha, \beta \in \mathbb{Z}[i] \) for every unit \( \eta \in \mathbb{Z}[i] \).

(h) We have \( \gcd_{\mathbb{Z}[i]}(\eta \alpha, \beta) \sim \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \) for all \( \alpha, \beta \in \mathbb{Z}[i] \) for every unit \( \eta \in \mathbb{Z}[i] \).

(i) If \( \alpha, \beta \in \mathbb{Z}[i] \) satisfy \( \alpha | \beta \), then \( \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \sim \gcd_{\mathbb{Z}[i]}(\alpha) \).

(j) If the greatest common Gaussian divisor of the empty list of Gaussian integers is \( \gcd_{\mathbb{Z}[i]}() = 0 \).

Theorem 2.9.14 still holds for Gaussian integers.

Theorem 2.9.16 still holds for Gaussian integers.

Theorem 2.9.18 still holds for Gaussian integers.

Corollary 2.9.19 has to be modified as follows:

**Corollary 4.2.34.** Let \( \sigma, \alpha, \beta \in \mathbb{Z}[i] \). Then,

\[
\gcd_{\mathbb{Z}[i]}(\sigma \alpha, \sigma \beta) \sim \sigma \gcd_{\mathbb{Z}[i]}(\alpha, \beta).
\]

Exercise 2.9.4 still holds for Gaussian integers.

Exercise 2.9.5 becomes the claim that if \( \alpha_1 \sim \alpha_2 \) and \( \beta_1 \sim \beta_2 \), then \( \gcd_{\mathbb{Z}[i]}(\alpha_1, \beta_1) \sim \gcd_{\mathbb{Z}[i]}(\alpha_2, \beta_2) \). The solution does not carry over, but you can easily prove this new claim by hand.

Greatest common Gaussian divisors of \( k \) Gaussian integers can also be defined.

The next definition is an analogue of Definition 2.10.1.

**Definition 4.2.35.** Let \( \alpha \) and \( \beta \) be two Gaussian integers. We say that \( \alpha \) is coprime to \( \beta \) if and only if \( \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \sim 1 \) (that is, \( \gcd_{\mathbb{Z}[i]}(\alpha, \beta) \) is a unit).

Thus, any two coprime integers are also two coprime Gaussian integers (because of Proposition 4.2.31). This is why we can afford speaking of "coprime Gaussian integers" and not just "Gaussian-coprime Gaussian integers".

Everything we said about coprinality of integers still holds for Gaussian integers. In particular, Proposition 2.10.4, Theorem 2.10.6, Theorem 2.10.7, Theorem 2.10.8 and Theorem 2.10.9 still hold if all integers are replaced by Gaussian integers (with the caveat that the gcd is no longer unique, so for example...
“$ab \equiv \gcd(a, n) \mod n$” must be interpreted as “$ab$ is congruent to some of the possible values of $\gcd_{\mathbb{Z}[i]}(a, n)$ modulo $n$”.

We could define Gaussian rationals (their set is called $\mathbb{Q}[i]$) as complex numbers $a + bi$ with $a, b \in \mathbb{Q}$. These are exactly the quotients of Gaussian integers.

Lowest common multiples of Gaussian integers still exist, but their definition has to be modified. For example, we can define $\text{lcm}_{\mathbb{Z}[i]}(\alpha, \beta)$ as the (unique up to unit-equivalence) Gaussian integer $\gamma$ such that the Gaussian common multiples of $\alpha$ and $\beta$ are the Gaussian multiples of $\gamma$. (We would have to prove that it actually is unique and exists.) Theorem 2.11.6 still holds, in the sense that $\gcd_{\mathbb{Z}[i]}(\alpha, \beta) \cdot \text{lcm}_{\mathbb{Z}[i]}(\alpha, \beta) \sim \alpha \beta$. Many other properties of lowest common multiplies extend to Gaussian integers.

The Chinese remainder theorem (Theorem 2.12.1) still holds for coprime Gaussian integers $\mu$ and $\nu$. Similarly for $k$ mutually coprime Gaussian integers.

4.2.5. Gaussian primes

The next definition is an analogue of Definition 2.13.1:

**Definition 4.2.36.** Let $\pi$ be a nonzero Gaussian integer that is not a unit. We say that $\pi$ is a Gaussian prime if each Gaussian divisor of $\pi$ is either a unit or unit-equivalent to $\pi$.

The letter “$\pi$” in this definition is unrelated to the irrational number $\pi = 3.14159\ldots$. It just happens to be the Greek letter corresponding to the Roman “$p$”.

The Gaussian primes are not a superset of the primes. For example:

**Example 4.2.37.** The Gaussian integer 2 is not a Gaussian prime.

*Proof.* We have $2 = (1 + i)(1 - i)$. The factors $1 + i$ and $1 - i$ have norms 2, which means that they are neither units themselves (since units would have norm 1) nor unit-equivalent to 2 (since 2 has norm 4, but unit-equivalent Gaussian integers have equal norms). Thus, $1 + i$ is a Gaussian divisor of 2 that is neither a unit nor is unit-equivalent to 2. Hence, 2 is not a Gaussian prime (by the definition of “Gaussian prime”).

So don’t forget the word “Gaussian” when you mean it!

Let us search for Gaussian primes. So we know that 2 is not a Gaussian prime. What about 3?

**Example 4.2.38.** The Gaussian integer 3 is a Gaussian prime.

*Proof.* Assume the contrary. Thus, there exists a Gaussian divisor $\alpha$ of 3 that is neither a unit nor unit-equivalent to 3 (since 3 is a nonzero Gaussian integer that is not a unit). Consider this $\alpha$. We have $3 = \alpha \gamma$ for some Gaussian integer $\gamma$ (since $\alpha$ is a Gaussian divisor $\gamma$ of 3). Consider this $\gamma$. 
If $\gamma$ was a unit, then we would have $3 \sim \alpha$ (since $3 = \alpha \gamma$); but this would contradict the fact that $\alpha$ is not unit-equivalent to 3. Hence, $\gamma$ is not a unit. Thus, $N(\gamma) \neq 1$ (by Proposition 4.2.9(b)). Also, $\alpha$ is not a unit; hence, $N(\alpha) \neq 1$ (by Proposition 4.2.9(b)).

We have $3 = \alpha \gamma$. Thus, $N(3) = N(\alpha \gamma) = N(\alpha) \cdot N(\gamma)$ (by Proposition 4.1.26(d)). Hence, $N(\alpha) \cdot N(\gamma) = N(3) = 3^2 + 0^2 = 9$. Since $N(\alpha)$ and $N(\gamma)$ are nonnegative integers, this would mean that

- either $N(\alpha) = 1$ and $N(\gamma) = 9$,
- or $N(\alpha) = 3$ and $N(\gamma) = 3$,
- or $N(\alpha) = 9$ and $N(\gamma) = 1$.

The first and the third of these three options are impossible, since $N(\alpha) \neq 1$ and $N(\gamma) \neq 1$. So the second option must be true. Thus, $N(\alpha) = 3$ and $N(\gamma) = 3$. But let us write the Gaussian integer $\alpha$ as $(a, b)$ for some $a, b \in \mathbb{Z}$. Thus, $N(\alpha) = a^2 + b^2$, so that $a^2 + b^2 = N(\alpha) = 3 \equiv 3 \mod 4$. This contradicts Exercise 2.7.2(c). This contradiction shows that our assumption was false. So 3 is a Gaussian prime.

2019-03-11 lecture

So we know that 3 is a Gaussian prime, but 2 is not. Is there a way to tell which integers are Gaussian primes, without checking all Gaussian divisors?

Let us first state a positive criterion:

**Lemma 4.2.39.** Let $p$ be a prime such that $p \equiv 3 \mod 4$. Then, $p$ is a Gaussian prime.

**Proof.** Assume the contrary. Thus, $p$ has a Gaussian divisor $\delta$ that is neither a unit nor unit-equivalent to $p$. Consider this $\delta$. Thus, there exists a Gaussian integer $\epsilon$ such that $p = \delta \epsilon$. Consider this $\epsilon$.

From $p = \delta \epsilon$, we obtain $N(p) = N(\delta \epsilon) = N(\delta) \cdot N(\epsilon)$. Hence, $N(\delta) \cdot N(\epsilon) = N(p) = p^2 + 0^2 = p^2$. Since $N(\delta)$ and $N(\epsilon)$ are nonnegative integers (and $p$ is prime), this leaves only three options:

- either $N(\delta) = 1$ and $N(\epsilon) = p^2$,
- or $N(\delta) = p$ and $N(\epsilon) = p$,
- or $N(\delta) = p^2$ and $N(\epsilon) = 1$.

The first of these three options would cause $\delta$ to be a unit, which is impossible (by the definition of $\delta$).

The third of these three options would cause $\delta$ to be unit-equivalent to $p$ (since $\epsilon$ would be a unit, and $p = \delta \epsilon$), which is impossible (by the definition of $\delta$).

Thus, the second of these three options must hold. In other words, $N(\delta) = p$ and $N(\epsilon) = p$. Now, write the Gaussian integer $\delta$ as $\delta = (a, b)$ with integers $a, b$. Then, $N(\delta) = a^2 + b^2$, so that $a^2 + b^2 = N(\delta) = p \equiv 3 \mod 4$. This contradicts Exercise 2.7.2(c). Thus, Lemma 4.2.39 is proven.
It is clear that no prime is divisible by 4. Thus, there are three types of primes:

- **Type 1**: Primes that are \(\equiv 1 \mod 4\): these are 5, 13, 17, 29, \ldots.
- **Type 2**: Primes that are even: there is only one of these, namely 2.
- **Type 3**: Primes that are \(\equiv 3 \mod 4\): these are 3, 7, 11, 19, 23, \ldots.

(One can show that there are infinitely many primes of Type 1 and infinitely many primes of Type 3. It can also be shown that there are “roughly the same amount” of Type-1 primes and of Type-3 primes “in theory”, but “in practice” the Type-3 primes are more frequent. For the concrete meaning of this weird paradoxical claim, google for “Chebyshev’s bias”.)

Lemma 4.2.39 says that all Type-3 primes are Gaussian primes. What about the other primes – are they Gaussian primes? We already know that 2 is not, since \(2 = (1 + i) (1 - i)\). Likewise, 5 is not, since \(5 = (1 + 2i) (1 - 2i)\). Likewise, 13 is not, since \(13 = (2 + 3i) (2 - 3i)\).

This may suggest that primes \(p\) satisfying \(p = 2\) or \(p \equiv 1 \mod 4\) (that is, primes of Type 1 or Type 2) not only factor nontrivially, but actually factor as

\[
p = (x + yi) (x - yi)
\]

for some integers \(x\) and \(y\).

Of course, this equation rewrites as \(p = x^2 + y^2\). Thus, we are back to asking Question 1.4.1, at least for primes.

We shall now answer this question, and actually prove a bit more:

**Theorem 4.2.40.** Let \(p\) be a prime such that either \(p = 2\) or \(p \equiv 1 \mod 4\).

(a) There exist integers \(x\) and \(y\) such that \(p = x^2 + y^2\).

(b) If \(p \equiv 1 \mod 4\), then there exist exactly 8 pairs \((x, y)\) of integers such that \(p = x^2 + y^2\). (For example, if \(p = 5\), then these 8 pairs are \((1, 2)\), \((2, 1)\), \((1, -2)\), \((-2, 1)\), \((-1, 2)\), \((2, -1)\), \((-1, -2)\) and \((-2, -1)\).)

(c) There exists a Gaussian prime \(\pi\) such that \(p = \pi \overline{\pi}\).

(d) The Gaussian integer \(p\) itself is not a Gaussian prime.

(e) Assume that \(p \equiv 1 \mod 4\). Consider the Gaussian prime \(\pi\) from Theorem 4.2.40 (c). Then, \(\overline{\pi}\) is also a Gaussian prime, and we do not have \(\pi \sim \overline{\pi}\).

For example, the Type-1 prime 17 satisfies

\[
17 = 1^2 + 4^2 = (1 + 4i) (1 - 4i) = (1 + 4i) (\overline{1 + 4i})
= (1 - 4i) (\overline{1 - 4i}) = (4 + i) (4 + i).
\]

Before we can prove Theorem 4.2.40, we will have to build up the theory of Gaussian primes a bit more. We first state the Gaussian-integer analogue of Proposition 2.13.5.
Proposition 4.2.41. Let $\pi$ be a Gaussian prime. Let $\alpha \in \mathbb{Z}[i]$. Then, either $\pi | \alpha$ or $\pi \perp \alpha$.

Proof of Proposition 4.2.41. Analogous to our proof of Proposition 2.13.5 above. $\square$

Next, we state the analogue to Theorem 2.13.6:

Theorem 4.2.42. Let $\pi$ be a Gaussian prime. Let $\alpha, \beta \in \mathbb{Z}[i]$ such that $\pi | \alpha \beta$. Then, $\pi | \alpha$ or $\pi | \beta$.

Proof of Theorem 4.2.42. Analogous to our proof of Theorem 2.13.6 above. $\square$

We also need the following simple facts:

Lemma 4.2.43. Let $\alpha$ be a Gaussian integer. If $N(\alpha)$ is prime, then $\alpha$ is a Gaussian prime.

This shows, for example, that $1 + i$ and $1 + 2i$ are Gaussian primes. The converse of Lemma 4.2.43 does not hold (e.g., since 3 is a Gaussian prime, but $N(3) = 9$ is not prime).

Proof of Lemma 4.2.43. Assume that $N(\alpha)$ is prime. We must prove that $\alpha$ is a Gaussian prime.

Assume the contrary. Thus, $\alpha$ is not a Gaussian prime, but $\alpha$ is neither zero nor a unit (since $N(\alpha)$ is prime and therefore $> 1$). Hence, $\alpha$ has a Gaussian divisor $\delta$ that is neither a unit nor unit-equivalent to $\alpha$. Consider this $\delta$. Thus, there exists a Gaussian integer $\epsilon$ such that $\alpha = \delta \epsilon$. Consider this $\epsilon$.

From $\alpha = \delta \epsilon$, we obtain $N(\alpha) = N(\delta \epsilon) = N(\delta) \cdot N(\epsilon)$. Hence, $N(\delta) \cdot N(\epsilon) = N(\alpha)$. Since $N(\delta)$ and $N(\epsilon)$ are nonnegative integers and $N(\alpha)$ is prime, this leaves only two options:

- either $N(\delta) = 1$ and $N(\epsilon) = N(\alpha)$,
- or $N(\delta) = N(\alpha)$ and $N(\epsilon) = 1$.

The first of these two options would cause $\delta$ to be a unit, which is impossible (by the definition of $\delta$). Thus, the second option must be true. In other words, we have $N(\delta) = N(\alpha)$ and $N(\epsilon) = 1$. But $N(\epsilon) = 1$ shows that $\epsilon$ is a unit, and thus $\alpha \sim \delta$ (since $\alpha = \delta \epsilon$). This contradicts the assumption that $\delta$ is not unit-equivalent to $\alpha$. This contradiction proves that our assumption was wrong. Lemma 4.2.43 is proven. $\square$

Lemma 4.2.44. Let $\pi$ be a Gaussian prime. Then, $\overline{\pi}$ is a Gaussian prime, too.
Proof. Easy: Conjugation reflects everything across the x-axis, without changing any properties. For example: If \( \delta \) is a Gaussian divisor of \( \pi \), then \( \overline{\delta} \) is a Gaussian divisor of \( \overline{\pi} \), and vice versa. If \( \delta \) is not a unit, then \( \overline{\delta} \) is not a unit, and vice versa. If \( \delta \) is not unit-equivalent to \( \pi \), then \( \overline{\delta} \) is not unit-equivalent to \( \overline{\pi} \), and vice versa. If \( \pi \) is nonzero and not a unit, then \( \overline{\pi} \) is nonzero and not a unit, and vice versa. Thus, if you compare what it means for \( \pi \) to be a Gaussian prime with what it means for \( \overline{\pi} \) to be a Gaussian prime, then you will see that it means the same thing. \( \square \)

Now, we can prove Theorem 4.2.40.

Proof of Theorem 4.2.40. \( \textbf{(d)} \) Assume the contrary. Thus, \( p \) is a Gaussian prime. But 2 is not a Gaussian prime (by Example 4.2.37). Hence, \( p \not\equiv 2 \mod 4 \) (since we assumed that either \( p = 2 \) or \( p \equiv 1 \mod 4 \)). Therefore, \( p = 2k + 1 \) for some even \( k \in \mathbb{N} \). Consider this \( k \).

Exercise 2.15.5 yields \( k^2 \equiv -1 \mod p \). Set \( u = k! \); thus, this becomes \( u^2 \equiv -1 \mod p \). In other words,

\[
p \mid u^2 - (-1) = u^2 - i^2 = (u + i)(u - i).
\]

This is a divisibility in \( \mathbb{Z} \), thus also a divisibility in \( \mathbb{Z}[i] \).

Hence, Theorem 4.2.42 (applied to \( \pi = p \), \( \alpha = u + i \) and \( \beta = u - i \)) yields that \( p \mid u + i \) or \( p \mid u - i \) (since \( p \) is a Gaussian prime). But if \( p \mid u - i \), then \( p \mid u + i \) holds as well (since Exercise 4.2.5 shows that \( p \mid u - i \) implies \( \overline{p} \mid u - i = u + i \), which means \( p = \overline{p} \mid u + i \)). Hence, we have \( p \mid u + i \) in both cases.

This means that there exists a Gaussian integer \( \gamma \) such that \( u + i = p \gamma \). Consider this \( \gamma \). Write \( \gamma \) as \( \gamma = (a, b) \) with \( a, b \in \mathbb{Z} \). Then, \( (u, 1) = u + i = p \gamma = (pa, pb) \). Thus, \( u = pa \) and \( 1 = pb \). But \( 1 = pb \) leads to \( p \mid 1 \) in \( \mathbb{Z} \) (since \( b \in \mathbb{Z} \)), which is absurd (since \( p \) is prime). This contradiction shows that our assumption was wrong. Thus, Theorem 4.2.40 (d) is proven.

\( \textbf{(a)} \) We have \( N(p) = p^2 + 0^2 = p^2 > 1 \) (since \( p > 1 \)). Thus, \( p \) is nonzero and not a unit. But Theorem 4.2.40 (d) shows that \( p \) is not a Gaussian prime. Since \( p \) is nonzero and not a unit, this shows that \( p \) has a Gaussian divisor \( \delta \) that is neither a unit nor unit-equivalent to \( p \). Consider this \( \delta \). Thus, there exists a Gaussian integer \( \epsilon \) such that \( p = \delta \epsilon \). Consider this \( \epsilon \).

From \( p = \delta \epsilon \), we obtain \( N(p) = N(\delta \epsilon) = N(\delta) \cdot N(\epsilon) \). Hence, \( N(\delta) \cdot N(\epsilon) = N(p) = p^2 + 0^2 = p^2 \). Since \( N(\delta) \) and \( N(\epsilon) \) are nonnegative integers (and \( p \) is prime), this leaves only three options:

- \( \textbf{either} \ N(\delta) = 1 \) and \( N(\epsilon) = p^2 \),
- \( \textbf{or} \ N(\delta) = p \) and \( N(\epsilon) = p \),
- \( \textbf{or} \ N(\delta) = p^2 \) and \( N(\epsilon) = 1 \).
The first of these three options would cause \( \delta \) to be a unit, which is impossible (by the definition of \( \delta \)).

The third of these three options would cause \( \delta \) to be unit-equivalent to \( p \) (since \( \varepsilon \) would be a unit, and \( p = \delta \varepsilon \)), which is impossible (by the definition of \( \delta \)).

Thus, the second of these three options must hold. In other words, \( N(\delta) = p \) and \( N(\varepsilon) = p \). Now, write the Gaussian integer \( \delta \) as \( \delta = (a,b) \) with integers \( a, b \).

Then, \( N(\delta) = a^2 + b^2 \), so that \( a^2 + b^2 = N(\delta) = p \). Hence, there exist integers \( x \) and \( y \) such that \( p = x^2 + y^2 \) (namely, \( x = a \) and \( y = b \)). This proves Theorem 4.2.40 (a).

(c) Theorem 4.2.40 (a) shows that there exist integers \( x \) and \( y \) such that \( p = x^2 + y^2 \). Consider these \( x \) and \( y \). Let \( \pi \) be the Gaussian integer \( x + iy \). Then, \( \pi \bar{\pi} = (x + iy)(x - iy) = x^2 + y^2 = p \). Thus, \( p = \pi \bar{\pi} \). It remains to prove that \( \pi \) is a Gaussian prime.

The norm of \( \pi \) is \( N(\pi) = x^2 + y^2 = p \), which is prime. Hence, Lemma 4.2.43 (applied to \( \alpha = \pi \)) shows that \( \pi \) is a Gaussian prime. This completes the proof of Theorem 4.2.40 (c).

(b) Assume that \( p \equiv 1 \mod 4 \). We must prove that there exist exactly 8 pairs \((x, y)\) of integers such that \( p = x^2 + y^2 \).

One such pair is provided by Theorem 4.2.40 (a). Let us call it \((a, b)\). To get the other 7, we notice that it must satisfy \( a \neq 0 \) (since \( p \) is not a perfect square) and \( b \neq 0 \) (for the same reason) and \( a \neq b \) (since \( p \) is not \( 2n^2 \) for any \( n \in \mathbb{Z} \)), and thus it leads to 7 other pairs

\[
(b, a), \quad (a, -b), \quad (-b, a), \quad (-a, b), \quad (b, -a), \quad (-a, -b), \quad (-b, -a),
\]

which are all distinct. It thus remains to prove that these altogether 8 pairs are the only pairs \((x, y)\) of integers such that \( p = x^2 + y^2 \).

In other words, we need to prove that if \((x, y)\) is a pair of integers such that \( p = x^2 + y^2 \), then \((x, y)\) is one of the above 8 pairs. So let us fix a pair \((x, y)\) of integers such that \( p = x^2 + y^2 \). We must prove that \((x, y)\) is one of the above 8 pairs. In other words, we must prove that \((x, y)\) equals \((a, b)\) up to order and signs. This is equivalent to proving that \( x + yi \sim a + bi \) or \( x - yi \sim a + bi \).

Set \( \pi = x + yi \) and \( \alpha = a + bi \). Then, \( \pi \) and \( \alpha \) are Gaussian integers having norms \( N(\pi) = x^2 + y^2 = p \) and \( N(\alpha) = a^2 + b^2 = p \) (by the definition of \((a, b)\)).

Thus, \( N(\alpha) = p \) is prime. Hence, Lemma 4.2.43 shows that \( \alpha \) is a Gaussian prime. Similarly, \( \pi \) is a Gaussian prime.

We must prove that \( x + yi \sim a + bi \) or \( x - yi \sim a + bi \). In other words, we must prove that \( \pi \sim \alpha \) or \( \bar{\pi} \sim \alpha \) (since \( x + yi = \pi \) and \( x - yi = \bar{\pi} \) and \( a + bi = \alpha \)).

Now,

\[
\alpha = a + bi \quad | \quad (a + bi)(a - bi) = a^2 + b^2 = p = N(\pi) = \pi \bar{\pi}.
\]

Since \( \alpha \) is a Gaussian prime, this yields that \( \alpha \mid \pi \) or \( \alpha \mid \bar{\pi} \) (by Theorem 4.2.42). Thus, we are in one of the following two Cases:

Case 1: We have \( \alpha \mid \pi \).
Case 2: We have \( \alpha | \overline{\pi} \).

In Case 1, we have \( \alpha | \pi \). In other words, there exists a Gaussian integer \( \zeta \) such that \( \pi = \alpha \zeta \). Consider this \( \zeta \). We have \( \pi = \alpha \zeta \), thus \( N(\pi) = N(\alpha \zeta) = \alpha N(\zeta) \).

Since both \( N(\pi) \) and \( N(\alpha) \) are \( p \) (because \( N(\pi) = x^2 + y^2 = p \) and \( N(\alpha) = a^2 + b^2 = p \)), this rewrites as \( p = p N(\zeta) \). We can cancel \( p \) from this equality, and obtain \( N(\zeta) = 1 \). Hence, \( \zeta \) is a unit. Therefore, \( \pi = \alpha \zeta \) yields \( \pi \sim \alpha \). In other words, \( x + yi \sim a + bi \) (since \( \pi = x + yi \) and \( \alpha = a + bi \)).

In Case 2, we similarly obtain \( x - yi \sim a + bi \) (since \( N(\pi) = N(\pi) = x^2 + y^2 = p \) and \( \overline{\pi} = x - yi \)).

Hence, in both Cases, we have proven that \( x + yi \sim a + bi \) or \( x - yi \sim a + bi \). This completes our proof of Theorem 4.2.40 (b).

(e) Conjugation is a symmetry which preserves divisibility (by Exercise 4.2.5). Thus, since \( \pi \) is a Gaussian prime, its conjugate \( \overline{\pi} \) is a Gaussian prime as well. It remains to prove that we do not have \( \pi \sim \overline{\pi} \).

Assume the contrary. Thus, \( \pi \sim \overline{\pi} \). Write the Gaussian prime \( \pi \) in the form \( \pi = x + yi \) for some \( x, y \in \mathbb{Z} \). Thus, \( \pi = x - yi \) and \( N(\pi) = x^2 + y^2 \); therefore, \( p = \pi\overline{\pi} = N(\pi) = x^2 + y^2 \). But \( \pi \sim \overline{\pi} \). Hence, there exists a unit \( \gamma \) such that \( \overline{\pi} = \gamma \pi \). Consider this \( \gamma \). Since \( \gamma \) is a unit, we have \( \gamma \in \{1, -1, i, -i\} \) (by Proposition 4.2.10). So we are in one of the following four cases:

Case 1: We have \( \gamma = 1 \).

Case 2: We have \( \gamma = -1 \).

Case 3: We have \( \gamma = i \).

Case 4: We have \( \gamma = -i \).

Let us first consider Case 1. In this case, we have \( \gamma = 1 \). Now, \( x - yi = \pi = \overline{\pi} = x + yi \). Hence, \( y = 0 \). Now, \( p = x^2 + y^2 = x^2 \) (since \( y = 0 \)). This contradicts the fact that \( p \) (being a prime) cannot be a square.

Let us first consider Case 2. In this case, we have \( \gamma = -1 \). Now, \( x - yi = \pi = \overline{\pi} = x + yi \). Hence, \( x = 0 \). Now, \( p = x^2 + y^2 = y^2 \) (since \( x = 0 \)). This contradicts the fact that \( p \) (being a prime) cannot be a square.

Let us first consider Case 3. In this case, we have \( \gamma = i \). Now, \( x - yi = \pi = \overline{\pi} = i(x + yi) = -y + xi \). Hence, \( x = -y \). Now, \( p = x^2 + y^2 = 2y^2 \) (since \( x = -y \)). This contradicts the fact that \( p \) is odd (since \( p \equiv 1 \mod 4 \)).

Let us first consider Case 4. In this case, we have \( \gamma = -i \). Now, \( x - yi = \pi = \overline{\pi} = -i(x + yi) = y - xi \). Hence, \( x = y \). Now, \( p = x^2 + y^2 = 2y^2 \) (since \( x = y \)). This contradicts the fact that \( p \) is odd (since \( p \equiv 1 \mod 4 \)).

We have thus obtained a contradiction in each of our four cases. Hence, we always get a contradiction. Thus, the proof of Theorem 4.2.40 (e) is complete.

We have thus answered Question 1.4.2 (b) in the case when \( n \) is a prime: We have shown that \( p \) is a sum of two perfect squares if and only if either \( p = 2 \) or
$p \equiv 1 \mod 4$; and we have shown that the number of pairs $(x, y) \in \mathbb{Z}^2$ satisfying $p = x^2 + y^2$ is 8 when $p \equiv 1 \mod 4$ and is 4 when $p = 2$ (the latter claim is easy to check).

What about the case of arbitrary $n$?

For $n = 21$, we have $n \equiv 1 \mod 4$, but $n$ is not a sum of two perfect squares. So the answer we gave for the case of prime $n$ does not generalize to arbitrary $n$.

2019-03-13 lecture

It turns out that the right answer for arbitrary $n$ will come from the analogue of prime factorization in $\mathbb{Z}[i]$.

**Proposition 4.2.45.** Let $\nu$ be a nonzero Gaussian integer that is not a unit. Then, there exists at least one Gaussian prime $\pi$ such that $\pi | \nu$.

*Proof of Proposition 4.2.45* This is an analogue of Proposition 2.13.8, and can be proven in the same way. Just replace $d$ (the smallest positive divisor of $n$) by $\delta$ (a Gaussian divisor of $\nu$ that is not a unit and has the smallest norm among all such divisors).

**Proposition 4.2.46.** Let $\nu$ be a nonzero Gaussian integer. Then, $\nu$ is unit-equivalent to a certain product of finitely many Gaussian primes.

*Proof of Proposition 4.2.46* This is an analogue of Proposition 2.13.10, and can be proven in the same way: Strong induction on $N(\nu)$. The main difference is that the case of $N(\nu) = 1$ leads to $\nu$ being a unit (hence unit-equivalent to an empty product of Gaussian primes) rather than $\nu$ being 1.

**Definition 4.2.47.** Let $\nu$ be a nonzero Gaussian integer. A **Gaussian prime factorization** of $\nu$ means a tuple $(\pi_1, \pi_2, \ldots, \pi_k)$ of Gaussian primes such that $\nu \sim \pi_1 \pi_2 \cdots \pi_k$.

Why did we require only $\nu \sim \pi_1 \pi_2 \cdots \pi_k$ and not $\nu = \pi_1 \pi_2 \cdots \pi_k$? Because we want $-1$ to have a Gaussian prime factorization, but there is no way to literally write $-1$ as a product of Gaussian primes.

**Exercise 4.2.6.** Let $\pi$ and $\kappa$ be two Gaussian primes that do not satisfy $\pi \sim \kappa$. Prove that $\pi \perp \kappa$.

*Solution sketch.* This is an analogue of Exercise 2.13.1, and its solution goes accordingly.

**Lemma 4.2.48.** Let $\pi$ be a Gaussian prime. Let $\alpha$ be a nonzero Gaussian integer. Then, there exists a largest $m \in \mathbb{N}$ such that $\pi^m | \alpha$. 
**Proof of Lemma 4.2.48.** This is an analogue of Lemma 2.13.22 for Gaussian integers (with \( \pi \) and \( \alpha \) playing the roles of \( p \) and \( n \)), and the proof also proceeds similarly. Here are the main differences: Instead of \( p > 1 \), we now have \( N(\pi) > 1 \) (which is because \( \pi \) is nonzero and not a unit). Again, let \( W \) be the set of all \( m \in \mathbb{N} \) satisfying \( \pi^m | \alpha \). Then, \( W \) is a nonempty set of integers (this is proven as in the proof of Lemma 2.13.22). Let \( u = N(\alpha) \). Thus, \( u \in \mathbb{N} \). It is easy to see that \( N(\pi^k) > k \) for each \( k \in \mathbb{N} \) (indeed, Corollary 4.1.27(b) yields \( N(\pi^k) = (N(\pi))^k > k \) by Exercise 2.13.4 (applied to \( p = N(\pi) \)). From this point, we proceed similarly as in the proof of Lemma 2.13.22.

Similarly to Definition 2.13.23, we can define \( \pi \)-adic valuations:

**Definition 4.2.49.** Let \( \pi \) be a Gaussian prime.

(a) Let \( \alpha \) be a nonzero Gaussian integer. Then, \( v_\pi(\alpha) \) shall denote the largest \( m \in \mathbb{N} \) such that \( \pi^m | \alpha \). This is well-defined (by Lemma 4.2.48). This nonnegative integer \( v_\pi(\alpha) \) will be called the \( \pi \)-valuation (or the \( \pi \)-adic valuation) of \( \alpha \).

(b) We extend this definition of \( v_\pi(\alpha) \) to the case of \( \alpha = 0 \) as follows: Set \( v_\pi(0) = \infty \).

Definition 4.2.49 does not conflict with Definition 2.13.23. Indeed, if a prime \( p \) happens to also be a Gaussian prime, and if \( n \) is an integer, then both definitions yield the same value of \( v_p(n) \) (since \( p^m | a \) means the same thing whether we treat \( p \) and \( a \) as integers or as Gaussian integers).

**Theorem 4.2.50.** Let \( \pi \) be a Gaussian prime.

(a) We have \( v_\pi(\alpha\beta) = v_\pi(\alpha) + v_\pi(\beta) \) for any two Gaussian integers \( \alpha \) and \( \beta \).

(b) We have \( v_\pi(\alpha + \beta) \geq \min\{v_\pi(\alpha), v_\pi(\beta)\} \) for any two Gaussian integers \( \alpha \) and \( \beta \).

(c) We have \( v_\pi(1) = 1 \). More generally, \( v_\pi(\alpha) = 0 \) for any unit \( \alpha \in \mathbb{Z}[i] \).

(d) We have \( v_\pi(\kappa) = \begin{cases} 1, & \text{if } \kappa \sim \pi; \\ 0, & \text{otherwise} \end{cases} \) for any Gaussian prime \( \kappa \).

**Proof.** This is an analogue of Theorem 2.13.28 and is proven similarly.

**Proposition 4.2.51.** Let \( v \) be a nonzero Gaussian integer. Let \( (\alpha_1, \alpha_2, \ldots, \alpha_k) \) be a Gaussian prime factorization of \( v \). Let \( \pi \) be a Gaussian prime. Then,

\[
(\text{the number of times a Gaussian integer unit-equivalent to } \pi \\
\text{appears in the tuple } (\alpha_1, \alpha_2, \ldots, \alpha_k))
\]

\[
= (\text{the number of times } [\pi]_\sim \text{ appears in the tuple } ([\alpha_1]_\sim, [\alpha_2]_\sim, \ldots, [\alpha_k]_\sim))
\]

\[
= (\text{the number of } i \in \{1, 2, \ldots, k\} \text{ such that } \alpha_i \sim \pi)
\]

\[
= (\text{the number of } i \in \{1, 2, \ldots, k\} \text{ such that } [\alpha_i]_\sim = [\pi]_\sim)
\]

\[
= v_\pi(v).
\]
Proof. This is an analogue of Proposition 2.13.30 and is proven similarly. \qed

**Theorem 4.2.52.** Let $\nu$ be a nonzero Gaussian integer.

(a) There exists a Gaussian prime factorization of $\nu$.

(b) Any two such factorizations differ only by reordering their entries and multiplying them by units. More precisely: If $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ and $(\beta_1, \beta_2, \ldots, \beta_\ell)$ are two Gaussian prime factorizations of $\nu$, then $(\lfloor \alpha_1 \rfloor, \lfloor \alpha_2 \rfloor, \ldots, \lfloor \alpha_k \rfloor)$ is a permutation of $(\lfloor \beta_1 \rfloor, \lfloor \beta_2 \rfloor, \ldots, \lfloor \beta_\ell \rfloor)$.

**Proof.** This is an analogue of Theorem 2.13.31 and is proven similarly. \qed

**Example 4.2.53.** We have $5 = (1 + 2i) (1 - 2i) = (2 + i) (2 - i)$.

Thus, both $(1 + 2i, 1 - 2i)$ and $(2 + i, 2 - i)$ are Gaussian prime factorizations of 5. They may look different, but actually you get the second one from the first by swapping the two entries and multiplying the first entry by the unit $i$ and multiplying the second entry by the unit $-i$. This perfectly agrees with Theorem 4.2.52.

In analogy to Exercise 2.13.5 (and with the same proof), we have:

**Exercise 4.2.7.** Let $\pi$ be a Gaussian prime. Let $\alpha, \beta \in \mathbb{Z}[i]$ be such that $\alpha \sim \beta$. Prove that $v_\pi(\alpha) = v_\pi(\beta)$.

We also have the following:

**Exercise 4.2.8.** Let $\pi$ be a Gaussian prime. Let $\alpha \in \mathbb{Z}[i]$. Then, $\overline{\pi}$ is a Gaussian prime as well, and satisfies $v_\pi(\overline{\alpha}) = v_\pi(\alpha)$.

**Solution sketch.** Conjugation is a symmetry, taking Gaussian integers to Gaussian integers and preserving divisibility. Thus, any $i \in \mathbb{N}$ satisfying $\pi^i | \alpha$ must also satisfy $\pi^i | \overline{\alpha}$, and vice versa. Hence, $v_\pi(\overline{\alpha}) = v_\pi(\alpha)$ easily follows. \qed

**Definition 4.2.54.** For the rest of this section, let $GP$ be the set of all Gaussian primes of the form $x + yi$ with $x \in \{1, 2, 3, \ldots\}$ and $y \in \{0, 1, 2, \ldots\}$.

The following is easy to see:

**Lemma 4.2.55.** Let $\pi$ be a Gaussian prime. Then, there exists exactly one $\sigma \in GP$ such that $\pi \sim \sigma$.

In other words, each Gaussian prime is unit-equivalent to exactly one $\sigma \in GP$. Thus, the set $GP$ contains exactly one element of each unit-equivalence class of Gaussian primes. (Thus, $GP$ is what is called a “system of distinct representatives” for the unit-equivalence classes of all Gaussian primes.)

In analogy to Corollary 2.13.34, we have:
Corollary 4.2.56. Let $\alpha$ be a nonzero Gaussian integer. Then,

$$\alpha \sim \prod_{\pi \in \text{GP}} \pi^{v_\pi(\alpha)}.$$ 

Here, the infinite product $\prod_{\pi \in \text{GP}} \pi^{v_\pi(\alpha)}$ is well-defined (according to the Gaussian-integer analogue of Lemma 2.13.32 (b)).

In analogy to Proposition 2.13.35, we have the following:

Proposition 4.2.57. Let $\alpha$ and $\beta$ be Gaussian integers. Then, $\alpha \mid \beta$ if and only if each Gaussian prime $\pi$ satisfies $v_\pi(\alpha) \leq v_\pi(\beta)$.

If $\alpha$ is a Gaussian integer, and $c$ is a unit-equivalence class of Gaussian integers, then either all elements of $c$ divide $\alpha$ or none of them does.

134 Thus, we can talk of unit-equivalence classes of Gaussian divisors of $\alpha$ (by which we mean unit-equivalence classes of Gaussian integers whose elements all divide $\alpha$).

Here is an analogue of Proposition 2.18.1 for Gaussian integers:

Proposition 4.2.58. Let $\alpha \in \mathbb{Z}[i]$ be a nonzero Gaussian integer. Then:

(a) The product $\prod_{\pi \in \text{GP}} (v_\pi(\alpha) + 1)$ is well-defined, since all but finitely many of its factors are 1.

(b) We have

\[
= \prod_{\pi \in \text{GP}} (v_\pi(\alpha) + 1).
\]

(c) We have

\[
= 4 \cdot \prod_{\pi \in \text{GP}} (v_\pi(\alpha) + 1).
\]

Proof. Same proof as for Proposition 2.18.1, but you have to be more careful with unit-equivalence (since in part (b), you are counting unit-equivalence classes rather than positive divisors). The analogue of Lemma 2.18.3 we need to use for this proof is the following lemma:

Lemma 4.2.59. Let $\pi_1, \pi_2, \ldots, \pi_u$ be finitely many Gaussian primes, no two of which are unit-equivalent. For each $i \in \{1, 2, \ldots, u\}$, let $a_i$ be a nonnegative integer. Let $\alpha = \pi_1^{a_1} \pi_2^{a_2} \cdots \pi_u^{a_u}$.

134 This is easy to check. Indeed, it boils down to the fact that any two elements of $c$ divide each other (because they are unit-equivalent).
Define a set $T$ by

$$T = \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\} = \{(b_1, b_2, \ldots, b_u) \mid b_i \in \{0, 1, \ldots, a_i\} \text{ for each } i \in \{1, 2, \ldots, u\}\} = \{(b_1, b_2, \ldots, b_u) \in \mathbb{N}^u \mid b_i \leq a_i \text{ for each } i \in \{1, 2, \ldots, u\}\}.$$ 

Then, the map

$$\Lambda : T \to \{\text{unit-equivalence classes of Gaussian divisors of } \alpha\},$$

$$(b_1, b_2, \ldots, b_u) \mapsto \left[\pi_1^{b_1} \pi_2^{b_2} \cdots \pi_u^{b_u}\right]_\sim$$

is well-defined and bijective.

Now, we can finally answer Question 1.4.2(b) (following [DumFoo04, §8.3, Corollary 19]):

**Theorem 4.2.60.** Let $n$ be a positive integer.

(a) If there is at least one prime $p \equiv 3 \mod 4$ such that $v_p(n)$ is odd, then there is no pair $(x, y) \in \mathbb{Z}^2$ such that $n = x^2 + y^2$.

(b) Assume that for each prime $p \equiv 3 \mod 4$, the number $v_p(n)$ is even. Then,

$$\left(\text{the number of pairs } (x, y) \in \mathbb{Z}^2 \text{ such that } n = x^2 + y^2\right) = 4 \cdot \prod_{\substack{p \text{ prime; } p \equiv 1 \mod 4 \atop v_p(n) + 1}} v_p(n).$$

**Example 4.2.61.** (a) Let $n = 35$. Then, Theorem 4.2.60(a) yields that there are no integers $x$ and $y$ such that $n = x^2 + y^2$. In fact, the prime $7 \equiv 3 \mod 4$ satisfies $v_7(n) = 1$.

(b) Let $n = 45$. Then, for each prime $p \equiv 3 \mod 4$, the number $v_p(n)$ is even. Indeed, $n = 45 = 3^2 \cdot 5$, so $v_3(n) = 2$ is even and $v_p(n) = 0$ for all other primes $p$ of Type 3. Hence, Theorem 4.2.60(b) yields

$$\left(\text{the number of pairs } (x, y) \in \mathbb{Z}^2 \text{ such that } n = x^2 + y^2\right) = 4 \cdot \prod_{\substack{p \text{ prime; } p \equiv 1 \mod 4 \atop v_5(n) + 1}} v_p(n) + 1 = 4 \cdot 2 = 8.$$
Proof of Theorem 4.2.60 (sketched). (a) Assume that there is at least one prime $p \equiv 3 \mod 4$ such that $v_p(n)$ is odd. We must prove that there is no pair $(x, y) \in \mathbb{Z}^2$ such that $n = x^2 + y^2$.

Indeed, let $(x, y) \in \mathbb{Z}^2$ be a pair such that $n = x^2 + y^2$. We must derive a contradiction.

Let $\alpha$ be the Gaussian integer $x + yi$. Thus, $\alpha \overline{\alpha} = x^2 + y^2 = n$.

We have assumed that there is at least one prime $p \equiv 3 \mod 4$ such that $v_p(n)$ is odd. Consider this $p$. Note that $p$ is a Gaussian prime (by Lemma 4.2.39).

Thus, Exercise 4.2.8 (applied to $\pi = p$) yields $v_p(\overline{\alpha}) = v_p(\alpha)$. In view of $\overline{p} = p$, this rewrites as $v_p(\overline{\alpha}) = v_p(\alpha)$. But

$$v_p \left( n \right) = v_p(\alpha \overline{\alpha}) = v_p(\alpha) + v_p(\overline{\alpha}) = v_p(\alpha) + v_p(\alpha) = 2v_p(\alpha).$$

Thus, $v_p(n)$ is even. This contradicts the fact that $v_p(n)$ is odd. Thus, we have found a contradiction for each pair $(x, y) \in \mathbb{Z}^2$ such that $n = x^2 + y^2$. Hence, there exists no such pair. This proves Theorem 4.2.60 (a).

(b) We have

$$(\text{the number of } \alpha \in \mathbb{Z}[i] \text{ such that } n = \alpha \overline{\alpha})$$

$$= \left( \text{the number of pairs } (x, y) \in \mathbb{Z}^2 \text{ such that } n = x^2 + y^2 \right),$$

since the map

$$\{(x, y) \in \mathbb{Z}^2 \mid n = x^2 + y^2\} \to \{\alpha \in \mathbb{Z}[i] \mid n = \alpha \overline{\alpha}\},$$

$$(x, y) \mapsto x + yi$$

is a bijection.

Write the canonical factorization of $n$ as

$$n = 2^c \cdot p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k} q_1^{b_1} q_2^{b_2} \cdots q_{\ell}^{b_{\ell}},$$

(140)

where all the exponents $c, a_i, b_j \in \mathbb{N}$, and where $p_1, p_2, \ldots, p_k$ are distinct primes of Type 1, and where $q_1, q_2, \ldots, q_{\ell}$ are distinct primes of Type 3. Note that $c = 0$ if $n$ is odd.

We have assumed that for each prime $p \equiv 3 \mod 4$, the number $v_p(n)$ is even. In other words, for each prime $p$ of Type 3, the number $v_p(n)$ is even. In other words, $b_1, b_2, \ldots, b_{\ell}$ are even (since $q_1, q_2, \ldots, q_{\ell}$ are primes of Type 3, and thus the corresponding exponents $b_j = v_{q_j}(n)$ must be even). Hence, $b_j/2 \in \mathbb{N}$ for each $j$.

Each $q_j$ is a prime of Type 3, and thus is a Gaussian prime (by Lemma 4.2.39). Meanwhile, each $p_h$ is a prime of Type 1, and thus can be written in the form $p_h = \pi_h \overline{\pi}_h$ for some Gaussian prime $\pi_h$ (by Theorem 4.2.40 (c)). Consider these
For every $h$, Theorem 4.2.40 (e) shows that the conjugate $\overline{\pi_h}$ is also a Gaussian prime, and that we do not have $\pi_h \sim \overline{\pi_h}$.

Finally, let $\rho$ be the Gaussian prime $1 + i$; thus $2 = \rho \overline{\rho}$. But note that $\rho \sim \overline{\rho}$ (indeed, $\overline{\rho} = 1 - i = (-i)(1+i) = (-i)\rho$). Now, (140) becomes

\[ n = \frac{2^c}{\rho^c \overline{\rho}^c} \cdot \left( \prod_{h=1}^{k} \left( \frac{p_h^{a_h}}{\pi_h^{d_h} \pi_{\overline{h}}^{\overline{a_h}}} \right) \cdot q_1^{b_1} q_2^{b_2} \cdots q_\ell^{b_\ell} \right) \]

and this is a decomposition of $n$ as a product of powers of Gaussian primes (albeit $\rho$ and $\overline{\rho}$ are unit-equivalent).

No two of the Gaussian primes $\rho, \pi_h, \pi_{\overline{h}}, q_j$ are unit-equivalent. (Proof: Compare their norms (since unit-equivalent Gaussian integers have equal norms). The only of these Gaussian primes that have equal norms are $\pi_h$ and $\overline{\pi_h}$. So we merely need to rule out $\pi_h \sim \overline{\pi_h}$. But this is clear, since we already showed that we do not have $\pi_h \sim \overline{\pi_h}$.)

Now, define a map

\[ F : \{1,i,-1,-i\} \times \prod_{h=1}^{k} \{0,1,\ldots,a_h\} \to \{\alpha \in \mathbb{Z}[i] \mid n = \alpha \overline{\alpha}\}, \]

\[ (\gamma, (d_1,d_2,\ldots,d_k)) \mapsto \gamma \cdot \rho^c \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \overline{\pi}_{\overline{h}}^{\overline{d}_h}}{\pi_{\overline{h}}^{a_h} \pi_h^{a_{\overline{h}}}} \right) \cdot q_1^{b_1/2} q_2^{b_2/2} \cdots q_\ell^{b_\ell/2} \right). \]

It is easy to check that this map $F$ is well-defined\(^{135}\). We claim that this map $F$ is a bijection.

Proof: To see that $F$ is injective, we must find a way to reconstruct $(\gamma, (d_1,d_2,\ldots,d_k)) \in \{1,i,-1,-i\} \times \prod_{h=1}^{k} \{0,1,\ldots,a_h\}$ from

\[ \alpha : = F ((\gamma, (d_1,d_2,\ldots,d_k))) \]

\[ = \gamma \cdot \rho^c \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \overline{\pi}_{\overline{h}}^{\overline{d}_h}}{\pi_{\overline{h}}^{a_h} \pi_h^{a_{\overline{h}}}} \right) \cdot q_1^{b_1/2} q_2^{b_2/2} \cdots q_\ell^{b_\ell/2} \right). \]

\(^{135}\)Just multiply out $\alpha \overline{\alpha}$ for $\alpha = \gamma \cdot \rho^c \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \overline{\pi}_{\overline{h}}^{\overline{d}_h}}{\pi_{\overline{h}}^{a_h} \pi_h^{a_{\overline{h}}}} \right) \cdot q_1^{b_1/2} q_2^{b_2/2} \cdots q_\ell^{b_\ell/2} \right)$ and check that you obtain $n$. Corollary 4.1.27 (a) needs to be used.
But this is easy: You first reconstruct the \(k\)-tuple \((d_1, d_2, \ldots, d_k)\) by observing that 
\[ d_h = v \pi_h(\alpha) \]
for each \(h\). Once you have that, you can reconstruct \(\gamma\) by
\[
\gamma = \frac{\alpha}{\rho^c \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \pi_h^{d_h} - d_h}{\pi_h^{d_h} \pi_h^{d_h}} \right) \right) \cdot q_1^{b_1/2} q_2^{b_2/2} \cdots q_{\ell}^{b_{\ell}/2}}.
\]
So \(F\) is injective.

To see that \(F\) is surjective, we must prove that each \(\alpha \in \mathbb{Z}[i]\) satisfying 
\[ u = \alpha \bar{\alpha} \]
has the form
\[
\alpha = \gamma \cdot \rho^c \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \pi_h^{d_h} - d_h}{\pi_h^{d_h} \pi_h^{d_h}} \right) \right) \cdot q_1^{b_1/2} q_2^{b_2/2} \cdots q_{\ell}^{b_{\ell}/2}
\]
for some \((\gamma, (d_1, d_2, \ldots, d_k))\). To prove this, use canonical factorization of \(\alpha\) inside 
\(\mathbb{Z}[i]\) to see that
\[
\alpha = \gamma \cdot \rho^c \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \pi_h^{d_h} - d_h}{\pi_h^{d_h} \pi_h^{d_h}} \right) \right) \cdot q_1^{b_1/2} q_2^{b_2/2} \cdots q_{\ell}^{b_{\ell}/2}
\]
(142)
for some \(c', d'_h, c'_h, b'_{j} \in \mathbb{N}\). Consider these \(c', d'_h, c'_h, b'_{j} \in \mathbb{N}\). From (142), we obtain
\[
\alpha \bar{\alpha} = \gamma \cdot \rho^c \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \pi_h^{d_h} - d_h}{\pi_h^{d_h} \pi_h^{d_h}} \right) \right) \cdot q_1^{b_1/2} q_2^{b_2/2} \cdots q_{\ell}^{b_{\ell}/2}
\]
\
\[
\begin{align*}
\alpha \bar{\alpha} &= \gamma \cdot \rho^c \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \pi_h^{d_h} - d_h}{\pi_h^{d_h} \pi_h^{d_h}} \right) \right) \cdot \prod_{h=1}^{k} \left( \frac{\pi_h^{d'_h} \pi_h^{d'_h} - d'_h}{\pi_h^{d'_h} \pi_h^{d'_h}} \right) \cdot q_1^{b'_1} q_2^{b'_2} \cdots q_{\ell}^{b'_{\ell}} \\
&= \gamma \cdot \rho^c \cdot \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \pi_h^{d_h} - d_h}{\pi_h^{d_h} \pi_h^{d_h}} \right) \cdot q_1^{b_1/2} q_2^{b_2/2} \cdots q_{\ell}^{b_{\ell}/2} \\
&= \gamma \cdot \rho^c \cdot \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \pi_h^{d_h} - d_h}{\pi_h^{d_h} \pi_h^{d_h}} \right) \cdot q_1^{b'_1} q_2^{b'_2} \cdots q_{\ell}^{b'_{\ell}}
\end{align*}
\]
(143)
Thus,
\[
\begin{align*}
n &= \alpha \bar{\alpha} = 2^{c'} \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \pi_h^{d_h} - d_h}{\pi_h^{d_h} \pi_h^{d_h}} \right) \right) \cdot q_1^{b_1/2} q_2^{b_2/2} \cdots q_{\ell}^{b_{\ell}/2} = 2^{c'} \cdot \left( \prod_{h=1}^{k} \left( \frac{\pi_h^{d_h} \pi_h^{d_h} - d_h}{\pi_h^{d_h} \pi_h^{d_h}} \right) \right) \cdot q_1^{b'_1} q_2^{b'_2} \cdots q_{\ell}^{b'_{\ell}}
\end{align*}
\]
This is a prime factorization of \( n \) as an integer. But so is \( (140) \). Since the prime factorization of an integer is unique (or by comparing \( p \)-valuations of the right hand sides on \( (143) \) and \( (140) \)), we thus conclude
\[
c' = c; \quad d'_h + e'_h = a_h \quad \text{for all } h;
\]
\[
2b'_j = b_j \quad \text{for all } j.
\]
In other words,
\[
c' = c; \quad e'_h = a_h - d'_h \quad \text{for all } h;
\]
\[
b'_j = b_j/2 \quad \text{for all } j.
\]
Hence, \( (142) \) rewrites as
\[
\alpha = \gamma \cdot p^c \cdot \left( \prod_{h=1}^k \left( \frac{d'_h}{\pi_h} - \frac{a_h}{\bar{\pi}_h} \right) \right) \cdot q_{1/2}^{b_1/2} q_{2/2}^{b_2/2} \cdots q_{k/2}^{b_k/2} = F \left( (\gamma, (d'_1, d'_2, \ldots, d'_k)) \right),
\]
(by the definition of \( F \)).

Thus, we have shown that \( \alpha \) is a value of \( F \). This proves that \( F \) is surjective.

Now, \( F \) is injective and surjective, hence bijective.]  
So \( F \) is a bijection. Thus,
\[
\left| \{1, i, -1, -i\} \times \prod_{h=1}^k \{0, 1, \ldots, a_h\} \right|
\]
\[
= \left| \{\alpha \in \mathbb{Z}[i] \mid n = a\bar{a} \} \right|
\]
\[
= (\text{the number of } \alpha \in \mathbb{Z}[i] \text{ such that } n = a\bar{a})
\]
\[
= (\text{the number of pairs } (x, y) \in \mathbb{Z}^2 \text{ such that } n = x^2 + y^2),
\]
so that
\[
\left( \text{the number of pairs } (x, y) \in \mathbb{Z}^2 \text{ such that } n = x^2 + y^2 \right)
\]
\[
= \left| \{1, i, -1, -i\} \times \prod_{h=1}^k \{0, 1, \ldots, a_h\} \right| = \left| \{1, i, -1, -i\} \right| \cdot \prod_{h=1}^k \{0, 1, \ldots, a_h\}
\]
\[
= 4 \cdot \prod_{h=1}^k \left( a_{1/2} \right) + 1 = 4 \cdot \prod_{h=1}^k (\nu_{p_h}(n) + 1) = 4 \cdot \prod_{\substack{p \text{ prime}; \\
p \equiv 1 \mod 4}} (\nu_p(n) + 1).
\]
(The last equality sign is a consequence of the fact that \( p_1, p_2, \ldots, p_h \) are distinct primes of Type 1, and that all other primes \( p \notin \{p_1, p_2, \ldots, p_h\} \) of Type 1 satisfy \( \nu_p(n) = 0 \).) This proves Theorem \( \boxed{4.2.60} \)(b).
4.2.6. What are the Gaussian primes?

We have so far seen the following Gaussian primes:

- Each prime of Type 3 is a Gaussian prime.
- $1 + i$ is a Gaussian prime.
- For each prime $p$ of Type 1, we have a Gaussian prime $\pi$ such that $p = \pi \overline{\pi}$, and then $\overline{\pi}$ is also a Gaussian prime.

**Theorem 4.2.62.** Each Gaussian prime is unit-equivalent to one of the Gaussian primes in this list.

*Proof.* HW5.

4.3. Brief survey of similar number systems

- Let us now see when a prime $p$ can be written as $x^2 + 2y^2$ with $x, y \in \mathbb{Z}$.

The set

$$\mathbb{Z} \left[ \sqrt{-2} \right] = \mathbb{Z} \left[ \sqrt{2i} \right]$$

is defined as the set of all complex numbers of the form $a + b\sqrt{2i}$ with $a, b \in \mathbb{Z}$. It is perhaps easier to regard it as its own variant of Gaussian integers, which I will call the “2-Gaussian integers”. These “2-Gaussian integers” can be defined as pairs $(a, b) \in \mathbb{Z}^2$ with addition and subtraction defined entrywise and multiplication defined by

$$(a, b) \cdot (c, d) = (ac - 2bd, ad + bc).$$

You can then write such pairs $(a, b)$ as $a + b\sqrt{2i}$, where $\sqrt{2i}$ is simply a symbol for the 2-Gaussian integer $(0, 1)$. Each 2-Gaussian integer $(a, b)$ has a norm, defined by $N((a, b)) = a^2 + 2b^2$.

Much of the theory of Gaussian integers still applies verbatim to 2-Gaussian integers. In particular, division with remainder still works for 2-Gaussian integers (like it does for Gaussian integers, i.e., non-uniquely), and the proof uses the same argument, but this time we have $N(\rho) \leq 3N(\beta)/4$ instead of $N(\rho) \leq N(\beta)/2$. Hence, 2-Gaussian integers have unique factorizations into “2-Gaussian primes”.

This can be used to show that a prime $p$ can be written as $x^2 + 2y^2$ if and only if there is an integer $u$ satisfying $u^2 \equiv -2 \mod p$. It can furthermore be shown that such an integer $u$ exists if and only if $p = 2$ or $p \equiv 1, 3 \mod 8$ (where “$p \equiv 1, 3 \mod 8$” is shorthand for “$p \equiv 1 \mod 8$ or $p \equiv 3 \mod 8$”). The proof uses a fact called quadratic reciprocity, which we may see later in this course.
• When can a prime $p$ be written as $x^2 + 3y^2$ with $x, y \in \mathbb{Z}$?

The logical continuation of the above pattern would be “when $p = 3$ or $p \equiv 1 \mod 3$”, since these are the cases when there is an integer $u$ satisfying $u^2 \equiv -3 \mod p$. And that is indeed true, but the proof is more complicated. Indeed, the “3-Gaussian integers” no longer have division with remainder, as $N(\rho) \leq N(\beta) / 2$ turns into $N(\rho) \leq N(\beta)$ which is not a strict inequality. Nevertheless we can prove our guess with more complicated reasoning: We need to use not $\mathbb{Z}[\sqrt{-3}]$ but rather the \textit{Eisenstein integers} $a + b\omega$ with $a, b \in \mathbb{Z}$ and $\omega = \frac{-1 + i\sqrt{3}}{2}$. These are best understood as pairs $(a, b) \in \mathbb{Z}^2$ with addition and subtraction defined entrywise and multiplication defined by

$$(a, b) (c, d) = (ac - bd, ad + bc - bd).$$

Their norm is $N((a, b)) = a^2 - ab + b^2$. They form a triangular lattice, not a rectangular one, and they do have division with remainder. Note that $N(a + b\omega) = a^2 - ab + b^2$, so some more work is needed to turn them into $x^2 + 3y^2$ solutions, but it’s doable.

• When can a prime $p$ be written as $x^2 + 4y^2$ with $x, y \in \mathbb{Z}$?

This is easy: $4y^2 = (2y)^2$, so we are looking for a way of writing $p$ as $x^2 + y^2$ with $y$ even.

I claim that the answer is “when $p \equiv 1 \mod 4$”. Do you see why?

• When can a prime $p$ be written as $x^2 + 5y^2$ with $x, y \in \mathbb{Z}$?

Our guess, by following the above pattern, would be “when $p = 2$ or $p = 5$ or $p \equiv 1, 3, 7, 9 \mod 20$”, since these are the cases when there is an integer $u$ satisfying $u^2 \equiv -5 \mod p$. But this is not true anymore. The right answer is “when $p = 2$ or $p = 5$ or $p \equiv 1, 9 \mod 20$”. And unsurprisingly, $\mathbb{Z}[\sqrt{-5}]$ does not have division with remainder.

• More generally, you can fix $n \in \mathbb{N}$ and ask when a prime can be written in the form $x^2 + ny^2$. There is a whole book [Cox13] devoted to this question! The answer becomes more complicated with $n$ getting large, and touches on a surprising number of different fields of mathematics (geometry, complex analysis, elliptic functions and elliptic curves).

• We can also ask when a prime $p$ can be written as $x^2 - ny^2$. The appropriate analogue of $\mathbb{Z}[i]$ tailored to this question is $\mathbb{Z}[\sqrt{n}]$, which however behaves much differently, since $\sqrt{n}$ is real. For example, as you saw on homework set #4 (in the Remark after Exercise 4), there are infinitely many units in $\mathbb{Z}[\sqrt{2}]$; the same is true for each $\mathbb{Z}[\sqrt{n}]$ with $n > 1$ and $n$ not being a perfect square (but this is much harder to prove).
• When can \( n \in \mathbb{N} \) be written as a sum of three squares? Legendre’s three-squares theorem: iff \( n \) is not of the form \( n = 4^a (8b + 7) \) for integers \( a, b \). Very hard to prove ([UspHea39, Chapter XIII] might have the only elementary proof).

• When can \( n \in \mathbb{N} \) be written as a sum of four squares? Lagrange’s four-squares theorem: always.\(^{136}\) This is easier to show, and there is even a formula for the number of representations: it is \( 8 \sum_{d|n; 4 \nmid d} \). The existence part can be proven using “Hurwitz integers”, which are certain quaternions.

2019-03-25 lecture

5. Rings and fields I: definitions and examples

5.1. Definition of a ring

We have seen several “number systems” in the above chapters:

• \( \mathbb{N} \) (the nonnegative integers);
• \( \mathbb{Z} \) (the integers);
• \( \mathbb{R} \) (the real numbers);
• \( \mathbb{Z}/n \) for a positive integer \( n \);
• \( \mathbb{C} \) (the complex numbers);
• \( \mathbb{D} \) (the dual numbers – see homework set #4 exercise 3);
• \( \mathbb{Z}[\sqrt{2}] = \{ a + b\sqrt{2} \mid a, b \in \mathbb{Z} \} \) (see homework set #4 exercise 4);
• \( \mathbb{Z}[\omega] = \{ a + b\omega \mid a, b \in \mathbb{Z} \} \) (the Eisenstein integers);
• \( \mathbb{Z}[\sqrt{-3}] \) (see homework set #5 exercise 6).

It may be a stretch to call the elements of some of these “numbers”, but it is not taboo (the word “number” has no precise meaning in mathematics), and these sets have a lot in common: We can add, subtract and multiply their elements (except for \( \mathbb{N} \), which does not allow subtraction); these operations satisfy the usual rules

\(^{136}\) An application (fortunately, no longer relevant):

“Warning: Due to a known bug, the default Linux document viewer evince prints \( N*N \) copies of a PDF file when \( N \) copies requested. As a workaround, use Adobe Reader acroread for printing multiple copies of PDF documents, or use the fact that every natural number is a sum of at most four squares.”
Math 4281 notes page 305

(e.g., associativity of multiplication, distributivity, etc.); these sets contain some element “behaving like 0” (that is, an element 0 such that \(a + 0 = 0 + a = a\) and \(a \cdot 0 = 0 \cdot a = 0\) for all \(a\)) and some element “behaving like 1” (that is, an element \(1\) such that \(a \cdot 1 = 1 \cdot a = a\) for all \(a\)). It turns out that just a few of these rules are sufficient to make “all the other rules” (in a certain appropriate sense) follow from them. Thus, it is reasonable to crystallize these few rules into a common, general notion (of which the above examples – excluding \(\mathbb{N}\) – will be particular cases); this notion will be called a “ring”. Hence, we shall define a ring to be (roughly speaking) a set with operations + and \(\cdot\) and elements 0 and 1 that satisfies these few rules. Let us be specific about what these rules are:

### Definition 5.1.1.

(a) A ring means a set \(K\) endowed with

- two binary operations called “addition” and “multiplication”, and denoted by \(+_K\) and \(\cdot_K\), respectively, and
- two elements called “zero” (or “origin”) and “unity” (or “one”), and denoted by \(0_K\) and \(1_K\), respectively

such that the following axioms are satisfied:

- **Commutativity of addition:** We have \(a +_K b = b +_K a\) for all \(a, b \in K\).
- **Associativity of addition:** We have \(a +_K (b +_K c) = (a +_K b) +_K c\) for all \(a, b, c \in K\).
- **Neutrality of zero:** We have \(a +_K 0_K = 0_K +_K a = a\) for all \(a \in K\).
- **Existence of additive inverses:** For any \(a \in K\), there exists an element \(a' \in K\) such that \(a +_K a' = a' +_K a = 0_K\). (It is not immediately obvious, but will be shown later, that such an \(a'\) is unique. Thus, \(a'\) is called the additive inverse of \(a\), and is denoted by \(-a\).)
- **Associativity of multiplication:** We have \(a (bc) = (ab) c\) for all \(a, b, c \in K\). Here and in the following, we use “xy” as an abbreviation for “\(x \cdot_K y\)”.
- **Neutrality of one:** We have \(a1_K = 1_K a = a\) for all \(a \in K\).
- **Annihilation:** We have \(a0_K = 0_K a = 0_K\) for all \(a \in K\).
- **Distributivity:** We have

\[
a (b +_K c) = ab +_K ac \quad \text{and} \quad (a +_K b) c = ac +_K bc
\]

for all \(a, b, c \in K\). Here and in the following, we are using the PEMDAS convention for order of operations; thus, for example, “\(ab +_K ac\)” must be understood as “\((ab) +_K (ac)\)”.

Recall the definition of a “binary operation” (Definition 1.6.1). In particular, a binary operation on a set \(S\) must have all its values in \(S\).
These eight axioms will be called the ring axioms.
(Note that we do not require the existence of a “subtraction” operation \(-K\). But we will later construct such an operation out of the existing operations and axioms; it is thus unnecessary to require it. We also do not require the existence of multiplicative inverses; nor do we require commutativity of multiplication yet.)

(b) A ring \(K\) (with operations \(+_K\) and \(\cdot_K\)) is called commutative if it satisfies the following extra axiom:

- **Commutativity of multiplication**: We have \(ab = ba\) for all \(a, b \in K\).

Note a few things:

- We shall abbreviate \(+_K\), \(\cdot_K\), \(0_K\) and \(1_K\) as \(+\), \(\cdot\), \(0\) and \(1\) unless there is a chance of confusion with the “usual” notions of addition, multiplication, zero and one. (The example of the ring \(\mathbb{Z}'\) shown below is a case where such confusion is possible; but most of the time, it is not.)

- We have not required our rings to be endowed with a “subtraction” operation. Nevertheless, each ring \(K\) automatically has a subtraction operation: Namely, for any \(a, b \in K\), we can define \(a - b\) to be \(a + b'\), where \(b'\) is the additive inverse of \(b\). (We will later see that this operation is well-defined (Definition 5.3.4) and satisfies the rules you would expect (Definition 5.3.5).)

- Some of the ring axioms we required in Definition 5.1.1 are redundant, i.e., they follow from other ring axioms. (For example, Annihilation follows from the other axioms.) We don’t mind this, as long as these axioms are natural and easy to check in real examples.

- We have required commutativity of addition to hold for all rings, but commutativity of multiplication only to hold for commutative rings. You may wonder what happens if we also omit the commutativity of addition. The answer is “nothing new”: Commutativity of addition follows from the other axioms! (Proving this is a fun, although inconsequential, puzzle.)

- By our definition, a ring consists of a set \(K\), two operations \(+\) and \(\cdot\) and two elements \(0\) and \(1\). Thus, strictly speaking, a ring is a 5-tuple \((K, +, \cdot, 0, 1)\). In reality, we will often just speak of the “ring \(K\)” (so we will mention only the set and not the other four pieces of data) and assume that the reader can figure out the rest of the 5-tuple. This is okay as long as the rest of the 5-tuple can be inferred from the context. For example, when we say “the ring \(\mathbb{Z}'\)”, it is clear that we mean the ring \((\mathbb{Z}, +, \cdot, 0, 1)\) with the usual addition and multiplication operations and the usual numbers 0 and 1. The same applies when we speak of “the ring \(\mathbb{R}\)” or “the ring \(\mathbb{C}\)” or “the ring \(\mathbb{Z}[i]\)”. In general, whenever a set \(S\) is equipped with two operations that are called \(+\) and \(\cdot\) and
two elements that are called 0 and 1 (even if these elements are not literally the numbers 0 and 1), we automatically understand “the ring $S$” to be the ring $(S, +, \cdot, 0, 1)$ that is defined using these operations and elements. If we want to make a different ring out of the set $S$, then we have to say this explicitly.

- Some authors do not require the element 1 as part of what it means to be a ring. But we do. Be careful when reading the literature, as the truth or falsehood of many results depends on whether the 1 is included in the definition of a ring or not. (When authors do not require the element 1 in the definition of a ring, they reserve the notion of a “unital ring” for a ring that does come equipped with a 1 that satisfies the “Neutrality of one” axiom; i.e., they call “unital ring” what we call “ring”.)

5.2. Examples of rings

Many of the “number systems” seen above, and several others, are examples of rings:

- The sets $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ (endowed with the usual addition, multiplication, 0 and 1) are commutative rings. In each case, the additive inverse of an element $a$ is what we know as $-a$ from high school (or undergraduate mathematics, in the case of $\mathbb{C}$).

- The set $\mathbb{N}$ (again endowed with the usual addition, multiplication, 0 and 1) is not a ring. Indeed, the “existence of additive inverses” axiom fails for $a = 1$, because the element 1 has no additive inverse in $\mathbb{N}$ (that is, there is no $1' \in \mathbb{N}$ such that $1 + 1' = 1' + 1 = 0$).

- The sets $\mathbb{C}$, $\mathbb{Z}[i]$, $\mathbb{D}$, $\mathbb{Z}[\sqrt{2}]$, $\mathbb{Z}[\omega]$ and $\mathbb{Z}[\sqrt{-3}]$ (from the previous chapter and from the homework sets) are commutative rings. All of the axioms are easy to check, and some of them we have checked. In each case, the element $a'$ in the “existence of additive inverses” axiom is $-a$.

- If you have seen polynomials: The set $\mathbb{Z}[x]$ of all polynomials in a single variable $x$ with integer coefficients is a commutative ring. Similarly for other kinds of coefficients, and several variables. But we will come back later to this, once we have rigorously defined polynomials.

- We can define a commutative ring $\mathbb{Z}'$ as follows:

  We define a binary operation $\tilde{\times}$ on $\mathbb{Z}$ by

  $$(a \tilde{\times} b = -ab \quad \text{for all } a, b \in \mathbb{Z}).$$

  Now, let $\mathbb{Z}'$ be the set $\mathbb{Z}$, endowed with the usual addition $+$ and the unusual multiplication $\tilde{\times}$ and the elements $0_{\mathbb{Z}'} = 0$ and $1_{\mathbb{Z}'} = -1$.

  Is this $\mathbb{Z}'$ a commutative ring? Let us check the axioms:
- The first four axioms involve only addition and 0 (but not multiplication and 1), and therefore still hold for \( \mathbb{Z}' \) (because \( \mathbb{Z}' \) has the same addition and 0 as \( \mathbb{Z} \)).

- Associativity of multiplication: We must check that
  \[
  a \times (b \times c) = (a \times b) \times c
  \]
  for all \( a, b, c \in \mathbb{Z}' \).

  (Note that we cannot omit the “multiplication sign” \( \times \) here and simply write “\( bc \)” for “\( b \times c \)”, because “\( bc \)” already means something different. Note also that “\( a, b, c \in \mathbb{Z}' \)” means the same as “\( a, b, c \in \mathbb{Z} \)”, because \( \mathbb{Z}' = \mathbb{Z} \) as sets.)

  Checking this is straightforward: Let \( a, b, c \in \mathbb{Z}' \). Then, comparing
  \[
  a \times (b \times c) = a \times (-bc) = -a (-bc) = abc
  \]
  with
  \[
  (a \times b) \times c = (-ab) \times c = -(-ab) c = abc,
  \]
  we obtain \( a \times (b \times c) = (a \times b) \times c \). Thus, associativity of multiplication holds for \( \mathbb{Z}' \).

- Neutrality of one in \( \mathbb{Z}' \): We must check that
  \[
  a \times 1_{\mathbb{Z}'} = 1_{\mathbb{Z}'} \times a = a \quad \text{for all} \quad a \in \mathbb{Z}'.
  \]

  This, too, is straightforward: If \( a \in \mathbb{Z}' \), then \( a \times 1_{\mathbb{Z}'} = a \times (-1) = -a (-1) = a \) and similarly \( 1_{\mathbb{Z}'} \times a = a \).

- Annihilation and commutativity of multiplication are just as easy to check.

- Distributivity for \( \mathbb{Z}' \): We must check that
  \[
  a \times (b + c) = a \times b + a \times c \quad \text{and}
  (a + b) \times c = a \times c + b \times c
  \]
  for all \( a, b, c \in \mathbb{Z}' \).

  So let \( a, b, c \in \mathbb{Z}' \). In order to verify \( a \times (b + c) = a \times b + a \times c \), we compare
  \[
  a \times (b + c) = -a (b + c) = -ab - ac
  \]
  with
  \[
  a \times b + a \times c = (-ab) + (-ac) = -ab - ac.
  \]
  Similarly we can check \( (a + b) \times c = a \times c + b \times c \).
So \( \mathbb{Z}' \) is a ring.

(Note that \( (\mathbb{Z}, +, \times, 0, 1) \) is not a ring.)

However, \( \mathbb{Z}' \) is not a new ring. It is just \( \mathbb{Z} \) with its elements renamed. Namely, if we rename each integer \( a \) as \( -a \), then the operations of \(+\) and \( \cdot \) and the elements 0 and 1 of \( Z \) turn into the operations \( +\) and \( \tilde{\times} \) and the elements 0 and 1_{\mathbb{Z}'} of \( \mathbb{Z}' \). This is a confusing thing to say (please don’t actually rename numbers as other numbers!); the rigorous (and hopefully not confusing) way to say this is as follows: The bijection

\[
\varphi : \mathbb{Z} \to \mathbb{Z}', \quad a \mapsto -a
\]

satisfies

\[
\begin{align*}
\varphi (a + b) & = \varphi (a) + \varphi (b) \quad \text{for all } a, b \in \mathbb{Z}; \\
\varphi (ab) & = \varphi (a) \tilde{\times} \varphi (b) \quad \text{for all } a, b \in \mathbb{Z}; \\
\varphi (0) & = 0; \\
\varphi (1) & = -1 = 1_{\mathbb{Z}'}.
\end{align*}
\]

Thus, we can view \( \varphi \) as a way of relabelling the integers so that data \(+, \cdot, 0, 1\) of \( \mathbb{Z} \) become the data \( +, \tilde{\times}, 0, 1_{\mathbb{Z}'} \) of \( \mathbb{Z}' \). We will later call bijections like \( \varphi \) ring isomorphisms.

- Recall: If \( A \) and \( B \) are two sets, then

\[
B^A := \{ \text{maps } A \to B \}.
\]

(This notation is not wantonly chosen to annoy you with its seeming backwardness; instead, it harkens back to the fact that \( |B^A| = |B|^{|A|} \).) The set \( \mathbb{Q}^\mathbb{Q} \) of all the maps from \( \mathbb{Q} \) to \( \mathbb{Q} \) is a commutative ring, where

- addition and multiplication are defined pointwise: i.e., if \( f, g \in \mathbb{Q}^\mathbb{Q} \) are two maps, then the maps \( f + g \) and \( f \cdot g \) are defined by

\[
( f + g ) (x) = f(x) + g(x), \quad \text{and} \quad ( f \cdot g ) (x) = f(x) \cdot g(x) \quad \text{for all } x \in \mathbb{Q};
\]

- 0 means the “constant 0” function (i.e., the map \( \mathbb{Q} \to \mathbb{Q}, x \mapsto 0 \));

- 1 means the “constant 1” function (i.e., the map \( \mathbb{Q} \to \mathbb{Q}, x \mapsto 1 \)).

All the axioms are easy to check.

Similarly for \( \mathbb{Q}^\mathbb{C} \) or \( \mathbb{Q}^\mathbb{N} \) or \( \mathbb{R}^\mathbb{R} \) (the set of “functions” you know from calculus) or \( \mathbb{C}^\mathbb{C} \) (or, more generally, for \( \mathbb{K}^S \), where \( \mathbb{K} \) is any commutative ring and \( S \) is any set), but not for \( \mathbb{N}^\mathbb{Q} \). The problem with \( \mathbb{N}^\mathbb{Q} \) is that “existence of additive inverses” is not satisfied, since \(-a \notin \mathbb{N} \) for positive \( a \in \mathbb{N} \).
• Recall that \( \mathbb{Z}[\sqrt{2}] = \{ a + b\sqrt{2} \mid a, b \in \mathbb{Z} \} \) is a ring.

But the set \( \{ a + b\sqrt{2} \mid a, b \in \mathbb{Z} \} \) (with the usual addition and multiplication) is not a ring. The reason is that multiplication is not a binary operation on this set, since it is possible that two numbers \( a \) and \( \beta \) lie in this set but their product \( a\beta \) does not. For example, \( 1 + \sqrt{2} \) lies in this set, but

\[
\left( 1 + \sqrt{2} \right) \left( 1 + \sqrt{2} \right) = 1 + 2\sqrt{2} + \sqrt{8}
\]

does not. (That said, this set does satisfy all the eight ring axioms.)

2019-03-27 lecture

• The set of \( 2 \times 2 \)-matrices with rational entries (endowed with matrix addition as \( + \), matrix multiplication as \( \cdot \), the zero matrix \( \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \) as \( 0 \), and \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) as \( 1 \)) is a ring, but not a commutative ring. Indeed, the ring axioms are true (this is known from linear algebra), but commutativity of multiplication is not (the product \( AB \) of two \( 2 \times 2 \)-matrices \( A \) and \( B \) is not always equal to \( BA \)). The same applies to \( n \times n \)-matrices for arbitrary \( n \). (See [Grinbe18, §2.9] for a proof of the associativity of multiplication of \( n \times n \)-matrices.)

• If you like the empty set, you will enjoy the zero ring. This is the one-element set \( \{ 0 \} \), endowed with the only possible addition (given by \( 0 + 0 = 0 \)), the only possible multiplication (given by \( 0 \cdot 0 = 0 \)), the only possible zero (namely, \( 0 \)) and the only possible unity (also \( 0 \)). This is a commutative ring, and is known as the zero ring. Resist the temptation of denoting its unity by \( 1 \), as this will quickly lead to painful confusion.

(Some authors choose to forbid this ring, usually for no good reasons.)

• If \( n \) is an integer, then \( \mathbb{Z}/n \) is a ring (with the operations \( + \) and \( \cdot \) that we defined, with the zero \( [0]_n \) and the unity \( [1]_n \)). When the integer \( n \) is positive, this ring \( \mathbb{Z}/n \) has \( n \) elements. (When \( n \) is prime, it can be shown that \( \mathbb{Z}/n \) is the only ring with exactly \( n \) elements, up to relabeling its elements. In general, however, there can be several rings with \( n \) elements.)

• In set theory, the symmetric difference \( A \triangle B \) of two sets \( A \) and \( B \) is defined to be the set

\[
(A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A) = \{ x \mid x \text{ belongs to exactly one of } A \text{ and } B \}.
\]
Now, let $S$ be any set. Let $\mathcal{P}(S)$ denote the power set of $S$ (that is, the set of all subsets of $S$). Then, it is easy to check that the following properties hold:

- $A \triangle B = B \triangle A$ for any sets $A$ and $B$;
- $A \cap B = B \cap A$ for any sets $A$ and $B$;
- $(A \triangle B) \triangle C = A \triangle (B \triangle C)$ for any sets $A, B, C$;
- $(A \cap B) \cap C = A \cap (B \cap C)$ for any sets $A, B, C$;
- $A \triangle \emptyset = \emptyset \triangle A = A$ for any set $A$;
- $A \cap S = S \cap A = A$ for any subset $A$ of $S$;
- $A \triangle A = \emptyset$ for any set $A$;
- $\emptyset \cap A = A \cap \emptyset = \emptyset$ for any set $A$;
- $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$ for any sets $A, B, C$;
- $(A \triangle B) \cap C = (A \cap C) \triangle (B \cap C)$ for any sets $A, B, C$.

Therefore, the set $\mathcal{P}(S)$, endowed with the addition $\triangle$, the multiplication $\cap$, the zero $\emptyset$ and the unity $S$ is a commutative ring. Furthermore, the additive inverse of any $A \in \mathcal{P}(S)$ is $A$ itself (since $A \triangle A = \emptyset$). Moreover, each $A \in \mathcal{P}(S)$ satisfies $A \cap A = A$, which means (in the language of ring operations) that its square is itself. Thus, $\mathcal{P}(S)$ is what is called a Boolean ring.

Let us now see some non-examples – i.e., examples of things that are not rings:

- How comes we cannot divide by 0?

   Let us make the question precise. Of course, we cannot find an integer $a$ that satisfies $0 \cdot a = 1$, or a real, or a complex number, etc. But could we perhaps find such a number $a$ in some larger “number system”? The answer, of course, depends on what “number system” means for you. If it means a ring, then we cannot find such an $a$ in any ring.

   Indeed, assume that we can. In other words, assume that there is a ring $\mathbb{K}$ that contains the usual set $\mathbb{Z}$ of integers as well as a new element $\infty$ such that $0 \cdot \infty = 1$. And assume (this is a very reasonable assumption) that the numbers 0 and 1 are indeed the zero and the unity of this ring. Then, the Annihilation axiom yields $0 \cdot \infty = 0$, so that $0 = 0 \cdot \infty = 1$, which is absurd. So such a ring $\mathbb{K}$ cannot exist. Thus, we cannot divide by 0, even if we extend our “number system”.

- Here is an “almost-ring” beloved to combinatorialists: the max-plus semiring $\mathbb{T}$ (also known as the tropical semiring\textsuperscript{138}).

   We introduce a new symbol $-\infty$, and we set $\mathbb{T} = \mathbb{Z} \cup \{-\infty\}$ as sets. But we do not “inherit” the addition and multiplication from $\mathbb{Z}$. Instead, let us

\textsuperscript{138}To be pedantic: The name “tropical semiring” refers to several different objects, of which $\mathbb{T}$ is but one.
define two new “addition” and “multiplication” operations $+_T$ and $\cdot_T$ (not to be mistaken for the original addition $+$ and multiplication $\cdot$ of integers) as follows:

\[
\begin{align*}
    a +_T b &= \max\{a, b\}; \\
    a \cdot_T b &= a + b \quad \text{(usual addition of integers),}
\end{align*}
\]

where we set

\[
\max\{-\infty, n\} = \max\{n, -\infty\} = n \quad \text{and} \quad (-\infty) + n = n + (-\infty) = -\infty \quad \text{for any } n \in \mathbb{Z} \cup \{-\infty\}.
\]

This set $T$ endowed with the “addition” $+_T$, “multiplication” $\cdot_T$, “zero” $-\infty$ and “unity” $0$ satisfies all but one of the ring axioms. The only one that it does not satisfy is the existence of additive inverses. Such a structure is called a semiring.

- Consider the set

\[
2\mathbb{Z} := \{2a \mid a \in \mathbb{Z}\} = \{\ldots, -4, -2, 0, 2, 4, \ldots\} = \{\text{all even integers}\}.
\]

Endowing this set with the usual addition and multiplication (and 0), we obtain a structure that is like a ring but has no unity. This is called a nonunital ring. There is no way to find a unity for it, because (for example) 2 is not a product of any two elements of $2\mathbb{Z}$.

5.3. Additive inverses, sums, powers and their properties

What can you do when you have a ring?

<table>
<thead>
<tr>
<th>Convention 5.3.1.</th>
<th>For the rest of this section, we fix a ring $K$, and we denote its addition, multiplication, zero and unity by $+$, $\cdot$, 0 and 1.</th>
</tr>
</thead>
</table>

One thing you can do is subtraction. This relies on the following fact:

<table>
<thead>
<tr>
<th>Theorem 5.3.2.</th>
<th>Let $a \in K$. Then, $a$ has exactly one additive inverse.</th>
</tr>
</thead>
</table>

Before we prove this, let us recall how additive inverses are defined:

\[
\begin{align*}
    a + \max\{b, c\} &= \max\{a + b, a + c\} \quad \text{and} \\
    \max\{a, b\} + c &= \max\{a + c, b + c\}.
\end{align*}
\]

\[^{139}\text{For example, the distributivity axiom for } T \text{ boils down to the two identities}\]
Definition 5.3.3. Let $a \in K$. An additive inverse of $a$ means an element $a'$ of $K$ such that $a + a' = a' + a = 0$.

Proof of Theorem 5.3.2. By the ring axioms, $a$ has at least one additive inverse. We must thus only show that $a$ has at most one additive inverse.

This can be done by the same argument that we used previously to prove that a residue class in $\mathbb{Z}/n$ has at most one inverse (in the proof of Proposition 3.5.4), but now we need to replace $\mathbb{Z}/n$, multiplication and $[1]_n$ by $K$, addition and 0, respectively.

In detail: Let $b$ and $c$ be two additive inverses of $a$. We must prove that $b = c$. We have $a + b = 0$ (since $b$ is an additive inverse of $a$) and $c + a = 0$ (since $c$ is an additive inverse of $a$). Hence, the associativity of addition yields

$$(c + a) + b = c + (a + b) = c + 0 = c$$

(by the neutrality of zero). Comparing this with

$$(c + a) + b = 0 + b = b$$

(by the neutrality of zero),

we obtain $b = c$, so our two additive inverses are equal. This shows that the additive inverse is unique. Thus, Theorem 5.3.2 is proven.

Definition 5.3.4. (a) If $a \in K$, then the additive inverse of $a$ will be called $-a$. (This is well-defined, since Theorem 5.3.2 shows that this additive inverse is unique.)

(b) If $a \in K$ and $b \in K$, then we define the difference $a - b$ to be the element $a + (-b)$ of $K$. This new binary operation $-$ on $K$ is called "subtraction".

Additive inverses and subtraction satisfy certain rules that should not surprise you:

Proposition 5.3.5. Let $a, b, c \in K$.

(a) We have $a - b = c$ if and only if $a = b + c$. (Roughly speaking, this means that subtraction undoes addition.)

(b) We have $- (a + b) = (-a) + (-b)$.

(c) We have $-0 = 0$.

(d) We have $0 - a = -a$.

(e) We have $- (-a) = a$.

(f) We have $- (ab) = (-a) b = a (-b)$.

(g) We have $a - b - c = a - (b + c)$. (Here and in the following, "$a - b - c$" should be read as "$(a - b) - c$".)

Proof of Proposition 5.3.5. All of this is fairly straightforward to prove:
(a) \[\Rightarrow\]: Assume \(a - b = c\). Thus, \(c = a - b = a + (-b)\) (by the definition of \(a - b\)). Adding \(b\) on both sides of this equation, we get
\[
c + b = (a + (-b)) + b = a + ((-b) + b) = a +0 = a
\]
(by the neutrality of zero), so that \(a = c + b = b + c\).

\[\Leftarrow\]: Assume \(a = b + c\). Adding \(-b\) to both sides of this equation, we get
\[
a + (-b) = (b + c) + (-b) = (c + b) + (-b) = c + (b + (-b)) = c +0 = c
\]
(by associativity of addition)
(by the neutrality of zero), so that \(c = a + (-b) = a - b\). Thus, Proposition 5.3.5 \(a\) is proven.

(b) We need to prove that \((-a) + (-b) = -(a + b)\). In other words, we need to prove that \((-a) + (-b)\) is the additive inverse of \(a + b\) (because that’s what \(- (a + b)\) is). In other words, we need to prove that
\[
(a + b) + ((-a) + (-b)) = ((-a) + (-b)) + (a + b) = 0.
\]

Associativity of addition yields
\[
(a + b) + ((-a) + (-b)) = a + \underbrace{b + ((-a) + (-b))}_{=0} = a + (b + ((-b) + (-a))) = a + ((-b) + (-a)) + (b + (-b) + (-a)) = a + (0 + (-a)) = a + (-a) = 0
\]
(since \(-a\) is the additive inverse of \(a\)). Also, \((a + b) + ((-a) + (-b)) = ((-a) + (-b)) + (a + b)\) (by commutativity of addition). Combining these two equalities, we obtain
\[
(a + b) + ((-a) + (-b)) = ((-a) + (-b)) + (a + b) = 0.
\]
This completes the proof of Proposition 5.3.5 \textit{(b)}.

\textbf{(c)} We have \(0 + 0 = 0\) (by the neutrality of 0). But this shows precisely that 0 is an additive inverse of 0. In other words, \(0 = \overline{0}\). This proves Proposition 5.3.5 \textit{(c)}.

\textbf{(d)} The definition of subtraction yields \(-a = 0 + (-a) = -a\) (by the neutrality of 0). This proves Proposition 5.3.5 \textit{(d)}.

\textbf{(e)} Since \(-a\) is an additive inverse of \(a\), we have \((-a) + a = 0\) and \(a + (-a) = 0\). But the same two equations say that \(a\) is an additive inverse of \(-a\). In other words, \(a = -(-a)\). This proves Proposition 5.3.5 \textit{(e)}.

\textbf{(f)} We have \((-a) + a = 0\) (since \(-a\) is an additive inverse of \(a\)). But distributivity yields \((-a) b + ab = ((-a) + a) b = 0 b = 0\) (by annihilation). Likewise, \(ab + \)

\((-a) b = 0\). Hence, \((-a) b\) is an additive inverse of \(ab\). In other words, \((-a) b = -\overline{(ab)}\).

We have \(b + (-b) = 0\) (since \(-b\) is an additive inverse of \(b\)). But distributivity yields \(\overline{ab} + a (-b) = a (b + (-b)) = a 0 = 0\) (by annihilation). Likewise, \(a (-b) +\)

\(ab = 0\). Hence, \(a (-b)\) is an additive inverse of \(ab\). In other words, \(a (-b) = -\overline{(ab)}\).

Combining this with \((-a) b = -\overline{(ab)}\), we obtain \(-\overline{(ab)} = (-a) b = a (-b)\). This proves Proposition 5.3.5 \textit{(f)}.

\textbf{(g)} The definition of subtraction yields

\[a - b - c = (a - b) + (-c) = (a + (-b)) + (-c) = a + ((-b) + (-c))\]

(by associativity of addition). But Proposition 5.3.5 \textit{(b)} (applied to \(b\) and \(c\) instead of \(a\) and \(b\)) yields \(-\overline{(b + c)} = (-b) + (-c)\). The definition of subtraction yields \(a - \overline{(b + c)} = a + (\overline{(-b + (-c))}) = a + ((-b) + (-c)).\)

Comparing this with \(a - b - c = a + ((-b) + (-c))\), we obtain \(a - b - c = a - (b + c)\). This proves Proposition 5.3.5 \textit{(g)}. \hfill \Box

If \(a, b \in \mathbb{K}\), then the expression “\(-ab\)” can be considered ambiguous, since it can be read either as “\(-a b\)” or as “\(- (ab)\)”.

But Proposition 5.3.5 \textit{(f)} shows that these two readings yield the same result; therefore, you need not fear this ambiguity.

Furthermore, we don’t need to parenthesize expressions like \(a + b + c\) or \(abc\). Indeed:

**Theorem 5.3.6.** Finite sums of elements of \(\mathbb{K}\) can be defined in the same way as finite sums of usual (i.e., real or rational) numbers (with the empty sum defined to be 0). That is, if \(S\) is a finite set, and if \(a_s \in \mathbb{K}\) for each \(s \in S\), then \(\sum_{s \in S} a_s\) is well-defined and satisfies the usual rules, such as

\[\sum_{s \in S} (a_s + b_s) = \sum_{s \in S} a_s + \sum_{s \in S} b_s.\]
Thus, in particular, sums like $\sum_{i=p}^{q} a_i$ or $a_1 + a_2 + \cdots + a_k$ are well-defined. We don’t need to put parentheses or specify the order of summation in order to make them non-ambiguous.

Proof. This is proven just as for numbers. (See [Grinbe15, §2.14 and §1.4] for how it is proven for numbers.)

What about finite products? Is $\prod_{s \in S} a_s$ well-defined? Not always, but only for commutative rings. Indeed, a product like $\prod_{s \in S} a_s$ has no pre-defined order of multiplication (in general), so for it to be well-defined, it would have to be independent of the order; but this would require the commutativity of multiplication.

**Theorem 5.3.7.** (a) Finite products of elements of $\mathbb{K}$ can be defined in the same way as finite products of usual (i.e., real or rational) numbers (with the empty product defined to be 1) as long as the ring $\mathbb{K}$ is commutative.

(b) For general (not necessarily commutative) rings $\mathbb{K}$, we can still define products with a pre-determined order, such as $a_1 a_2 \cdots a_k$ (where $a_1, a_2, \ldots, a_k \in \mathbb{K}$). These products can be defined recursively as follows:

$$a_1 a_2 \cdots a_k = 1 \quad \text{if } k = 0;$$

otherwise,

$$a_1 a_2 \cdots a_k = (a_1 a_2 \cdots a_{k-1}) a_k.$$

These products still satisfy reasonable rules, such as

$$a_1 a_2 \cdots a_k = (a_1 a_2 \cdots a_i) (a_{i+1} a_{i+2} \cdots a_k) \quad \text{for all } i \in \{0, 1, \ldots, k\}.$$

Proof. (a) This is proven just as for numbers. (See [Grinbe15, §2.14 and §1.4] for how it is proven for numbers.)

(b) This will be proven on HW0. (For now, you can find proofs in various texts on algebra, or at [https://groupprops.subwiki.org/wiki/Associative_implies_generalized_associative](https://groupprops.subwiki.org/wiki/Associative_implies_generalized_associative).)

Theorem 5.3.7 (b) is called the **general associativity theorem**. Note that Theorem 5.3.7 (b) entails that if we have $k$ elements $a_1, a_2, \ldots, a_k$ of a ring $\mathbb{K}$, then any two ways of parenthesizing the product $a_1 a_2 \cdots a_k$ yield the same result. For example, for $k = 4$, we have

$$((a_1 a_2) a_3) a_4 = (a_1 (a_2 a_3)) a_4 = (a_1 a_2) (a_3 a_4) = a_1 ((a_2 a_3) a_4) = a_1 (a_2 (a_3 a_4)).$$

(It is not hard to prove this particular chain of identities by multiplying the associativity of multiplication in the appropriate places; but for higher values of $k$, such a manual approach becomes more and more cumbersome.)
What else can we do with our ring $K$?

By definition, we know how to multiply two elements of $K$. But there is also a natural way to multiply an element of $K$ with an integer. This is defined as follows:

**Definition 5.3.8.** Let $a \in K$ and $n \in \mathbb{Z}$. Then, we define an element $na$ of $K$ by

$$na = \begin{cases} 
a + a + \cdots + a, & \text{if } n \geq 0; \\
- (a + a + \cdots + a), & \text{if } n < 0 
\end{cases} \text{, \ n \ times}$$

The “$na$” that we have just defined has nothing to do with the multiplication $\cdot$ of $K$, since $n$ is not (generally) an element of $K$. However, when $K$ is one of the usual rings of numbers (like $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$), then this kind of multiplication coincides with the multiplication $\cdot$ of $K$ (that is, $na$ means the same thing). Indeed, Definition 5.3.8 clearly generalizes the definition of $na$ for rational numbers $a$. Furthermore, when $K = \mathbb{Z}/n$ for some integer $n$, Definition 5.3.8 agrees with Definition 3.4.18 (in the sense that both definitions yield the same result for $r \alpha$ when $r \in \mathbb{Z}$ and $\alpha \in \mathbb{Z}/n$). (This is easy to prove by induction.)

2019-03-29 lecture

We can also define powers of elements of a commutative ring:

**Definition 5.3.9.** Let $a \in K$ and $n \in \mathbb{N}$. Then, we define an element $a^n$ of $K$ by

$$a^n = a \cdot a \cdot \cdots \cdot a \text{. \ n \ times}$$

This definition clearly generalizes the definition of $a^n$ for rational numbers $a$. Furthermore, when $K = \mathbb{Z}/n$ for some integer $n$, Definition 5.3.9 agrees with Definition 3.4.20 (in the sense that both definitions yield the same result for $\alpha^k$ when $\alpha \in \mathbb{Z}/n$ and $k \in \mathbb{N}$). (This follows from Theorem 3.4.26 (c).) Furthermore, when $K = \mathbb{C}$, Definition 5.3.9 agrees with Definition 4.1.17.

Standard rules for addition, subtraction, multiplication and taking powers hold in every ring:
Proposition 5.3.10. (a) We have

\[(n + m) a = na + ma \quad \text{for all } a \in K \text{ and } n, m \in \mathbb{Z};\]  
\[n (a + b) = na + nb \quad \text{for all } a, b \in K \text{ and } n \in \mathbb{Z};\]  
\[−(na) = (−n) a = n (−a) \quad \text{for all } a \in K \text{ and } n \in \mathbb{Z};\]  
\[(nm) a = n (ma) \quad \text{for all } a \in K \text{ and } n, m \in \mathbb{Z};\]  
\[n (ab) = (na) b = a (nb) \quad \text{for all } a, b \in K \text{ and } n \in \mathbb{Z};\]  
\[1a = a \quad \text{for all } a \in K;\]  
\[−1a = −a \quad \text{for all } a \in K;\]  
\[1^n = 1 \quad \text{for all } n \in \mathbb{N};\]  
\[0^n = \begin{cases} 0, & \text{if } n > 0 \\ 1, & \text{if } n = 0 \end{cases} \quad \text{for all } n \in \mathbb{N};\]  
\[a^{n+m} = a^n a^m \quad \text{for all } a \in K \text{ and } n, m \in \mathbb{N};\]  
\[(a^n)^m = a^{nm} \quad \text{for all } a \in K \text{ and } n, m \in \mathbb{N};\]  
\[a^0 = 1 \quad \text{for all } a \in K.\]

In particular:

- The equality (146) shows that the expression “$−na$” (with $a \in K$ and $n \in \mathbb{Z}$) is unambiguous (since its two possible interpretations, namely $−(na)$ and $(−n) a$, yield equal results).

- The equality (147) shows that the expression “$nma$” (with $a \in K$ and $n, m \in \mathbb{Z}$) is unambiguous.

- The equality (148) shows that the expression “$nab$” (with $a, b \in K$ and $n \in \mathbb{Z}$) is unambiguous.

(b) For any $a, b \in K$, we have

\[(a + b)^2 = (a + b) (a + b) = a (a + b) + b (a + b) = a a + a b + b a + b b = a^2 + a b + b a + b^2.\]

This further equals $a^2 + 2ab + b^2$ if $K$ is commutative.

(c) Let $a, b \in K$ satisfy $ab = ba$. (This holds automatically when $K$ is commutative.) Then:

\[ab^n = b^n a \quad \text{for all } n \in \mathbb{N};\]  
\[a^i b^j = b^j a^i \quad \text{for all } i, j \in \mathbb{N};\]  
\[(ab)^n = a^n b^n \quad \text{for all } n \in \mathbb{N};\]  
\[(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n−k} \quad \text{for all } n \in \mathbb{N}.\]
(d) Let $a, b \in \mathbb{K}$ satisfy $ab = ba$. Then,

$$a^n - b^n = (a - b) \left( a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1} \right)$$

for all $n \in \mathbb{N}$.

Proof. (a) Analogous to the proofs for rationals (at least if you know the right proofs). For example, to prove (144), consider six cases corresponding to the six possible constellations of signs of $n, m, n + m$.
(b) Expand using distributivity.
(c) Induction.
(d) Analogous to the proof for rationals.

5.4. Multiplicative inverses and fields

Convention 5.4.1. For the rest of this section, we fix a ring $\mathbb{K}$, and we denote its addition, multiplication, zero and unity by $+$, $\cdot$, 0 and 1.

Each element $a$ of the ring $\mathbb{K}$ has an additive inverse $-a$, which satisfies $(-a) + a = a + (-a) = 0$. What about a “multiplicative inverse”?

Definition 5.4.2. Let $a \in \mathbb{K}$. A multiplicative inverse of $a$ means an element $a'$ of $\mathbb{K}$ such that $aa' = a'a = 1$.

Of course, multiplicative inverses don’t always exist. In $\mathbb{Q}$, the number 0 has none. In $\mathbb{Z}$, the number 2 has none. But when they do exist, they are unique:

Theorem 5.4.3. Let $a \in \mathbb{K}$. Then, $a$ has at most one multiplicative inverse.

Proof. This is analogous to Theorem 5.3.2, but we have to replace $+$ and 0 by $\cdot$ and 1.

Warning: In Definition 5.3.3, we could have replaced “$a + a' = a' + a = 0$” by “$a + a' = 0$”, since $a + a' = a' + a$ already follows from commutativity of addition. But in Definition 5.4.2, we cannot replace “$aa' = a'a = 1$” by “$aa' = 1$”, since $\mathbb{K}$ need not be commutative. If we require $aa' = 1$ only, then $a'$ is just a right inverse of $a$; such a right inverse is not necessarily unique.

Definition 5.4.4. (a) An element $a \in \mathbb{K}$ is said to be invertible if it has a multiplicative inverse. An invertible element is also called a unit.
(b) If $a \in \mathbb{K}$ is invertible, then the multiplicative inverse of $a$ will be called $a^{-1}$.
(This is well-defined, since Theorem 5.4.3 shows that this multiplicative inverse is unique.)
(c) Assume that $\mathbb{K}$ is commutative. If $a \in \mathbb{K}$ and $b \in \mathbb{K}$ are such that $b$ is invertible, then we define the quotient $a/b$ (also called $\frac{a}{b}$) to be the element $ab^{-1}$ of $\mathbb{K}$. This new binary partial operation $/$ on $\mathbb{K}$ is called “division.”
The word “partial” in “partial operation” means that it is not always defined. We already have seen this for rational numbers: We cannot divide by 0.

Again, we follow PEMDAS rules as far as division is concerned. Do not use the ambiguous expression “$a/bc$”; it can mean either $a/(bc)$ or $(a/b)c$, depending on whom you ask, and thus should always be parenthesized.

The notion of “unit” we have just defined generalizes the units of $\mathbb{Z}[i]$. Don’t confuse “unit” (= invertible element) with “unity” (= $1_K$). The unity is always a unit (by Exercise 5.4.1 (a) further below), but often not the only unit.

Definition 5.4.4 (c) generalizes the usual meaning of $a/b$ in $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$.

Please do not use Definition 5.4.4 (c) when $K$ is not commutative; that would cause confusion, since $ab^{-1}$ and $b^{-1}a$ would have equal rights to the name “$\frac{a}{b}$”.

If $K = \mathbb{Z}/n$ for a positive integer $n$, and if $a \in K$, then the multiplicative inverse of $a$ is the same as an inverse of $a$ (as defined in Definition 3.5.2). Thus, multiplicative inverses in arbitrary rings generalize the concept of inverses in $\mathbb{Z}/n$. Likewise, they generalize inverses in $\mathbb{C}$; that is, an inverse of a complex number $a \in \mathbb{C}$ (as defined in Definition 4.1.11) is the same as a multiplicative inverse of $a$.

Again, it is not hard to check that multiplicative inverses and division have the properties you would hope them to have:

**Exercise 5.4.1.** Prove the following:

(a) The element $1_K$ of $K$ is always invertible.

(b) The element $-1_K$ of $K$ is always invertible. (Note that $-1_K$ is not always distinct from $1_K$.)

(c) Let $a \in K$ be invertible. Then, its inverse $a^{-1}$ is invertible as well, and its inverse is $(a^{-1})^{-1} = a$.

(d) Let $a, b \in K$ be invertible. Then, their product $ab$ is invertible as well, and its inverse is $(ab)^{-1} = b^{-1}a^{-1}$. (Mind the order of multiplication: it is $b^{-1}a^{-1}$, not $a^{-1}b^{-1}$.)

(e) Assume that $K$ is commutative. Let $a, b, c, d \in K$ be such that $b$ and $d$ are invertible. Then,

$$a/b + c/d = (ad + bc) / (bd) \quad \text{and} \quad (a/b) (c/d) = (ac) / (bd).$$

Some rings have many invertible elements (such as $\mathbb{Q}$, where each nonzero element is invertible), while others have few (such as $\mathbb{Z}$, whose only invertible elements are 1 and $-1$). The extreme case on the former end is called a *skew field* or a *field*, depending on its commutativity:

**Definition 5.4.5.** (a) An element $a \in K$ is said to be **nonzero** if $a \neq 0$. (Here, of course, 0 means the zero of $K$.)

(b) We say that $K$ is a *skew field* if $0 \neq 1$ in $K$ and if every nonzero $a \in K$ is invertible. (Here, “$0 \neq 1$ in $K$” means “$0_K \neq 1_K$”; we are clearly not requiring the integers 0 and 1 to be distinct.)

(c) We say that $K$ is a *field* if $K$ is a commutative skew field.
The condition “0 ≠ 1 in \( K \)” is technical. It is easy to see that if a ring \( K \) satisfies 0 = 1 in \( K \), then it has only one element (to wit: any \( a \in K \) must satisfy \( a = \frac{1}{0} \cdot a = 0 \cdot a = 0 \)), which entails that \( K \) is the zero ring (up to relabeling of its element 0). We do not want the zero ring to count as a skew field; thus we require 0 ≠ 1 in \( K \).

Some authors call skew fields division rings.

Example 5.4.6. (a) The rings \( \mathbb{Q} \), \( \mathbb{R} \) and \( \mathbb{C} \) are fields.

(b) The rings \( \mathbb{Z} \), \( \mathbb{Z}[i] \) and \( \mathbb{Z}[\sqrt{2}] \) are not fields (since, for example, 2 is not invertible in any of these rings). However, \( \mathbb{Z}[i] \) and \( \mathbb{Z}[\sqrt{2}] \) would become fields if we had used \( \mathbb{Q} \) instead of \( \mathbb{Z} \) in their definitions.

(c) The polynomial ring \( \mathbb{Z}[x] \) is not a field (since, for example, \( x \) is not invertible in it). There is a way to get a field out of it, similarly to how \( \mathbb{Q} \) is obtained from \( \mathbb{Z} \) (more on that later perhaps).

(d) Recall the commutative ring \( \mathbb{Q}^Q \); the elements of this ring are functions from \( \mathbb{Q} \) to \( \mathbb{Q} \), and the operations + and \( \cdot \) are defined pointwise. Is this ring a field?

Let us see what the multiplicative inverse of a function \( f \in \mathbb{Q}^Q \) is. If \( f, g \in \mathbb{Q}^Q \) are two functions, then we have the following chain of equivalences:

\[
(g \text{ is the multiplicative inverse of } f) \iff (fg = gf = 1_{\mathbb{Q}^Q}) \iff ((fg)(x) = (gf)(x) = 1_{\mathbb{Q}^Q}(x) \text{ for all } x \in \mathbb{Q}) \iff (f(x) \cdot g(x) = g(x) \cdot f(x) = 1 \text{ for all } x \in \mathbb{Q})
\]

since each \( x \in \mathbb{Q} \) satisfies \((fg)(x) = f(x) \cdot g(x) \) and \((gf)(x) = g(x) \cdot f(x) \) and \( 1_{\mathbb{Q}^Q}(x) = 1 \).

\[
\iff (g(x) = \frac{1}{f(x)} \text{ for all } x \in \mathbb{Q})
\]

(Note that this is not the same as saying that \( f \) and \( g \) are inverse maps! The multiplication of \( \mathbb{Q}^Q \) is not given by composition of maps.)

This shows that a function \( f \in \mathbb{Q}^Q \) is invertible in \( \mathbb{Q}^Q \) if and only if it never takes the value 0 (because its multiplicative inverse \( g \) would have to satisfy \( g(x) = \frac{1}{f(x)} \) for all \( x \in \mathbb{Q} \)). But a function \( f \in \mathbb{Q}^Q \) can be 0 at some point and \( \neq 0 \) at another. Then, it is not invertible (since it is 0 at some point) yet nonzero (since it is \( \neq 0 \) at another). For example, the function \( \text{id} \in \mathbb{Q}^Q \) is not invertible yet nonzero. Thus, \( \mathbb{Q}^Q \) is not a field.

(e) The ring \( \mathbb{Q}^{2\times 2} \) of \( 2 \times 2 \)-matrices with rational entries is not a skew field. Indeed, the \( 2 \times 2 \)-matrix \( \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \) is nonzero but not invertible. (More generally:
For each \( n \in \mathbb{N} \), the \( n \times n \)-matrices over \( \mathbb{Q} \) form a ring, which we will study later. Our notion of “invertible” for elements of this ring coincides with the usual notion of “invertible” for \( n \times n \)-matrices in linear algebra.)

What about \( \mathbb{Z}/n \)?

**Theorem 5.4.7.** Let \( n \) be a positive integer. The ring \( \mathbb{Z}/n \) is a field if and only if \( n \) is prime.

**Proof of Theorem 5.4.7** \( \iff \): Assume that \( n \) is prime. We must prove that \( \mathbb{Z}/n \) is a field.

First of all, \( n > 1 \) (since \( n \) is prime). Thus, \( n \nmid 1 \), so that \( 1 \neq 0 \mod n \), so that \( 0 \neq 1 \mod n \). In other words, \([0]_n \neq [1]_n\). In other words, \( 0 \neq 1 \) in \( \mathbb{Z}/n \) (since \( 0_{\mathbb{Z}/n} = [0]_n \) and \( 1_{\mathbb{Z}/n} = [1]_n \)).

Also, the ring \( \mathbb{Z}/n \) is commutative.

So it remains to prove that every nonzero \( \alpha \in \mathbb{Z}/n \) is invertible (because then, we will immediately conclude that \( \mathbb{Z}/n \) is a skew field and therefore a field).

Indeed, let \( \alpha \in \mathbb{Z}/n \) be nonzero. We must prove that \( \alpha \) is invertible.

Proposition 3.4.6 (b) shows that there exists a unique \( a \in \{0, 1, \ldots, n-1\} \) satisfying \( \alpha = [a]_n \). Consider this \( a \). If we had \( a = 0 \), then \( \alpha = [a]_n \) would become \( \alpha = [0]_n = 0_{\mathbb{Z}/n} \), which would contradict the assumption that \( \alpha \) is nonzero. So \( a \neq 0 \). Thus, \( a \in \{1, 2, \ldots, n-1\} \) (since \( a \in \{0, 1, \ldots, n-1\} \)). Hence, Proposition 2.13.4 (applied to \( i = a \) and \( p = n \)) shows that \( a \) is coprime to \( n \). In other words, \( a \perp n \).

Now, define a set \( U_n \) as in Corollary 3.5.5. Then, Corollary 3.5.5 (a) yields \([a]_n \in U_n \) (since \( a \perp n \)). In other words, \([a]_n \) has an inverse (by the definition of \( U_n \)). In other words, \( \alpha \) has an inverse (since \( \alpha = [a]_n \)). In other words, \( \alpha \) has a multiplicative inverse (because inverses in \( \mathbb{Z}/n \) are precisely what we now call multiplicative inverses). In other words, \( \alpha \) is invertible. So we have proven the “\( \iff \)” direction of Theorem 5.4.7.

\( \implies \): Rough idea: Assume that \( \mathbb{Z}/n \) is a field. We must prove that \( n \) is a prime. Assume the contrary.

We have \( 0 \neq 1 \) in \( \mathbb{Z}/n \) (since \( \mathbb{Z}/n \) is a field); in other words, \( 0 \neq 1 \mod n \). This quickly yields \( n > 1 \). Hence, there must exist two elements \( d, e \in \{1, 2, \ldots, n-1\} \) such that \( n = de \) (since \( n \) is not a prime). Consider these \( d \) and \( e \). Theorem 3.4.4 shows that the \( n \) residue classes \([0]_n, [1]_n, \ldots, [n-1]_n\), are distinct. Hence, the residue classes \([d]_n \) and \([e]_n \) are nonzero (since \( d, e \in \{1, 2, \ldots, n-1\} \) are distinct from 0). Thus, these two residue classes are invertible (since \( \mathbb{Z}/n \) is a field). Thus, by Exercise 5.4.1 (d), their product \([d]_n [e]_n \) is invertible as well. But this product is \([d]_n [e]_n = [de]_n = [n]_n = [0]_n = 0_{\mathbb{Z}/n} \), which is not invertible. Contradiction. Thus, the “\( \implies \)” direction of Theorem 5.4.7 is proven.

\[ 2019-04-01 \text{ lecture} \]

Now, for an example of a skew field that is not a field.
Example 5.4.8. Informally, we have obtained \( \mathbb{C} \) from \( \mathbb{R} \) by throwing in a new number \( i \) that satisfies \( i^2 = -1 \). In order for \( i \) not to feel alone, let us introduce yet another new “number” \( j \) such that \( j^2 = -1 \) and \( ji = -ij \). Now we try to calculate with these \( i \) and \( j \). Of course, \( i \) and \( j \) cannot belong to a commutative ring together, but let us assume that they (and the further numbers we obtain from them) at least satisfy the ring axioms.

We have

\[
i \cdot ij = ii \cdot j = (-1) j = -j \quad \text{and} \quad \frac{j \cdot ij = ji \cdot j = -i \cdot jj = -i (-1) = i}{= -ij \quad = i^2 = -1}
\]

and (using the distributivity laws)

\[
(1 + 2i + 3ij) (2 - 3j) = 2 + 4i + 6ij - 3j - 6ij - 9i \cdot j^2 = 2 + 4i - 3j + 9i = 2 + 13i - 3j.
\]

Similarly, any of these new “numbers” can be written in the form \( a + bi + cj + dij \) for reals \( a, b, c, d \).

Blithely introducing new “numbers” like this can be risky. It could happen that (just as with defining \( \infty \) to be \( \frac{1}{0} \)) our new numbers would lead to contradictions. For example, what if we have some expression that involves \( i \) and \( j \) and that can be simplified to 0 in one way and simplified to 1 in another; would that mean that \( 0 = 1 \)? No; it would simply mean that the new “numbers” we have introduced do not actually exist. (Or, speaking more abstractly: that the new numbers are just the zero ring in a complicated disguise.)

So it makes sense to look for a rigorous definition of our new numbers. There is a direct (though rather painful) way of doing this: We can rigorously define our new numbers as 4-tuples \((a, b, c, d)\) of real numbers, with addition and subtraction defined entrywise, and with multiplication given by

\[
(\begin{array}{c}
(x_1, x_2, x_3, x_4) \\
(y_1, y_2, y_3, y_4)
\end{array}) \cdot (\begin{array}{c}
(x_1, x_2, x_3, x_4) \\
(y_1, y_2, y_3, y_4)
\end{array}) = (x_1y_1 - x_2y_2 - x_3y_3 - x_4y_4, x_1y_2 + x_2y_1 + x_3y_4 - x_4y_3, x_1y_3 - x_2y_4 + x_3y_1 + x_4y_2, x_1y_4 + x_2y_3 - x_3y_2 + x_4y_1).
\]

These new numbers are \textbf{quaternions}. It turns out that they form a skew field, albeit not a field (since commutativity is lacking). They have several properties that make them useful in physics and space geometry. For one, they encode both the scalar product and the cross product of two vectors in \( \mathbb{R}^3 \): Namely,
If \( \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \in \mathbb{R}^3 \) and \( \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in \mathbb{R}^3 \) are two vectors, then the quaternion

\[
(0, a_1, a_2, a_3) \cdot (0, b_1, b_2, b_3)
\]

\[
= \begin{pmatrix} -a_1b_1 - a_2b_2 - a_3b_3, a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1 \end{pmatrix}.
\]

(where \( \cdot \) stands for the scalar product)

Also, the quaternions can be used to encode rotations in 3-dimensional space.

5.5. Hunting for finite fields I

**Definition 5.5.1.** (a) The *ground set* of a ring \((\mathbb{K}, +, \cdot, 0, 1)\) is defined to be the set \(\mathbb{K}\).

(b) The *elements* of a ring are defined to be the elements of its ground set.

(c) The *size* (or *cardinality*) of a ring is defined to be the size of its ground set.

(d) A ring is said to be *finite* if its size is finite (i.e., if it has only finitely many elements).

We have seen a bunch of finite rings. For example, if \(S\) is a finite set, then the commutative ring \((\mathcal{P}(S), \triangle, \cap, \emptyset, S)\) (see one of the examples above) has size \(|\mathcal{P}(S)| = 2^{|S|}\), and thus is finite.

We also have seen infinitely many finite fields:

\[\mathbb{Z}/2, \quad \mathbb{Z}/3, \quad \mathbb{Z}/5, \quad \mathbb{Z}/7, \quad \mathbb{Z}/11, \quad \ldots\]

Indeed, Theorem 5.4.7 yields that \(\mathbb{Z}/p\) is a finite field whenever \(p\) is a prime.

**Question 5.5.2.** Are there any further finite fields?

**Remark 5.5.3.** Recall Shamir’s Secret Sharing Scheme, which we introduced in Subsection 1.6.7. The way we defined the Scheme, it had a problem: It relied on a spurious notion of a “uniformly random rational number”, which does not exist in nature. Now we can fix this problem: Replace rational numbers by elements of a finite field. More precisely, let \(N\) again be the length of the bitstring that we want to encrypt. Pick a prime \(p\) that satisfies \(p > N\); this exists due to Theorem 2.13.43. Then, the finite field \(\mathbb{Z}/p\) has more than \(p\) elements. Now, use elements of \(\mathbb{Z}/p\) instead of integers. (Thus, a bitstring \(a_{N-1}a_{N-2} \cdots a_0\) will be encoded as the residue class \([a_{N-1}a_{N-2} \cdots a_0]_p \in \mathbb{Z}/p\) rather than as the number \(a_{N-1}a_{N-2} \cdots a_0 \in \mathbb{Z}\).) Instead of picking two uniformly random bitstrings \(c\) and
b and transforming them into numbers c and b, just pick two uniformly random residue classes c, b ∈ ℤ/p. (This is possible, since ℤ/p is a finite set.)

This relies on having a well-behaved notion of polynomials over ℤ/p. We will later give a rigorous definition of this notion.

Finite fields have many uses – not just in making Shamir’s Secret Sharing Scheme work. One great source of applications is coding theory, which we might see later.

Let us take some first steps towards addressing Question 5.5.2. We have found a field of size p for each prime p. Are there fields of other sizes?

First idea: Let us try to get such a field by “duplicating” the known field ℤ/p. Thus, we fix a prime p, and consider the Cartesian product (ℤ/p × ℤ/p). Define addition, subtraction and multiplication on this Cartesian product elementwise (so, e.g., (a,b) + (c,d) = (ac, bd)). This will yield a commutative ring with zero ((0)p, (0)p) and unity ((1)p, (1)p).

However, the element ((0)p, (1)p) of this ring is nonzero (because it is not ((0)p, (0)p)) but has no inverse (since multiplying it by anything will never make its first entry anything other than (0)p). So this ring is not a field.

Second idea: We obtained C from ℜ by “adjoining” a square root of −1. (In abstract algebra, the verb “adjoin” means “insert” or “add” – not in the sense of the addition operation +, but in the sense of throwing in something new into an existing collection.)

Let’s try to do this with ℤ/p instead of ℜ.

More generally, let us start with an arbitrary commutative ring ℜ, and try to “adjoin” a square root of −1 to it. We are bold and don’t care whether there might already be such a square root in ℜ; if there is, then we will get a second one!

Let 0 and 1 stand for the zero and the unity of ℜ. If ℜ = ℤ/n for some integer n, then these are the residue classes [0]n and [1]n.

Now, we want to define a new commutative ring ℜ′ by “adjoining” a square root of −1 to ℜ. A way to make this rigorous is as follows (just as we defined ℂ rigorously in Definition 4.1.1):

Definition 5.5.4. Let ℜ be a commutative ring.

(a) Let ℜ′ be the set of all pairs (a, b) ∈ ℜ × ℜ.

(b) For each r ∈ ℜ, we denote the pair (r, 0) ∈ ℜ′ by rℜ′. We identify r ∈ ℜ with rℜ′ = (r, 0) ∈ ℜ′, so that ℜ becomes a subset of ℜ′.

(c) We let i be the pair (0, 1) ∈ ℜ′.

(d) We define three binary operations +, − and · on ℜ′ by setting

(a, b) + (c, d) = (a + c, b + d),

(a, b) − (c, d) = (a − c, b − d),

(a, b) · (c, d) = (ac − bd, ad + bc)

for all (a, b) ∈ ℜ′ and (c, d) ∈ ℜ′.
(e) If $\alpha, \beta \in K'$, then we write $\alpha \beta$ for $\alpha \cdot \beta$.

You will, of course, recognize this definition to be a calque of Definition 4.1.1 with $\mathbb{R}$ and $\mathbb{C}$ replaced by $K$ and $K'$. The elements of $K'$ are like complex numbers, but built upon $K$ instead of $\mathbb{R}$.

**Proposition 5.5.5.** The set $K'$ defined in Definition 5.5.4 (equipped with the operations $+$ and $\cdot$ and the elements $0_{K'}$ and $1_{K'}$) is a commutative ring.

**Proof.** Same argument as we did for $\mathbb{C}$ in the proof of Theorem 4.1.2. □

**Convention 5.5.6.** For the rest of this section, we let $K'$ be the commutative ring constructed in Proposition 5.5.5 (i.e., the set $K'$ equipped with the operations $+$ and $\cdot$ and the elements $0_{K'}$ and $1_{K'}$).

Thus, if $K = \mathbb{Z}/p$, then $K'$ is a commutative ring with $p^2$ elements.

**Question 5.5.7.** When is $K'$ a field?

Assume that $0 \neq 1$ in $K$; thus, $0 \neq 1$ in $K'$ as well (since $0_{K'} = (0,0) \neq (1,0) = 1_{K'}$). Hence, in order for $K'$ to be a field, every nonzero $\xi \in K'$ needs to have a multiplicative inverse. Thus, in particular, every nonzero element of $K$ must have a multiplicative inverse in $K'$. It is easy to see that such an inverse, if it exists, must belong to $K$ as well (i.e., it must have the form $r_{K'}$ for some $r \in K$); thus, this means that every nonzero element of $K$ must have a multiplicative inverse in $K$. In other words, $K$ itself must be a field.

Thus, we assume from now on that $K$ is a field. But we are not done yet. It is definitely not always true that $K'$ is a field. For example, if $K = \mathbb{Z}/2$, then the element $(1,1)$ of $K'$ has no inverse (check this!), and so $K'$ is not a field in this case. What must $K$ satisfy in order for $K'$ to be a field?

We know what it must satisfy: The condition is that every nonzero $\xi \in K'$ has a multiplicative inverse. We just need to see when this condition holds.

So let $\xi = (x,y) \in K'$ (with $x,y \in K$) be nonzero. Thus, $(x,y) \neq (0,0)$.

How to find $\xi^{-1}$? Notice that $\xi = (x,y) = x + yi$ (this is proven just as for complex numbers). Thus, you can try to compute $\xi^{-1}$ by rationalizing the denominator (just as we learned to divide complex numbers):

$$
\frac{1}{\xi} = \frac{1}{x + yi} = \frac{x - yi}{(x + yi)(x - yi)} = \frac{x - yi}{x^2 + y^2}
$$

(since $(x + yi)(x - yi) = (x,y)(x,-y) = (x^2 + y^2,0)$, as you can easily see using the definition of $\cdot$ on $K'$).

We need $x^2 + y^2 \neq 0$ in $K$ for this to work. In other words, we need the following condition to hold:
Condition 1: For every pair \((x, y) \in K \times K\) satisfying \((x, y) \neq (0, 0)\), we have \(x^2 + y^2 \neq 0\) in \(K\).

Thus, \(K'\) is a field if Condition 1 holds. Conversely, if \(K'\) is a field, then Condition 1 holds (because if \((x, y) \in K \times K\) satisfies \((x, y) \neq (0, 0)\), then \((x, y)(x, -y) = (x^2 + y^2, 0)\) would have to be \(\neq (0, 0)\) in order for \(K'\) to be a field). So \(K'\) is a field if and only if Condition 1 holds.

If \(K = \mathbb{Z}/p\) for some prime \(p\), then Condition 1 can be restated as follows:

Condition 1': For every pair \((x, y) \in (\mathbb{Z}/p) \times (\mathbb{Z}/p)\) satisfying \((x, y) \neq (0, 0)\), we have \(x^2 + y^2 \neq 0\) in \(\mathbb{Z}/p\).

We can further restate Condition 1' in terms of integers by replacing the residue classes \(x\) and \(y\) with their representatives \(a\) and \(b\):

Condition 2: For every pair \((a, b) \in \mathbb{Z} \times \mathbb{Z}\) such that \(\text{not both} a\) and \(b\) are divisible by \(p\), the sum \(a^2 + b^2\) is not divisible by \(p\).

So the ring \(K'\) constructed from \(K = \mathbb{Z}/p\) is a field if and only if Condition 2 holds. When does Condition 2 hold?

Example 5.5.8. Let \(K = \mathbb{Z}/p\).

(a) If \(p = 2\), then Condition 2 fails for \((a, b) = (1, 1)\). So \(K'\) is not a field for \(p = 2\).

(b) If \(p = 3\), then Condition 2 holds. So \(K'\) is a field for \(p = 3\). Thus we have found a field with \(3^2 = 9\) elements.

(c) If \(p = 5\), then Condition 2 fails for \((a, b) = (1, 2)\). So \(K'\) is not a field for \(p = 5\).

This suggests that the following:

Proposition 5.5.9. A prime \(p\) satisfies Condition 2 if and only if \(p \equiv 3 \mod 4\).

Proof. \(\implies\): Assume that a prime \(p\) satisfies Condition 2. Assume (for contradiction) that \(p \not\equiv 3 \mod 4\). So \(p\) is a prime of Type 1 or 2. Thus, \(p = x^2 + y^2\) for two integers \(x, y\) (by Theorem 4.2.40(a)). Now, \((x, y)\) is a pair in \(\mathbb{Z} \times \mathbb{Z}\) such that \(\text{not both} x\) and \(y\) are divisible by \(p\) (why not?), but the sum \(x^2 + y^2 = p\) is divisible by \(p\). So Condition 2 fails for \((a, b) = (x, y)\).

\(\impliedby\): Assume that a prime \(p\) satisfies \(p \equiv 3 \mod 4\). Thus, \((p - 1)/2\) is an odd nonnegative integer.

We must prove that Condition 2 holds. In other words, we must prove that for every pair \((a, b) \in \mathbb{Z} \times \mathbb{Z}\) such that \(\text{not both} a\) and \(b\) are divisible by \(p\), the sum \(a^2 + b^2\) is not divisible by \(p\).

Let \((a, b) \in \mathbb{Z} \times \mathbb{Z}\) be a pair such that \(\text{not both} a\) and \(b\) are divisible by \(p\). We must prove that the sum \(a^2 + b^2\) is not divisible by \(p\).
Assume the contrary. Thus, \(a^2 + b^2 \equiv 0 \mod p\).

If we had \(p \mid a\), then we would have \(a \equiv 0 \mod p\) and thus \(a^2 + b^2 \equiv 0^2 + b^2 = b^2 \mod p\), so that \(b^2 \equiv a^2 + b^2 \equiv 0 \mod p\) and thus \(p \mid b^2\) and therefore \(p \mid b\); but this would contradict our assumption that not both \(a\) and \(b\) are divisible by \(p\). Hence, we cannot have \(p \mid a\). Thus, we have \(p \nmid a\). Hence, Fermat’s little theorem yields \(a^{p-1} \equiv 1 \mod p\). Similarly, \(b^{p-1} \equiv 1 \mod p\).

From \(a^2 + b^2 \equiv 0 \mod p\), we get \(a^2 \equiv -b^2 \mod p\). Taking this congruence to the \((p-1)/2\)-th power\(^{140}\), we find

\[
\left( a^2 \right)^{(p-1)/2} \equiv \left( -b^2 \right)^{(p-1)/2} = \left( -1 \right)^{(p-1)/2} \left( b^2 \right)^{(p-1)/2} \equiv -1 \mod p.
\]

Hence,

\[
-1 \equiv \left( a^2 \right)^{(p-1)/2} = a^{p-1} \equiv 1 \mod p.
\]

Hence, \(p \mid (-1) - 1 = -2 \mid 2\), so \(p = 2\). This contradicts \(p \equiv 3 \mod 4\). This contradiction shows that our assumption was false; thus, Condition 2 holds. \(\square\)

Thus, if we set \(K = \mathbb{Z}/p\) where \(p\) is a prime of Type 3, then \(K'\) will be a field. So we have found a field \(K'\) with \(p^2\) elements for any prime \(p\) of Type 3. What about the other primes?

We can try to vary the construction above: Instead of adjoining a square root of \(-1\), we adjoin a square root of some other element \(\eta \in \mathbb{Z}/p\).

**Definition 5.5.10.** Let \(K\) be a ring. A square (in \(K\)) means an element of the form \(a^2\) for some \(a \in K\).

Now, we generalize Definition 5.5.10 as follows:

**Definition 5.5.11.** Let \(K\) be a commutative ring. Let \(\eta \in K\).

(a) Let \(K'_\eta\) be the set of all pairs \((a, b) \in K \times K\).

(b) For each \(r \in K\), we denote the pair \((r, 0) \in K'_\eta\) by \(r_{K'_\eta}\). We identify \(r \in K\) with \(r_{K'_\eta} = (r, 0) \in K'_\eta\), so that \(K\) becomes a subset of \(K'_\eta\).

(c) We let \(i_{\eta} \) be the pair \((0, 1) \in K'_\eta\).

(d) We define three binary operations \(+, -\), and \(\cdot\) on \(K'_\eta\) by setting

\[
(a, b) + (c, d) = (a + c, b + d), \quad (a, b) - (c, d) = (a - c, b - d), \quad \text{and} \quad (a, b) \cdot (c, d) = (ac + \eta bd, ad + bc)
\]

for all \((a, b) \in K'_\eta\) and \((c, d) \in K'_\eta\).

(e) If \(\alpha, \beta \in K'_\eta\), then we write \(\alpha \beta\) for \(\alpha \cdot \beta\).

\(\text{\footnotesize \(^{140}\)We can do this, since \((p-1)/2\) is a nonnegative integer.}\)
Note that \( K'_\eta \) differs from \( K' \) only in how the multiplication is defined.

**Theorem 5.5.12.** (a) The set \( K'_\eta \) defined in Definition 5.5.11 (equipped with the operations \(+\) and \(\cdot\) and the elements 0\(_{K'_\eta}\) and 1\(_{K'_\eta}\)) is a commutative ring.

(b) If \( K \) is a field and \( \eta \) is not a square in \( K \), then \( K'_\eta \) is a field.

(c) Let \( p \) be a prime. There always exists an element \( \eta \in \mathbb{Z}/p \) that is not a square, unless \( p = 2 \).

**Proof.** (a) Similar to the proof for \( C \).

(b) Assume that \( K \) is a field and that \( \eta \) is not a square in \( K \). We need to prove that \( K'_\eta \) is a field. In other words, we need to prove that each nonzero \( \xi \in K'_\eta \) is invertible.

So let \( \xi \in K'_\eta \) be nonzero, and write \( \xi \) as \( \xi = (x, y) \) with \( x, y \in K \). We must show that \( \xi \) is invertible.

We have \( (x, y)(x, -y) = (x^2 - \eta y^2, 0) \) (by the definition of the operation \(\cdot\) on \( K'_\eta \)). If \( x^2 - \eta y^2 \) is nonzero, then this shows quickly that \( \left( \frac{x}{x^2 - \eta y^2}, \frac{-y}{x^2 - \eta y^2} \right) \) is an inverse of \( (x, y) = \xi \), and thus \( \xi \) is invertible. So we need to prove that \( x^2 - \eta y^2 \) is nonzero.

Assume the contrary. Thus, \( x^2 - \eta y^2 = 0 \), so that \( x^2 = \eta y^2 \). If \( y \) is nonzero, then this can be rewritten as \( \frac{x^2}{y^2} = \eta \), whence \( \eta = \left( \frac{x}{y} \right)^2 \), which contradicts the fact that \( \eta \) is not a square. So \( y \) cannot be nonzero. Thus, \( y = 0 \). Hence, \( x^2 = \eta \cdot \frac{y^2}{0^2} = 0 \). Since \( \xi = \left( x, \frac{y}{0} \right) = (x, 0) \), we know that \( x \) is nonzero (since \( \xi \) is nonzero). Hence, \( x \) has a multiplicative inverse (since \( K \) is a field). Hence, multiplying \( x^2 = 0 \) by \( x^{-1} \), we obtain \( x = 0 \), which contradicts \( x \) being nonzero. So our assumption was wrong. Part (b) is proven.

(c) Assume that \( p \neq 2 \). Thus, \( p > 2 \).

Consider the map

\[
\mathbb{Z}/p \rightarrow \mathbb{Z}/p, \quad \alpha \mapsto \alpha^2.
\]

This map is not injective (since it sends the two distinct residue classes \([1]_p\) and \([p - 1]_p\) both to \([1]_p\)). Hence, it cannot be surjective either (since otherwise, the Pigeonhole Principle for Surjections would entail that it is bijective, hence injective). In other words, there exists some \( \eta \in \mathbb{Z}/p \) that is not in its image. In other words, there exists an element \( \eta \in \mathbb{Z}/p \) that is not a square.

Now, if \( p \) is a prime with \( p > 2 \), then part (c) of the above Theorem yields that there exists an element \( \eta \in \mathbb{Z}/p \) that is not a square; therefore, part (b) of the above Theorem shows that \( K'_\eta \) is a field where \( K = \mathbb{Z}/p \). This is a field with \( p^2 \) elements.

Is there a field of size 4, too?
We cannot get such a field by adjoining a square root to \( \mathbb{Z}/2 \). So let us instead try to adjoin an element \( j \) such that \( j^2 = j + 1 \). So we define \( \mathbb{K}'' \) as the set of all pairs \( (a, b) \in (\mathbb{Z}/2) \times (\mathbb{Z}/2) \), and we define \(+, -, \cdot\) on \( \mathbb{K}'' \) by
\[
(a, b) + (c, d) = (a + c, b + d), \\
(a, b) - (c, d) = (a - c, b - d), \quad \text{and} \\
(a, b) \cdot (c, d) = (ac + bd, ad + bc + bd).
\]
You can check that this is a field with 4 elements.

Thus, for each prime \( p \), we have found a field with \( p^2 \) elements.

For the sake of completeness, let me mention a third idea for constructing fields of size \( p^2 \): Recall that our field \( \mathbb{Z}/p \) of size \( p \) consisted of residue classes of integers modulo \( p \). What happens if we take the residue classes of Gaussian integers modulo a Gaussian prime \( \pi \)?

I will not go into details, but here is a summary:

- The result is always a field of size \( N(\pi) \).
- If \( \pi \) is not unit-equivalent to an integer, then this is a field that we already know (namely, \( \mathbb{Z}/p \) for \( p = N(\pi) \)) with its elements relabelled.
- If \( \pi \) is unit-equivalent to an integer, then \( \pi \) is unit-equivalent to a prime \( p \) of Type 3, and the field of residue classes modulo \( \pi \) will be a field with \( p^2 \) elements. Namely, it will be the field \( \mathbb{K}' \) we constructed above (for \( \mathbb{K} = \mathbb{Z}/p \)), with its elements relabelled.

So this approach only gets us fields of size \( p^2 \) when \( p \) is a prime of Type 3; it is thus inferior to the second idea above. Nevertheless, it illustrates a general idea: that residue classes make sense not only for integers.

\textbf{2019-04-03 lecture}

\textbf{Warning:} When \( p \) is a prime, \( \mathbb{Z}/p^2 \) is not a field; thus, the field with \( p^2 \) elements that we constructed is not \( \mathbb{Z}/p^2 \).

Now, what about finite fields of size \( p^3, p^4, \ldots \) ? What about finite fields of size 6 ?

\textbf{Spoiler:} It turns out that the former exist, while the latter do not. We will hopefully prove this later. More generally, for an integer \( n > 1 \), there exists a field of size \( n \) if and only if \( n \) is a prime power (= power of a prime). Even better, if \( n \) is a prime power, then a field of size \( n \) is unique up to relabeling. We hope to see a proof of this (at least of the existence part) further on in this class.

\section*{5.6. Cartesian products}
Definition 5.6.1. Let $\mathbb{K}_1, \mathbb{K}_2, \ldots, \mathbb{K}_n$ be $n$ rings. Consider the set $\mathbb{K}_1 \times \mathbb{K}_2 \times \cdots \times \mathbb{K}_n$, whose elements are $n$-tuples $(k_1, k_2, \ldots, k_n)$ with $k_i \in \mathbb{K}_i$.

We define operations $+$ and $\cdot$ on $\mathbb{K}_1 \times \mathbb{K}_2 \times \cdots \times \mathbb{K}_n$ by

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$

and

$$(a_1, a_2, \ldots, a_n) \cdot (b_1, b_2, \ldots, b_n) = (a_1 b_1, a_2 b_2, \ldots, a_n b_n).$$

Proposition 5.6.2. Let $\mathbb{K}_1, \mathbb{K}_2, \ldots, \mathbb{K}_n$ be $n$ rings.

(a) The set $\mathbb{K}_1 \times \mathbb{K}_2 \times \cdots \times \mathbb{K}_n$, endowed with the operations $+$ and $\cdot$ we just defined and with the zero $(0, 0, \ldots, 0)$ and the unity $(1, 1, \ldots, 1)$, is a ring.

(b) If the rings $\mathbb{K}_1, \mathbb{K}_2, \ldots, \mathbb{K}_n$ are commutative, then so is the ring $\mathbb{K}_1 \times \mathbb{K}_2 \times \cdots \times \mathbb{K}_n$.

Proof. All axioms are checked entrywise: For example, associativity of multiplication follows from comparing

$$((a_1, a_2, \ldots, a_n) \cdot (b_1, b_2, \ldots, b_n)) \cdot (c_1, c_2, \ldots, c_n) = (a_1 b_1, a_2 b_2, \ldots, a_n b_n) \cdot (c_1, c_2, \ldots, c_n) = (a_1 b_1 c_1, a_2 b_2 c_2, \ldots, a_n b_n c_n)$$

with

$$(a_1, a_2, \ldots, a_n) \cdot ((b_1, b_2, \ldots, b_n) \cdot (c_1, c_2, \ldots, c_n)) = (a_1, a_2, \ldots, a_n) \cdot (b_1 c_1, b_2 c_2, \ldots, b_n c_n) = (a_1 b_1 c_1, a_2 b_2 c_2, \ldots, a_n b_n c_n).$$

The additive inverse of $(a_1, a_2, \ldots, a_n)$ is $(-a_1, -a_2, \ldots, -a_n)$. \qed

Definition 5.6.3. The ring $\mathbb{K}_1 \times \mathbb{K}_2 \times \cdots \times \mathbb{K}_n$ constructed in Proposition 5.6.2 is called the Cartesian product (or direct product) of the rings $\mathbb{K}_1, \mathbb{K}_2, \ldots, \mathbb{K}_n$.

We have already seen a Cartesian product. Indeed, recall the binary operations $\text{XOR}$ defined back in Subsection 1.6.4. We first defined an operation $\text{XOR}$ on bits (Definition 1.6.3), and then defined an operation $\text{XOR}$ on bitstrings (Definition 1.6.4). It is easy to see that

$$\{0, 1\}, \text{XOR}, \cdot, 0, 1$$

is a commutative ring. Let me call this ring $\mathbb{X}$ for now. Note that this ring $\mathbb{X}$ can be seen as $\mathbb{Z}/2$ with its elements relabeled (more precisely, the elements $[0]_2$ and $[1]_2$ of $\mathbb{Z}/2$ need to be relabeled as 0 and 1 in order to get $\mathbb{X}$); for example, the
correspondence between the XOR operation on $X$ and the addition on $\mathbb{Z}/2$ can be seen by comparing their results face to face:

$$
\begin{align*}
0 \text{ XOR } 0 &= 0 & \text{ and } & [0]_2 + [0]_2 = [0]_2, \\
0 \text{ XOR } 1 &= 1 & \text{ and } & [0]_2 + [1]_2 = [1]_2, \\
1 \text{ XOR } 0 &= 1 & \text{ and } & [1]_2 + [0]_2 = [1]_2, \\
1 \text{ XOR } 1 &= 0 & \text{ and } & [1]_2 + [1]_2 = [0]_2.
\end{align*}
$$

In Definition 1.6.4, we defined a binary operation XOR on $\{0, 1\}^n$, i.e., on bitstrings. This gives a ring

$$
(\{0, 1\}^n, \text{XOR, entrywise multiplication, } 00 \cdots 0, 11 \cdots 1)
$$

of bitstrings. This ring is precisely the Cartesian product

$$
\underbrace{X \times X \times \cdots \times X}_{m \text{ times}}.
$$

### 5.7. Matrices and matrix rings

**Convention 5.7.1.** In this section, we fix a ring $K$.

We take the familiar concept of matrices, and generalize it in a straightforward way, allowing matrices with entries in $K$:

**Definition 5.7.2.** Given $n, m \in \mathbb{N}$, we define an $n \times m$-matrix over $K$ to be a rectangular table with $n$ rows and $m$ columns whose entries are elements of $K$. When $K$ is clear from the context (or irrelevant), we just say “$n \times m$-matrix” instead of “$n \times m$-matrix over $K$”.

For example, if $K = \mathbb{Q}$, then

$$
\begin{pmatrix}
0 & 1/3 & -6 \\
-1 & -2/5 & 1
\end{pmatrix}
$$

is a $2 \times 3$-matrix over $\mathbb{K}$.

(Formally, an $n \times m$-matrix is defined as a map from $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$ to $K$. Its entry in row $i$ and column $j$ is then defined to be the image of the pair $(i, j)$ under this map.)

Note that the “$\times$” symbol in the notion of an “$n \times m$-matrix” is just a symbol, not an invitation to actually multiply the numbers $n$ and $m$ together! For example, $2 \cdot 3 = 3 \cdot 2$, yet a $2 \times 3$-matrix is not the same as a $3 \times 2$-matrix.

Let us define two pieces of notation:
Definition 5.7.3. Let $A$ be an $n \times m$-matrix over $K$. Let $i \in \{1, 2, \ldots, n\}$ and $j \in \{1, 2, \ldots, m\}$. The $(i, j)$-th entry of $A$ is defined to be the entry of $A$ in row $i$ and column $j$.

Definition 5.7.4. Let $n, m \in \mathbb{N}$. Assume that we are given some element $a_{i,j} \in K$ for every $(i, j) \in \{1, 2, \ldots, n\} \times \{1, 2, \ldots, m\}$. Then, we shall use the notation

$$(a_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$$ (156)

for the $n \times m$-matrix

$$
\begin{pmatrix}
  a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\
  a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n,1} & a_{n,2} & \cdots & a_{n,m}
\end{pmatrix}
$$

(this is the $n \times m$-matrix whose $(i, j)$-th entry is $a_{i,j}$ for all $i$ and $j$).

For example,

$$(i + j)_{1 \leq i \leq 3, 1 \leq j \leq 4} = \begin{pmatrix}
  2 & 3 & 4 & 5 \\
  3 & 4 & 5 & 6 \\
  4 & 5 & 6 & 7
\end{pmatrix} \quad \text{and} \quad
(i - j)_{1 \leq i \leq 3, 1 \leq j \leq 4} = \begin{pmatrix}
  0 & -1 & -2 & -3 \\
  1 & 0 & -1 & -2 \\
  2 & 1 & 0 & -1
\end{pmatrix}.$$

The letters $i$ and $j$ in the notation (156) are not set in stone; we can use any other letters instead. For example,

$$(i - j)_{1 \leq i \leq 3, 1 \leq j \leq 4} = (x - y)_{1 \leq x \leq 3, 1 \leq y \leq 4} = (j - i)_{1 \leq j \leq 3, 1 \leq i \leq 4}.$$ 

Definition 5.7.5. Let $n, m \in \mathbb{N}$. Then, $K^{n \times m}$ will denote the set of all $n \times m$-matrices. (Some call it $M_{n,m}(K)$ instead.)

Again, the “$\times$” symbol in this notation is just a symbol; it does not stand for a product of numbers.

Definition 5.7.6. (a) A **matrix** means an $n \times m$-matrix for some $n, m \in \mathbb{N}$. 

(b) A **square matrix** means an $n \times n$-matrix for some $n \in \mathbb{N}$.

For example, \( \begin{pmatrix} 1 & 2 & 6 \\ 3 & 4 & 5 \end{pmatrix} \) is a matrix, and \( \begin{pmatrix} 2 & 6 \\ 4 & 5 \end{pmatrix} \) is a square matrix.

We now define various operations with matrices:
Definition 5.7.7. Fix $n, m \in \mathbb{N}$.

(a) The sum $A + B$ of two $n \times m$-matrices $A$ and $B$ is defined entrywise: i.e., if $A = (a_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$ and $B = (b_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$, then

$$A + B = (a_{i,j} + b_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}.$$

(b) The difference $A - B$ of two $n \times m$-matrices $A$ and $B$ is defined entrywise: i.e., if $A = (a_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$ and $B = (b_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$, then

$$A - B = (a_{i,j} - b_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}.$$

(c) We define scaling of $n \times m$-matrices as follows: If $\lambda \in \mathbb{K}$ and $A \in \mathbb{K}^{n \times m}$, then the matrix $\lambda A \in \mathbb{K}^{n \times m}$ is defined by multiplying each entry of $A$ by $\lambda$. Formally speaking: if $A = (a_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$, then

$$\lambda A = (\lambda a_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}.$$

To be more honest, the operation we defined in Definition 5.7.7 (c) should have been called “left scaling” rather than “scaling”. And we should have defined an analogous operation called “right scaling”, which takes an element $\lambda \in \mathbb{K}$ and a matrix $A = (a_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m} \in \mathbb{K}^{n \times m}$, and returns a new matrix

$$A \lambda = (a_{i,j} \lambda)_{1 \leq i \leq n, 1 \leq j \leq m}.$$

But we will mostly be dealing with the case when the ring $\mathbb{K}$ is commutative; and in this case, we have $A \lambda = \lambda A$ always (meaning that “right scaling” and “left scaling” are the same operation). Thus, we take the liberty to put the “right scaling” operation on the backburner. (Its properties are analogous to the corresponding properties of “left scaling” anyway.)

Definition 5.7.8. Let $n, m, p \in \mathbb{N}$. Let $A = (a_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$ be an $n \times m$-matrix.

Let $B = (b_{i,j})_{1 \leq i \leq m, 1 \leq j \leq p}$ be an $m \times p$-matrix. Then, we define the product $AB$ of the two matrices $A$ and $B$ by

$$AB = \left( \sum_{k=1}^{m} a_{i,k} b_{k,j} \right)_{1 \leq i \leq n, 1 \leq j \leq p}.$$

This is an $n \times p$-matrix.

So you can add together two $n \times m$-matrices, but only multiply an $n \times m$-matrix with an $m \times p$-matrix. (You cannot multiply two $n \times m$-matrices, unless $n = m$.)
Definition 5.7.9. (a) If \( n, m \in \mathbb{N} \), then the \( n \times m \) zero matrix is defined to be the \( n \times m \)-matrix
\[
(0)_{1 \leq i \leq n, 1 \leq j \leq m} = \begin{pmatrix}
0 & 0 & \cdots & 0 \\
0 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 0
\end{pmatrix}.
\]
It is called \( 0_{n \times m} \).

(b) If \( n \in \mathbb{N} \), then the \( n \times n \) identity matrix is defined to be the \( n \times n \)-matrix
\[
(\delta_{ij})_{1 \leq i \leq n, 1 \leq j \leq n} = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{pmatrix},
\]
where
\[
\delta_{ij} = \begin{cases}
1, & \text{if } i = j; \\
0, & \text{if } i \neq j.
\end{cases}
\]
(Note that using the Iverson bracket notation we introduced in Exercise 2.17.2, we have \( \delta_{ij} = [i = j] \).)

The \( n \times n \) identity matrix is called \( I_n \).

Note that the 0 and the 1 here are the zero and the unity of \( \mathbb{K} \).

Thus, a zero matrix can be of any size, but an identity matrix has to be a square matrix.

The following rules hold for addition, subtraction, multiplication and scaling of matrices:

Theorem 5.7.10. Let \( n, m, p, q \in \mathbb{N} \).

(a) We have \( A + B = B + A \) for any \( A, B \in \mathbb{K}^{n \times m} \).

(b) We have \( A + (B + C) = (A + B) + C \) for any \( A, B, C \in \mathbb{K}^{n \times m} \).

(c) We have \( A + 0_{n \times m} = 0_{n \times m} + A = A \) for any \( A \in \mathbb{K}^{n \times m} \).

(d) We have \( A \cdot I_m = I_n \cdot A = A \) for any \( A \in \mathbb{K}^{n \times m} \).

(e) In general, we do not have \( AB = BA \). In fact, it can happen that one of \( AB \) and \( BA \) is defined and the other is not; but even if both are defined, they can be distinct (even if \( \mathbb{K} \) is commutative).

(f) We have \( A(BC) = (AB)C \) for any \( A \in \mathbb{K}^{n \times m}, B \in \mathbb{K}^{m \times p} \) and \( C \in \mathbb{K}^{p \times q} \).

(g) We have \( A(B + C) = AB + AC \) for any \( A \in \mathbb{K}^{n \times m} \) and \( B, C \in \mathbb{K}^{m \times p} \).

We have \( (A + B)C = AC + BC \) for any \( A, B \in \mathbb{K}^{n \times m} \) and \( C \in \mathbb{K}^{m \times p} \).

(h) We have \( A \cdot 0_{m \times p} = 0_{n \times p} \) and \( 0_{p \times n} \cdot A = 0_{p \times m} \) for any \( A \in \mathbb{K}^{n \times m} \).

(i) If \( A, B, C \in \mathbb{K}^{n \times m} \), then we have the equivalence \( (A - B = C) \iff (A = B + C) \).

(j) We have \( r(A + B) = rA + rB \) for any \( r \in \mathbb{K} \) and \( A, B \in \mathbb{K}^{n \times m} \).

(k) We have \( (r + s)A = rA + sA \) for any \( r, s \in \mathbb{K} \) and \( A \in \mathbb{K}^{n \times m} \).
We have $r(sA) = (rs)A$ for any $r, s \in K$ and $A \in K^{n \times m}$.

We have $r(AB) = (rA)B = A(rB)$ for any $r \in K$ and $A, B \in K^{n \times m}$ if $K$ is commutative. The first equality also holds in general.

We have $r(-A) = (-r)A = r(-A)$ for any $r \in K$ and $A \in K^{n \times m}$.

We have $1A = A$ for any $A \in K^{n \times m}$.

We have $-(A + B) = (-A) + (-B)$ for any $A, B \in K^{n \times m}$.

We have $-0_{n \times m} = 0_{n \times m}$.

We have $-(−A) = A$ for any $A \in K^{n \times m}$.

We have $−(AB) = (−A)B = A(−B)$ for any $A \in K^{n \times m}$ and $B \in K^{m \times p}$.

Proof. Most of these are trivial. The hardest one is (f). See [Grinbe18, §2.9] for its proof.

Corollary 5.7.11. Let $n \in \mathbb{N}$. The set $K^{n \times n}$ of all $n \times n$-matrices (endowed with addition $+$, multiplication $\cdot$, zero $0_{n \times n}$ and unity $I_n$) is a ring.

Proof. Follows from the rules above.

2019-04-05 lecture

So we know that $K^{n \times n}$ is a ring whenever $n \in \mathbb{N}$. Hence, Proposition 5.3.6 shows that can define finite sums and finite products in $K^{n \times n}$ (but finite products need to have the order of their factors specified: i.e., we can make sense of “$A_1 A_2 \cdots A_k$” but not of “$\prod s \in S A_s$”). These also make sense for non-square matrices whenever “their sizes match”: e.g., you can define a sum of finitely many $n \times m$-matrices, and a product $A_1 A_2 \cdots A_k$ where each $A_i$ is an $n_i \times n_{i+1}$-matrix (for any $n_1, n_2, \ldots, n_k+1 \in \mathbb{N}$). Standard rules for sums and products hold, at least to the extent they don’t rely on commutativity of multiplication.

But $K^{n \times n}$ is not the only ring we can make out of matrices:

Definition 5.7.12. Let $n \in \mathbb{N}$. Let $A = (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq n}$ be an $n \times n$-matrix.

(a) We say that $A$ is lower-triangular if and only if $a_{ij} = 0$ whenever $i < j$.

(b) We say that $A$ is upper-triangular if and only if $a_{ij} = 0$ whenever $i > j$.

(c) We say that $A$ is diagonal if and only if $a_{ij} = 0$ whenever $i \neq j$. 
For example, the $2 \times 2$-matrix \[
\begin{pmatrix}
1 & 2 \\
0 & 3
\end{pmatrix}
\] is upper-triangular (but not lower-triangular), while the $2 \times 2$-matrix \[
\begin{pmatrix}
1 & 0 \\
2 & 3
\end{pmatrix}
\] is lower-triangular (but not upper-triangular).

**Proposition 5.7.13.** Let $n \in \mathbb{N}$.

(a) The set of all lower-triangular $n \times n$-matrices is a ring.

(b) The set of all upper-triangular $n \times n$-matrices is a ring.

(c) The set of all diagonal $n \times n$-matrices is a ring.

**Example 5.7.14.** For $n = 2$, the multiplication of lower-triangular $n \times n$-matrices looks as follows:
\[
\begin{pmatrix}
a & b \\
0 & c
\end{pmatrix}
\begin{pmatrix}
x & y \\
0 & z
\end{pmatrix}
= \begin{pmatrix}
ax & ay + bz \\
0 & cz
\end{pmatrix},
\]
and the multiplication of diagonal $n \times n$-matrices looks as follows:
\[
\begin{pmatrix}
a & 0 \\
0 & c
\end{pmatrix}
\begin{pmatrix}
x & 0 \\
0 & z
\end{pmatrix}
= \begin{pmatrix}
ax & 0 \\
0 & cz
\end{pmatrix}.
\]

**Proof of Proposition 5.7.13.** The main “difficulty” is showing that the product of two upper-triangular matrices is upper-triangular (and similarly for lower-triangular matrices). This is [Grinbe18, Theorem 3.23 (a)].

Note that diagonal $n \times n$-matrices are “essentially” the same as $n$-tuples of elements of $\mathbb{K}$; the ring they form is $\mathbb{K} \times \mathbb{K} \times \cdots \times \mathbb{K}$ in disguise. We will see soon how to make this precise.

One of the most important operations on matrices is taking the transpose:

**Definition 5.7.15.** Let $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Let $A = (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq m}$ be an $n \times m$-matrix. Then, we define an $m \times n$-matrix $A^T$ by
\[
A^T = (a_{ji})_{1 \leq i \leq m, 1 \leq j \leq n}.
\]
Thus, for each $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$, the $(i, j)$-th entry of $A^T$ is the $(j, i)$-th entry of $A$. This matrix $A^T$ is called the **transpose** of $A$.

For example,
\[
\begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6
\end{pmatrix}^T = \begin{pmatrix}
1 & 4 \\
2 & 5 \\
3 & 6
\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}
1 & 2 \\
1 & 0
\end{pmatrix}^T = \begin{pmatrix}
1 & 1 \\
2 & 0
\end{pmatrix}.
\]

Let us use this occasion to define column vectors and row vectors:
**Definition 5.7.16.** Let \( n \in \mathbb{N} \).

(a) A **column vector of size** \( n \) will mean an \( n \times 1 \)-matrix.

(b) A **row vector of size** \( n \) will mean a \( 1 \times n \)-matrix.

For example, \( \begin{pmatrix} 1 \\ 2 \end{pmatrix} \) is a column vector of size 2, while \( \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \) is a row vector of size 3. We will often identify a row vector \( \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \in K^{1 \times n} \) with the corresponding \( n \)-tuple \((a_1, a_2, \ldots, a_n)\).

If \( v \) is a column vector of size \( n \), then \( v^T \) is a row vector of size \( n \).

### 5.8. Ring homomorphisms

**Definition 5.8.1.** Let \( K \) and \( L \) be two rings. A **ring homomorphism** from \( K \) to \( L \) means a map \( f : K \to L \) that satisfies the following four axioms:

(a) We have \( f(a + b) = f(a) + f(b) \) for all \( a, b \in K \). (This is called "\( f \) respects addition" or "\( f \) preserves addition").

(b) We have \( f(0) = 0 \). (This, of course, means \( f(0_K) = 0_L \).)

(c) We have \( f(ab) = f(a)f(b) \) for all \( a, b \in K \). (This is called "\( f \) respects multiplication" or "\( f \) preserves multiplication".)

(d) We have \( f(1) = 1 \). (This, of course, means \( f(1_K) = 1_L \).)

**Remark 5.8.2.** The axiom (b) in Definition 5.8.1 is redundant – it follows from axiom (a).

**Proof.** Assume that axiom (a) holds. Apply axiom (a) to \( a = 0 \) and \( b = 0 \). Thus, you get

\[
f(0 + 0) = f(0) + f(0).
\]

Since \( 0 + 0 = 0 \), this rewrites as

\[
f(0) = f(0) + f(0).
\]

Subtracting \( f(0) \) on both sides (we can do this, since \( L \) is a ring), we obtain \( 0 = f(0) \), thus \( f(0) = 0 \). Thus, axiom (b) holds.

If the axiom (b) in Definition 5.8.1 is redundant, then why did we require it? One reason to do so is purely aesthetic: It ensures that each of the two “multiplicative” axioms (viz., axioms (c) and (d)) is matched by a corresponding “additive” axiom (viz., axioms (a) and (b)). We cannot omit axiom (d)\(^{141}\); thus, to avoid breaking the symmetry, I prefer not to omit axiom (b) either. There is another reason to keep axiom (b). Indeed, if we want to define **semiring homomorphisms** (i.e., the analogue of ring homomorphisms in which rings are replaced by semirings), then axiom (b) is no longer redundant (since we cannot subtract elements in a semiring); thus, if we omitted axiom (b), our definition of ring homomorphisms would become less robust with respect to replacing “ring” by “semiring”.

\(^{141}\)More precisely: if we did, then we would obtain a weaker, less useful notion of ring homomorphism.
Example 5.8.3. Let $K$ be any ring. The map $\text{id} : K \to K$ is a ring homomorphism.

Proof. Let’s check property (c): It means $\text{id}(ab) = \text{id}(a)\text{id}(b)$. This means $ab = ab$. Obvious. Similarly, the other properties hold. \qed

Example 5.8.4. Let $K$ be any ring, and let $M$ be the zero ring $\{0\}$. Then, the map $K \to M, \quad a \mapsto 0$
is a ring homomorphism.

Example 5.8.5. Let $n$ be an integer. Consider the projection

$$\pi_n : \mathbb{Z} \to \mathbb{Z}/n, \quad s \mapsto [s]_n.$$ 

This is a ring homomorphism.

Proof. Again, let us check axiom (c) only. So let $a, b \in \mathbb{Z}$. We must prove that $\pi_n(ab) = \pi_n(a)\cdot \pi_n(b)$.

The left hand side is $[ab]_n$, while the right hand side is $[a]_n \cdot [b]_n$. So they are equal, because this is how $[a]_n \cdot [b]_n$ was defined. Thus, axiom (c) holds. \qed

Example 5.8.6. Let $n$ be a positive integer. Consider the map

$$R_n : \mathbb{Z}/n \to \mathbb{Z}, \quad [s]_n \mapsto s \mod n.$$ 

(This is the map sending $[0]_n, [1]_n, \ldots, [n-1]_n$ to the numbers $0, 1, \ldots, n-1$.)

This map $R_n$ is not a ring homomorphism.

Proof. Assume the contrary. Thus, $R_n$ is a ring homomorphism. We want a contradiction.

We are in one of the following two cases:

Case 1: We have $n > 1$.

Case 2: We have $n = 1$.

Let us first consider Case 1. In this case, we have $n > 1$.

We have assumed that $R_n$ is a ring homomorphism. Thus, axiom (a) can be applied to $a = [1]_n$ and $b = [n-1]_n$, and thus we get $R_n([1]_n + [n-1]_n) = R_n([1]_n) + R_n([n-1]_n)$. But comparing

$$R_n([1]_n + [n-1]_n) = R_n([n]_n) = R_n([0]_n) = 0$$

with

$$R_n([1]_n) + R_n([n-1]_n) = 1 + (n-1) = n,$$

we get a contradiction. Thus, $R_n$ is not a ring homomorphism.
we see that this is not true. So we have found a contradiction in Case 1.

Let us now consider Case 2. In this case, we have \( n = 1 \). Thus, \([1]_1 = [0]_1\). But the map \( R_n \) maps \([0]_1\) to 0 (by its definition). However, axiom (d) forces \( R_n ([1]_1) = 1 \), which contradicts \( R_n ([1]_1) = R_n ([0]_1) = 0 \). So we have found a contradiction in Case 2.

Thus, we always get a contradiction.

**Warning:** The same people who don’t require rings to have a unity, of course, do not require ring homomorphisms to satisfy axiom (d). So for them, \( R_n \) would be a ring homomorphism for \( n = 1 \).

**Example 5.8.7.** Let \( n \) and \( d \) be integers such that \( d \mid n \). Then, the map

\[
\pi_{n,d} : \mathbb{Z}/n \to \mathbb{Z}/d, \\
[s]_n \mapsto [s]_d
\]

is a ring homomorphism.

**Proof.** Let us check axiom (c). So we must prove that \( \pi_{n,d} (\alpha \beta) = \pi_{n,d} (\alpha) \cdot \pi_{n,d} (\beta) \) for all \( \alpha, \beta \in \mathbb{Z}/n \).

Fix \( \alpha, \beta \in \mathbb{Z}/n \). Write \( \alpha \) as \( \alpha = [a]_n \) with \( a \in \mathbb{Z} \). Write \( \beta \) as \( \beta = [b]_n \) with \( b \in \mathbb{Z} \).

Thus, \( \pi_{n,d} (\alpha) = \pi_{n,d} ([a]_n) = [a]_d \) and \( \pi_{n,d} (\beta) = \pi_{n,d} ([b]_n) = [b]_d \). Multiplying these two equalities, we obtain

\[
\pi_{n,d} (\alpha) \cdot \pi_{n,d} (\beta) = [a]_d \cdot [b]_d = [ab]_d. \tag{157}
\]

But \( \alpha \beta = [a]_n \cdot [b]_n = [ab]_n \) and thus \( \pi_{n,d} (\alpha \beta) = \pi_{n,d} ([ab]_n) = [ab]_d \).

Comparing this equality to (157), we conclude \( \pi_{n,d} (\alpha \beta) = \pi_{n,d} (\alpha) \cdot \pi_{n,d} (\beta) \). Thus, axiom (c) is proven. The rest is LTTR.

**Remark 5.8.8.** Let \( n \) and \( d \) be integers. Then:

(a) If \( d \mid n \), then the only ring homomorphism from \( \mathbb{Z}/n \) to \( \mathbb{Z}/d \) is \( \pi_{n,d} \).

(b) If \( d \nmid n \), then there is no ring homomorphism from \( \mathbb{Z}/n \) to \( \mathbb{Z}/d \).

We won’t prove this.

**Example 5.8.9.** Consider the map \( \mu : \mathbb{C} \to \mathbb{R}^{2 \times 2} \) defined in Proposition 4.1.30. It is given by

\[
\mu (a,b) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}
\]

This map \( \mu \) is a ring homomorphism.
Proof. Proposition 4.1.30 yields that the map $\mu$ satisfies axioms (a) and (c). It is easy to see that it satisfies the other two.\hfill\Box

**Example 5.8.10.** Let $\iota$ be the map

$$
\mathbb{R} \to \mathbb{C},
$$

$$r \mapsto r_\mathbb{C} = (r, 0).
$$

This is a ring homomorphism.

Proof. Theorem 4.1.5 shows that $\iota$ satisfies axioms (a) and (c). As for (b) and (d), these follow from the fact that the zero of $\mathbb{C}$ is $0_\mathbb{C} = (0, 0)$ and the unity of $\mathbb{C}$ is $1_\mathbb{C} = (1, 0)$.\hfill\Box

**Example 5.8.11.** Let $\mathbb{K}$ be a commutative ring.

Let $\mathbb{K}^{\le 2}$ be the ring of upper-triangular $2 \times 2$-matrices. (This is a ring, by Proposition 5.7.13)

Let $\mathbb{K}^{\ge 2}$ be the ring of lower-triangular $2 \times 2$-matrices. (This is a ring, by Proposition 5.7.13)

(a) Consider the map

$$
\mathbb{K}^{\le 2} \to \mathbb{K}^{\ge 2},
$$

$$
\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mapsto \begin{pmatrix} a & 0 \\ b & c \end{pmatrix}.
$$

In other words, this is the map sending each $A$ to $A^T$ (the transpose of $A$). Is this a ring homomorphism? No, because $(AB)^T$ is $B^T A^T$, not $A^T B^T$ (in general). This is called a *ring antihomomorphism*. Note that if $\mathbb{K}$ was an arbitrary (not commutative) ring, then $(AB)^T$ would (in general!) equal neither $B^T A^T$ nor $A^T B^T$.

(b) Consider the map

$$
\mathbb{K}^{\le 2} \to \mathbb{K}^{\ge 2},
$$

$$
\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mapsto \begin{pmatrix} c & 0 \\ b & a \end{pmatrix}.
$$

In other words, this is the map that reverses the order of the rows and reverses the order of the columns. You can check that this is a ring homomorphism. This holds even if $\mathbb{K}$ is an arbitrary (not commutative) ring.

**Proposition 5.8.12.** Let $\mathbb{K}$ and $\mathbb{L}$ be two rings. Let $f : \mathbb{K} \to \mathbb{L}$ be a ring homomorphism.

(a) We have $f(-a) = -f(a)$ for all $a \in \mathbb{K}$. (In other words, $f$ “preserves additive inverses”.)
(b) If \( a \in \mathbb{K} \) is invertible, then \( f(a) \in \mathbb{L} \) is also invertible, and we have \( f(a^{-1}) = (f(a))^{-1} \). (In other words, \( f \) “preserves multiplicative inverses”.)

(c) We have \( f(a - b) = f(a) - f(b) \) for all \( a, b \in \mathbb{K} \).

(d) If the rings \( \mathbb{K} \) and \( \mathbb{L} \) are commutative, then we have \( f\left(\frac{a}{b}\right) = \frac{f(a)}{f(b)} \) for all \( a, b \in \mathbb{K} \) for which \( b \) is invertible.

(e) We have \( f\left(\sum_{s \in S} a_s\right) = \sum_{s \in S} f(a_s) \) whenever \( S \) is a finite set and \( a_s \in \mathbb{K} \) for all \( s \in S \).

(f) We have \( f(a_1 a_2 \cdots a_k) = f(a_1) f(a_2) \cdots f(a_k) \) whenever \( a_1, a_2, \ldots, a_k \in \mathbb{K} \).

(g) If the rings \( \mathbb{K} \) and \( \mathbb{L} \) are commutative, then \( f\left(\prod_{s \in S} a_s\right) = \prod_{s \in S} f(a_s) \) whenever \( S \) is a finite set and \( a_s \in \mathbb{K} \) for all \( s \in S \).

(h) We have \( f(a^n) = (f(a))^n \) for each \( a \in \mathbb{K} \) and each \( n \in \mathbb{N} \).

Proof. (b) Let \( a \in \mathbb{K} \) be invertible. We have
\[
f\left(a^{-1}a\right) = f\left(a^{-1}\right) f(a) \quad \text{(by axiom (c))}.
\]
Thus,
\[
f\left(a^{-1}\right) f(a) = f\left(a^{-1}a\right) = f(1) = 1 \quad \text{by axiom (d)).}
\]
Similarly, \( f(a) f(a^{-1}) = 1 \). These two equations are saying that \( f(a^{-1}) \) is a multiplicative inverse of \( f(a) \). Thus, \( f(a) \) is invertible, and \( f(a^{-1}) = (f(a))^{-1} \). This proves part (b).

(a) Repeat the proof we gave for part (b), but replace multiplication and 1 by addition and 0 (and forget about invertibility, because every element of \( \mathbb{K} \) or \( \mathbb{L} \) has an additive inverse).

(c) Let \( a, b \in \mathbb{K} \). Then,
\[
f(a - b) = f(a + (-b)) = f(a) + f\left(-b\right) = f(a) - f(b) \quad \text{(by axiom (a))}
\]
\[
\quad = f(a) + (-f(b)) = f(a) - f(b) \quad \text{(by part (a) of the proposition)}
\]
\[
= f(a) - f(b) = f(a) - f(b).
\]

(d) Similar to part (c), but addition is replaced by multiplication.

(e) Induction on \( |S| \). The induction base uses axiom (b); the induction step uses axiom (a).

(f) Induction on \( k \). The induction base uses axiom (d); the induction step uses axiom (c).

(g) Induction on \( |S| \). The induction base uses axiom (d); the induction step uses axiom (c).

(h) Follows from (f), applied to \( k = n \) and \( a_i = a \). \( \square \)
5.9. Ring isomorphisms

**Definition 5.9.1.** Let $\mathbb{K}$ and $\mathbb{L}$ be two rings. Let $f : \mathbb{K} \to \mathbb{L}$ be a map. Then, $f$ is called a *ring isomorphism* if and only if $f$ is invertible (i.e., bijective) and both $f$ and $f^{-1}$ are ring homomorphisms.

**Example 5.9.2.** Let $\mathbb{K}$ be a ring. The identity map $\text{id} : \mathbb{K} \to \mathbb{K}$ is a ring isomorphism.

**Example 5.9.3.** Let $\mathbb{K}$ be a ring. Let $n \in \mathbb{N}$. Consider the map

$$d_n : \underbrace{\mathbb{K} \times \mathbb{K} \times \cdots \times \mathbb{K}}_{n \text{ times}} \to \{\text{diagonal } n \times n\text{-matrices over } \mathbb{K}\},$$

$$(d_1, d_2, \ldots, d_n) \mapsto \begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{pmatrix}.$$ 

Note that both $\underbrace{\mathbb{K} \times \mathbb{K} \times \cdots \times \mathbb{K}}_{n \text{ times}}$ and $\{\text{diagonal } n \times n\text{-matrices over } \mathbb{K}\}$ are rings (the former by Definition 5.6.3, the latter by Proposition 5.7.13(c)).

The map $d_n$ is invertible. I claim that furthermore, $d_n$ is a ring isomorphism. This is easiest to check using Proposition 5.9.5 further below. Note that this claim is a rigorous version of our earlier informal statement that the ring formed by the diagonal $n \times n$-matrices is just $\underbrace{\mathbb{K} \times \mathbb{K} \times \cdots \times \mathbb{K}}_{n \text{ times}}$ in disguise. The isomorphism $d_n$ is responsible for the disguise!

**Example 5.9.4.** The map from $\mathbb{K}^{2 \leq 2}$ to $\mathbb{K}^{2 \geq 2}$ introduced in Example 5.8.11(b) is a ring isomorphism. Its inverse is the map

$$\mathbb{K}^{2 \geq 2} \to \mathbb{K}^{2 \leq 2},$$

$$\begin{pmatrix} c \\ b \\ a \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}.$$ 

**Proposition 5.9.5.** Let $\mathbb{K}$ and $\mathbb{L}$ be two rings. Let $f : \mathbb{K} \to \mathbb{L}$ be an invertible ring homomorphism. Then, $f$ is a ring isomorphism.

**Proof.** We just need to show that $f^{-1}$ is a ring homomorphism.

Let us verify axiom (c) for $f^{-1}$. This means that we must prove that

$$f^{-1}(ab) = f^{-1}(a)f^{-1}(b) \quad \text{for all } a, b \in \mathbb{L}.$$
So let \(a, b \in L\). We know that \(f\) is a ring homomorphism; thus it satisfies axiom (c). Applying this axiom to \(f^{-1}(a)\) and \(f^{-1}(b)\) instead of \(a\) and \(b\), we find

\[
f\left(f^{-1}(a) f^{-1}(b)\right) = f\left(f^{-1}(a)\right) \cdot f\left(f^{-1}(b)\right) = ab = f\left(f^{-1}(ab)\right).\]

Since \(f\) is injective (because \(f\) is invertible), we thus conclude

\[
f^{-1}(a) f^{-1}(b) = f^{-1}(ab),\]

which is precisely what we wanted to prove.

So axiom (c) for \(f^{-1}\) is verified. Axiom (a) follows by the same argument with \(+\) instead of \(\cdot\).

Since \(f\) satisfies axiom (d) (being a ring homomorphism), we have \(f(1) = 1\). But this yields \(f^{-1}(1) = 1\); thus, \(f^{-1}\) satisfies axiom (d). Similarly, \(f^{-1}\) satisfies axiom (b).

Thus, the map \(f^{-1} : L \to K\) satisfies all four axioms for a ring homomorphism. Hence, \(f^{-1}\) is a ring homomorphism. Thus, \(f\) is a ring isomorphism (by the definition of a ring isomorphism).

2019-04-08 lecture

**Example 5.9.6.** Let \(m\) and \(n\) be two coprime positive integers. Then, \((\mathbb{Z}/m) \times (\mathbb{Z}/n)\) is a ring (according to Definition 5.6.3). Theorem 3.6.2 says that the map

\[
S_{m,n} : \mathbb{Z}/(mn) \to (\mathbb{Z}/m) \times (\mathbb{Z}/n),
\alpha \mapsto (\pi_{mn,m}(\alpha), \pi_{mn,n}(\alpha))
\]

is well-defined and is a bijection. We claim that this map \(S_{m,n}\) is a ring isomorphism.

**Proof.** The map \(S_{m,n}\) is a bijection, thus invertible.

Let us first prove that the map \(S_{m,n}\) is a ring homomorphism.

For each \(s \in \mathbb{Z}\), we have

\[
S_{m,n} ([s]_{mn}) = \left( \frac{\pi_{mn,m}([s]_{mn})}{= [s]_m}, \frac{\pi_{mn,n}([s]_{mn})}{= [s]_n} \right) = ([s]_m, [s]_n). \tag{158}
\]

Let us check axiom (c) from Definition 5.8.1 for \(f = S_{m,n}\). Let \(\alpha, \beta \in \mathbb{Z}/(mn)\). We must prove that \(S_{m,n}(\alpha \beta) = S_{m,n}(\alpha) \cdot S_{m,n}(\beta)\).

Write \(\alpha\) and \(\beta\) in the form \(\alpha = [a]_{mn}\) and \(\beta = [b]_{mn}\) for some \(a, b \in \mathbb{Z}\). Then, \(\alpha \beta = [a]_{mn} [b]_{mn} = [ab]_{mn}\), so

\[
S_{m,n}(\alpha \beta) = S_{m,n}([ab]_{mn}) = ([ab]_m, [ab]_n) \quad \text{(by (158))}.
\]
Comparing this with

\[ S_{m,n} \left( \left[ \begin{array}{c} \alpha \\ = [a]_{mn} \end{array} \right] \right) \cdot S_{m,n} \left( \left[ \begin{array}{c} \beta \\ = [b]_{mn} \end{array} \right] \right) \]

\[ = S_{m,n} \left( \left[ \begin{array}{c} a \\ = [a]_{mn} \end{array} \right] \right) \cdot S_{m,n} \left( \left[ \begin{array}{c} b \\ = [b]_{mn} \end{array} \right] \right) \]

\[ = \left( [a]_m, [a]_n \right) \cdot \left( [b]_m, [b]_n \right) = \left( \left[ \begin{array}{c} a_m [b]_m \\ = [ab]_m \end{array} \right], \left[ \begin{array}{c} a_n [b]_n \\ = [ab]_n \end{array} \right] \right) \]

(since the multiplication \( \cdot \) on the Cartesian product \((\mathbb{Z}/m) \times (\mathbb{Z}/n)\) is defined entrywise)

\[ = \left( [ab]_m, [ab]_n \right), \]

we obtain \( S_{m,n} (\alpha \beta) = S_{m,n} (\alpha) \cdot S_{m,n} (\beta) \). This proves axiom (c) for our map \( S_{m,n} \).

Similarly, the other axioms can be shown. Thus, \( S_{m,n} \) is a ring homomorphism. Therefore, Proposition 5.9.5 shows that \( S_{m,n} \) is a ring isomorphism (since \( S_{m,n} \) is invertible).

\( \square \)

Note one more simple general fact:

**Proposition 5.9.7.** Let \( K \) and \( L \) be two rings. Let \( f : K \rightarrow L \) be a ring isomorphism. Then, \( f^{-1} : L \rightarrow K \) is also a ring isomorphism.

**Proof.** Clearly, \( f^{-1} \) is a ring homomorphism (since \( f \) is a ring isomorphism). Furthermore, \( f^{-1} \) is invertible (with inverse \((f^{-1})^{-1} = f\)) and its inverse \((f^{-1})^{-1} = f\) is a ring homomorphism as well. Thus, \( f^{-1} \) is a ring isomorphism. \( \square \)

Let me attempt to discuss the use of ring isomorphisms; unfortunately, I will have to be vague at this point. Ring homomorphisms allow us to transfer some things from one ring into another. For example, if \( f : K \rightarrow L \) is a ring homomorphism from a ring \( K \) to a ring \( L \), then \( f \) sends any invertible element of \( K \) to an invertible element of \( L \) (by Proposition 5.8.12 (b)). However, they are generally only “one-way roads”. For instance, if \( f : K \rightarrow L \) is a ring homomorphism from a ring \( K \) to a ring \( L \), and if \( a \in K \) is such that \( f(a) \in L \) is invertible, then \( a \) may and may not be invertible. A ring homomorphism does not determine either ring in terms of the other. You can have homomorphisms between completely different rings, such as from \( Z \) to the zero ring, or from \( Z \) to \( C \).

On the other hand, ring isomorphisms let us go “back and forth” between the rings they connect; if we have a ring isomorphism \( f : K \rightarrow L \), we can regard \( L \) as being “the same ring as \( K \), with its elements renamed”. (The isomorphism \( f \) does the renaming: you should think of each \( a \in K \) being renamed as \( f(a) \).)
Thus, when you have a ring isomorphism $f : K \to L$, you can take any “intrinsic” property of $K$ and obtain the corresponding property of $L$, and vice versa. Here is an example:

**Proposition 5.9.8.** Let $K$ and $L$ be two rings. Let $f : K \to L$ be a ring isomorphism.

(a) If $K$ is commutative, then $L$ is commutative.
(b) If $0 \neq 1$ in $K$, then $0 \neq 1$ in $L$.
(c) If $K$ is a skew field, then $L$ is a skew field.
(d) If $K$ is a field, then $L$ is a field.

**Proof of Proposition 5.9.8.** The proofs given below are examplary; you should be able to similarly transfer any other property of $K$ to $L$ and vice versa.

We first recall that $f$ is a ring isomorphism. Thus, the map $f$ is invertible, and both $f$ and $f^{-1}$ are ring homomorphisms (by the definition of a ring isomorphism).

(a) Assume that $K$ is commutative. We must prove that $L$ is commutative. In other words, we must prove that $ab = ba$ for any $a, b \in L$.

Fix $a, b \in L$. We know that $f$ is invertible. Hence, $f^{-1}(a)$ and $f^{-1}(b)$ are well-defined. Since $K$ is commutative, these satisfy

$$f^{-1}(a) f^{-1}(b) = f^{-1}(b) f^{-1}(a).$$

Applying $f$ to this equality, we get

$$f \left( f^{-1}(a) f^{-1}(b) \right) = f \left( f^{-1}(b) f^{-1}(a) \right).$$

But since $f$ is a ring homomorphism, we can apply axiom (c) of Definition 5.8.1 to $f^{-1}(a)$ and $f^{-1}(b)$ instead of $a$ and $b$. We thus obtain

$$f \left( f^{-1}(a) f^{-1}(b) \right) = f \left( f^{-1}(a) \right) f \left( f^{-1}(b) \right) = ab$$

and similarly $f \left( f^{-1}(b) f^{-1}(a) \right) = ba$; thus, the equality (159) becomes $ab = ba$. This shows that $L$ is commutative. This proves Proposition 5.9.8 (a).

Before I move on to the next part of Proposition 5.9.8, let me explain how the above proof could be found straightforwardly, without any creative input. The point of this is to show how to prove not just Proposition 5.9.8 (a), but any similar claim as well.

---

What do we mean by “intrinsic”? Roughly speaking, an intrinsic property of a ring is a property that can be stated entirely in terms of its structure (i.e., its ground set and its operations $+$ and $\cdot$ and its elements 0 and 1), without referring to outside objects. For instance, “every element $a$ of the ring satisfies $a^3 = a^2$” is an intrinsic property (since $a^3 = a a a$ and $a^2 = a a$ are defined purely in terms of the operation $\cdot$), and “the ring has two nonzero elements $a$ and $b$ such that $ab = 0$” is an intrinsic property as well (provided that “nonzero” and “0” refer to the zero of the ring, rather than the number 0), but “the ring contains the number $\sqrt{2}$” is not an intrinsic property (since it refers to an outside object – namely, the number $\sqrt{2}$).
The (only) idea involved in the above proof was the following: The two mutually inverse bijections \( f : \mathbb{K} \to \mathbb{L} \) and \( f^{-1} : \mathbb{L} \to \mathbb{K} \) provide a “railway system” that can be used to transport anything (elements, equalities, subsets, etc.) between \( \mathbb{K} \) and \( \mathbb{L} \). Since these bijections are ring homomorphisms, the structure of the objects that we are transporting does not get “damaged in transit”: Products remain products (i.e., if we have three elements \( a, b \) and \( c \) of \( \mathbb{K} \) satisfying \( ab = c \), and if we transport these three elements to \( \mathbb{L} \) via \( f \), then the resulting three elements of \( \mathbb{L} \) will still satisfy \( f(a)f(b) = f(c) \)), sums remain sums, etc. Thus, we can move back and forth between \( \mathbb{K} \) and \( \mathbb{L} \) without keeping track of where precisely we take our sums and products.

With this in mind, our above proof of Proposition 5.9.8(a) can be discovered as follows:

Assume that \( \mathbb{K} \) is commutative. We must prove that \( \mathbb{L} \) is commutative. In other words, we must prove that \( ab = ba \) for any \( a, b \in \mathbb{L} \). So let us fix \( a, b \in \mathbb{L} \). We want to prove \( ab = ba \), but all we have is an analogous identity for elements of \( \mathbb{K} \) (since we know that \( \mathbb{K} \) is commutative). In other words, we have

\[
a'b' = b'a' \quad \text{for all } a', b' \in \mathbb{K}. \tag{160}
\]

So we transport our two elements \( a, b \) of \( \mathbb{L} \) to \( \mathbb{K} \) (by our “railway system”—specifically, using the map \( f^{-1} \)), in order to be able to apply (160) to them. The result are the two elements \( f^{-1}(a), f^{-1}(b) \) of \( \mathbb{K} \). Applying the identity (160) to \( a' = f^{-1}(a) \) and \( b' = f^{-1}(b) \), we obtain \( f^{-1}(a)f^{-1}(b) = f^{-1}(b)f^{-1}(a) \). This is an equality inside \( \mathbb{K} \), whereas our goal is to prove an equality inside \( \mathbb{L} \) (namely, the equality \( ab = ba \)). So we transport this equality back into \( \mathbb{L} \) by applying \( f \) to its two sides. We thus obtain \( f(f^{-1}(a)f^{-1}(b)) = f(f^{-1}(b)f^{-1}(a)) \). But recalling that \( f \) is a ring homomorphism and thus no structure gets “damaged in transit”, we see that

\[
f\left( f^{-1}(a)f^{-1}(b) \right) = f\left( f^{-1}(a) \right) \cdot f\left( f^{-1}(b) \right) = ab
\]

and similarly \( f(f^{-1}(b)f^{-1}(a)) = ba \). Hence, the equality \( f(f^{-1}(a)f^{-1}(b)) = f(f^{-1}(b)f^{-1}(a)) \) that we have just proven rewrites as \( ab = ba \), which is precisely what we wanted to prove. Thus, we have proven Proposition 5.9.8(a) by merely going back and forth between \( \mathbb{K} \) and \( \mathbb{L} \).

Let us now prove the rest of Proposition 5.9.8.

(b) Assume \( 0 \neq 1 \) in \( \mathbb{K} \). We must prove \( 0 \neq 1 \) in \( \mathbb{L} \).

The map \( f \) is an isomorphism, thus invertible, thus bijective, thus injective. Hence, from \( 0 \neq 1 \) in \( \mathbb{K} \), we conclude \( f(0) \neq f(1) \) in \( \mathbb{L} \). But since \( f \) is a ring homomorphism, we have \( f(0) = 0 \) and \( f(1) = 1 \); so this becomes \( 0 \neq 1 \) in \( \mathbb{L} \). This proves Proposition 5.9.8(b).

(c) Assume that \( \mathbb{K} \) is a skew field. We must prove that \( \mathbb{L} \) is a skew field.

Since \( \mathbb{K} \) is a skew field, we have \( 0 \neq 1 \) in \( \mathbb{K} \), thus \( 0 \neq 1 \) in \( \mathbb{L} \) (by Proposition 5.9.8(b)). Hence, it remains to prove that every nonzero element \( a \in \mathbb{L} \) has a multiplicative inverse.
Let $a \in L$ be nonzero. Then, $f^{-1}(a) \in \mathbb{K}$ is nonzero (because if it was zero, then we would have $f^{-1}(a) = 0$ and thus $f \left( f^{-1}(a) \right) = f(0) = 0$ (since $f$ is a ring homomorphism); but this would contradict the fact that $f \left( f^{-1}(a) \right) = a$ is nonzero). Hence, $f^{-1}(a) \in \mathbb{K}$ has a multiplicative inverse $b$ (since $\mathbb{K}$ is a skew field). Consider this $b$. Thus, $f^{-1}(a) b = bf^{-1}(a) = 1$ (since $b$ is a multiplicative inverse of $f^{-1}(a)$). Applying $f$ to this equation, we obtain
\[ f \left( f^{-1}(a) b \right) = f \left( bf^{-1}(a) \right) = f(1). \]
This quickly rewrites as
\[ af(b) = f(b)a = 1 \]
(since $f$ is a ring homomorphism). Thus, $f(b)$ is a multiplicative inverse of $a$. Hence, $a$ has a multiplicative inverse. Thus, we have shown that every nonzero element $a \in L$ has a multiplicative inverse. Since $0 \neq 1$ in $L$, this shows that $L$ is a skew field. This proves Proposition 5.9.8(c).

(d) Assume that $\mathbb{K}$ is a field. We must prove that $L$ is a field.

Since $\mathbb{K}$ is a field, $\mathbb{K}$ is commutative, and thus $L$ is commutative (by Proposition 5.9.8(a)).

But $\mathbb{K}$ is a field, and thus a skew field. Hence, Proposition 5.9.8(c) shows that $L$ is a skew field. Since $L$ is commutative, this yields that $L$ is a field. This proves Proposition 5.9.8(d).

The idea of the above proof (and of many similar proofs, which we will omit) is that if you have a ring isomorphism $f : \mathbb{K} \to L$, you can transport any equality or element from $\mathbb{K}$ to $L$ (via $f$) or vice versa (via $f^{-1}$); and each time, the ring operations ($+,-,\cdot,\sum, 0,1$) do not get damaged on the way (since $f$ and $f^{-1}$ are ring homomorphisms).

Here is another example:

**Proposition 5.9.9.** Let $\mathbb{K}$ and $L$ be two rings. Let $f : \mathbb{K} \to L$ be a ring isomorphism. Then:

(a) We have
\[ |\{\text{invertible elements of } \mathbb{K}\}| = |\{\text{invertible elements of } L\}|. \]

(b) We have
\[ |\{\text{idempotent elements of } \mathbb{K}\}| = |\{\text{idempotent elements of } L\}|. \]

Here, an element $a$ of a ring $\mathbb{K}$ is said to be idempotent if $a^2 = a$.

**Proof of Proposition 5.9.9** Proposition 5.9.9 is another instance of the “anything can be transported along a ring isomorphism” principle. Here is a proof in more detail:
(a) If $a$ is an invertible element of $K$, then $f(a)$ is an invertible element of $L$ (since we can pick a multiplicative inverse $b$ of $a$ in $K$, and then $f(b)$ will be a multiplicative inverse of $f(a)$ in $L$). Hence, the map

$$\{\text{invertible elements of } K\} \rightarrow \{\text{invertible elements of } L\}, \quad a \mapsto f(a)$$

is well-defined. Similarly, the map

$$\{\text{invertible elements of } L\} \rightarrow \{\text{invertible elements of } K\}, \quad a \mapsto f^{-1}(a)$$

is also well-defined (since Proposition 5.9.7 shows that the map $f^{-1} : L \rightarrow K$ is also a ring isomorphism). These two maps are clearly mutually inverse, and therefore are bijections. Hence, we have found a bijection from $\{\text{invertible elements of } K\}$ to $\{\text{invertible elements of } L\}$. Thus,

$$|\{\text{invertible elements of } K\}| = |\{\text{invertible elements of } L\}|.$$

This proves Proposition 5.9.9 (a).

(b) If $a$ is an idempotent element of $K$, then $f(a)$ is an idempotent element of $L$ (since $f$ is a ring homomorphism and thus $(f(a))^2 = f\left(\frac{a^2}{a} = a\right) = f(a)$). Hence, the map

$$\{\text{idempotent elements of } K\} \rightarrow \{\text{idempotent elements of } L\}, \quad a \mapsto f(a)$$

is well-defined. Similarly, the map

$$\{\text{idempotent elements of } L\} \rightarrow \{\text{idempotent elements of } K\}, \quad a \mapsto f^{-1}(a)$$

is also well-defined (since Proposition 5.9.7 shows that the map $f^{-1} : L \rightarrow K$ is also a ring isomorphism). These two maps are clearly mutually inverse, and therefore are bijections. Hence, we have found a bijection from $\{\text{idempotent elements of } K\}$ to $\{\text{idempotent elements of } L\}$. Thus,

$$|\{\text{idempotent elements of } K\}| = |\{\text{idempotent elements of } L\}|.$$

This proves Proposition 5.9.9 (b). \qed

Now let us see some applications of ring isomorphisms. Recall that we proved Theorem 2.14.4 using the Chinese Remainder Theorem in Section 3.6. Let us redo this proof in a shorter way:
New version of our Second proof of Theorem 2.14.4: Example 5.9.6 says that the map
\[ S_{m,n} : \mathbb{Z}/(mn) \to (\mathbb{Z}/m) \times (\mathbb{Z}/n), \]
\[ \alpha \mapsto (\pi_{mn,m}(\hat{\alpha}), \pi_{mn,n}(\hat{\alpha})) \]
is a ring isomorphism. Thus,
\[
|\{\text{invertible elements of } \mathbb{Z}/(mn)\}| = |\{\text{invertible elements of } (\mathbb{Z}/m) \times (\mathbb{Z}/n)\}| \tag{161}
\]
(by Proposition 5.9.9 (a)).

But if \( K \) and \( L \) are any two rings, then
\[
\{\text{invertible elements of } K \times L\} = \{\text{invertible elements of } K\} \times \{\text{invertible elements of } L\}
\]
(since multiplication on \( K \times L \) is defined entrywise, so an element \((a,b) \in K \times L\) is invertible if and only if both \(a \in K\) and \(b \in L\) are invertible). Hence,
\[
\{\text{invertible elements of } (\mathbb{Z}/m) \times (\mathbb{Z}/n)\} = \{\text{invertible elements of } \mathbb{Z}/m\} \times \{\text{invertible elements of } \mathbb{Z}/n\},
\]
so that
\[
|\{\text{invertible elements of } (\mathbb{Z}/m) \times (\mathbb{Z}/n)\}| = |\{\text{invertible elements of } \mathbb{Z}/m\}| \cdot |\{\text{invertible elements of } \mathbb{Z}/n\}|.
\]
Hence, (161) becomes
\[
|\{\text{invertible elements of } \mathbb{Z}/(mn)\}| = |\{\text{invertible elements of } (\mathbb{Z}/m) \times (\mathbb{Z}/n)\}| = |\{\text{invertible elements of } \mathbb{Z}/m\}| \cdot |\{\text{invertible elements of } \mathbb{Z}/n\}| \cdot |\{\text{invertible elements of } \mathbb{Z}/(mn)\}|. \tag{162}
\]

On the other hand, we know that
\[
\varphi(n) = |\{\text{invertible elements of } \mathbb{Z}/n\}|
\]
(in fact, this is Corollary 3.5.5 (b), since what we called \( U_n \) in this corollary is exactly \{invertible elements of \( \mathbb{Z}/n \)\}). Similarly,
\[
\varphi(m) = |\{\text{invertible elements of } \mathbb{Z}/m\}| \quad \text{and} \quad \varphi(mn) = |\{\text{invertible elements of } \mathbb{Z}/(mn)\}|.
\]
So the equality (162) rewrites as \( \varphi(mn) = \varphi(m) \cdot \varphi(n) \). So Theorem 2.14.4 is proven again.

The next exercise offers another example of the same strategy:
Exercise 5.9.1. Let $p$ and $q$ be two distinct primes. How many idempotent elements does the ring $\mathbb{Z}/(pq)$ have?

Solution. The primes $p$ and $q$ are distinct, so they are coprime. Hence, Example 5.9.6 (applied to $m = p$ and $n = q$) says that the map

$$S_{p,q} : \mathbb{Z}/(pq) \to (\mathbb{Z}/p) \times (\mathbb{Z}/q),$$

$$\alpha \mapsto (\pi_{pq,p}(\alpha), \pi_{pq,q}(\alpha))$$

is a ring isomorphism. Hence, Proposition 5.9.9 (b) yields

$$|\{\text{idempotent elements of } \mathbb{Z}/(pq)\}| = |\{\text{idempotent elements of } (\mathbb{Z}/p) \times (\mathbb{Z}/q)\}| = |\{\text{idempotent elements of } \mathbb{Z}/p\} \times \{\text{idempotent elements of } \mathbb{Z}/q\}|$$

(since for any two rings $K$ and $L$, we have

$$\{\text{idempotent elements of } K \times L\} = \{\text{idempotent elements of } K\} \times \{\text{idempotent elements of } L\},$$

because of the entrywise multiplication on $K \times L$). Thus, it remains to find the number of idempotent elements of $\mathbb{Z}/p$ and the number of idempotent elements of $\mathbb{Z}/q$.

How many idempotent elements does $\mathbb{Z}/p$ have? For any $a \in \mathbb{Z}$, we have the following chain of equivalences:

$$\begin{align*}
([a]_p \text{ is idempotent}) & \iff \left( ([a]_p)^2 = [a]_p \right) \quad \text{(by the definition of “idempotent“)} \\
& \iff \left( [a^2]_p = [a]_p \right) \quad \text{(since } ([a]_p)^2 = [a^2]_p) \\
& \iff (a^2 \equiv a \mod p) \\
& \iff \left( p \mid \frac{a^2 - a}{a(a-1)} \right) \iff (p \mid a(a-1)) \\
& \iff (p \mid a \text{ or } p \mid a - 1) \quad \text{(since } p \text{ is prime)} \\
& \iff (a \equiv 0 \mod p \text{ or } a \equiv 1 \mod p) \\
& \iff \left( [a]_p = [0]_p \text{ or } [a]_p = [1]_p \right).
\end{align*}$$

In other words, for a given $a \in \mathbb{Z}$, the residue class $[a]_p$ is idempotent if and only if $[a]_p$ equals $[0]_p$ or $[1]_p$. Since every residue class $\alpha \in \mathbb{Z}/p$ has the form $[a]_p$ for
some $a \in \mathbb{Z}$, we can restate this as follows: A residue class $\alpha \in \mathbb{Z}/p$ is idempotent if and only if it equals $[0]_p$ or $[1]_p$. Thus, the ring $\mathbb{Z}/p$ has exactly two idempotent elements (namely, $[0]_p$ and $[1]_p$). In other words,

$$|\{\text{idempotent elements of } \mathbb{Z}/p\}| = 2.$$  

Similarly,

$$|\{\text{idempotent elements of } \mathbb{Z}/q\}| = 2.$$  

Now, the above computation becomes

$$|\{\text{idempotent elements of } \mathbb{Z}/(pq)\}| = |\{\text{idempotent elements of } \mathbb{Z}/p\} \times \{\text{idempotent elements of } \mathbb{Z}/q\}| = 2 \cdot 2 = 4.$$  

In other words, the ring $\mathbb{Z}/(pq)$ has 4 idempotent elements.

Side-note: What are these 4 idempotent elements? Two of them are easy to find: $[0]_{pq}$ and $[1]_{pq}$ (in fact, 0 and 1 are idempotent elements in any ring). But how to get the other two?

Here is a systematic approach: Recall that $S_{p,q} : \mathbb{Z}/(pq) \to (\mathbb{Z}/p) \times (\mathbb{Z}/q)$ is a ring isomorphism. Thus, looking back at the proof of Proposition 5.9.9 (b), we see that the idempotent elements of $\mathbb{Z}/(pq)$ are the preimages of the idempotent elements of $(\mathbb{Z}/p) \times (\mathbb{Z}/q)$ under this isomorphism $S_{p,q}$.

The 4 idempotent elements of $(\mathbb{Z}/p) \times (\mathbb{Z}/q)$ are

$$([0]_p, [0]_q), \quad ([1]_p, [1]_q), \quad ([0]_p, [1]_q), \quad ([1]_p, [0]_q).$$  

To find the 4 idempotent elements in $\mathbb{Z}/(pq)$, we thus have to apply the inverse $(S_{p,q})^{-1}$ of the isomorphism $S_{p,q}$ to them.

- The first gets sent to $[0]_{pq}$.
- The second gets sent to $[1]_{pq}$.
- The last two get sent to $[px]_{pq}$ and $[qy]_{pq}$, where $x$ is a modular inverse of $p$ modulo $q$, and where $y$ is a modular inverse of $q$ modulo $p$. (It does not matter which inverses we choose; we get the same elements.)

Here is a slightly different way to get the last two idempotent elements: Bezout’s theorem yields that there exist integers $x$ and $y$ such that $xp + yq = 1$. Then, $[xp]_{pq}$ and $[yq]_{pq}$ are the two missing idempotents.

\[\square\]

2019-04-10 lecture
Example 5.9.10. Let $A$ be the $2 \times 2$-matrix $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \in \mathbb{Z}^{2 \times 2}$.

On MT2 exercise 5, you have encountered the ring

$$\mathcal{F} = \{aA + bI_2 \mid a, b \in \mathbb{Z}\} = \left\{ \begin{pmatrix} b & a \\ a & a+b \end{pmatrix} \mid a, b \in \mathbb{Z} \right\}.$$  

On HW5 exercise 5, you have encountered the ring

$$\mathbb{Z}[\phi] = \{a + b\phi \mid a, b \in \mathbb{Z}\},$$

where $\phi = \frac{1 + \sqrt{5}}{2} = 1.618\ldots$ is the golden ratio.

I claim that there is an isomorphism from $\mathbb{Z}[\phi]$ to $\mathcal{F}$. Namely, the map

$$f : \mathbb{Z}[\phi] \to \mathcal{F},$$

$$a + b\phi \mapsto bA + aI_2 = \begin{pmatrix} a & b \\ b & a+b \end{pmatrix}$$

is a ring isomorphism (but not the only one!).

(Check this by hand.)

Definition 5.9.11. Let $K$ and $L$ be two rings. We say that the rings $K$ and $L$ are isomorphic if there exists a ring isomorphism $f : K \to L$.

We write “$K \cong L$ (as rings)” to say that the rings $K$ and $L$ are isomorphic.

5.10. An overview of matrix algebra over fields

I assume you have all seen some basic matrix algebra: Gaussian elimination, ranks of matrices, inverses of matrices, determinants, etc. (If not, see [Heffer17].)

Usually, these things are done for matrices over $\mathbb{R}$ or $\mathbb{C}$. But we can try doing the same with matrices over an arbitrary commutative ring $K$.

5.10.1. Matrices over fields

Let us first study the situation when $K$ is a field.

Example: Let $K = \mathbb{Z}/3$, and let $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \in K^{3 \times 3}$. (Here, of course, 0 and 1 mean $[0]_3$ and $[1]_3$.) Let $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \in K^{3 \times 1}$. We want to find a column vector
\( x \in K^{3 \times 1} \) such that \( Ax = b \). This means, explicitly, to find \( x_1, x_2, x_3 \in K \) such that

\[
\begin{align*}
0x_1 + 1x_2 + 1x_3 &= 1; \\
1x_1 + 0x_2 + 1x_3 &= 1; \\
1x_1 + 1x_2 + 0x_3 &= 1.
\end{align*}
\]

Can we do this? Well, we can try: Augment the matrix \( A \) with the column \( b \), obtaining the augmented matrix

\[
(A \mid b) = \begin{pmatrix}
0 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1
\end{pmatrix}.
\]

Now, we shall transform this matrix into reduced row echelon form (see [Strick13, §5] or [Heffer17, Chapter One, §III] by a series of row operations (this is called Gauss-Jordan reduction in [Heffer17, Chapter One, §III], and also appears as Method 6.3 in [Strick13]):

\[
(A \mid b) = \begin{pmatrix}
0 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 \\
1 & 1 & 0 & 1
\end{pmatrix} \xrightarrow{\text{swap row 2 with row 1}} \begin{pmatrix}
1 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 \\
1 & 1 & 0 & 1
\end{pmatrix}
\]

\[
\xrightarrow{\text{subtract row 1 from row 3}} \begin{pmatrix}
1 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 \\
0 & 1 & 2 & 0
\end{pmatrix} \quad \text{(since } -1 = 2 \text{ in } \mathbb{Z}/3\text{)}
\]

\[
\xrightarrow{\text{subtract row 2 from row 3}} \begin{pmatrix}
1 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 \\
0 & 0 & 1 & 2
\end{pmatrix}
\]

\[
\text{(this is a row echelon form, but not a reduced one)}
\]

\[
\xrightarrow{\text{subtract row 3 from row 1}} \begin{pmatrix}
1 & 0 & 0 & 2 \\
0 & 1 & 1 & 1 \\
0 & 0 & 1 & 2
\end{pmatrix}
\]

\[
\xrightarrow{\text{subtract row 3 from row 2}} \begin{pmatrix}
1 & 0 & 0 & 2 \\
0 & 1 & 0 & 2 \\
0 & 0 & 1 & 2
\end{pmatrix}.
\]

\[^{143}\text{The reduced row echelon form is called “reduced echelon form” in [Heffer17].}\]
So for any vector \( x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in K^{3 \times 1} \), we have the following chain of equivalences:

\[
(Ax = b) \iff (Ax - b = 0_{3 \times 1}) \iff \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0_{3 \times 1} \end{pmatrix} \iff \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}.
\]

So our linear system has the unique solution

\[
x = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}.
\]

Next, try doing the same for \( K = \mathbb{Z}/2 \), with the “same” matrix. (It will not be literally the same matrix, of course, since 0 and 1 will now mean \([0]_2\) and \([1]_2\).)

Thus, let \( K = \mathbb{Z}/2 \), and let \( A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \in K^{3 \times 3} \). (Here, of course, 0 and 1 mean \([0]_2\) and \([1]_2\).) Let \( b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \in K^{3 \times 1} \). We want to find a column vector \( x \in K^{3 \times 1} \) such that \( Ax = b \). This means, explicitly, to find \( x_1, x_2, x_3 \in K \) such that

\[
\begin{align*}
0x_1 + 1x_2 + 1x_3 &= 1; \\
1x_1 + 0x_2 + 1x_3 &= 1; \\
1x_1 + 1x_2 + 0x_3 &= 1.
\end{align*}
\]

Can we do this? Again, we can try: Augment the matrix \( A \) with the column \( b \),
obtaining

\[(A \mid b) = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}.
\]

Now, we shall transform this matrix into reduced row echelon form by a series of row operations:

\[(A \mid b) = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix} \quad \xrightarrow{\text{swap row 2 with row 1}} \quad \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}
\]

\[\xrightarrow{\text{subtract row 1 from row 3}} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{(since } -1 = 1 \text{ in } \mathbb{Z}/2)\]

\[\xrightarrow{\text{subtract row 2 from row 3}} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\]

\[\xrightarrow{\text{subtract row 3 from row 1}} \begin{pmatrix} x_1 + x_3 = 0 \\ x_2 + x_3 = 0 \\ -1 = 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \xrightarrow{\text{false}}.
\]

So for any vector \(x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{K}^{3 \times 1}\), we have the following chain of equivalences:

\[(Ax = b) \quad \iff (Ax - b = 0_{3 \times 1}) \quad \iff (A \mid b) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0_{3 \times 1} \quad \iff \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \quad \iff (false).
\]

So our linear system has no solution.
By the way, you could have easily seen this from the system itself:

\[
\begin{aligned}
0x_1 + 1x_2 + 1x_3 &= 1; \\
1x_1 + 0x_2 + 1x_3 &= 1; \\
1x_1 + 1x_2 + 0x_3 &= 1.
\end{aligned}
\]

Adding together the three equations, we get 0 = 1 (since 1 + 1 = 0 and 1 + 1 + 1 = 1 in \(\mathbb{Z}/2\)), which is absurd. So the system has no solution.

**Upshot:** We can do linear algebra over any field more or less in the same as we did over real/complex numbers. But the result will depend on the field.

Let me recall a couple theorems from linear algebra that hold (with the same proofs) over any field:

**Theorem 5.10.1.** Let \( \mathbb{K} \) be a field.

(a) Any matrix over \( \mathbb{K} \) has a reduced row echelon form (RREF).

(b) If \( A \in \mathbb{K}^{n \times m} \) is any matrix and \( R \) is its RREF, then the row space, kernel (= nullspace) and rank of \( A \) are equal to those of \( R \). (Here, the row space, kernel and rank of a matrix is defined as for real/complex matrices.)

(c) If \( A \in \mathbb{K}^{n \times m} \) is any matrix, and if \( b \in \mathbb{K}^{n \times 1} \) is any column vector, then the equation \( Ax = b \) (for an unknown column vector \( x \in \mathbb{K}^{m \times 1} \)) can be solved using the Gaussian elimination algorithm (e.g., by forming the augmented matrix \((A \mid b)\), then transforming it into RREF, and reading off the solutions from this RREF by the same method as you learned in Linear Algebra).

(d) If \( A \in \mathbb{K}^{n \times m} \) is a matrix with \( n < m \), then there exists a nonzero \( x \in \mathbb{K}^{m \times 1} \) such that \( Ax = 0_{n \times 1}. \) (“Nonzero” means “distinct from \( 0_{m \times 1} \)” here.)

(e) Let \( A \in \mathbb{K}^{n \times n} \). Then, the following are equivalent:

- The matrix \( A \) is invertible.
- The matrix \( A \) is row-equivalent to \( I_n \). (Two matrices are said to be row-equivalent if one can be transformed into the other via row operations: swapping rows, scaling rows and adding a multiple of one row to another.)
- The matrix \( A \) is column-equivalent to \( I_n \). (The definition of “column-equivalent” is the same as of “row-equivalent”, but with columns instead of rows.)
- The RREF of \( A \) is \( I_n \).
- The RREF of \( A \) has \( n \) pivots.
- The rank of \( A \) is \( n \).
- The equation \( Ax = 0_{n \times 1} \) (for an unknown \( x \in \mathbb{K}^{n \times 1} \)) has only the trivial solution (that is, \( x = 0_{n \times 1} \)).
- For each vector \( b \in \mathbb{K}^{n \times 1} \), the equation \( Ax = b \) has a solution.
• For each vector \( b \in K^{n \times 1} \), the equation \( Ax = b \) has a unique solution.
• The columns of \( A \) are linearly independent.
• The rows of \( A \) are linearly independent.
• There is a matrix \( B \in K^{n \times n} \) such that \( AB = I_n \).
• There is a matrix \( B \in K^{n \times n} \) such that \( BA = I_n \).
• We have \( \det A \neq 0 \). (We will later define determinants and study them in some detail.)

(Matrices satisfying these equivalent conditions are called nonsingular.)

5.10.2. What if \( K \) is not a field?

Things get weird when \( K \) is not a field. For an example, set \( K = \mathbb{Z}/26 \). This is not a field, since 26 is not prime (after all, \( 26 = 2 \cdot 13 \)). The ring \( \mathbb{Z}/26 \) has been used in classical cryptography, since its elements are in bijection with the letters of the (modern) Roman alphabet:

\[
0 \mapsto A, \quad 1 \mapsto B, \quad 2 \mapsto C, \quad \ldots
\]

For example, the Hill cipher lets you encrypt a word using a \( 3 \times 3 \)-matrix over \( \mathbb{Z}/26 \) as a key. The idea is simple: You split the word into 3-letter pieces; you turn each piece into a column vector in \( (\mathbb{Z}/26)^{3 \times 1} \); and you multiply each of these column vectors by your key matrix. To decrypt, you would have to invert the key matrix.

So we want to know how to invert a matrix over \( \mathbb{Z}/26 \).

If \( \mathbb{Z}/26 \) was a field, you would know how to do this via Gaussian elimination. Most of Theorem 5.10.1 collapses when \( K \) is not a field. For example, let \( K = \mathbb{Z}/26 \) and

\[
A = \begin{pmatrix}
2 & 13 \\
13 & 20
\end{pmatrix} \in K^{2 \times 2}.
\]

(We are abusing notation here: In truth, the entries of \( A \) are not the integers 2, 13, 13, 20 but rather their residue classes \( [2]_{26}, [13]_{26}, [13]_{26}, [20]_{26} \). But we shall simply write the integers instead and hope that the reader knows what we mean.)

Is this matrix \( A \) invertible?

Let us first try to find the RREF of \( A \). If we would blindly follow the Gaussian elimination algorithm, we would fail very quickly: None of the 4 entries of \( A \) has a multiplicative inverse; thus we could not transform any entry of \( A \) into 1 by scaling a row of \( A \). But we can try to loosen Gaussian elimination by allowing more strategic row operations: Instead of trying to get a 1 in a pivot position immediately by scaling a row, we can attempt to obtain a 1 by row addition operations. For
example, we can transform our matrix $A$ above as follows:

\[
\begin{pmatrix}
2 & 13 \\
13 & 20
\end{pmatrix}
\]

subtract 6 times row 1 from row 2
\[
\begin{pmatrix}
2 & 13 \\
1 & 20
\end{pmatrix}
\]

swap row 1 with row 2
\[
\begin{pmatrix}
1 & 20 \\
2 & 13
\end{pmatrix}
\]

subtract 2 times row 1 from row 2
\[
\begin{pmatrix}
1 & 20 \\
0 & 25
\end{pmatrix}
\]

scale row 2 by $-1$
\[
\begin{pmatrix}
1 & 20 \\
0 & 1
\end{pmatrix}
\]

subtract 20 times row 2 from row 1
\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} = I_2.
\]

So our matrix $A$ does have a RREF, and even is invertible! (We can find an inverse of $A$ by computing an RREF of the block matrix $(A | I_2)$; see, e.g., [Strick13] Method 11.11 for this procedure.)

What exactly was the method behind our above row-reduction procedure? Let us see how the first column was being transformed:

\[
\begin{pmatrix}
2 \\
13
\end{pmatrix} \text{ subtract 6 times row 1 from row 2 } \begin{pmatrix}
2 \\
1
\end{pmatrix} \text{ swap row 1 with row 2 } \begin{pmatrix}
1 \\
2
\end{pmatrix} \text{ subtract 2 times row 1 from row 2 } \begin{pmatrix}
1 \\
0
\end{pmatrix}.
\]

So what we did was progressively making the entries of the first column smaller by subtracting a multiple of the first entry from the second entry (and swapping the two entries, in order to move the smaller entry into the first position). This is exactly the Euclidean algorithm! (Or, rather, it would be the Euclidean algorithm if we had used honest integers instead of residue classes in $\mathbb{Z}/26$.)

What happens in general? In general, when $K = \mathbb{Z}/n$, the Gaussian elimination algorithm as defined in linear algebra does not always work. Nevertheless, a variant of it works, in which you do not directly scale rows to turn entries into 1, but instead “minimize” the whole column using the Euclidean algorithm as we did with our matrix $A$ above. You will not always be able to get 1’s in pivot positions, because the gcd (which the Euclidean algorithm computes) may not be 1; thus, the result will not always be an RREF in the classical sense, but rather something loosely resembling it.

For details, look up the Smith normal form. Note that for $n = 0$, we have $\mathbb{Z}/n \cong \mathbb{Z}$, so this applies to matrices with integer entries.

2019-04-12 lecture

5.10.3. Review of basic notions from linear algebra
Convention 5.10.2. For the rest of this section, we fix a field \( \mathbb{K} \). The elements of \( \mathbb{K} \) will be referred to as scalars.

In the linear algebra you have seen before, the scalars are usually real numbers (i.e., we have \( \mathbb{K} = \mathbb{R} \)), but much of the theory works in the same way for every field.

Definition 5.10.3. Let \( n \in \mathbb{N} \). Recall that \( \mathbb{K}^{1 \times n} \) is the set of all row vectors of size \( n \).

A subspace of \( \mathbb{K}^{1 \times n} \) means a subset \( S \subseteq \mathbb{K}^{1 \times n} \) satisfying the following axioms:

(a) We have \( 0_{1 \times n} \in S \).
(b) If \( a, b \in S \), then \( a + b \in S \).
(c) If \( a \in S \) and \( \lambda \in \mathbb{K} \), then \( \lambda a \in S \).

In other words, a subspace of \( \mathbb{K}^{1 \times n} \) is a subset of \( \mathbb{K}^{1 \times n} \) that contains the zero vector and is closed under addition and scaling.

Subspaces are often called vector subspaces.

A similar definition defines subspaces of \( \mathbb{K}^{n \times 1} \) (column vectors).

There is a more general version of this definition, which extends it to subspaces of arbitrary vector spaces (see below); we may come to it later.

Definition 5.10.4. Let \( n \in \mathbb{N} \). Let \( v_1, v_2, \ldots, v_k \) be some row vectors in \( \mathbb{K}^{1 \times n} \).

(a) A linear combination of \( v_1, v_2, \ldots, v_k \) means a row vector of the form

\[ \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k, \quad \text{with } \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{K}. \]

(b) The span of \( v_1, v_2, \ldots, v_k \) is defined to be the subset

\[ \{ \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k \mid \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{K} \} = \{ \text{linear combinations of } v_1, v_2, \ldots, v_k \} \]

of \( \mathbb{K}^{1 \times n} \). This span is a subspace of \( \mathbb{K}^{1 \times n} \). (This is easy to check.)

(c) The vectors \( v_1, v_2, \ldots, v_k \) are said to be linearly independent if the only \( k \)-tuple

\[ (\lambda_1, \lambda_2, \ldots, \lambda_k) \in \mathbb{K}^k \]

satisfying \( \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = 0_{1 \times n} \) is \( \begin{pmatrix} 0, 0, \ldots, 0 \end{pmatrix} \) (\( k \) times).

(d) Let \( U \) be a subspace of \( \mathbb{K}^{1 \times n} \). We say that \( v_1, v_2, \ldots, v_k \) form a basis of \( U \) (or, more formally, \( (v_1, v_2, \ldots, v_k) \) is a basis of \( U \)) if and only if the vectors \( v_1, v_2, \ldots, v_k \) are linearly independent and their span is \( U \).

All the terminology we have just introduced depends on \( \mathbb{K} \). Whenever the field \( \mathbb{K} \) is not clear from the context, you can insert it into this terminology to make it unambiguous: e.g., say “\( \mathbb{K} \)-linear combination” instead of “linear combination”, and “\( \mathbb{K} \)-span” instead of “span”. 
Theorem 5.10.5. Let \( n \in \mathbb{N} \). Let \( U \) be a subspace of \( K^{1 \times n} \).

(a) There exists at least one basis of \( U \).

(b) Any two bases of \( U \) have the same size (= number of vectors).

(c) Given \( k \) linearly independent vectors in \( U \), and given \( \ell \) vectors that span \( U \), we always have \( k \leq \ell \).

(d) Any list of \( k \) linearly independent vectors in \( U \) can be extended to a basis of \( U \).

(e) Any list of \( \ell \) vectors that span \( U \) can be shrunk to a basis of \( U \) (i.e., we can remove some vectors from this list to get a basis of \( U \)).

Proof sketch. This all is proven just as in standard linear algebra. (For a specific reference: See [Heffer17, Chapter Two, Theorem III.2.5] for part (b); [Heffer17, Chapter Two, Corollary III.2.13] for part (d); [Heffer17, Chapter Two, Corollary III.2.14] for part (e). Part (a) follows by applying part (d) to the empty list (with \( k = 0 \)). Part (c) follows from parts (b), (d) and (e) once you extend your list of \( k \) linearly independent vectors in \( U \) to a basis of \( U \) and shrink your list of \( \ell \) vectors spanning \( U \) to a basis of \( U \).)

Definition 5.10.6. Let \( U \) be a subspace of \( K^{1 \times n} \).

The dimension of \( U \) is defined to be the size of a basis of \( U \). (Parts (a) and (b) of Theorem 5.10.5 show that this is indeed well-defined.) The dimension of \( U \) is denoted by \( \text{dim} \ U \).

Proposition 5.10.7. Let \( U \) and \( V \) be two subspaces of \( K^{1 \times n} \) such that \( U \subseteq V \).

(a) We have \( \text{dim} \ U \leq \text{dim} \ V \).

(b) If \( \text{dim} \ U = \text{dim} \ V \), then \( U = V \).

Proof sketch. Pick a basis of \( U \). This basis is a list of \( \text{dim} \ U \) many linearly independent vectors in \( V \) (since \( U \subseteq V \)). Thus, Theorem 5.10.5 (d) (applied to \( \text{dim} \ U \) and \( V \) instead of \( k \) and \( U \)) shows that this list can be extended to a basis of \( V \). The latter basis, of course, has size \( \text{dim} \ V \). Thus, \( \text{dim} \ U \leq \text{dim} \ V \) (since we have extended a list of size \( \text{dim} \ U \) and obtained a list of size \( \text{dim} \ V \)). This proves Proposition 5.10.7 (a).

(b) Assume that \( \text{dim} \ U = \text{dim} \ V \). We have just found a basis of \( V \) by extending a basis of \( U \). In light of \( \text{dim} \ U = \text{dim} \ V \), this extension must have been trivial – i.e., we must have extended our basis of \( U \) by no further vectors. This means that our basis of \( U \) was already a basis of \( V \) to begin with. From this, it is easy to see that \( U = V \) (because the span of a basis of \( U \) is always \( U \), whereas the span of a basis of \( V \) is always \( V \)). This proves Proposition 5.10.7 (b).

Now, let us connect this with matrices:
Definition 5.10.8. Let \( A \in \mathbb{K}^{n \times m} \) be a matrix.

(a) The **row space** of \( A \) is defined to be the span of the rows of \( A \). This is a subspace of \( \mathbb{K}^{1 \times m} \), and is called \( \text{Row} \ A \).

(b) The **column space** of \( A \) is defined to be the span of the columns of \( A \). This is a subspace of \( \mathbb{K}^{n \times 1} \), and is called \( \text{Col} \ A \).

Theorem 5.10.9. Let \( A \in \mathbb{K}^{n \times m} \) be a matrix. Then, \( \dim \text{Row} \ A = \dim \text{Col} \ A \).

*Proof of Theorem 5.10.9 (sketched).* One way to prove this is to transform \( A \) into RREF, and argue that both \( \dim \text{Row} \ A \) and \( \dim \text{Col} \ A \) equal the number of pivots in the RREF. There are other, more abstract ways. See linear algebra textbooks for this proof: for example, [Heffer17, Chapter Two, Theorem III.3.11] (where \( \dim \text{Row} \ A \) is called the “row rank” of \( A \), and \( \dim \text{Col} \ A \) is called the “column rank” of \( A \)). □

Definition 5.10.10. Let \( A \in \mathbb{K}^{n \times m} \) be a matrix. Theorem 5.10.9 shows that \( \dim \text{Row} \ A = \dim \text{Col} \ A \). This number \( \dim \text{Row} \ A = \dim \text{Col} \ A \) is called the **rank** of \( A \) and is denoted by \( \text{rank} \ A \).

The following is easy to see:

Proposition 5.10.11. Let \( A \in \mathbb{K}^{n \times m} \) be a matrix. Then, \( \text{rank} \ A \) is an integer between 0 and \( \min \{n, m\} \).

So we have seen that a matrix gives rise to two subspaces: its row space and its column space. But there is more:

Definition 5.10.12. Let \( A \in \mathbb{K}^{n \times m} \) be a matrix.

(a) The **kernel** (or nullspace) of \( A \) is defined to be the set of all column vectors \( v \in \mathbb{K}^{m \times 1} \) such that \( Av = 0_{n \times 1} \). This is a subspace of \( \mathbb{K}^{m \times 1} \), and is called \( \text{Ker} \ A \).

(b) The **left kernel** (or left nullspace) of \( A \) is defined to be the set of all row vectors \( w \in \mathbb{K}^{1 \times n} \) such that \( wA = 0_{1 \times m} \). This is a subspace of \( \mathbb{K}^{1 \times n} \).

Altogether, we have thus found four subspaces coming out of a matrix \( A \). These are the famous “four fundamental subspaces” (in Gilbert Strang’s terminology). One result that connects two of them is the following fact, known as [the rank-nullity theorem](#).

Theorem 5.10.13. Let \( A \in \mathbb{K}^{n \times m} \) be a matrix. Then,

\[
\text{rank} \ A + \dim \text{Ker} \ A = m.
\]

*Proof of Theorem 5.10.13 (sketched).* Most textbooks state Theorem 5.10.13 not in terms of matrices, but rather in the (equivalent) language of linear maps. For example, this is how it is stated in [Heffer17, Chapter Three, Theorem II.2.14]. □

Note that the number \( \dim \text{Ker} \ A \) is known as the **nullity** of a matrix \( A \).
5.10.4. Linear algebra over $\mathbb{Z}/2$: “button madness” / “lights out”

We now discuss an old puzzle, which is known as “button madness” or “lights out” (more precisely, these are two slightly different variants of the same puzzle). You can try it out on

http://bz.var.ru/comp/web/js/floor.html

(see also https://www.win.tue.nl/~aeb/ca/madness/madrect.html for a list of mathematical sources on this puzzle).

One version of this puzzle gives you 16 lamps arranged into a $4 \times 4$-grid. Each lamp comes with a lightswitch; but this lightswitch toggles not just this lamp, but also its four adjacent lamps (or three or two adjacent lamps, if the lamp you have toggled is at the border of the grid). For example, if your grid looks like this:

\[
\begin{array}{cccc}
1 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 \\
1 & 1 & 0 & 1 \\
\end{array}
\]

(where an entry 1 means a lamp turned on, and an entry 0 means a lamp turned off), and you flip the lightswitch in cell $(2,3)$ (that is, the third cell from the left in the second row from the top), then you obtain the grid

\[
\begin{array}{cccc}
1 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 \\
\end{array}
\]

(A total of 5 lamps have changed their state: three have been turned off, and two have been turned on.) If you then flip the lightswitch in cell $(1,3)$ of this new grid, then you obtain the grid

\[
\begin{array}{cccc}
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 \\
\end{array}
\]

At the beginning, all lamps are turned off. Your goal is to achieve the opposite state (i.e., all lamps being on at the same time) by flipping a sequence of lightswitches. Is this possible, and how? (In some versions of this puzzle – such as the “lights out” version – it’s exactly the other way round: The lights are all on
initially, and you must turn them all off. Of course, this makes no difference to the solution.)

In some versions of this puzzle, the grid is “toroidal”, in the sense that it is understood to wrap around – for example, the cells \((1, 4)\) and \((1, 1)\) are considered to be adjacent, and so are the cells \((4, 1)\) and \((1, 1)\). We shall not consider this case here, but it can be solved by the same method.

Of course, you can play the same game on larger grids, triangular grids, etc. But in order to get a grip on how to solve such a puzzle, we shall first analyze a much simpler version: the “1-dimensional version” of the puzzle.

Here is this “1-dimensional version”: We have 4 lamps in a row (numbered 1, 2, 3, 4), each equipped with a lightswitch. The lightswitch at lamp \(i\) toggles lamp \(i\), lamp \(i - 1\) (if it exists) and lamp \(i + 1\) (if it exists). Initially, all 4 lamps are off. Can we turn them all on by flipping a sequence of lightswitches?

Yes, of course: we just have to flip the lightswitches at lamps 1 and 4. But let us pretend that we aren’t that smart, and instead try to solve the puzzle systematically.

We model the states of our lamps by a row vector in \((\mathbb{Z}/2)^{1\times 4}\). We write a row vector \((a_1, a_2, \ldots, a_4)\) as \((a_1, a_2, a_3, a_4)\) ∈ \((\mathbb{Z}/2)^{1\times 4}\), where

\[
a_i = \begin{cases} 
[0]_2, & \text{if lamp } i \text{ is off;} \\
[1]_2, & \text{if lamp } i \text{ is on}
\end{cases}

\]

We shall write 0 and 1 for \([0]_2\) and \([1]_2\) throughout this section (except in Proposition 5.10.14), so we can rewrite this as

\[
a_i = \begin{cases} 
0, & \text{if lamp } i \text{ is off;} \\
1, & \text{if lamp } i \text{ is on}
\end{cases} = \left[\text{lamp } i \text{ is on}\right]_{2},
\]

but keep in mind that these values are understood to be in \(\mathbb{Z}/2\).

The initial state is \((0, 0, 0, 0)\). The final state that we want to achieve is \((1, 1, 1, 1)\). Flipping a lightswitch corresponds to adding a certain row vector to our state. Namely:

- Flipping lightswitch 1 means adding \((1, 1, 0, 0)\).
- Flipping lightswitch 2 means adding \((1, 1, 1, 0)\).
- Flipping lightswitch 3 means adding \((0, 1, 1, 1)\).
- Flipping lightswitch 4 means adding \((0, 0, 1, 1)\).
Thus, flipping a lightswitch means adding the corresponding row of the matrix

\[
A := \begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 \\
\end{pmatrix} \in (\mathbb{Z}/2)^{4 \times 4}.
\]

The reachable states are thus exactly the elements of Row \(A\), the row space of \(A\).

Hence, our goal is to show that \((1, 1, 1, 1) \in \text{Row } A\).

This is quite easy for the concrete matrix \(A\) above, but let us try a theoretical argument. It will rely on the following general fact:

**Proposition 5.10.14.** Let \(\mathbb{K}\) be a field. Let \(A \in \mathbb{K}^{n \times m}\) and \(b \in \mathbb{K}^{1 \times m}\). Assume the following:

**Assumption 1:** If \(c \in \mathbb{K}^{m \times 1}\) satisfies \(Ac = 0\), then \(bc = 0\).

Then, \(b \in \text{Row } A\).

**Proof of Proposition 5.10.14.** Let \(\begin{pmatrix} A \\ b \end{pmatrix}\) denote the \((n + 1) \times m\)-matrix formed from \(A\) by attaching the row vector \(b\) to its bottom. For example, if \(A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}\) and \(b = (b_1 \ b_2)\), then \(\begin{pmatrix} A \\ b \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ b_1 & b_2 \end{pmatrix}\).

Theorem 5.10.13 yields \(\text{rank } A + \text{dim Ker } A = m\). Thus,

\[
\text{rank } A = m - \text{dim Ker } A. \quad (163)
\]

The same argument can be applied to the matrix \(\begin{pmatrix} A \\ b \end{pmatrix}\) instead of \(A\). We thus obtain

\[
\text{rank } \begin{pmatrix} A \\ b \end{pmatrix} = m - \text{dim Ker } \begin{pmatrix} A \\ b \end{pmatrix}. \quad (164)
\]

But each \(c \in \text{Ker } A\) satisfies \(Ac = 0\) and thus \(bc = 0\) (by Assumption 1), and therefore \(c \in \text{Ker } \begin{pmatrix} A \\ b \end{pmatrix}\) (because the “row-by-column” nature of matrix multiplication shows that \(\begin{pmatrix} A \\ b \end{pmatrix}c = \begin{pmatrix} Ac \\ bc \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0\)). So we have \(\text{Ker } A \subseteq \text{Ker } \begin{pmatrix} A \\ b \end{pmatrix}\).

Also, clearly, \(\text{Ker } \begin{pmatrix} A \\ b \end{pmatrix} \subseteq \text{Ker } A\) (because if \(c \in \text{Ker } \begin{pmatrix} A \\ b \end{pmatrix}\), then \(\begin{pmatrix} A \\ b \end{pmatrix}c = 0\), so that \(0 = \begin{pmatrix} A \\ b \end{pmatrix}c = \begin{pmatrix} Ac \\ bc \end{pmatrix}\) and thus \(Ac = 0\), so that \(c \in \text{Ker } A\)). Combining these two relations, we obtain \(\text{Ker } A = \text{Ker } \begin{pmatrix} A \\ b \end{pmatrix}\). Hence, the right hand sides of (163)
and (164) are equal. Thus, the left hand sides are equal as well. In other words, rank $A = \text{rank} \begin{pmatrix} A \\ b \end{pmatrix}$. In other words,

$$\dim \text{Row} A = \dim \text{Row} \begin{pmatrix} A \\ b \end{pmatrix}$$

(since rank $B = \dim \text{Row} B$ for any matrix $B$). But $\text{Row} A \subseteq \text{Row} \begin{pmatrix} A \\ b \end{pmatrix}$ (by definition of a row space). Combining these facts, we obtain

$$\text{Row} A = \text{Row} \begin{pmatrix} A \\ b \end{pmatrix}$$

(by Proposition 5.10.7 (b), applied to $U = \text{Row} A$ and $V = \text{Row} \begin{pmatrix} A \\ b \end{pmatrix}$). Thus,

$$b \in \text{Row} \begin{pmatrix} A \\ b \end{pmatrix} = \text{Row} A, \text{ qed.} \square$$

Over the field $\mathbb{Z}/2$, this fact has the following consequence:

**Corollary 5.10.15.** Let $A \in (\mathbb{Z}/2)^{n \times n}$ be a symmetric matrix. ("Symmetric" means that the $(i,j)$-th entry of $A$ equals the $(j,i)$-th entry of $A$ for all $i$ and $j$. In other words, it means that $A^T = A$.)

Let $d$ be the diagonal of $A$, written as a row vector. (In other words, let $D = (a_{1,1}, a_{2,2}, \ldots, a_{n,n})$, where $a_{i,j}$ is the $(i,j)$-th entry of $A$.)

Then, $d \in \text{Row} A$.

Note that Corollary 5.10.15 brutally fails over fields different from $\mathbb{Z}/2$. For example, if we allow $A$ to be a matrix in $\mathbb{Z}^{n \times n}$ instead, then $A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ is symmetric but its diagonal $d = (1,1)$ does not belong to $\text{Row} A$.

**Proof of Corollary 5.10.15** Set $\mathbb{K} = \mathbb{Z}/2$. By Proposition 5.10.14 (applied to $m = n$ and $b = d$), it suffices to show the following:

**Assumption 1:** If $c \in \mathbb{K}^{n \times 1}$ satisfies $Ac = 0$, then $dc = 0$.

[**Proof of Assumption 1:** Let $c \in \mathbb{K}^{n \times 1}$ satisfy $Ac = 0$. We must prove that $dc = 0$. I claim that $dc = c^T Ac$.]
To see this, write \( c \) as \( c = (c_1, c_2, \ldots, c_n)^T \) and \( A \) as \( A = (a_{ij})_{1 \leq i \leq n, 1 \leq j \leq n} \). Then, expanding \( c^T A c \) yields

\[
c^T A c = \sum_{i,j} c_i a_{ij} c_j = \sum_{i} a_{ii} c_i^2 - \sum_{i < j} c_i a_{ij} c_j + \sum_{i > j} a_{ij} c_i c_j.
\]

(Here, we have renamed the indices \( i \) and \( j \) as \( j \) and \( i \))

\[
= \sum_{j<i} a_{j,i} c_i c_j + \sum_{i} a_{ii} c_i - \sum_{j<i} a_{j,i} c_i c_j = 2 \sum_{j<i} a_{j,i} c_i c_j = 0
\]

(since \( A \) is symmetric and its diagonal is \( (1, 1, \ldots, 1) \).)

The same argument works for the “proper” (2-dimensional) lights-out puzzle; we just have to use row vectors of size 16 (not 4) and 16 \( \times \) 16-matrices (not 4 \( \times \) 4-matrices). More generally, the same argument works for any such puzzle on any “grid” as long as:

- each lamp \( i \) has a lightswitch which toggles at least lamp \( i \);
- if the lightswitch at lamp \( i \) toggles lamp \( j \), then the lightswitch at lamp \( j \) toggles lamp \( i \).

These conditions guarantee that the corresponding matrix \( A \) will be symmetric and its diagonal will be \( (1, 1, \ldots, 1) \) (and thus we can apply Corollary 5.10.15).

How to find the exact sequence of flips that results in all lights being on? This is tantamount to finding the coefficient of a linear combination of the rows of \( A \) that equals \( (1, 1, \ldots, 1) \). This boils down to solving a system of linear equations over \( \mathbb{Z}/2 \), which can be achieved using Gaussian elimination.

What other states can be achieved by flipping lightswitches? Again, for each specific grid and each specific state, this can be solved by Gaussian elimination;
but characterizing the reachable states more explicitly is a hard problem with no unified answer. (See the link above.)

2019-04-15 lecture

5.10.5. A warning about orthogonality and positivity

I have said above that “more or less” all linear algebra over \( \mathbb{R} \) works identically over any field \( \mathbb{K} \). There is an exception: Anything that uses positivity will break down over some fields \( \mathbb{K} \).

One thing that uses positivity is QR-decomposition. And indeed, not every matrix over an arbitrary field has a QR-decomposition. You can still define dot products and orthogonal complements of subspaces. But it is no longer true that \( V = U \oplus U^\perp \) for a subspace \( U \) of \( V \). It can happen that \( U \cap U^\perp \neq 0 \). For example, there are column vectors \( v \neq 0 \) that are orthogonal to themselves with respect to the dot product (that is, \( v^T v = 0 \)).

**Example:** In \( \mathbb{Z}/3 \), we have

\[
\begin{pmatrix}
1 \\
1 \\
1
\end{pmatrix}^T
\begin{pmatrix}
1 \\
1 \\
1
\end{pmatrix} = (1, 1, 1)
\begin{pmatrix}
1 \\
1
\end{pmatrix} = 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 1 = 3 = 0.
\]

So the vector \( \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \) is orthogonal to itself.

5.11. Modules

Roughly speaking, a module is the same as a vector space, except that it is over a commutative ring instead of a field:

**Definition 5.11.1.** Let \( \mathbb{K} \) be a commutative ring.

A \( \mathbb{K} \)-module means a set \( M \) equipped with a binary operation \( + \) (called “addition”), a map

\[
\cdot : \mathbb{K} \times M \to M
\]

(called “scaling”) and an element \( 0_M \) satisfying the following axioms (where, again, we write \( a \cdot b \) or \( ab \) for \( \cdot (a,b) \)):

(a) We have \( a + b = b + a \) for all \( a, b \in M \).

(b) We have \( a + (b + c) = (a + b) + c \) for all \( a, b, c \in M \).

(c) We have \( a + 0_M = 0_M + a = a \) for all \( a \in M \).

(d) Each \( a \in M \) has an additive inverse (i.e., there is an \( a' \in M \) such that \( a + a' = a' + a = 0_M \)).

(e) We have \( \lambda (a + b) = \lambda a + \lambda b \) for all \( \lambda \in \mathbb{K} \) and \( a, b \in M \).

(f) We have \( (\lambda + \mu) a = \lambda a + \mu a \) for all \( \lambda, \mu \in \mathbb{K} \) and \( a \in M \).
(g) We have $0a = 0_M$ for all $a \in M$.
(h) We have $(\lambda \mu) a = \lambda (\mu a)$ for all $\lambda, \mu \in \mathbb{K}$ and $a \in M$.
(i) We have $1a = a$ for all $a \in M$.

Definition 5.11.2. If $\mathbb{K}$ is a commutative ring and $M$ is a $\mathbb{K}$-module, then the elements of $M$ are called vectors, while the elements of $\mathbb{K}$ are called scalars.

Definition 5.11.3. If $\mathbb{K}$ is a field, then $\mathbb{K}$-modules are called $\mathbb{K}$-vector spaces. (When $\mathbb{K} = \mathbb{R}$, these are the usual real vector spaces known from undergraduate linear algebra classes.)

Example 5.11.4. (a) If $\mathbb{K}$ is a commutative ring, and $n \in \mathbb{N}$, then $\mathbb{K}^n$ (that is, the set of all $n$-tuples of elements of $\mathbb{K}$) is a $\mathbb{K}$-module, if you equip it with entrywise addition (that is,

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$

) and entrywise scaling (that is,

$$\lambda (a_1, a_2, \ldots, a_n) = (\lambda a_1, \lambda a_2, \ldots, \lambda a_n)$$

).

(b) If $\mathbb{K}$ is a commutative ring, and $n, m \in \mathbb{N}$, then $\mathbb{K}^{n \times m}$ is a $\mathbb{K}$-module, if you equip it with the addition and the scaling that we defined above.

(c) For each integer $n$, the set $\mathbb{Z}/n$ is a $\mathbb{Z}$-module, if you equip it with the addition and the scaling that we defined above.

6. Solutions to the exercises

6.1. Solution to Exercise 2.2.1

Solution to Exercise 2.2.1 The definition of $|a|$ shows that $|a|$ equals either $a$ or $-a$. In other words, $|a|$ equals either $1a$ or $(-1) a$. In other words, $|a| = qa$ for some $q \in \{1, -1\}$. Consider this $q$. Clearly, $q$ is an integer. Now, from $|a| = qa = aq$, we conclude that $a \mid |a|$ (since $q$ is an integer). This solves Exercise 2.2.1 (a).

(b) From $q \in \{1, -1\}$, we obtain $q^2 \in \{1^2, (-1)^2\} = \{1, 1\} = \{1\}$, so that $q^2 = 1$. Now, multiplying the equality $|a| = qa$ by $q$, we obtain $q |a| = \underbrace{qq}_{=q^2=1} a = a$. Hence, $a = q |a| = |a| \cdot q$. Thus, $|a| \mid a$ (since $q$ is an integer). This solves Exercise 2.2.1 (b).

6.2. Solution to Exercise 2.2.2

Solution to Exercise 2.2.2 We are in one of the following two cases:

Case 1: We have $b \neq 0$. 


Case 2: We have $b = 0$.
Let us first consider Case 1. In this case, we have $b \neq 0$. Thus, Proposition \ref{prop:2.2.3} (b) yields $|a| \leq |b|$ (since $a \mid b$).

We have $a \mid b$. In other words, there exists an integer $c$ such that $b = ac$. Consider this $c$. If we had $a = 0$, then we would have $b = \lim_{c \to 0} ac = 0$, which would contradict $b \neq 0$. Thus, we cannot have $a = 0$. Hence, Proposition \ref{prop:2.2.3} (b) (applied to $b$ and $a$ instead of $a$ and $b$) yields $|b| \leq |a|$ (since $b \mid a$). Combining this with $|a| \leq |b|$, we obtain $|a| = |b|$. Thus, Exercise \ref{ex:2.2.2} is solved in Case 1.

Let us now consider Case 2. In this case, we have $b = 0$. But we have $b \mid a$. In other words, there exists an integer $c$ such that $a = bc$. Consider this $c$. Hence, $a = bc = 0c = 0 = b$ (since $b = 0$). Thus, $|a| = |b|$. Hence, Exercise \ref{ex:2.2.2} is solved in Case 2.

Now, we have solved Exercise \ref{ex:2.2.2} in both Cases 1 and 2. Hence, Exercise \ref{ex:2.2.2} always holds.

6.3. Solution to Exercise \ref{ex:2.2.3}

Solution to Exercise \ref{ex:2.2.3} $\iff$: Assume that $a \mid b$ holds. We must prove that $ac \mid bc$.

It is easy to do this straight from the definition of divisibility, but here is a shorter argument: Proposition \ref{prop:2.2.4} (a) (applied to $c$ instead of $a$) yields $c \mid c$. Also, $a \mid b$. Hence, Proposition \ref{prop:2.2.4} (c) (applied to $a_1 = a, b_1 = b, a_2 = c$ and $b_2 = c$) yields $ac \mid bc$. This proves the $\iff$ direction of Exercise \ref{ex:2.2.3}

$\iff$: Assume that $ac \mid bc$ holds. We must prove that $a \mid b$.

We have $ac \mid bc$. In other words, there exists an integer $d$ such that $bc = (ac)d$ (by Definition \ref{def:2.2.1}). Consider this $d$. We have $bc = (ac)d = adc$. We can divide both sides of this equality by $c$ (since $c \neq 0$), and thus obtain $b = ad$. Thus, there exists an integer $e$ such that $b = ae$ (namely, $e = d$). In other words, $a \mid b$ (by Definition \ref{def:2.2.1}). This proves the $\iff$ direction of Exercise \ref{ex:2.2.3}

6.4. Solution to Exercise \ref{ex:2.2.4}

Solution to Exercise \ref{ex:2.2.4} We have $b - a \geq 0$ (since $a \leq b$), thus $b - a \in \mathbb{N}$. Hence, $n^{b-a}$ is a well-defined integer. Now, $n^b = n^an^{b-a}$ (since $n^a n^{b-a} = n^{a+(b-a)} = n^b$). Hence, there exists an integer $c$ such that $n^b = n^bc$ (namely, $c = n^{b-a}$). In other words, $n^a \mid n^b$ (by the definition of divisibility). This solves Exercise \ref{ex:2.2.4}

6.5. Solution to Exercise \ref{ex:2.2.5}

Solution to Exercise \ref{ex:2.2.5} Assume the contrary. Thus, $g \neq 1$. But Proposition \ref{prop:2.2.3} (b) (applied to $g$ and 1 instead of $a$ and $b$) yields $|g| \leq |1|$ (since $g \mid 1$ and 1 $\neq 0$). But $g$ is nonnegative; hence, $|g| = g$, so that $g = |g| \leq |1| = 1$. Combining this with $g \neq 1$, we obtain $g < 1$. Hence, $g = 0$ (since $g$ is a nonnegative integer).

But $g \mid 1$. In other words, there exists an integer $c$ such that $1 = gc$ (by Definition \ref{def:2.2.1}). Consider this $c$. Now, $1 = \lim_{c \to 0} gc = 0c = 0$. This contradicts $1 \neq 0$. This contradiction...
shows that our assumption was wrong. Hence, Exercise 2.2.5 is solved.

### 6.6. Solution to Exercise 2.2.6

**Solution to Exercise 2.2.6.** We have \( a \mid b \). In other words, there exists an integer \( d \) such that \( b = ad \) (by Definition 2.2.1). Consider this \( d \). Clearly, \( d^k \) is an integer (since \( d \) is an integer and \( k \in \mathbb{N} \)). From \( b = ad \), we obtain \( b^k = (ad)^k = a^kd^k \). Hence, there exists an integer \( c \) such that \( b^k = a^kc \) (namely, \( c = d^k \)). In other words, \( a^k \mid b^k \) (by Definition 2.2.1). This solves Exercise 2.2.6.

### 6.7. Solution to Exercise 2.3.1

**Solution to Exercise 2.3.1.** According to Definition 2.3.1, we have \( a + b \equiv a - b \mod 2 \) if and only if \( 2 \mid (a + b) - (a - b) \). Thus, it remains to prove that \( 2 \mid (a + b) - (a - b) \). But this follows immediately from \( (a + b) - (a - b) = 2b \). Thus Exercise 2.3.1 is solved.

### 6.8. Solution to Exercise 2.3.2

**Solution to Exercise 2.3.2.** There are many such examples. Here is one:

\[
\begin{align*}
n & = 8, \\
a_1 & = 10, \\
a_2 & = 2, \\
b_1 & = 10, \\
b_2 & = 10.
\end{align*}
\]

These satisfy \( a_1 \equiv b_1 \mod n \) and \( a_2 \equiv b_2 \mod n \) but neither \( a_1/a_2 \equiv b_1/b_2 \mod n \) nor \( a_1^{a_2} \equiv b_1^{b_2} \mod n \).

It is much easier to find examples which fail only one of the two congruences \( a_1/a_2 \equiv b_1/b_2 \mod n \). In other words, \( a_1^k \equiv b_1^k \mod n \).

### 6.9. Solution to Exercise 2.3.3

**Solution to Exercise 2.3.3.** We have \( a \equiv b \mod n \). In other words, \( n \mid a - b \) (by the definition of congruence). Note that all of \( a/d, b/d \) and \( n/d \) are integers (since \( d \) divides each of \( a, b, n \)). Hence, \( (a - b)/d = a/d - b/d \) is an integer as well. Hence, Exercise 2.2.3 (applied to \( n/d, (a - b)/d \) and \( d \) instead of \( a, b \) and \( c \)) shows that \( n/d \mid (a - b)/d \) holds if and only if \( (n/d)/d \mid ((a - b)/d)/d \). Since \( (n/d)/d \mid ((a - b)/d)/d \) does hold (indeed, this is just a complicated way to say \( n \mid a - b \)), we thus conclude that \( n/d \mid (a - b)/d \) holds. In other words, \( n/d \mid a/d - b/d \) (since \( (a - b)/d = a/d - b/d \)). In other words, \( a/d \equiv b/d \mod n/d \) (by the definition of congruence). This solves Exercise 2.3.3.

### 6.10. Solution to Exercise 2.3.4

**First solution to Exercise 2.3.4.** We want to prove that

\[
a^k \equiv b^k \mod n \quad \text{for each } k \in \mathbb{N}.
\]

We shall prove this by induction on \( k \):
**Induction base.** Proposition 2.3.4 (a) yields $1 \equiv 1 \mod n$. In view of $a^0 = 1$ and $b^0 = 1$, this rewrites as $a^0 \equiv b^0 \mod n$. In other words, (165) holds for $k = 0$. This completes the induction base.

**Induction step.** Let $\ell \in \mathbb{N}$. Assume that (165) holds for $k = \ell$. We must prove that (165) holds for $k = \ell + 1$.

We have assumed that (165) holds for $k = \ell$. In other words, we have $a^\ell \equiv b^\ell \mod n$. Also, recall that $a \equiv b \mod n$. Hence, (165) (applied to $c = a^\ell$ and $d = b^\ell$) yields $aa^\ell \equiv bb^\ell \mod n$. In other words, $a^{\ell+1} \equiv b^{\ell+1} \mod n$ (since $aa^\ell = a^{\ell+1}$ and $bb^\ell = b^{\ell+1}$). In other words, (165) holds for $k = \ell + 1$. This completes the induction step. Thus, (165) is proven by induction. Therefore, Exercise 2.3.4 is solved.

**Second solution to Exercise 2.3.4** Recall that
\[
(a - b) \left( a^{k-1} + a^{k-2}b + a^{k-3}b^2 + \cdots + ab^{k-2} + b^{k-1} \right) = a^k - b^k \tag{166}
\]
for every $k \in \mathbb{N}$. (This is a well-known identity, and it appears (with $k$ renamed as $n$) as the first half of Exercise 1 on homework set #0.)

Now, let $k \in \mathbb{N}$. We have assumed that $a \equiv b \mod n$. In other words, $n \mid a - b$. In other words, there exists an integer $c$ such that $a - b = nc$. Consider this $c$. Now, (166) yields
\[
a^k - b^k = (a - b) \left( a^{k-1} + a^{k-2}b + a^{k-3}b^2 + \cdots + ab^{k-2} + b^{k-1} \right) = nc \left( a^{k-1} + a^{k-2}b + a^{k-3}b^2 + \cdots + ab^{k-2} + b^{k-1} \right).
\]

The right hand side of this equality is clearly divisible by $n$. Hence, so is the left hand side. In other words, $n \mid a^k - b^k$. In other words, $a^k \equiv b^k \mod n$. Hence, Exercise 2.3.4 is solved again.

### 6.11. Solution to Exercise 2.3.5

**Solution to Exercise 2.3.5 (a)** We shall solve Exercise 2.3.5 (a) by induction on $|S|$:

- **Induction base:** Exercise 2.3.5 (a) holds whenever $|S| = 0$. This completes the induction base.

- **Induction step:** Fix $k \in \mathbb{N}$. Assume that Exercise 2.3.5 (a) holds whenever $|S| = k$. We must prove that Exercise 2.3.5 (a) holds whenever $|S| = k + 1$.

We have assumed that Exercise 2.3.5 (a) holds whenever $|S| = k$. In other words, the following statement is true:

**Statement 1:** Let $n$, $S$, $a_s$ and $b_s$ be as in Exercise 2.3.5. Assume that $|S| = k$.

Then, $\sum_{s \in S} a_s \equiv \sum_{s \in S} b_s \mod n$.

**Proof.** Let $n$, $S$, $a_s$ and $b_s$ be as in Exercise 2.3.5 and assume that $|S| = 0$. Then, the set $S$ is empty (since $|S| = 0$), and thus we have $\sum_{s \in S} a_s = (\text{empty sum}) = 0$. Similarly, $\sum_{s \in S} b_s = 0$. Now, Proposition 2.3.4 (a) yields $0 \equiv 0 \mod n$. In view of $\sum_{s \in S} a_s = 0$ and $\sum_{s \in S} b_s = 0$, this rewrites as $\sum_{s \in S} a_s \equiv \sum_{s \in S} b_s \mod n$. Thus, Exercise 2.3.5 (a) holds in our case.

So we have shown that Exercise 2.3.5 (a) holds whenever $|S| = 0$. 

---

144 **Proof.** Let $n$, $S$, $a_s$ and $b_s$ be as in Exercise 2.3.5 and assume that $|S| = 0$. Then, the set $S$ is empty (since $|S| = 0$), and thus we have $\sum_{s \in S} a_s = (\text{empty sum}) = 0$. Similarly, $\sum_{s \in S} b_s = 0$. Now, Proposition 2.3.4 (a) yields $0 \equiv 0 \mod n$. In view of $\sum_{s \in S} a_s = 0$ and $\sum_{s \in S} b_s = 0$, this rewrites as $\sum_{s \in S} a_s \equiv \sum_{s \in S} b_s \mod n$. Thus, Exercise 2.3.5 (a) holds in our case.

So we have shown that Exercise 2.3.5 (a) holds whenever $|S| = 0$. 

---

Math 4281 notes page 372
Now, we must prove that Exercise 2.3.5(a) holds whenever \(|S| = k + 1\). In other words, we must prove the following statement:

**Statement 2:** Let \(n, S, a_s\) and \(b_s\) be as in Exercise 2.3.5 Assume that \(|S| = k + 1\). Then, \(\sum_{s \in S} a_s \equiv \sum_{s \in S} b_s \mod n\).

**Proof of Statement 2:** We have \(|S| = k + 1 > k \geq 0\); thus, the set \(S\) is nonempty. Hence, there exists some \(t \in S\). Pick such a \(t\). Thus, \(|S \setminus \{t\}| = |S| - 1 = k\) (since \(|S| = k + 1\)). Moreover, from (7), we immediately obtain that

\[a_s \equiv b_s \mod n\quad \text{for each } s \in S \setminus \{t\}\]

(since each \(s \in S \setminus \{t\}\) belongs to \(S\)). Hence, we can apply Statement 1 to \(S \setminus \{t\}\) instead of \(S\). We thus obtain

\[\sum_{s \in S \setminus \{t\}} a_s \equiv \sum_{s \in S \setminus \{t\}} b_s \mod n\]

Also, we have

\[a_t \equiv b_t \mod n\]

(by (7), applied to \(s = t\)). Adding these two congruences together, we obtain

\[\sum_{s \in S \setminus \{t\}} a_s + a_t \equiv \sum_{s \in S \setminus \{t\}} b_s + b_t \mod n\]

In view of

\[\sum_{s \in S} a_s = \sum_{s \in S \setminus \{t\}} a_s + a_t \quad \text{ (here, we have split off the addend for } s = t \text{ from the sum)}\]

and

\[\sum_{s \in S} b_s = \sum_{s \in S \setminus \{t\}} b_s + b_t \quad \text{ (here, we have split off the addend for } s = t \text{ from the sum)}\],

this can be rewritten as

\[\sum_{s \in S} a_s \equiv \sum_{s \in S} b_s \mod n\]

This proves Statement 2.

We have now proven Statement 2; this means that Exercise 2.3.5(a) holds whenever \(|S| = k + 1\). This completes the induction step; thus, Exercise 2.3.5(a) is solved.

**b)** The solution to Exercise 2.3.5(b) is analogous to the one we gave above for Exercise 2.3.5(a); the main difference is that we have to replace sums by products (and 0 by 1).

### 6.12. Solution to Exercise 2.3.6

**Solution to Exercise 2.3.6** No, it is not true. For example, \(a_1 = 1, a_2 = 1, b_1 = 1, b_2 = 0, n_1 = 0\) and \(n_2 = 1\) yield a counterexample.
6.13. Solution to Exercise 2.3.7

Solution to Exercise 2.3.7. If \( a \equiv b \mod n \), then \( b \equiv a \mod n \) (by Proposition 2.3.4 (c)). In other words, the implication \( (a \equiv b \mod n) \implies (b \equiv a \mod n) \) holds. The same argument (but with the roles of \( a \) and \( b \) swapped) shows that the implication \( (b \equiv a \mod n) \implies (a \equiv b \mod n) \) holds. Combining these two implications, we obtain the logical equivalence \( (a \equiv b \mod n) \iff (b \equiv a \mod n) \). Thus, we have the following chain of logical equivalences:

\[
(a \equiv b \mod n) \iff (b \equiv a \mod n) \\
\iff (n \mid b - a) \quad \text{(by the definition of congruence)} \\
\iff (\text{there exists an integer } d \text{ such that } b - a = nd) \\
\iff (\text{there exists an integer } d \text{ such that } b = a + nd)
\]

(since the equation \( b - a = nd \) for an integer \( d \) is equivalent to \( b = a + nd \)). In other words, \( a \equiv b \mod n \) if and only if there exists some \( d \in \mathbb{Z} \) such that \( b = a + nd \). This solves Exercise 2.3.7.

6.14. Solution to Exercise 2.3.8

Solution to Exercise 2.3.8. We have \( a - b \equiv c \mod n \) if and only if \( n \mid (a - b) - c \) (by the definition of congruence). Thus, we have the logical equivalence

\[
(a - b \equiv c \mod n) \iff (n \mid (a - b) - c).
\]

(167)

On the other hand, we have \( a \equiv b + c \mod n \) if and only if \( n \mid a - (b + c) \) (by the definition of congruence). Thus, we have the logical equivalence

\[
(a \equiv b + c \mod n) \iff (n \mid a - (b + c)).
\]

(168)

Now, we have the following chain of logical equivalences:

\[
(a - b \equiv c \mod n) \iff \left( n \mid \frac{(a - b) - c}{a - (b + c)} \right) \quad \text{(by (167))} \\
\iff (n \mid a - (b + c)) \iff (a \equiv b + c \mod n) \quad \text{(by (168))}.
\]

In other words, we have \( a - b \equiv c \mod n \) if and only if \( a \equiv b + c \mod n \). This solves Exercise 2.3.8.

6.15. Solution to Exercise 2.5.1

Solution to Exercise 2.5.1. We have \( 3^{2n+1} = \left( \underbrace{3^2}_{=9 \mod 7} \right)^n \cdot 3 \equiv 2^n \cdot 3 \mod 7 \). (This follows from the PSC, in its extended form that allows \( k \)-th powers in the expression \( A \). Alternatively, you can argue by hand as follows: We have \( 3^2 = 9 \equiv 2 \mod 7 \). Thus, Exercise 2.3.4).
(applied to $7$, $3^2$, $2$ and $n$ instead of $n$, $a$, $b$ and $k$) yields \((3^2)^n \equiv 2^n \mod 7\). Multiplying this congruence by the obvious congruence $3 \equiv 3 \mod 7$, we obtain \((3^2)^n \cdot 3 \equiv 2^n \cdot 3 \mod 7\).

Thus, \(3^{2n+1} = (3^2)^n \cdot 3 \equiv 2^n \cdot 3 \mod 7\).

Hence, again using the PSC, we obtain

\[
\frac{3^{2n+1}}{2^n \cdot 3} + \frac{2^{n+2}}{2^n \cdot 3} \equiv 2^n \cdot 3 + 2^n \cdot 4 = 2^n \cdot (3 + 4) = 2^n \cdot 7 \equiv 0 \mod 7
\]

(since \(2^n \cdot 7\) is clearly divisible by 7). In other words, \(7 \mid 3^{2n+1} + 2^{n+2}\). This solves Exercise 2.5.1

[Remark: Here is a sketch of a different solution: If we set \(a_n = 3^{2n+1} + 2^{n+2}\) for each \(n \in \mathbb{N}\), then we must prove that \(7 \mid a_n\) for all \(n \in \mathbb{N}\). But a straightforward computation reveals that

\[
a_n = 11a_{n-1} - 18a_{n-2}
\]

for each \(n \geq 2\). (169)

Thus, once we check that \(7 \mid a_0\) and \(7 \mid a_1\), we can use a straightforward strong induction on \(n\) to see that \(7 \mid a_n\) for all \(n \in \mathbb{N}\), which is exactly the claim of Exercise 2.5.1. Of course, finding the relation (169) was the main trick in this solution; it becomes somewhat natural once you know the theory of linear recurrences (such as the Fibonacci sequence).]

\[\square\]

6.16. Solution to Exercise 2.6.1

Solution to Exercise 2.6.1 \iff: Assume that \(u \equiv v \mod n\). We must prove that \(u \% n = v \% n\).

Corollary 2.6.9 (a) yields that \(u \% n \in \{0,1,\ldots,n-1\}\) and \(u \% n \equiv u \mod n\). Hence, \(u \% n \equiv u \equiv v \mod n\).

But Corollary 2.6.9 (c) (applied to \(v\) instead of \(u\)) yields that if \(c \in \{0,1,\ldots,n-1\}\) is such that \(c \equiv v \mod n\), then \(c = v \% n\). Applying this to \(c = u \% n\), we obtain \(u \% n = v \% n\) (since \(u \% n \in \{0,1,\ldots,n-1\}\) and \(v \% n \equiv v \mod n\)). This proves the “\(\Rightarrow\)” direction of Exercise 2.6.1

\[\iff: \text{Assume that } u \% n = v \% n. \text{ We must prove that } u \equiv v \mod n. \]

Corollary 2.6.9 (a) yields that \(u \% n \in \{0,1,\ldots,n-1\}\) and \(u \% n \equiv u \mod n\). Corollary 2.6.9 (a) (applied to \(v\) instead of \(u\)) yields that \(v \% n \in \{0,1,\ldots,n-1\}\) and \(v \% n \equiv v \mod n\).

From \(u \% n \equiv u \mod n\), we obtain \(u \equiv u \% n = v \% n \equiv v \mod n\). Thus, we have proven \(u \equiv v \mod n\). This proves the “\(\Leftarrow\)” direction of Exercise 2.6.1

\[\square\]

6.17. Solution to Exercise 2.6.2

Solution to Exercise 2.6.2 (a) Theorem 2.6.1 shows that there exists a unique pair \((q, r) \in \mathbb{Z} \times \{0,1,\ldots,n-1\}\) such that \(u = qn + r\). Consider this pair \((q, r)\), and denote it by \((s, t)\). Thus, \((s, t) \in \mathbb{Z} \times \{0,1,\ldots,n-1\}\) is a pair satisfying \(u = sn + t\). From \((s, t) \in \mathbb{Z} \times \{0,1,\ldots,n-1\}\), we obtain \(s \in \mathbb{Z}\) and \(t \in \{0,1,\ldots,n-1\} \subseteq \mathbb{Z}\).

We are in one of the following two cases:

Case 1: We have \(t \leq n/2\).

Case 2: We have \(t > n/2\).

Let us first consider Case 1. In this case, we have \(t \leq n/2\). But \(t\) is nonnegative (since \(t \in \{0,1,\ldots,n-1\}\)); thus, \(|t| = t \leq n/2\). So we have \((s, t) \in \mathbb{Z} \times \mathbb{Z}\) (since \(s \in \mathbb{Z}\) and \(t \in \mathbb{Z}\)
and \( u = sn + t \) and \(|t| \leq n/2\). Hence, there exists a pair \((q,r) \in \mathbb{Z} \times \mathbb{Z}\) such that \( u = qn + r \) and \(|r| \leq n/2\) (namely, \((q,r) = (s,t)\)). Thus, Exercise 2.6.2(a) is solved in Case 1.

Let us now consider Case 2. In this case, we have \( t > n/2 \). But \( t \in \{0,1,\ldots,n-1\} \), thus \( t \leq n-1 \leq n \). Hence, \(|t-n| = -(t-n) = n-t < n-n/2 = n/2\).

Therefore, \(|t-n| \leq n/2\). Furthermore, \( t-n \in \mathbb{Z}\) (since \( t \in \mathbb{Z} \) and \( n \in \mathbb{Z}\)) and \( s + 1 \in \mathbb{Z}\) (since \( s \in \mathbb{Z}\)). So we have \((s+1,t-n) \in \mathbb{Z} \times \mathbb{Z}\) (since \( s+1 \in \mathbb{Z} \) and \( t-n \in \mathbb{Z}\)) and \( u = (s+1)n + (t-n) \) (since \( s+1 \) \( n \) \((t-n) = sn + t = u\)) and \(|t-n| \leq n/2\).

Hence, there exists a pair \((q,r) \in \mathbb{Z} \times \mathbb{Z}\) such that \( u = qn + r \) and \(|r| \leq n/2\) (namely, \((q,r) = (s+1,t-n)\)). Thus, Exercise 2.6.2(a) is solved in Case 2.

We have now solved Exercise 2.6.2(a) in each of the two Cases 1 and 2. Since these two Cases cover all possibilities, we thus conclude that Exercise 2.6.2(a) always holds.

(b) For example, if \( n = 2 \) and \( u = 5 \), then both \((2,1)\) and \((3,-1)\) are pairs \((q,r) \in \mathbb{Z} \times \mathbb{Z}\) such that \( u = qn + r \) and \(|r| \leq n/2\).

More generally: If \( n = 2k \) for some positive integer \( k \), and if \( u \equiv k \mod n \), then both \((u-k)/n,k)\) and \((u+k)/n,-k)\) are pairs \((q,r) \in \mathbb{Z} \times \mathbb{Z}\) such that \( u = qn + r \) and \(|r| \leq n/2\).

[It is not hard to see that these are the only cases in which the pair \((q,r)\) from Exercise 2.6.2(a) is not unique.]

\(\square\)

6.18. Solution to Exercise 2.6.3

Solution to Exercise 2.6.3(a) Corollary 2.6.9(a) yields that \( u \% n \in \{0,1,\ldots,n-1\} \) and \( u \% n \equiv u \mod n \). From \( u \% n \in \{0,1,\ldots,n-1\} \), we conclude that \( u \% n \) is an integer satisfying \( 0 \leq u \% n \leq n-1 \).

Corollary 2.6.9(a) (applied to \( v \) instead of \( u \)) yields that \( v \% n \in \{0,1,\ldots,n-1\} \) and \( v \% n \equiv v \mod n \). From \( v \% n \in \{0,1,\ldots,n-1\} \), we conclude that \( v \% n \) is an integer satisfying \( 0 \leq v \% n \leq n-1 \).

Corollary 2.6.9(a) (applied to \( u+v \) instead of \( u \)) yields that \( (u+v) \% n \in \{0,1,\ldots,n-1\} \) and \( (u+v) \% n \equiv u+v \mod n \). From \( (u+v) \% n \in \{0,1,\ldots,n-1\} \), we conclude that \( (u+v) \% n \) is an integer satisfying \( 0 \leq (u+v) \% n \leq n-1 \).

Adding the congruences \( u \% n \equiv u \mod n \) and \( v \% n \equiv v \mod n \) together, we obtain \( u \% n + v \% n \equiv u + v \mod n \). Subtracting the congruence \((u+v) \% n \equiv u+v \mod n\) from this congruence, we obtain \( u \% n + v \% n - (u+v) \% n \equiv (u+v) - (u+v) \equiv 0 \mod n \). By Proposition 2.3.3 (applied to \( a = u \% n + v \% n - (u+v) \% n \)), this entails \( n \mid (u \% n + v \% n - (u+v) \% n) \). In other words, there exists an integer \( c \) such that \( u \% n + v \% n - (u+v) \% n = nc \). Consider this \( c \).

Hence,

\[
nc = u \% n \bigg|_{\leq n-1<n} + v \% n \bigg|_{\leq n-1<n} - (u+v) \% n \bigg|_{\geq 0} < n+n-0 = 2n = n \cdot 2.
\]

We can divide this inequality by \( n \) (since \( n \) is positive). We thus obtain \( c < 2 \). Hence, \( c \leq 1 \) (since \( c \) is an integer).

On the other hand,

\[
nc = u \% n \bigg|_{\leq n-1<n} + v \% n \bigg|_{\leq n-1<n} - (u+v) \% n \bigg|_{\geq 0} > 0 + 0 - n = -n = n \cdot (-1).
\]
We can divide this inequality by \( n \) (since \( n \) is positive). We thus obtain \( c > -1 \). Hence, \( c \geq 0 \) (since \( c \) is an integer).

Combining \( c \geq 0 \) with \( c \leq 1 \), we obtain \( c \in \{0, 1\} \) (since \( c \) is an integer). In other words, we have \( nc = n \cdot 0 = 0 \) or \( nc = n \cdot 1 = n \). In other words, \( nc \in \{0, n\} \). Now, recall that \( u\%n + v\%n - (u + v) \%n = nc \in \{0, n\} \). This solves Exercise 2.6.3 (a).

(b) Exercise 2.6.3 (a) yields \( u\%n + v\%n - (u + v) \%n \in \{0, n\} \).

The integer \( n \) is positive and thus nonzero. Corollary 2.6.9 (d) yields \( u = (u / n)n + (u\%n) \). Solving this equation for \( u / n \), we find

\[
u / n = \frac{u - u\%n}{n} \tag{170}
\]

(since \( n \) is nonzero). The same argument (applied to \( v \) instead of \( u \)) yields

\[
v / n = \frac{v - v\%n}{n} \tag{171}
\]

Finally, the same argument that we used to prove (170) can be applied to \( u + v \) instead of \( u \), and thus we obtain

\[
(u + v) / n = \frac{(u + v) - (u + v) \%n}{n} \tag{172}
\]

Now,

\[
\frac{(u + v) / n}{n} = \frac{u / n}{n} - \frac{v / n}{n} = \frac{u - u\%n}{n} - \frac{v - v\%n}{n} = \frac{(u + v) - (u + v) \%n}{n} \tag{by (172)}
\]

\[
= \frac{1}{n} \left( (u + v) - (u + v) \%n - (u - u\%n) - (v - v\%n) \right) = \frac{1}{n} \left( u\%n + v\%n - (u + v) \%n \right)
\]

\[
\in \left\{ \frac{1}{n}, \frac{n}{n} \right\} \quad \text{(since \( u\%n + v\%n - (u + v) \%n \in \{0, n\} \))}
\]

\[
= \{0, 1\} \quad \text{(since \( \frac{1}{n} = 0 \) and \( \frac{n}{n} = 1 \)).}
\]

This solves Exercise 2.6.3 (b).

6.19. Solution to Exercise 2.7.1

Solution to Exercise 2.7.1 Corollary 2.6.9 (b) (applied to \( n = 2 \)) shows that we have \( 2 \mid u \) if and only if \( u\%2 = 0 \). In other words, we have the logical equivalence

\[
(2 \mid u) \iff (u\%2 = 0). \tag{173}
\]
Corollary 2.6.9 (a) (applied to \( n = 2 \)) yields that \( u \% 2 \in \{0,1,\ldots,2-1\} \) and \( u \% 2 \equiv u \mod 2 \). Thus, in particular, \( u \% 2 \in \{0,1,\ldots,2-1\} = \{0,1\} \). Hence, \( u \% 2 \) is either 0 or 1. Thus, the number \( u \% 2 \) is 1 if and only if it is not 0. In other words, we have the equivalence

\[
(u \% 2 = 1) \iff (u \% 2 \neq 0) \tag{174}
\]

Proposition 2.3.3 (applied to \( a = u \) and \( n = 2 \)) shows that \( u \equiv 0 \mod 2 \) if and only if \( 2 \mid u \). In other words, we have the equivalence

\[
(u \equiv 0 \mod 2) \iff (2 \mid u) \tag{175}
\]

(a) We have the following chain of equivalences:

\[
(u \text{ is even}) \iff (u \text{ is divisible by 2}) \quad \text{(by the definition of “even”)}
\iff (2 \mid u) \iff (u \% 2 = 0) \quad \text{(by (173)).}
\]

In other words, \( u \) is even if and only if \( u \% 2 = 0 \). This solves Exercise 2.7.1 (a).

(b) We have the following chain of equivalences:

\[
(u \text{ is odd}) \iff (u \text{ is not divisible by 2}) \quad \text{(by the definition of “odd”)}
\iff (\text{we don’t have } 2 \mid u) \iff (\text{we don’t have } u \% 2 = 0)
\iff (\text{because of the equivalence } (2 \mid u) \iff (u \% 2 = 0))
\iff (u \% 2 \neq 0)
\iff (u \% 2 = 1) \quad \text{(by (174)).}
\]

In other words, \( u \) is odd if and only if \( u \% 2 = 1 \). This solves Exercise 2.7.1 (b).

(c) We have the following chain of equivalences:

\[
(u \text{ is even}) \iff (u \text{ is divisible by 2}) \quad \text{(by the definition of “even”)}
\iff (2 \mid u) \iff (u \equiv 0 \mod 2) \quad \text{(by (175)).}
\]

In other words, \( u \) is even if and only if \( u \equiv 0 \mod 2 \). This solves Exercise 2.7.1 (c).

(d) \( \implies \): Assume that \( u \equiv 1 \mod 2 \). We must prove that \( u \equiv 1 \mod 2 \).

We know that \( u \) is odd. In other words, \( u \% 2 = 1 \) (by Exercise 2.7.1 (b)). But recall that \( u \% 2 \equiv u \mod 2 \). Thus, \( u \equiv u \% 2 = 1 \mod 2 \). This proves the “\( \implies \)” direction of Exercise 2.7.1 (d).

\( \iff \): Assume that \( u \equiv 1 \mod 2 \). We must prove that \( u \) is odd.

We have \( 1 \equiv u \mod 2 \) (since \( u \equiv 1 \mod 2 \) and \( 1 \in \{0,1,\ldots,2-1\} \)). But Corollary 2.6.9 (c) (applied to \( n = 2 \)) says that if \( c \in \{0,1,\ldots,2-1\} \) satisfies \( c \equiv u \mod 2 \), then \( c = u \% 2 \).

Applying this to \( c = 1 \), we find \( 1 = u \% 2 \) (since \( 1 \in \{0,1,\ldots,2-1\} \) and \( 1 \equiv u \mod 2 \)). In other words, \( u \% 2 = 1 \). According to Exercise 2.7.1 (b), this means that \( u \) is odd. This proves the “\( \iff \)” direction of Exercise 2.7.1 (d).

(e) \( \implies \): Assume that \( u \) is odd. We must prove that \( u + 1 \) is even.

We have assumed that \( u \) is odd. According to Exercise 2.7.1 (d), this means that \( u \equiv 1 \mod 2 \). On the other hand, \( 1 \equiv -1 \mod 2 \) (since \( 2 \mid 1 - (-1) \)). Adding these two congruences together, we find \( u + 1 \equiv 1 + (-1) = 0 \mod 2 \).

But Exercise 2.7.1 (e) (applied to \( u + 1 \) instead of \( u \)) shows that \( u + 1 \) is even if and only if \( u + 1 \equiv 0 \mod 2 \). Hence, \( u + 1 \) is even (since \( u + 1 \equiv 0 \mod 2 \)). This proves the “\( \implies \)” direction of Exercise 2.7.1 (e).
\[(u + 1) \text{ is even} \]

We know that \(u + 1\) is even. But Exercise 2.7.1(e) (applied to \(u + 1\) instead of \(u\)) shows that \(u + 1\) is even if and only if \(u + 1 \equiv 0 \pmod{2}\). Hence, \(u + 1 \equiv 0 \pmod{2}\) (since \(u + 1\) is even). On the other hand, \(-1 \equiv 1 \pmod{2}\) (since \(2 \mid (-1) - 1\)). Adding these two congruences together, we obtain \((u + 1) + (-1) \equiv 0 + 1 \equiv 1 \pmod{2}\). In view of \((u + 1) + (-1) = u\), this rewrites as \(u \equiv 1 \pmod{2}\). According to Exercise 2.7.1(d), this means that \(u\) is odd. This proves the \(\Leftarrow\) direction of Exercise 2.7.1(e).

(f) We have the equivalence \((u \text{ is divisible by } 2) \iff (u \text{ is even})\) (by the definition of “even”).

Exercise 2.7.1(e) shows that \(u\) is odd if and only if \(u + 1\) is even. Thus, we have the following chain of equivalences:

\[
(u + 1 \text{ is even}) \\
\iff (u \text{ is odd}) \iff (u \text{ is not divisible by } 2) \text{ (by the definition of “odd”)} \\
\iff (u \text{ is not even})
\]

(because of the equivalence \((u \text{ is divisible by } 2) \iff (u \text{ is even})\)). In other words, \(u + 1\) is even if and only if \(u\) is not. In other words, exactly one of the two numbers \(u\) and \(u + 1\) is even. This solves Exercise 2.7.1(f).

(g) Exercise 2.7.1(f) shows that exactly one of the two numbers \(u\) and \(u + 1\) is even. Thus, in particular, at least one of these two numbers is even. Hence, the product \(u \cdot (u + 1)\) has at least one even factor. But a product of any even integer with any integer is even. Hence, a product that has at least one even factor is always even. Thus, \(u \cdot (u + 1)\) is even (since \(u \equiv 1 \pmod{2}\) is a product that has at least one even factor). In other words, \(2 \mid u \cdot (u + 1)\). In other words, \(u \equiv 0 \pmod{2}\). This solves Exercise 2.7.1(g).

(h) We have \(u^2 - (-u) = u^2 + u = u \cdot (u + 1) \equiv 0 \pmod{2}\) (by Exercise 2.7.1(g)). In other words, \(2 \mid u^2 - (-u)\). In other words, \(u^2 \equiv -u \equiv u \pmod{2}\). This solves Exercise 2.7.1(h).

(i) Exercise 2.6.1 (applied to \(n = 2\)) shows that \(u \equiv v \pmod{2}\) if and only if \(u \equiv v \equiv 0\).

We are in one of the following four cases:

\begin{itemize}
  \item \text{Case 1}: \ We have \(u \equiv 0\) and \(v \equiv 0\).
  \item \text{Case 2}: \ We have \(u \equiv 0\) and \(v \equiv 0\).
  \item \text{Case 3}: \ We have \(u \equiv 0\) and \(v \equiv 0\).
  \item \text{Case 4}: \ We have \(u \equiv 0\) and \(v \equiv 0\).
\end{itemize}

Let us first consider Case 1. In this case, we have \(u \equiv 0\) and \(v \equiv 0\). Thus, \(u \equiv v \equiv 0\) and therefore \(u \equiv v \pmod{2}\) (since we know that \(u \equiv v \pmod{2}\) if and only if \(u \equiv v \equiv 0\)). But recall that \(u \equiv 0\). Equivalently, \(u\) is even (because of Exercise 2.7.1(a)). Similarly, from \(v \equiv 0\), we conclude that \(v\) is even. Thus, \(u\) and \(v\) are either both odd or both even (namely, they are both even).

Thus, \(u \equiv v \pmod{2}\) holds if and only if \(u\) and \(v\) are either both odd or both even (because both statements “\(u \equiv v \pmod{2}\)” and “\(u\) and \(v\) are either both odd or both even” hold). Hence, Exercise 2.7.1(i) is solved in Case 1.

---

145 Proof. We must prove that if \(a\) is an even integer, and if \(b\) is an integer, then the product \(ab\) is even.

So let \(a\) be an even integer, and let \(b\) be an integer. Then, \(a\) is even; in other words, \(2 \mid a\) (by the definition of “even”). But \(a \mid ab\). Hence, \(2 \mid a \mid ab\); in other words, \(ab\) is even. Qed.
Let us now consider Case 2. In this case, we have $u \% 2 = 0$ and $v \% 2 \neq 0$. Thus, $u \% 2 = 0 \neq v \% 2$. In other words, “$u \% 2 = v \% 2$” is false. Thus, “$u \equiv v \mod 2$” is false as well (since we know that $u \equiv v \mod 2$ if and only if $u \% 2 = v \% 2$). But recall that $u \% 2 = 0$. Equivalently, $u$ is even (because of Exercise 2.7.1(a)). Hence, $u$ is not odd. Thus, $u$ and $v$ are not both odd. Also, Exercise 2.7.1(a) (applied to $v$ instead of $u$) shows that $v$ is even if and only if $v \% 2 = 0$. Since we don’t have $v \% 2 = 0$ (because $v \% 2 \neq 0$), we thus conclude that $v$ is not even. Thus, $u$ and $v$ are not both even.

So $u$ and $v$ are neither both odd nor both even. In other words, the statement “$u$ and $v$ are either both odd or both even” is false.

Thus, $u \equiv v \mod 2$ holds if and only if $u$ and $v$ are either both odd or both even (because both statements “$u \equiv v \mod 2$” and “$u$ and $v$ are either both odd or both even” are false).

Hence, Exercise 2.7.1(i) is solved in Case 2.

Case 3 is analogous to Case 2 (it differs from Case 2 only in that $u$ and $v$ trade places).

Let us finally consider Case 4. In this case, we have $u \% 2 \neq 0$ and $v \% 2 \neq 0$. By (176), we have the logical equivalence $(u \text{ is odd}) \iff (u \% 2 \neq 0)$. Hence, $u$ is odd (since $u \% 2 \neq 0$).

Similarly, $v$ is odd. Thus, $u$ and $v$ are both odd. Thus, $u$ and $v$ are either both odd or both even (namely, they are both odd). Moreover, we know that $u$ is odd; equivalently, $u \% 2 = 1$ (by Exercise 2.7.1(b)). Similarly, $v \% 2 = 1$. Hence, $u \% 2 = 1 = v \% 2$. Therefore, $u \equiv v \mod 2$ (since we know that $u \equiv v \mod 2$ if and only if $u \% 2 = v \% 2$).

Thus, $u \equiv v \mod 2$ holds if and only if $u$ and $v$ are either both odd or both even (because both statements “$u \equiv v \mod 2$” and “$u$ and $v$ are either both odd or both even” hold).

Hence, Exercise 2.7.1(i) is solved in Case 4.

We have now solved Exercise 2.7.1(i) in all four Cases 1, 2, 3 and 4. Hence, Exercise 2.7.1(i) is solved.

6.20. Solution to Exercise 2.7.2

Solution to Exercise 2.7.2 (a) Let $u$ be an even integer. Thus, $u$ is even. In other words, $u$ is divisible by 2. In other words, there exists some integer $c$ such that $u = 2c$. Consider this $c$.

From $u = 2c$, we obtain $u^2 = (2c)^2 = 4c^2$, which is clearly divisible by 4. So we have $4 \mid u^2 = u^2 - 0$. In other words, $u^2 \equiv 0 \mod 4$. This solves Exercise 2.7.2(a).

(b) Let $u$ be an odd integer. Thus, $u$ is odd. Equivalently, $u \equiv 1 \mod 2$ (by Exercise 2.7.1(d)). In other words, $2 \mid u - 1$. In other words, there exists some integer $c$ such that $u - 1 = 2c$. Consider this $c$.

From $u - 1 = 2c$, we obtain $u = 2c + 1$ and thus $u^2 = (2c + 1)^2 = 4c^2 + 4c + 1$. Hence, $u^2 - 1 = 4c^2 + 4c = 4(c^2 + c)$, which is clearly divisible by 4. So we have $4 \mid u^2 - 1$. In other words, $u^2 \equiv 1 \mod 4$. This solves Exercise 2.7.2(b).

(c) Let $x$ and $y$ be two integers such that $x^2 + y^2 \equiv 3 \mod 4$. We shall derive a contradiction.

Recall that an integer is always either even or odd. Thus, $x$ is either even or odd. Similarly, $y$ is either even or odd. Thus, we are in one of the following four cases:

Case 1: The integer $x$ is even, and the integer $y$ is even.
Case 2: The integer $x$ is even, and the integer $y$ is odd.
Case 3: The integer $x$ is odd, and the integer $y$ is even.
Case 4: The integer $x$ is odd, and the integer $y$ is odd.

\[\text{because an integer is either even or odd (but not both at the same time)}\]
Let us first consider Case 1. In this case, the integer \( x \) is even, and the integer \( y \) is even. Hence, Exercise 2.7.2 (a) (applied to \( u = x \)) yields \( x^2 \equiv 0 \mod 4 \) (since \( x \) is even). Also, Exercise 2.7.2 (a) (applied to \( u = y \)) yields \( y^2 \equiv 0 \mod 4 \) (since \( y \) is even). Thus, \( x^2 + y^2 \equiv 0 + 0 = 0 \mod 4 \). Hence, 0 \( \equiv 3 \) \mod 4. But Exercise 2.6.1 (applied to \( n = 4, u = 0 \) and \( v = 3 \)) shows that 0 \( \equiv 3 \mod 4 \) if and only if \( 0 \equiv 3 \mod 4 \). Hence, 0 \( \equiv 3 \mod 4 \) (since 0 \( \equiv 3 \mod 4 \)). This contradicts the fact that 0 \( \equiv 0 \neq 3 \equiv 3 \mod 4 \). Hence, we have obtained a contradiction in Case 1.

Let us next consider Case 2. In this case, the integer \( x \) is even, and the integer \( y \) is odd. Hence, Exercise 2.7.2 (a) (applied to \( u = x \)) yields \( x^2 \equiv 0 \mod 4 \) (since \( x \) is even). Also, Exercise 2.7.2 (b) (applied to \( u = y \)) yields \( y^2 \equiv 1 \mod 4 \) (since \( y \) is odd). Thus, \( x^2 + y^2 \equiv 0 + 1 = 1 \mod 4 \). Hence, 1 \( \equiv 3 \mod 4 \). But Exercise 2.6.1 (applied to \( n = 4, u = 1 \) and \( v = 3 \)) shows that 1 \( \equiv 3 \mod 4 \) if and only if \( 1 \equiv 3 \mod 4 \). Hence, 1 \( \equiv 3 \mod 4 \) (since 1 \( \equiv 3 \mod 4 \)). This contradicts the fact that 1 \( \equiv 1 \neq 3 \equiv 3 \mod 4 \). Hence, we have obtained a contradiction in Case 2.

The arguments in Cases 3 and 4 are completely analogous (in Case 3, we obtain \( x^2 + y^2 \equiv 1 \mod 4 \) again, whereas in Case 4 we obtain \( x^2 + y^2 \equiv 2 \mod 4 \)). Thus, we have obtained a contradiction in each of the four Cases 1, 2, 3 and 4. Hence, we always have a contradiction.

Now, forget that we fixed \( x \) and \( y \). We thus have obtained a contradiction whenever \( x \) and \( y \) are two integers such that \( x^2 + y^2 \equiv 3 \mod 4 \). Thus, there are no such two integers. This solves Exercise 2.7.2 (c).

(d) The solution of Exercise 2.7.2 (d) is very similar to the above solution of Exercise 2.7.2 (c) (indeed, we have to consider the same four cases, but this time we don’t get a contradiction in Case 4) and is left to the reader. \( \square \)

6.21. Solution to Exercise 2.9.1

Solution to Exercise 2.9.1 Assume that \( \{b_1, b_2, \ldots, b_k\} = \{c_1, c_2, \ldots, c_l\} \).

Let \( a \) be an integer. Then, \( a \) is a common divisor of \( b_1, b_2, \ldots, b_k \) if and only if \( a \) satisfies \( (a \mid b_i \text{ for all } i \in \{1, 2, \ldots, k\}) \) (by the definition of “common divisor”). Hence, we have the following chain of equivalences:

\[
\begin{align*}
(a \text{ is a common divisor of } b_1, b_2, \ldots, b_k) & \iff (a \mid b_i \text{ for all } i \in \{1, 2, \ldots, k\}) \\
& \iff (a \mid b_1 \text{ and } a \mid b_2 \text{ and } \cdots \text{ and } a \mid b_k) \\
& \iff (a \mid u \text{ for each } u \in \{b_1, b_2, \ldots, b_k\}).
\end{align*}
\]

But \( \text{Div} \ (b_1, b_2, \ldots, b_k) \) is the set of all common divisors of \( b_1, b_2, \ldots, b_k \). Hence, we have \( a \in \text{Div} \ (b_1, b_2, \ldots, b_k) \) if and only if \( a \) is a common divisor of \( b_1, b_2, \ldots, b_k \). Thus, we have the following chain of equivalences:

\[
\begin{align*}
(a \in \text{Div} \ (b_1, b_2, \ldots, b_k)) & \iff (a \text{ is a common divisor of } b_1, b_2, \ldots, b_k) \\
& \iff (a \mid u \text{ for each } u \in \{b_1, b_2, \ldots, b_k\}).
\end{align*}
\] (177)
The same argument (applied to \( \ell \) and \((c_1, c_2, \ldots, c_\ell)\) instead of \(k\) and \((b_1, b_2, \ldots, b_\ell)\)) yields the equivalence

\[
(a \in \text{Div} (c_1, c_2, \ldots, c_\ell)) \iff (a \mid u \text{ for each } u \in \{c_1, c_2, \ldots, c_\ell\}).
\] (178)

Now, we have the following chain of equivalences:

\[
(a \in \text{Div} (b_1, b_2, \ldots, b_\ell))
\iff \left( \begin{array}{l}
\text{for each } u \in \{b_1, b_2, \ldots, b_\ell\} \text{ (by (177))}
\end{array} \right)
\iff (a \mid u \text{ for each } u \in \{c_1, c_2, \ldots, c_\ell\})
\iff (a \in \text{Div} (c_1, c_2, \ldots, c_\ell)) \quad \text{(by (178))}.
\]

Now, forget that we fixed \(a\). We thus have proven the equivalence

\[
(a \in \text{Div} (b_1, b_2, \ldots, b_\ell)) \iff (a \in \text{Div} (c_1, c_2, \ldots, c_\ell))
\]

for each integer \(a\). In other words, an integer \(a\) belongs to the set \(\text{Div} (b_1, b_2, \ldots, b_\ell)\) if and only if it belongs to the set \(\text{Div} (c_1, c_2, \ldots, c_\ell)\). In other words, the two sets \(\text{Div} (b_1, b_2, \ldots, b_\ell)\) and \(\text{Div} (c_1, c_2, \ldots, c_\ell)\) contain the exact same integers. Since both of these sets consist of integers only, this entails that these two sets are equal. In other words, \(\text{Div} (b_1, b_2, \ldots, b_\ell) = \text{Div} (c_1, c_2, \ldots, c_\ell)\). This solves Exercise \ref{ex:2.9.1} \hfill \Box

6.22. Solution to Exercise \ref{ex:2.9.2}

Solution to Exercise \ref{ex:2.9.2} Assume that \(\{b_1, b_2, \ldots, b_\ell\} = \{c_1, c_2, \ldots, c_\ell\}\). Then, Exercise \ref{ex:2.9.1} yields \(\text{Div} (b_1, b_2, \ldots, b_\ell) = \text{Div} (c_1, c_2, \ldots, c_\ell)\). Hence, Lemma \ref{lem:2.9.9} yields \(\gcd (b_1, b_2, \ldots, b_\ell) = \gcd (c_1, c_2, \ldots, c_\ell)\). This solves Exercise \ref{ex:2.9.2} \hfill \Box

6.23. Solution to Exercise \ref{ex:2.9.3}

Solution to Exercise \ref{ex:2.9.3} \hspace{1em} (a) Let \(a \in \mathbb{N}\) and \(b \in \mathbb{N}\) be such that \(b \geq a\).

We have \(b - a \geq 0\) (since \(b \geq a\)), hence \(b - a \in \mathbb{N}\). Thus, \(u^{b-a}\) is an integer. We have

\[
\left(u^b - 1\right) - \left(u^a - 1\right) = u^b - u^a = u^{(b-a)+a} - u^a = u^{b-a}u^a - u^a = \left(u^{b-a} - 1\right)u^a.
\]

Thus, \(u^{b-a} - 1 \mid \left(u^b - 1\right) - \left(u^a - 1\right)\) (since \(u^a\) is an integer). In other words, \(u^b - 1 \equiv u^a - 1 \mod u^{b-a} - 1\). This solves Exercise \ref{ex:2.9.3} \hspace{1em} (a).

(b) The following argument will imitate our proof of Lemma \ref{lem:2.9.12} above.

We use strong induction on \(a + b\). Thus, we fix an \(n \in \mathbb{N}\), and assume (as induction hypothesis) that Exercise \ref{ex:2.9.3} \hspace{1em} (b) holds whenever \(a + b < n\). We must now prove that Exercise \ref{ex:2.9.3} \hspace{1em} (b) holds whenever \(a + b = n\).

We have assumed that Exercise \ref{ex:2.9.3} \hspace{1em} (b) holds whenever \(a + b < n\). In other words, the following statement holds:
Statement 1: Let \( a \in \mathbb{N} \) and \( b \in \mathbb{N} \) be such that \( a + b < n \). Then, \( \gcd(a - 1, b - 1) = \left| u^{\gcd(a,b)} - 1 \right| \).

Now, we must prove that Exercise 2.9.3 (b) holds whenever \( a + b = n \). Let us first prove this in the case when \( b \geq a \):

Statement 2: Let \( a \in \mathbb{N} \) and \( b \in \mathbb{N} \) be such that \( a + b = n \) and \( b \geq a \). Then, \( \gcd(a - 1, b - 1) = \left| u^{\gcd(a,b)} - 1 \right| \).

[Proof of Statement 2: We are in one of the following two cases:
Case 1: We have \( a = 0 \).
Case 2: We have \( a \neq 0 \).
Let us first consider Case 1. In this case, we have \( a = 0 \). Hence, \( a^b = u^0 = 1 \) and thus \( u^a - 1 = 0 \). Thus,

\[
gcd\left(u^a - 1, u^b - 1\right) = \gcd\left(0, u^b - 1\right) = \gcd\left(u^b - 1, 0\right)
\]

(since Proposition 2.9.7 (a) (applied to \( u^b - 1 \) instead of \( a \)) yields \( \gcd(u^b - 1, 0) = \gcd(u^b - 1) = \left| u^b - 1 \right| \)).

But Proposition 2.9.7 (b) yields \( \gcd(a, b) = \gcd\left(b, -\frac{a}{b}\right) = \gcd(b, 0) \). Now, Proposition 2.9.7 (a) (applied to \( b \) instead of \( a \)) yields \( \gcd(b, 0) = \gcd(b) = \left| b \right| \) (since \( b \) is nonnegative). Hence, \( \left| u^{\gcd(a,b)} - 1 \right| = \left| u^b - 1 \right| \). Comparing this equality with (179), we obtain \( \gcd(a - 1, b - 1) = \left| u^{\gcd(a,b)} - 1 \right| \). Thus, Statement 2 is proven in Case 1.

Let us next consider Case 2. In this case, we have \( a \neq 0 \). Hence, \( a > 0 \) (since \( a \in \mathbb{N} \)), so that \( a + b > b \). Hence, \( b < a + b = n \).

From \( b \geq a \), we obtain \( b - a \in \mathbb{N} \). Moreover, \( a \in \mathbb{N} \) and \( b - a \in \mathbb{N} \) satisfy \( a + (b - a) = b < n \). Therefore, we can apply Statement 1 to \( b - a \) instead of \( b \). Thus we obtain that \( \gcd(a - 1, b - 1) = \left| u^{\gcd(a,b-a)} - 1 \right| \).

But Proposition 2.9.7 (c) (applied to \( u = -1 \)) yields \( \gcd(a, (-1) a + b) = \gcd(a, b) \). This rewrites as \( \gcd(a, b - a) = \gcd(a, b) \) (since \( (-1) a + b = b - a \)).

Recall that \( b - a \in \mathbb{N} \). Also, \( b \geq b - a \) (since \( a \geq 0 \)). Hence, Exercise 2.9.3 (a) (applied to \( b - a \) instead of \( a \)) yields \( u^{b - 1} \equiv u^{b-a - 1} \mod u^{b - (b - a) - 1} \). Since \( b - (b - a) = a \), this rewrites as \( u^{b - 1} \equiv u^{b-a - 1} \mod u^{a - 1} \). Hence, Proposition 2.9.7 (d) (applied to \( u^a - 1, u^b - 1 \) and \( u^{b-a} - 1 \) instead of \( a, b \) and \( c \)) yields

\[
gcd\left(u^a - 1, u^b - 1\right) = \gcd\left(u^a - 1, u^{b-a} - 1\right) = \left| u^{\gcd(a,b-a)} - 1 \right| = \left| u^{\gcd(a,b)} - 1 \right|
\]

(since \( \gcd(a, b - a) = \gcd(a, b) \)).
Thus, Statement 2 is proven in Case 2.

We have now proven Statement 2 in both Cases 1 and 2. Hence, Statement 2 is always proven.

Now, we can prove that Exercise 2.9.3 (b) holds whenever \( a + b = n \):

Statement 3: Let \( a \in \mathbb{N} \) and \( b \in \mathbb{N} \) be such that \( a + b = n \). Then, \( \gcd(u^a - 1, u^b - 1) = \left| u^{\gcd(a,b)} - 1 \right| \).

[Proof of Statement 3: We are in one of the following two cases:

Case 1: We have \( b \geq a \).

Case 2: We have \( b < a \).

Let us first consider Case 1. In this case, we have \( b \geq a \). Hence, Statement 2 shows that \( \gcd(u^a - 1, u^b - 1) = \left| u^{\gcd(a,b)} - 1 \right| \). Thus, Statement 3 is proven in Case 1.

Let us next consider Case 2. In this case, we have \( b < a \). Hence, \( a > b \), so that \( a \geq b \).

This shows that we can apply Statement 2 to \( b \) and \( a \) instead of \( a \) and \( b \). Thus we obtain \( \gcd(u^b - 1, u^a - 1) = \left| u^{\gcd(b,a)} - 1 \right| \). But Proposition 2.9.7 (b) yields \( \gcd(a, b) = \gcd(b, a) \).

Moreover, Proposition 2.9.7 (b) (applied to \( u^a - 1 \) and \( u^b - 1 \) instead of \( a \) and \( b \)) yields

\[
\gcd(u^a - 1, u^b - 1) = \gcd(u^b - 1, u^a - 1) = \left| u^{\gcd(b,a)} - 1 \right| = \left| u^{\gcd(a,b)} - 1 \right|
\]

(since \( \gcd(b,a) = \gcd(a,b) \)). Thus, Statement 3 is proven in Case 2.

We have now proven Statement 3 in both Cases 1 and 2. Hence, Statement 3 is always proven.

By proving Statement 3, we have shown that Exercise 2.9.3 (b) holds whenever \( a + b = n \). This completes the induction step. Thus, Exercise 2.9.3 (b) is proven by strong induction.

[See also https://math.stackexchange.com/questions/7473/ for various solutions of Exercise 2.9.3 (b).]

\[ \square \]

6.24. Solution to Exercise 2.9.4

Solution to Exercise 2.9.4 Proposition 2.9.7 (f) (applied to \( a = a_1 \) and \( b = a_2 \)) yields that we have \( \gcd(a_1, a_2) \mid a_1 \) and \( \gcd(a_1, a_2) \mid a_2 \). Thus, \( \gcd(a_1, a_2) \mid a_1 \mid b_1 \) and \( \gcd(a_1, a_2) \mid a_2 \mid b_2 \).

So we know that \( \gcd(a_1, a_2) \mid b_1 \) and \( \gcd(a_1, a_2) \mid b_2 \). Hence, Lemma 2.9.15 (applied to \( m = \gcd(a_1, a_2) \), \( a = b_1 \) and \( b = b_2 \)) yields \( \gcd(a_1, a_2) \mid \gcd(b_1, b_2) \). This solves Exercise 2.9.4

\[ \square \]

6.25. Solution to Exercise 2.9.5

Solution to Exercise 2.9.5 (a) If \( b \geq 0 \), then \( |b| = b \). Hence, if \( b \geq 0 \), then Exercise 2.9.5 (a) holds (since \( \gcd(a, |b|) = \gcd(a, b) \)). Thus, for the rest of this solution to Exercise 2.9.5 (a), we WLOG assume that we don’t have \( b \geq 0 \). Hence, we have \( b < 0 \). Thus, \( |b| = -b \) and
therefore \( \gcd \left( \frac{a}{|a|}, \frac{|b|}{b} \right) = \gcd (a, -b) = \gcd (a, b) \) (by Proposition 2.9.7 (h)). This solves Exercise 2.9.5 (a).

(b) If \( a \geq 0 \), then \( |a| = a \). Hence, if \( a \geq 0 \), then Exercise 2.9.5 (b) holds (since \( \gcd \left( \frac{|a|}{a}, \frac{b}{b} \right) = \gcd (a, b) \)). Thus, for the rest of this solution to Exercise 2.9.5 (b), we WLOG assume that we don’t have \( a \geq 0 \). Hence, we have \( a < 0 \). Thus, \( |a| = -a \) and therefore \( \gcd \left( \frac{|a|}{a}, \frac{b}{b} \right) = \gcd (-a, b) = \gcd (a, b) \) (by Proposition 2.9.7 (g)). This solves Exercise 2.9.5 (b).

(c) Exercise 2.9.5 (b) (applied to \( |b| \) instead of \( b \)) yields \( \gcd (|a|, |b|) = \gcd (a, |b|) = \gcd (a, b) \) (by Exercise 2.9.5 (a)). This solves Exercise 2.9.5 (c).

6.26. Solution to Exercise 2.9.6

**Solution to Exercise 2.9.6** Corollary 2.9.19 yields

\[
\gcd (sa, sb) = |s| \gcd (a, b). \tag{180}
\]

But \( \gcd (a, b) \) is a nonnegative integer (by the definition of \( \gcd (a, b) \)). The equality \( \text{5} \) (applied to \( x = s \) and \( y = \gcd (a, b) \)) yields

\[
|s \gcd (a, b)| = |s| \cdot \gcd (a, b) = |s| \gcd (a, b) \quad \text{(since } \gcd (a, b) \text{ is nonnegative)}
\]

\[
= \gcd (sa, sb) \quad \text{(by (180))}. \tag{181}
\]

Now, Theorem 2.9.20 (d) (applied to \( 3 \) and \( (a, b, c) \) instead of \( k \) and \( (b_1, b_2, \ldots, b_k) \)) yields

\[
\gcd (a, b, c) = \gcd (\gcd (a, b), c). \tag{182}
\]

The same argument (applied to \( sa, sb, sc \) instead of \( a, b, c \)) yields

\[
\gcd (sa, sb, sc)
\]

\[
= \gcd \left( \frac{\gcd (sa, sb, sc)}{|s \gcd (a, b)|} \right) = \gcd \left( \gcd (|s \gcd (a, b)|, sc) \right)
\]

\[
= \gcd (s \gcd (a, b), sc) \quad \text{(by Exercise 2.9.5 (b), applied to } s \gcd (a, b) \text{ and } sc \text{ instead of } a \text{ and } b)
\]

\[
= |s| \gcd (\gcd (a, b), c) \quad \text{(by Corollary 2.9.19, applied to } \gcd (a, b) \text{ and } c \text{ instead of } a \text{ and } b)
\]

\[
= |s| \gcd (a, b, c).
\]

This solves Exercise 2.9.6. \( \square \)
6.27. Solution to Exercise 2.9.7

Solution to Exercise 2.9.7  We shall show that

\[ \gcd (s a_1, s a_2, \ldots, s a_i) = |s| \gcd (a_1, a_2, \ldots, a_i) \] (183)

for each \( i \in \{0, 1, \ldots, k\} \).

[Proof of (183): We proceed by induction on \( i \):

Induction base: Proposition 2.9.7(j) shows that the greatest common divisor of the empty list of integers is \( \gcd () = 0 \). Now, comparing \( \gcd (s a_1, s a_2, \ldots, s a_0) = \gcd () = 0 \) with \( |s| \gcd (a_1, a_2, \ldots, a_0) = 0 \), we obtain \( \gcd (s a_1, s a_2, \ldots, s a_0) = |s| \gcd (a_1, a_2, \ldots, a_0) \). In other words, (183) holds for \( i = 0 \). This completes the induction base.

Induction step: Let \( j \in \{1, 2, \ldots, k\} \). Assume that (183) holds for \( i = j - 1 \). We must prove that (183) holds for \( i = j \).

We have assumed that (183) holds for \( i = j - 1 \). In other words, we have

\[ \gcd (s a_1, s a_2, \ldots, s a_{j-1}) = |s| \gcd (a_1, a_2, \ldots, a_{j-1}) \] (184)

But \( \gcd (a_1, a_2, \ldots, a_{j-1}) \) is a nonnegative integer (by the definition of \( \gcd (a_1, a_2, \ldots, a_{j-1}) \)). The equality (3) (applied to \( x = s \) and \( y = \gcd (a_1, a_2, \ldots, a_{j-1}) \)) yields

\[ |s \gcd (a_1, a_2, \ldots, a_{j-1})| = |s| \cdot |\gcd (a_1, a_2, \ldots, a_{j-1})| = |s| \gcd (a_1, a_2, \ldots, a_{j-1}) \]

\[ = \gcd (a_1, a_2, \ldots, a_{j-1}) \quad \text{(since \( \gcd (a_1, a_2, \ldots, a_{j-1}) \) is nonnegative)} \]

\[ = \gcd (s a_1, s a_2, \ldots, s a_{j-1}) \quad \text{(by (184))}. \] (185)

Theorem 2.9.20(d) (applied to \( j \) and \( a_i \) instead of \( k \) and \( b_i \)) yields

\[ \gcd (a_1, a_2, \ldots, a_j) = \gcd (\gcd (a_1, a_2, \ldots, a_{j-1}), a_j) \] (186)

Theorem 2.9.20(d) (applied to \( j \) and \( s a_i \) instead of \( k \) and \( b_i \)) yields

\[ \gcd (s a_1, s a_2, \ldots, s a_j) \]

\[ = \gcd \left( \begin{array}{c} \gcd (s a_1, s a_2, \ldots, s a_{j-1}), s a_j \\ = |s \gcd (a_1, a_2, \ldots, a_{j-1})| \quad \text{(by (185))} \end{array} \right) \]

\[ = \gcd (|s \gcd (a_1, a_2, \ldots, a_{j-1})|, s a_j) \]

\[ = \gcd (s \gcd (a_1, a_2, \ldots, a_{j-1}), s a_j) \quad \text{(by Exercise 2.9.5(b), applied to \( s \gcd (a_1, a_2, \ldots, a_{j-1}) \) and \( s a_j \) instead of \( a \) and \( b \))} \]

\[ = |s| \gcd (\gcd (a_1, a_2, \ldots, a_{j-1}), a_j) \]

\[ = \gcd (a_1, a_2, \ldots, a_j) \quad \text{(by Corollary 2.9.19, applied to \( \gcd (a_1, a_2, \ldots, a_{j-1}) \) and \( a_j \) instead of \( a \) and \( b \))} \]

\[ = |s| \gcd (a_1, a_2, \ldots, a_j). \]
In other words, (183) holds for \( i = j \). This completes the induction step. Thus, (183) is proven.

Now, (183) (applied to \( i = k \)) yields \( \gcd(sa_1, sa_2, \ldots, sa_k) = |s| \gcd(a_1, a_2, \ldots, a_k) \). This solves Exercise 2.9.7.

### 6.28. Solution to Exercise 2.10.1

**Solution to Exercise 2.10.1** (a) We have \( 1 \mid a \) (since \( a = 1 \cdot a \)). Thus, Proposition 2.9.7 (i) (applied to 1 and \( a \) instead of \( a \) and \( b \)) yields \( \gcd(1, a) = |1| = 1 \). In other words, \( 1 \perp a \) (by the definition of “coprime”). This solves Exercise 2.10.2 (a).

(b) Proposition 2.9.7 (a) yields \( \gcd(a, 0) = \gcd(a) = |a| \). But Proposition 2.9.7 (b) (applied to \( b = 0 \)) yields \( \gcd(a, 0) = \gcd(0, a) \). Hence, \( \gcd(0, a) = \gcd(a, 0) = |a| \). Now, we have the following chain of logical equivalences:

\[
(0 \perp a) \iff (\gcd(0, a) = 1) \iff (|a| = 1) \quad (\text{since } \gcd(0, a) = |a|).
\]

In other words, we have \( 0 \perp a \) if and only if \( |a| = 1 \). This solves Exercise 2.10.1 (b).

### 6.29. Solution to Exercise 2.10.2

**Solution to Exercise 2.10.2** Let us prove that

\[ a_1 a_2 \cdots a_i \perp c \quad \text{for each } i \in \{0, 1, \ldots, k\} \tag{187} \]

**Proof of (187):** We shall prove (187) by induction on \( i \):

**Induction base:** Exercise 2.10.1 (a) (applied to \( a = c \)) yields \( 1 \perp c \). Now, \( a_1 a_2 \cdots a_0 = (\text{empty product}) = 1 \perp c \). Hence, (187) holds for \( i = 0 \). This completes the induction base.

**Induction step:** Let \( j \in \{1, 2, \ldots, k\} \). Assume that (187) holds for \( i = j - 1 \). We must now prove that (187) holds for \( i = j \).

We have assumed that (187) holds for \( i = j - 1 \). In other words, \( a_1 a_2 \cdots a_{j-1} \perp c \).

We have assumed that each \( i \in \{1, 2, \ldots, k\} \) satisfies \( a_i \perp c \). Applying this to \( i = j \), we find \( a_j \perp c \).

Now we know that \( a_1 a_2 \cdots a_{j-1} \perp c \) and \( a_j \perp c \). Hence, Theorem 2.10.9 (applied to \( a = a_1 a_2 \cdots a_{j-1} \) and \( b = a_j \)) yields \( (a_1 a_2 \cdots a_{j-1}) a_j \perp c \). In other words, \( a_1 a_2 \cdots a_j \perp c \) (since \( a_1 a_2 \cdots a_j = (a_1 a_2 \cdots a_{j-1}) a_j \)). In other words, (187) holds for \( i = j \). This completes the induction step. Thus, (187) is proven by induction.

Now, we can apply (187) to \( i = k \). We thus obtain \( a_1 a_2 \cdots a_k \perp c \). This proves Exercise 2.10.2.

### 6.30. Solution to Exercise 2.10.3

**Solution to Exercise 2.10.3** We assumed that the integers \( b_1, b_2, \ldots, b_k \) are mutually coprime. In other words, we have

\[ b_i \perp b_j \quad \text{for all } i, j \in \{1, 2, \ldots, k\} \text{ satisfying } i \neq j. \tag{188} \]
Let us prove that
\[ b_1 b_2 \cdots b_i \mid c \quad \text{for each } i \in \{0, 1, \ldots, k\}. \] (189)

**Proof of (189):** We shall prove (189) by induction on \( i \):
- **Induction base:** We have \( b_1 b_2 \cdots b_0 = (\text{empty product}) = 1 \mid c \). Hence, (189) holds for \( i = 0 \). This completes the induction base.
- **Induction step:** Let \( j \in \{1, 2, \ldots, k\} \). Assume that (189) holds for \( i = j - 1 \). We must now prove that (189) holds for \( i = j \).

We have assumed that (189) holds for \( i = j - 1 \). In other words, \( b_1 b_2 \cdots b_{j-1} \mid c \).

We have assumed that each \( i \in \{1, 2, \ldots, k\} \) satisfies \( b_i \mid c \). Applying this to \( i = j \), we find \( b_j \mid c \).

For each \( i \in \{1, 2, \ldots, j-1\} \), we have \( i \leq j - 1 < j \) and thus \( i \neq j \) and therefore \( b_i \perp b_j \) (by (188)). Hence, Exercise 2.10.3 (applied to \( j - 1, b_j \) and \( (b_1, b_2, \ldots, b_{j-1}) \) instead of \( k, c \) and \( (a_1, a_2, \ldots, a_k) \)) yields \( b_1 b_2 \cdots b_{j-1} \perp b_j \).

Now we know that \( b_1 b_2 \cdots b_{j-1} \mid c \) and \( b_j \mid c \) and \( b_1 b_2 \cdots b_{j-1} \perp b_j \). Hence, Theorem 2.10.7 (applied to \( a = b_1 b_2 \cdots b_{j-1} \) and \( b = b_j \)) yields \( (b_1 b_2 \cdots b_{j-1}) b_j \mid c \). In other words, \( b_1 b_2 \cdots b_j \mid c \) (since \( b_1 b_2 \cdots b_{j-1} = (b_1 b_2 \cdots b_{j-1}) b_j \)). In other words, (189) holds for \( i = j \). This completes the induction step. Thus, (189) is proven by induction.

Now, we can apply (189) to \( i = k \). We thus obtain \( b_1 b_2 \cdots b_k \mid c \). This proves Exercise 2.10.3. \( \square \)

### 6.31. Solution to Exercise 2.10.4

**Solution to Exercise 2.10.4** We have \( a \perp b \). Thus, Exercise 2.10.3 (applied to \( n, b \) and \( \left( \frac{a, a, \ldots, a}{n \text{ times}} \right) \) instead of \( k, c \) and \( (a_1, a_2, \ldots, a_k) \)) yields that \( \frac{a a \cdots a}{n \text{ times}} \perp b \). In other words, \( a^n \perp b \).

According to Proposition 2.10.4 (applied to \( a^n \) instead of \( a \)), we have \( a^n \perp b \) if and only if \( b \perp a^n \). Thus, \( b \perp a^n \) (since \( a^n \perp b \)). Hence, Exercise 2.10.3 (applied to \( m, a^n \) and \( \left( \frac{b, b, \ldots, b}{m \text{ times}} \right) \) instead of \( k, c \) and \( (a_1, a_2, \ldots, a_k) \)) yields that \( \frac{bb \cdots b}{m \text{ times}} \perp a^n \). In other words, \( b^m \perp a^n \).

According to Proposition 2.10.4 (applied to \( a^n \) and \( b^m \) instead of \( a \) and \( b \)), we have \( a^n \perp b^m \) if and only if \( b^m \perp a^n \). Hence, \( a^n \perp b^m \) (since \( b^m \perp a^n \)). This solves Exercise 2.10.4. \( \square \)

### 6.32. Solution to Exercise 2.10.5

**Solution to Exercise 2.10.5** Exercise 2.10.2 and Exercise 2.10.5 say the same thing: They say that if \( c \) is a fixed integer, then a product of finitely many integers that are coprime to \( c \) will also be coprime to \( c \). The difference between these two exercises is merely how the product is indexed. Thus, deriving Exercise 2.10.5 from Exercise 2.10.2 is merely a matter of bookkeeping. Let us do this bookkeeping:

By assumption, we have \( b_i \perp c \) for each \( i \in I \). (190)
Thus, (191) can be rewritten as
\[ \prod_{i \in I} b_i = \prod_{j \in \{1,2,\ldots,k\}} b_{f(j)} = \prod_{i=1}^{k} b_{f(i)} = b_{f(1)}b_{f(2)}\cdots b_{f(k)}. \]

The map \( f : \{1,2,\ldots,k\} \to I \) is a bijection. Hence, we can substitute \( f(j) \) for \( i \) in the product \( \prod_{i \in I} b_i \). We thus find
\[ \prod_{i \in I} b_i = \prod_{j \in \{1,2,\ldots,k\}} b_{f(j)} = \prod_{i=1}^{k} b_{f(i)} = b_{f(1)}b_{f(2)}\cdots b_{f(k)}. \]

Thus, (191) can be rewritten as \( \prod_{i \in I} b_i \perp c \). This solves Exercise 2.10.5 \( \square \)

### 6.33. Solution to Exercise 2.10.6

**Solution to Exercise 2.10.6** Assume that \( a \perp c \). But Proposition 2.10.4 (applied to \( c \) instead of \( b \)) shows that \( a \perp c \) if and only if \( c \perp a \). Thus, we have \( c \perp a \) (since \( a \perp c \)). In other words, \( c \) is coprime to \( a \). In other words, \( \gcd(c,a) = 1 \) (by the definition of “coprime”).

But \( a \equiv b \mod c \). Hence, Proposition 2.9.7 (e) (applied to \( c, a \) and \( b \) instead of \( a, b \) and \( c \)) yields \( \gcd(c,a) = \gcd(c,b) \). Hence, \( \gcd(c,b) = \gcd(c,a) = 1 \). In other words, \( c \) is coprime to \( b \). In other words, \( c \perp b \). But Proposition 2.10.4 (applied to \( c \) instead of \( a \)) shows that \( c \perp b \) if and only if \( b \perp c \). Hence, \( b \perp c \) (since \( c \perp b \)). This solves Exercise 2.10.6 \( \square \)

### 6.34. Solution to Exercise 2.10.7

**Solution to Exercise 2.10.7** Proposition 2.9.7 (b) (applied to \( b - a \) instead of \( a \)) yields
\[ \gcd(b-a,b) = \gcd\left(\underbrace{b, b-a}_{= 1b+(-a)}\right) = \gcd(b, 1b + (-a)) \]
\[ = \gcd(b, -a) \quad \text{(by Proposition 2.9.7 (a), applied to } b, -a \text{ and } 1 \text{ instead of } a, b \text{ and } u) \]
\[ = \gcd(b, a) \quad \text{(by Proposition 2.9.7 (h), applied to } b \text{ and } a \text{ instead of } a \text{ and } u) \]
\[ = \gcd(a, b) \quad \text{(by Proposition 2.9.7 (b)).} \]

Now, we have the following chain of logical equivalences:
\[ (b-a \perp b) \iff \left( \gcd(b-a,b) = 1 \right) \quad \text{(by the definition of “coprime”) } \]
\[ \iff \left( \gcd(a,b) = 1 \right) \iff (a \perp b) \quad \text{(by the definition of “coprime”).} \]

In other words, \( b-a \perp b \) holds if and only if \( a \perp b \). This solves Exercise 2.10.7 \( \square \)
6.35. Solution to Exercise 2.10.8

Solution to Exercise 2.10.8  Proposition 2.10.12 yields $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$. Thus, we need to prove that

$$\frac{n(n+1)}{2} | 1^d + 2^d + \cdots + n^d.$$ 

This is equivalent to

$$n(n+1) | 2 \left( 1^d + 2^d + \cdots + n^d \right) \tag{192}$$ 

(by Exercise 2.2.3, applied to $a = \frac{n(n+1)}{2}$, $b = 1^d + 2^d + \cdots + n^d$ and $c = 2$). Hence, it suffices to prove (192).

In order to prove (192), it suffices to show that

$$n | 2 \left( 1^d + 2^d + \cdots + n^d \right) \quad \text{and} \quad n + 1 | 2 \left( 1^d + 2^d + \cdots + n^d \right). \tag{193} \tag{194}$$

Indeed, the integers $n$ and $n+1$ are coprime (by Example 2.10.2 (c), applied to $a = n$); in other words, $n \perp n+1$. Hence, if we can prove (193) and (194), then Theorem 2.10.7 (applied to $a = n$, $b = n + 1$ and $c = 2$) will yield $n(n+1) | 2 \left( 1^d + 2^d + \cdots + n^d \right)$; this will prove (192) and therefore complete our solution.

We shall prove (194) first:

Proof of (194): We have

$$2 \left( 1^d + 2^d + \cdots + n^d \right) = \left( 1^d + 2^d + \cdots + n^d \right) + \left( 1^d + 2^d + \cdots + n^d \right)$$

$$= \left( 1^d + 2^d + \cdots + n^d \right) + \left( n^d + (n-1)^d + \cdots + 1^d \right)$$

$$= \sum_{k=1}^{n} k^d + \sum_{k=1}^{n} (n+1-k)^d$$

$$= \sum_{k=1}^{n} \left( k^d + (n+1-k)^d \right). \tag{195}$$

But if $k \in \mathbb{Z}$, then Lemma 2.10.11 (b) (applied to $x = k$ and $y = n+1-k$) shows that $k^d + (n+1-k)^d$ is divisible by $k + (n+1-k) = n + 1$. Hence, each addend in the sum on the right hand side of (195) is divisible by $n + 1$. Therefore, the whole sum is divisible by $n + 1$ as well. Thus, the left hand side is divisible by $n + 1$, too. In other words, $n + 1 | 2 \left( 1^d + 2^d + \cdots + n^d \right)$. Thus, (194) is proven.

Proof of (193): If $n = 0$, then (193) boils down to $0 \mid 2 \cdot 0$ (since empty sums are 0); this is obvious. Thus, for the rest of this proof, we WLOG assume that $n \neq 0$. Hence, $n$ is a positive integer, and thus $n - 1 \in \mathbb{N}$. Therefore, we can apply (194) to $n - 1$ instead of $n$ (since we have already proven (194) for each $n \in \mathbb{N}$). We thus obtain

$$n | 2 \left( 1^d + 2^d + \cdots + (n-1)^d \right).$$
In other words, $2 \left( 1^d + 2^d + \cdots + (n - 1)^d \right) \equiv 0 \mod n$. Now,

$$2 \left( 1^d + 2^d + \cdots + n^d \right) - 2 \left( 1^d + 2^d + \cdots + (n - 1)^d \right) = 2 \cdot \left( \left( 1^d + 2^d + \cdots + n^d \right) - \left( 1^d + 2^d + \cdots + (n - 1)^d \right) \right) = 2n^d = n \cdot 2n^{d-1}$$

is clearly divisible by $n$. In other words,

$$2 \left( 1^d + 2^d + \cdots + n^d \right) \equiv 2 \left( 1^d + 2^d + \cdots + (n - 1)^d \right) \mod n.$$

Hence,

$$2 \left( 1^d + 2^d + \cdots + n^d \right) \equiv 2 \left( 1^d + 2^d + \cdots + (n - 1)^d \right) \equiv 0 \mod n.$$ 

That is, $n \mid 2 \left( 1^d + 2^d + \cdots + n^d \right)$. This proves (193).

We have now proven both (193) and (194). As we have explained, this yields (192), which in turn solves Exercise 2.10.8.

### 6.36. Solution to Exercise 2.10.9

Solution to Exercise 2.10.9  Proposition 2.9.27 yields $\gcd(a, b) \mid xa + yb = 1$. But $\gcd(a, b)$ is a nonnegative integer. Hence, Exercise 2.2.5 (applied to $g = \gcd(a, b)$) yields $\gcd(a, b) = 1$ (since $\gcd(a, b) \mid 1$). In other words, $a$ is coprime to $b$. In other words, $a \perp b$. This solves Exercise 2.10.9.

### 6.37. Solution to Exercise 2.10.10

Solution to Exercise 2.10.10  Let $g = \gcd(x, y)$. Then, $g$ is a nonnegative integer (since any $\gcd$ is a nonnegative integer); thus, $|g| = g$.

Theorem 2.9.25 (applied to $2$, $(ux, uy)$, $2$ and $(vx, vy)$ instead of $k$, $(b_1, b_2, \ldots, b_k)$, $\ell$ and $(c_1, c_2, \ldots, c_\ell)$) yields

$$\gcd(ux, uy, vx, vy) = \gcd(\gcd(ux, uy), \gcd(vx, vy)).$$

Corollary 2.9.19 (applied to $s = u$, $a = x$ and $b = y$) yields

$$\gcd(ux, uy) = |u| \gcd(x, y) = |u| g = g |u|.$$ 

The same argument (applied to $v$ instead of $u$) yields

$$\gcd(vx, vy) = g |v|.$$
Now, (196) becomes
\[
\gcd (ux, uy, vx, vy) = \gcd \left( \frac{\gcd (ux, uy), \gcd (vx, vy)}{-g|u| -g|v|} \right) = \gcd (g |u|, g |v|) \\
= \frac{-g|u| -g|v|}{=gcd(x,y)} \quad (by Exercise 2.9.5(c), applied to a=\ell and b=\ell)
\]
(by Corollary 2.9.19 applied to s = g, a = |u| and b = |v|)

\[= \gcd (x, y) \cdot \gcd (u, v) = \gcd (u, v) \cdot \gcd (x, y).\]

This solves Exercise 2.10.10

6.38. Solution to Exercise 2.10.11

Solution to Exercise 2.10.11 (a) Let \( g = \gcd (a, b, c) \). Then, \( g = \gcd (a, b, c) \geq 0 \) (since any \( \gcd \) is a nonnegative integer) and thus \( |g| = g \). But Exercise 2.9.6 (applied to \( s = a \)) yields
\[
\gcd (aa, ab, ac) = |a| \gcd (a, b, c) = |a| \cdot |g| = |ag|
\]
(since (\ref{196}) yields \( |ag| = |a| \cdot |g| \)).

Exercise 2.10.10 (applied to \( u = a, v = c, x = a \) and \( y = b \)) yields
\[
gcd (a, c) \cdot \gcd (a, b)
\]
\[
= \gcd \left( \frac{aa, ab, ca, cb}{=ac =bc} \right) = \gcd (aa, ab, ac, bc) = \gcd \left( \frac{\gcd (aa, ab, ac), bc}{=|ag|} \quad (by Theorem 2.9.20(d)) \right)
\]
\[
= \gcd (|ag|, bc) = \gcd (ag, bc)
\]
(by Exercise 2.9.5(b), applied to \( ag \) and \( bc \) instead of \( a \) and \( b \)). In other words, \( \gcd (a, b) \cdot \gcd (a, c) = \gcd (ag, bc) \). This solves Exercise 2.10.11(a).

(b) Assume that \( b \perp c \). Thus, \( \gcd (b, c) = 1 \).

Let \( g = \gcd (a, b, c) \). Theorem 2.9.25 (applied to 1, (a), 2 and (b, c) instead of \( k \), \( b_1, b_2, \ldots, b_k \), \( \ell \) and \( \ell_1, \ell_2, \ldots, \ell_\ell \)) yields
\[
gcd (a, b, c) = \gcd \left( \frac{\gcd (a), \gcd (b, c)}{=1} \right) = \gcd (\gcd (a), 1) \mid 1
\]
(by Proposition 2.9.7(f), applied to \( \gcd (a) \) and 1 instead of \( a \) and \( b \)). This rewrites as \( g \mid 1 \) (since \( g = \gcd (a, b, c) \)).
but \( g = \gcd(a, b, c) \) is a nonnegative integer (since any \( \gcd \) is a nonnegative integer). Hence, Exercise \textbf{2.2.5} yields \( g = 1 \) (since \( g \mid 1 \)). Now, Exercise \textbf{2.10.11}(a) yields

\[ \gcd(a, b) \cdot \gcd(a, c) = \gcd \left( \frac{a}{g}, bc \right) = \gcd(a, bc). \]

This solves Exercise \textbf{2.10.11}(b).

\[ \square \]

6.39. Solution to Exercise \textbf{2.10.12}

\textbf{Solution to Exercise \textbf{2.10.12}}. We have assumed that \( a \) and \( b \) are not both zero. In other words, the two integers \( a, b \) are not all zero. Hence, Definition \textbf{2.9.6} shows that \( \gcd(a, b) \) is defined as the largest element of the set \( \text{Div}(a, b) \) and is a positive integer.

Now, \( g = \gcd(a, b) \). Hence, \( g \) is a positive integer (since \( \gcd(a, b) \) is a positive integer). Thus, \( |g| = g \). Also, \( g \neq 0 \) (since \( g \) is positive).

Proposition \textbf{2.9.7}(f) yields \( \gcd(a, b) \mid a \) and \( \gcd(a, b) \mid b \). Hence, \( g = \gcd(a, b) \mid a \). But Proposition \textbf{2.2.3}(c) (applied to \( g \) and \( a \) instead of \( a \) and \( b \)) shows that \( g \mid a \) if and only if \( \frac{a}{g} \in \mathbb{Z} \). Hence, we have \( \frac{a}{g} \in \mathbb{Z} \) (since \( g \mid a \)). Similarly, \( \frac{b}{g} \in \mathbb{Z} \). Thus, \( a \) and \( b \) are integers.

It remains to prove that \( \frac{a}{g} \perp \frac{b}{g} \). But Corollary \textbf{2.9.19} (applied to \( g \), \( \frac{a}{g} \) and \( \frac{b}{g} \) instead of \( s \), \( a \) and \( b \)) shows that

\[ \gcd \left( g \cdot \frac{a}{g}, g \cdot \frac{b}{g} \right) = \gcd \left( \frac{a}{g}, \frac{b}{g} \right) = g \gcd \left( \frac{a}{g}, \frac{b}{g} \right). \]

Comparing this with

\[ \gcd \left( g \cdot \frac{a}{g}, g \cdot \frac{b}{g} \right) = g, \]

we obtain \( g \gcd \left( \frac{a}{g}, \frac{b}{g} \right) = g \). We can cancel \( g \) from this equality (since \( g \neq 0 \)), and thus obtain \( \gcd \left( \frac{a}{g}, \frac{b}{g} \right) = 1 \). In other words, \( \frac{a}{g} \perp \frac{b}{g} \). Thus, the solution of Exercise \textbf{2.10.12} is finished.

\[ \square \]

6.40. Solution to Exercise \textbf{2.10.13}

\textbf{Solution to Exercise \textbf{2.10.13}}. If \( k = 0 \), then Exercise \textbf{2.10.13} holds.\[147\] Hence, for the rest of this solution, we WLOG assume that \( k \neq 0 \). Thus, \( k \) is a positive integer (since \( k \in \mathbb{N} \)); therefore, \( 0^k = 0 \).

\[147\text{Proof. Assume that } k = 0. \text{ Thus, } (\gcd(a, b))^k = (\gcd(a, b))^0 = 1. \]
If the integers $a$ and $b$ are both zero, then Exercise 2.10.13 holds. Thus, for the rest of this solution, we WLOG assume that $a$ and $b$ are not both zero. Let $g = \gcd(a, b)$. Then, $g \geq 0$ (since any gcd is nonnegative) and therefore $g^k \geq 0$. But Exercise 2.10.12 yields that $\frac{a}{g}$ and $\frac{b}{g}$ are integers satisfying $\frac{a}{g} \perp \frac{b}{g}$. Therefore, Exercise 2.10.4 (applied to $\frac{a}{g}, \frac{b}{g}, k$ and $k$ instead of $a, b, n$ and $m$) yields $\gcd\left(\left(\frac{a}{g}\right)^k, \left(\frac{b}{g}\right)^k\right) = 1$.

Note that $\left(\frac{a}{g}\right)^k$ and $\left(\frac{b}{g}\right)^k$ are integers (since $\frac{a}{g}$ and $\frac{b}{g}$ are integers). Thus, Corollary 2.9.19 (applied to $g^k, \left(\frac{a}{g}\right)^k$ and $\left(\frac{b}{g}\right)^k$ instead of $s, a$ and $b$) yields

$$
\gcd\left(g^k \left(\frac{a}{g}\right)^k, g^k \left(\frac{b}{g}\right)^k\right) = \begin{cases} g^k & \text{(since } g^k \geq 0) \\ 1 & \text{otherwise} \end{cases} = g^k = (\gcd(a, b))^k
$$

(since $g = \gcd(a, b)$). Comparing this with

$$
\gcd\left(g^k \left(\frac{a}{g}\right)^k, g^k \left(\frac{b}{g}\right)^k\right) = \gcd\left(\left(\frac{a}{g}\right)^k, \left(\frac{b}{g}\right)^k\right) = \gcd(a^k, b^k),
$$

we obtain $\gcd(a^k, b^k) = (\gcd(a, b))^k$. This solves Exercise 2.10.13. \qed

6.41. Solution to Exercise 2.10.14

Solution to Exercise 2.10.14 We have $r \in \mathbb{Q}$. In other words, $r$ is a rational number. Thus, $r$ can be written in the form $r = x/y$ for some $x \in \mathbb{Z}$ and some nonzero $y \in \mathbb{Z}$ (by the

But $1 \mid 1$. Thus, Proposition 2.9.7 (i) (applied to 1 and 1 instead of $a$ and $b$) yields $\gcd(1, 1) = 1$. Therefore, Exercise 2.10.13 holds (under the assumption that $k = 0$).

148Proof. Assume that $a$ and $b$ are both zero. In other words, $a = 0$ and $b = 0$. Thus,

$$
\gcd(a^k, b^k) = \gcd\left(\left(\frac{a}{g}\right)^k, \left(\frac{b}{g}\right)^k\right) = \gcd(0, 0) = 0
$$

(by Proposition 2.9.7 (i), applied to 0 and 0 instead of $a$ and $b$). Thus, $\gcd(a^k, b^k) = 0 = 0$.

But the integers $a, b$ are all zero (since $a = 0$ and $b = 0$). Thus, the definition of gcd yields $\gcd(a, b) = 0$. Hence, $(\gcd(a, b))^k = 0^k = 0$. Comparing this with $\gcd(a^k, b^k) = 0$, we obtain $\gcd(a^k, b^k) = (\gcd(a, b))^k$. Hence, Exercise 2.10.13 holds (under the assumption that $a$ and $b$ are both zero).
Thus, \( \sqrt{r} \) is rational. Similarly, \( \sqrt{u} \) must be a perfect square (since \( a \) is an integer). This contradicts the fact that \( u \) is not a perfect square. Hence, \( \sqrt{u} \) is irrational. This solves Exercise 2.10.14 (a).

(b) Let \( u \) and \( v \) be two positive integers that are not both perfect squares. We must prove that \( \sqrt{u} + \sqrt{v} \) is irrational.

Assume the contrary. Thus, \( \sqrt{u} + \sqrt{v} \) is rational. Denote this rational number \( \sqrt{u} + \sqrt{v} \) by \( x \). Thus, \( x = \sqrt{u} + \sqrt{v} \). Squaring both sides of this equality, we obtain \((x - \sqrt{u})^2 = v \). Hence,

\[
v = (x - \sqrt{u})^2 = x^2 - 2x\sqrt{u} + (\sqrt{u})^2 = x^2 - 2x\sqrt{u} + u.
\]

Subtracting \( x^2 + u \) from both sides of this equation, we obtain

\[
v - (x^2 + u) = -2x\sqrt{u}.
\]

But \( x = \sqrt{u} + \sqrt{v} > 0 \) (since \( u \) and \( v \) are positive) and thus \( x \neq 0 \), so that \(-2x \neq 0\). Hence, we can solve the equation (198) for \( \sqrt{u} \); we thus obtain

\[
\sqrt{u} = \frac{v - (x^2 + u)}{-2x}.
\]

Thus, \( \sqrt{u} \) is rational (since \( v, x \) and \( u \) are rational). Therefore, \( u \) must be a perfect square (since otherwise, Exercise 2.10.15 (a) would yield that \( \sqrt{u} \) is irrational). Similarly, \( v \) must
be a perfect square. This shows that both $u$ and $v$ are perfect squares; but this contradicts the fact that $u$ and $v$ are not both perfect squares. This contradiction shows that our assumption was false. Hence, $\sqrt{u} + \sqrt{v}$ is irrational. This solves Exercise 2.10.15(b).

6.43. Solution to Exercise 2.10.16

Solution to Exercise 2.10.16. A gcd of a list of integers is always a nonnegative integer (by the definition of a gcd). Hence, in particular, $\gcd(a_1, a_2, \ldots, a_k)$ is a nonnegative integer. Thus, we can define a nonnegative integer $h$ by $h = \gcd(a_1, a_2, \ldots, a_k)$. Consider this $h$. We have $|h| = h$ (since $h$ is nonnegative).

Theorem 2.9.25 (applied to $(a_1 x, a_2 x, \ldots, a_k x), (a_1 y, a_2 y, \ldots, a_k y)$ instead of $(b_1, b_2, \ldots, b_k), (c_1, c_2, \ldots, c_\ell)$) yields

$$\gcd(a_1 x, a_2 x, \ldots, a_k x, a_1 y, a_2 y, \ldots, a_k y) = \gcd(\gcd(a_1 x, a_2 x, \ldots, a_k x), \gcd(a_1 y, a_2 y, \ldots, a_k y)).$$

(199)

But $(a_1 x, a_2 x, \ldots, a_k x) = (xa_1, xa_2, \ldots, xa_k)$ (since $a_i x = xa_i$ for each $i \in \{1, 2, \ldots, k\}$) and thus

$$\gcd(a_1 x, a_2 x, \ldots, a_k x) = \gcd(xa_1, xa_2, \ldots, xa_k) = |x| \gcd(a_1, a_2, \ldots, a_k)$$

(by Exercise 2.9.7 applied to $s = x$).

Comparing this with

$$|xh| = |x| \cdot |h| = |x| \cdot \underbrace{h}_{=h} = |x| \underbrace{\gcd(a_1, a_2, \ldots, a_k)}_{=\gcd(a_1, a_2, \ldots, a_k)}$$

we obtain

$$\gcd(a_1 x, a_2 x, \ldots, a_k x) = |xh|.$$ 

The same argument (applied to $y$ instead of $x$) yields

$$\gcd(a_1 y, a_2 y, \ldots, a_k y) = |yh|.$$
Thus, (199) becomes

\[
\gcd(a_1x, a_2x, \ldots, a_kx, a_1y, a_2y, \ldots, a_ky) = \gcd(\frac{\gcd(a_1x, a_2x, \ldots, a_kx), \gcd(a_1y, a_2y, \ldots, a_ky)}{\gcd(|xh|, |yh|)}) = \gcd(|xh|, |yh|)
\]

\[
= \gcd(\frac{xh}{-hx}, \frac{yh}{-hy}) \quad \text{(by Exercise 2.9.5(e) (applied to } xh \text{ and } yh \text{ instead of } a \text{ and } b)}
\]

\[
= \gcd(hx, hy) = |h| \gcd(x, y) \quad \text{(by Corollary 2.9.19 (applied to } h, x \text{ and } y \text{ instead of } s, a \text{ and } b)}
\]

\[
= \frac{h}{\gcd(a_1, a_2, \ldots, a_k)} \gcd(x, y) = \gcd(a_1, a_2, \ldots, a_k) \cdot \gcd(x, y).
\]

This solves Exercise 2.10.16 \[\square\]

6.44. Solution to Exercise 2.10.17

Solution to Exercise 2.10.17. Theorem 2.9.20(d) (applied to 3 and \((x, y, z)\) instead of \(k\) and \((b_1, b_2, \ldots, b_k)\)) yields \(\gcd(x, y, z) = \gcd(\gcd(x, y), z)\). But Exercise 2.9.5(a) (applied to \(\gcd(x, y)\) and \(z\) instead of \(a\) and \(b\)) yields \(\gcd(\gcd(x, y), |z|) = \gcd(\gcd(x, y), z)\). Comparing these two equalities, we obtain

\[
\gcd(\gcd(x, y), |z|) = \gcd(x, y, z). \quad (200)
\]

A gcd of a list of integers is always a nonnegative integer (by the definition of a gcd). Hence, in particular, \(\gcd(a_1, a_2, \ldots, a_k)\) is a nonnegative integer. Thus, we can define a nonnegative integer \(h\) by \(h = \gcd(a_1, a_2, \ldots, a_k)\). Consider this \(h\). We have \(|h| = h\) (since \(h\) is nonnegative).

Multiplying both sides of the equality \(h = \gcd(a_1, a_2, \ldots, a_k)\) by \(\gcd(x, y)\), we obtain

\[
h \gcd(x, y) = \gcd(a_1, a_2, \ldots, a_k) \cdot \gcd(x, y)
\]

\[
= \gcd(a_1x, a_2x, \ldots, a_kx, a_1y, a_2y, \ldots, a_ky) \quad (201)
\]

(by Exercise 2.10.16).

But \((a_1z, a_2z, \ldots, a_kz) = (za_1, za_2, \ldots, za_k)\) (since \(a_iz = za_i\) for each \(i \in \{1, 2, \ldots, k\}\)) and thus

\[
\gcd(a_1z, a_2z, \ldots, a_kz) = \gcd(za_1, za_2, \ldots, za_k) = |z| \gcd(a_1, a_2, \ldots, a_k)
\]

(by Exercise 2.9.7 applied to \(s = z\)). Thus,

\[
\gcd(a_1z, a_2z, \ldots, a_kz) = |z| \gcd(a_1, a_2, \ldots, a_k) = |z| h = h |z|. \quad (202)
\]
Theorem 2.9.25 (applied to 2k, \((a_1x, a_2x, \ldots, a_kx, a_1y, a_2y, \ldots, a_ky)\), \(k\) and \((a_1z, a_2z, \ldots, a_kz)\) instead of \(k, (b_1, b_2, \ldots, b_k)\), \(\ell\) and \((c_1, c_2, \ldots, c_\ell)\)) yields
\[
gcd(a_1, a_2, \ldots, a_kx, a_1y, a_2y, \ldots, a_ky, a_1z, a_2z, \ldots, a_kz) = \gcd(h \gcd(x, y), h |z|) = \gcd(\gcd(x, y), |z|) = \gcd(\gcd(x, y), |z|) = \gcd(\gcd(x, y), |z|)
\]
(by Corollary 2.9.19 (applied to \(h, \gcd(x, y)\) and \(|z|\) instead of \(s, a\) and \(b\))
\[
= \gcd(a_1, a_2, \ldots, a_k) \cdot \gcd(x, y, z).
\]

This solves Exercise 2.10.17.

6.45. Solution to Exercise 2.10.18

Solution to Exercise 2.10.18 Exercise 2.10.10 (applied to \(u = b, v = c, x = c\) and \(y = a\)) yields
\[
gcd(b, c) \cdot \gcd(c, a) = \gcd(bc, ba, cc, ca).
\]

Multiplying both sides of this equality by \(\gcd(a, b)\), we find
\[
gcd(b, c) \cdot \gcd(c, a) \cdot \gcd(a, b) = \gcd(bc, ba, cc, ca) \cdot \gcd(a, b) = \gcd(bca, baa, cca, caa, bab, ccb, bab, ccaab, bab, ccb, cab).
\]

(by Exercise 2.10.16 applied to \(a, b\) and \((bc, ba, cc, ca)\) instead of \(x, y, k\) and \((a_1, a_2, \ldots, a_k)\)).

On the other hand, Exercise 2.10.17 (applied to \(bc, ca, ab\) and \((a, b, c)\) instead of \(x, y, z, k\) and \((a_1, a_2, \ldots, a_k)\)) yields
\[
gcd(a, b, c) \cdot \gcd(bc, ca, ab) = \gcd(abc, bbc, cbc, aca, bca, cca, aab, bab, ccb, cab).
\]

But comparing
\[
\begin{align*}
\begin{cases}
  bca, baa, cca, cca, bab, ccb, cab \\
  \quad = abc, = a^2b, = c^2a, = a^2c, = b^2c, = b^2a, = abc \\
\end{cases}
= \{abc, a^2b, c^2a, b^2c, b^2a, c^2b, abc\} = \{abc, b^2c, c^2b, a^2c, a^2b, b^2a\}
\end{align*}
\]

with
\[
\begin{align*}
\begin{cases}
  abc, bbc, cbc, aca, bca, cca, aab, bab, ccb, cab \\
  \quad = b^2c, = c^2b, = a^2c, = abc, = c^2a, = a^2b, = b^2a, = abc \\
\end{cases}
= \{abc, b^2c, c^2b, a^2c, abc, c^2a, a^2b, b^2a, abc\} = \{abc, b^2c, c^2b, a^2c, a^2b, b^2a\},
\end{align*}
\]

\(\square\)
we obtain

\[ \{ bca, baa, cca, caa, bcb, bab, ccb, cab \} = \{ abc, bbc, cbc, abc, cca, aab, bab, cab \} . \]

Hence, Exercise 2.9.2 (applied to 8, \( bca, baa, cca, caa, bcb, bab, ccb, cab \)), 9 and 
\((abc, bbc, cbc, apa, bca, cca, aab, bab, cab)\) instead of \(k, (b_1, b_2, \ldots, b_k), \ell \) and \((c_1, c_2, \ldots, c_i)\) yields

\[ \gcd (bca, baa, cca, caa, bcb, bab, ccb, cab) = \gcd (abc, bbc, cbc, abc, cca, aab, bab, cab) . \]

Comparing this with (205), we obtain

\[ \gcd (a, b, c) \cdot \gcd (bc, ca, ab) = \gcd (bca, baa, cca, caa, bcb, bab, ccb, cab) . \]

Comparing this with (204), we obtain

\[ \gcd (b, c) \cdot \gcd (c, a) \cdot \gcd (a, b) = \gcd (a, b, c) \cdot \gcd (bc, ca, ab) . \]

This solves Exercise 2.10.18 \(\square\)

### 6.46. Solution to Exercise 2.11.1

**Solution to Exercise 2.11.1** The lowest common multiple of any set of integers is a nonnegative integer (by Definition 2.11.4). Thus, in particular, \(\lcm (a, b)\) and \(\lcm (b, a)\) and \(\lcm (-a, b)\) are nonnegative integers.

(a) Proposition 2.11.5 (c) (applied to \(b\) and \(a\) instead of \(a\) and \(b\)) yields \(b \mid \lcm (b, a)\) and \(a \mid \lcm (b, a)\). Thus, \(a \mid \lcm (b, a)\) and \(b \mid \lcm (b, a)\). Hence, Lemma 2.11.8 (applied to \(m = \lcm (b, a)\)) yields \(\lcm (a, b) \mid \lcm (b, a)\). The same argument (applied to \(b\) and \(a\) instead of \(a\) and \(b\)) yields \(\lcm (b, a) \mid \lcm (a, b)\). Hence, Exercise 2.2.2 (applied to \(\lcm (a, b)\) and \(\lcm (b, a)\) instead of \(a\) and \(b\)) yields \(|\lcm (a, b)| = |\lcm (b, a)| = \lcm (b, a)\) (since \(\lcm (b, a)\) is nonnegative). But \(\lcm (a, b)\) is nonnegative; thus, \(|\lcm (a, b)| = \lcm (a, b)\). Hence, \(\lcm (a, b) = |\lcm (a, b)| = \lcm (b, a)\). This solves Exercise 2.11.1 (a).

(b) Proposition 2.11.5 (c) (applied to \(-a\) instead of \(a\)) yields \(-a \mid \lcm (-a, b)\) and \(b \mid \lcm (-a, b)\). But \(a \mid -a\) (since \(-a = a \cdot (-1)\)). Thus, \(-a \mid \lcm (-a, b)\) and \(b \mid \lcm (-a, b)\). Hence, Lemma 2.11.8 (applied to \(m = \lcm (-a, b)\)) yields \(\lcm (a, b) \mid \lcm (-a, b)\). The same argument (applied to \(-a\) instead of \(a\)) yields \(\lcm (-a, b) \mid \lcm (-(-a), b)\). In view of \(-(-a) = a\), this rewrites as \(\lcm (-a, b) \mid \lcm (a, b)\). Hence, Exercise 2.2.2 (applied to \(\lcm (a, b)\) and \(\lcm (-a, b)\) instead of \(a\) and \(b\)) yields \(|\lcm (a, b)| = |\lcm (-a, b)| = \lcm (-a, b)\) (since \(\lcm (-a, b)\) is nonnegative). But \(\lcm (a, b)\) is nonnegative; thus, \(|\lcm (a, b)| = \lcm (a, b)\). Hence, \(\lcm (a, b) = |\lcm (a, b)| = \lcm (-a, b)\). This solves Exercise 2.11.1 (b).

(c) Exercise 2.11.1 (a) (applied to \(-b\) instead of \(b\)) yields \(\lcm (a, -b) = \lcm (-b, a) = \lcm (b, a)\) (by Exercise 2.11.1 (b), applied to \(-b\) instead of \(a\) and \(b\)). But Exercise 2.11.1 (a) yields \(\lcm (a, b) = \lcm (b, a) = \lcm (a, -b)\) (since \(\lcm (a, -b) = \lcm (b, a)\)). This solves Exercise 2.11.1 (c).

(d) Assume that \(a \mid b\). Thus, \(a \mid b\) and \(b \mid b\). Hence, Lemma 2.11.8 (applied to \(m = b\)) yields \(\lcm (a, b) \mid b\). But Proposition 2.11.5 (c) yields \(a \mid \lcm (a, b)\) and \(b \mid \lcm (a, b)\). Hence, Exercise 2.2.2 (applied to \(b\) and \(\lcm (a, b)\) instead of \(a\) and \(b\)) yields \(|b| = |\lcm (a, b)| = \lcm (a, b)\) (since \(\lcm (a, b)\) is nonnegative). In other words, \(\lcm (a, b) = |b|\). This solves Exercise 2.11.1 (d).
(e) If \( s = 0 \), then Exercise 2.11.1 (e) holds\(^{149}\). Hence, for the rest of this solution, we WLOG assume that \( s \neq 0 \).

Recall that any \( \text{lcm} \) is a nonnegative integer. Thus, \( \text{lcm} (sa, sb) \) is a nonnegative integer.

Proposition 2.11.5 (e) (applied to \( sa \) and \( sb \) instead of \( a \) and \( b \)) yields \( sa \mid \text{lcm} (sa, sb) \) and \( sb \mid \text{lcm} (sa, sb) \). Hence, \( s \mid sa \mid \text{lcm} (sa, sb) \). In other words, there exists some integer \( d \) such that \( \text{lcm} (sa, sb) = sd \). Consider this \( d \).

Now, \( as \mid \text{lcm} (sa, sb) = sd = ds \). But Exercise 2.11.1 (c) (applied to \( a, d \) and \( s \)) yields that \( a \mid d \) holds if and only if \( as \mid ds \). Thus, we have \( a \mid d \) (since \( as \mid ds \)).

Also, \( bs \mid \text{lcm} (sa, sb) = sd = ds \). But Exercise 2.11.1 (c) (applied to \( b, d \) and \( s \)) yields \( b \mid d \) holds if and only if \( bs \mid ds \). Thus, we have \( b \mid d \) (since \( bs \mid ds \)).

From \( a \mid d \) and \( b \mid d \), we obtain \( \text{lcm} (a, b) \mid d \) (by Lemma 2.11.8 applied to \( m = d \)). Hence, Proposition 2.2.4 (c) (applied to \( a_1 = s, a_2 = \text{lcm} (a, b), b_1 = s \) and \( b_2 = d \)) yields \( s \mid \text{lcm} (a, b) \mid sd \) (since \( s \mid s \)). In view of \( \text{lcm} (sa, sb) = sd \), this rewrites as \( s \mid \text{lcm} (sa, sb) \).

Proposition 2.11.5 (c) yields \( a \mid \text{lcm} (a, b) \) and \( b \mid \text{lcm} (a, b) \). Hence, Proposition 2.2.4 (c) (applied to \( a_1 = s, a_2 = a, b_1 = s \) and \( b_2 = \text{lcm} (a, b) \)) yields \( sa \mid s \text{lcm} (a, b) \) (since \( s \mid s \) and \( a \mid \text{lcm} (a, b) \)). Similarly, we obtain \( sb \mid s \text{lcm} (a, b) \) (since \( s \mid s \) and \( b \mid \text{lcm} (a, b) \)). Hence, Lemma 2.11.8 (applied to \( sa, sb \) and \( s \text{lcm} (a, b) \) instead of \( a, b \) and \( m \)) yields that \( \text{lcm} (sa, sb) \mid s \text{lcm} (a, b) \).

Now, we know that \( s \text{lcm} (a, b) \mid \text{lcm} (sa, sb) \) and \( \text{lcm} (sa, sb) \mid s \text{lcm} (a, b) \). Hence, Exercise 2.2.2 (applied to \( s \text{lcm} (a, b) \) and \( \text{lcm} (sa, sb) \) instead of \( a \) and \( b \)) yields \( s \mid \text{lcm} (a, b) \mid |s \text{lcm} (a, b)| = s \text{lcm} (sa, sb) \) (since \( \text{lcm} (sa, sb) \)) is nonnegative).

Hence,

\[
\text{lcm} (sa, sb) = |s \text{lcm} (a, b)| = |s| \cdot \frac{|\text{lcm} (a, b)|}{\text{lcm} (a, b)} = |s| \text{lcm} (a, b).
\]

This solves Exercise 2.11.1 (e). \( \square \)

6.47. Solution to Exercise 2.11.2

First solution to Exercise 2.11.2 Let us prove a more general fact:

Claim 1: Let \( x, y, z, N \) be four integers such that \( ax = by = cz = N \). Then, \( \gcd (a, b, c) \cdot \text{lcm} (x, y, z) = |N| \).

Once we have proven Claim 1, we will immediately obtain Exercise 2.11.2 (a) by applying Claim 1 to \( x = bc, y = ca, z = ab \) and \( N = abc \); and we will obtain Exercise 2.11.2 (b) in a similar way (see below for the details). Thus, let us focus on proving Claim 1.

\(^{149}\text{Proof.} \) Assume that \( s = 0 \). Thus, \( s = 0 \). Hence, the integers \( sa, sb \) are not all nonzero. Hence,

\[
\text{lcm} (sa, sb) = 0 \quad \text{(by Definition 2.11.4).Comparing this with} \quad s \mid \text{lcm} (a, b) = 0 \text{ \quad \text{lcm} (a, b) = 0},
\]

we obtain \( \text{lcm} (sa, sb) = s \mid \text{lcm} (a, b) \). Hence, Exercise 2.11.1 (e) holds. Qed.
Proof of Claim 1: If the integers \(x, y, z\) are not all nonzero, then Claim 1 holds.\(^{150}\) Thus, for the rest of this proof, we WLOG assume that the integers \(x, y, z\) are all nonzero. Hence, \(\text{lcm} (x, y, z)\) is the smallest positive element of the set \(\text{Mul} (x, y, z)\) (by Definition 2.11.4). Thus, \(\text{lcm} (x, y, z)\) is a positive integer.

If the integers \(a, b, c\) are all zero, then Claim 1 holds.\(^{151}\) Hence, for the rest of this proof, we WLOG assume that the integers \(a, b, c\) are not all zero. Hence, \(\gcd (a, b, c)\) is a positive integer (by Definition 2.9.6). Denote this positive integer by \(g\). Hence, \(g = \gcd (a, b, c)\).

Definition 2.9.6 also shows that \(\gcd (a, b, c)\) is the largest element of the set \(\text{Div} (a, b, c)\) (since \(a, b, c\) are not all zero). Hence, \(\gcd (a, b, c) \in \text{Div} (a, b, c)\). In other words, \(g \in \text{Div} (a, b, c)\) (since \(g = \gcd (a, b, c)\)). In other words, \(g\) is a common divisor of \(a, b, c\) (by the definition of \(\text{Div} (a, b, c)\)). In other words, \(g\) is an integer satisfying \(g \, | \, a \) and \(g \, | \, b \) and \(g \, | \, c\). Thus, \(g \, | \, a \mid ax = N\). In other words, there exists an integer \(h\) such that \(N = gh\). Consider this \(h\).

It is easy to see that \(N \neq 0\).\(^{152}\) Now, \(gh = N \neq 0\) and thus \(h \neq 0\). Hence, \(|h|\) is a positive integer (since \(h\) is an integer). Denote this positive integer by \(m\). Thus, \(m = |h|\).

Also, set \(N' = |N|\). Thus, \(N'\) is an integer satisfying

\[
N' = \left\lfloor \frac{N}{gh} \right\rfloor = |gh| = \frac{|g|}{g} \cdot \frac{|h|}{m} = gm.
\]

(206)

Our next goal is to prove that \(m = \text{lcm} (x, y, z)\). First, we shall prove that \(m \in \text{Mul} (x, y, z)\). Indeed, we have \(h \neq 0\). Hence, Exercise 2.2.3 (applied to \(g\), \(a\) and \(h\) instead of \(a\), \(b\) and \(c\)) shows that \(g \, | \, a\) holds if and only if \(gh \mid ah\). Hence, \(gh \mid ah\) holds (since \(g \mid a\) holds).

Now, \(xa = ax = N = gh \mid ah = ha\). But \(a \neq 0\) (since \(ax = N \neq 0\)). Thus, Exercise 2.2.3 (applied to \(x\), \(h\) and \(a\) instead of \(a\), \(b\) and \(c\)) shows that \(x \mid h\) holds if and only if \(xa \mid ha\). Hence, \(x \mid h\) holds (since \(xa \mid ha\) holds). But Exercise 2.2.1 (applied to \(h\) instead

---

\(^{150}\) Proof. Assume that the integers \(x, y, z\) are not all nonzero. In other words, \(x = 0\) or \(y = 0\) or \(z = 0\). We thus WLOG assume that \(x = 0\) (since the proofs in the two cases \(y = 0\) and \(z = 0\) are analogous).

The integers \(x, y, z\) are not all nonzero. Hence, Definition 2.11.4 yields that their lowest common multiple is 0. In other words, \(\text{lcm} (x, y, z) = 0\).

But \(ax = N\), thus \(N = a \frac{x}{a} = 0\). Hence, \(|N| = |0| = 0\). Comparing this with \(\gcd (a, b, c) \cdot \text{lcm} (x, y, z) = 0\), we obtain \(\gcd (a, b, c) \cdot \text{lcm} (x, y, z) = |N|\). Hence, Claim 1 holds, qed.

\(^{151}\) Proof. Assume that the integers \(a, b, c\) are all zero. Hence, \(\gcd (a, b, c) = 0\) (by Definition 2.9.6). Also, \(a = 0\) (since \(a, b, c\) are all zero).

But \(ax = N\), thus \(N = a \frac{a}{a} x = 0\). Hence, \(|N| = |0| = 0\). Comparing this with \(\gcd (a, b, c) \cdot \text{lcm} (x, y, z) = 0\), we obtain \(\gcd (a, b, c) \cdot \text{lcm} (x, y, z) = |N|\). Hence, Claim 1 holds, qed.

\(^{152}\) Proof. Assume the contrary. Thus, \(N = 0\). But \(x \neq 0\) (since \(x, y, z\) are nonzero). Hence, from \(ax = N = 0\), we obtain \(a = 0\). Similarly, \(b = 0\) and \(c = 0\). Thus, the integers \(a, b, c\) are all zero. This contradicts the fact that the integers \(a, b, c\) are not all zero. This contradiction shows that our assumption was wrong, qed.
of a) yields \( h \mid |h| = m \). Thus, \( x \mid h \mid m \). Similarly, \( y \mid m \) and \( z \mid m \). Thus, we have \( x \mid m \) and \( y \mid m \) and \( z \mid m \). In other words, \( m \) is a common multiple of \( x, y, z \). In other words, \( m \in \text{Mul}(x, y, z) \). So we know that \( m \) is a positive element of the set \( \text{Mul}(x, y, z) \) (since \( m \) is positive).

We shall now show that \( m \) is the smallest positive element of this set. Indeed, let \( w \) be any positive element of \( \text{Mul}(x, y, z) \). We are going to prove that \( w \geq m \).

In fact, \( w \in \text{Mul}(x, y, z) \). In other words, \( w \) is a common multiple of \( x, y, z \). In other words, we have \( (x \mid w \) and \( y \mid w \) and \( z \mid w \). Also, \( w \neq 0 \) (since \( w \) is positive).

We have \( wa \neq 0 \) (since \( w \neq 0 \) and \( a \neq 0 \)). Hence, the integers \( wa, wb, wc \) are not all zero. Thus, Definition 2.9.6 shows that \( \gcd(wa, wb, wc) \) is the largest element of the set \( \text{Div}(wa, wb, wc) \).

We have \( a \neq 0 \). Hence, Exercise 2.2.3 (applied to \( x, w \) and \( a \) instead of \( b, c \)) shows that \( x \mid w \) holds if and only if \( xa \mid wa \). Hence, \( xa \mid wa \) holds (since \( x \mid w \). Thus, \( N = ax = xa \mid wa \). But Exercise 2.2.1(b) (applied to \( N \) instead of \( a \)) yields \( |N| \mid N \). In other words, \( N' \mid N \) (since \( N' = |N| \)). Hence, \( N' \mid wa \). Similarly, \( N' \mid wb \) and \( N' \mid wc \). Thus, \( (N' \mid wa \) and \( N' \mid wb \) and \( N' \mid wc \). In other words, \( N' \) is a common divisor of \( wa, wb, wc \). In other words, \( N' \in \text{Div}(wa, wb, wc) \). Hence, \( N' \leq \gcd(wa, wb, wc) \) (since \( \gcd(wa, wb, wc) \) is the largest element of the set \( \text{Div}(wa, wb, wc) \)). Now, (206) yields

\[
\begin{align*}
gm &= N' \leq \gcd(wa, wb, wc) = \underbrace{|w|}_{\text{(since w is positive)}} \underbrace{\gcd(a, b, c)}_{=g} \\
&= \gcd(wa, wb, wc) = \gcd(w, g).
\end{align*}
\]

We can divide both sides of this inequality by \( g \) (since \( g \) is positive), and thus obtain \( m \leq w \).

In other words, \( w \geq m \).

Now, forget that we fixed \( w \). We thus have proven that each positive element \( w \) of the set \( \text{Mul}(x, y, z) \) satisfies \( w \geq m \). Hence, \( m \) is the smallest positive element of the set \( \text{Mul}(x, y, z) \) (since we already know that \( m \) is a positive element of the set \( \text{Mul}(x, y, z) \)). In other words, \( m \) is \( \text{lcm}(x, y, z) \) (since \( \text{lcm}(x, y, z) \) is the smallest positive element of the set \( \text{Mul}(x, y, z) \)). In other words, \( m = \text{lcm}(x, y, z) \). Hence, (206) becomes

\[
N' = \underbrace{g}_{= \gcd(a, b, c)} \underbrace{m}_{= \text{lcm}(x, y, z)} = \gcd(a, b, c) \cdot \text{lcm}(x, y, z).
\]

Thus, \( \gcd(a, b, c) \cdot \text{lcm}(x, y, z) = N' = |N| \). This proves Claim 1.

We can now solve the actual exercise:

(a) We have \( a(bc) = b(ca) = c(ab) = abc \). Hence, Claim 1 (applied to \( x = bc, y = ca, z = ab \) and \( N = abc \)) yields \( \gcd(a, b, c) \cdot \text{lcm}(bc, ca, ab) = |abc| \). This solves Exercise 2.11.2(a).

(b) We have \( (bc)a = (ca)b = (ab)c = abc \). Hence, Claim 1 (applied to \( bc, ca, ab, a, b, c \) and \( abc \) instead of \( a, b, c, x, y, z \) and \( N \)) yields \( \gcd(bc, ca, ab) \cdot \text{lcm}(a, b, c) = |abc| \). Thus, \( \text{lcm}(a, b, c) \cdot \gcd(bc, ca, ab) = \gcd(bc, ca, ab) \cdot \text{lcm}(a, b, c) = |abc| \). This solves Exercise 2.11.2(b).
6.48. Solution to Exercise 2.13.1

Solution to Exercise 2.13.1. We have $p \neq q$ (since $p$ and $q$ are distinct). Hence, $p \nmid q$.

But Proposition 2.13.5 (applied to $a = q$) shows that either $p \mid q$ or $p \perp q$. Since $p \mid q$ cannot hold (because we have $p \nmid q$), we thus conclude that $p \perp q$. This solves Exercise 2.13.1.

6.49. Solution to Exercise 2.13.2

Solution to Exercise 2.13.2. The integer $p$ is positive (since $p > 1 > 0$). Thus, $|p| = p$.

Let $d$ be a positive divisor of $p$ other than 1 and $p$. We shall derive a contradiction.

We know that $d$ is a divisor of $p$ other than 1 and $p$. Hence, $d \neq 1$ and $d \neq p$.

But $d$ is a divisor of $p$. In other words, there exists an integer $c$ such that $p = dc$. Consider this $c$.

The integer $d$ is positive, therefore nonzero. Hence, we can solve the equality $p = dc$ for $c$; thus we find $c = p/d > 0$ (since both $p$ and $d$ are positive). Thus, the integer $c$ is positive; hence, $c \geq 1$. Also, $d \geq 1$ (since $d$ is a positive integer). Combining this with $d \neq 1$, we obtain $d > 1$.

Since $c > 0$, we can multiply the inequality $d > 1$ by $c$. We thus find $cd > c \cdot 1 = c$.

Hence, $c < cd = dc = p$. Since $c$ is positive, we have $|c| = c < p$. But $p$ is positive; thus, $|p| = p > |c|$ (since $|c| < p$).

Since $d > 0$, we can multiply the inequality $c \geq 1$ by $d$. We thus find $dc \geq d \cdot 1 = d$.

Hence, $d \leq dc = p$. Combining this with $d \neq p$, we obtain $d < p$. Since $d$ is positive, we have $|d| = d < p$. But $p$ is positive; thus, $|p| = p > |d|$ (since $|d| < p$).

We have $p \mid p = dc$. But let us recall that for every $a, b \in \mathbb{Z}$ satisfying $p \mid ab$, we must have $p \mid a$ or $p \mid b$. Applying this to $a = d$ and $b = c$, we conclude that $p \mid d$ or $p \mid c$.

We have $d \neq 0$ (since $d > 0$). Hence, if we had $p \mid d$, then we would have $|p| \leq |d|$ (by Proposition 2.2.3(b), applied to $a = p$ and $b = d$); but this would contradict $|p| > |d|$. Hence, we cannot have $p \mid d$.

We have $c \neq 0$ (since $c > 0$). Hence, if we had $p \mid c$, then we would have $|p| \leq |c|$ (by Proposition 2.2.3(b), applied to $a = p$ and $b = c$); but this would contradict $|p| > |c|$. Hence, we cannot have $p \mid c$.

Thus, we have neither $p \mid d$ nor $p \mid c$. This contradicts the fact that $p \mid d$ or $p \mid c$.

Now, forget that we have fixed $d$. We thus have found a contradiction for each positive divisor $d$ of $p$ other than 1 and $p$. Thus, there exists no positive divisor $d$ of $p$ other than 1 and $p$. In other words, each positive divisor of $p$ is either 1 or $p$. Thus, the only positive divisors of $p$ are 1 and $p$ (since 1 and $p$ are indeed positive divisors of $p$). In other words, $p$ is prime (by the definition of “prime”). This solves Exercise 2.13.2.

\[\Box\]

153 Proof. Assume the contrary. Thus, $p \mid q$. In other words, $p$ is a divisor of $q$.

But $q$ is a prime. According to the definition of a prime, this means that $q > 1$ and that the only positive divisors of $q$ are 1 and $q$.

Also, $p$ is a prime; thus, $p > 1$ (by the definition of a prime); hence, $p > 1 > 0$. Thus, $p$ is positive. So we know that $p$ is a positive divisor of $q$. Hence, $p$ must be either 1 or $q$ (since the only positive divisors of $q$ are 1 and $q$). Since $p$ cannot be 1 (because $p > 1$), we thus conclude that $p$ must be $q$. In other words, $p = q$. This contradicts $p \neq q$. This contradiction shows that our assumption was false, qed.
6.50. Solution to Exercise 2.13.3

Solution to Exercise 2.13.3 \(\Rightarrow\): Assume that \(a \perp p^k\) holds. We must prove that \(p \nmid a\).

Assume the contrary. Thus, \(p \mid a\). But \(k \geq 1\) (since \(k\) is a positive integer), so that \(1 \leq k\). Hence, Exercise 2.2.4 (applied to \(p\) of the \(p\) words, \(p\) of the \(p\) words, \(p\) of the \(p\) words)\(\Rightarrow\) yields \(p \mid p^k\) (since \(p^1 = p\)). Now, Lemma 2.9.15 (applied to \(m = p\) and \(b = p^k\)) yields \(p \mid \gcd(a, p^k)\) (since \(p \mid a\) and \(p \mid p^k\)). But from \(a \perp p^k\), we obtain \(\gcd(a, p^k) = 1\). Hence, \(p \mid \gcd(a, p^k) = 1\). But \(p\) is prime; hence, \(p > 1 > 0\), so that \(p\) is nonnegative. Hence, Exercise 2.2.5 (applied to \(g = p\)) yields \(p = 1\) (since \(p \mid 1\)). This contradicts \(p > 1\). This contradiction shows that our assumption was false. Hence, \(p \nmid a\) is proven. This concludes the proof of the \(\Rightarrow\) direction of Exercise 2.13.3

\(\Leftarrow\): Assume that \(p \nmid a\). We must prove that \(a \perp p^k\).

Proposition 2.13.5 yields that either \(p \mid a\) or \(p \perp a\). Since \(p \mid a\) does not hold (because \(p \nmid a\)), we thus conclude that \(p \perp a\). But Proposition 2.10.4 (applied to \(b = p\)) yields that \(a \perp p\) if and only if \(p \perp a\). Hence, \(a \perp p\) (since \(p \perp a\)). Thus, Exercise 2.10.4 (applied to \(b = p\), \(n = 1\) and \(m = k\)) yields \(a^1 \perp p^k\). In other words, \(a \perp p^k\). This concludes the proof of the \(\Leftarrow\) direction of Exercise 2.13.3

\(\square\)

6.51. Solution to Exercise 2.13.4

Solution to Exercise 2.13.4 This exercise can easily be solved by induction on \(k\); but here is a more artful proof:

Let \(k \in \mathbb{N}\). We have \(p \geq 2\) (since \(p\) is an integer and satisfies \(p > 1\)). Thus, \(p - 1 \geq 1\).

Recall the identity (25), which holds for every \(a, b \in \mathbb{Q}\). Let us apply this identity to \(a = p\) and \(b = 1\). We thus obtain

\[
(p - 1) \left( p^{k-1} + p^{k-2} \cdot 1 + p^{k-3} \cdot 1^2 + \ldots + p \cdot 1^{k-2} + 1^{k-1} \right) = p^k - \sum_{i=1}^{k-1} i^k = p^k - 1.
\]

Thus,

\[
p^k - 1 = (p - 1) \left( p^{k-1} + p^{k-2} \cdot 1 + p^{k-3} \cdot 1^2 + \ldots + p \cdot 1^{k-2} + 1^{k-1} \right)
\]

\[
= (p - 1) \sum_{i=0}^{k-1} p^i 1^{k-i} \geq 1 \sum_{i=0}^{k-1} p^i 1^{k-i} \quad \text{(since } \sum_{i=0}^{k-1} p^i 1^{k-i} \text{ is clearly } \geq 0)\]

\[
= \sum_{i=0}^{k-1} p^i 1^{k-i} \geq \sum_{i=0}^{k-1} 1^{k-i} = 1^k = k.
\]

Hence, \(p^k \geq k + 1 > k\). This solves Exercise 2.13.4

\(\square\)

6.52. Solution to Exercise 2.13.5

Solution to Exercise 2.13.5 If \(n \geq 0\), then we have \(|n| = n\) and thus \(v_p(|n|) = v_p(n)\). Hence, if \(n \geq 0\), then Exercise 2.13.5 holds. Thus, for the rest of this solution, we WLOG assume
that \( n < 0 \). Hence, \( |n| = -n \).

We have \(-1 \mid p \) (since \( p = (-1) \cdot (-p) \)). Thus, Proposition 2.9.7 (i) (applied to \( a = -1 \) and \( b = p \)) yields \( \gcd (-1, p) = |1| = 1 \). In other words, \(-1 \perp p \). Also, \(-1 = (a) \cdot p^0 \) (since \( p^0 = 1 \)). Thus, Lemma 2.13.27 (b) (applied to \(-1, 0 \) and \( 1 \) instead of \( n, i \) and \( w \)) yields \( v_p (-1) = 0 \). Now, Theorem 2.13.28 (a) (applied to \( a = -1 \) and \( b = n \)) yields \( v_p ((-1) n) = v_p (-1) + v_p (n) \). In view of \((-1) n = -n = |n| \), this rewrites as \( v_p (|n|) = v_p (n) \). This solves Exercise 2.13.5.

\[ \square \]

6.53. Solution to Exercise 2.13.6

Solution to Exercise 2.13.6 Corollary 2.13.29 (applied to \( a_i = a \)) yields

\[
v_p \left( \frac{a \cdots a}{k \text{ times}} \right) = v_p (a) + v_p (a) + \cdots + v_p (a) = kv_p (a).
\]

In view of \( \frac{a \cdots a}{k \text{ times}} = a^k \), this rewrites as \( v_p (a^k) = kv_p (a) \). This solves Exercise 2.13.6. \[ \square \]

6.54. Solution to Exercise 2.13.7

Solution to Exercise 2.13.7 For each \( i \in \{ 1, 2, \ldots, u \} \), the number \( p_i^{a_i} \) is a well-defined positive integer (since \( p_i \) is a prime, and since \( a_i \) is a nonnegative integer). Thus, \( p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u} \) is a product of positive integers, and therefore itself a positive integer.

Now, let \( p \) be a prime. Note that \( p_1^{a_1}, p_2^{a_2}, \ldots, p_u^{a_u} \) are \( u \) integers. Hence, Corollary 2.13.29 (applied to \( u \) and \( p_j^{a_j} \) instead of \( k \) and \( a_j \)) yields

\[
v_p \left( p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u} \right) = v_p \left( p_1^{a_1} \right) + v_p \left( p_2^{a_2} \right) + \cdots + v_p \left( p_u^{a_u} \right)
= \sum_{j=1}^{u} a_j \left[ v_p \left( p_j \right) \right] \quad \text{(by Exercise 2.13.6 applied to } a=p_j \text{ and } k=a_j) = \sum_{j=1}^{u} a_j \begin{cases} 1, & \text{if } p_j = p; \\ 0, & \text{if } p_j \neq p \end{cases}
\]

(by Theorem 2.13.28 (d), applied to \( a=p_j \) (since \( p_j \) is prime))

\[
= \sum_{j \in \{1,2,\ldots,u\}} a_j \begin{cases} 1, & \text{if } p_j = p; \\ 0, & \text{if } p_j \neq p \end{cases}
= a_i \begin{cases} 1, & \text{if } p_i = p; \\ 0, & \text{if } p_i \neq p \end{cases} + \sum_{j \in \{1,2,\ldots,u\}; j \neq i} a_j \begin{cases} 1, & \text{if } p_j = p; \\ 0, & \text{if } p_j \neq p \end{cases}
\]

(207)

Now, forget that we fixed \( p \). We thus have proven the equality (207) for each prime \( p \).

(a) Let \( i \in \{1,2,\ldots,u\} \). Then, \( p_i \) is a prime. Hence, (207) (applied to \( p = p_i \)) yields

\[
v_p \left( p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u} \right) = a_i \begin{cases} 1, & \text{if } p_i = p; \\ 0, & \text{if } p_i \neq p \end{cases} + \sum_{j \in \{1,2,\ldots,u\}; j \neq i} a_j \begin{cases} 1, & \text{if } p_j = p; \\ 0, & \text{if } p_j \neq p \end{cases}
\]

(208)
(here, we have split off the addend for \( j = i \) from the sum). But if \( j \in \{1, 2, \ldots, u\} \) satisfies \( j \neq i \), then it must satisfy \( p_j \neq p_i \) (since the primes \( p_1, p_2, \ldots, p_u \) are distinct) and therefore
\[
\begin{cases}
1, & \text{if } p_j = p_i; \\
0, & \text{if } p_j \neq p_i.
\end{cases}
\] (209)

Hence, (208) becomes
\[
v_p \left( p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u} \right) = a_i \underbrace{1}_{(\text{since } p_i = p_i)} + \sum_{j \neq i} a_j \underbrace{0}_{\text{by (209)}} = a_i + \sum_{j \neq i} a_j 0 = a_i 1 = a_i.
\]

This solves Exercise 2.13.7 (a).

(b) Let \( p \) be a prime satisfying \( p \notin \{p_1, p_2, \ldots, p_u\} \). Then, for each \( j \in \{1, 2, \ldots, u\} \), we have \( p_j \neq p \) (since otherwise, we would have \( p_j = p \) and therefore \( p = p_j \in \{p_1, p_2, \ldots, p_u\} \), which would contradict \( p \notin \{p_1, p_2, \ldots, p_u\} \)) and therefore
\[
\begin{cases}
1, & \text{if } p_j = p_i; \\
0, & \text{if } p_j \neq p_i.
\end{cases}
\] (210)

Hence, (207) becomes
\[
v_p \left( p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u} \right) = \sum_{j \in \{1, 2, \ldots, u\}} a_j \underbrace{1}_{\text{by (210)}} = \sum_{j \in \{1, 2, \ldots, u\}} a_j 0 = 0.
\]

This solves Exercise 2.13.7 (b). \( \square \)

6.55. Solution to Exercise 2.13.8

Solution to Exercise 2.13.8 Applying (41) to \( p = 2 \), we obtain \( v_2 (n) = v_2 (m) \) (since 2 is a prime).

If \( n \) is nonzero, then \( v_p (n) \) is a nonnegative integer for every prime \( p \) (by Definition 2.13.23 (a)). Applying this to \( p = 2 \), we conclude the following: If \( n \) is nonzero, then \( v_2 (n) \) is a nonnegative integer and thus satisfies \( v_2 (n) \neq \infty \). Thus, we have shown that
\[
\text{if } n \text{ is nonzero, then } v_2 (n) \neq \infty.
\] (211)

The same argument (applied to \( m \) instead of \( n \)) shows that
\[
\text{if } m \text{ is nonzero, then } v_2 (m) \neq \infty.
\] (212)
We are in one of the following two cases:

**Case 1:** We have \( n = 0 \).

**Case 2:** We have \( n \neq 0 \).

Let us first consider Case 1. In this case, we have \( n = 0 \). Thus, \( v_2(n) = v_2(0) = \infty \) (by Definition 2.13.23 (b)). Comparing this with \( v_2(n) = v_2(m) \), we obtain \( v_2(m) = \infty \). But if \( m \) was nonzero, then we would have \( v_2(m) \neq \infty \) (by 2.12), which would contradict \( v_2(m) = \infty \). Hence, \( m \) cannot be nonzero. In other words, \( m \) must be 0. Thus, \( m = 0 \). Comparing this with \( n = 0 \), we obtain \( n = m \). Hence, Exercise 2.13.8 is solved in Case 1.

Let us now consider Case 2. In this case, we have \( n \neq 0 \). Thus, \( n \) is nonzero; hence, \( v_2(n) \neq \infty \) (by 2.11). We had \( m = 0 \), then we would have \( v_2(n) = v_2 \left( \frac{m}{n} \right) = v_2(0) = \infty \) (by Definition 2.13.23 (b)), which would contradict \( v_2(n) \neq \infty \). Thus, we cannot have \( m = 0 \). Hence, \( m \) is nonzero.

The integer \( n \) is nonzero and nonnegative; thus, the integer \( n \) is positive. Thus, Corollary 2.13.33 yields

\[
\prod_{p \text{ prime}} v_p(n) = \prod_{p \text{ prime}} v_p(m) \quad \text{(since (41) yields } v_p(n) = v_p(m))
\]

The integer \( m \) is nonzero and nonnegative; thus, the integer \( m \) is positive. Hence, Corollary 2.13.33 (applied to \( m \) instead of \( n \)) yields

\[
m = \prod_{p \text{ prime}} p^{v_p(m)}.
\]

Comparing this with (213), we obtain \( n = m \). Thus, Exercise 2.13.8 is solved in Case 2.

We have now solved Exercise 2.13.8 in both of the Cases 1 and 2. Hence, Exercise 2.13.8 always holds.

### 6.56. Second solution to Exercise 2.11.2

**Second solution to Exercise 2.11.2 (sketched).** WLOG assume that \( a, b, c \) are nonzero (since otherwise, the claim of Exercise 2.11.2 easily reduces to \( 0 = 0 \)). Then, \( abc \) is nonzero as well. Hence, Corollary 2.13.34 yields

\[
|abc| = \prod_{p \text{ prime}} p^{v_p(abc)}.
\]
We have
\[
\frac{\gcd(a, b, c)}{\lcm(bc, ca, ab)} = \prod_{p \text{ prime}} \frac{p^\min\{v_p(a), v_p(b), v_p(c)\}}{p^\max\{v_p(bc), v_p(ca), v_p(ab)\}} = \prod_{p \text{ prime}} \frac{p^\min\{v_p(a), v_p(b), v_p(c)\}}{p^\max\{v_p(bc), v_p(ca), v_p(ab)\}}.
\]
(214)

Let us now fix a prime \(p\), and try to simplify
\[
\min \{v_p(a), v_p(b), v_p(c)\} + \max \{v_p(bc), v_p(ca), v_p(ab)\}.
\]
Indeed, set \(u = v_p(abc)\). Note that \(u \in \mathbb{N}\) (since \(abc\) is nonzero).

Theorem 2.13.28 (a) (applied to \(ca\) and \(b\) instead of \(a\) and \(b\)) yields \(v_p(cab) = v_p(ca) + v_p(b)\). Comparing this with \(v_p\left(\frac{cab}{abc}\right) = v_p(abc) = u\), we obtain \(u = v_p(ca) + v_p(b)\). Subtracting \(v_p(b)\) from this equality, we obtain \(u - v_p(b) = v_p(ca)\). Thus, \(v_p(ca) = u - v_p(b)\). Similarly, \(v_p(ab) = u - v_p(c)\) and \(v_p(bc) = u - v_p(a)\). Now,
\[
\begin{align*}
\min \{v_p(a), v_p(b), v_p(c)\} + &\max \left\{v_p(bc), v_p(ca), v_p(ab)\right\} \\
= &\min \{v_p(a), v_p(b), v_p(c)\} + \max \left\{u - v_p(a), u - v_p(b), u - v_p(c)\right\} \\
= &\min \{v_p(a), v_p(b), v_p(c)\} + (u - \min \{v_p(a), v_p(b), v_p(c)\}) \\
= &u = v_p(abc).
\end{align*}
\]
(215)

Now, forget that we fixed \(p\). We thus have proven (215) for each prime \(p\). Thus, (214) becomes
\[
\gcd(a, b, c) \cdot \lcm(bc, ca, ab) = \prod_{p \text{ prime}} \frac{p^\min\{v_p(a), v_p(b), v_p(c)\}+\max\{v_p(bc), v_p(ca), v_p(ab)\}}{p^\max\{v_p(bc), v_p(ca), v_p(ab)\}} = \prod_{p \text{ prime}} p^{v_p(abc)} = |abc|.
\]
This solves Exercise 2.11.2 (a) again. Similarly we can re-solve Exercise 2.11.2 (b). \(\square\)

\textsuperscript{154}This is allowed, since \(v_p(b) \in \mathbb{N}\) (because \(b\) is nonzero).
Remark 6.56.1. Similarly, we could show that any four integers \(a, b, c, d\) satisfy

\[
\gcd(a, b, c, d) \cdot \text{lcm}(bcd, cda, dab, abc) = |abcd| \quad \text{and} \quad \text{lcm}(a, b, c, d) \cdot \gcd(bcd, cda, dab, abc) = |abcd|.
\]

Indeed, the last equality holds since each prime \(p\) satisfies

\[
\min\{v_p(a), v_p(b), v_p(c), v_p(d)\} + \max\left\{v_p(bcd) - v_p(a), v_p(cda) - v_p(a), v_p(dab) - v_p(a), v_p(abc) - v_p(a)\right\} = \min\{v_p(a), v_p(b), v_p(c), v_p(d)\}
\]

(assuming that \(a, b, c, d\) are nonzero). Similarly, the first equality holds. You can likewise prove generalizations to \(k\) integers\(^{155}\).

6.57. Solution to Exercise 2.13.9

First solution to Exercise 2.13.9 We must prove that \(a \equiv b \mod n\). If \(a = b\), then this is true (because if \(a = b\), then \(a = b \equiv b \mod n\)). Thus, for the rest of this proof, we WLOG assume that we don’t have \(a = b\). Hence, \(a \neq b\), so that \(a - b \neq 0\). Thus, \(a - b\) is a nonzero integer. Hence, \(v_p(a - b) \in \mathbb{N}\) for every prime \(p\).

Let \(p\) be any prime. Then, \(^{43}\) yields \(a \equiv b \mod p^{v_p(n)}\). In other words, \(p^{v_p(n)} \mid a - b\). But Lemma 2.13.25 (applied to \(v_p(n)\) and \(a - b\) instead of \(i\) and \(n\)) yields that \(p^{v_p(n)} \mid a - b\) if and only if \(v_p(a - b) \geq v_p(n)\). Hence, we have \(v_p(a - b) \geq v_p(n)\) (since \(p^{v_p(n)} \mid a - b\)). In other words, \(v_p(n) \leq v_p(a - b)\).

Now, forget that we fixed \(p\). We thus have proven that each prime \(p\) satisfies \(v_p(n) \leq v_p(a - b)\). But Proposition 2.13.35 (applied to \(a - b\) instead of \(m\)) shows that \(n \mid a - b\) if and only if each prime \(p\) satisfies \(v_p(n) \leq v_p(a - b)\). Hence, \(n \mid a - b\) (since each prime \(p\) satisfies \(v_p(n) \leq v_p(a - b)\)). In other words, \(a \equiv b \mod n\). This solves Exercise 2.13.9 \(\Box\)

\(^{155}\)That said, it is probably better (and easier) to generalize Claim 1 from the first solution of Exercise 2.11.2 to \(k\) integers:

**Claim 2:** Let \(k > 0\). Let \(N\) be an integer. Let \(a_1, a_2, \ldots, a_k\) be \(k\) integers, and let \(x_1, x_2, \ldots, x_k\) be \(k\) integers such that \(a_1 x_1 = a_2 x_2 = \cdots = a_k x_k = N\). Then,

\[
\gcd(a_1, a_2, \ldots, a_k) \cdot \text{lcm}(x_1, x_2, \ldots, x_k) = |N|.
\]

This Claim 2 can be proven either by generalizing the proof of Claim 1 from the solution of Exercise 2.11.2 or (again) using Proposition 2.13.38.
Second solution to Exercise 2.13.9. Define an integer \( c \) by \( c = a - b \).

We shall show that

\[ d \mid c \text{ for every positive divisor } d \text{ of } n. \]  \hspace{1cm} (216)

[Proof of (216): We shall prove (216) by strong induction on \( d \):

Let \( e \) be a positive divisor of \( n \). Assume that (216) holds for all positive divisors \( d \) of \( n \) satisfying \( d < e \). We must prove that (216) holds for \( d = e \). In other words, we must prove that \( e \mid c \).

We have assumed that (216) holds for all positive divisors \( d \) of \( n \) satisfying \( d < e \). In other words, if \( d \) is a positive divisor of \( n \) satisfying \( d < e \), then

\[ d \mid c. \]  \hspace{1cm} (217)

Note that the integer \( e \) is nonzero (since \( e \) is positive); thus, \( v_p(e) \in \mathbb{N} \). Also, \( v_p(n) \in \mathbb{N} \) (since \( n \) is nonzero). We know that \( e \mid n \) (since \( e \) is a divisor of \( n \)). But Proposition 2.13.35 (applied to \( e \) and \( n \) instead of \( n \) and \( m \)) shows that \( e \mid n \) if and only if each prime \( p \) satisfies \( v_p(e) \leq v_p(n) \). Hence,

\[ \text{each prime } p \text{ satisfies } v_p(e) \leq v_p(n) \]  \hspace{1cm} (218)

(since \( e \mid n \)).

We must prove that \( e \mid c \). If \( e = 1 \), then this follows from the (obvious) fact that \( 1 \mid c \).

Thus, for the rest of this proof, we WLOG assume that \( e \neq 1 \). Hence, \( e > 1 \) (since \( e \) is a positive integer). Thus, Proposition 2.13.35 (applied to \( e \) instead of \( n \)) shows that there exists at least one prime \( p \) such that \( p \mid e \). Consider this \( p \). Theorem 2.13.27 (applied to \( e \) instead of \( n \)) shows that there exists a nonzero integer \( u \) such that \( u \perp p \) and \( e = up^{v_p(e)} \).

Consider this \( u \). From \( e = up^{v_p(e)} \), we obtain \( u = e/p^{v_p(e)} \); therefore, \( u \) is positive (since \( e \) and \( p^{v_p(e)} \) are positive).

Lemma 2.13.25 (applied to 1 and \( e \) instead of \( i \) and \( n \)) shows that \( p^1 \mid e \) if and only if \( v_p(e) \geq 1 \). Hence, \( v_p(e) \geq 1 \) (since \( p^1 \mid p \mid e \)). Hence, \( p^{v_p(e)} \geq p^1 = p > 1 \). Now, \( e = up^{v_p(e)} \geq u \) (since \( u \) is positive). Thus, \( u < e \). Furthermore, \( p^{v_p(e)} \) is an integer; hence, the equality \( e = up^{v_p(e)} \) yields that \( u \mid e \). Thus, \( u \mid e \mid n \). Hence, \( u \) is a divisor of \( n \). Thus, \( u \) is a positive divisor of \( n \) (since \( u \) is positive). Since \( u < e \), we can thus apply (217) to \( d = u \). We thus obtain \( u \mid c \).

On the other hand, (48) yields \( a \equiv b \mod p^{v_p(n)} \). In other words, \( p^{v_p(n)} \mid a - b \). This rewrites as \( p^{v_p(n)} \mid c \) (since \( c = a - b \)). But \( v_p(e) \leq v_p(n) \) (by (218)). Hence, Exercise 2.2.4 (applied to \( p \), \( v_p(e) \), and \( v_p(n) \) instead of \( a \), \( b \), and \( n \)) yields \( p^{v_p(e)} \mid p^{v_p(n)} \mid c \).

Recall that \( u \perp p \). Thus, Exercise 2.10.4 (applied to \( u \), \( p \), 1 and \( v_p(e) \) instead of \( a \), \( b \), \( n \) and \( m \)) yields \( u^1 \perp p^{v_p(e)} \). In view of \( u^1 = u \), this rewrites as \( u \perp p^{v_p(e)} \).

Now we know that \( u \in \mathbb{Z} \) and \( p^{v_p(e)} \in \mathbb{Z} \) and \( c \in \mathbb{Z} \) and \( u \mid c \) and \( p^{v_p(e)} \mid c \) and \( u \perp p^{v_p(e)} \). Hence, Theorem 2.10.7 (applied to \( u \) and \( p^{v_p(e)} \) instead of \( a \) and \( b \)) yields \( up^{v_p(e)} \mid p^{v_p(e)} \mid c \). In view of \( e = up^{v_p(e)} \), this rewrites as \( e \mid c \). In other words, (216) holds for \( d = e \). This completes the induction step. Hence, (216) is proven by strong induction.]

Let \( n' = |n| \). Then, \( n' \neq |n| > 0 \) (since \( n \) is nonzero). Moreover, Exercise 2.2.1 (b) (applied to \( n \) instead of \( a \)) yields \( |n| \mid n \). Since \( n' = |n| \), this rewrites as \( n' \mid n \). Hence, \( n' \) is a divisor of \( n \). Since \( n' \) is positive, we thus conclude that \( n' \) is a positive divisor of \( n \). Hence, (216) (applied to \( d = n' \)) yields \( n' \mid c \). But Exercise 2.2.1 (a) (applied to \( n \) instead of \( a \)) yields...
\( n \mid |n| \). Since \( n' = |n| \), this rewrites as \( n \mid n' \). Hence, \( n \mid n' \mid c = a - b \). In other words, \( a \equiv b \mod n \). This solves Exercise 2.13.9 again. \( \square \)

### 6.58. Solution to Exercise 2.13.10

**Solution to Exercise 2.13.10** (a) Forget that we fixed \( p \). We thus must prove that
\[
 v_p (\gcd(n, m)) = \min \{ v_p(n), v_p(m) \} \quad \text{for each prime } p. \tag{219}
\]

We are in one of the following two cases:

**Case 1:** The integers \( n \) and \( m \) are both nonzero.

**Case 2:** The integers \( n \) and \( m \) are not both nonzero.

Let us first consider Case 1. In this case, the integers \( n \) and \( m \) are both nonzero. Thus, \(^{156}\) yields
\[
 \gcd(n, m) = \prod_{p \text{ prime}} p^{\min \{v_p(n), v_p(m)\}}.
\]

In particular, the product \( \prod_{p \text{ prime}} p^{\min \{v_p(n), v_p(m)\}} \) is well-defined. In other words:

- The number \( \min \{v_p(n), v_p(m)\} \) is a nonnegative integer for each prime \( p \);
- all but finitely many primes \( p \) satisfy \( \min \{v_p(n), v_p(m)\} = 0 \).

(Both of these facts were proven during our proof of Proposition 2.13.38.)

Hence, Corollary 2.13.37 (applied to \( \gcd(n, m) \) and \( \min \{v_p(n), v_p(m)\} \) instead of \( n \) and \( b_p \)) shows that
\[
 v_q (\gcd(n, m)) = \min \{ v_q(n), v_q(m) \} \quad \text{for each prime } q.
\]

Renaming \( q \) as \( p \) in this statement, we obtain
\[
 v_p (\gcd(n, m)) = \min \{ v_p(n), v_p(m) \} \quad \text{for each prime } p.
\]

Thus, (219) is proven in Case 1.

Now, let us consider Case 2. In this case, the integers \( n \) and \( m \) are not both nonzero. In other words, we have \( n = 0 \) or \( m = 0 \). Thus, we can WLOG assume that \( m = 0 \) (because in the case \( n = 0 \), we can swap \( n \) with \( m \) without changing the meaning of our claim, since \( n \) and \( m \) play symmetric roles\(^{156}\)). Assume this.

Fix a prime \( p \). Now, \( \gcd \left( \left\lfloor \frac{n}{m} \right\rfloor = 0 \right) = \gcd(n, 0) = |n| \) (since Proposition 2.9.7 (a) (applied to \( a = n \)) yields \( \gcd(n, 0) = \gcd(n) = |n| \)). Hence,
\[
 v_p (\gcd(n, m)) = v_p(|n|) = v_p(n) \quad \text{by Exercise 2.13.5}. \tag{213.5}
\]

On the other hand, from \( m = 0 \), we obtain \( v_p(m) = v_p(0) = \infty \) (by Definition 2.13.23 (b)). Thus,
\[
 \min \left\{ v_p(n), v_p(m) \right\}_{=\infty} = \min \{ v_p(n), \infty \} = v_p(n)
\]

\(^{156}\)Here we are using the fact that \( \gcd(n, m) = \gcd(m, n) \) (which follows from Proposition 2.9.7 (b)).
In particular, the product \( \prod_{\text{prime } p} p^{\max\{v_p(n), v_p(m)\}} \)
(by our rules for the symbol \( \infty \)). Comparing this with \( v_p(\gcd(n,m)) = v_p(n) \), we obtain
\( v_p(\gcd(n,m)) = \min\{v_p(n), v_p(m)\} \).

Now, forget that we fixed \( p \). We thus have proven that \( v_p(\gcd(n,m)) = \min\{v_p(n), v_p(m)\} \)
for each prime \( p \). Thus, \((219)\) is proven in Case 2.

We have now proven \((219)\) in each of the two Cases 1 and 2. Thus, \((219)\) always holds.

This solves Exercise 2.13.10 \( a \).

\( b \) Forget that we fixed \( p \). We thus must prove that
\[ v_p(\lcm(n,m)) = \max\{v_p(n), v_p(m)\} \quad \text{for each prime } p. \quad (220) \]

We are in one of the following two cases:

Case 1: The integers \( n \) and \( m \) are both nonzero.

Case 2: The integers \( n \) and \( m \) are not both nonzero.

Let us first consider Case 1. In this case, the integers \( n \) and \( m \) are both nonzero. Thus, \((50)\) yields
\[ \lcm(n,m) = \prod_{\text{prime } p} p^{\max\{v_p(n), v_p(m)\}}. \]

In particular, the product \( \prod_{\text{prime } p} p^{\max\{v_p(n), v_p(m)\}} \) is well-defined. In other words:

- The number \( \max\{v_p(n), v_p(m)\} \) is a nonnegative integer for each prime \( p \);
- all but finitely many primes \( p \) satisfy \( \max\{v_p(n), v_p(m)\} = 0 \).

(Both of these facts were proven during our proof of Proposition 2.13.38.)

Hence, Corollary 2.13.37 \( \text{applied to } \lcm(n,m) \) and \( \max\{v_p(n), v_p(m)\} \) instead of \( n \) and \( b_p \) shows that
\[ v_q(\lcm(n,m)) = \max\{v_q(n), v_q(m)\} \quad \text{for each prime } q. \]

Renaming \( q \) as \( p \) in this statement, we obtain
\[ v_p(\lcm(n,m)) = \max\{v_p(n), v_p(m)\} \quad \text{for each prime } p. \]

Thus, \((220)\) is proven in Case 1.

Now, let us consider Case 2. In this case, the integers \( n \) and \( m \) are not both nonzero. In other words, we have \( n = 0 \) or \( m = 0 \). Thus, we can WLOG assume that \( m = 0 \) (because in the case \( n = 0 \), we can swap \( n \) with \( m \) without changing the meaning of our claim, since \( n \) and \( m \) play symmetric roles\(^{157} \)). Assume this.

Fix a prime \( p \). The numbers \( n, m \) are not all nonzero (since \( m = 0 \)). Hence, \( \lcm(n,m) = 0 \) (by the definition of \( \lcm(n,m) \)). Thus,
\[ v_p(\lcm(n,m)) = v_p(0) = \infty \quad \text{(by Definition 2.13.23 \( b \))}. \]

On the other hand, from \( m = 0 \), we obtain \( v_p(m) = v_p(0) = \infty \) (by Definition 2.13.23 \( b \)). Thus,
\[ \max\{v_p(n), v_p(m)\}_{n=\infty} = \max\{v_p(n), \infty\} = \infty \]

\(^{157}\)Here we are using the fact that \( \lcm(n,m) = \lcm(m,n) \) (which follows from Exercise 2.11.1 \( a \)).
(by our rules for the symbol $\infty$). Comparing this with $v_p(\text{lcm}(n,m)) = \infty$, we obtain $v_p(\text{lcm}(n,m)) = \max\{v_p(n), v_p(m)\}$.

Now, forget that we fixed $p$. We thus have proven that $v_p(\text{lcm}(n,m)) = \max\{v_p(n), v_p(m)\}$ for each prime $p$. Thus, (220) is proven in Case 2.

We have now proven (220) in each of the two Cases 1 and 2. Thus, (220) always holds. This solves Exercise 2.13.10(b).

\[
\square
\]

### 6.59. Solution to Exercise 2.13.11

We shall solve Exercise 2.13.11 in two different ways. The first solution uses $p$-valuations (and Exercise 2.13.10 in particular) to reduce the claim of the exercise to a simple identity between minima and maxima of sets of numbers. This solution (just as our Second proof of Theorem 2.11.6 above) illustrates how properties of gcds and lcm's of integers can be proven in a straightforward way using $p$-valuations. The second solution (due to Bill Dubuque), in contrast, completely avoids the use of prime numbers.

**First solution to Exercise 2.13.11** (a) Let us first show an auxiliary claim:

**Claim 1:** Let $i,j,k \in \mathbb{N} \cup \{\infty\}$ be arbitrary. Then,

\[
\min\{i, \max\{j,k\}\} = \max\{\min\{i,j\}, \min\{i,k\}\}.
\]

**Proof of Claim 1:** We have $j \leq k$ or $j \geq k$. Since $j$ and $k$ play symmetric roles in Claim 1, we can always swap $j$ and $k$; thus, we can WLOG assume that $j \leq k$. Assume this. From $j \leq k$, we obtain $\max\{j,k\} = k$. Moreover, any element of a set must be $\geq$ to the minimum of this set; hence, $i \geq \min\{i,j\}$ and $j \geq \min\{i,j\}$. The minimum $\min\{i,k\}$ must be one of the two elements $i$ and $k$ (since the minimum of a set must always be an element of this set); but both of these elements $i$ and $k$ are $\geq \min\{i,j\}$ (because $i \geq \min\{i,j\}$ and $k \geq j \geq \min\{i,j\}$). Hence, $\min\{i,k\}$ must be $\geq \min\{i,j\}$. Hence, $\max\{\min\{i,j\}, \min\{i,k\}\} = \min\{i,k\}$. Comparing this with $\min\{\max\{i,j\}, k\} = \min\{i,k\}$, we obtain $\min\{i, \max\{j,k\}\} = \max\{\min\{i,j\}, \min\{i,k\}\}$.

This proves Claim 1.

Now, fix a prime $p$. Then, Exercise 2.13.10(a) (applied to $n = a$ and $m = \text{lcm}(b,c)$) yields

\[
v_p(\gcd(a, \text{lcm}(b,c))) = \min\left\{ v_p(a), v_p(\text{lcm}(b,c)) \right\} = \max\left\{ v_p(b), v_p(c) \right\} \quad \text{(by Exercise 2.13.10(b), applied to } n = a \text{ and } m = \infty) \right.
\]

\[
= \min\{v_p(a), \max\{v_p(b), v_p(c)\}\} = \max\{\min\{v_p(a), v_p(b)\}, \min\{v_p(a), v_p(c)\}\}.
\]

(221)
by Claim 1, applied to \( i = v_p(a), j = v_p(b) \) and \( k = v_p(c) \). On the other hand, Exercise 2.13.10(b) (applied to \( n = \text{gcd}(a, b) \) and \( m = \text{gcd}(a, c) \)) yields

\[
v_p(\text{lcm}(\text{gcd}(a, b), \text{gcd}(a, c))) = \max \left\{ \frac{v_p(\text{gcd}(a, b))}{\min\{v_p(a), v_p(b)\}}, \frac{v_p(\text{gcd}(a, c))}{\min\{v_p(a), v_p(c)\}} \right\}
\]

(by Exercise 2.13.10(a), applied to \( n=a \) and \( m=b \))

\[
= \max \left\{ \min\{v_p(a), v_p(b)\}, \min\{v_p(a), v_p(c)\} \right\}.
\]

Comparing this with (221), we obtain

\[
v_p(\text{gcd}(a, \text{lcm}(b, c))) = v_p(\text{lcm}(\text{gcd}(a, b), \text{gcd}(a, c))).
\]

Now, forget that we fixed \( p \). We thus have proven that

\[
v_p(\text{gcd}(a, \text{lcm}(b, c))) = v_p(\text{lcm}(\text{gcd}(a, b), \text{gcd}(a, c)))
\]

for every prime \( p \).

Thus, Exercise 2.13.8 (applied to \( n = \text{gcd}(a, \text{lcm}(b, c)) \) and \( m = \text{lcm}(\text{gcd}(a, b), \text{gcd}(a, c)) \)) yields that \( \text{gcd}(a, \text{lcm}(b, c)) = \text{lcm}(\text{gcd}(a, b), \text{gcd}(a, c)) \) (since \( \text{gcd}(a, \text{lcm}(b, c)) \) and \( \text{lcm}(\text{gcd}(a, b), \text{gcd}(a, c)) \) are nonnegative integers\(^{158}\)). This solves Exercise 2.13.11(a).

(b) Exercise 2.13.11(b) can be above in the same way as we solved Exercise 2.13.11(a) above, after making the obvious changes (i.e., all minima should be replaced by maxima and vice versa; all inequalities in the proof of Claim 1 need to be flipped; all gcds should be replaced by lcms and vice versa). For example, instead of Claim 1, we now need the following claim (with an analogous proof):

**Claim 2:** Let \( i, j, k \in \mathbb{N} \cup \{\infty\} \) be arbitrary. Then,

\[
\max \{i, \min\{j, k\}\} = \min \{\max\{i, j\}, \max\{i, k\}\}.
\]

\[\square\]

**Second solution to Exercise 2.13.11 (a)** We are in one of the following two cases:

**Case 1:** The two integers \( b, c \) are all 0.

**Case 2:** The two integers \( b, c \) are not all 0.

Let us first consider Case 1. In this case, the two integers \( b, c \) are all 0. In other words, \( b = 0 \) and \( c = 0 \). Clearly, \( |a| = |a| \). Thus, Exercise 2.11.1(e) (applied to \( |a| \) and \( |a| \) instead of \( a \) and \( b \)) yields \( \text{lcm}(|a|, |a|) = |a| = |a| \) (since \( |a| \geq 0 \)).

**Proposition 2.9.7(a)** yields \( \text{gcd}(a, 0) = \text{gcd}(a) = |a| \). Now,

\[
\text{lcm} \left( \begin{array}{c}
gcd \left( a, \frac{b}{0} \right) \\
gcd \left( a, \frac{c}{0} \right)
\end{array} \right) = \text{lcm} \left( \begin{array}{c}
\text{gcd}(a, 0) \\
\text{gcd}(a, 0)
\end{array} \right) = \text{lcm} \left( \begin{array}{c}
\frac{|a|}{0} \\
\frac{|a|}{0}
\end{array} \right) = |a|.
\]

\(^{158}\)because any gcd and any lcm is a nonnegative integer
On the other hand, \( \text{lcm} \left( \begin{array}{c}
\frac{b}{0} \\
\frac{c}{0}
\end{array} \right) = 0 \) (by Definition 2.11.4, because the integers 0, 0 are not all nonzero). Hence,
\[
gcd \left( a, \text{lcm} \left( \begin{array}{c}
\frac{b}{0} \\
\frac{c}{0}
\end{array} \right) \right) = gcd \left( a, 0 \right) = |a|. \]
Comparing this with \( \text{lcm} \left( gcd \left( a, b \right), gcd \left( a, c \right) \right) = |a| \), we find \( gcd \left( a, \text{lcm} \left( b, c \right) \right) = \text{lcm} \left( gcd \left( a, b \right), gcd \left( a, c \right) \right) \). Thus, Exercise 2.13.11(a) is solved in Case 1.

Let us now consider Case 2. In this case, the two integers \( b, c \) are not all 0. Thus, \( gcd \left( b, c \right) \) is a positive integer (by Definition 2.9.6). Hence, \( gcd \left( b, c \right) > 0 \), so that \( gcd \left( b, c \right) \neq 0 \).

Corollary 2.9.19 (applied to \( a, b \) and \( c \) instead of \( s, a \) and \( b \)) yields \( gcd \left( ab, ac \right) = |a| \cdot gcd \left( b, c \right) \). But (3) (applied to \( x = a \) and \( y = gcd \left( b, c \right) \)) yields \( |a| \cdot gcd \left( b, c \right) | = |a| \cdot gcd \left( b, c \right) \). Comparing this with \( gcd \left( ab, ac \right) = |a| \cdot gcd \left( b, c \right) \), we obtain
\[
|a| gcd \left( b, c \right) = gcd \left( ab, ac \right) = gcd \left( ab, ca \right). \tag{222}
\]
On the other hand, Theorem 2.11.6 (applied to \( b \) and \( c \) instead of \( a \) and \( b \)) yields \( gcd \left( b, c \right) \cdot \text{lcm} \left( b, c \right) = |bc| \).

Now, Corollary 2.9.19 (applied to \( gcd \left( b, c \right), a \) and \( \text{lcm} \left( b, c \right) \) instead of \( s, a \) and \( b \)) yields
\[
gcd \left( gcd \left( b, c \right) a, gcd \left( b, c \right) \text{lcm} \left( b, c \right) \right) = \left| gcd \left( b, c \right) \right| \cdot gcd \left( a, \text{lcm} \left( b, c \right) \right)
\]
\[
= gcd \left( a, gcd \left( b, c \right), bc \right) \tag{223}
\]
(by Exercise 2.9.5(a), applied to \( a \cdot gcd \left( b, c \right) \) and \( bc \) instead of \( a \) and \( b \)).

On the other hand, Exercise 2.9.5(b) (applied to \( a \cdot gcd \left( b, c \right) \) and \( bc \) instead of \( a \) and \( b \)) yields
\[
gcd \left( |a \cdot gcd \left( b, c \right)|, bc \right) = gcd \left( a \cdot gcd \left( b, c \right), bc \right).
\]
Comparing this with (223), we obtain
\[
gcd \left( b, c \right) \cdot gcd \left( a, lcm \left( b, c \right) \right) = gcd \left( \left| a \cdot gcd \left( b, c \right) \right|, bc \right) = gcd \left( gcd \left( ab, ca \right), bc \right). \tag{224}
\]
On the other hand, Theorem 2.9.20(d) (applied to 3 and \((ab, ca, bc)\) instead of \(k\) and \((b_1, b_2, \ldots, b_k)\)) yields

\[
gcd(ab, ca, bc) = gcd(gcd(ab, ca), bc).
\]

Comparing these two equalities, we obtain

\[
gcd(b, c) gcd(a, lcm(b, c)) = gcd(ab, ca, bc).
\]  \hspace{1cm} (224)

But \(\{bc, ca, ab\} = \{ab, ca, bc\}\). Thus, Exercise 2.9.1 (applied to 3, \((bc, ca, ab)\), 3 and \((ab, ca, bc)\) instead of \(k\), \((b_1, b_2, \ldots, b_k)\), \(\ell\) and \((c_1, c_2, \ldots, c_\ell)\)) yields \(gcd(bc, ca, ab) = gcd(ab, ca, bc)\).

Comparing this with (224), we obtain

\[
gcd(b, c) gcd(a, lcm(b, c)) = gcd(bc, ca, ab).
\]

We can divide both sides of this equality by \(gcd(b, c)\) (since \(gcd(b, c) \neq 0\)); thus we obtain

\[
gcd(a, lcm(b, c)) = \frac{gcd(bc, ca, ab)}{gcd(b, c)}.
\]  \hspace{1cm} (225)

On the other hand, the three integers \(a, b, c\) are not all 0 (since the two integers \(b, c\) are not all 0). Thus, \(gcd(a, b, c)\) is a positive integer (by Definition 2.9.6). Hence, \(gcd(a, b, c) > 0\), so that \(gcd(a, b, c) \neq 0\). From \(gcd(a, b, c) \neq 0\) and \(gcd(b, c) \neq 0\), we obtain \(gcd(a, b, c) \cdot gcd(b, c) \neq 0\).

Recall that a \(gcd\) of a finite list of integers is always nonnegative. Hence, the two numbers \(gcd(a, b)\) and \(gcd(a, c)\) are nonnegative. Thus, their product \(gcd(a, b) gcd(a, c)\) is nonnegative as well.

Now, Theorem 2.11.6 (applied to \(gcd(a, b)\) and \(gcd(a, c)\) instead of \(a\) and \(b\)) yields

\[
gcd(gcd(a, b), gcd(a, c)) \cdot lcm(gcd(a, b), gcd(a, c)) = \frac{|gcd(a, b) gcd(a, c)|}{gcd(a, b)} = gcd(a, c).
\]

(by Proposition 2.9.7(b), applied to \(c\) instead of \(b\))

(since \(gcd(a, b) gcd(a, c)\) is nonnegative)

\[
= gcd(a, b) \cdot gcd(a, c) = gcd(a, b, c).
\]  \hspace{1cm} (226)

But \(\{a, b, a, c\} = \{a, b, c\}\). Hence, Exercise 2.9.1 (applied to 4, \((a, b, a, c)\), 3 and \((a, b, c)\) instead of \(k\), \((b_1, b_2, \ldots, b_k)\), \(\ell\) and \((c_1, c_2, \ldots, c_\ell)\)) yields

\[
gcd(a, b, a, c) = gcd(a, b, c).
\]

But Theorem 2.9.25 (applied to 2, \((a, b)\), 2 and \((a, c)\) instead of \(k\), \((b_1, b_2, \ldots, b_k)\), \(\ell\) and \((c_1, c_2, \ldots, c_\ell)\)) yields

\[
gcd(a, b, a, c) = gcd(gcd(a, b), gcd(a, c)).
\]

Comparing these two equalities, we obtain

\[
gcd(gcd(a, b), gcd(a, c)) = gcd(a, b, c).
\]
Now, \((226)\) yields
\[
gcd (c, a) \cdot gcd (a, b) = gcd(gcd (a, b), gcd (a, c)) \cdot lcm (gcd (a, b), gcd (a, c))
\]
\[
= gcd (a, b, c) \cdot lcm (gcd (a, b), gcd (a, c)).
\]
Dividing both sides of this equality by \(gcd (a, b, c)\) (we can do this, since \(gcd (a, b, c) \neq 0\)), we obtain
\[
gcd (c, a) \cdot gcd (a, b) = gcd (a, b, c) \cdot lcm (gcd (a, b), gcd (a, c)).
\]
(\(227\))

However, Exercise \([2.10.18]\) yields
\[
gcd (b, c) \cdot gcd (c, a) \cdot gcd (a, b) = gcd (a, b, c) \cdot gcd bc ca ab.
\]

We can divide both sides of this equality by \(gcd (a, b, c)\) (since \(gcd (a, b, c) \neq 0\)); we thus obtain
\[
gcd (c, a) \cdot gcd (a, b) = gcd (b, c) \cdot gcd (bc, ca, ab).
\]
Comparing this with \((227)\), we find
\[
lcm (gcd (a, b), gcd (a, c)) = \frac{gcd (bc, ca, ab)}{gcd (b, c)} = gcd (a, lcm (b, c))
\]
(by \((225)\)). Thus, Exercise \([2.13.11] (a)\) is solved in Case 2.

We have now solved Exercise \([2.13.11] (a)\) in each of the two Cases 1 and 2. Thus, Exercise \([2.13.11] (a)\) is solved in all cases.

[Remark: The above solution to Exercise \([2.13.11] (a)\) is (an expanded version of) Bill Dubuque’s post \(https://math.stackexchange.com/a/147992/\). Note that Dubuque uses the notations \((x_1, x_2, \ldots, x_k)\) and \([x_1, x_2, \ldots, x_k]\) for what we call \(gcd (x_1, x_2, \ldots, x_k)\) and \(lcm (x_1, x_2, \ldots, x_k)\).]

(b) We are in one of the following two cases:

Case 1: We have \(a = 0\).

Case 2: We have \(a \neq 0\).

Let us first consider Case 1. In this case, we have \(a = 0\). Thus, the two integers \(a, gcd (b, c)\) are not all nonzero. Hence, Definition \([2.11.4]\) yields \(lcm (a, gcd (b, c)) = 0\). Also, the two integers \(a, b\) are not all nonzero (since \(a = 0\)). Hence, \(lcm (a, b) = 0\) (again by Definition \([2.11.4]\)). Similarly, \(lcm (a, c) = 0\). Now,
\[
gcd \left( \frac{lcm (a, b), lcm (a, c)}{0, 0} \right) = gcd (0, 0) = 0
\]
(by Definition \([2.9.6]\) since all of the integers \(0, 0\) are 0). Comparing this with \(lcm (a, gcd (b, c)) = 0\), we obtain \(lcm (a, gcd (b, c)) = gcd (lcm (a, b), lcm (a, c))\). Thus, Exercise \([2.13.11] (b)\) is solved in Case 1.

Let us now consider Case 2. In this case, we have \(a \neq 0\). Hence, the two integers \(a, b\) are not all 0. Thus, \(gcd (a, b)\) is a positive integer (by Definition \([2.9.6]\)). Similarly, \(gcd (a, c)\) is a
positive integer. Also, the three integers \( a, b, c \) are not all 0 (since \( a \neq 0 \)). Thus, \( \gcd (a, b, c) \) is a positive integer (by Definition 2.9.6). Hence, \( \gcd (a, b, c) \neq 0 \).

The numbers \( \gcd (a, b) \) and \( \gcd (a, c) \) are positive integers. Hence, their product \( \gcd (a, b) \cdot \gcd (a, c) \) is a positive integer as well. Let us denote this positive integer by \( g \). Thus,

\[
g = \gcd (a, b) \cdot \gcd (a, c) = \gcd (a, b) \cdot \gcd (c, a)
\]

(by Proposition 2.9.7 (b), applied to \( c \) instead of \( b \))

\[
g = \gcd (c, a) \cdot \gcd (a, b) .
\]

Also, \( g \neq 0 \) (since \( g \) is positive).

Theorem 2.11.6 yields \( \gcd (a, b) \cdot \operatorname{lcm} (a, b) = |ab| \). Also, Theorem 2.11.6 (applied to \( c \) instead of \( b \)) yields \( \gcd (a, c) \cdot \operatorname{lcm} (a, c) = |ac| \).

Corollary 2.9.19 (applied to \( g \), \( \operatorname{lcm} (a, b) \) and \( \operatorname{lcm} (a, c) \) instead of \( s, a \) and \( b \)) yields

\[
\gcd (g \operatorname{lcm} (a, b), g \operatorname{lcm} (a, c)) = \left\lceil \frac{|g| \gcd (\operatorname{lcm} (a, b), \operatorname{lcm} (a, c))}{g} \right\rceil
\]

(since \( g \) is positive)

\[
g \gcd (\operatorname{lcm} (a, b), \operatorname{lcm} (a, c)) .
\]

Hence,

\[
g \gcd (\operatorname{lcm} (a, b), \operatorname{lcm} (a, c))
\]

\[
= \gcd \left( g \operatorname{lcm} (a, b), g \operatorname{lcm} (a, c) \right)
\]

\[
= \gcd \left( \gcd (c, a) \cdot \gcd (a, b) \cdot \operatorname{lcm} (a, b), \gcd (a, b) \cdot \gcd (a, c) \cdot \operatorname{lcm} (a, c) \right)
\]

\[
= \gcd \left( \gcd (c, a) \cdot |ab|, \gcd (a, b) \cdot |ac| \right) .
\]

(229)

On the other hand, Corollary 2.9.19 (applied to \( ab, c \) and \( a \) instead of \( s, a \) and \( b \)) yields

\[
\gcd (abc, aba) = |ab| \gcd (c, a) = \gcd (c, a) \cdot |ab| .
\]

(230)

Also, Corollary 2.9.19 (applied to \( ac, a \) and \( b \) instead of \( s, a \) and \( b \)) yields

\[
\gcd (aca, acb) = |ac| \gcd (a, b) = \gcd (a, b) \cdot |ac| .
\]

(231)

Also, \( \{abc, aba, aca, acb \} = \{abc, aab, aca, abc \} = \{abc, aca, aab \} \). Hence, Exercise 2.9.1 (applied to \( 4, (abc, aba, aca, acb), 3 \) and \( (abc, aca, aab) \) instead of \( k, (b_1, b_2, \ldots, b_k), \ell \) and \( (c_1, c_2, \ldots, c_\ell) \)) yields

\[
\gcd (abc, aba, aca, acb) = \gcd (abc, aca, aab) = |a| \gcd (bc, ca, ab)
\]
(by Exercise 2.9.7 applied to $a, 3$ and $(bc, ca, ab)$ instead of $s, k$ and $(a_1, a_2, \ldots, a_k)$). Hence,

$$|a| \gcd (bc, ca, ab) = \gcd (abc, aba,aca, acb) = \gcd \left( \frac{\gcd (abc, aba), \gcd (aca, acb)}{= \gcd (c, a) - |ab| \text{ (by (230))}} \right)$$

(by Theorem 2.9.25 applied to 2, $(abc, aba), 2$ and $(aca, acb)$) instead of $k, (b_1, b_2, \ldots, b_k), \ell$ and $(c_1, c_2, \ldots, c_\ell)$

$$= \gcd (\gcd (c, a) \cdot |ab|, \gcd (a, b) \cdot |ac|) = g \gcd (\text{lcm}(a, b), \text{lcm}(a, c)) \quad \text{(by (229)).}$$

(232)

We can divide both sides of this equality by $g$ (since $g \neq 0$); thus we obtain

$$\frac{|a| \gcd (bc, ca, ab)}{g} = \gcd (\text{lcm}(a, b), \text{lcm}(a, c)). \quad \text{(232)}$$

**Proposition 2.9.7 (b)** (applied to gcd $(b, c)$ instead of $b$) yields

$$\gcd (a, \gcd (b, c)) = \gcd (g \gcd (b, c), a). \quad \text{But Theorem 2.9.20 (d) (applied to 3 and $(b, c, a)$ instead of $k$ and $(b_1, b_2, \ldots, b_k)$) yields $\gcd (b, c, a) = \gcd (\gcd (b, c), a)$}. \quad \text{Comparing these two equalities, we obtain}

$$\gcd (a, \gcd (b, c)) = \gcd (b, c, a). \quad \text{(233)}$$

But $\{b, c, a\} = \{a, b, c\}$. Thus, Exercise 2.9.1 (applied to 3, $(b, c, a), 3$ and $(a, b, c)$ instead of $k, (b_1, b_2, \ldots, b_k), \ell$ and $(c_1, c_2, \ldots, c_\ell)$) yields $\gcd (b, c, a) = \gcd (a, b, c)$. Thus, (233) becomes

$$\gcd (a, \gcd (b, c)) = \gcd (b, c, a) = \gcd (a, b, c). \quad \text{(234)}$$

The gcd of any list of integers is a nonnegative integer. Thus, $\gcd (b, c)$ is a nonnegative integer. Now, Theorem 2.11.6 (applied to gcd $(b, c)$ instead of $b$) yields

$$\gcd (a, \gcd (b, c)) \cdot \text{lcm} (a, \gcd (b, c))$$

$$= |a \gcd (b, c)| = |a| \cdot \left| \gcd (b, c) \right| \quad \text{(by (3), applied to $x = a$ and $y = \gcd (b, c)$)}$$

$$= |a| \gcd (b, c).$$

In view of (234), this rewrites as

$$\gcd (a, b, c) \cdot \text{lcm} (a, \gcd (b, c)) = |a| \gcd (b, c). \quad \text{(235)}$$

Multiplying both sides of this equality by $g$, we obtain

$$\gcd (a, b, c) \cdot \text{lcm} (a, \gcd (b, c)) \cdot g = |a| \gcd (b, c) \cdot g$$

(by Exercise 2.10.18)

$$= |a| \gcd (b, c) \cdot \gcd (c, a) \cdot \gcd (a, b) \quad \text{(by Exercise 2.10.18)}$$

$$= |a| \gcd (b, c) \cdot \gcd (bc, ca, ab) \quad \text{(by Exercise 2.10.18)}$$

$$= \gcd (a, b, c) \cdot |a| \gcd (bc, ca, ab) \cdot |a| \gcd (bc, ca, ab).$$
We can divide both sides of this equality by \( \text{gcd} (a, b, c) \) (since \( \text{gcd} (a, b, c) \neq 0 \)); we thus obtain
\[
\text{lcm} (a, \text{gcd} (b, c)) \cdot g = |a| \text{gcd} (bc, ca, ab).
\]
We can divide both sides of this equality by \( g \) (since \( g \neq 0 \)); thus we find
\[
\text{lcm} (a, \text{gcd} (b, c)) = \frac{|a| \text{gcd} (bc, ca, ab)}{g} = \text{gcd} (\text{lcm} (a, b), \text{lcm} (a, c))
\]
(by (232)). Thus, Exercise 2.13.11(b) is solved in Case 2.

We have now solved Exercise 2.13.11(b) in each of the two Cases 1 and 2. Thus, Exercise 2.13.11(b) is solved in all cases.

### 6.60. Solution to Exercise 2.13.12

**Solution to Exercise 2.13.12** We have \( a^2 \equiv 1 \mod p \). In other words, \( p \mid a^2 - 1 = (a - 1)(a + 1) \). Hence, Theorem 2.13.6 (applied to \( a - 1 \) and \( a + 1 \) instead of \( a \) and \( b \)) yields that \( p \mid a - 1 \) or \( p \mid a + 1 \). In view of the logical equivalences
\[
(p \mid a - 1) \iff (a \equiv 1 \mod p)
\]
and
\[
\left( p \mid \frac{a + 1}{a-1} \right) \iff (p \mid a - (-1)) \iff (a \equiv -1 \mod p),
\]
this rewrites as follows: \( a \equiv 1 \mod p \) or \( a \equiv -1 \mod p \). This solves Exercise 2.13.12.

### 6.61. Solution to Exercise 2.14.1

**Solution to Exercise 2.14.1** Let \( M \) be the set \( \{1p, 2p, \ldots, pk-1p\} = \{cp \mid c \in \{1, 2, \ldots, pk-1\}\} \). This set \( M \) has \( pk-1 \) elements (since the \( pk-1 \) numbers \( 1p, 2p, \ldots, pk-1p \) are all distinct). In other words, \( |M| = pk-1 \). Also, \( M \subseteq \{1, 2, \ldots, pk\} \).\(^\text{159}\) Hence,
\[
\left|\left\{1, 2, \ldots, pk\right\} \setminus M\right| = \left|\left\{1, 2, \ldots, pk\right\} - \left\{1, 2, \ldots, pk-1\right\}\right| = pk - pk-1. \quad (236)
\]

Next, we claim that
\[
\left\{i \in \{1, 2, \ldots, pk\} \mid i \perp pk\right\} \subseteq \left\{1, 2, \ldots, pk\right\} \setminus M. \quad (237)
\]

\(^{159}\)Proof. Let \( m \in M \). Thus, \( m \in M = \{1p, 2p, \ldots, pk-1p\} \); in other words, \( m = cp \) for some \( c \in \{1, 2, \ldots, pk-1\} \). Consider this \( c \). From \( c \in \{1, 2, \ldots, pk-1\} \), we obtain \( c \leq pk-1 \). We can multiply this inequality by \( p \) (since \( p > 0 \)) and thus obtain \( cp \leq pk-1p = pk \). Also, \( cp \) is a positive integer (since \( c \) and \( p \) are positive integers). Thus, \( cp \) is a positive integer and \( \leq pk \). In other words, \( cp \in \{1, 2, \ldots, pk\} \). Thus, \( m = cp \in \{1, 2, \ldots, pk\} \).

Now, forget that we fixed \( m \). We thus have shown that \( m \in \{1, 2, \ldots, pk\} \) for each \( m \in M \). In other words, \( M \subseteq \{1, 2, \ldots, pk\} \).
[Proof of (237):] Let \( a \in \{ i \in \{1, 2, \ldots, p^k \} \mid i \perp p^k \} \). In other words, \( a \) is an element of \( \{1, 2, \ldots, p^k \} \) and satisfies \( a \perp p^k \).

Exercise 2.13.3 shows that \( a \perp p^k \) holds if and only if \( p \nmid a \). Hence, \( p \nmid a \) (since \( a \perp p^k \)), and therefore \( a \notin M \) Combining \( a \in \{1, 2, \ldots, p^k \} \) with \( a \notin M \), we obtain \( a \in \{1, 2, \ldots, p^k \} \setminus M \).

Now, forget that we fixed \( a \). Thus, we have shown that \( a \in \{1, 2, \ldots, p^k \} \setminus M \) for each \( a \in \{ i \in \{1, 2, \ldots, p^k \} \mid i \perp p^k \} \). In other words, \( \{ i \in \{1, 2, \ldots, p^k \} \mid i \perp p^k \} \subseteq \{1, 2, \ldots, p^k \} \setminus M \). This proves (237).

Furthermore, we have

\[
\{1, 2, \ldots, p^k \} \setminus M \subseteq \{ i \in \{1, 2, \ldots, p^k \} \mid i \perp p^k \}.
\] (238)

[Proof of (238):] Let \( a \in \{1, 2, \ldots, p^k \} \setminus M \). In other words, \( a \in \{1, 2, \ldots, p^k \} \) and \( a \notin M \).

We have \( p \nmid a \) but Exercise 2.13.3 shows that \( a \perp p^k \) holds if and only if \( p \nmid a \). Hence, \( a \perp p^k \) (since \( p \nmid a \)). Now, we know that \( a \) is an \( i \in \{1, 2, \ldots, p^k \} \) satisfying \( i \perp p^k \) (since \( a \in \{1, 2, \ldots, p^k \} \) and \( a \perp p^k \)). In other words, \( a \in \{ i \in \{1, 2, \ldots, p^k \} \mid i \perp p^k \} \).

Now, forget that we fixed \( a \). Thus, we have shown that \( a \in \{ i \in \{1, 2, \ldots, p^k \} \mid i \perp p^k \} \) for each \( a \in \{1, 2, \ldots, p^k \} \setminus M \). In other words, \( \{1, 2, \ldots, p^k \} \setminus M \subseteq \{ i \in \{1, 2, \ldots, p^k \} \mid i \perp p^k \} \).

This proves (238).

Combining (237) with (238), we obtain

\[
\{ i \in \{1, 2, \ldots, p^k \} \mid i \perp p^k \} = \{1, 2, \ldots, p^k \} \setminus M.
\]

Now, (54) (applied to \( n = p^k \)) yields

\[
\phi\left(p^k\right) = \left| i \in \{1, 2, \ldots, p^k \} \mid i \perp p^k \right| = \left| \{1, 2, \ldots, p^k \} \setminus M \right|
\]

\[
= p^k - p^{k-1} \quad \text{(by (236))}
\]

\[
= pp^{k-1} - p^{k-1} = (p - 1) p^{k-1}.
\]

This solves Exercise 2.14.1 \( \square \)

\[160\text{Proof.} \] Assume the contrary. Thus, \( a \in M = \{1p, 2p, \ldots, p^{k-1}p\} \). In other words, \( a = cp \) for some \( c \in \{1, 2, \ldots, p^{k-1}\} \). Consider this \( c \). Clearly, \( c \) is an integer; thus, from \( a = cp = pc \), we obtain \( p \mid a \). But this contradicts \( p \nmid a \). This contradiction shows that our assumption was false, qed.

\[161\text{Proof.} \] Assume the contrary. Thus, \( p \mid a \). In other words, \( a = pc \) for some integer \( c \). Consider this \( c \). We have \( a = pc \), thus \( c = a/p \) (since \( p \) is nonzero). Also, \( c = a/p > 0 \) (since \( a \) and \( p \) are positive). Hence, \( c \) is a positive integer. Furthermore, from \( a \in \{1, 2, \ldots, p^k \} \), we obtain \( a \leq p^k \) and thus \( a/p \leq p^k/p = p^{k-1} \), so that \( c = a/p \leq p^{k-1} \). Thus, \( c \in \{1, 2, \ldots, p^{k-1}\} \) (since \( c \) is a positive integer) and therefore \( cp \in \{1p, 2p, \ldots, p^{k-1}p\} = M \) (by the definition of \( M \)). Now, \( a = pc = cp \in M \); but this contradicts \( a \notin M \). This contradiction shows that our assumption was false, qed.
6.62. Solution to Exercise 2.14.2

Solution to Exercise 2.14.2 (a) If \( I \) is any set, and if we are given a statement \( A(i) \) for each \( i \in I \), then
\[
\{ i \in I \mid \text{we don’t have } A(i) \} = I \setminus \{ i \in I \mid A(i) \}.
\]
(This is one of the basic rules of sets and logic.)

Now,
\[
\begin{align*}
\{ i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \} \\
= \{ \{1,2,\ldots,n\} \setminus \{ i \in \{1,2,\ldots,n\} \mid i \perp n \} \} \\
= |\{1,2,\ldots,n\} \setminus \{ i \in \{1,2,\ldots,n\} \mid i \perp n \}| \\
= \phi(n) \\
= n - \phi(n).
\end{align*}
\]

This solves Exercise 2.14.2 (a).

(b) Exercise 2.14.2 (a) yields
\[
|\{ i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \}| \geq 0
\]
(since \( |X| \geq 0 \) for any finite set \( X \)). This solves Exercise 2.14.2 (b).

(c) We have \( d \mid n \) and \( n \neq 0 \) (since \( n \) is positive). Thus, Proposition 2.2.3 (a) (applied to \( a = d \) and \( b = n \)) yields \( |d| \leq |n| = n \) (since \( n \) is positive). But \( |d| = d \) (since \( d \) is positive).
Hence, \( d = |d| \leq n \).

It is easy to see that
\[
\{ i \in \{1,2,\ldots,d\} \mid \text{we don’t have } i \perp d \} \\
\subseteq \{ i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \}.
\]

[Proof of (240): Let \( j \in \{ i \in \{1,2,\ldots,d\} \mid \text{we don’t have } i \perp d \} \). Thus, \( j \) is an \( i \in \{1,2,\ldots,d\} \) such that we don’t have \( i \perp d \). In other words, \( j \) is an element of \( \{1,2,\ldots,d\} \) and has the property that we don’t have \( j \perp d \).

Now, recall that \( j \) is an element of \( \{1,2,\ldots,d\} \). Hence, \( j \in \{1,2,\ldots,d\} \subseteq \{1,2,\ldots,n\} \) (since \( d \leq n \)). In other words, \( j \) is an element of \( \{1,2,\ldots,n\} \).

Also, we don’t have \( j \perp d \). Hence, we don’t have \( j \perp n \).]

\footnote{Proof: Assume the contrary. Thus, \( j \perp n \). In other words, \( \gcd(j,n) = 1 \). But \( j \mid j \) and \( d \mid n \). Hence, Exercise 2.9.4 (applied to \( j, d \), \( j \) and \( n \) instead of \( a_1, a_2, b_1 \) and \( b_2 \)) yields \( \gcd(j,d) \mid \gcd(j,n) = 1 \). But \( \gcd(j,d) \) is a nonnegative integer (since the \( \gcd \) of any list of integers is a nonnegative integer). Hence, Exercise 2.2.5 (applied to \( g = \gcd(j,d) \)) yields \( \gcd(j,d) = 1 \) (since \( \gcd(j,d) \mid 1 \)). In other words, \( j \perp d \). This contradicts the fact that we don’t have \( j \perp d \). This contradiction shows that our assumption was wrong. \( \q.e.d. \).} Now, we know that \( j \) is an element of \( \{1,2,\ldots,n\} \) and has the property that we don’t have \( j \perp n \).
other words, \( j \) is an \( i \in \{1,2,\ldots,n\} \) such that we don’t have \( i \perp n \). In other words, \( j \in \{i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \} \).

Now, forget that we fixed \( j \). We thus have proven that \( j \in \{i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \} \) for each \( j \in \{i \in \{1,2,\ldots,d\} \mid \text{we don’t have } i \perp d \} \). In other words,

\[
\{i \in \{1,2,\ldots,d\} \mid \text{we don’t have } i \perp d \} \subseteq \{i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \}.
\]

This proves (240).]

Now, Exercise 2.14.2 (a) (applied to \( d \) instead of \( n \)) yields

\[
d - \phi(d) = \left\lfloor \frac{i \in \{1,2,\ldots,d\} \mid \text{we don’t have } i \perp d} \right\lfloor \subseteq \left\lfloor i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \right\rfloor \quad (\text{by } 240)
\]

\[
\leq \left\lfloor \{i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \} \right\rfloor = n - \phi(n)
\]

(by Exercise 2.14.2 (a)). Thus, Exercise 2.14.2 (c) is solved.

(d) In our solution to Exercise 2.14.2 (c) above, we have shown that \( d \leq n \). Combining this with \( d \neq n \), we obtain \( d < n \). Thus, \( n > d \). Also, \( d \geq 1 \) (since \( d \) is a positive integer).

Hence, \( n > d \geq 1 \).

Define two finite sets

\[
D = \{i \in \{1,2,\ldots,d\} \mid \text{we don’t have } i \perp d \} \quad \text{and} \quad N = \{i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \}.
\]

In our solution to Exercise 2.14.2 (c) above, we have proven (240). Thus,

\[
D = \{i \in \{1,2,\ldots,d\} \mid \text{we don’t have } i \perp d \} \subseteq \{i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \} \quad (\text{by } 240)
\]

\[
= N.
\]

Now, we have \( n \in N \).

[Proof: We have \( n \in \{1,2,\ldots,n\} \) (since \( n \geq 1 \)), but we don’t have \( n \perp n \).] Thus, \( n \) is an element of \( \{1,2,\ldots,n\} \) such that we don’t have \( n \perp n \). In other words, \( n \) is an \( i \in \{1,2,\ldots,n\} \) such that we don’t have \( i \perp n \). Hence,

\[
n \in \{i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \}.
\]

This rewrites as \( n \in N \) (since \( N = \{i \in \{1,2,\ldots,n\} \mid \text{we don’t have } i \perp n \} \)). Qed.]

Now, we claim that \( D \neq N \).

[Proof: Assume the contrary. Thus, \( D = N \). Hence, \( N = D \). Now,

\[
n \in N = D = \{i \in \{1,2,\ldots,d\} \mid \text{we don’t have } i \perp d \} \subseteq \{1,2,\ldots,d\},
\]

\]

[Proof. Assume the contrary. Thus, \( n \perp n \). In other words, \( \gcd(n,n) = 1 \). But Proposition 2.9.7 (i) (applied to \( a = n \) and \( b = n \)) yields \( \gcd(n,n) = |n| \) (since \( n \mid n \)). Hence, \( \gcd(n,n) = |n| = n \) (since \( n > 1 > 0 \)). Therefore, \( n = \gcd(n,n) = 1 \). This contradicts \( n > 1 \). This contradiction shows that our assumption was wrong. Qed.]
so that $n \leq d$. This contradicts $d < n$. This contradiction shows that our assumption was false. Hence, $D \neq N$ is proven.]

Now, $D$ is a subset of $N$ (since $D \subseteq N$), and hence is a proper subset of $N$ (since $D \neq N$). Hence, $|D| < |N|$ (because if $X$ is a proper subset of a finite set $Y$, then $|X| < |Y|$). But Exercise 2.14.2 (a) applies to $d$ instead of $n$) yields

$$d - \phi(d) = \left| \left\{ i \in \{1, 2, \ldots, d\} \mid \text{we don't have } i \perp d \right\} \right| = |D|$$

(by Exercise 2.14.2 (a)). Thus, Exercise 2.14.2 (d) is solved. 

### 6.63. Solution to Exercise 2.14.4

**Solution to Exercise 2.14.4** The following solution is mostly a calque of the proof of Proposition 2.14.7 we gave above, but using subtraction instead of division and using the relation “coprime” instead of “divides”.

Clearly, $n$ is a positive integer (since $n > 2$), so that $\phi(n)$ is well-defined. Let $C = \{i \in \{1, 2, \ldots, n\} \mid i \perp n\}$. An element $c$ of $C$ is said to be

- **small** if $c < n - c$;
- **medium** if $c = n - c$;
- **large** if $c > n - c$.

Now, it is easy to see that the set $C$ has no medium elements. In other words,

$$|\{\text{medium elements of } C\}| = 0.$$

Furthermore, if $c \in C$, then $n - c \in C$. This allows us to define a map

$$F : C \to C,$$

$$c \mapsto n - c.$$

**Proof.** Let $c$ be a medium element of $C$. We shall derive a contradiction.

We have $c \in C = \{i \in \{1, 2, \ldots, n\} \mid i \perp n\}$. In other words, $c$ is an element of $\{1, 2, \ldots, n\}$ and satisfies $c \perp n$. Clearly, the integer $c$ is positive (since $c \in \{1, 2, \ldots, n\}$) and we have $\gcd(c, n) = 1$ (since $c \perp n$).

Furthermore, $c$ is medium; in other words, $c = n - c$. Thus, $2c = n$, so that $n = 2c = c \cdot 2$. Hence, $c \mid n$. Thus, Proposition 2.9.7 (i) (applied to $a = c$ and $b = n$) yields $\gcd(c, n) = |c| = c$ (since $c$ is positive). But from $2c = n$, we also obtain $c = n/2 > 2/2$ (since $n > 2$). Thus, $\gcd(c, n) = c > 2/2 = 1$. This contradicts $\gcd(c, n) = 1$.

Forget that we fixed $c$. We thus have obtained a contradiction for each medium element $c$ of $C$. Thus, there are no such elements. In other words, the set $C$ has no medium elements.

**Proof.** Let $c \in C$.

We have $c \in C = \{i \in \{1, 2, \ldots, n\} \mid i \perp n\}$. In other words, $c$ is an element of $\{1, 2, \ldots, n\}$ and satisfies $c \perp n$. From $c \perp n$, we obtain $n \perp c$ (by Proposition 2.10.4) and thus $\gcd(n, c) = 1$. 

---
This map $F$ has the property that $F \circ F = \text{id}$, because each $c \in C$ satisfies

$$(F \circ F)(c) = F\left(\begin{array}{c} F(c) \\ \overline{\text{by the definition of } F} \end{array}\right) = F(n - c)$$

$$= n - (n - c) \quad \text{(by the definition of } F)$$

$$= c = \text{id}(c).$$

Hence, the map $F$ is inverse to itself. Thus, the map $F$ is invertible, i.e., is a bijection. Hence, the map

$$F^+ : \{\text{small elements of } C\} \to \{\text{large elements of } C\},$$

$$c \mapsto F(c)$$

is well-defined. Similarly, the map

$$F^- : \{\text{large elements of } C\} \to \{\text{small elements of } C\},$$

$$c \mapsto F(c)$$

is well-defined. These two maps $F^+$ and $F^-$ are both restrictions of the map $F$, and thus are mutually inverse (since the map $F$ is inverse to itself). Hence, the map $F^+$ is invertible, i.e., is a bijection. Thus, we have found a bijection from $\{\text{small elements of } C\}$ to $\{\text{large elements of } C\}$ (namely, $F^+$). Therefore,

$$|\{\text{small elements of } C\}| = |\{\text{large elements of } C\}|.$$
Thus, 2 \mid \phi(n) (since \{| \text{large elements of } C \} \) is an integer). In other words, \( \phi(n) \) is even. This solves Exercise 2.14.4.


### 6.64. Solution to Exercise 2.14.5

Our below solution to Exercise 2.14.5 imitates the proof of Proposition 2.10.12.

**Solution to Exercise 2.14.5** We do not have \( n \perp n \) \(^{167} \) Hence,

\[
\{ i \in \{1, 2, \ldots, n \} \mid i \perp n \} = \{ i \in \{1, 2, \ldots, n-1 \} \mid i \perp n \} \tag{241}
\]

\(^{168} \) Now, (54) yields

\[
\phi(n) = \left| \{ i \in \{1, 2, \ldots, n \} \mid i \perp n \} \right| = \left| \{ i \in \{1, 2, \ldots, n-1 \} \mid i \perp n \} \right| \tag{242}
\]

\(^{167} \) Proof. Assume the contrary. Thus, \( n \perp n \). In other words, \( \gcd(n, n) = 1 \). But Proposition 2.9.7 (i) (applied to \( a = n \) and \( b = n \)) yields \( \gcd(n, n) = |n| \) (since \( n \mid n \)). Hence, \( \gcd(n, n) = |n| = n \) (since \( n > 1 > 0 \)). Therefore, \( n = \gcd(n, n) = 1 \). This contradicts \( n > 1 \). This contradiction shows that our assumption was wrong. QED.

\(^{168} \) Proof of (241): Let \( j \in \{ i \in \{1, 2, \ldots, n \} \mid i \perp n \} \). Thus, \( j \) is an element \( i \) of \( \{1, 2, \ldots, n\} \) satisfying \( i \perp n \). In other words, \( j \) is an element of \( \{1, 2, \ldots, n\} \) and satisfies \( j \perp n \). We have \( j \perp n \). If we had \( j = n \), then this would rewrite as \( n \perp n \), which would contradict the fact that we do not have \( n \perp n \). Thus, we cannot have \( j = n \). In other words, we have \( j \neq n \). Combining \( j \in \{1, 2, \ldots, n\} \) with \( j \neq n \), we obtain \( j \in \{1, 2, \ldots, n\} \setminus \{n\} = \{1, 2, \ldots, n-1\} \). Hence, \( j \) is an element of \( \{1, 2, \ldots, n-1\} \) and satisfies \( j \perp n \). In other words, \( j \) is an element \( i \) of \( \{1, 2, \ldots, n-1\} \) satisfying \( i \perp n \). In other words, \( j \in \{1, 2, \ldots, n-1\} \) and \( i \perp n \).

Now, forget that we fixed \( j \). We thus have proven that \( j \in \{ i \in \{1, 2, \ldots, n-1\} \mid i \perp n \} \) for each \( j \in \{ i \in \{1, 2, \ldots, n\} \mid i \perp n \} \). In other words,

\[
\{ i \in \{1, 2, \ldots, n\} \mid i \perp n \} \subseteq \{ i \in \{1, 2, \ldots, n-1\} \mid i \perp n \}.
\]

Combining this with

\[
\left\{ i \in \{1, 2, \ldots, n-1\} \mid i \perp n \right\} \subseteq \{ i \in \{1, 2, \ldots, n\} \mid i \perp n \},
\]

we obtain \( \{ i \in \{1, 2, \ldots, n\} \mid i \perp n \} = \{ i \in \{1, 2, \ldots, n-1\} \mid i \perp n \} \). This proves (241).
(by (241)).

On the other hand, for each $k \in \mathbb{Z}$, we have the logical equivalence
\[(n - k \perp n) \iff (k \perp n)\]  \hspace{1cm} (243)
(because Exercise 2.10.7 (applied to $a = k$ and $b = n$) shows that $n - k \perp n$ holds if and only if $k \perp n$). Now,

\[
2 \cdot \sum_{i \in \{1, 2, \ldots, n\}; \ i \perp n} i = 2 \cdot \sum_{k \in \{1, 2, \ldots, n\}; \ k \perp n} k \hspace{1cm} \text{(here, we have renamed the summation index } i \text{ as } k) \hspace{1cm} \text{(by (241))}
\]

\[
= \sum_{k \in \{i \in \{1, 2, \ldots, n\}; \ i \perp n\}} k \hspace{1cm} \text{(by (241))}
\]

\[
= 2 \cdot \sum_{k \in \{i \in \{1, 2, \ldots, n-1\}; \ i \perp n\}} k \hspace{1cm} \text{(by the equivalence (243))}
\]

\[
= \sum_{k \in \{1, 2, \ldots, n-1\}; \ k \perp n} k + \sum_{k \in \{1, 2, \ldots, n-1\}; \ k \perp n} (n - k) \hspace{1cm} \text{(here, we have substituted } n - k \text{ for } k \text{ in the second sum)}
\]

\[
= \sum_{k \in \{1, 2, \ldots, n-1\}; \ k \perp n} n = \sum_{i \in \{1, 2, \ldots, n-1\}; \ i \perp n} \phi(n) \cdot n = n \phi(n).
\]

Dividing this equality by 2, we obtain
\[
\sum_{i \in \{1, 2, \ldots, n\}; \ i \perp n} i = n \phi(n) / 2.
\]

This solves Exercise 2.14.5.

[Remark: Alternatively, instead of doubling the sum we wanted to compute, we could have paired its addends up with each other: every $i$ gets paired with the respective $n - i$. But this is slightly messy, since $n/2$ can happen to be a term of the sum (this happens when $n = 2$), which necessitates a separate argument.]
6.65. Solution to Exercise 2.15.2

Solution to Exercise 2.15.2 From $u \equiv v \pmod{p - 1}$, we obtain $v \equiv u \pmod{p - 1}$. Likewise, the congruence $a^v \equiv a^u \pmod{p}$, which we need to prove, is clearly equivalent to $a^v \equiv a^u \pmod{p}$. Hence, the numbers $u$ and $v$ play symmetric roles in Exercise 2.15.2 Thus, we can WLOG assume that $u \leq v$ (since otherwise, we can simply swap $u$ with $v$). Assume this.

We have $v \equiv u \pmod{p - 1}$. In other words, $p - 1 \mid v - u$. Thus, there exists an integer $c$ such that $v - u = (p - 1) c$. Consider this $c$. Thus, $(p - 1) c = v - u \geq v - v = 0$. But $p > 1$ (since $p$ is a prime), thus $p - 1 > 0$. Hence, we can divide the inequality $(p - 1) c \geq 0$ by $p - 1$, and obtain $c \geq 0$. Hence, $c \in \mathbb{N}$.

Now, Theorem 2.15.7 (a) yields $a^{p - 1} \equiv 1 \pmod{p}$. We can take this congruence to the $c$-th power (since $c \in \mathbb{N}$), and obtain $(a^{p - 1})^c \equiv 1^c = 1 \pmod{p}$. But $(a^{p - 1})^c = a^{(p - 1)c} = a^{v - u}$ (since $(p - 1) c = v - u$). Hence, $a^{v - u} \equiv (a^{p - 1})^c \equiv 1 \pmod{p}$. Multiplying this congruence by the obvious congruence $a^u \equiv a^u \pmod{p}$, we obtain $a^{v - u} a^u \equiv a^1 a^u$ mod $p$. Hence, $a^u \equiv a^{v - u} a^u = a^{(v - u) + u} = a^v$ mod $p$. This solves Exercise 2.15.2 

6.66. Solution to Exercise 2.15.4

Solution to Exercise 2.15.4 Theorem 2.15.7 yields

$$(p - 1)! \equiv -1 \equiv p - 1 \pmod{p}$$

(since $p - 1 \equiv -1 \pmod{p}$). In other words, $p \mid (p - 1)! - (p - 1)$.

On the other hand, $p > 1$ (since $p$ is a prime). Now, the definition of $(p - 1)!$ yields

$$(p - 1)! = 1 \cdot 2 \cdot \cdots \cdot (p - 1) = (1 \cdot 2 \cdot \cdots \cdot (p - 2)) \cdot (p - 1).$$

Subtracting $p - 1$ from both sides of this equality, we obtain

$$(p - 1)! - (p - 1) = (1 \cdot 2 \cdot \cdots \cdot (p - 2)) \cdot (p - 1) - (p - 1)$$
$$= (1 \cdot 2 \cdot \cdots \cdot (p - 2) - 1) \cdot (p - 1)$$
$$= (p - 1) \cdot (1 \cdot 2 \cdot \cdots \cdot (p - 2) - 1).$$

Hence, $p - 1 \mid (p - 1)! - (p - 1)$ (since $1 \cdot 2 \cdot \cdots \cdot (p - 2) - 1$ is an integer).

Now, it is easy to see that $p - 1 \perp p$. Furthermore, recall that $p - 1 \mid (p - 1)! - (p - 1)$ and $p \mid (p - 1)! - (p - 1)$. Hence, Theorem 2.10.7 (applied to $a = p - 1$, $b = p$ and

\textbf{Proof.} Proposition 2.9.7 (e) (applied to $a = p - 1$, $b = 1$ and $u = 1$) yields $\gcd(p - 1, 1(p - 1) + 1) = \gcd(p - 1, 1)$. But Proposition 2.9.7 (f) (applied to $a = p - 1$ and $b = 1$) yields $\gcd(p - 1, 1) \mid p - 1$ and $\gcd(p - 1, 1) \mid 1$. Since $\gcd(p - 1, 1)$ is a nonnegative integer satisfying $\gcd(p - 1, 1) \mid 1$, we obtain $\gcd(p - 1, 1) = 1$ (by Exercise 2.25 applied to $g = \gcd(p - 1, 1)$). Hence,

$$\gcd\left(p - 1, \frac{p}{p - 1 - 1}ight) = \gcd(p - 1, 1(p - 1) + 1) = \gcd(p - 1, 1) = 1.$$ 

In other words, $p - 1 \perp p$. 

\textit{Math 4281 notes page 428}
c = (p - 1)! - (p - 1)) yields

\[ (p - 1) \divides (p - 1)! - (p - 1). \]  \hspace{1cm} (244)

But Proposition \textbf{2.10.12} (applied to \( n = p - 1 \)) yields

\[ 1 + 2 + \cdots + (p - 1) = \frac{(p - 1)((p - 1) + 1)}{2} = \frac{(p - 1)p}{2}. \]

Hence, \((p - 1)p = (1 + 2 + \cdots + (p - 1)) \cdot 2\), so that \((p - 1)! \equiv (p - 1)p \mod 1 + 2 + \cdots + (p - 1)\) (by (244)). In other words, \((p - 1)! \equiv p - 1 \mod 1 + 2 + \cdots + (p - 1)\). This solves Exercise \textbf{2.15.4}.

\section*{6.67. Solution to Exercise \textbf{2.15.5}}

\textit{Solution to Exercise \textbf{2.15.5}} From \( p = 2k + 1 \), we obtain \( p - 1 = 2k \). But Theorem \textbf{2.15.7} yields \((p - 1)! \equiv -1 \mod p\). In view of \( p - 1 = 2k \), this rewrites as \((2k)! \equiv -1 \mod p\).

But the definition of \((2k)!\) yields

\[ (2k)! = 1 \cdot 2 \cdot \cdots \cdot (2k) = \prod_{j=1}^{2k} j = \left( \prod_{j=1}^{k} j \right) \cdot \left( \prod_{j=k+1}^{2k} j \right) \] \hspace{1cm} (245)

(here, we have split the product at \( j = k \), because \( 0 \leq k \leq 2k \)). But \( \prod_{j=1}^{k} j = 1 \cdot 2 \cdot \cdots \cdot k = k! \) (by the definition of \( k! \)). Furthermore

\[ \prod_{j=k+1}^{2k} j = \prod_{j=p-2k}^{p-(k+1)} (p-j) \left( \text{ here, we have substituted } p-j \right) \]

\[ = \prod_{j=1}^{k} (p-j) \left( \text{ since } p-2k = (2k+1) - 2k = 1 \right) \]

\[ \text{ and } p - (k+1) = (2k+1) - (k+1) = k \]

\[ \equiv \prod_{j=1}^{k} (-j) \mod p. \]
(Here, the last congruence sign is a consequence of (9)\(^{170}\) Thus,
\[
\prod_{j=k+1}^{2k} j \equiv \prod_{j=1}^{k} (-j) = (-1)^k \prod_{j=1}^{k} j = (-1)^k \cdot k! \pmod{p}.
\]
Hence, (245) becomes
\[
(2k)! = \left( \prod_{j=1}^{k} j \right) \cdot \left( \prod_{j=k+1}^{2k} j \right) \equiv k! \cdot (-1)^k \cdot k! = k!^2 \cdot (-1)^k \pmod{p}.
\]
Therefore,
\[
k!^2 \cdot (-1)^k \equiv (2k)! \equiv -1 \pmod{p}.
\]
Multiplying this congruence by the obvious congruence \((-1)^k \equiv (-1)^k \pmod{p}\), we find
\[
k!^2 \cdot (-1)^k \cdot (-1)^k \equiv (-1) \cdot (-1)^k = - (-1)^k \pmod{p}.
\]
In view of
\[
k!^2 \cdot (-1)^k \cdot (-1)^k = k!^2,
\]
this rewrites as \(k!^2 \equiv - (-1)^k \pmod{p}\). This solves Exercise 2.15.5 \(\square\)

6.68. Solution to Exercise 2.16.1

Solution to Exercise 2.16.1. We assumed that the integers \(n_1, n_2, \ldots, n_k\) are mutually coprime. In other words, we have
\[
n_i \perp n_j \text{ for all } i, j \in \{1, 2, \ldots, k\} \text{ satisfying } i \neq j.
\]
(246)
We claim that
\[
\phi (n_1 n_2 \cdots n_i) = \phi (n_1) \cdot \phi (n_2) \cdots \phi (n_i)
\]
(247)
for each \(i \in \{0, 1, \ldots, k\}\).

[Proof of (247): We shall prove (247) by induction on \(i\):]

\(^{170}\) In more detail: We have \(p - s \equiv -s \pmod{p}\) for each \(s \in \{1, 2, \ldots, k\}\). Hence, (9) (applied to \(n = p, S = \{1, 2, \ldots, k\}, a_s = p - s\) and \(b_s = -s\)) yields \(\prod_{s \in \{1, 2, \ldots, k\}} (p - s) \equiv \prod_{s \in \{1, 2, \ldots, k\}} (-s) \pmod{p}\). If we rename the index \(s\) (of the product sign “\(\prod_{s \in \{1, 2, \ldots, k\}} \)” as \(j\) on both sides of this congruence, then we obtain \(\prod_{j \in \{1, 2, \ldots, k\}} (p - j) \equiv \prod_{j \in \{1, 2, \ldots, k\}} (-j) \pmod{p}\). This rewrites as \(\prod_{j=1}^{k} (p - j) \equiv \prod_{j=1}^{k} (-j) \pmod{p}\) (since the product sign \(\prod_{j \in \{1, 2, \ldots, k\}} \) is equivalent to the product sign \(\prod_{j=1}^{k}\)).
Induction base: It is easy to see that \( \phi(1) = 1 \) \(^{[71]}\) But applying the map \( \phi \) to the equality
\[ n_1 n_2 \cdots n_0 = (\text{empty product}) = 1, \]
we obtain
\[ \phi(n_1 n_2 \cdots n_0) = \phi(1) = 1. \]
Comparing this with \( \phi(n_1) \cdot \phi(n_2) \cdots \cdot \phi(n_0) = (\text{empty product}) = 0, \)
we obtain \( \phi(n_1 n_2 \cdots n_0) = \phi(n_1) \cdot \phi(n_2) \cdots \cdot \phi(n_0). \) In other words, \( \text{(247)} \) holds for \( i = 0. \) This completes the induction base.

Induction step: Let \( i \in \{0, 1, \ldots, k\} \) be positive. Assume that \( \text{(247)} \) holds for \( i = j - 1. \) We must prove that \( \text{(247)} \) holds for \( i = j. \)

For each \( i \in \{1, 2, \ldots, j - 1\}, \) we have \( i \leq j - 1 < j \) and thus \( i \neq j \) and therefore \( n_i \perp n_j \) (by \( \text{(246)} \)). Hence, Exercise 2.10.3 (applied to \( j - 1, n_i \) and \( n_1 n_2 \cdots n_{j-1} \) instead of \( k, c \) and \( a_1, a_2, \ldots, a_k \)) yields \( n_1 n_2 \cdots n_{j-1} \perp n_j. \) In other words, the two positive integers \( n_1 n_2 \cdots n_{j-1} \) and \( n_j \) are coprime. Therefore, Theorem 2.14.4 yields
\[ \phi(\n_1 n_2 \cdots n_{j-1} \cdot n_j) = \phi(n_1 n_2 \cdots n_{j-1}) \cdot \phi(n_j). \]

But we have assumed that \( \text{(247)} \) holds for \( i = j - 1. \) In other words, we have
\[ \phi(n_1 n_2 \cdots n_{j-1}) = \phi(n_1) \cdot \phi(n_2) \cdots \phi(n_{j-1}). \]

Now,
\[
\phi \left( \frac{n_1 n_2 \cdots n_j}{(n_1 n_2 \cdots n_{j-1}) n_j} \right) = \phi(\n_1 n_2 \cdots n_{j-1} n_j) = \phi(n_1 n_2 \cdots n_{j-1}) \cdot \phi(n_j) = \phi(n_1) \cdot \phi(n_2) \cdots \phi(n_{j-1}) \cdot \phi(n_j) = \phi(n_1) \cdot \phi(n_2) \cdots \phi(n_{j-1}) \cdot \phi(n_j).
\]
In other words, \( \text{(247)} \) holds for \( i = j. \) This completes the induction step. Hence, the induction proof of \( \text{(247)} \) is complete. \(^{[7]}\)

Now, \( \text{(247)} \) (applied to \( i = k \)) yields \( \phi(n_1 n_2 \cdots n_k) = \phi(n_1) \cdot \phi(n_2) \cdots \cdot \phi(n_k). \) This solves Exercise 2.16.1 \( \square \)

\(^{[71]}\) Proof. We have \( 1 \mid 1. \) Hence, Proposition 2.9.7 (ii) (applied to 1 and 1 instead of \( a \) and \( b \)) yields \( \gcd(1, 1) = |1| = 1. \) Hence, \( 1 \perp 1. \) Thus, \( 1 \) is an \( i \in \{1, 2, \ldots, 1\} \) satisfying \( i \perp 1. \) In other words, \( 1 \) is an element of the set \( \{i \in \{1, 2, \ldots, 1\} \mid i \perp 1\}. \) Since \( 1 \) is the only element of this set (because every element of \( \{i \in \{1, 2, \ldots, 1\} \mid i \perp 1\} \) must belong to the set \( \{1, 2, \ldots, 1\} = \{1\} \) and thus must equal to 1), we can thus conclude that \( \{i \in \{1, 2, \ldots, 1\} \mid i \perp 1\} = \{1\}. \)

But the equality \( (54) \) (applied to \( n = 1 \)) yields
\[ \phi(1) = \left| \left\{ i \in \{1, 2, \ldots, 1\} \mid i \perp 1 \right\} \right| = |\{1\}| = 1. \]
6.69. Solution to Exercise 2.16.2

Solution to Exercise 2.16.2 Exercise 2.16.1 and Exercise 2.16.2 say the same thing: They say that applying the function \( \phi \) to a product of finitely many mutually coprime positive integers yields the same result as applying \( \phi \) to each of these integers separately and then taking the product. The difference between these two exercises is merely how the product is indexed. Thus, deriving Exercise 2.16.2 from Exercise 2.16.1 is merely a matter of bookkeeping (and this is pretty much the same sort of bookkeeping that we used to derive Exercise 2.10.5 from Exercise 2.10.2 above). Let us do this bookkeeping:

The set \( I \) is finite; thus, we can define some \( k \in \mathbb{N} \) by \( k = |I| \). Consider this \( k \). There exists an \( f : \{1, 2, \ldots, k\} \to I \) (since \( k = |I| \)). Pick such an \( f \). Thus, \( f(1), f(2), \ldots, f(k) \) are the \( k \) elements of \( I \); hence, \( n_{f(1)}, n_{f(2)}, \ldots, n_{f(k)} \) are \( k \) positive integers. Moreover, these \( k \) integers \( n_{f(1)}, n_{f(2)}, \ldots, n_{f(k)} \) are mutually coprime. Hence, Exercise 2.16.1 (applied to \( n_{f(i)} \) instead of \( n_i \) yields \( \phi \left( n_{f(1)} n_{f(2)} \cdots n_{f(k)} \right) = \phi \left( n_{f(1)} \right) \cdot \phi \left( n_{f(2)} \right) \cdots \phi \left( n_{f(k)} \right) \).

The map \( f : \{1, 2, \ldots, k\} \to I \) is a bijection. Hence, we can substitute \( f(j) \) for \( i \) in the product \( \prod_{i \in I} n_i \). We thus find

\[
\prod_{i \in I} n_i = \prod_{j \in \{1, 2, \ldots, k\}} n_{f(j)} = \prod_{j=1}^{k} n_{f(j)} = n_{f(1)} n_{f(2)} \cdots n_{f(k)}.
\]

Applying the map \( \phi \) to both sides of this equality, we obtain

\[
\phi \left( \prod_{i \in I} n_i \right) = \phi \left( n_{f(1)} n_{f(2)} \cdots n_{f(k)} \right) = \phi \left( n_{f(1)} \right) \cdot \phi \left( n_{f(2)} \right) \cdots \phi \left( n_{f(k)} \right)
\]

\[
= \prod_{j=1}^{k} \phi \left( n_{f(j)} \right) = \prod_{j \in \{1, 2, \ldots, k\}} \phi \left( n_{f(j)} \right) = \prod_{i \in I} \phi \left( n_i \right)
\]

(here, we have substituted \( i \) for \( f(j) \) in the product, since the map \( f : \{1, 2, \ldots, k\} \to I \) is a bijection). This solves Exercise 2.16.2. \qed

6.70. Solution to Exercise 2.16.3

Solution to Exercise 2.16.3 The integer \( n \) is positive and thus nonzero. Hence, \( v_p(n) \in \mathbb{N} \) for each prime \( p \).

We shall next prove that

\[
a^n \equiv a^{n-\phi(n)} \mod p^{v_p(n)} \quad \text{for every prime } p.
\]  \hspace{1cm} (249)

\text{Proof.} Let \( x \) and \( y \) be two distinct elements of \( \{1, 2, \ldots, k\} \). We claim that \( n_{f(x)} \perp n_{f(y)} \).

Indeed, the map \( f \) is a bijection, and thus is injective. But \( x \) and \( y \) are distinct; thus, \( x \neq y \). Therefore, \( f(x) \neq f(y) \) (since \( f \) is injective). Hence, \( f(x) \) and \( f(y) \) are two distinct elements of \( I \). Thus, \( (75) \) (applied to \( i = f(x) \) and \( j = f(y) \)) yields \( n_{f(x)} \perp n_{f(y)} \).

Now, forget that we fixed \( x \) and \( y \). We thus have proven that every two distinct elements \( x \) and \( y \) of \( \{1, 2, \ldots, k\} \) satisfy \( n_{f(x)} \perp n_{f(y)} \). In other words, the \( k \) integers \( n_{f(1)}, n_{f(2)}, \ldots, n_{f(k)} \) are mutually coprime.
Once this congruence is proven, the claim of Exercise 2.16.3 will easily follow using Exercise 2.13.9.

[Proof of (249): Let \( p \) be a prime. Then, \( p > 1 \). Hence, \( p \neq 1 \), and the integer \( p \) is positive.

Also, Lemma 2.13.27(a) shows that there exists a nonzero integer \( u \) such that \( u \perp p \) and \( n = up^v \). Consider this \( u \).

We have \( v_p (n) \in \mathbb{N} \) (since \( n \) is nonzero). Thus, we can define \( r \in \mathbb{N} \) by \( r = v_p (n) \). Consider this \( r \). Note that \( p^r \) is nonzero (since \( p \) is positive and thus nonzero).

We have \( n = up^v = up^r \) (since \( v_p (n) = r \)). Solving this equation for \( u \), we obtain \( u = n/p^r \) (since \( p^r \) is nonzero). Thus, \( u \) is positive (since both \( n \) and \( p \) are positive). Now, we claim that

\[
up^i - \phi (up^i) \geq i \quad \text{for each} \quad i \in \mathbb{N}.
\]

Let us give three proofs of this inequality:

- [First proof of (250): We shall prove (250) by induction on \( i \):

  Induction base: Exercise 2.14.2(b) (applied to \( u \) instead of \( n \)) yields \( u - \phi (u) \geq 0 \).

  Now, \( u p^0 - \phi \left( u p^0 \right) = u - \phi (u) \geq 0 \). In other words, (250) holds for \( i = 0 \).

  This completes the induction base.

  Induction step: Let \( j \) be a positive integer. Assume that (250) holds for \( i = j - 1 \). We must prove that (250) holds for \( i = j \).

  We have assumed that (250) holds for \( i = j - 1 \). In other words, we have \( up^{j-1} - \phi (up^{j-1}) \geq j - 1 \).

  We have \( j - 1 \in \mathbb{N} \) (since \( j \) is a positive integer); hence, \( p^{j-1} \) is an integer. Thus, \( up^{j-1} \) is an integer. Clearly, \( up^{j-1} \mid up^j \) (since \( up^j = up^{j-1}p \)); thus, \( up^{j-1} \) is a divisor of \( up^j \). Also, \( up^{j-1} \) is positive \(^{173} \) and satisfies \( up^{j-1} \neq up^j \) (since \( \frac{up^j}{up^{j-1}} = p \neq 1 \)). Hence, Exercise 2.14.2(d) (applied to \( up^{j-1} \) and \( up^j \) instead of \( d \) and \( n \)) yields

  \[
  up^{j-1} - \phi (up^{j-1}) < up^j - \phi (up^j).
  \]

  Hence,

  \[
  up^j - \phi (up^j) > up^{j-1} - \phi (up^{j-1}) \geq j - 1.
  \]

  Since \( up^j - \phi (up^j) \) and \( j - 1 \) are integers, this leads to

  \[
  up^j - \phi (up^j) \geq (j - 1) + 1 = j.
  \]

  In other words, (250) holds for \( i = j \). This completes the induction step. Hence, (250) is proven by induction.]

- [Second proof of (250) (sketched): Let \( j \in \mathbb{N} \). We shall show that \( up^j - \phi (up^j) \geq j \).

  Recall that \( u \) is a positive integer. Hence, \( u \geq 1 \) and thus \( u \geq 1 \), \( p^j \geq p^i \) (since \( p^j \) is a positive integer); in other words, \( p^i \leq up^i \). Also, from \( p > 1 \), we obtain \( p^1 < p^2 < \cdots < p^i \). Hence, the \( j \) integers \( p^1, p^2, \ldots, p^j \) are distinct.]

\(^{173}\)since \( u \) and \( p \) are positive.
On the other hand, it is easy to see that

\[ \{p^1, p^2, \ldots, p^j\} \subseteq \{i \in \{1, 2, \ldots, up^j\} \mid \text{we don't have } i \perp up^j\} \]

[174] Hence,

\[ \left| \{p^1, p^2, \ldots, p^j\} \right| \leq \left| \{i \in \{1, 2, \ldots, up^j\} \mid \text{we don't have } i \perp up^j\} \right|. \] (251)

But Exercise 2.14.2 (applied to \( up^j \) instead of \( n \)) yields

\[ up^j - \phi \left( up^j \right) = \left| \{i \in \{1, 2, \ldots, up^j\} \mid \text{we don't have } i \perp up^j\} \right| \]
\[ \geq \left| \{p^1, p^2, \ldots, p^j\} \right| \quad \text{(by (251))} \]
\[ = j \quad \text{(since the } j \text{ integers } p^1, p^2, \ldots, p^j \text{ are distinct)}. \]

Now, forget that we fixed \( j \). We thus have proven that \( up^j - \phi \left( up^j \right) \geq j \) for each \( j \in \mathbb{N} \). Renaming \( j \) as \( i \) in this statement, we conclude that \( up^i - \phi \left( up^i \right) \geq i \) for each \( i \in \mathbb{N} \). Thus, (250) is proven.]

- [Third proof of (250):] The following proof is a typical estimation argument (the kind you see in analysis).

Fix \( i \in \mathbb{N} \). We must prove that \( up^i - \phi \left( up^i \right) \geq i \). This is obvious in the case when \( i = 0 \) (because Exercise 2.14.2 (b) (applied to \( up^0 \) instead of \( n \)) yields \( up^0 - \phi \left( up^0 \right) \geq 0 \)). Hence, for the rest of this proof, we WLOG assume that \( i \neq 0 \). Hence, \( i \) is a

174Proof. Let \( s \in \{1, 2, \ldots, j\} \). Thus, \( s \geq 1 \) and \( s \leq j \). From \( s \geq 1 \), we conclude that \( p^s \) is a positive integer. Moreover, from \( s \geq 1 \), we obtain \( p^s \geq p^1 \) (since \( p > 1 \)), thus \( p^s \geq p^1 = p > 1 \). From \( s \leq j \), we obtain \( p^s \leq p^j \) (since \( p > 1 \)) and thus \( p^s \leq p^j \leq up^j \). Thus, \( p^j \in \{1, 2, \ldots, up^j\} \) (since \( p^s \) is a positive integer). Furthermore, from \( s \leq j \), we obtain \( p^s \mid p^j \) (by Exercise 2.2.4 applied to \( p, s \) and \( j \) instead of \( n, a \) and \( b \) ). Thus, \( p^s \mid p^j \mid up^j \) (since \( u \) is an integer). Hence, Proposition 2.9.7 (i) (applied to \( p^j \) and \( up^j \) instead of \( a \) and \( b \)) yields \( \gcd \left( p^s, up^j \right) = \left| p^s \right| = p^s \) (since \( p^s \) is positive). But \( p > 1 \), so that \( p^s > 1 \) and thus \( p^s \neq 1 \). Hence, \( \gcd \left( p^s, up^j \right) = p^s \neq 1 \). In other words, we don't have \( p^s \perp up^j \).

So we have shown that \( p^s \) is an element of \( \{1, 2, \ldots, up^j\} \) (since \( p^s \in \{1, 2, \ldots, up^j\} \)) with the property that we don't have \( p^s \perp up^j \). In other words, \( p^s \) is an \( i \in \{1, 2, \ldots, up^j\} \) such that we don't have \( i \perp up^j \). In other words,

\[ p^s \in \{i \in \{1, 2, \ldots, up^j\} \mid \text{we don't have } i \perp up^j\}. \]

Now, forget that we fixed \( s \). We thus have proven that

\[ p^s \in \{i \in \{1, 2, \ldots, up^j\} \mid \text{we don't have } i \perp up^j\} \]

for each \( s \in \{1, 2, \ldots, j\} \). In other words,

\[ \{p^1, p^2, \ldots, p^j\} \subseteq \{i \in \{1, 2, \ldots, up^j\} \mid \text{we don't have } i \perp up^j\}\],

qed.
positive integer (since \( i \in \mathbb{N} \)). Therefore, Exercise 2.14.1 (applied to \( k = i \)) yields \( \phi (p^i) = (p - 1) p^{i-1} \). Hence,

\[
\frac{p^i - \phi (p^i)}{p^{i-1}} = \frac{pp^{i-1} - (p - 1) p^{i-1}}{(p-1)p^{i-1}} = p^{i-1}. \tag{252}
\]

Also, \( i - 1 \in \mathbb{N} \) (since \( i \) is a positive integer). Hence, Exercise 2.13.4 (applied to \( k = i - 1 \)) yields \( p^{i-1} > i - 1 \). Since \( p^{i-1} \) and \( i - 1 \) are integers (because \( i - 1 \in \mathbb{N} \)), this leads to \( p^{i-1} \geq (i - 1) + 1 = i \).

Note that \( u \) is a positive integer. Thus, \( u \geq 1 \). Exercise 2.14.2 (b) (applied to \( u \) instead of \( n \)) yields \( u - \phi (u) \geq 0 \). In other words, \( \phi (u) \leq u \). Furthermore, \( p^i \) is a positive integer; thus, \( \phi (p^i) \geq 0 \) (because clearly, \( \phi (m) = 0 \) for every positive integer \( m \)).

But \( u \perp p^i \). Hence, Exercise 2.10.4 (applied to \( u, p, 1 \) and \( i \) instead of \( a, b, n \) and \( m \)) yields \( u^i \perp p^i \) (since \( i \in \mathbb{N} \)). In other words, \( u \perp p^i \) (since \( u^i = u \)). In other words, the integers \( u \) and \( p^i \) are coprime. Thus, Theorem 2.14.4 (applied to \( u \) and \( p^i \) instead of \( m \) and \( n \)) yields \( \phi (up^i) = \phi (u) \cdot \phi (p^i) \) (since \( u \) and \( p^i \) are positive integers). Thus,

\[
\phi (up^i) = \phi (u) \cdot \phi (p^i) \leq u \cdot \phi (p^i) \tag{since \( \phi (p^i) \geq 0 \)}.
\]

Hence,

\[
up^i - \phi (up^i) \geq up^i - u \cdot \phi (p^i) = u \cdot \left( p^i - \phi (p^i) \right) = u \cdot \frac{p^{i-1}}{p^{i-1}} = \frac{up^{i-1}}{p^{i-1}} \tag{by 252}
\]

\[
\geq 1p^{i-1} \tag{since \( p^{i-1} \geq 0 \)}
\]

\[
= p^{i-1} \geq i.
\]

Thus, 250 is proven again.]

Now, we can apply 250 to \( i = r \) (since \( r \in \mathbb{N} \)). We thus obtain \( up^r - \phi (up^r) \geq r \). Hence,

\[
\frac{n}{up^r} - \phi \left( \frac{n}{up^r} \right) = up^r - \phi (up^r) \geq r. \tag{253}
\]

Now, we are in one of the following two cases:

**Case 1**: We have \( p \mid a \).

**Case 2**: We don’t have \( p \mid a \).

Let us first consider Case 1. In this case, \( p \mid a \). Now, 253 yields \( n - \phi (n) \geq r \geq 0 \) (since \( r \in \mathbb{N} \)). Hence, \( n - \phi (n) \in \mathbb{N} \) and \( r \in \mathbb{N} \). From \( n - \phi (n) \geq r \), we also obtain \( r \leq n - \phi (n) \). Hence, Exercise 2.2.4 (applied to \( p, r \) and \( n - \phi (n) \) instead of \( n, a \) and \( b \)) yields \( p^r \mid p^{n-\phi(n)} \). But \( p \mid a \). Hence, Exercise 2.2.6 (applied to \( p, a \) and \( n - \phi (n) \) instead of
This rewrites as \( a^n \equiv a^{n-\phi(n)} \mod p^{\nu_p(n)} \) (since \( r = \nu_p(n) \)). Hence, (249) is proven in Case 1.

Let us now consider Case 2. In this case, we don’t have \( p \mid a \). But Proposition 2.13.5 yields that either \( p \mid a \) or \( p \perp a \). Hence, \( p \perp a \) (since we don’t have \( p \mid a \)). Due to Proposition 2.10.4, this leads to \( a \perp p \). Thus, Exercise 2.10.4 (applied to \( p, 1 \) and \( r \) instead of \( b, n \) and \( m \)) yields \( a^1 \perp p^r \) (since \( r \in \mathbb{N} \)). In other words, \( a \perp p^r \) (since \( a^1 = a \)). In other words, \( a \) is coprime to \( p^r \). Hence, Theorem 2.13.3 (applied to \( p^r \) instead of \( n \)) yields \( a^{\phi(p^r)} \equiv 1 \mod p^r \).

But \( u \perp p \). Hence, Exercise 2.10.4 (applied to \( u, p, 1 \) and \( r \) instead of \( a, b, n \) and \( m \)) yields \( u^1 \perp p^r \) (since \( r \in \mathbb{N} \)). In other words, \( u \perp p^r \) (since \( u^1 = u \)). In other words, the integers \( u \) and \( p^r \) are coprime. Thus, Theorem 2.14.4 (applied to \( u \) and \( p^r \) instead of \( m \) and \( n \)) yield \( \phi(u p^r) = \phi(u) \cdot \phi(p^r) \) (since \( u \) and \( p^r \) are positive integers). Now, applying the map \( \phi \) to both sides of the equality \( n = u p^r \), we obtain \( \phi(n) = \phi(u p^r) = \phi(u) \cdot \phi(p^r) = \phi(p^r) \cdot \phi(u) \).

Hence,

\[
\begin{align*}
a^{\phi(n)} &= a^{\phi(p^r) \cdot \phi(u)} = \left( a^{\phi(p^r)} \right)^{\phi(u)} = a^{\phi(p^r) \cdot \phi(u) \mod p^r} \equiv 1^{\phi(u)} = 1 \mod p^r. \\
\end{align*}
\]

Hence,

\[
\begin{align*}
a^n &= a^{\phi(n)+(n-\phi(n))} = a^{\phi(n)} a^{n-\phi(n)} = a^{n-\phi(n) \mod p^r} \equiv a^{n-\phi(n)} \mod p^r. \\
\end{align*}
\]

This rewrites as \( a^n \equiv a^{n-\phi(n)} \mod p^{\nu_p(n)} \) (since \( r = \nu_p(n) \)). Therefore, (249) is proven in Case 2.

We have thus shown (249) in both Cases 1 and 2. These cases cover all possibilities, and so (249) is always proven.]\]

Hence, Exercise 2.13.9 (applied to \( a^n \) and \( a^{n-\phi(n)} \) instead of \( a \) and \( b \)) yields \( a^n \equiv a^{n-\phi(n)} \mod n \). This solves Exercise 2.16.3 \( \square \)

### 6.71. Solution to Exercise 2.17.1

**Solution to Exercise 2.17.1** A product of \( k \) consecutive integers always has the form \((a+1) \cdot (a+2) \cdot \cdots (a+k)\) for some \( a \in \mathbb{Z} \). Thus, we must prove that \((a+1) \cdot (a+2) \cdot \cdots (a+k)\) is divisible by \( k! \) for each \( a \in \mathbb{Z} \).

Let \( a \in \mathbb{Z} \). We must prove that \((a+1) \cdot (a+2) \cdot \cdots (a+k)\) is divisible by \( k! \).

Proposition 2.17.12 (applied to \( n = a+k \)) yields that \( \binom{a+k}{k} \) is an integer. Now, the
definition of \( \binom{a+k}{k} \) yields
\[
\binom{a+k}{k} = \frac{(a+k) (a+k-1) (a+k-2) \cdots (a+k-k+1)}{k!}.
\]

Multiplying both sides of this equality by \( k! \), we find
\[
k! \binom{a+k}{k} = (a+k) (a+k-1) (a+k-2) \cdots (a+k-k+1)
= (a+k) (a+k-1) (a+k-2) \cdots (a+1)
= (a+1) (a+2) \cdots (a+k)
\]
(here, we have reversed the order of the factors in the product). Thus, \((a+1) (a+2) \cdots (a+k) = k! \binom{a+k}{k}\). Since \(\binom{a+k}{k}\) is an integer, this equality yields that \((a+1) (a+2) \cdots (a+k)\) is divisible by \(k!\). This solves Exercise 2.17.1.

\[\square\]

### 6.72. Solution to Exercise 2.17.2

**Solution to Exercise 2.17.2** (a) Let \( n \in \mathbb{N} \). Let \( k \) be a positive integer. Theorem 2.6.1 (applied to \( n \) and \( k \) instead of \( u \) and \( n \)) shows that there exists a unique pair \((q,r) \in \mathbb{Z} \times \{0,1,\ldots,k-1\}\) such that \( n = qk + r \). Consider this pair. Then, \( n/k = q \) (by the definition of \( n/k \)). From \((q,r) \in \mathbb{Z} \times \{0,1,\ldots,k-1\}\), we obtain \( q \in \mathbb{Z} \) and \( r \in \{0,1,\ldots,k-1\}\). From \( r \in \{0,1,\ldots,k-1\}\), we obtain \( r \geq 0 \) and \( r \leq k-1 \). From \( n = qk + r \), we obtain \( qk = n - r \geq -r \geq -k = (-1)k \). If we had \( q \leq -1 \), then we would have \( qk \leq (-1)k \) (since \( k \) is positive), which would contradict \( qk > (-1)k \). Thus, we cannot have \( q \leq -1 \). Hence, \( q > -1 \). Therefore, \( q \geq 0 \) (since \( q \in \mathbb{Z} \)). Also, \( n = qk + r \geq 0 \) and \( n = qk + r \leq k-1 < k \).

We have \( 1 < 2 < \cdots < q \). Since \( k \) is positive, we can multiply this chain of inequalities by \( k \), and obtain \( 1k < 2k < \cdots < qk \). Thus, the \( q \) numbers \( 1k, 2k, \ldots, qk \) are distinct; therefore,
\[
|\{1k, 2k, \ldots, qk\}| = q.
\]

We now shall show the following:

**Claim 1:** We have
\[
\{ i \in \{1,2,\ldots,n\} \text{ satisfying } k \mid i \} = \{1k, 2k, \ldots, qk\}.
\]

(We are writing the word “satisfying” out, since the “|” symbol is already being used for divisibility here.)

**Proof of Claim 1:** Let \( j \in \{ i \in \{1,2,\ldots,n\} \text{ satisfying } k \mid i \} \). We shall show that \( j \in \{1k, 2k, \ldots, qk\} \).
Indeed, we have \( j \in \{ i \in \{1, 2, \ldots, n \} \text{ satisfying } k \mid i \} \). In other words, \( j \) is an element of \( \{1, 2, \ldots, n\} \) satisfying \( k \mid j \). From \( k \mid j \), we conclude that there exists an integer \( c \) such that \( j = kc \). Consider this \( c \). From \( j \in \{1, 2, \ldots, n\} \), we obtain \( j \geq 1 \) and \( j \leq n \). Now, \( ck = kc = j \leq n < (q + 1)k \). We can divide this inequality by \( k \) (since \( k \) is positive) and thus obtain \( c < q + 1 \). Hence, \( c \leq q \) (since \( c \) and \( q \) are integers).

Also, \( ck = kc = j \geq 1 > 0 \). We can divide this inequality by \( k \) (since \( k \) is positive) and thus obtain \( c > 0 \). Hence, \( c \geq 1 \) (since \( c \) is an integer). Combining this with \( c \leq q \), we obtain \( c \in \{1, 2, \ldots, q\} \) (since \( c \) is an integer) and thus \( ck \in \{1, 2k, \ldots, qk\} \). Hence, \( j = kc = ck \in \{1k, 2k, \ldots, qk\} \).

Now, forget that we fixed \( j \). We thus have proven that \( j \in \{1k, 2k, \ldots, qk\} \) for each \( j \in \{i \in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i \} \). In other words,

\[
\{ i \in \{1, 2, \ldots, n \} \text{ satisfying } k \mid i \} \subseteq \{1k, 2k, \ldots, qk\}. \tag{254}
\]

On the other hand, let \( h \in \{1k, 2k, \ldots, qk\} \). We shall show that \( h \in \{i \in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i \} \).

Indeed, \( h \in \{1k, 2k, \ldots, qk\} \). In other words, \( h = dk \) for some \( d \in \{1, 2, \ldots, q\} \). Consider this \( d \). From \( d \in \{1, 2, \ldots, q\} \), we obtain \( d \geq 1 \) and \( d \leq q \). We can multiply the inequality \( d \geq 1 \) by \( k \) (since \( k \) is positive) and thus obtain \( dk \geq 1k = k \geq 1 \) (since \( k \) is a positive integer). Also, we can multiply the inequality \( d \leq q \) by \( k \) (since \( k \) is positive) and thus obtain \( dk \leq qk \leq n \) (since \( n \geq qk \)). Combining this with \( dk \geq 1 \), we obtain \( dk \in \{1, 2, \ldots, n\} \) (since \( dk \) is an integer). In other words, \( h \in \{1, 2, \ldots, n\} \) (since \( h = dk \)). Moreover, \( k \mid h \) (since \( h = dk = kd \)). Hence, \( h \) is an \( i \in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i \) (since \( h \in \{1, 2, \ldots, n\} \) and \( k \mid h \)). In other words, \( h \in \{i \in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i \} \).

Now, forget that we fixed \( h \). We thus have proven that \( h \in \{i \in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i \} \) for each \( h \in \{1k, 2k, \ldots, qk\} \). In other words,

\[
\{1k, 2k, \ldots, qk\} \subseteq \{ i \in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i \}.
\]

Combining this with (254), we obtain \( \{ i \in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i \} = \{1k, 2k, \ldots, qk\} \).

This proves Claim 1.]

Now,

\[
\sum_{i=1}^{n} \left[ k \mid i \right] = \sum_{i \in \{1, 2, \ldots, n\}} \left[ k \mid i \right] = \sum_{i \in \{1, 2, \ldots, n\}} \sum_{k|i} \left[ k \mid i \right] = \sum_{i \in \{1, 2, \ldots, n\}} \sum_{k|i \text{; \(k \mid i\)}} \left[ k \mid i \right] = 1 + \sum_{i \in \{1, 2, \ldots, n\}} \sum_{k|i \text{; \(k \not\mid i\)}} \left[ k \mid i \right] = 0 = 1
\]

(since each \( i \in \{1, 2, \ldots, n\} \) either satisfies \( k \mid i \) or does not)

\[
= \sum_{i \not\in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i} 0 = \sum_{i \in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i} 1
\]

= (the number of all \( i \in \{1, 2, \ldots, n\} \) satisfying \( k \mid i \)) \cdot 1

= (the number of all \( i \in \{1, 2, \ldots, n\} \) satisfying \( k \mid i \))

= \left| \{i \in \{1, 2, \ldots, n\} \text{ satisfying } k \mid i\} \right|

= \left| \{1k, 2k, \ldots, qk\} \right| \quad \text{(by Claim 1)}

= q = n / k.
In other words, \( n / / k = \sum_{i=1}^{\infty} [k \mid i] \). This solves Exercise 2.17.2 (a).

(b) Let \( p \) be a prime. Let \( n \) be a nonzero integer.

Then, \( v_p (n) \) is the largest \( m \in \mathbb{N} \) such that \( p^m \mid n \) (by Definition 2.13.23 (a)). Hence, \( v_p (n) \in \mathbb{N} \). Thus, \( \{1, 2, \ldots, v_p (n)\} \) is a finite set and has size \( \mid \{1, 2, \ldots, v_p (n)\} \mid = v_p (n) \). Every \( i \in \{1, 2, \ldots, v_p (n)\} \) satisfies

\[
[p^i \mid n] = 1
\]

(255)

Moreover, every positive integer \( i \) satisfying \( i > v_p (n) \) satisfies

\[
[p^i \mid n] = 0
\]

(256)

Hence, all but finitely many positive integers \( i \) satisfy \( [p^i \mid n] = 0 \) (since all but finitely many positive integers \( i \) satisfy \( i > v_p (n) \)). In other words, all but finitely many addends of the sum \( \sum_{i=1}^{\infty} [p^i \mid n] \) are zero. In other words, this sum has only finitely many nonzero addends. Hence, this sum is well-defined.

Now, we can split this sum at \( i = v_p (n) \) (since \( v_p (n) \geq 0 \)), and thus obtain

\[
\sum_{i=1}^{\infty} [p^i \mid n] = \sum_{i=1}^{v_p (n)} [p^i \mid n] + \sum_{i=v_p (n)+1}^{\infty} [p^i \mid n] = \sum_{i=1}^{v_p (n)} 1 + \sum_{i=v_p (n)+1}^{\infty} 0 = \sum_{i=1}^{v_p (n)} 1 = v_p (n) \cdot 1 = v_p (n) .
\]

(by 255)

(by 256)

In other words, \( v_p (n) = \sum_{i=1}^{\infty} [p^i \mid n] \). This solves Exercise 2.17.2 (b).

(c) Let \( p \) be a prime. Let \( n \in \mathbb{N} \). For every positive integer \( m \), we have

\[
\begin{align*}
v_p (m) &= \sum_{i=1}^{\infty} [p^i \mid m] \\
&= \sum_{i=1}^{\infty} [p^i \mid m] \\
&= \sum_{j=1}^{\infty} [p^j \mid m]
\end{align*}
\]

(by Exercise 2.17.2 (b),

applied to \( m \) instead of \( n \) )

(257)

(here, we have renamed the summation index \( i \) as \( j \)). Note that the sum \( \sum_{j=1}^{\infty} [p^j \mid m] \) on the right hand side of this equality is well-defined, i.e., has only finitely many nonzero addends.

Corollary 2.13.29 (applied to \( k = n \) and \( a_i = i \)) yields

\[
v_p (1 \cdot 2 \cdot \ldots \cdot n) = v_p (1) + v_p (2) + \ldots + v_p (n) = \sum_{m=1}^{n} v_p (m) .
\]

Proof. Let \( i \in \{1, 2, \ldots, v_p (n)\} \). Thus, \( i \) is a positive integer satisfying \( i \leq v_p (n) \). Hence, \( v_p (n) \geq i \). But Lemma 2.13.25 shows that \( p^i \mid n \) if and only if \( v_p (n) \geq i \). Thus, we have \( p^i \mid n \) (since we have \( v_p (n) \geq i \)). Therefore, \( [p^i \mid n] = 1 \), \text{qed.}

Proof. Let \( i \) be a positive integer satisfying \( i > v_p (n) \). We must prove that \( [p^i \mid n] = 0 \).

We have \( i > v_p (n) \). Thus, we don’t have \( v_p (n) \geq i \). But Lemma 2.13.25 shows that \( p^i \mid n \) if and only if \( v_p (n) \geq i \). Hence, we don’t have \( p^i \mid n \) (since we don’t have \( v_p (n) \geq i \)). Thus, \( [p^i \mid n] = 0 \). \text{Qed.}
In view of \(1 \cdot 2 \cdot \cdots \cdot n = n!\), this rewrites as

\[
v_p(n!) = \sum_{m=1}^{n} v_p(m) = \sum_{m=1}^{n} \sum_{j \geq 1} \left\lfloor \frac{p^j}{m} \right\rfloor = \sum_{j \geq 1} \sum_{m=1}^{n} \left\lfloor \frac{p^j}{m} \right\rfloor . \tag{258}
\]

(Here, we have interchanged the two summation signs \(\sum_{m=1}^{n}\) and \(\sum_{j \geq 1}\). This is legitimate, since the first summation is finite whereas the second summation has already been proven to be well-defined.) But each positive integer \(j\) satisfies

\[
n/p^j = \sum_{i=1}^{n} \left\lfloor \frac{p^j}{i} \right\rfloor \quad \text{(by Exercise 2.17.2(a), applied to } k = p^j)\]

\[
= \sum_{m=1}^{n} \left\lfloor \frac{p^j}{m} \right\rfloor \tag{259}
\]

(here, we have renamed the summation index \(i\) as \(m\)). Thus, (258) becomes

\[
v_p(n!) = \sum_{j \geq 1} \sum_{m=1}^{n} \left\lfloor \frac{p^j}{m} \right\rfloor = \sum_{j \geq 1} n/p^j = \sum_{i \geq 1} n/p^i
\]

(here, we have renamed the summation index \(j\) as \(i\)). This solves Exercise 2.17.2(c).

**d) Second proof of Corollary 2.17.11** We must prove that \(\binom{n}{k}\) is a nonnegative integer. If \(k \notin \mathbb{N}\), then this holds (because if \(k \notin \mathbb{N}\), then Definition 2.17.1(b) yields \(\binom{n}{k} = 0\)). Thus, for the rest of this proof, we WLOG assume that \(k \in \mathbb{N}\).

If \(k > n\), then we have \(\binom{n}{k} = 0\) (by Theorem 2.17.4). Hence, if \(k > n\), then \(\binom{n}{k}\) is clearly a nonnegative integer. Thus, for the rest of this proof, we WLOG assume that we don’t have \(k > n\). Hence, \(k \leq n\). Therefore, \(n \geq k\) and thus \(n - k \in \mathbb{N}\). Hence, Theorem 2.17.3 yields

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{260}
\]

This yields that the number \(\binom{n}{k}\) is positive (since the numbers \(n!, k!\) and \(n-k)!\) are all positive) and therefore nonnegative. It remains to prove that \(\binom{n}{k}\) is an integer.

Fix a prime \(p\). We shall show that \(v_p(k! \cdot (n-k)!) \leq v_p(n!)\). (This will then yield \(k! \cdot (n-k)! \ | \ n!,\) and this will in turn yield the integrality of \(\binom{n}{k}\) by way of (260).)

Fix a positive integer \(i\). Exercise 2.6.3(b) (applied to \(k, n-k\) and \(p^i\) instead of \(u, v\) and \(n\)) yields \((k + (n-k)) / / p^i - k / / p^i - (n-k) / / p^i \in \{0,1\} \subseteq \mathbb{N}\). This rewrites as \(n / / p^i - k / / p^i - (n-k) / / p^i \in \mathbb{N}\) (since \(k + (n-k) = n\)). Hence,

\[
n / / p^i - k / / p^i - (n-k) / / p^i \geq 0. \tag{261}
\]
Now, forget that we fixed \( i \). We thus have proven the inequality (261) for each positive integer \( i \).

Theorem 2.13.28 (a) (applied to \( a = k! \) and \( b = (n - k)! \)) yields

\[
v_p \left( k! (n - k)! \right) = v_p \left( k! \right) + v_p \left( (n - k)! \right).
\]

Hence,

\[
v_p \left( n! \right) - v_p \left( k! (n - k)! \right) = v_p \left( n! \right) - (v_p \left( k! \right) + v_p \left( (n - k)! \right))
\]

\[
= \sum_{i \geq 1} \frac{n}{p^i} - \sum_{i \geq 1} \frac{k}{p^i} - \sum_{i \geq 1} \frac{(n - k)}{p^i} = \sum_{i \geq 1} \left( \frac{n}{p^i} - \frac{k}{p^i} - \frac{(n - k)}{p^i} \right)
\]

\[
\geq \sum_{i \geq 1} 0 = 0.
\]

In other words,

\[v_p \left( k! (n - k)! \right) \leq v_p \left( n! \right).\]

Now, forget that we fixed \( p \). We thus have proven that each prime \( p \) satisfies \( v_p \left( k! (n - k)! \right) \leq v_p \left( n! \right) \). But Proposition 2.13.35 (applied to \( k! (n - k)! \) and \( n! \) instead of \( n \) and \( m \)), we have \( k! (n - k)! \mid n! \) if and only if each prime \( p \) satisfies \( v_p \left( k! (n - k)! \right) \leq v_p \left( n! \right) \). Thus, we have \( k! (n - k)! \mid n! \) (since each prime \( p \) satisfies \( v_p \left( k! (n - k)! \right) \leq v_p \left( n! \right) \)). In other words, \( \frac{n!}{k! (n - k)!} \) is an integer (since \( k! (n - k)! \neq 0 \)). In view of (260), this rewrites as follows:

\[
\left( \frac{n}{k} \right) \text{ is an integer. Hence, } \left( \frac{n}{k} \right) \text{ is a nonnegative integer (since we already know that } \left( \frac{n}{k} \right) \text{ is nonnegative). Hence, Corollary 2.17.11 is proven again.}
\]

Thus, Exercise 2.17.2 (d) is solved. \( \square \)

6.73. Solution to Exercise 2.18.1

Solution to Exercise 2.18.1 (a) For every prime \( p > |n| \), we have \( v_p (n) = 0 \) (by Lemma 2.13.32 (a)) and thus

\[
p^0k + p^1k + \cdots + p^{v_p(n) - k} = p^0k + p^1k + \cdots + p^{0k} = p^0k = p^0 = 1.
\]

Thus, all but finitely many primes \( p \) satisfy \( p^0k + p^1k + \cdots + p^{v_p(n) - k} = 1 \) (since all but finitely many primes \( p \) satisfy \( p > |n| \)). Therefore, all but finitely many factors of the product \( \prod_{p \text{ prime}} \left( p^0k + p^1k + \cdots + p^{v_p(n) - k} \right) \) are 1. In other words, the product
\[ \prod_{p \text{ prime}} \left( p^{0k} + p^{1k} + \cdots + p^{v_p(n)k} \right) \] has only finitely many factors different from 1. Hence, this product is well-defined. This solves Exercise 2.18.1(a).

(b) Forget that we fixed \( k \). Instead, fix \( w \in \mathbb{Z} \). (The only reason we are doing this is that we will have to use the letter "\( k \)" for a different purpose.)

For every prime \( p > |n| \), we have \( v_p(n) = 0 \) (by Lemma 2.13.32(a)). Thus, all but finitely many primes \( p \) satisfy \( v_p(n) = 0 \) (since all but finitely many primes \( p \) satisfy \( p > |n| \)). In other words, the set of all primes \( p \) satisfying \( v_p(n) \neq 0 \) is finite. Let \( P \) be this set. Thus, \( P \) is finite.

Let \((p_1, p_2, \ldots, p_u)\) be a list of elements of \( P \), with no repetitions.\(^{177}\) Thus, \( \{p_1, p_2, \ldots, p_u\} = P \). Now, the elements \( p_1, p_2, \ldots, p_u \) belong to \( \{p_1, p_2, \ldots, p_u\} = P \), and thus are primes (since \( P \) is a set of primes). Furthermore, the elements \( p_1, p_2, \ldots, p_u \) are distinct (since \( (p_1, p_2, \ldots, p_u) \) was defined to be a list with no repetitions).

For each \( i \in \{1, 2, \ldots, u\} \), define a nonnegative integer \( a_i \) by
\[
a_i = v_{p_i}(n). \tag{262}
\]
This is well-defined, since \( p_i \) is a prime (because \( p_1, p_2, \ldots, p_u \) are primes) and since \( n \) is nonzero.

The following facts have been proven in the proof of Proposition 2.18.1:

- The map \( \{1, 2, \ldots, u\} \to \{p_1, p_2, \ldots, p_u\}, i \mapsto p_i \) is a bijection.
- If \( p \) is a prime such that \( p \not\in \{p_1, p_2, \ldots, p_u\} \), then
\[
v_p(n) = 0. \tag{263}
\]
- We have \( n = p_1^{a_1} p_2^{a_2} \cdots p_u^{a_u} \).

If \( p \) is a prime such that \( p \not\in \{p_1, p_2, \ldots, p_u\} \), then
\[
\begin{align*}
p^{0w} + p^{1w} + \cdots + p^{v_p(n)w} &= p^{0w} + p^{1w} + \cdots + p^{0w} \quad \text{(since (263) yields } v_p(n) = 0) \\
&= p^{0w} = p^0 = 1. \tag{264}
\end{align*}
\]

Define a set \( T \) as in Lemma 2.18.3. Then, Lemma 2.18.3 says that the map
\[
\Lambda : T \to \{\text{positive divisors of } n\},
\]
\[
(b_1, b_2, \ldots, b_u) \mapsto p_1^{b_1} p_2^{b_2} \cdots p_u^{b_u}
\]
is well-defined and bijective. Thus, this map \( \Lambda \) is a bijection.

Now, the summation sign \( \sum_{d|n} \) stands for a sum over all positive divisors of \( n \), and thus

\(^{177}\)Such a list exists, since \( P \) is finite.
is equivalent to the summation sign "\( \sum_{d \in \{ \text{positive divisors of } n \}} \)". Hence,

\[
\sum_{d|n} d^w = \sum_{d \in \{ \text{positive divisors of } n \}} d^w = \sum_{(b_1, b_2, \ldots, b_u) \in T} \left( \Lambda \left( b_1, b_2, \ldots, b_u \right) \right)^w \\
\text{(by the definition of } \Lambda \text{)}
\]

\[
= \sum_{(b_1, b_2, \ldots, b_u) \in T} \left( \prod_{i=1}^{u} p_i^{b_i w} \right)
\]

(265)

Now, Lemma 2.18.6 (applied to \( u, \{0, 1, \ldots, a_i\} \) and \( p_i^{k_i w} \) instead of \( n, Z_i \) and \( p_i^{k_i} \)) yields

\[
\prod_{i=1}^{u} \sum_{k \in \{0, 1, \ldots, a_i\}} p_i^{k_i w} = \sum_{(k_1, k_2, \ldots, k_u) \in \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\}} \prod_{i=1}^{u} p_i^{k_i w}
\]

\[
= \sum_{(k_1, k_2, \ldots, k_u) \in \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\}} \prod_{i=1}^{u} p_i^{k_i w}
\]

(since \( \{0, 1, \ldots, a_1\} \times \{0, 1, \ldots, a_2\} \times \cdots \times \{0, 1, \ldots, a_u\} = T \)). Comparing this with (265), we find

\[
\sum_{d|n} d^w = \prod_{i=1}^{u} \sum_{k \in \{0, 1, \ldots, a_i\}} p_i^{k_i w} = \prod_{i=1}^{u} \left( p_i^0 w + p_i^1 w + \cdots + p_i^{a_i w} \right).
\]
Comparing this with

\[
\prod_{p \text{ prime}} \left( p^{0w} + p^{1w} + \cdots + p^{i_p(n)\cdot w} \right)
\]

\[
= \left( \prod_{p \text{ prime}; p \in \{p_1, p_2, \ldots, p_u\}} \left( p^{0w} + p^{1w} + \cdots + p^{i_p(n)\cdot w} \right) \right) \left( \prod_{p \text{ prime}; p \notin \{p_1, p_2, \ldots, p_u\}} \left( p^{0w} + p^{1w} + \cdots + p^{i_p(n)\cdot w} \right) \right)
\]

(by 264)

\[
\text{(since each prime } p \text{ satisfies either } p \in \{p_1, p_2, \ldots, p_u\} \text{ or } p \notin \{p_1, p_2, \ldots, p_u\} \text{ (but not both simultaneously)})
\]

\[
= \left( \prod_{p \text{ prime}; p \in \{p_1, p_2, \ldots, p_u\}} \left( p^{0w} + p^{1w} + \cdots + p^{i_p(n)\cdot w} \right) \right) \left( \prod_{p \text{ prime}; p \notin \{p_1, p_2, \ldots, p_u\}} 1 \right)
\]

\[
\text{(since each } p \in \{p_1, p_2, \ldots, p_u\} \text{ is a prime)}
\]

\[
= \prod_{p \in \{p_1, p_2, \ldots, p_u\}} \left( p^{0w} + p^{1w} + \cdots + p^{i_p(n)\cdot w} \right) = \prod_{i=1}^{u} \left( p_i^{0w} + p_i^{1w} + \cdots + p_i^{i_p(n)\cdot w} \right)
\]

\[
\text{(since (262) yields } v_i(n)=a_i)\]

\[
\text{here, we have substituted } p_i \text{ for } p \text{ in the product,}
\]

\[
\text{since the map } \{1, 2, \ldots, u\} \to \{p_1, p_2, \ldots, p_u\}, \ i \mapsto p_i \text{ is a bijection}
\]

\[
= \prod_{i=1}^{u} \left( p_i^{0w} + p_i^{1w} + \cdots + p_i^{i_p(n)\cdot w} \right),
\]

we obtain

\[
\sum_{d|n} d^w = \prod_{p \text{ prime}} \left( p^{0w} + p^{1w} + \cdots + p^{i_p(n)\cdot w} \right).
\]

Now, forget that we fixed \(w\). We thus have proven that every \(w \in \mathbb{Z}\) satisfies

\[
\sum_{d|n} d^w = \prod_{p \text{ prime}} \left( p^{0w} + p^{1w} + \cdots + p^{i_p(n)\cdot w} \right).
\]

Renaming the index \(w\) as \(k\) in this statement, we conclude that every \(k \in \mathbb{Z}\) satisfies

\[
\sum_{d|n} d^k = \prod_{p \text{ prime}} \left( p^{0k} + p^{1k} + \cdots + p^{i_p(n)\cdot k} \right).
\]

This solves Exercise 2.18.1(b).
6.74. Solution to Exercise 3.3.1

Solution to Exercise 3.3.1 Let \( \alpha \) and \( \beta \) be two equivalence classes of \( \sim \). Thus, \( \alpha = [x]_\sim \) and \( \beta = [y]_\sim \) for two elements \( x \) and \( y \) of \( S \) (by the definition of “equivalence classes of \( \sim \)”). Consider these \( x \) and \( y \).

If \( x \sim y \), then the classes \([x]_\sim\) and \([y]_\sim\) are identical (by Theorem 3.3.5(a)). Otherwise, they are disjoint (by Theorem 3.3.5(b)). Thus, in either case, the classes \([x]_\sim\) and \([y]_\sim\) are either identical or disjoint. In view of \( \alpha = [x]_\sim \) and \( \beta = [y]_\sim \), this rewrites as follows: The classes \( \alpha \) and \( \beta \) are either identical or disjoint.

Now, forget that we fixed \( \alpha \) and \( \beta \). We thus have shown that if \( \alpha \) and \( \beta \) are two equivalence classes of \( \sim \), then \( \alpha \) and \( \beta \) are either identical or disjoint. This solves Exercise 3.3.1. \( \square \)

6.75. Solution to Exercise 3.3.2

Solution to Exercise 3.3.2 Indeed:

- The relation \( \sim \) is reflexive.
  
  [Proof: Informally, this is obvious, because each \( k \)-tuple is a permutation of itself (just permute it by leaving all its entries in place). The formal version of this argument proceeds as follows:]

  Let \( a \in A^k \). Write the \( k \)-tuple \( a \) in the form \( a = (a_1, a_2, \ldots, a_k) \) for some \( a_1, a_2, \ldots, a_k \in A \). Then, \( a = (a_1, a_2, \ldots, a_k) = (a_{id(1)}, a_{id(2)}, \ldots, a_{id(k)}) \). Hence, the \( k \)-tuple \( a \) has the form \( (a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(k)}) \) for some permutation \( \sigma \) of the set \( \{1, 2, \ldots, k\} \) (namely, for \( \sigma = id \)). In other words, \( a \) is a permutation of the \( k \)-tuple \( (a_1, a_2, \ldots, a_k) \) (by Definition 2.13.16). In other words, \( a \) is a permutation of the \( k \)-tuple \( a \) (since \( a = (a_1, a_2, \ldots, a_k) \)). In other words, \( a \sim a \) (by the definition of the relation \( \sim \)).

  Now, forget that we fixed \( a \). We thus have proven that every \( a \in A^k \) satisfies \( a \sim a \). In other words, the relation \( \sim \) is reflexive.]

- The relation \( \sim \) is symmetric.
  
  [Proof: Let \( a, b \in A^k \) be such that \( a \sim b \). We shall prove that \( b \sim a \). Write the \( k \)-tuple \( a \) in the form \( a = (a_1, a_2, \ldots, a_k) \) for some \( a_1, a_2, \ldots, a_k \in A \). Write the \( k \)-tuple \( b \) in the form \( b = (b_1, b_2, \ldots, b_k) \) for some \( b_1, b_2, \ldots, b_k \in A \). We have \( a \sim b \). In other words, \( a \) is a permutation of \( b \) (by the definition of the relation \( \sim \)). In other words, \( (a_1, a_2, \ldots, a_k) \) is a permutation of \( (b_1, b_2, \ldots, b_k) \) (since \( a = (a_1, a_2, \ldots, a_k) \) and \( b = (b_1, b_2, \ldots, b_k) \)). Hence, Proposition 2.13.18 (applied to \( p_i = b_i \) and \( q_i = a_i \)) shows that \( (b_1, b_2, \ldots, b_k) \) is a permutation of \( (a_1, a_2, \ldots, a_k) \). In other words, \( b \) is a permutation of \( a \) (since \( a = (a_1, a_2, \ldots, a_k) \) and \( b = (b_1, b_2, \ldots, b_k) \)). In other words, \( b \sim a \).]
Forget that we fixed \( a \) and \( b \). We thus have shown that every \( a, b \in A^k \) satisfying \( a \sim b \) satisfy \( b \sim a \). In other words, the relation \( \sim \) is symmetric.

- The relation \( \sim \) is transitive.

[Proof: Let \( a, b, c \in A^k \) be such that \( a \sim b \) and \( b \sim c \). We shall prove that \( a \sim c \).

Write the \( k \)-tuple \( a \) in the form \( a = (a_1, a_2, \ldots, a_k) \) for some \( a_1, a_2, \ldots, a_k \in A \). Write the \( k \)-tuple \( b \) in the form \( b = (b_1, b_2, \ldots, b_k) \) for some \( b_1, b_2, \ldots, b_k \in A \). Write the \( k \)-tuple \( c \) in the form \( c = (c_1, c_2, \ldots, c_k) \) for some \( c_1, c_2, \ldots, c_k \in A \).

We have \( a \sim b \). In other words, \( a \) is a permutation of \( b \) (by the definition of the relation \( \sim \)). In other words, \( (a_1, a_2, \ldots, a_k) \) is a permutation of \( (b_1, b_2, \ldots, b_k) \) (since \( a = (a_1, a_2, \ldots, a_k) \) and \( b = (b_1, b_2, \ldots, b_k) \)). In other words, the \( k \)-tuple \( (a_1, a_2, \ldots, a_k) \) has the form \( (b_{\sigma(1)}, b_{\sigma(2)}, \ldots, b_{\sigma(k)}) \) for some permutation \( \sigma \) of the set \( \{1, 2, \ldots, k\} \) (by Definition 2.13.16). Consider this \( \sigma \), and denote it by \( \lambda \). Thus, \( \lambda \) is a permutation of \( \{1, 2, \ldots, k\} \) and has the property that \( (a_1, a_2, \ldots, a_k) = (b_{\lambda(1)}, b_{\lambda(2)}, \ldots, b_{\lambda(k)}) \). Likewise, we can find a permutation \( \mu \) of \( \{1, 2, \ldots, k\} \) with the property that \( (b_1, b_2, \ldots, b_k) = (c_{\mu(1)}, c_{\mu(2)}, \ldots, c_{\mu(k)}) \) (because of our assumption that \( b \sim c \)). Consider this \( \mu \) as well.

Now \( \mu \) and \( \lambda \) are permutations of the set \( \{1, 2, \ldots, k\} \), that is, bijective maps \( \{1, 2, \ldots, k\} \rightarrow \{1, 2, \ldots, k\} \). Hence, their composition \( \mu \circ \lambda \) is a bijective map \( \{1, 2, \ldots, k\} \rightarrow \{1, 2, \ldots, k\} \) as well, i.e., is a permutation of the set \( \{1, 2, \ldots, k\} \).

Recall that \( (b_1, b_2, \ldots, b_k) = (c_{\mu(1)}, c_{\mu(2)}, \ldots, c_{\mu(k)}) \). In other words, each \( j \in \{1, 2, \ldots, k\} \) satisfies

\[
b_j = c_{\mu(j)}.
\]

Also, \( (a_1, a_2, \ldots, a_k) = (b_{\lambda(1)}, b_{\lambda(2)}, \ldots, b_{\lambda(k)}) \). Hence, each \( i \in \{1, 2, \ldots, k\} \) satisfies

\[
a_i = b_{\lambda(i)} = c_{\mu(\lambda(i))} \quad \text{(by (266), applied to } j = \lambda(i))
\]

\[
= c_{(\mu \circ \lambda)(i)}.
\]

In other words, we have \( (a_1, a_2, \ldots, a_k) = (c_{(\mu \circ \lambda)(1)}, c_{(\mu \circ \lambda)(2)}, \ldots, c_{(\mu \circ \lambda)(k)}) \). Hence, the \( k \)-tuple \( (a_1, a_2, \ldots, a_k) \) has the form \( (c_{\sigma(1)}, c_{\sigma(2)}, \ldots, c_{\sigma(k)}) \) for some permutation \( \sigma \) of the set \( \{1, 2, \ldots, k\} \) (namely, for \( \sigma = \mu \circ \lambda \)). In other words, the \( k \)-tuple \( (a_1, a_2, \ldots, a_k) \) is a permutation of \( (c_1, c_2, \ldots, c_k) \) (by Definition 2.13.16). In other words, \( a \) is a permutation of \( c \) (since \( a = (a_1, a_2, \ldots, a_k) \) and \( c = (c_1, c_2, \ldots, c_k) \)). In other words, \( a \sim c \).

Forget that we fixed \( a, b, c \). We thus have shown that every \( a, b, c \in A^k \) satisfying \( a \sim b \) and \( b \sim c \) satisfy \( a \sim c \). In other words, the relation \( \sim \) is transitive.

We now know that the relation \( \sim \) is reflexive, symmetric and transitive. In other words, \( \sim \) is an equivalence relation. This solves Exercise 3.3.2.

\( \Box \)
6.76. Solution to Exercise 3.3.3

Solution to Exercise 3.3.3. Let \( T \) be the quotient set \( S / \sim \), and let \( f : S \to T \) be the canonical projection \( \pi : S \to S / \sim \). We shall prove that the relation \( \sim \) equals the relation \( \equiv_f \).

Let \( a \in S \) and \( b \in S \). Recall that \( f \) is the map \( \pi_\sim \). Thus, \( f(a) = \pi_\sim(a) = [a]_\sim \) (by the definition of \( \pi_\sim \)). The same argument (applied to \( b \) instead of \( a \)) yields \( f(b) = [b]_\sim \).

Theorem 3.3.5(e) (applied to \( x = a \) and \( y = b \)) shows that we have \( a \sim b \) if and only if \( [a]_\sim = [b]_\sim \). In other words, we have the logical equivalence

\[
(a \sim b) \iff ([a]_\sim = [b]_\sim) .
\]

We have the following chain of logical equivalences:

\[
\begin{align*}
(a \equiv_f b) & \iff \left( f(a) = f(b) \right) \quad \text{(by the definition of the relation } \equiv_f) \\
& \iff \left( [a]_\sim = [b]_\sim \right) \iff (a \sim b) \quad \text{(by (267))} .
\end{align*}
\]

Now, forget that we fixed \( a \) and \( b \). We thus have proven the equivalence (268) for all \( a \in S \) and \( b \in S \). Now, recall that we have defined a relation on the set \( S \) to be a subset of \( S \times S \) (namely, the subset of all pairs \((a, b) \in S \times S \) satisfying this relation). Hence, the relation \( \sim \) is actually the subset \( \{(a, b) \in S \times S \mid a \sim b\} \) of \( S \times S \), whereas the relation \( \equiv_f \) is actually the subset \( \{(a, b) \in S \times S \mid a \equiv_f b\} \) of \( S \times S \). But these two subsets are clearly identical, because we have shown the equivalence (268) for all \( a \in S \) and \( b \in S \). In other words, the two relations \( \sim \) and \( \equiv_f \) are identical. In other words, the relation \( \sim \) equals the relation \( \equiv_f \). This solves Exercise 3.3.3.

6.77. Solution to Exercise 3.4.1

Solution to Exercise 3.4.1. (a) Let \( r \) be a nonzero rational number. We must prove that the integer \( w_p(r) \) is well-defined. Recall that we have defined \( w_p(r) \) by setting \( w_p(r) = v_p(a) - v_p(b) \), where we write \( r \) in the form \( r = a/b \) for two nonzero integers \( a \) and \( b \). In order to prove that \( w_p(r) \) is well-defined, we must thus verify the following three claims:

Claim 1: It is possible to write \( r \) in the form \( r = a/b \) for two nonzero integers \( a \) and \( b \).
Claim 2: If we write \( r = a/b \) for two nonzero integers \( a \) and \( b \), then \( v_p (a) - v_p (b) \) is a well-defined integer.\(^{178}\)

Claim 3: If we write \( r = a/b \) for two nonzero integers \( a \) and \( b \), then the integer \( v_p (a) - v_p (b) \) depends only on \( p \) and \( r \) (but not on \( a \) and \( b \)).

Claim 1 and Claim 2 are easy to verify:

[Proof of Claim 1: We know that \( r \) is a rational number. Hence, we can write \( r = c/d \) for some integer \( c \) and some nonzero integer \( d \). Consider these \( c \) and \( d \). If we had \( c = 0 \), then we would have \( r = -\infty \); but this would contradict the fact that \( r \) is nonzero. Hence, we cannot have \( c = 0 \). Thus, \( c \) is nonzero. Thus, there exist two nonzero integers \( a \) and \( b \) such that \( r = a/b \) (namely, \( a = c \) and \( b = d \)). In other words, it is possible to write \( r \) in the form \( r = a/b \) for two nonzero integers \( a \) and \( b \). This proves Claim 1.]

[Proof of Claim 2: Assume that \( r \) is written in the form \( r = a/b \) for two nonzero integers \( a \) and \( b \). Definition \([2.13.23](a)\) shows that \( v_p (n) \in \mathbb{N} \) for each nonzero integer \( n \). Thus, \( v_p (a) \in \mathbb{N} \) (since \( a \) is nonzero) and \( v_p (b) \in \mathbb{N} \) (since \( b \) is nonzero). Hence, \( v_p (a) - v_p (b) \in \mathbb{N} \).

In other words, \( v_p (a) - v_p (b) \) is a well-defined integer. This proves Claim 2.]

It remains to prove Claim 3. Clearly, Claim 3 can be restated as follows:

Claim 4: Let \((a_1, b_1)\) and \((a_2, b_2)\) be two pairs \((a, b)\) of nonzero integers \( a \) and \( b \) satisfying \( r = a/b \). Then, \( v_p (a_1) - v_p (b_1) = v_p (a_2) - v_p (b_2) \).

[Proof of Claim 4: We have assumed that \((a_1, b_1)\) is a pair \((a, b)\) of nonzero integers \( a \) and \( b \) satisfying \( r = a/b \). In other words, \((a_1, b_1)\) is a pair of nonzero integers satisfying \( r = a_1/b_1 \). Similarly, \((a_2, b_2)\) is a pair of nonzero integers satisfying \( r = a_2/b_2 \).

We have \( r = a_1/b_1 \), thus \( a_1/b_1 = r = a_2/b_2 \). Multiplying this equality by \( b_1 b_2 \), we find \( a_1 b_2 = a_2 b_1 \). Theorem \([2.13.28](a)\) (applied to \( a = a_1 \) and \( b = b_2 \)) yields \( v_p (a_1 b_2) = v_p (a_1) + v_p (b_2) \). Theorem \([2.13.28](a)\) (applied to \( a = a_2 \) and \( b = b_1 \)) yields \( v_p (a_2 b_1) = v_p (a_2) + v_p (b_1) \). Now, from \( v_p (a_1 b_2) = v_p (a_1) + v_p (b_2) \), we obtain

\[
\begin{align*}
v_p (a_1) + v_p (b_2) &= v_p \left( \frac{a_1 b_2}{a_2 b_1} \right) = v_p (a_2 b_1) = v_p (a_2) + v_p (b_1).
\end{align*}
\]

But \( b_1 \) is a nonzero integer (since \((a_1, b_1)\) is a pair of nonzero integers); thus, \( v_p (b_1) \in \mathbb{N} \) (since Definition \([2.13.23](a)\) shows that \( v_p (n) \in \mathbb{N} \) for each nonzero integer \( n \)). Similarly, \( v_p (b_2) \in \mathbb{N} \). Hence, \( v_p (b_1) + v_p (b_2) \in \mathbb{N} \). Thus, we can subtract \( v_p (b_1) + v_p (b_2) \) from both sides of the equality \([269]\)\(^{179}\). We thus obtain \( v_p (a_1) - v_p (b_1) = v_p (a_2) - v_p (b_2) \). This proves Claim 4.]

\(^{178}\)This needs saying, because \( p \)-valuations can be \( \infty \) and thus their differences may fail to be well-defined integers (for example, \( \infty - \infty \) is not even well-defined).

\(^{179}\)The reason why we are so circumspect about this is that \( p \)-valuations can be \( \infty \), and \( \infty \) cannot be subtracted. So when subtracting a \( p \)-valuation, it is important to ensure that this \( p \)-valuation is an element of \( \mathbb{N} \) (that is, it is not \( \infty \).
As we recall, Claim 4 is just a restatement of Claim 3. Hence, Claim 3 is proven (since Claim 4 is proven). From Claims 1, 2 and 3, we conclude that $w_p(r)$ is well-defined. Thus, Exercise 3.4.1 (a) is solved.

Let us state a consequence of the definition of $w_p(r)$: If $r$ is a nonzero rational number, and if $a$ and $b$ are two nonzero integers satisfying $r = a/b$, then

$$w_p(r) = v_p(a) - v_p(b).$$

(b) Let $n$ be a nonzero integer. We must prove that $w_p(n) = v_p(n)$.

We know that $n$ and 1 are two nonzero integers satisfying $n = n/1$. Hence, (270) (applied to $r = n$, $a = n$ and $b = 1$) yields

$$w_p(n) = v_p(n) - v_p(1) = v_p(n).$$

(by Theorem 2.13.28 (c))

This solves Exercise 3.4.1 (b).

(c) Let $a$ and $b$ be two nonzero rational numbers. We must show that $w_p(ab) = w_p(a) + w_p(b)$.

We know that $a$ is a rational number. Thus, we can write $a$ in the form $a = n_1/d_1$ for some integer $n_1$ and some nonzero integer $d_1$. Consider these $n_1$ and $d_1$. If we had $n_1 = 0$, then we would have $a = n_1/d_1 = 0$, which would contradict the assumption that $a$ is nonzero. Hence, we do not have $n_1 = 0$. In other words, $n_1$ is nonzero. Thus, (270) (applied to $a$, $n_1$ and $d_1$ instead of $r$, $a$ and $b$) yields

$$w_p(a) = v_p(n_1) - v_p(d_1).$$

We know that $b$ is a rational number. Thus, we can write $b$ in the form $b = n_2/d_2$ for some integer $n_2$ and some nonzero integer $d_2$. Consider these $n_2$ and $d_2$. If we had $n_2 = 0$, then we would have $b = n_2/d_2 = 0$, which would contradict the assumption that $b$ is nonzero. Hence, we do not have $n_2 = 0$. In other words, $n_2$ is nonzero. Thus, (270) (applied to $b$, $n_2$ and $d_2$ instead of $r$, $a$ and $b$) yields

$$w_p(b) = v_p(n_2) - v_p(d_2).$$

From $a = n_1/d_1$ and $b = n_2/d_2$, we obtain

$$ab = (n_1/d_1)(n_2/d_2) = (n_1n_2)/(d_1d_2).$$

Moreover, the integer $n_1n_2$ is nonzero (since $n_1$ and $n_2$ are nonzero), and the integer $d_1d_2$ is nonzero (since $d_1$ and $d_2$ are nonzero). Hence, (270) (applied to $ab$, $n_1n_2$ and $d_1d_2$ instead
of \( r, a \) and \( b \) yields

\[
\begin{align*}
  w_p(ab) &= v_p(n_1n_2) - v_p(d_1d_2) \\
  &= v_p(n_1) + v_p(n_2) - v_p(d_1) - v_p(d_2) \\
  &= w_p(a) + w_p(b).
\end{align*}
\]

This solves Exercise 3.4.1 (c).

(d) Let \( a \) and \( b \) be two nonzero rational numbers such that \( a + b \neq 0 \). We must show that \( w_p(a+b) \geq \min\{w_p(a), w_p(b)\} \).

Note that \( a+b \) is a nonzero rational number (since \( a+b \neq 0 \)); thus, \( w_p(a+b) \) is well-defined.

We know that \( a \) is a rational number. Thus, we can write \( a \) in the form \( a = n_1/d_1 \) for some integer \( n_1 \) and some nonzero integer \( d_1 \). Consider these \( n_1 \) and \( d_1 \). Then, \( n_1 \) is nonzero (this is proven in the same way as in our solution to Exercise 3.4.1 (a)). Thus, (270) (applied to \( a, n_1 \) and \( d_1 \) instead of \( r, a \) and \( b \)) yields

\[
w_p(a) = v_p(n_1) - v_p(d_1). \tag{273}
\]

We know that \( b \) is a rational number. Thus, we can write \( b \) in the form \( b = n_2/d_2 \) for some integer \( n_2 \) and some nonzero integer \( d_2 \). Consider these \( n_2 \) and \( d_2 \). Then, \( n_2 \) is nonzero (this is proven in the same way as in our solution to Exercise 3.4.1 (d)). Thus, (270) (applied to \( b, n_2 \) and \( d_2 \) instead of \( r, a \) and \( b \)) yields

\[
w_p(b) = v_p(n_2) - v_p(d_2). \tag{274}
\]

From \( a = n_1/d_1 \) and \( b = n_2/d_2 \), we obtain

\[
a + b = (n_1/d_1) + (n_2/d_2) = (n_1d_2 + n_2d_1) / (d_1d_2).
\]

Moreover, the integer \( n_1d_2 + n_2d_1 \) is nonzero (because otherwise, we would have \( n_1d_2 + n_2d_1 = 0 \) and therefore \( a + b = (n_1d_2 + n_2d_1) / (d_1d_2) = 0 \), which would contradict \( a + b \neq 0 \)), and the integer \( d_1d_2 \) is nonzero (since \( d_1 \) and \( d_2 \) are nonzero). Hence, (270) (applied to
\[ w_p (a + b) = v_p (n_1 d_2 + n_2 d_1) - v_p (d_1 d_2) \]
\[ \geq \min \left\{ v_p (n_1 d_2), v_p (n_2 d_1) \right\} \]
(by Theorem 2.13.28(b), applied to \( n_1 d_2 \) and \( n_2 d_1 \) instead of \( a \) and \( b \))
\[ = v_p (n_1) + v_p (d_2) \]
(by Theorem 2.13.28(a), applied to \( n_1 \) and \( d_2 \) instead of \( a \) and \( b \))
\[ = v_p (n_2) + v_p (d_1) \]
(by Theorem 2.13.28(a), applied to \( n_2 \) and \( d_1 \) instead of \( a \) and \( b \))
\[ = v_p (d_1) + v_p (d_2) \]
(by Theorem 2.13.28(a), applied to \( d_1 \) and \( d_2 \) instead of \( a \) and \( b \))
\[ = \min \left\{ v_p (n_1) + v_p (d_2), v_p (n_2) + v_p (d_1) \right\} - (v_p (d_1) + v_p (d_2)) \]
(275)

But it is easy to see that any three numbers \( i, j, k \in \mathbb{N} \) satisfy
\[ \min \{ i, j \} - k = \min \{ i - k, j - k \} \]
(276)

Also, it is easy to see that the three numbers \( v_p (n_1) + v_p (d_2), v_p (n_2) + v_p (d_1), v_p (d_1) + v_p (d_2) \) belong to \( \mathbb{N} \) (since \( n_1, n_2, d_1, d_2 \) are all nonzero). Hence, (276) (applied to \( i = v_p (n_1) + v_p (d_2), j = v_p (n_2) + v_p (d_1) \) and \( k = v_p (d_1) + v_p (d_2) \)) yields
\[ \min \left\{ v_p (n_1) + v_p (d_2), v_p (n_2) + v_p (d_1) \right\} - (v_p (d_1) + v_p (d_2)) \]
\[ = \min \left\{ v_p (n_1) + v_p (d_2) - (v_p (d_1) + v_p (d_2)), v_p (n_2) + v_p (d_1) - (v_p (d_1) + v_p (d_2)) \right\} \]
(by 273)
\[ = \min \{ w_p (a), w_p (b) \} \]
(by 274)

Thus, (275) becomes
\[ w_p (a + b) \geq \min \left\{ v_p (n_1) + v_p (d_2), v_p (n_2) + v_p (d_1) \right\} - (v_p (d_1) + v_p (d_2)) \]
\[ = \min \{ w_p (a), w_p (b) \} . \]

This solves Exercise 3.4.1(d). \( \Box \)

Proof of (276): Let \( i, j, k \in \mathbb{N} \) be three numbers. We must prove the equality (276). We can WLOG assume that \( i \leq j \) (since \( i \) and \( j \) play symmetric roles in our claim, and thus swapping \( i \) with \( j \) will not change anything). Assume this. Hence, \( i - k \leq j - k \), thus \( \min \{ i - k, j - k \} = i - k \).

Comparing this with \( \min \{ i, j \} - k = i - k \), we obtain \( \min \{ i, j \} - k = \min \{ i - k, j - k \} \). This proves (276).
6.78. Solution to Exercise 3.5.1

Solution to Exercise 3.5.1 (a) The inverse $\alpha^{-1}$ of $\alpha$ exists (since $\alpha$ has an inverse). We have $\alpha \cdot \alpha^{-1} = [1]_n$ (since $\alpha^{-1}$ is an inverse of $\alpha$). But Theorem 3.4.23(e) (applied to $\alpha^{-1}$ and $\alpha$ instead of $\alpha$ and $\beta$) yields $\alpha^{-1} \cdot \alpha = \alpha \cdot \alpha^{-1} = [1]_n$. In other words, $\alpha$ is an inverse of $\alpha^{-1}$. Thus, $\alpha^{-1}$ has an inverse (namely, $\alpha$). Therefore, we can speak of “the inverse of $\alpha^{-1}$.

Moreover, $(\alpha^{-1})^{-1} = (\text{the inverse of } \alpha^{-1}) = \alpha$ (since $\alpha$ is an inverse of $\alpha^{-1}$). This solves Exercise 3.5.1(a).

(b) The inverse $\alpha^{-1}$ of $\alpha$ exists (since $\alpha$ has an inverse). We have $\alpha \cdot \alpha^{-1} = [1]_n$ (since $\alpha^{-1}$ is an inverse of $\alpha$). The inverse $\beta^{-1}$ of $\beta$ exists (since $\beta$ has an inverse). We have $\beta \cdot \beta^{-1} = [1]_n$ (since $\beta^{-1}$ is an inverse of $\beta$). Theorem 3.4.23(e) yields $\alpha \cdot \beta = \beta \cdot \alpha$. In other words, $\alpha \beta = \beta \alpha$.

Theorem 3.4.23(f) (applied to $\beta$, $\alpha$ and $\alpha^{-1}$) yields $\beta \cdot (\alpha \cdot \alpha^{-1}) = (\beta \cdot \alpha) \cdot \alpha^{-1} = (\beta \alpha) \cdot \alpha^{-1}$, so that

$$
(\beta \alpha) \cdot \alpha^{-1} = \beta \cdot \left( \alpha \cdot \alpha^{-1} \right) = \beta \cdot [1]_n = \beta
$$

(by Theorem 3.4.23(d)).

Now, Theorem 3.4.23(f) (applied to $\alpha \beta$, $\alpha^{-1}$ and $\beta^{-1}$ instead of $\alpha$, $\beta$ and $\gamma$) yields

$$
(\alpha \beta) \cdot (\alpha^{-1} \beta^{-1}) = \left( \underbrace{(\alpha \beta) \cdot \alpha^{-1}}_{= \beta \alpha} \right) \cdot \beta^{-1} = \left( \underbrace{(\beta \alpha) \cdot \alpha^{-1}}_{= \beta} \right) \cdot \beta^{-1} = \beta \cdot \beta^{-1} = [1]_n
$$

In other words, $\alpha^{-1} \beta^{-1}$ is an inverse of $\alpha \beta$. Thus, $\alpha \beta$ has an inverse (namely, $\alpha^{-1} \beta^{-1}$). Therefore, we can speak of “the inverse of $\alpha \beta$”. Moreover, $(\alpha \beta)^{-1} = (\text{the inverse of } \alpha \beta) = \alpha^{-1} \beta^{-1}$ (since $\alpha^{-1} \beta^{-1}$ is an inverse of $\alpha \beta$). This solves Exercise 3.5.1(b).

[Remark: In the above solution, we have avoided writing products of the form $\alpha_1 \alpha_2 \cdots \alpha_k$ with more than 2 factors without explicitly placing parentheses. This was done for the purpose of making each single use of associativity explicit. Had we instead written such products without parenthesization, our solution would have become much simpler; namely, we could simply argue that

$$
(\alpha \beta) \cdot (\alpha^{-1} \beta^{-1}) = \left( \underbrace{\alpha \beta}_{= \beta \alpha} \right) \cdot \left( \underbrace{\alpha^{-1} \beta^{-1}}_{= \beta} \right) = \beta \cdot [1]_n = \beta^{-1} = \beta \beta^{-1} = [1]_n
$$

(by Theorem 3.4.23(d)).

Computations of this kind are perfectly legitimate, because Proposition 3.4.25 shows that products of the form $\alpha_1 \alpha_2 \cdots \alpha_k$ are well-defined and satisfy the standard rules (which include the one saying that $\alpha_1 \alpha_2 \cdots \alpha_k = (\alpha_1 \alpha_2 \cdots \alpha_i) (\alpha_{i+1} \alpha_{i+2} \cdots \alpha_k)$ for each $i \in \{0, 1, \ldots, k\}$).]
6.79. Solution to Exercise 4.1.1

Proof of Proposition 4.1.19 (a) Let $\alpha \in \mathbb{C}$ and $n \in \mathbb{N}$. The definition of $\alpha^n$ yields $\alpha^n = \overbrace{\alpha \cdot \cdots \cdot \alpha}^{n \text{ times}}$. The definition of $\alpha^{n+1}$ yields

$$\alpha^{n+1} = \underbrace{\alpha \cdot \cdots \cdot \alpha}_{n+1 \text{ times}}.$$  

Comparing this with

$$\alpha \cdot \underbrace{\alpha \cdot \cdots \cdot \alpha}_{n \text{ times}} = \overbrace{\alpha \cdot \cdots \cdot \alpha}^{n \text{ times}} = \alpha^n,$$

we obtain $\alpha^{n+1} = \alpha^n \alpha$. This proves Proposition 4.1.19 (a).

(b) First proof of Proposition 4.1.19 (b): Let $\alpha \in \mathbb{C}$ and $n, m \in \mathbb{N}$. Definition 4.1.17 (a) yields $\alpha^n = \overbrace{\alpha \cdot \cdots \cdot \alpha}^{n \text{ times}}$ and $\alpha^m = \overbrace{\alpha \cdot \cdots \cdot \alpha}^{m \text{ times}}$. Multiplying these two equalities, we obtain

$$\alpha^n \alpha^m = \overbrace{\alpha \cdot \cdots \cdot \alpha}^{n \text{ times}} \cdot \overbrace{\alpha \cdot \cdots \cdot \alpha}^{m \text{ times}} = \overbrace{\alpha \cdot \cdots \cdot \alpha}^{n+m \text{ times}}.$$

Comparing this with

$$\alpha^{n+m} = \overbrace{\alpha \cdot \cdots \cdot \alpha}^{n+m \text{ times}}$$  

we obtain $\alpha^{n+m} = \alpha^n \alpha^m$. This proves Proposition 4.1.19 (b).

Second proof of Proposition 4.1.19 (b): We shall prove Proposition 4.1.19 (b) by induction on $n$:

Induction base: If $\alpha \in \mathbb{C}$ and $m \in \mathbb{N}$, then $\alpha^0 \cdot \alpha^m = 1 \cdot \alpha^m = \alpha^m = \alpha^{0+m}$ (since $m = 0 + m$).

In other words, we have $\alpha^{0+m} = \alpha^0 \alpha^m$ for all $\alpha \in \mathbb{C}$ and $m \in \mathbb{N}$. In other words, Proposition 4.1.19 (b) holds for $n = 0$. This completes the induction base.

Induction step: Let $k \in \mathbb{N}$. Assume that Proposition 4.1.19 (b) holds for $n = k$. We must prove that Proposition 4.1.19 (b) holds for $n = k + 1$.

Let $\alpha \in \mathbb{C}$ and $m \in \mathbb{N}$. We have assumed that Proposition 4.1.19 (b) holds for $n = k$. Hence, we can apply Proposition 4.1.19 (b) to $n = k$. We thus obtain $\alpha^{k+m} = \alpha^k \alpha^m$.

Proposition 4.1.19 (a) (applied to $n = k$) yields $\alpha^{k+1} = \alpha \alpha^k$. Also, Proposition 4.1.19 (a) (applied to $n = k+m$) yields $\alpha^{(k+m)+1} = \alpha \alpha^{k+m}$.

But $(k+1) + m = (k+m) + 1$ and hence

$$\alpha^{(k+1)+m} = \alpha^{(k+m)+1} = \alpha \alpha^{k+m} = \alpha \underbrace{(\alpha^k \alpha^m)}_{=\alpha^k \alpha^m} = \underbrace{(\alpha^k)}_{=\alpha^k} \underbrace{\alpha^m}_{=\alpha^m} = \alpha^{k+1} \alpha^m.$$  

Now, forget that we fixed $\alpha$ and $m$. We thus have shown that $\alpha^{(k+1)+m} = \alpha^{k+1} \alpha^m$ for all $\alpha \in \mathbb{C}$ and $m \in \mathbb{N}$. In other words, Proposition 4.1.19 (b) holds for $n = k + 1$. This completes the induction step. Thus, Proposition 4.1.19 (b) is proven by induction.
(c) We shall prove Proposition 4.1.19 (c) by induction on \( n \):

**Induction base:** Let \( \alpha, \beta \in C \). Then, \((\alpha\beta)^0 = 1 = \alpha^0 \beta^0 \) (since \( \alpha^0 = \beta^0 = 1 \cdot 1 = 1 \)).

Forget that we fixed \( \alpha, \beta \). We thus have shown that \((\alpha\beta)^0 = \alpha^0 \beta^0 \) for all \( \alpha, \beta \in C \). In other words, Proposition 4.1.19 (c) holds for \( n = 0 \). This completes the induction base.

**Induction step:** Let \( k \in \mathbb{N} \). Assume that Proposition 4.1.19 (c) holds for \( n = k \). We must prove that Proposition 4.1.19 (c) holds for \( n = k + 1 \).

Let \( \alpha, \beta \in C \). We have assumed that Proposition 4.1.19 (c) holds for \( n = k \). Hence, \((\alpha\beta)^k = \alpha^k \beta^k \). But Proposition 4.1.19 (a) (applied to \( n = k \)) yields \( \alpha^{k+1} = \alpha^k \). Similarly, \( \beta^{k+1} = \beta^k \). Multiplying these two equalities together, we obtain

\[
\alpha^{k+1} \beta^{k+1} = (\alpha \alpha^k) \left( \beta \beta^k \right) = \left( (\alpha \alpha^k) \beta \right) \beta^k
\]

(by Theorem 4.1.2 (f)). But Proposition 4.1.19 (a) (applied to \( \alpha \beta \) and \( k \) instead of \( \alpha \) and \( n \)) yields\(^{181}\)

\[
(\alpha \beta)^{k+1} = (\alpha \beta) (\alpha \beta)^k = (\alpha \beta) (\alpha^k \beta^k) = \alpha \left( \beta \left( \alpha^k \beta^k \right) \right) = \alpha \left( \beta \left( \alpha \alpha^k \beta \right) \right) = \alpha \left( \beta \left( \alpha \alpha^k \beta^k \right) \right)
\]

\[
= \alpha \left( \beta \alpha^k \right) \beta^k = \alpha \left( \beta \alpha^k \right) \beta^k = \alpha \left( \beta \alpha^k \right) \beta^k.
\]

Comparing this with \( (277) \), we obtain \((\alpha \beta)^{k+1} = \alpha^{k+1} \beta^{k+1} \).

Now, forget that we fixed \( \alpha, \beta \). We thus have shown that \((\alpha \beta)^{k+1} = \alpha^{k+1} \beta^{k+1} \) for all \( \alpha, \beta \in C \). In other words, Proposition 4.1.19 (c) holds for \( n = k + 1 \). This completes the induction base.

(d) We shall prove Proposition 4.1.19 (d) by induction on \( m \):

**Induction base:** For all \( \alpha \in C \) and \( n \in \mathbb{N} \), we have \((\alpha^n)^0 = 1 = \alpha^0 = \alpha^0 \) (since \( 0 = n \cdot 0 \)). In other words, Proposition 4.1.19 (d) holds for \( m = 0 \). This completes the induction base.

**Induction step:** Let \( k \in \mathbb{N} \). Assume that Proposition 4.1.19 (d) holds for \( m = k \). We must prove that Proposition 4.1.19 (d) holds for \( m = k + 1 \).

The following computation could be much shorter if I used unparenthesized products (i.e., if I wrote “\( \alpha \beta \alpha^k \beta^k \)” instead of explicitly invoking the associativity of multiplication to move between various ways of parenthesizing this product). I am avoiding unparenthesized products here for the purpose of illustrating how to live without them; in the future I will simply use them wherever they can help.

\[^{181}\text{The following computation could be much shorter if I used unparenthesized products (i.e., if I wrote “}\alpha \beta \alpha^k \beta^k\text{” instead of explicitly invoking the associativity of multiplication to move between various ways of parenthesizing this product). I am avoiding unparenthesized products here for the purpose of illustrating how to live without them; in the future I will simply use them wherever they can help.}\]
Let \( \alpha \in \mathbb{C} \) and \( n \in \mathbb{N} \). Then, \((\alpha^n)^k = \alpha^{nk}\) (since Proposition 4.1.19(d) holds for \( m = k \)).

Now, Proposition 4.1.19(a) (applied to \( \alpha^n \) and \( k \) instead of \( \alpha \) and \( n \)) yields
\[
(\alpha^n)^{k+1} = \alpha^{nk} \cdot \alpha^{nk} = \alpha^{nk+1}.
\]

Comparing this with
\[
\alpha^{n(k+1)} = \alpha^{n+nk} = \alpha^n \cdot \alpha^n = \alpha^n \cdot \alpha^{nk},
\]
we obtain \((\alpha^n)^{k+1} = \alpha^{n(k+1)}\).

Now, forget that we fixed \( \alpha \) and \( n \). We thus have shown that \((\alpha^n)^{k+1} = \alpha^{n(k+1)}\) for all \( \alpha \in \mathbb{C} \) and \( n \in \mathbb{N} \). In other words, Proposition 4.1.19(d) holds for \( m = k + 1 \). Hence, Proposition 4.1.19(d) is proven by induction.

(e) We shall prove Proposition 4.1.19(e) by induction on \( n \):

**Induction base:** Proposition 4.1.19(e) holds for \( n = 0 \) (since \( 1^0 = 1 \)). This completes the induction base.

**Induction step:** Let \( k \in \mathbb{N} \). Assume that Proposition 4.1.19(e) holds for \( n = k \). We must prove that Proposition 4.1.19(e) holds for \( n = k + 1 \).

We have assumed that Proposition 4.1.19(e) holds for \( n = k \). In other words, \( 1^k = 1 \).

Now, Proposition 4.1.19(a) (applied to \( \alpha = 1 \) and \( n = k \)) yields \( 1^{k+1} = 1 \cdot 1^k = 1 \cdot 1 = 1 \).

In other words, Proposition 4.1.19(e) holds for \( n = k + 1 \). This completes the induction step. Hence, Proposition 4.1.19(e) is proven by induction.

(f) Let \( \alpha \in \mathbb{C} \) be nonzero. Let \( n \in \mathbb{Z} \). We must prove that \( \alpha^{n+1} = \alpha \alpha^n \).

Recall that \( \alpha^{-1} \) is the inverse of \( \alpha \); hence, \( \alpha \alpha^{-1} = 1 \) (by the definition of “inverse”).

We are in one of the following three cases:

**Case 1:** We have \( n > -1 \).

**Case 2:** We have \( n = -1 \).

**Case 3:** We have \( n < -1 \).

Let us first consider Case 1. In this case, we have \( n > -1 \). Thus, \( n \geq 0 \) (since \( n \) is an integer), so that \( n \in \mathbb{N} \). Hence, Proposition 4.1.19(a) yields \( \alpha^{n+1} = \alpha \alpha^n \). Thus, \( \alpha^{n+1} = \alpha \alpha^n \) is proven in Case 1.

Let us next consider Case 2. In this case, we have \( n = -1 \). Thus, \( n + 1 = 0 \), so that \( \alpha^{n+1} = \alpha^0 = 1 \). On the other hand, from \( n = -1 \), we obtain \( \alpha \alpha^n = \alpha \alpha^{-1} = 1 \). Comparing this with \( \alpha^{n+1} = 1 \), we find \( \alpha^{n+1} = \alpha \alpha^n \). Hence, \( \alpha^{n+1} = \alpha \alpha^n \) is proven in Case 2.

Let us finally consider Case 3. In this case, we have \( n < -1 \). Hence, \( n + 1 < 0 \). Therefore, Definition 4.1.18 (applied to \( n + 1 \) instead of \( n \)) yields
\[
\alpha^{n+1} = \left( \alpha^{-1} \right)^{-(n+1)}.
\]

On the other hand, \( n + 1 < 0 \), hence \( -(n + 1) > 0 \) and thus \( -(n + 1) \in \mathbb{N} \). Hence, Proposition 4.1.19(a) (applied to \( \alpha^{-1} \) and \( -(n + 1) \) instead of \( \alpha \) and \( n \)) yields \( \left( \alpha^{-1} \right)^{-(n+1)+1} = \alpha^{-1} \).
\(\alpha^{-1}(\alpha^{-1})^{-(n+1)}\). On the other hand, \(n < -1 < 0\). Hence, Definition 4.1.18 yields
\[
\alpha^n = (\alpha^{-1})^{-n} = (\alpha^{-1})^{-(n+1)+1} = \alpha^{-1} \alpha^{n+1}.
\]

Thus,
\[
\alpha \alpha^{n} = \alpha \alpha^{-1} \alpha^{n+1} = \alpha^{n+1}.
\]

In other words, \(\alpha^{n+1} = \alpha \alpha^n\). Hence, \(\alpha^{n+1} = \alpha \alpha^n\) is proven in Case 3.

We have now proven \(\alpha^{n+1} = \alpha \alpha^n\) in all three Cases 1, 2 and 3. Thus, \(\alpha^{n+1} = \alpha \alpha^n\) always holds. This proves Proposition 4.1.19 (f).

(g) Let \(\alpha \in \mathbb{C}\) be nonzero. Let \(n \in \mathbb{Z}\). We must prove that \(\alpha^{-n} = (\alpha^{-1})^n\).

We are in one of the following three Cases:

Case 1: We have \(n > 0\).
Case 2: We have \(n = 0\).
Case 3: We have \(n < 0\).

Let us first consider Case 1. In this case, we have \(n > 0\). Hence, \(-n < 0\). In other words, \(-n\) is negative. Hence, Definition 4.1.18 (applied to \(-n\) instead of \(n\)) yields \(\alpha^{-n} = (\alpha^{-1})^{-(n)} = (\alpha^{-1})^n\) (since \(- (n) = n\)). Hence, \(\alpha^{-n} = (\alpha^{-1})^n\) is proven in Case 1.

Let us next consider Case 2. In this case, we have \(n = 0\). Hence, \(\alpha^{-n} = \alpha^0 = \alpha^0 = 1\) and \((\alpha^{-1})^n = (\alpha^{-1})^0 = 1\). Comparing these two equalities, we obtain \(\alpha^{-n} = (\alpha^{-1})^n\). Hence, \(\alpha^{-n} = (\alpha^{-1})^n\) is proven in Case 2.

Let us next consider Case 3. In this case, we have \(n < 0\). In other words, \(n\) is negative.

Recall that \(\alpha^{-1}\) is the inverse of \(\alpha\); hence, \(a \alpha^{-1} = 1\) (by the definition of "inverse"). Hence, \(\alpha^{-1} \neq 0\) (since otherwise, we would have \(\alpha^{-1} = 0\) and thus \(1 = \alpha \alpha^{-1} = 0\), which would be absurd). Hence, Definition 4.1.18 (applied to \(\alpha^{-1}\) instead of \(\alpha\)) yields \((\alpha^{-1})^n = \left((\alpha^{-1})^{-1}\right)^{-n} = \alpha^{-n}\), since \(n\) is negative.

But Proposition 4.1.15 (a) yields \((\alpha^{-1})^{-1} = \alpha\). Thus, \((\alpha^{-1})^n = \left((\alpha^{-1})^{-1}\right)^{-n} = \alpha^{-n}\), so that \(\alpha^{-n} = (\alpha^{-1})^n\). Hence, \(\alpha^{-n} = (\alpha^{-1})^n\) is proven in Case 3.

We have now proven that \(\alpha^{-n} = (\alpha^{-1})^n\) in all three Cases 1, 2 and 3. Thus, \(\alpha^{-n} = (\alpha^{-1})^n\) always holds. This completes the proof of Proposition 4.1.19 (g).

(h) We must prove that \(\alpha^{n+m} = \alpha^n \alpha^m\) for all nonzero \(\alpha \in \mathbb{C}\) and all \(n, m \in \mathbb{Z}\). We shall first prove the following less general result:

**Claim 1:** We have \(\alpha^{n+m} = \alpha^n \alpha^m\) for all nonzero \(\alpha \in \mathbb{C}\) and all \(n \in \mathbb{N}\) and \(m \in \mathbb{Z}\).

**Proof of Claim 1:** The following proof is very similar to our Second proof of Proposition 4.1.19 (b) above.

We shall prove Claim 1 by induction on \(n\):
**Induction base:** If \( \alpha \in \mathbb{C} \) is nonzero, and if \( m \in \mathbb{Z} \), then \( \alpha^0 \cdot \alpha^m = 1 \cdot \alpha^m = \alpha^m = \alpha^{0+m} \) (since \( m = 0 + m \)). In other words, we have \( \alpha^{0+m} = \alpha^0 \alpha^m \) for all nonzero \( \alpha \in \mathbb{C} \) and all \( m \in \mathbb{Z} \). In other words, Claim 1 holds for \( n = 0 \). This completes the induction base.

**Induction step:** Let \( k \in \mathbb{N} \). Assume that Claim 1 holds for \( n = k \). We must prove that Claim 1 holds for \( n = k + 1 \).

Let \( \alpha \in \mathbb{C} \) be nonzero, and let \( m \in \mathbb{Z} \). We have assumed that Claim 1 holds for \( n = k \). Hence, we can apply Claim 1 to \( \alpha \cdot \alpha^{k+m} = \alpha^{k+1} \cdot \alpha^m \). This completes the induction step. Thus, Claim 1 is proven by induction.

Now, forget that we fixed \( m \). We thus have shown that \( \alpha^{(k+1)+m} = \alpha^{k+1} \cdot \alpha^m \) for all nonzero \( \alpha \in \mathbb{C} \) and all \( m \in \mathbb{Z} \). In other words, Claim 1 holds for \( n = k + 1 \). This completes the induction step. Thus, Claim 1 is proven by induction.

Now, let us prove Proposition 4.1.19 (h) in full generality: Fix a nonzero \( \alpha \in \mathbb{C} \). Fix \( n, m \in \mathbb{Z} \). We must prove that \( \alpha^{n+m} = \alpha^n \cdot \alpha^m \). If \( n \in \mathbb{N} \), then this follows from Claim 1. Hence, for the rest of this proof, we WLOG assume that we don’t have \( n \in \mathbb{N} \). Combining this with \( n \in \mathbb{Z} \), we obtain \( n \in \mathbb{Z} \setminus \mathbb{N} = \{ -1, -2, -3, \ldots \} \). In other words, \( n \) is negative. Hence, \( -n \) is positive. Thus, \( -n \in \mathbb{N} \).

Recall that \( -1 \) is the inverse of \( \alpha \); hence, \( \alpha \cdot \alpha^{-1} = 1 \) (by the definition of “inverse”). Hence, \( \alpha^{-1} \neq 0 \) (since otherwise, we would have \( \alpha^{-1} = 0 \) and thus \( 1 = \alpha \cdot \alpha^{-1} = 0 \), which would be absurd). So we have \( \alpha^{-1} \neq 0 \) and \( -n \in \mathbb{N} \). Hence, Claim 1 (applied to \( \alpha^{-1}, -n \) and \(-m\) instead of \( \alpha, n \) and \( m \)) yields

\[
\left( \alpha^{-1} \right)^{-n} \cdot \left( \alpha^{-1} \right)^{-m} = \left( \alpha^{-1} \right)^{-(n+m)} = \left( \alpha^{-1} \right)^{-(n+m)}.
\]

(279)

Definition 4.1.18 yields \( \alpha^n = (\alpha^{-1})^{-n} \) (since \( n \) is negative). Proposition 4.1.19 (g) (applied to \(-m\) instead of \( n \)) yields \( \alpha^{-n} \cdot \alpha^{-m} = (\alpha^{-1})^{-n} \cdot \alpha^{-m} \). In view of \( -(-m) = m \), this rewrites as \( \alpha^{-n} \cdot \alpha^{-m} = (\alpha^{-1})^{-n} \cdot \alpha^{-m} \). The same argument (applied to \( n + m \) instead of \( m \)) yields \( \alpha^{n+m} = (\alpha^{-1})^{-(n+m)} \).

Hence,

\[
\alpha^{n+m} = (\alpha^{-1})^{-n} \cdot (\alpha^{-1})^{-m} = (\alpha^{-1})^{-n} \cdot (\alpha^{-1})^{-m} \quad \text{(since } n + m = -n + -m) \quad \text{(by 279)}.
\]

Comparing this with

\[
\frac{\alpha^n}{\alpha^{-n}} = (\alpha^{-1})^{-n} = (\alpha^{-1})^{-n},
\]

we obtain \( \alpha^{n+m} = \alpha^n \cdot \alpha^m \). This completes our proof of Proposition 4.1.19 (h).

(i) Let \( \alpha, \beta \in \mathbb{C} \) be nonzero. Let \( n \in \mathbb{Z} \). We must prove that \( (\alpha \beta)^n = \alpha^n \beta^n \).
If \( n \in \mathbb{N} \), then this follows from Proposition 4.1.19 (c). Hence, for the rest of this proof, we WLOG assume that we don’t have \( n \in \mathbb{N} \). Combining this with \( n \in \mathbb{Z} \), we obtain
\( n \in \mathbb{Z} \setminus \mathbb{N} = \{ -1, -2, -3, \ldots \} \). In other words, \( n \) is negative. Hence, \( -n \) is positive. Thus, \( -n \in \mathbb{N} \). Hence, Proposition 4.1.19 (c) (applied to \(-n\), \( \alpha^{-1} \) and \( \beta^{-1} \) instead of \( n\), \( \alpha \) and \( \beta \)) yields
\[
(\alpha^{-1}\beta^{-1})^{-n} = (\alpha^{-1})^{-n}(\beta^{-1})^{-n}.
\]

Proposition 4.1.15 (b) yields that \( (\alpha\beta)^{-1} = \alpha^{-1}\beta^{-1} \). Also, Corollary 4.1.16 shows that \( \alpha\beta \) is nonzero. Recall also that \( n \) is negative. Thus, Definition 4.1.18 (applied to \( \alpha \)) yields \( (\alpha\beta)^n = \left( (\alpha\beta)^{-1} \right)^{-n} = (\alpha^{-1}\beta^{-1})^{-n} = (\alpha^{-1})^{-n}(\beta^{-1})^{-n} \).

(280)

On the other hand, Definition 4.1.18 yields \( \alpha^n = (\alpha^{-1})^{-n} \) (since \( n \) is negative). Similarly, \( \beta^n = (\beta^{-1})^{-n} \). Multiplying these two equalities, we obtain
\[
\alpha^n\beta^n = (\alpha^{-1})^{-n}(\beta^{-1})^{-n}.
\]
Comparing this with (280), we find \( (\alpha\beta)^n = \alpha^n\beta^n \). This completes our proof of Proposition 4.1.19 (i).

(j) Let \( n \in \mathbb{Z} \). We must prove that \( 1^n = 1 \). If \( n \in \mathbb{N} \), then this follows from Proposition 4.1.19 (e). Hence, for the rest of this proof, we WLOG assume that we don’t have \( n \in \mathbb{N} \). Combining this with \( n \in \mathbb{Z} \), we obtain \( n \in \mathbb{Z} \setminus \mathbb{N} = \{ -1, -2, -3, \ldots \} \). In other words, \( n \) is negative. Hence, \( -n \) is positive. Thus, \( -n \in \mathbb{N} \). Hence, Proposition 4.1.19 (e) (applied to \(-n\) instead of \( n\)) yields \( 1^{-n} = 1 \).

Also, \( 1 \cdot 1 = 1 \). Hence, 1 is an inverse of 1 (by the definition of “inverse”). Thus, the inverse of 1 is 1. In other words, \( 1^{-1} = 1 \).

But \( n \) is negative. Thus, Definition 4.1.18 (applied to \( \alpha = 1 \)) yields \( 1^n = \left( 1^{-1} \right)^{-n} = 1^{-n} = 1 \). Hence, Proposition 4.1.19 (j) is proven.

(k) Let \( \alpha \in \mathbb{C} \) be nonzero. Let \( n \in \mathbb{Z} \).

Recall that \( \alpha^{-1} \) is the inverse of \( \alpha \); hence, \( \alpha\alpha^{-1} = 1 \) (by the definition of “inverse”). Proposition 4.1.19 (i) (applied to \( \beta = \alpha^{-1} \)) yields \( (\alpha\alpha^{-1})^n = \alpha^n (\alpha^{-1})^n \). Hence, \( \alpha^n (\alpha^{-1})^n = \left( \alpha^{-1} \right)^{-n} = 1^n = 1 \) (by Proposition 4.1.19 (j)). If we had \( \alpha^n = 0 \), then we would have \( \alpha^n (\alpha^{-1})^n = 0 \), which would contradict \( \alpha^n (\alpha^{-1})^n = 1 \neq 0 \). Hence, we cannot have \( \alpha^n = 0 \). In other words, \( \alpha^n \neq 0 \). Thus, \( \alpha^n \) is nonzero. Hence, the inverse \((\alpha^n)^{-1}\) of \( \alpha^n \) is well-defined.

Moreover, the equality \( \alpha^n (\alpha^{-1})^n = 1 \) shows that \( (\alpha^{-1})^n \) is an inverse of \( \alpha^n \) (by the definition of “inverse”). In other words, the inverse of \( \alpha^n \) is \( (\alpha^{-1})^n \). In other words, \( (\alpha^n)^{-1} = (\alpha^{-1})^n \). But Proposition 4.1.19 (g) yields \( \alpha^{-n} = (\alpha^{-1})^n \). Comparing these two equalities, we obtain \( (\alpha^n)^{-1} = \alpha^{-n} \). Thus, Proposition 4.1.19 (k) is proven.

(l) We must prove that \( (\alpha^n)^m = \alpha^{nm} \) for all nonzero \( \alpha \in \mathbb{C} \) and all \( n,m \in \mathbb{Z} \).

We shall first prove this in lesser generality:

**Claim 2:** We have \( (\alpha^n)^m = \alpha^{nm} \) for all nonzero \( \alpha \in \mathbb{C} \) and all \( n,m \in \mathbb{Z} \) and \( m \in \mathbb{N} \).
Thus, and that \( \alpha \) is invertible, and thus \( n \in \alpha \). We must prove that \( (a^{n})^{-1} = a^{-n} \). In particular, \( (a^{n})^{-1} \) is well-defined, so that \( a^{n} \) is invertible, and thus \( a^{n} \) is nonzero. Hence, Proposition 4.1.19 (k) (applied to \( a^{n} \) and \( -m \) instead of \( \alpha \) and \( n \)) yields

\[
\left( (a^{n})^{-m} \right)^{-1} = (a^{n})^{-(m)} = (a^{n})^{m} \quad \text{ (since } -m = m \text{)}.
\]

Thus,

\[
(a^{n})^{m} = \left( \left( a^{n} \right)^{-m} \right)^{-1} = (a^{m})^{-1}.
\]

But Proposition 4.1.19 (k) (applied to \( -nm \) instead of \( n \)) yields

\[
(a^{m})^{-1} = a^{-(nm)} = a^{nm} \quad \text{ (since } -(nm) = nm \text{)}.
\]

Hence, \( (a^{n})^{m} = (a^{m})^{-1} = a^{nm} \). This completes the proof of Proposition 4.1.19 (l).

(m) This is proven in the same way as the usual binomial formula (i.e., Theorem 2.17.13) is proven. (The only difference is that we are now calling our two numbers \( \alpha \) and \( \beta \) rather than \( x \) and \( y \), and that we now need to use Theorem 4.1.2 and Proposition 4.1.19 (a) instead of the analogous rules for real numbers.)
References


[daSilv12] Patrick Da Silva, *Polynomial in Q[x] sending integers to integers?*, math.stackexchange answer #108318
https://www.math.vt.edu/people/day/ProofsBook/IPaMV.pdf.


https://www.cis.upenn.edu/~jean/RSA-primality-testing.pdf

http://homepage.divms.uiowa.edu/~goodman/algebrabook.dir/algebrabook.html


The numbering of theorems and formulas in this link might shift when the project gets updated; for a “frozen” version whose numbering is guaranteed to match that in the citations above, see https://github.com/darijgr/detnotes/releases/tag/2019-01-10.

http://www.cip.ifi.lmu.de/~grinberg/floor.pdf

http://www.cip.ifi.lmu.de/~grinberg/lucascong.pdf
https://neil-strickland.staff.shef.ac.uk/courses/MAS201/MAS201.pdf

[Swanso18] Irena Swanson, *Introduction to Analysis*, with construction of the number systems, 21 August 2018.
http://people.reed.edu/~iswanson/analysisconstructR.pdf


