Computer Science

Dr. Y. Liow (March 10, 2023)

Contents

Chapter 202

Elementary number theory

```
File: chap.tex
```

```
File: basic-number-theory.tex
```

202.1 Elementary number theory

The goal of this chapter is to study elementary number theory, the study of the theory of $\mathbb{Z}$.

The area of number theory – the study of whole numbers – is unbelievably huge. There are actually many "types" of number theory. Elementary number theory is one of them. There are also algebraic number theory, analytic number theory, combinatorial number theory, probabilistic number theory, diophantine geometry, combinatorial number theory, etc. There are many areas of study which does not contains the phrase "number theory" which also arise from number theory. Examples include the theory of modular forms, theory of automorphic forms, arithmetic algebraic geometry, etc.

"Elementary" number theory means you use basic algebraic rules of $\mathbb{Z}$, in fact frequently only $\mathbb{N}$. Just to let you know "elementary" number theory does not mean it is easy! But what we'll cover is considered basic. For sure you study some elementary number theory before you can even talk about algebraic number theory or analytic number theory.

I will only cover enough facts for us to understand RSA. RSA is built on the difficulty of prime factorization. That's why ... we need to look at number theory.

You should think of $\mathbb{Z}$ not as a set alone, but that it comes with the addition $+$ and multiplication $\cdot$ operation. Furthermore you should also think of the number 0 and 1 as being significant for these two operations. So as a whole we want to think of

$$(\mathbb{Z}, +, \cdot, 0, 1)$$

as a whole system. (If you like, you can think of it as a C++ class: Think of objects of $\mathbb{Z}$ as whole numbers with the $+$ and $\cdot$ operations and 0 and 1 as special static objects.) Also for simplicity I will write xy for $x \cdot y$.

Anyway here are the facts for $\mathbb{Z}$ ...

The following lists some basic algebraic information about $\mathbb{Z}$. Note that I will not list specific facts such as $1 + 1 = 2$. I'll be listed general algebraic rules.

1. If $x, y \in \mathbb{Z}$ then $x + y \in \mathbb{Z}$
2. If $x, y, z \in \mathbb{Z}$, then $(x + y) + z = x + (y + z)$
3. If $x \in \mathbb{Z}$, then $x + 0 = x = 0 + x$
4. If $x \in \mathbb{Z}$, there is some $y \in \mathbb{Z}$ such that $x + y = 0 = y + x$.
5. If $x, y \in \mathbb{Z}$, then $x + y = y + x$.
6. If $x, y \in \mathbb{Z}$, then $xy \in \mathbb{Z}$.
7. If $x, y, z \in \mathbb{Z}$, then $(xy)z = x(yz)$.
8. If $x \in \mathbb{Z}$, then $x1 = x = 1x$.
9. If $x, y \in \mathbb{Z}$, then $xy = yx$.
10. If $x, y, z \in \mathbb{Z}$, then $x(y + z) = xy + xz$.
11. If $x, y \in \mathbb{Z}$, then $xy = 0$ implies $x = 0$ or $y = 0$.

If you look at the above carefully, you'll see that you can break up the above to three parts. Part 1 has to do with addition:

1. If $x, y \in \mathbb{Z}$ then $x + y \in \mathbb{Z}$
2. If $x, y, z \in \mathbb{Z}$, then $(x + y) + z = x + (y + z)$
3. If $x \in \mathbb{Z}$, then $x + 0 = x = 0 + x$
4. If $x \in \mathbb{Z}$, there is some $y \in \mathbb{Z}$ such that $x + y = 0 = y + x$.
5. If $x, y \in \mathbb{Z}$, then $x + y = y + x$.

Note that the above involves $+$ and 0. Part 2 has to do with multiplication:

1. If $x, y \in \mathbb{Z}$, then $xy \in \mathbb{Z}$.
2. If $x, y, z \in \mathbb{Z}$, then $(xy)z = x(yz)$.
3. If $x \in \mathbb{Z}$, then $x1 = x = 1x$.
4. If $x, y \in \mathbb{Z}$, then $xy = yx$.

(Compare this with the above and you'll see that the pair up. Except for one. Do you see which one is the odd one?) And there's one that's involves $+$ and $\cdot$:

1. If $x, y, z \in \mathbb{Z}$, then $x(y + z) = xy + xz$.

And there's one that's by it's own because it involves $\cdot$ but also involves 0:

1. If $x, y \in \mathbb{Z}$, then $xy = 0$ implies $x = 0$ or $y = 0$.

It's actually more efficient to study the operations separately. I'm going to start by focusing on $+$ and focus on

$$(\mathbb{Z}, +, 0)$$

(Note that except for the last rule, 0 is always tied to $+$.) By compartmentalizing the different parts of $\mathbb{Z}$, the facts and the dependencies between the facts become clearer.

But the study of $\mathbb{Z}$ with only just $+$ is not just to make the theory clearer. A set with only one operation actually happens very frequently so that the generalized theory of $\mathbb{Z}$ with $+$, called group theory, can be applied to many other scenarios.

Although we are moving toward the RSA cipher which depends on elementary number theory of $\mathbb{Z}$, the concept of groups is also used in another type of ciphers called group-based cipher or discrete log ciphers.

Breaking up an object of study into small parts in done not just for number theory, but it happens in any area of study. You even see that in writing software when you organize functions into different libraries or when you create classes.

So we'll start with basic group theory.

File: introduction-to-group-theory.tex

202.2 Introduction to group theory

202.2.1 Definition of a group

Definition 202.2.1. A **group** is a triple $(G, *, e)$ where G is a set such that group

1. If $x, y \in G$, then $x * y \in G$, i.e., $*$ is a binary operator of G, i.e., $* : G \times G \to G$ is a function.
2. If $x, y, z \in G$, then $(x * y) * z = x * (y * z)$.
3. If $x \in G$, then there is some $y \in G$ such that $x * y = e = y * x$. y is called an **inverse** of x. (In fact the inverse of x is unique to x. See proposition below. So I can say x^{-1} is *the* inverse of x.) inverse
4. If $x \in G$, then $e * x = x = x * e$.

The element e is called an **identity** of G. (In fact e is unique to G. See proposition below. Therefore I can say e is *the* identity element of G.) e is also called a **neutral** element of G. □ identity neutral

Here's a memory aid. Label the group axioms for a group in the following way:

1. Closure: If $x, y \in G$, then $x * y \in G$.
2. Associativity: If $x, y, z \in G$, then $(x * y) * z = x * (y * z)$.
3. Inverse: If $x \in G$, then there is some $y \in G$ such that $x * y = e = y * x$. y is called an **inverse** of x. (In fact the inverse of x is unique to x. See proposition below.) inverse
4. Neutral: If $x \in G$, then $e * x = x = x * e$.

and you get CAIN.

Of course you can use other notations for the group operation $*$. For instance you can use $\cdot$, $\oplus$, $+$, etc. If you use $+$, it's common to use 0 to denote the identity element so that G looks more or less like $\mathbb{Z}$ or $\mathbb{Z}/N$. If you use $*$ or $\cdot$, we say that you are using multiplicative notation and if you use $+$ we say you are using additive notation. If you use the multiplicative notation such as $*$ and $\cdot$, it's also common to write xy as a shorthand instead of $x \cdot y$ or $x * y$. Just read the context carefully.

Definition 202.2.2. $(G, *, e)$ is an **abelian group** if $(G, *, e)$ is a group and abelian group if $x, y \in G$, then

$$x * y = y * x$$

Frequently I want to write a long chain of group operation on the same element g, such as $g * g * g * g * g$. Here's a shorthand:

Definition 202.2.3. Let $(G, *, e)$ be a group and let $x \in G$. For $n \geq 0$, We define x^n as follows:

$$x^n = \begin{cases} e & \text{if } n = 0 \\ x & \text{if } n = 1 \\ x^{n-1} * x & \text{if } n > 1 \end{cases}$$

If $n < 0$, then

$$x^n = \left(x^{-1}\right)^{-n}$$

for instance x^{-3} means $(x^{-1})^3$.

If you use the additive notation, then instead of writing x^n, you would write nx.

Here are some very simple examples of groups:

1. $(\mathbb{Z}, +, 0)$ is an abelian group.
2. $(\mathbb{Z}/N, +, 0)$ is an abelian group where $\mathbb{Z}/N$ is the set of $\mathbb{Z}$ modulo N where $N > 0$.
3. $(\mathbb{Q}, +, 0)$ is an abelian group.
4. $(\mathbb{R}, +, 0)$ is an abelian group.
5. $(\mathbb{C}, +, 0)$ is an abelian group.

It can be slightly more complicated.

1. $(\mathbb{R}[X], +, 0)$ is an abelian group where $\mathbb{R}[X]$ is the set of polynomials with $\mathbb{R}$ coefficients.
2. $(\mathbb{Z}/5 - \{0\}, \cdot, 1)$ is an abelian group where $\cdot$ is the usual multiplication.

And these are even more complicated because they are non-abelian:

1. $(\mathrm{GL}_2(\mathbb{R}), \cdot, I)$. Recall that $\mathrm{GL}_2(\mathbb{R})$ is the set of invertible 2-by-2 matrices with $\mathbb{R}$ coefficients. Here $\cdot$ is matrix multiplication and I is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
2. Let $n \geq 1$ be an integer. Let S_n denote the set of bijections $\{1, ..., n\} \to \{1, \ldots, n\}$ Then $(S_n, \circ, \mathrm{id})$ is a group where $\circ$ is function composition and id is the identity map, i.e., $\mathrm{id}(i) = i$ for $i \in \{1, ..., n\}$. S_n is called

a **symmetric group**. See notes on permutation cipher for the cyclic notation for describing permutations of $\{1, ..., n\}$.

symmetric group

When the group operation and the identity is known, usually we'll reference a group by mentioning the set of values. For instance I might say "Consider the group $\mathbb{Z}$" instead of "Consider the group $(\mathbb{Z}, +, 0)$".

You can fully describe the behavior of a finite group by drawing the complete description of the group operation. For instance suppose $G = \{e\}$ is a group of size 1. Then of course we must have

$$e * e = e$$

Here's a picture of $*$:

$*$	e
e	e

The above is called the **group table** for the group. Make sure you check that this $G = \{e\}$ satisfies all the group axioms.

group table

Suppose G is a group of size 2. Let me call it C_2. Say $C_2 = \{e, a\}$. Note that in drawing the group table, it's common to list the identity e first. Here's the unfilled group table for $C_2 = \{e, a\}$:

$*$	e	a
e		
a		

Using the identity group axiom, I get

$*$	e	a
e	e	a
a	a	

The last square is either e or a. I claim that it must be e:

$*$	e	a
e	e	a
a	a	e

Why? (You have 2 minutes.)

Exercise 202.2.1. Is the above group abelian?

202.2.2 Group isomorphism

Now let's write $(\mathbb{Z}/2, +, 0)$:

$+$	0	1
0	0	1
1	1	0

Here's C_2 again:

$*$	e	a
e	e	a
a	a	e

So you see that C_2 is essentially the same as $\mathbb{Z}/2$. The only difference is in the naming of the elements of the two groups and the symbol for the group operation. The technical term is C_2 is **(group) isomorphic** to $\mathbb{Z}/2$. In other words there is a bijection $f : C_2 \rightarrow \mathbb{Z}/2$ such that

(group) isomorphic

$$f(x * y) = f(x) + f(y)$$

for all $x, y \in C_2$.

Definition 202.2.4. Let $(G, *, e)$ and $(G', *', e')$ be two groups. G and G' are **isomorphic** if there is a bijection $f : G \rightarrow G'$ such that if $x, y \in G$,

isomorphic

$$f(x * y) = f(x) *' f(y)$$

In that case we write $G \simeq G'$. The function f itself is called a **group isomorphism**.

group isomorphism

Exercise 202.2.2.

1. Show that S_2 is isomorphic to $\mathbb{Z}/2$.
2. Write down the group table of S_3. Look at the permutation cipher for the cyclic notation for describing permutations of $\{1, 2, 3\}$. Is S_3 abelian?

Exercise 202.2.3.

1. Find all non-isomorphic groups of size 3. One of them is $\mathbb{Z}/3$.
2. Find all non-isomorphic groups of size 4. One of them is $\mathbb{Z}/4$. (Hint: There are two.)
3. Find all non-isomorphic groups of size 5. One of them is $\mathbb{Z}/5$.
4. Find all non-isomorphic groups of size 6. One of them is $\mathbb{Z}/6$. (Hint: There is a non-abelian group of order 6 and it has something to do with $3! = 6$ and has something to do with the permutations of 3 symbols. You might want to read the notes on permutation cipher.)
5. Find all non-isomorphic groups of size 7. One of them is $\mathbb{Z}/7$.

(The general case when the size is any positive integer is an extremely difficult problem.)

Recall that $(\mathbb{Z}/N)^\times$ is the set of integers x such that $0 \le x \le n-1$ and such that x has a multiplicative inverse in mod N, i.e., there is some y such that $xy \equiv 1 \pmod{N}$.

In fact $(\mathbb{Z}/N)^\times$ is not just a set, $(\mathbb{Z}/N)^\times$ is a subset of $\mathbb{Z}/N$ (which has addition and multiplication). Using the multiplication of $\mathbb{Z}/N$, $((\mathbb{Z}/N)^\times, \cdot, 1)$ forms an abelian group! Make sure you prove that!

Exercise 202.2.4. Show that $((\mathbb{Z}/N)^\times, \cdot, 1)$ is a finite abelian group.

Proof. Exercise. □

$(\mathbb{Z}/N)^\times$ is also denoted by $U(\mathbb{Z}/N)$.

It can be shown that if x is an integer and $0 \le x \le N-1$, then $x \in (\mathbb{Z}/N)^\times$ (has a multiplicative inverse) iff $\gcd(x, N) = 1$. Define the Euler-phi function ϕ as follows:

$$\phi(N) = \{x \mid 0 \le x \le N-1, \;\; \gcd(x, N) = 1\}$$

In other words

$$\phi(N) = |(\mathbb{Z}/N)^\times|$$

$\phi(N)$ can be computed using the following two formulas:

1. $\phi(mn) = \phi(m)\phi(n)$ if $\gcd(m, n) = 1$
2. $\phi(p^\alpha) = p^\alpha - p^{\alpha-1}$ where p is a prime and $\alpha > 0$ is an integer.

For instance

$$\phi(300) = \phi(2^2 \cdot 3 \cdot 5^2) = \phi(2^2) \cdot \phi(3) \cdot \phi(5^2) = (2^2 - 2^1)(3^1 - 3^0)(5^2 - 5^1)$$

which is 80. In other words in $\mathbb{Z}/300$ there are 80 values x such that $0 \le x < 300$ and have multiplicative inverses. Hence $|(\mathbb{Z}/300)^\times| = 80$. Of course the above depends on performing prime factorization on 300.

Exercise 202.2.5.

1. Write down the group table of $((\mathbb{Z}/5)^\times, \cdot, 1)$.
2. Write down the group table of $((\mathbb{Z}/6)^\times, \cdot, 1)$.
3. Write down the group table of $((\mathbb{Z}/8)^\times, \cdot, 1)$.

Exercise 202.2.6. Of course you can easily model a group using strings or integers as group elements and using functions can be modeled using dictionary. For instance

$*$	e	a
e	e	a
a	a	e

can be easily described by

```
e = 'e'
a = 'a'
G0 = [e, a]
d0 = {}
d0[(e, e)] = e
d0[(e, a)] = a
d0[(a, e)] = a
d0[(a, a)] = e
e0 = e
```

In the above `(G0, d0, e0)` describes a group.

1. Write a function `is_group` so that `is_group(G0, d0, e0)` returns true iff `(G0, d0, e0)` is a group.
2. Modify the above so that when you can also call the above function as `is_group(G0, d0)` in case the identity is not known.
3. Write a function `identity(G0, d0)` that returns the identity element of `G0` where the group operation of described by `d0`.
4. Write a function `inverse(G0, d0, x)` that returns the inverse of `x`.
5. Write a function `is_isomorphic(G0, d0, G1, d1)` that returns true if `G0` is isomorphic to `G1`. (You might also want to write a function that returns an isomorphism.)
6. Write a function that returns non-isomorphic groups of a given order. The function should accept a size for the groups and return a list of `(G0, d0, e0)`.

Proposition 202.2.1. *Let $(G, *, e)$ be a group. There is only one identity element in G. In other words, if $e, e' \in G$ satisfies the identity axiom:*

(a) *If $x \in G$, $e * x = x = x * e$.*
(b) *If $x \in G$, $e' * x = x = x * e'$.*

Then $e = e'$. □

Proof. Exercise. □

Proposition 202.2.2. *Let* $(G, *, e)$ *be a group. The inverse of* $x \in G$ *is unique. In other words, if* $y, y' \in G$ *satisfies the inverse axiom for* x*:*

(a) $y * x = e = x * y$
(b) $y' * x = e = x * y'$

i.e., y, y' *are inverses of* x*, then* $y = y'$*.* □

Proof. Exercise. □

Since the inverse of x is unique to x, we can write x^{-1} for the inverse of x. If the group operation is written as $+$, then it's more common to write the inverse as $-x$.

Proposition 202.2.3. *Let* $(G, *, e)$ *be a group. Let* $a, x, y \in G$*. If*

$$a * x = a * y$$

then

$$x = y$$

The above is sometimes called the cancellation property for groups.

Proof. Exercise. □

Exercise 202.2.7. Let $(G, *, e)$ be a group and let $x, y \in G$.

(a) $e^{-1} = e$
(b) $(x^{-1})^{-1} = x$
(c) $(x * y)^{-1} = y^{-1} * x^{-1}$
(d) $(x^n)^{-1} = (x^{-1})^n$

As an example of the power of abstract group theory, if M and N are invertible square matrices (of the same n-by-n size), then $(MN)^{-1} = N^{-1}M^{-1}$. This follows immediately from (c) above since the set of invertible n-by-n matrices forms a group.

Here's the proof of (b). Recall that uniqueness of inverse: if

$$x * y = e = y * x$$

then $y = x^{-1}$. Replacing x with x^{-1} and y by x, since

$$x^{-1} * x = e = x * x^{-1}$$

I get $x = (x^{-1})^{-1}$. All the others can be proven using the uniqueness of inverse except that for (d) you'll also need induction.

Exercise 202.2.8. Let $(G, *, e)$ be a group such that $x^2 = e$ for all $x \in G$. Show that G is an abelian group.

202.2.3 Subgroups

It's common to study a mathematical object by studying its sub-structures. For instance we study a graph by studying its subgraphs. This is the same for group theory:

Definition 202.2.5. Let $(G, *, e)$ be a group. Let H be a subset of G containing e. Using the same binary operator $*$ when restricted to H, assuming $x * y \in H$ for all $x, y \in H$, we say that $(H, *, e)$ is a **subgroup** of $(G, *, e)$ if $(H, *, e)$ is also a group. In that case I'll write $H \leq G$.

subgroup

For instance in $(\mathbb{Z}, +, 0)$, there's an obvious subgroup, the subgroup of even integers $(2\mathbb{Z}, +, 0)$ where

$$2\mathbb{Z} = \{..., 2 \cdot (-3), 2 \cdot (-2), 2 \cdot (-1), 2 \cdot 0, 2 \cdot 1, 2 \cdot 2, 2 \cdot 3, ...\} = \{2n \mid n \in \mathbb{Z}\}$$

For instance $2 \cdot 3, 2 \cdot 17 \in 2\mathbb{Z}$ and

$$2 \cdot 3 + 2 \cdot 17 = 2 \cdot 20 \in 2\mathbb{Z}$$

You should verify that $2\mathbb{Z}$ is a group and hence is a subgroup of $\mathbb{Z}$.

Exercise 202.2.9. Prove that $(2\mathbb{Z}, +, 0)$ is an abelian subgroup of $(\mathbb{Z}, +, 0)$.

Exercise 202.2.10. What are all the subgroups of $\mathbb{Z}$? (To prove that your list is correct will require facts that I'll prove later. Do you know which fact I'm talking about?)

Definition 202.2.6. Note that if $(G, *, e)$ is a group, then $\{e\}$ and G are subgroups of G. So you get two subgroups of G for free. $\{e\}$ is called the **trivial subgroup** of G. A **proper subgroup** H of G is a subgroup of G such that $|H| < |G|$.

trivial subgroup
proper subgroup

So of course the interesting subgroups of G are the nontrivial proper subgroups of G.

Proposition 202.2.4. *Let $(G, *, e)$ be a group and H be a nonempty subset of G. If for all $x, y \in H$, $x * y^{-1} \in H$, then H is a subgroup of G, i.e., $e \in H$ and $(H, *, e)$ satisfies the group axioms.*

Proof. Exercise.

Exercise 202.2.11. Find all the subgroups of $\mathbb{Z}/6$.

Exercise 202.2.12. Find all the subgroups of S_3.

202.2.4 Cyclic subgroups

Definition 202.2.7. Let $(G, *, e)$ be a group and let $g \in G$. Let $\langle g \rangle$ denote the set

$$\langle g \rangle = \{g^n \mid n \in \mathbb{Z}\}$$

where $g^0 = e$, $g^1 = g$, $g^2 = g * g$, $g^3 = g * g * g$, $g^{-3} = (g^{-1})^3$, etc. (See definition of g^n.)

Proposition 202.2.5. *Let $(G, *, e)$ be a group and let $g \in G$. Then $\langle g \rangle$ is a subgroup of G. It is the smallest subgroup of G containing g.*

Proof. Exercise.

Definition 202.2.8. In the above, $\langle g \rangle$ is called the **cyclic subgroup** of G generated by g. In general a **cyclic group** is a group G such that $G = \langle g \rangle$ for some $g \in G$ and we would say that G is **generated** by g and that g is a **generator** of G.

cyclic subgroup

cyclic group

generated

generator

Exercise 202.2.13. Consider $(\mathbb{Z}, +, 0)$.

1. What is the cyclic subgroups of $(\mathbb{Z}, +, 0)$ generated by 0, i.e. $\langle 0 \rangle$? Describe all its values.
2. What about $\langle 5 \rangle$? Describe all its values.
3. What about $\langle 6 \rangle$? Describe all its values.
4. Is $(\mathbb{Z}, +, 0)$ itself is a cyclic group? □

Let's look at the case of $(\mathbb{Z}/N)^\times$. $(\mathbb{Z}/2)^\times = \{1\}$. Therefore $(\mathbb{Z}/2)^\times$ is the trivial group $\langle 1 \rangle$. So it's cyclic. $(\mathbb{Z}/3)^\times = \{1, 2\}$. Note that in $(\mathbb{Z}/3)^\times$

$$\begin{aligned} &1^1 = 1 \\ &2^1 = 2, 2^2 = 4 \equiv 1 \pmod{3} \end{aligned}$$

Therefore

$$(\mathbb{Z}/3)^\times = \{2^0, 2^1\} = \langle 2 \rangle$$

is cyclic with one generator, i.e., 2.

Exercise 202.2.14. For each of the following, check if the group is cyclic. If it is, find all the generators. If not, find elements with the largest order.

1. $((\mathbb{Z}/4)^\times, \cdot, 1)$
2. $((\mathbb{Z}/5)^\times, \cdot, 1)$
3. $((\mathbb{Z}/6)^\times, \cdot, 1)$
4. $((\mathbb{Z}/7)^\times, \cdot, 1)$
5. $((\mathbb{Z}/8)^\times, \cdot, 1)$

Is there a pattern in the above?

Definition 202.2.9. g is said to be a **primitive root** mod N if g generates $(\mathbb{Z}/N)^\times$, i.e., if $\langle g \rangle = (\mathbb{Z}/N)^\times$.

primitive root

Later we'll see that if g a primitive root mod n, then it can give rise to a group-based public key cryptography, also called a discrete log cryptography. Understanding this will help understand ECC (elliptic curve cryptography). Although there are now other types of public key ciphers, RSA and ECC are generally considered the two main public key ciphers.

Primitive roots mod n are very important not just for cryptography.

Exercise 202.2.15. Write a program and check when $(\mathbb{Z}/N)^\times$ is cyclic and for each $(\mathbb{Z}/N)^\times$ find the generators. How many generators are there in $(\mathbb{Z}/N)^\times$? How often is 2 a generator of $(\mathbb{Z}/N)^\times$? How about just checking if 2 is a generator of $(\mathbb{Z}/p)^\times$ where p runs through all the primes? To be precise, if p_n denotes the n–th prime, what is $\lim_{n\to\infty}(N_n/n)$ where N_n is the number of primes p in $\{p_1, ..., p_n\}$ such that 2 is a primitive root mod p.

202.2.5 Lagrange's theorem

Theorem 202.2.1. Lagrange's theorem *Let H be a subgroup of G where G is a finite group. Then $|H|$ divides $|G|$.*

Lagrange's theorem

(The above can be generalized to infinite groups.)

Proof. Exercise. (Hint: If $g \in G$, define the sets $gH = \{g * h \mid h \in H\} \subseteq G$. gH is called a **left coset** of G with respect to H. Show G can be partitioned into sets of the form gH. Next show that $|gH| = |g'H|$ for $g, g' \in G$.) □

left coset

You should try to complete the proof using the above strategy outlined above. The answer (the proof) is on the next page. If you discovered another proof let me know.

Proof of Lagrange's theorem.

First I'll show that if $g, g' \in G$, then either gH and $g'H$ are disjoint or $gH = g'H$. Let's called this Lemma A.

If $gH \cap g'H = \emptyset$, then we are done. Now assume that $gH \cap g'H \neq \emptyset$. Then there is some $x \in gH \cap g'H$. Note that $x = gh = g'h'$ for some $h, h' \in H$. Therefore $g = g'h'h^{-1}$.

I will now prove that $gH \subseteq g'H$. Let $y \in gH$. Then $y = gh''$ for some $h'' \in H$. Then $y = (g'h'h^{-1})h'' = g'(h'h^{-1}h'') \in g'H$. Hence $gH \subseteq g'H$. By symmetry, $g'H \subseteq gH$. Hence $gH = g'H$.

Lemma A is now proven.

Next I'll show that there exist $g_1, ..., g_n \in G$ such that $g_1H, ..., g_nH$ is a partition of G, i.e.,

$$G = g_1H \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} g_nH$$

i.e., that G is a disjoint union of left cosets of H.

Observe the following: If $g_{n+1} \in G - (g_1H \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} g_nH)$, then $g_{n+1}H$ is disjoint from $g_1H \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} g_nH$. For otherwise if $g_{n+1}H$ intersects g_iH Then from Lemma A above, $g_{n+1}H = g_iH$ and hence $g_{n+1} = g_{n+1}e \in g_iH \subseteq g_1H \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} g_nH)$, which is a contradiction. This is simply saying that if a disjoint union of left coset of H is not G, then I can add another left coset to that union and obtain another larger disjoint union. I will called this Lemma B.

Now I will construct the disjoint union of cosets for G.

First I choose any $g_1 \in G$. I set $X_1 = g_1H$.

If $G = X_1$, I'm done: $G = g_1H$. Otherwise I choose $g_2 \in G - X_1$. From Lemma B, g_2H and X_1 are disjoint. Let $X_2 = g_1H \mathbin{\dot\cup} g_2H$ which is a disjoint union.

If $G = X_2$, I'm done. Otherwise I choose $g_3 \in G - X_2$. Again g_3H is disjoint from X_2. Let $X_3 = X_2 \mathbin{\dot\cup} g_3H$ which is a disjoint union.

The general fact is this: Let $g_1 \in G$ and define $X_1 = g_1H$. Recursively if $X_n \neq G$, let $g_n \in G - X_{n-1}$ and define $X_n = X_{n-1} \mathbin{\dot\cup} g_nH$. Then X_n is a disjoint union. The proof of this (by induction) follows the same argument above.

Note that $|X_1| < |X_2| < \ldots$. Since G is finite, we cannot possibly have $X_n \neq G$ for infinitely many n, otherwise there exists infinitely many distinct elements $g_1, g_2, \ldots$ in G. Therefore at some point $X_n = G$, i.e.,

$$G = g_1H \,\dot{\cup}\, \cdots \,\dot{\cup}\, g_nH$$

Alternatively, and more concisely, you can prove the above in the following way. Consider sets of the form $\dot{\bigcup}_{g \in I} gH$ where I is a subset of G. Let $\dot{\bigcup}_{g \in I} gH$ be one such set with maximal size ($|G|$ is finite). Assume the size is not $|G|$. Then there is some $g' \in G - \dot{\bigcup}_{g \in I} gH$. By Lemma B, $\dot{\bigcup}_{g \in I \cup g'} gH$ is a disjoint union of left cosets of H with size larger that $\dot{\bigcup}_{g \in I} gH$. This is a contradiction since we assumed $\dot{\bigcup}_{g \in I} gH$ has the largest possible size for a disjoint union of left cosets of H. Hence $\dot{\bigcup}_{g \in I} gH = G$.

(Note that the two different proofs above that G is a disjoint union have totally different flavors. The first proof is constructive and builds a maximal structure. The second is non-constructive and is by a maximal-contradiction argument. A pair of proofs like the above is very common. Pay attention to it.)

Now I claim that $|gH| = |H|$. Let's call this Lemma C. Define $f : H \to gH$ by $f(h) = gh$. f is a bijection since the map $f^{-1} : gH \to H$ defined by $f^{-1}(gh) = g^{-1}(gh) = h$ is the inverse of f:

$$\begin{aligned} f^{-1}(f(h)) &= f^{-1}(gh) = g^{-1}(gh) = h \\ f(f^{-1}(gh)) &= f(g^{-1}gh) = f(h) = gh \end{aligned}$$

Hence $|H| = |gH|$. Lemma C is proven.

Since G is a disjoint union of left cosets of H, i.e.,

$$G = g_1H \,\dot{\cup}\, \cdots \,\dot{\cup}\, g_nH$$

I have

$$|G| = \sum_{i=1}^{n} |g_iH|$$

By Lemma C, I obtain

$$|G| = \sum_{i=1}^{n} |g_iH| = \sum_{i=1}^{n} |H| = n|H|$$

Hence $|H|$ divides G. □

By the way the number $|G|/|H|$ is frequently denoted by $[G : H]$ and is called the **index** of H in G. This is basically the number of disjoint cosets of H to completely cover G. index

Let $(G, *, e)$ be a group and $g \in G$. Recall that $\langle g \rangle$ is the cyclic subgroup of G generated by g. You can think of $\langle g \rangle$ as the smallest subgroup of G containing g or concretely $\langle g \rangle$ is made up of e (i.e. g^0) and g^n, $(g^{-1})^n$ for all integers $n > 0$. Note that if G is finite, if you write down $g, g^2, g^3, ...$, at some point, $g^n = e$. In other words

$$\langle g \rangle = \{g^0 = e, g^1, g^2, \ldots, g^{n-1}\}$$

Hence $|\langle g \rangle| = n$. This n is the called the **order** of g. This is frequently denoted by $o(g)$ or $|g|$. Note that the order of e is always 1.

order

It's an easy exercise to show that if $o(g) = n$, the $\langle g \rangle \simeq \mathbb{Z}/n$.

As an example, consider $((\mathbb{Z}/5)^\times, \cdot, 1)$. Let's look at the order of 2 in $(\mathbb{Z}/5)^\times$.

$$\begin{aligned}
2^1 &= 2 \\
2^2 &= 2 \cdot 2 = 4 \\
2^3 &= 2 \cdot 2 \cdot 2 = 8 \equiv 3 \pmod{5} \\
2^4 &= 2 \cdot 2 \cdot 2 \cdot 2 = 16 \equiv 1 \pmod{5}
\end{aligned}$$

Therefore 2 has order 4 in $(\mathbb{Z}/5)^\times$. Here's another example. Consider $(\mathbb{Z}/6, +, 0)$. In this case the identity element e is 0 and the group operation $*$ is $+$. Instead of writing g^n, I'll write ng. Let's figure out the order of 2. I have

$$\begin{aligned}
1 \cdot 2 &= 2 \\
2 \cdot 2 &= 2 + 2 = 4 \\
3 \cdot 2 &= 2 + 2 + 2 = 6 \equiv 0 \pmod{6}
\end{aligned}$$

So 2 has order 3 in $\mathbb{Z}/6$. The order of g is the number of g's you need to self-operator to get to e.

Exercise 202.2.16.

1. What is the order of 1 in $(\mathbb{Z}/3)^\times$?
2. What is the order of 2 in $(\mathbb{Z}/3)^\times$?
3. What are the orders of $1, 2, 3, 4$ in $(\mathbb{Z}/5)^\times$? Who has the largest order? Do you notice something about the orders?
4. What is the order of $1, 3, 5$ in $(\mathbb{Z}/6)^\times$? Who has the largest order? Do you notice something about the orders?
5. What is the order of $1, 2, 3, 4, 5, 6$ in $(\mathbb{Z}/7)^\times$? Who has the largest order? Do you notice something about the orders?
6. Write a program to compute the order of n in $(\mathbb{Z}/N)^\times$ if $\gcd(n, N) = 1$. Who has the largest order? Do you notice something about the orders?

Corollary 202.2.1. *Let $(G, *, e)$ be a finite group and let $g \in G$. Then $o(g)$ divides $|G|$.*

Proof. Exercise. □

Corollary 202.2.2. *Every group of prime order must be cyclic. In particular if G has order prime p, then $G \simeq \mathbb{Z}/p$. Therefore, up to isomorphism, there is only one group of prime order.*

Proof. Exercise. □

Corollary 202.2.3. Fermat-Euler theorem *Let N be a positive integer. If* $\gcd(a, N) = 1$*, then*

$$a^{\phi(N)} \equiv 1 \pmod{N}$$

Fermat-Euler theorem

Proof. Exercise. □

Corollary 202.2.4. Fermat's little theorem *Let p be a prime. If* $\gcd(a, p) = 1$*, then*

$$a^{p-1} \equiv 1 \pmod{p}$$

Fermat's little theorem

Proof. Exercise. □

Lagrange's theorem says that a subgroup of a group G must have order dividing $|G|$. The converse of Lagrange's theorem is this question: Let $(G, *, e)$ be a finite group. Suppose n divides $|G|$. Is there a subgroup of G with order n? This is not true in general.

File: semigroup-with-identity.tex

202.3 Introduction to semigroups with identity

Recall that a group is a triple $(G, *, e)$ where G is a set, $e \in G$, and such that

1. Closure: If $x, y \in G$, then $x * y \in G$. This means that $* : G \times G \to G$ is a binary operator.
2. Associativity: If $x, y, z \in G$, then $(x * y) * z = x * (y * z)$.
3. Inverse: If $x \in G$, then there is some $y \in G$ such that $x * y = e = y * x$.
4. Neutral: If $x \in G$, then $e * x = x = x * e$.

A semigroup with identity is almost a group except that inverses need not exist:

Definition 202.3.1. A **semigroup with identity** is a triple $(G, *, e)$ where G is a set, $e \in G$ and the following are satisfied:

semigroup with identity

1. Closure: If $x, y \in G$, then $x * y \in G$. This means that $* : G \times G \to G$ is a binary operator.
2. Associativity: If $x, y, z \in G$, then $(x * y) * z = x * (y * z)$.
3. Neutral: If $x \in G$, then $e * x = x = x * e$.

As in groups, you can think of $*$ is a binary operator on G. And of course a **commutative semigroup with identity** $(G, *, e)$ is a semigroup with identity such that $*$ is commutative, i.e., if $x, y \in G$, then

commutative semigroup with identity

$$x * y = y * x$$

Proposition 202.3.1. (Uniqueness of identity) *Prove that the identity (or neural) element e of G is unique, i.e., if $e, e' \in G$, satisfy the following:*

(a) *If $x \in G$, then $x * e = x = e * x$.*
(b) *If $x \in G$, then $x * e' = x = e' * x$.*

Then $e = e'$.

Proof. If you look at the proof of this fact for groups, you would notice that in fact the proof actually did not use the inverse axiom. □

File: introduction-to-rings.tex

202.4 Introduction to rings and fields

202.4.1 Definition of rings and fields

Recall that we want to view $\mathbb{Z}$ as $(\mathbb{Z}, +, \cdot, 0, 1)$, i.e., it a set $\mathbb{Z}$ of values together with two operations $+$ and $\cdot$ and with two special values 0 and 1. With the language of groups and semigroups with identity, you can say that $(\mathbb{Z}, +, \cdot, 0, 1)$ satisfies

1. $(\mathbb{Z}, +, 0)$ is an abelian group.
2. $(\mathbb{Z}, \cdot, 1)$ is a commutative semigroup with identity.
3. If $x, y, z \in \mathbb{Z}$, then $x(y + z) = xy + xz$.
4. If $x, y \in \mathbb{Z}$, then $xy = 0$ implies $x = 0$ or $y = 0$.

All the above can be generalized.

ring

Definition 202.4.1. $(R, +_R, \cdot_R, 0_R, 1_R)$ is a **ring** if

1. $(R, +_R, 0_R)$ is an abelian group
2. $(R, \cdot_R, 1_R)$ is a semigroup with identity
3. Distribution: If $x, y, z \in R$, then

$$\begin{aligned} x \cdot_R (y +_R z) &= x \cdot_R y +_R x \cdot_R z \\ (y +_R z) \cdot_R x &= y \cdot_R x +_R z \cdot_R x \end{aligned}$$

commutative ring

$(R, +_R, \cdot_R, 0_R, 1_R)$ is a **commutative ring** if it is a ring and if $x, y \in R$, then

$$x \cdot_R y = y \cdot_R x$$

integral domain

In other words the semigroup with identity $(R, \cdot_R, 1_R)$ is a commutative semigroup. Furthermore $(R, +_R, \cdot_R, 0_R, 1_R)$ is an **integral domain** if it is a commutative ring and for $x, y \in R$,

$$x \cdot_R y = 0_R \implies x = 0_R \text{ or } y = 0_R$$

I've written the ring operations as $+_R$ and $\cdot_R$. And I've written 0_R for the identity for $+_R$ and the 1_R for the identity for $\cdot_R$. When the context is clear I will omit the R subscripts and write $+, \cdot, 0, 1$ with the understanding that these are abstract operations and abtract identities for $+$ and $\cdot$ for the abstract ring R. Also, when there is no confusion, I will write xy instead of $x \cdot y$.

Anyway, a ring R is a set of "things" with two operations abstractly denoted by $+_R$ and $\cdot_R$ ("addition" and "multiplication"). Furthermore there are two special "things" in R which we will call 0_R and 1_R. 0_R is the **additive identity** and 1_R is the **multiplicative identity**. Expanding the above, the properties of $+_R$ (written as $+$) with its identity 0_R (written as 0) are:

additive identity
multiplicative identity

1. If $x, y \in R$, then $x + y$ is in R.
2. If $x, y, z \in R$, then $(x + y) + z = x + (y + z)$
3. If $x \in R$, then $x + 0 = x = 0 + x$
4. If $x \in R$, then there is some $y \in R$ such that $x + y = 0 = y + x$. y is called the **additive inverse** of x.

additive inverse

(From group theory, we know that the additive inverse of x is unique to x. That's why I can call it *the* additive inverse of x.)

For properties of $\cdot_R$ (written as $\cdot$) with its identity 1_R (written as 1) are:

1. If $x, y \in R$, then $x \cdot y \in R$.
2. If $x, y, z \in R$, then $(x \cdot y) \cdot z = x \cdot (y \cdot z)$
3. If $r \in R$, then $r \cdot 1 = r = 1 \cdot r$.

The property involving both $+$ and $\cdot$ is

1. If $x, y, z \in R$, then $x \cdot (y + z) = x \cdot y + x \cdot z$

In addition a ring is a commutative ring if

1. If $r, s \in R$, then $r \cdot s = s \cdot r$.

Just remember this: A ring is a set of things with addition and multiplication. And when you're lost just think of the set of integers and its operations.

The following are commutative rings

1. $(\mathbb{Z}, +, \cdot, 0, 1)$. In fact it's an integral domain.
2. $(\mathbb{Z}/N, +, \cdot, 0, 1)$ for any $N > 0$
3. $(\mathbb{Q}, +, \cdot, 0, 1)$.
4. $(\mathbb{R}, +, \cdot, 0, 1)$.
5. $(\mathbb{C}, +, \cdot, 0, 1)$.
6. $(\mathbb{R}[X], +, \cdot, 0, 1)$.

The following is a ring but it's not a commutative ring

1. $\left(\mathrm{M}_2(\mathbb{R}), +, \cdot, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$

Recall that there are 2-by-2 matrices M, N such that $MN \neq NM$. Therefore $(\mathrm{M}_2(\mathbb{R})$ is *not* a commutative ring. In general, for any positive integer n, $\mathrm{M}_n(\mathbb{R})$ is a ring.

More generally, if R is a ring, then

1. $R[X]$ is a commutative ring where $R[X]$ is the set of polynomials with coefficients in R.

2. $\mathrm{M}_n(R)$ is a non-commutative ring.

For details on polynomial rings, see section "Polynomial rings". [TODO: create label]

Note that if $f(X) \in R[X]$, besides being an element of a ring, you can also talk about

1. Evaluation: If $r \in R$, then $f(r)$ is the evaluation fo $f(X)$ when X is substituted by r.

2. Root: A ring element $r \in R$ is a root of $f(X)$ if $f(r) = 0_R$.

Exercise 202.4.1. Is $\mathbb{Z}/N$ an integral domain for all $N > 1$?

Note that subtraction is technically a shorthand in the following sense:

$$x - y = x + (-y)$$

In other words subtraction is defined in terms of the ring addition together with additive inverse. In terms of programming, the above hints at the fact that when you write a `LongInt` class (i.e. integer with arbitrary number of digits) in C++, you should probably write the "negative of", i.e.,

```
    LongInt operator-() const;
```

and "addition", i.e.,

```
    LongInt operator+(const LongInt &) const;
```

and then define subtraction in terms of the above:

```
    LongInt operator-(const LongInt & x) const
    {
        return ((*this) + (-x));
    }
```

(Of course it's probably best to implement `operator-=` before "subtraction".)

Note that the properties of $(R, +, 0)$ is very similar to the properties of $(R, \cdot, 1)$. One main difference is that in $(R, +, 0)$, every element $x \in R$ as an additive inverse. However in $(R, \cdot, 1)$, not every element $x \in R$ has a multiplicative inverse. For instance in $\mathbb{Z}$, 2 has an additive inverse (i.e., -2), but 2 does not have a multiplicative inverse in $\mathbb{Z}$. However note that in the ring $(\mathbb{Q}, +, \cdot, 0, 1)$, *every* element other than 0 has a multiplicative inverse. This is an important and common idea so we'll give it a name:

Definition 202.4.2. $(R, +, \cdot, 0, 1)$ is a **field** if $(R, +, \cdot, 0, 1)$ is a commutative ring where every $x \in R$ which is not 0 has a multiplicative inverse. field

The following are fields which you are familiar with:

1. $(\mathbb{Q}, +, \cdot, 0, 1)$.
2. $(\mathbb{R}, +, \cdot, 0, 1)$.
3. $(\mathbb{C}, +, \cdot, 0, 1)$.

However these fields are not directly useful in cryptography. The follow fields are extremely important in cryptography and electrical/computer engineering for instance in the area of coding theory:

1. $(\mathbb{Z}/p, +, \cdot, 0, 1)$ where p is a prime.

And there are more. In particular $\mathbb{Z}/2$ is a field called the **binary field**. binary field

Exercise 202.4.2. Prove that if R is a field, then R is an integral domain.

From now on instead of saying "Let $(R, +, \cdot, 0, 1)$ be a ring", I will say "Let R be a ring" with the understanding that the operations and identities are denoted by $+, \cdot, 0, 1$. The context will make everything clear. But *do* pause and read slowing so that you get the context.

Proposition 202.4.1. *Let R be a ring and $x \in R$. Then $0 \cdot x = 0$.*

First Proof.

$$\begin{aligned}
1 \cdot x &= x && \text{by the neutral axiom of } \cdot \\
\therefore\ (1+0) \cdot x &= x && \text{by the neutral axiom of } + \\
\therefore\ 1 \cdot x + 0 \cdot x &= x && \text{by the distributivity axiom} \\
\therefore\ x + 0 \cdot x &= x && \text{by the identity axiom of } \cdot \\
\therefore\ (-x) + (x + 0 \cdot x) &= (-x) + x && \\
\therefore\ ((-x) + x) + 0 \cdot x &= (-x) + x && \text{by the associativity axiom of } + \\
\therefore\ 0 + 0 \cdot x &= (-x) + x && \text{by the inverse axiom of } + \\
\therefore\ 0 \cdot x &= (-x) + x && \text{by the neutral axiom of } + \\
&= 0 && \text{by the inverse axiom of } +
\end{aligned}$$

□

In the above you notice one step (line 5 above) without justification. If

$$a = b$$

you can *always* say

$$a + x = b + x, \qquad x + a = x + b, \qquad a \cdot x = b \cdot x, \qquad x \cdot a = x \cdot b$$

simply because $+$ is a well-defined function: if $f(x, y)$ is a function, $a = a'$, and $b = b'$, then of course $f(a, b) = f(a', b')$. This has nothing to do with rings.

The above proof requires nine steps. Can we do better? (Yes ... like programming we do talk about shorter and slicker proofs you know.)

Here's another proof:

Second Proof.

$$
\begin{aligned}
0 \cdot x &= (0+0) \cdot x && \text{by the neutrality axiom of } + \\
\therefore\ \ 0 \cdot x &= 0 \cdot x + 0 \cdot x && \text{by the distributivity axiom} \\
\therefore\ \ 0 \cdot x + -(0 \cdot x) &= (0 \cdot x + 0 \cdot x) + -(0 \cdot x) && \\
\therefore\ \ 0 &= (0 \cdot x + 0 \cdot x) + -(0 \cdot x) && \text{by the inverse axiom of } + \\
&= 0 \cdot x + (0 \cdot x + -(0 \cdot x)) && \text{by the associativity axiom of } + \\
&= 0 \cdot x + 0 && \text{by the inverse axiom of } + \\
&= 0 \cdot x && \text{by the neutrality axiom of } +
\end{aligned}
$$

□

By the way, in the above I created $-(0 \cdot x)$ at the third step. Note that in this ring R the "negative of" (i.e. additive inverse) can be applied only to values in R. We can do this to $0 \cdot x$ because of the closure axiom of $\cdot$: Since $0 \in R$ and $x \in R$, we know that $0 \cdot x \in R$. Therefore $-(0 \cdot x)$ makes sense and is an element of R. So we stay within the system (i.e. of R).

The above proof contains 7 steps. The following proof (6 steps) is very similar to the second proof.

Third Proof.

$$
\begin{aligned}
0 &= (0 \cdot x) + -(0 \cdot x) && \text{by the inverse axiom of } + \\
&= ((0+0) \cdot x) + -(0 \cdot x) && \text{by the neutrality axiom of } + \\
&= (0 \cdot x + 0 \cdot x) + -(0 \cdot x) && \text{by the distributivity axiom} \\
&= 0 \cdot x + (0 \cdot x + -(0 \cdot x)) && \text{by the associativity axiom of } + \\
&= 0 \cdot x + 0 && \text{by the inverse axiom of } + \\
&= 0 \cdot x && \text{by the neutrality axiom of } +
\end{aligned}
$$

□

Let me know if you can find a shorter proof.

Exercise 202.4.3. Analyze this “proof” of the above fact:

Evil “Proof”.

$$\begin{aligned} 0 \cdot x &= (0 + -0) \cdot x \\ &= 0 \cdot x + (-0) \cdot x \\ &= 0 \cdot x + -(0 \cdot x) \\ &= 0 \end{aligned}$$

At which point do we make an unjustified claim. (I'm *not* saying that the claim is *incorrect.*) □

We can add a little bit more to the above theorem:

Proposition 202.4.2. *Let R be a ring and $x \in R$. Then $0 \cdot x = 0 = x \cdot 0$.*

Evil proof. We have already shown that $0 \cdot x = 0$.

$$\begin{aligned} & 0 \cdot x = 0 && \text{by Proposition 202.4.1} \\ \therefore\ & x \cdot 0 = 0 \end{aligned}$$

□

Exercise 202.4.4. Why is the above proof evil? □

Exercise 202.4.5. Give a correct proof of the above proposition. □

Here's an important trick for proving something is a negative. Note that $-x$ has the property:

$$x + (-x) = 0$$

The following shows that if y satisfies the property

$$x + y = 0$$

then it must be true that

$$y = -x$$

Theorem 202.4.1. (Uniqueness of additive inverse) *Let R be a ring and let $x, y \in R$.*

(a) *If $x + y = 0$ then $y = -x$*
(b) *If $y + x = 0$ then $y = -x$*

Proof. (a) From $x + y = 0$ we have:

$$\begin{aligned}
x + y &= 0 \\
\therefore\ (-x) + (x + y) &= (-x) + 0 \\
\therefore\ ((-x) + x) + y &= (-x) + 0 && \text{by associativity axiom of addition} \\
\therefore\ 0 + y &= (-x) + 0 && \text{by inverse axiom of addition} \\
\therefore\ y &= (-x) + 0 && \text{by neutrality axiom of addition} \\
\therefore\ y &= -x && \text{by neutrality axiom of addition}
\end{aligned}$$

□

Exercise 202.4.6.

1. Can you give me a shorter proof of (a)?
2. Also prove part (b) above.
3. When you are done proving (a) and (b), tell me why the proofs are actually redundant. □

Proposition 202.4.3. *Let R be a ring and $x \in R$. Then $-(-x) = x$*

Proof. Exercise. □

Proposition 202.4.4. *Let R be a ring and $x \in R$.* $-1 \cdot x = -x = x \cdot (-1)$

Proof. Exercise. □

Proposition 202.4.5. *Let R be a ring and $x, y \in R$. Then $(-y) \cdot x = -(y \cdot x) = y \cdot (-x)$*

Look at what we want to prove:

$$(-y) \cdot x = -(y \cdot x)$$

This says that the additive inverse of $y \cdot x$ is $(-y) \cdot x$. The uniqueness of additive inverse is a tool about proving something is an additive inverse. Specifically, Theorem 3 says that if we can show

$$y \cdot x + (-y) \cdot x = 0$$

then $(-y) \cdot x$ *is* the negative of $y \cdot x$. (You can call it "backward reasoning" if you like.) So let's prove it:

Proof. Exercise. □

Proposition 202.4.6. *Let R be a ring and $x, y \in R$.*

(a) $(-x) \cdot (-y) = xy$.
(b) $(-1) \cdot (-1) = 1$

Proof.

Proposition 202.4.7. *Let R be a ring. Then $-0 = 0$.*

Proof.

The following says that you can perform left or right "cancellation" of similar terms.

Proposition 202.4.8. *Let R be a ring and $a, x, y \in R$.*

(a) *If $a + x = a + y$ then $x = y$*
(b) *If $x + a = y + a$ then $x = y$*

There's also a multiplicative version of cancellation where you can cancel similar factors.

Theorem 202.4.2. *Let R be an integral domain and $a, x, y \in R$.*

(a) *If $ax = ay$ and $a \neq 0$, then $x = y$.*
(b) *If $xa = ya$ and $a \neq 0$, then $x = y$.*

Proof. Exercise.

Exercise 202.4.7. Using $\mathbb{Z}$ as a model, think of some algebraic facts (involving $+$ and $\cdot$) and see if you can prove them. □

202.4.2 The unit group

If $(R, +, \cdot, 0, 1)$ is a ring, $(R, \cdot, 1)$ is a semigroup with identity – it need not be a group. However, the set of ring elements of R with the multiplicatively invertible elements of R, denoted by $R^\times$ or $U(R)$, is a group. An element of $R^\times$ is called a **unit** of R. unit

Note that 1 and -1 are units: $\{1, -1\} \subseteq R^\times$.

Proposition 202.4.9. *If* $(R, +, \cdot, 0, 1)$ *is a ring, then* $(R^\times, \cdot, 1)$ *is a group.*

Proof. Exercise. □

See next page for the answer (the proof). Therefore the invertible elements of $\mathbb{Z}/N$, $(\mathbb{Z}/N)^\times$, forms a group. And of course $(\mathbb{Z}/N)^\times$ is extremely important in cryptography.

Proof. Note that $1 \in R$ since by the identity axiom of R, $1 \cdot 1 = 1 = 1 \cdot 1$, i.e., 1 is multiplicatively invertible. Now to show that $R^\times$ satisfies the group axioms.

The first thing to show is closure: let $r, s \in R^\times$. Note that $rs \in R$. I need to show $rs \in R^\times$, i.e., I need to show rs is invertible. Since $r, s \in R^\times$, we see that r^{-1}, s^{-1} exist in R. Note that $s^{-1}r^{-1} \in R$ by closure of $\cdot$ of R. Now we note that

$$(rs)(s^{-1}r^{-1}) = ((rs)s^{-1})r^{-1}) = (r(ss^{-1}))r^{-1}) = (r1)r^{-1} = rr^{-1} = 1$$

and

$$(s^{-1}r^{-1})(rs) = s^{-1}(r^{-1}(rs)) = s^{-1}((r^{-1}r)s) = s^{-1}(1s) = s^{-1}s = 1$$

Hence rs is invertible in R and hence is in $R^\times$. The associativity of $\cdot$ on $R^\times$ is immediate since $\cdot$ is associative on R. Every element r in $R^\times$ by definition has an inverse r^{-1}. This r^{-1} is also invertible in R (since its inverse is r) and hence $r^{-1} \in R^\times$. Therefore for each $r \in R^\times$, there is some $r^{-1} \in R^\times$ such that $rr^{-1} = 1 = r^{-1}r$. The 1 in R is the multiplicative identity element on R and therefore is the group identity element on $R^\times$ under the $\cdot$ operation. □

Exercise 202.4.8. Prove the following:

1. If $u \in R^\times$, then $u \mid r$ for all $r \in R$.
2. Let $r, s \in R$ where R is an integral domain. If $r \mid s$ and $s \mid r$, then $r = us$ for some $u \in R^\times$.

202.4.3 Polynomial rings

The following is a way to construction more rings, through polynomials.

Proposition 202.4.10. *Let R be a commutative ring. Let $R[X]$ be the set of polynomials with coefficients in R and "variable" X. Let $r = \sum_{i=0}^{n} r_i X^i$ and $s = \sum_{i=0}^{n} s_i X^i$. Define equality as $r = s$ if $r_i = s_i$ for $i = 0, ..., n$. Also, the addition and multiplication for $R[X]$ are defined in the obvious way as:*

$$r + s = \sum_{i=0}^{n} (r_i + s_i) X^i$$

and

$$r \cdot s = \sum_{i=0}^{2n} t_i X^i, \text{ where } t_i = \sum_{k=0}^{i} r_k s_{i-k}$$

Note that $R \subseteq R[X]$. $0_{R[X]} = 0_R$ and $1_{R[X]} = 1_R$. Then $(R[X], +, \cdot, 0, 1)$ becomes a commutative ring where the additive inverse of r is given by

$$-r = \sum_{i=0}^{n} (-r_i) X^i$$

In particular note that since $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/N$ are commutative rings, the following are also commutative rings:

- $\mathbb{Z}[X]$
- $\mathbb{Q}[X]$
- $\mathbb{R}[X]$
- $\mathbb{C}[X]$
- $(\mathbb{Z}/N)[X]$

Wait a minute ... since $\mathbb{Z}[X]$ is a commutative ring ... that means I can also construct

$$(\mathbb{Z}[X])[Y]$$

where Y is a "variable". See that? I'll let you think about this: $(\mathbb{Z}[X])[Y]$ is just $\mathbb{Z}[X, Y]$ the ring of polynomials of two variables with coefficients in $\mathbb{Z}$. The same idea can be applied to all rings R and therefore you also have $R[X, Y, Z]$ for any commutative ring R. In the above, the technical term of "is just" is "is isormorphic as rings to":

Definition 202.4.3. Two rings R and S are **isomorphic** (as rings) if there is a bijection $f : R \to S$ such that

isomorphic

$$
\begin{aligned}
f(x +_R y) &= f(x) +_S f(y) \\
f(x \cdot_R y) &= f(x) \cdot_S f(y)
\end{aligned}
$$

Stop and think about the above definition carefully: there are two rings involved and there are two sets of ring operators.

Exercise 202.4.9. Prove that if $f : R \to S$ is a ring isomorphism, then

$$f(0_R) = 0_S$$
$$f(1_R) = 1_S$$
$$f(-x) = -f(x)$$

and if x has a multiplicative inverse in R (i.e., x^{-1} exists in R), then $f(x)$ also has a multiplicative inverse (in S) and

$$f(x^{-1}) = f(x)^{-1}$$

Proof. Exercise. □

Make sure you look at the above carefully!!! For instance look at the two "$-$" symbols. They are different. The "$-$" on the left refers to additive inverse in R while the other "$-$" refers to the additive inverse on the right. Similar there are two "${}^{-1}$" symbols as well.

Informally, the above facts basically tells you that if two rings are isomorphic, then their ring theoretic properties are the same.

Exercise 202.4.10. Prove that if R is an integral domain, then $R[X]$ is also an integral domain.

Proof. Exercise. □

202.4.4 Divisibility, congruences, quotient ring

Here's another way to construct rings: we are going to take the cue from $\mathbb{Z}/N$. For this section, I'll assume R is a commutative ring.

First let me define divisibility:

Definition 202.4.4. Let R be a commutative ring and let $r, s \in R$. r is **divisible** by s, and we write $s \mid r$, if there is some $c \in R$ such that divisible

$$r = c \cdot s$$

There's another way to say the above. Let

$$\langle s \rangle = R \cdot s = \{c \cdot n \mid c \in R\}$$

Therefore $\langle s \rangle$ is the set of all R–multiples of s. (Make sure compare this with the notation $\langle g \rangle$ used in group theory – the two are different.) Then I can redefine divisibility as follows: r is **divisible** by s if divisible

$$r \in \langle s \rangle$$

Of course the two definitions are the same.

This generalizes the idea of divisibility in $\mathbb{Z}$ to any commutative ring.

Notice that in the above I use the shorthand

$$R \cdot s = \{c \cdot n \mid c \in R\}$$

In general if X is a subset of R and $s \in R$, then I'll use the following shorthand

$$\begin{aligned} X \cdot s &= \{x \cdot s \mid x \in X\} \\ X + s &= \{x + s \mid x \in X\} \end{aligned}$$

I will also write Xs for $X \cdot s$. Likewise

$$\begin{aligned} s \cdot X &= \{s \cdot x \mid x \in X\} \\ s + X &= \{s + x \mid x \in X\} \end{aligned}$$

I'll also write sX for $s \cdot X$.

Exercise 202.4.11.

1. In $\mathbb{Z}$, is 10 divisible by 5? (You have 1 second.)
2. In $\mathbb{Z}/8$, is 10 divisible by 5? (10 seconds.)
3. In $\mathbb{Z}[X]$, is $X^2 - 1$ divisible by $X + 1$? (2 seconds.)
4. In $\mathbb{Z}[X]$, is $X^2 - 1$ divisible by $X + 5$? (15 seconds.)
5. In $(\mathbb{Z}/4)[X]$, is $X^2 - 1$ divisible by $X + 5$? (2 seconds.)
6. In $\mathbb{R}$, is π divisible by e? (You have 1 second.) And ... in C++ why is there no `%` (mod operator) for `double` type? And in general, why is the concept of divisibility not useful for fields.

Proposition 202.4.11. *Let R be a commutative ring and $r, s, t \in R$.*

(a) $1 \mid r$, $-1 \mid r$.
(b) $r \mid 0$.
(c) *Linearity: If $r \mid s$ and $r \mid t$, then $r \mid x \cdot s + y \cdot t$ for all $x, y \in R$.*
(d) *Reflexivity: $r \mid r$*
(e) *Transitivity: If $r \mid s$ and $s \mid t$, then $r \mid t$.*

Proof. Exercise. □

Now for congruence relation (the "mod" relation):

Definition 202.4.5. Let $n \in R$ where R is a commutative ring. Define the relation $\equiv \pmod{n}$ on R as follows: If $r, s \in R$, then

$$r \equiv s \pmod{n}$$

if $n \mid r - s$, i.e., if $r - s \in \langle n \rangle$. (Remember that $r - s$ is just a shorthand for $r + (-s)$.) If

$$r \equiv s \pmod{n}$$

holds you would say "r is **congruent** to s mod n". congruent

Proposition 202.4.12. *The relation* $\equiv \pmod{n}$ *is an equivalence relation on* R*, i.e., if* $r, s, t \in R$*, then*

(a) *Reflexive:* $r \equiv r \pmod{n}$
(b) *Symmetric: If* $r \equiv s \pmod{n}$*, then* $s \equiv r \pmod{n}$
(c) *Transitive: If* $r \equiv s \pmod{n}$ *and* $s \equiv t \pmod{n}$*, then* $r \equiv t \pmod{n}$

Proof. Exercise. □

Therefore R can be partitioned into a collection of equivalence classes, i.e.,

$$R = X_1 \mathbin{\dot\cup} X_2 \mathbin{\dot\cup} X_3 \mathbin{\dot\cup} \cdots$$

The "$\dot\cup$" means disjoint union, i.e., R is the union of the X_i and pairs of the X_i's intersect to give the empty set. Each of the X_i is a set of ring elements of R related to each other. Usually you pick some ring element, say $r_1 \in R$, to be a representative of everyone in X_i and write X_1 as $[r_1]$. If you want to emphasize the relation $\equiv \pmod{n}$, you can also write $[r_1]_n$. When the context is clear, usually the subscript "$_n$" is omitted.

Note that the above is actually not quite right because in general the number of equivalence classes need not be countable.

In the case of for instance $\equiv \pmod{N}$ for $R = \mathbb{Z}$ (with $N > 0$), the number of equivalence classes is indeed countable and in fact is finite. (The proof is by Euclidean algorithm.)

As an example, consider $\equiv \pmod{6}$ relation on $\mathbb{Z}$. $\mathbb{Z}$ is a disjoint union of 6 equivalence classes:

$$\mathbb{Z} = [0]_6 \mathbin{\dot\cup} [1]_6 \mathbin{\dot\cup} [2]_6 \mathbin{\dot\cup} [3]_6 \mathbin{\dot\cup} [4]_6 \mathbin{\dot\cup} [5]_6$$

Here, $[0]_6$ means the set of $x \in \mathbb{Z}$ related to 0 through the $\equiv \pmod{6}$ relation, i.e.,

$$[0]_6 = \{x \in \mathbb{Z} \mid x \equiv 0 \pmod{6}\}$$

This means that

$$[0]_6 = \{..., -18, -12, -6, 0, 6, 12, 18, ...\}$$

In other words $[0]_6$ is the set of multiples of 6, which can also be written

$$[0]_6 = \{..., -18, -12, -6, 0, 6, 12, 18, ...\} = 6 \cdot \mathbb{Z} = \langle 6 \rangle$$

And it's easy to see that

$$\begin{aligned}
[1]_6 &= \{..., -17, -11, -5, 1, 7, 13, 19, ...\} = 1 + 6\mathbb{Z} = 1 + \langle 6 \rangle \\
[2]_6 &= \{..., -16, -10, -4, 2, 8, 14, 20, ...\} = 2 + 6\mathbb{Z} = 2 + \langle 6 \rangle \\
[3]_6 &= \{..., -15, -9, -3, 3, 9, 15, 21, ...\} = 3 + 6\mathbb{Z} = 3 + \langle 6 \rangle \\
[4]_6 &= \{..., -14, -8, -2, 4, 10, 16, 22, ...\} = 4 + 6\mathbb{Z} = 4 + \langle 6 \rangle \\
[5]_6 &= \{..., -13, -7, -1, 5, 11, 17, 23, ...\} = 5 + 6\mathbb{Z} = 5 + \langle 6 \rangle
\end{aligned}$$

etc. At this point, it should be very clear why

$$\mathbb{Z} = [0]_6 \mathbin{\dot\cup} [1]_6 \mathbin{\dot\cup} [2]_6 \mathbin{\dot\cup} [3]_6 \mathbin{\dot\cup} [4]_6 \mathbin{\dot\cup} [5]_6$$

By the way, note that

$$\begin{aligned}
[0]_6 &= 0 + \langle 6 \rangle \\
[1]_6 &= 1 + \langle 6 \rangle \\
[2]_6 &= 2 + \langle 6 \rangle \\
[3]_6 &= 3 + \langle 6 \rangle \\
[4]_6 &= 4 + \langle 6 \rangle \\
[5]_6 &= 5 + \langle 6 \rangle
\end{aligned}$$

And going back to the proof of Lagrange's theorem, if the group is $(\mathbb{Z}, +, 0)$ and the subgroup is $\langle 6 \rangle$, then the $g + \langle 6 \rangle$ are left cosets of $\langle 6 \rangle$ in $\mathbb{Z}$.

Wait a minute ... what about $[6]_6$? Well by definition

$$[6]_6 = \{x \in \mathbb{Z} \mid x \equiv 6 \pmod{6}\}$$

But since $6 \equiv 0 \pmod{6}$, if $x \equiv 6 \pmod{6}$, by transitivity, $x \equiv 0 \pmod{6}$. And vice versa, if $x \equiv 0 \pmod{6}$, then $x \equiv 6 \pmod{6}$. Hence

$$\begin{aligned}
[6]_6 &= \{x \in \mathbb{Z} \mid x \equiv 6 \pmod{6}\} \\
&= \{x \in \mathbb{Z} \mid x \equiv 0 \pmod{6}\} \\
&= [0]_6
\end{aligned}$$

At this point, it's not difficult to see that

$$[0]_6 = [6]_6 = [-6]_6 = [12]_6 = [-12]_6 = [18]_6 = [-18]_6 = \cdots$$

Likewise

$$[1]_6 = [7]_6 = [-5]_6 = [13]_6 = [-11]_6 = [19]_6 = [-17]_6 = \cdots$$

Etc. This means that the set of equivalence classes

$$\{[x]_6 \mid x \in \mathbb{Z}\} = \{..., [-3]_6, [-2]_6, [-1]_6, [0]_6, [1]_6, [2]_6, [3]_6, ...\}$$

is actually finite:

$$\{[x]_6 \mid x \in \mathbb{Z}\} = \{[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6\}$$

Why is it important to know that $\mathbb{Z}$ is a disjoint union of equivalence classes:

$$\mathbb{Z} = [0]_6 \,\dot{\cup}\, [1]_6 \,\dot{\cup}\, [2]_6 \,\dot{\cup}\, [3]_6 \,\dot{\cup}\, [4]_6 \,\dot{\cup}\, [5]_6$$

Because the *set* of these equivalence classes

$$\{[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6\}$$

actually forms a commutative ring. The additive identity is $[0]_g$, the multiplicative identity is $[1]_6$, the addition is

$$[r]_6 + [s]_6 = [r + s]_6$$

and the multiplication is given by

$$[r]_6 \cdot [s]_6 = [r \cdot s]_6$$

In this case, in fact this ring is $\mathbb{Z}/6$:

$$\mathbb{Z}/6 = \{[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6\}$$

Informally I have been saying

$$\mathbb{Z}/6 = \{0, 1, 2, 3, 4, 5\}$$

and then I say "well ... 1 is the same as 7 in mod 6" which is not really absolutely correct. It's more correct to say 1 is "related" to 7 through the $\equiv \pmod{6}$ relation, i.e.,

$$1 \equiv 7 \pmod{6}$$

which is the same as saying the equivalence classes

$$[1]_6 = \{..., -12, -6, \underline{1}, 7, 13, ...\} = 1 + \langle 6 \rangle$$

and

$$[7]_6 = \{..., -6, 1, \underline{7}, 13, 19, ...\} = 7 + \langle 6 \rangle$$

are the same:

$$[1]_6 = [7]_6$$

See it? Or you can say the "1" of $\mathbb{Z}/6$ is really another notation for $[1]_6$.

As an example of computing with $\mathbb{Z}/6$, if you have the following computation

$$\begin{aligned}
5x - 4 &\equiv 5 \pmod{6} \\
\therefore\ \ 5x &\equiv 5 + 4 \pmod{6} \\
\therefore\ \ 5x &\equiv 9 \pmod{6} \\
\therefore\ \ 5x &\equiv 3 \pmod{6} \\
\therefore\ \ 5^{-1} \cdot 5x &\equiv 5^{-1} \cdot 3 \pmod{6} \\
\therefore\ \ 1x &\equiv 5 \cdot 3 \pmod{6} \\
\therefore\ \ x &\equiv 15 \pmod{6} \\
\therefore\ \ x &\equiv 3 \pmod{6}
\end{aligned}$$

It's essentially the same as the following where $[\ ]$ is $[\ ]_6$:

$$\begin{aligned}
[5]x - [4] &= [5] \\
\therefore\ \ [5]x &= [5] + [4] \\
\therefore\ \ [5]x &= [9] \\
\therefore\ \ [5]x &= [3] \\
\therefore\ \ [5]^{-1}[5]x &= [5]^{-1} \cdot [3] \\
\therefore\ \ [1]x &= [5] \cdot [3] \\
\therefore\ \ x &= [15] \\
\therefore\ \ x &= [3]
\end{aligned}$$

You do have to distinguish between the x of

$$x \equiv 3 \pmod{6}$$

where x is an *integer* and

$$x = [3]$$

where x is an *equivalence class* of integers.

The first sequence of computations is how I usually present computations involving modular arithmetic. The second sequence of computations is more for theoretical understanding.

Make sure you stare at this very carefully and see the difference:

$$\begin{aligned}
\mathbb{Z} &= [0]_6 \mathbin{\dot\cup} [1]_6 \mathbin{\dot\cup} [2]_6 \mathbin{\dot\cup} [3]_6 \mathbin{\dot\cup} [4]_6 \mathbin{\dot\cup} [5]_6 \\
\mathbb{Z}/6 &= \{[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6\}
\end{aligned}$$

This is important so let me record this interpretation of $\mathbb{Z}/N$:

Proposition 202.4.13. *Let*

$$\mathbb{Z}/N = \{[0]_N, [1]_N, ..., [N-1]_N\}$$

Define the operations $+$ *(or* $+_N$*) and* $\cdot$ *(or* $\cdot_N$*) on* $\mathbb{Z}/N$ *by*

$$\begin{aligned}[][a] + [b] &= [a + b] \\ [a] \cdot [b] &= [a \cdot b]\end{aligned}$$

Then $(\mathbb{Z}/N, +, \cdot, [0]_N, [1]_N)$ *is a commutative ring.*

Most of the facts are obvious except for the following fact: Each $[a]$ is a set. You can choose an element a of $\mathbb{Z}$ to represent the set $[a]$. Whenever you choose a representative for a set, you have to make sure that functions/operations/operators are well-defined. As an example suppose I define

$$\begin{aligned}A &= \{x \in \mathbb{R} \mid x < 0\} \\ B &= \{x \in \mathbb{R} \mid 0 \leq x < 1\} \\ C &= \{x \in \mathbb{R} \mid 1 \leq x\}\end{aligned}$$

For A, suppose I choose a value in A to be a representative, say -1, i.e., $A = [-1]$. Note that I could have chosen another value, say -42, in A to represent A, i.e., $A = [-1] = [-42]$. For B, I choose 0.5, i.e., $B = [0.5]$. For C, I choose 1, i.e., $C = [1]$. Therefore I have $\{A, B, C\} = \{[-1], [0.5], [1]\}$. Suppose I define this function $f : \{A, B, C\} \rightarrow \{A, B, C\}$ based on the computation of representatives as follows: Let $x \in \{A, B, C\}$. Suppose $x = [y]$ where y is any representative of x.

$$f([y]) = [y + 1]$$

Is this a well-defined function? Well let's see what happens for A.

$$f(A) = f([-1]) = [-1 + 1] = [0]$$

Note that 0 is in B. Therefore $[0] = B$. In other words

$$f(A) = B$$

But wait a minute ... -42 is also in A. If I chose -42 to represent A, then

$$f(A) = f([-42]) = [-42 + 1] = [-41]$$

but -41 is in A, i.e., $[-41] = A$. Therefore

$$f(A) = [-41] = A$$

Yike!!! The definition above actually yields two different functional value at A:

$$\begin{aligned} f(A) &= f([-1]) = B \\ f(A) &= f([-42]) = A \end{aligned}$$

So f is not a well-defined function at A, i.e., it's not even a function! (However the above definition is actually well-defined for $f(B)$ and $f(C)$. I'll let you think about that.)

In general if someone claimed to define a function

$$f : X \to Y$$

where X is a set of sets and such that $f(x)$ (for $x \in X$, a set) is computed based on any element a in set x representing x, there is always this issue of whether $f(x)$ is really well–defined in the sense of whether the definition is independent of the choice of representative. Of course you do have to check that $f(x) \in Y$. Make sure you read the above several times!!!

Inside the proof below, you will see that I have to prove that if

$$\begin{aligned} a &\equiv a' \pmod{N} \\ b &\equiv b' \pmod{N} \end{aligned}$$

then

$$\begin{aligned} a + b &\equiv a' + b' \pmod{N} \\ ab &\equiv a'b' \pmod{N} \end{aligned}$$

Make sure you watch for it and why I have to prove it.

Proof. Exercise. Before even checking the ring axioms, you need to show that $+$ and $\cdot$ are well–defined. In other words, you need to show that if

$$\begin{aligned} [a] &= [a'] \\ [b] &= [b'] \end{aligned}$$

then

$$\begin{aligned}[][a+b] &= [a'+b'] \\ [ab] &= [a'b']\end{aligned}$$

Here, I'm writing $[\,]$ for $[\,]_N$ for simplicity. Translating the above to congruence relation, you have to show if

$$\begin{aligned}a &\equiv a' \pmod{N} \\ b &\equiv b' \pmod{N}\end{aligned}$$

then

$$\begin{aligned}a+b &\equiv a'+b' \pmod{N} \\ ab &\equiv a'b' \pmod{N}\end{aligned}$$

After you have proved the above, *then* you show that this $+$ and $\cdot$ satisfy the ring axioms for $\mathbb{Z}/N$. Let's prove the above. If

$$\begin{aligned}a &\equiv a' \pmod{N} \\ b &\equiv b' \pmod{N}\end{aligned}$$

then

$$\begin{aligned}N &\mid a-a' \\ N &\mid b-b'\end{aligned}$$

By linearity of divisibility,

$$N \mid (a-a') + (b-b') = (a+b) - (a'+b')$$

(which ring axioms and properties of ring did I use?) i.e.,

$$a+b \equiv a'+b' \pmod{N}$$

Also, from

$$\begin{aligned}N &\mid a-a' \\ N &\mid b-b'\end{aligned}$$

I obtain

$$N \mid (a - a')b$$
$$N \mid (b - b')a'$$

By linearity again, I get

$$N \mid (a - a')b + (b - b')a' = ab - a'b + ba' - b'a' = ab - a'b + a'b - a'b' = ab - a'b'$$

(check the above very carefully ...) which means

$$ab \equiv a'b' \pmod{N}$$

Now we know that $+$ and $\cdot$ are well-defined on $\mathbb{Z}/N$. We can proceed to check that $+$ and $\cdot$ satisfies the ring axioms on $\mathbb{Z}/N$. But ... □

Wait a minute ...

Maybe the above is true for R/n where R is a a general commutative ring and $n \in R$. Recall from the above that I define a relation $\equiv \pmod{n}$ on R so that

$$r \equiv s \pmod{n}$$

if $n \mid r - s$ (or $r - s \in \langle n \rangle$). This is an equivalence relation on R. Therefore we have equivalence classes of R. Let R/n be the set of equivalence classes:

$$R/n = \{[r]_n \mid r \in R\}$$

(remember: each $[r]_n$ is a set.) Here $[r]_n$ is the equivalence class of r, i.e.,

$$[r]_n = \{s \in R \mid r \equiv s \pmod{n}\}$$

I'm now going to make R/n into a ring: For $[r], [s] \in R/n$, define

$$[r] + [s] = [r + s]$$
$$[r] \cdot [s] = [rs]$$

I claim:

Proposition 202.4.14. *Let R be a commutative ring and $n \in R$. Then $(R/n, +, \cdot, [0]_n, [1]_n)$ is a commutative ring.*

R/n is an example of a **quotient ring**. (I say "an example", because this

quotient ring

concept of "R/n" can be generalized.)

Proof. First I should prove that the $+$ and $\cdot$ on R/n are well-defined. I'll write $[\,]$ for $[\,]_n$. Let

$$\begin{aligned} [r] &= [r'] \\ [s] &= [s'] \end{aligned}$$

I want to prove

$$\begin{aligned} [r] + [s] &= [r'] + [s'] \\ [r] \cdot [s] &= [r'] + [s'] \end{aligned}$$

i.e.,

$$\begin{aligned} [r + s] &= [r' + s'] \\ [rs] &= [r's'] \end{aligned}$$

From

$$\begin{aligned} [r] &= [r'] \\ [s] &= [s'] \end{aligned}$$

I get

$$\begin{aligned} n &\mid r - r' \\ n &\mid s - s' \end{aligned}$$

By linearity,

$$n \mid (r - r') + (s + s') = (r + s) - (r' + s')$$

Hence

$$r + s \equiv r' + s' \pmod{n}$$

From

$$\begin{aligned} n &\mid r - r' \\ n &\mid s - s' \end{aligned}$$

again by linearity

$$n \mid (r - r')s + (s - s')r' = rs - r's + sr' - s'r' = rs - s'r' = rs - r's'$$

(I just used the fact that R is a commutative ring. See it?) Hence

$$rs \equiv r's' \pmod{n}$$

Therefore both $+$ and $\cdot$ are well-defined on R/n.

Next, you are now ready to prove the ring axioms hold on R/n. To show associativity of $+$ on R/n, let $r, s, t \in R$. Then

$$([r] + [s]) + [t] = [r + s] + [t] = [(r + s) + t] = [r + (s + t)]$$

In the last step I just used the associativity of $+$ on R. Continuing,

$$([r] + [s]) + [t] = [r + (s + t)] = [r] + [s + t] = [r] + ([s] + [t])$$

You see that associativity of $+$ on R/n is essentially due the associativity of $+$ on R. The proofs of other ring axioms are the same. Make sure you complete the proof. □

With the above proposition, we see in particular that since $\mathbb{Z}$ is a commutative ring, $\mathbb{Z}/N$ is a commutative ring. This in turn implies that $(\mathbb{Z}/N)[X]$ is also a commutative ring. And if $n \in (\mathbb{Z}/N)[X]$ is a polynomial with coefficients in $(\mathbb{Z}/N)$, then $((\mathbb{Z}/N)[X])/n$ is also a commutative ring ... etc.

Exercise 202.4.12. Let $R = (\mathbb{Z}/6)[X]$ and $n = X^3 + 2X + 3$.

(a) Write down some elements of $[0]_n$.
(b) Write down some elements of $[1]_n$.
(c) Write down some elements of $[X + 1]_n$.
(d) Let $r = X^3 + X^2 + X + 1 \in (\mathbb{Z}/6)[X]$. Find a polynomial $s \in (\mathbb{Z}/6)[X]$ of smallest degree such that $r \equiv s \pmod{n}$, i.e., $[r]_n = [s]_n$ as elements in $((\mathbb{Z}/6)[X])/n$.
(e) Let $r = X^3 + X^2 + X + 1 \in (\mathbb{Z}/6)[X]$. Does $[r]_n$ have a multiplicative inverse in $((\mathbb{Z}/6)[X])/n$? By definition, this is the same as asking if there is some $[s]_n$ such that $[r]_n[s]_n = [1]_n$, which is the same as asking if there is some $s \in (\mathbb{Z}/6)[X]$ such that $rs \equiv 1 \pmod{n}$.
(f) Let $r = X^2 + X + 1 \in (\mathbb{Z}/6)[X]$. Does $[r]_n$ have a multiplicative inverse in $((\mathbb{Z}/6)[X])/n$?

Exercise 202.4.13. Let $R = (\mathbb{Z}/2)[X]$ and $n = X^3 + X + 1$.

(a) Write down some elements of $[0]_n$.
(b) Write down some elements of $[1]_n$.
(c) Write down some elements of $[X]_n$.
(d) Write down some elements of $[X+1]_n$.
(e) How many equivalence classes are there for $((\mathbb{Z}/2)[X])/n$? If it's finite, write down all of them.
(f) Let $r = X^3 + X^2 + X + 1 \in (\mathbb{Z}/2)[X]$. Find a polynomial $s \in (\mathbb{Z}/2)[X]$ of smallest degree such that $r \equiv s \pmod{n}$, i.e., $[r]_n = [s]_n$ as elements in $((\mathbb{Z}/2)[X])/n$.
(g) Let $r = X^3 + X^2 + X + 1 \in (\mathbb{Z}/2)[X]$. Does $[r]_n$ have a multiplicative inverse in $((\mathbb{Z}/2)[X])/n$?
(h) Let $r = X^2 + X + 1 \in (\mathbb{Z}/2)[X]$. Does $[r]_n$ have a multiplicative inverse in $((\mathbb{Z}/2)[X])/n$?
(i) For each $r \in (\mathbb{Z}/2)[X]$ of degree at most 2, write down $[r]_n^{-1}$ as $[s]_n$ where s has the smallest degree.
(j) Is $((\mathbb{Z}/2)[X])/n$ a field?

Elements of this R/n are essentially of the form $aX^2 + bX + c$ where $a, b, c \in \mathbb{Z}/2$. Technically, I should say every element of $(\mathbb{Z}/2)[X]$ is congruence to some $aX^2 + bX + c \pmod{X^3 + X + 1}$. $aX^2 + bX + c$ (where $a, b, c \in \mathbb{Z}/2$) are essentially all the possible remainder mod $X^3 + X + 1$. For instance one element is $1X^2 + 1X + 0$. Instead of writing $1X^2 + 1X + 0$, a shorthand is to write $(1, 1, 0)$ (the coefficients) or even 110. This is a field – your first example of a (binary) finite field. Finite fields are extremely important in CS and engineering in areas such as cryptography, information theory, coding theory, etc. In these areas such finite fields are more important than $\mathbb{R}$. There are researchers specializing just in finite fields.

Exercise 202.4.14. Let $R = (\mathbb{Z}/3)[X]$ and $n = X^2 + 1$. The elements of R/n are of the form $aX + b$ where $a, b \in \mathbb{Z}/3$. (More accurately, I should say $[aX + b]_n$ or that every polynomial in R is congruence to some $aX + b$.) Therefore there are 9 values in R. For $aX + b$, as a shorthand, you can write ab. For instance $2X + 1$ can be written $(2, 1)$.

1. Write down the + table of R/n.

2. Write down the $\cdot$ table of R/n.

3. Write down the multiplicative inverse table of R/n. (I.e., for each element of R/n, write down the multiplicative inverse of that element).

4. Is R/n a field?

5. Write a python program that allows you to add, multiple, compute the multiplicative inverse in R/n. Each element of R/n can be represented by a tuple `(a,b)` or list `[a, b]`.

Exercise 202.4.15. Here's a curious bonus ring for you: Consider the commutative ring $\mathbb{R}[X]$ of polynomials with real coefficients. Let $n = X^2 + 1$. Let's take a look at this R.

1. Write down some values, $[r]$ of R. Remember: These are equivalence classes of $\mathbb{R}[X]$ under the $\equiv \pmod{n}$ relation. Simplify them as much as you can.
2. What is the general (and simplest) form of $[r]$?
3. Pick a random pair of values in R and add them and multiple them and then simplify as much as you can.
4. Have you see this R before? In particular have you seen the value $[X]$ (the equivalence class of X in $\mathbb{R}[X]$) before? Does $[X]$ behave (algebraically) like something you have seen before?

202.4.5 A minimalistic polynomial library

Exercise 202.4.16. Write a program that allows you to work with $((\mathbb{Z}/N)[X])/n$. First of all a polynomial such as $2X^3+X+3$ can be represented by `[2, 0, 1, 3]`. You'll need to add and multiply polynomials. You'll also need to mod the coefficients of a polynomial and you'll also need to mod a polynomial by another. All the functions are done except for modding by a polynomial. Note for instance that

$$X^3 + X + 1 \equiv 0 \pmod{X^3 + X + 1}$$

and therefore

$$X^3 \equiv -X - 1 \pmod{X^3 + X + 1}$$

This means that

$$aX^3 \equiv -aX - a \pmod{X^3 + X + 1}$$

In general, if $n = X^3 + bX^2 + cX + d$, we have

$$aX^3 \equiv -abX^2 - acX - ad \pmod{X^3 + bX^2 + cX + d}$$

and more generally

$$aX^{k+3} \equiv -abX^{k+2} - acX^{k+1} - adX^k \pmod{X^3 + bX^2 + cX + d}$$

for $k \geq 0$. As an example:

$$aX^{10} \equiv -abX^9 - acX^8 - adX^7 \pmod{X^3 + bX^2 + cX + d}$$

This means that if you have a polynomial $aX^{10} + (...)$, then

$$aX^{10} + (...) \equiv (-abX^9 - acX^8 - adX^7) + (...) \pmod{X^3 + bX^2 + cX + d}$$

Therefore you can replace an X^{10} terms by terms of lower degree. You then apply the same process to the X^9 term of $(-abX^9 - acX^8 - adX^7) + (...)$, etc. until you get a polynomial of degree 2. The same idea works when $n = X^3+bX^2+cX+d$ has any degree other than 3. To be absolutely concrete, suppose $n = X^3 + X + 1 \in (\mathbb{Z}/2)[X]$ and you want to reduce $X^8 + X^5 + X$ by modding with n. Note that from

$$X^3 + X + 1 \equiv 0 \pmod{X^3 + X + 1}$$

I get

$$X^3 \equiv -X - 1 \pmod{X^3 + X + 1}$$

Note that $-1 \equiv 1 \pmod 2$, therefore

$$X^3 \equiv X + 1 \pmod{X^3 + X + 1}$$

Here we go (the term being replaced is underlined):

$$\begin{aligned}
&\underline{X^8} + X^5 + X^1 \\
&\equiv X^6 + X^5 + X^5 + X^1 = X^6 + 2X^5 + X^1 \equiv \underline{X^6} + X^1 \pmod{X^3 + X + 1} \\
&\equiv \underline{X^4} + X^3 + X^1 \\
&\equiv X^2 + X^1 + X^3 + X^1 = X^3 + X^2 + 2X^1 = \underline{X^3} + X^2 \\
&\equiv X^1 + X^0 + X^2 = X^2 + X^1 + X^0
\end{aligned}$$

You are basically doing long division to get the remainder. If you do the actual long division, you'll also get the quotient as well. Notice that the continual replacement of X^k by smaller powers is very similar to the process of computing the remainder of an integer when modding by N. It's because of this that in fact the whole theory of extended euclidean algorithm of $\mathbb{Z}$ actually works for polynomial rings with coefficients in a field (and $\mathbb{Z}/2$ is a field). More formally, if K is a field and $K[X]$ is the polynomial ring over K, then both $\mathbb{Z}$ and $K[X]$ are Euclidean domains, i.e., integral domain satisfying the Euclidean property. For Z, this means given a, b, there are integer q and r such that

$$a = bq + r, \ \ 0 \leq r < |b|$$

and for $K[X]$, given polynomials a, b, there are polynomials q, r such that

$$a = bq + r, \ \ r = 0 \text{ or } \deg(r) < \deg(b)$$

Both concepts can be combined and generalize to an Euclidean domain: a ring R is an Eucliean domain if is it an integral domain and there is a function $f : R - \{0\} \rightarrow \{0, 1, 2, ...\}$ such that given $a, b \in R$, there exist $q, r \in R$ such that

$$a = bq + r, \ \ r = 0 \text{ or } f(r) < f(b)$$

In terms of implementation, for the above, if you have the following array

```
[0,1,0,0,0,1,0,0,1]
```

where the value at index i is the coefficient of X^i, you remove that `1` at index 8 (for X^8)

```
[0,1,0,0,0,1,0,0,0]
```

and add this (i.e., $X^6 + X^5$)

```
[0,0,0,0,0,1,1,0,0]
```

to get this

```
[0,1,0,0,0,0,1,0,0]
```

Continuing with the same process, it becomes

```
[0,1,0,1,1,0,0,0,0]
```

and then

```
[0,0,1,1,0,0,0,0,0]
```

and finally

```
[1,1,1,0,0,0,0,0,0]
```

The above should remind you of LFSR (linear feedback shift register).

```
# A simple polynomial library

def modpoly(p, n, N):
    ret = None
    return ret

def degree(p, n, N):
    return len(q) - 1

def coef(p, i):
    if i < len(p):
        return p[i]
    else:
        return 0

def addpoly(p, q, n, N):
    ret = None
    return None

def multpoly(p, q, n, N):
    ret = None
    return ret

def invpoly(p, n, N):
    ret = None
    return ret
```

For the brute force search for inverses, try this:

```
import itertools
for p in itertools.product(list(range(2)), repeat=3):
    print(p)
```

which gives you

```
(0, 0, 0)
(0, 0, 1)
(0, 1, 0)
(0, 1, 1)
(1, 0, 0)
(1, 0, 1)
(1, 1, 0)
(1, 1, 1)
```

□

Note that when $R[X]/n$ is a ring, you also want to find inverses of invertible elements. For instance since $(\mathbb{Z}/2[X]/(X^3+X+1)$ is a field, every nonzero element of $(\mathbb{Z}/2[X]/(X^3+X+1)$ (i.e., non-zero binary polynomial of degree ≤ 2) are invertible. In $\mathbb{Z}/N$, the inverse of a (if it exists) can be computed either by brute force or by using the Extended Euclidean Algorithm. Because $(\mathbb{Z}/2)[X]$ is also an Euclidean domain, it also has an Extended Euclidean Algorithm and therefore inverses in $(\mathbb{Z}/2)[X]/(X^3+X+1)$ can also be computed using its Extended Euclidean Algorithm. However in the above python library, to simplify the code, we'll just do brute force.

In our example of $(\mathbb{Z}/2[X]/(X^3+X+1)$, which is a field, there are 8 elements. A field where elements are of the form aX^2+bX+c with $a, b, c \in \mathbb{Z}/2$ is denoted by $\mathbb{F}_{2^3}$ or $\text{GF}(2^3)$ (GF = Galois field, in honor of Galois). In general if $(\mathbb{Z}/p[X]/n)$ is a field where n has degree d, then the number of elements is p^d. And such a field is denoted by $\mathbb{F}_{p^d}$ or $\text{GF}(p^d)$.

```
File:  euclidean-property.tex
```

202.5 Euclidean property

$\mathbb{Z}$ satisfies this very important property: Given integers a and $b \neq 0$, there are always integers q and r such that

$$a = bq + r$$

with

$$0 \leq r < |b|$$

Of course q is called the **quotient** when a is divided by b; r is the **remainder**. For instance if $a = 25$ and $b = 3$, then

quotient
remainder

$$25 = 3 \cdot 8 + 1, \quad 0 \leq 1 < 3$$

The computation

$$a, b \to q, r$$

is called a **division algorithm**.

division algorithm

In Python, you can do this:

```
a = 25
b = 8
q, r = divmod(25, 8)
print("%s = %s * %s + %s" % (a, b, q, r))
```

```
[student@localhost elementary-number-theory] python divmod.py
25 = 8 * 3 + 1
```

Algorithmically, when a and b have a huge number of digits and they are represented using arrays of digits, the division algorithm to compute q, r is basically long division you learnt in middle school. Note that this is the same when a and b are polynomials with coefficients in a field. Which is why I've said many times that the theory of $\mathbb{Z}$ is very similar to for instance $(\mathbb{Z}/p)[X]$ (where p is a prime, since $\mathbb{Z}/p$ is a field).

If we peek ahead and pretend for the time being that fractions such as $\frac{a}{b}$ exists, then for $a > 0$ and $b > 0$, we have

$$q = \left\lfloor \frac{a}{b} \right\rfloor, \quad r = a - bq$$

where $\lfloor x \rfloor$ means the floor of x. If we write `(a/b)` for the *integer* quotient of a by b (i.e. this is the / in C++ for integers) and `a%b` for the corresponding remainder, then of course we have

```
a = b * (a/b) + (a%b)
```

Again, this is an *extremely* important theorem.

Theorem 202.5.1. (**Euclidean property**) *If $a, b \in \mathbb{Z}$ with $b \neq 0$, then there are integers q and r satisfying*

Euclidean property

$$a = bq + r, \quad 0 \leq |r| < |b|$$

The above theorem is the general version that can be generalized to general rings. Below is the version for $\mathbb{Z}$. The only difference is the $|r|$ is replaced by r:

Theorem 202.5.2. (**Euclidean property 2**) *If $a, b \in \mathbb{Z}$ with $b \neq 0$, then there are integers q and r satisfying*

Euclidean property 2

$$a = bq + r, \quad 0 \leq r < |b|$$

In many cases, one is interested in the case when $a \geq 0$. So this version is the one you'll find in most textbooks:

Theorem 202.5.3. (**Euclidean property 3**) *If $a, b \in \mathbb{Z}$ with $a \geq 0, b > 0$, then there are integers $q \geq 0$ and $r \geq 0$ satisfying*

Euclidean property 3

$$a = bq + r, \quad 0 \leq r < b$$

Recall that $\mathbb{Z}$ is a domain in the sense that it is a commutative ring and for $r, s \in \mathbb{Z}$, if $r \cdot s = 0$, then $r = 0$ or $s = 0$. We say that $\mathbb{Z}$ is an **Euclidean domain** in the sense that

Euclidean domain

1. $\mathbb{Z}$ is a domain
2. $\mathbb{Z}$ satisfies the Euclidean property

(The above can be generalized to Euclidean rings and Euclidean domains, but

I won't do it here.)

Although the above Euclidean property is for $\mathbb{Z}$, I'll prove it for $a \geq 0$ and $b > 0$. And the q, r will satisfy $q \geq 0$, $r \geq 0$. (Furthermore in this setup q, r are unique.) Once you have proven the Euclidean property for integer $a \geq 0$, it will not be difficult to extend the result to the whole of $\mathbb{Z}$.

To prove the Euclidean property of $\mathbb{Z}$, we'll bring in the big guns:

1. Well-ordering principle
2. Principle of weak mathematical induction
3. Principle of string mathematical induction

Here we go ...

Well-ordering principle for $\mathbb{N}$: Let X be a subset of $\mathbb{N}$. Then X has a minimal element. In other words there is some $m \in X$ such that $m \leq x$ for all $x \in X$.

Well-ordering principle for $\mathbb{N}$

Note that the well-ordering principle is really a statement about the subsets $\mathbb{N}$. This is an axiom about $\mathbb{N}$ that we assume holds. You can prove the following version of well-ordering principle on $\mathbb{Z}$:

Well-ordering principle for $\mathbb{Z}$: Let X be a subset of $\mathbb{Z}$ that is *bounded below*. Then X has a minimal element. In other words there is some $m \in X$ such that $m \leq x$ for all $x \in X$.

Well-ordering principle for $\mathbb{Z}$

$\mathbb{R}$ does not satisfy the second version well-ordering principle with $\mathbb{Z}$ replaced by $\mathbb{R}$. For instance the open interval $X = (0, 1)$ is bounded below (for instance by -42). However there is no m in X such that $m \leq x$ for all x in X. For instance $m = 0.01 \in X$ is not a minimum element of X since $0.0001 \in X$ is smaller than m. Also, $m = 0.0000001 \in X$ is also not a minimum of X since $0.0000000001 \in X$ is less than m. In fact for any $m \in X$, $(1/2)m$ is in X and is less than m. In other words no value in X can be a minimum value of X.

Here's the weak mathematical induction on $\mathbb{Z}$:

Weak mathematical induction for $\mathbb{Z}$: Let n_0 be an integer in $\mathbb{Z}$ and let $P(n)$ be a proposition for integer $n \in \mathbb{Z}$ and $n \geq n_0$. If

Weak mathematical induction for $\mathbb{Z}$

1. $P(n_0)$ holds
2. $P(n)$ holds implies $P(n+1)$ holds for all $n \geq n_0$

then $P(n)$ holds for all $n \geq n_0$.

And here's the strong mathematical induction for $\mathbb{Z}$:

Strong mathematical induction for $\mathbb{Z}$

Strong mathematical induction for $\mathbb{Z}$: Let n_0 be an integer in $\mathbb{Z}$ and let $P(n)$ be a proposition for integer $n \in \mathbb{Z}$ and $n \geq n_0$. If

1. $P(n_0)$ holds
2. $P(n_0), P(n_0+1), ..., P(n)$ holds implies $P(n+1)$ holds for all $n \geq n_0$

then $P(n)$ holds for all $n \geq n_0$.

It can be shown that the following are equivalent:

1. Well-ordering principle of $\mathbb{Z}$
2. Weak mathematical induction of $\mathbb{Z}$
3. Strong mathematical induction of $\mathbb{Z}$

In other words, if you believe the well-ordering principle is true, then any of the two induction principles also hold, and vice versa. And of course we all believe in the well-ordering principle for $\mathbb{Z}$ and hence both induction principles.

I'll give you two proofs of Theorem 202.5.3. The standard one in most books uses the well-ordering principle. For the first proof, I'll use induction.

Proof 1. We want to prove that if $a \geq 0$ and $b > 0$, there exists $q \geq 0, r \geq 0$ such that

$$a = bq + r, \ \ 0 \leq r < b$$

We'll prove the theorem by mathematical induction. So let's form our $P(n)$. Let b be fixed. Let $P(n)$ be the statement:

$$P(n) : \text{There are integer } q, r \text{ such that } n = bq + r,\ 0 \leq r < b$$

(I'm thinking of the original a as a variable in the $P(n)$.)

The base case is easy: If $n = 0$, then if we set $q = 0$ and $r = 0$, then we do have

$$0 = b \cdot 0 + 0, \quad 0 \leq 0 < b$$

Hence $P(0)$ holds.

Now for the inductive case. Assume that $P(n)$ holds. Now we consider the statement $P(n+1)$. Note that $P(n)$ is true. Therefore there are integers q, r satisfying

$$n = bq + r, \quad 0 \leq r < b$$

Therefore

$$n + 1 = bq + r + 1$$

Either $r = b - 1$ or $r < b - 1$. (Note that r cannot be greater than $b - 1$ since $r < b$.)

<u>Case: $r = b - 1$.</u> Then we have

$$\begin{aligned} n &= bq + r \\ \therefore \quad n + 1 &= bq + r + 1 \\ \therefore \quad n + 1 &= bq + b \\ \therefore \quad n + 1 &= b(q + 1) \\ \therefore \quad n + 1 &= b(q + 1) + 0, 0 \leq 0 < b \end{aligned}$$

So if we set $q' = q + 1$ and $r' = 0$ we do have

$$n + 1 = bq' + r', \quad 0 \leq r' < b$$

In other words $P(n + 1)$ holds.

<u>Case: $r < b - 1$.</u> Then we have

$$\begin{aligned} n &= bq + r, \quad 0 \leq r, \quad r < b - 1 \\ \therefore \quad n + 1 &= bq + r + 1, \quad 0 \leq r, \quad r < b - 1 \\ \therefore \quad n + 1 &= bq + r + 1, \quad 0 \leq r, \quad r + 1 < b \\ \therefore \quad n + 1 &= bq + r + 1, \quad 0 \leq r + 1, \quad r + 1 < b \\ \therefore \quad n + 1 &= bq + r + 1, \quad 0 \leq r + 1 < b \end{aligned}$$

So if we set $q' = q$ and $r' = r + 1$, we have

$$n + 1 = bq' + r', \quad 0 \leq r' < b$$

<u>Case: $r > b - 1$.</u> This case cannot occur since we have the fact that $r < b$.

Therefore in all cases, $P(n + 1)$ holds.

Since $P(0)$ is true and if $P(n)$ holds, then so does $P(n + 1)$ for $n \geq 0$, by mathematical induction, $P(n)$ must be true for all $n \geq 0$. □

Now let me prove Theorem 202.5.3 using the well-ordering principle. Remember that I'm proving it for $a \geq 0, b > 0, q \geq 0, r \geq 0$.

Proof 2. Recall that we are assuming that $b > 0$. Let $X = \{a - bx \mid x \in \mathbb{Z}, a - bx \geq 0\} \subseteq \mathbb{N} \cup \{0\}$. X is not empty since $a = a - b \cdot 0 \geq 0$ is in X. X is bounded below by 0 (by definition). By the well–ordering principle X has a minimal element. Call it r. Note that $r \in \mathbb{N} \cup \{0\}$ (hence $r \geq 0$). Furthermore $r = a - bq$ for some $q \in \mathbb{Z}$ and hence

$$a = bq + r, \;\; 0 \leq r$$

I will now prove that $r < b$.

Suppose on the contrary that $r \geq b$. Then $0 \leq r - b$.

$$a = bq + r = bq + (r - b + b) = b(q + 1) + (r - b)$$

and therefore

$$a - b(q + 1) = (r - b) < r$$

Note that $0 \leq r - b < r$. Hence

$$0 \leq a - b(q + 1) = (r - b) < r$$

In other words $a - b(q + 1) \in X$ and is smaller than $a - bq$ which contradicts the minimality of $a - bq$.

Note that $q \geq 0$. For otherwise $q < 0$ implies that $bq + r \leq b(-1) + r < 0$ since $r < b$. □

Proposition 202.5.1. *Given a, b, the q, r in Theorem 202.5.3 are unique. In other words, if*

$$\begin{aligned} a &= bq + r, \;\; 0 \leq r < |b| \\ a &= bq' + r', \;\; 0 \leq r' < |b| \end{aligned}$$

then

$$q = q', \quad r = r'$$

Proof. From $bq + r = a = bq' + r'$, we have

$$bq + r = bq' + r'$$

If $q = q'$, then $r = r'$. We now assume $q \neq q'$. Without loss of generality, we'll assume that $q > q'$. We have

$$r' = b(q - q') + r > b + r \geq b$$

which contradicts $r' < b$. □

Now I'm going to prove Theorem 202.5.1 which allows a to be any integer.

Proof of Theorem 202.5.1. Now I'll use Euclidean Property 3 to prove Euclidean Property 1. We just need to handle the case when $a < 0$. Let u be ± 1 so that $ua \geq 0$. Let v be ± 1 so that $vb > 0$. Note that $(\pm 1)^2 = 1$, i.e., $u^{-1} = u, v^{-1} = v$. Using Euclidean Property 3, there exist $q' \geq 0, r'$ such that

$$a' = b'q' + r', \ \ 0 \leq r' < b'$$

Then

$$ua = vbq' + r', \ \ 0 \leq r' < vb = |b|$$

Multiplying by u^{-1}

$$a = uvbq' + ur', \ \ 0 \leq r' < vb = |b|$$

and hence

$$a = b(uvq') + ur', \ \ 0 \leq |ur'| < vb = |b|$$

(Note that $r' \geq 0$ and hence $|ur'| = |u||r'| = r'$.) Hence if $q = uvq'$ and $r = ur'$, then

$$a = bq + r, \ \ 0 \leq |r| < |b|$$

and we are done. □

Exercise 202.5.1. Prove: If $a \in \mathbb{Z}$ and $b \in \mathbb{Z}$ and $b \neq 0$, then there are unique integers q, r such that $a = bq + r$ and $b \leq r < 2b$.

Exercise 202.5.2. Using the Euclidean property, prove that every integer is congruence to $0, 1, 2,$ or $3 \mod 4$.

Exercise 202.5.3. Prove that squares are 0 or 1 mod 4. In other words if $a \in \mathbb{Z}$, then $a^2 \equiv 0 \text{ or } 1 \pmod 4$.

Exercise 202.5.4. Solve $4x^3 + y^2 = 5z^2 + 6$ (in $\mathbb{Z}$).

Exercise 202.5.5. Prove that $11, 111, 1111, 11111, 111111, ...$ are all not perfect squares. (An integer is a perfect square is it's of the form a^2 where a is an integer.)

Exercise 202.5.6. How many of $3, 23, 123, 1123, 11123, 111123, 1111123, ...$ are perfect squares?

File: gcd.tex

202.6 Euclidean algorithm – GCD

Now let me use the Euclidean property to compute the gcd of two integers.

Let's use the division algorithm on 20 and 6.

$$20 = 6 \cdot 3 + 2, \quad 0 \leq 2 < 6$$

Suppose I want to compute $\gcd(20, 6)$. Of course the example is small enough that we know that it is 2. But notice something about this:

$$20 = 6 \cdot 3 + 2, \quad 0 \leq 2 < 6$$

Note that if d is a divisor of 20 and 6, then it must also divide 2. Therefore $\gcd(20, 6)$ must divide 2. The converse might not be true. In general suppose you have an equation of the form

$$a = bq + r$$

If d divides a and b, then d divides $a - bq = r$ (by linearity of divisibility). And if d divides b and r, then d divides $bq + r = a$ (again by linearity of divisibility). In other words,

$$d \text{ is a common divisor of } a, b \iff d \text{ is a common divisor of } b, r$$

Therefore we have this crucial fact, the bridge between Euclidean property and common divisors:

Proposition 202.6.1. *If $a, b, q, r \in \mathbb{Z}$ such that*

$$a = bq + r$$

then

1. $\{d \mid d \textit{ is a common divisor of } a, b\} = \{d \mid d \textit{ is a common divisor of } b, r\}$
2. $\gcd(a, b) = \gcd(b, r)$

Note that in the above, I only require $a = bq + r$. For instance for to $\gcd(120, 15)$, I can use $120 = 1 \cdot 15 + (120 - 15)$, i.e., $a = 120, b = 15, q = 1, r =$

$120 - 15$. Then $\gcd(120, 15) = \gcd(15, 120 - 15) = \gcd(15, 105)$.

However if I use the division algorithm, then r is “small”:

$$0 \leq r < b$$

So if you want to compute $\gcd(a, b)$, make sure $a \geq b$ (otherwise swap them). Then $\gcd(a, b) = \gcd(b, r)$ and you would have $a \geq b > r$. So instead of computing $\gcd(a, b)$, you are better off computing $\gcd(b, r)$.

But like I said, we do not need the q and r to be the quotient and remainder. For instance suppose I want to compute the GCD of 514 and 24.

$$514 = 24 \cdot 1 + (514 - 24)$$

Then

$$\gcd(514, 24) = \gcd(24, 514 - 24)$$

which gives us

$$\gcd(514, 24) = \gcd(24, 490)$$

Of course we can shrink the number 514 to something smaller than 490. From

$$514 = 24 \cdot 10 + (514 - 24 \cdot 10)$$

we have

$$\gcd(514, 24) = \gcd(24, 514 - 24 \cdot 10) = \gcd(24, 514 - 240) = \gcd(24, 274)$$

Looks like we can subtract 10 more of 24.

$$\gcd(24, 274) = \gcd(24, 274 - 24 \cdot 10) = \gcd(24, 274 - 240) = \gcd(24, 34)$$

One more ...

$$\gcd(24, 34) = \gcd(24, 34 - 24) = \gcd(24, 10)$$

Now we subtract from 24 multiples of 10:

$$\gcd(24, 10) = \gcd(24 - 2 \cdot 10, 10) = \gcd(4, 10)$$

Now we subtract multiples of 4 from 10:

$$\gcd(4, 10) = \gcd(4, 10 - 2 \cdot 4) = \gcd(4, 2)$$

Now we subtract multiples of 2 from 4:

$$\gcd(4 - 2 \cdot 2, 2) = \gcd(0, 2)$$

Of course $\gcd(0, 2) = 2$.

Note that $\gcd(0, n) = n$ for any positive integer n. I'll let you think about that one. (Remember what I said before: 0 is in some sense a big number, like a black hole. Because every positive number divides 0.)

Of course this gives rise to the following algorithm

```
ALGORITHM: GCD
Inputs: a, b
Output: gcd(a, b)

if b > a:
    swap a, b

if b == 0:
    return a
else:
    return GCD(a - b, b)
```

This only subtract a one copy of b from a. Suppose we can compute

$$a = bq + r, \quad 0 \leq r < b$$

Then

$$\gcd(a, b) = \gcd(b, r)$$

Of course r is the remainder when a is divided by b. Using this we rewrite the above code to get the **Euclidean Algorithm**:

Euclidean Algorithm

```
ALGORITHM: GCD (Euclidean algorithm)
Inputs: a, b
Output: gcd(a, b)

if b > a:
    # To make sure that for gcd(a,b), a >= b
    swap a, b

if b == 0:
    return a
else:
    return GCD(b, a % b)
```

Note that if `a` $<$ `b`, then

```
GCD(a, b) = GCD(b, a % b) = GCD(b, a)
```

Therefore the swap is not necessary:

```
ALGORITHM: GCD (Euclidean algorithm)
Inputs: a, b
Output: gcd(a, b)

if b == 0:
    return a
else:
    return GCD(b, a % b)
```

In this case, I'm assuming that `a % b` is available. Or if you prefer using a loop:

```
ALGORITHM: GCD (Euclidean algorithm)
Inputs: a, b
Output: gcd(a, b)

while 1:
    if b == 0:
        return a
    else:
        a, b = b, a % b
```

As an example:

$$\begin{aligned}
\text{gcd}(514, 24) &= \text{gcd}(24, 514\%24) = \text{gcd}(24, 10) \\
&= \text{gcd}(10, 24\%10) = \text{gcd}(10, 4) \\
&= \text{gcd}(10, 10\%4) = \text{gcd}(10, 2) \\
&= \text{gcd}(2, 10\%2) = \text{gcd}(2, 0) \\
&= 2
\end{aligned}$$

Exercise 202.6.1. Compute the following using the Euclidean Algorithm explicitly.

1. $\gcd(0, 10)$
2. $\gcd(10, 0)$
3. $\gcd(10, 1)$
4. $\gcd(10, 10)$
5. $\gcd(107, 5)$
6. $\gcd(107, 26)$
7. $\gcd(84, 333)$
8. $\gcd(F_{10}, F_{11})$ where F_n is the n–th Fibonacci number. $(F_0 = 0, F_1 = 1, F_{n+2} = F_{n+1} + F_n.)$
9. $\gcd(ab, b)$
10. $\gcd(a, a + 1)$
11. $\gcd(ab + a, b)$ where $0 < a < b$. Go as far as you can.
12. $\gcd(a(a + 1) + a, (a + 1))$ where $0 < a < b$. Go as far as you can.

Exercise 202.6.2. Leetcode 650. `https://leetcode.com/problems/2-keys-keyboard/` There is only one character `'A'` on the screen of a notepad. You can perform one of two operations on this notepad for each step:

Copy All: You can copy all the characters present on the screen (a partial copy is not allowed).

Paste: You can paste the characters which are copied last time.

Given an integer `n`, return the minimum number of operations to get the character `'A'` exactly `n` times on the screen.

Exercise 202.6.3. Leetcode 2447. `https://leetcode.com/problems/number-of-subarrays-with-g` Given an integer array `nums` and an integer `k`, return the number of subarrays of `nums` where the greatest common divisor of the subarray's elements is `k`. A subarray is a contiguous non-empty sequence of elements within an array. The greatest common divisor of an array is the largest integer that evenly divides all the array elements.

Exercise 202.6.4. Leetcode 1998 `https://leetcode.com/problems/gcd-sort-of-an-array/` You are given an integer array `nums`, and you can perform the following operation any number of times on nums:

Swap the positions of two elements `nums[i]` and `nums[j]` if `gcd(nums[i], nums[j]) > 1` where `gcd(nums[i], nums[j])` is the greatest common divisor of `nums[i]` and `nums[j]`. Return true if it is possible to sort `nums` in non-decreasing order using the above swap method, or false otherwise.

File: extended-euclidean-algorithm.tex

202.7 Extended Euclidean algorithm: GCD as linear combination

Here's another super important fact:

Theorem 202.7.1. (**Extended Euclidean Algorithm**) *If a and b be integers which are not both zero, then there are integers x, y such that*

Extended Euclidean Algorithm

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

The x, y in the above theorem are called **Bézout's coefficients** of a, b. They are not unique.

Bézout's coefficients

Exercise 202.7.1. Prove that $a \neq 0$, then there are many possible x, y such that $ax + by = \gcd(a, b)$. □

First let me prove that there are x, y such that

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

The theorem does not give you the algorithm. Then I'll do a computational example that compute the gcd of a, b as a linear combination of a and b. The example actually contains the idea behind the algorithm to compute the Bézout's coefficients. The algorithm is called the Extended Euclidean Algorithm.

Proof. For convenience, let me write (a, b) as the set of linear combinations of a and b, i.e.,

$$(a, b) = \{ax + by \mid x, y \in \mathbb{Z}\}$$

I will also write (g) for the linear combination of g, i.e.,

$$(g) = \{gx \mid x \in \mathbb{Z}\}$$

(Such linear combinations are called ideals. They are extremely important in of themselves. Historically, they were created to study Fermat's last theorem.

Since then they are crucial in the the study of ring theory.)

The proof proceeds in two steps:

1. Given a, b not both zero, there is some g such that $(a, b) = (g)$. If a, b are not both zero, g can be chosen to be > 0.
2. The g in the above is in fact $\gcd(a, b)$.

Now let's prove step 1, i.e., given a, b, there is some g such that

$$(a, b) = (g)$$

First of all if $b = 0$, then by definition

$$(a, 0) = (a)$$

and we're done. Next, we now assume $b \neq 0$. Then $|b| > 0$. The set

$$X = \{ax + by \mid x, y \in \mathbb{Z} \text{ and } ax + by > 0\} \subseteq \mathbb{N}$$

is nonempty since it contains $0 \cdot a + 1 \cdot |b|$. By the well-ordering principle of $\mathbb{N}$, X has a minimum element, say g. I will now show that $(a, b) = (g)$.

Since g is a minimum element of X, g is in X. Therefore $g = ax + by$. Hence $gz = a(xz) + b(yz) \in (a, b)$ for all $z \in \mathbb{Z}$. This implies that $(g) \subseteq (a, b)$.

Now I will prove that $(a, b) \subseteq (g)$. Let $c \in (a, b)$, i.e., $c = ax + by$ for some $x, y \in \mathbb{Z}$. Therefore by the Euclidean property of $\mathbb{Z}$, there exists $q, r \in \mathbb{Z}$ such that

$$c = gq + r, \ \ 0 \leq r < g$$

(Look at the definition of X again. X is a subset of $\mathbb{N}$ so that $g \geq 1$). If $r \neq 0$, then

$$r = c - gq$$

Note that $c = ax + by$ by our assumption. We have already shown that $(g) \subseteq (a, b)$, i.e., $g = ax' + by'$. Therefore, altogether we have

$$r = c - gq = ax + by - (ax' + b'y)q = a(x - x'q) + b(y - y'q)$$

Hence $r \in X$. But $0 \leq r < d$ implies that

$$r = a(x - x'q) + b(y - b'q)$$

is an element of X which is less than g which contradicts the minimality of g.

Hence $r = 0$ and we have

$$c = gq + r = gq \in (g)$$

I have shown that $(a, b) \subseteq (g)$.

Altogether, I've shown $(a, b) = (g)$. Step 1 is now completed.

For step 2, I will show that g is the gcd of a and b. Since $(a, b) = (g)$, we have

$$a \in (a, b) = (g)$$

i.e., $a = xg$ which means g divides a. Likewise g divides b. Hence g is a common divisor of a and b. Since $(g) \subseteq (a, b)$, $g = ax_0 + by_0$. Suppose d is any divisor of a and b. Then $d \mid ax_0 + by_0$ by the linearity of divisibility. Hence $d \mid g$. Therefore g is the largest common divisor of a and b, i.e., $g = \gcd(a, b)$. □

The above does not give you an algorithm to compute the x and y. First let me do an example to show you that it's possible to compute $\gcd(a, b)$ as a linear combination of a and b. Then I'll give you the algorithm.

Recall that I have computed $\gcd(514, 24) = 2$. Extended Euclidean Algorithm says that it's possible to find x and y such that

$$2 = \gcd(514, 24) = 514x + 24y$$

How do we compute the x and y? Just like the gcd computation (the Euclidean Algorithm), the x, y are computed using the Euclidean property. First we have

$$514 = 21 \cdot 24 + 10$$

This implies that

$$514 \cdot 1 + 24 \cdot (-21) = 10$$

Now it would be nice if the pesky 10 goes away and is replaced by 2. How would we do that? Well look at 24 and 10 now. We have

$$24 = 2 \cdot 10 + 4$$

again by Euclidean algorithm. Multiplying the equation

$$514 \cdot 1 + 24 \cdot (-21) = 10$$

throughout by 2 gives us

$$514 \cdot 2 + 24 \cdot (-42) = 2 \cdot 10$$

The previous equation

$$24 = 2 \cdot 10 + 4$$

say that $2 \cdot 10$ can be replaced by $24 - 4$. This means that

$$514 \cdot 2 + 24 \cdot (-42) = 24 - 4$$

Hmmm ... this says that we have now

$$514 \cdot 2 + 24 \cdot (-43) = -4$$

or

$$514 \cdot (-2) + 24 \cdot 43 = 4$$

What about 4? Well, if we look at 10 and 4 just like what we did with 24 and 10 we would get

$$10 = 2 \cdot 4 + 2$$

and the remainder 2 gives us the GCD!!! Rearranging it a bit we have

$$1 \cdot 10 + (-2) \cdot 4 = 2$$

i.e. 2 is a linear combination of 10 and 4. But earlier we say that 4 is a linear combination of 514 and 24 ...

$$514 \cdot (-2) + 24 \cdot 43 = 4$$

and even earlier we saw that 10 is also a linear combination of 514 and 24 ...

$$514 \cdot 1 + 24 \cdot (-21) = 10$$

Surely if we substitute all these values into the equation

$$1 \cdot 10 + (-2) \cdot 4 = 2$$

we would get 2 as a linear combination of 514 and 24. Let's do it ...

$$\begin{aligned}
2 &= 1 \cdot 10 + (-2) \cdot 4 \\
&= 1 \cdot (514 \cdot 1 + 24 \cdot (-21)) + (-2)(514 \cdot (-2) + 24 \cdot 43) \\
&= 514 \cdot 1 + 24 \cdot (-21) + 514 \cdot 4 + 24 \cdot (-86) \\
&= 514 \cdot 5 + 24 \cdot (-107)
\end{aligned}$$

Vóila!

Exercise 202.7.2. Using the above idea, compute the gcd and Bézout's coefficients of 210 and 78, i.e., compute x and y such that $210x+78y = \gcd(210, 78)$.

Exercise 202.7.3. Analyze the above and design an algorithm so that when given a and b, the algorithm computes x and y such that $ax + by = \gcd(a, b)$.

To help you analyze the above computation, let me organize our computations a little. If we can make the process systematic, then there is hope that we can make the idea work for all a and b, i.e., then we would have an algorithm and hence can program it and compute it's runtime performance.

We know for sure that we have to continually use Euclidean property on pairs of numbers. So here we go:

$$\begin{aligned} 514 &= 21 \cdot 24 + 10 \\ 24 &= 2 \cdot 10 + 4 \\ 10 &= 2 \cdot 4 + 2 \\ 4 &= 2 \cdot 2 + 0 \end{aligned}$$

Note that this corresponds to the gcd computation

$$\begin{aligned} \gcd(514, 24) &= \gcd(24, 514 - 21 \cdot 24) = \gcd(24, 10) \\ &= \gcd(10, 24 - 2 \cdot 10) = \gcd(10, 4) \\ &= \gcd(4, 10 - 2 \cdot 4) = \gcd(4, 2) \\ &= \gcd(2, 4 - 2 \cdot 2) = \gcd(2, 0) \\ &= 2 \end{aligned}$$

So in the computation

$$\begin{aligned} 514 &= 21 \cdot 24 + 10 \\ 24 &= 2 \cdot 10 + 4 \\ 10 &= 2 \cdot 4 + 2 \\ 4 &= 2 \cdot 2 + 0 \end{aligned}$$

if the remainder is 0 (see the last line), then the previous line's remainder must be the gcd.

Let's look at our computation of the gcd of 514 and 24:

$$\begin{aligned} 514 &= 21 \cdot 24 + 10 \\ 24 &= 2 \cdot 10 + 4 \\ 10 &= 2 \cdot 4 + 2 \\ 4 &= 2 \cdot 2 + 0 \end{aligned}$$

Recall that the above computation means that the gcd is 2. Note only that through backward substitution, we can rewrite 2 as a linear combination of 514 and 24.

Let's try to do this in a more organized way. So here's our facts again:

$$\begin{aligned}
514 &= 21 \cdot 24 + 10 \\
24 &= 2 \cdot 10 + 4 \\
10 &= 2 \cdot 4 + 2
\end{aligned}$$

Let me put the remainders on one side:

$$\begin{aligned}
10 &= 514 - 21 \cdot 24 && (1) \\
4 &= 24 - 2 \cdot 10 && (2) \\
2 &= 10 - 2 \cdot 4 && (3)
\end{aligned}$$

Note that (1) tells you that 10 is a linear combination of $514, 24$. (2) tells you that 4 is a linear combination of $24, 10$. If I substitute (1) into (2), 4 will become a linear combination of $514, 24$. (3) says that 2 is a linear combination of $10, 4$. But 10 is a linear combination of $514, 24$ and 4 is a linear combination of $514, 24$. Hence 2 is also a linear combination of $514, 24$. See it?

OK. Let's do it. From

$$\begin{aligned}
10 &= 514 - 21 \cdot 24 && (1) \\
4 &= 24 - 2 \cdot 10 && (2) \\
2 &= 10 - 2 \cdot 4 && (3)
\end{aligned}$$

if I substitute (1) into (2) and (3) (i.e., rewrite 10 as a linear combination of $514, 24$):

$$\begin{aligned}
10 &= 514 - 21 \cdot 24 && (1) \\
4 &= 24 - 2 \cdot (514 - 21 \cdot 24) && (2) \\
2 &= (514 - 21 \cdot 24) - 2 \cdot 4 && (3)
\end{aligned}$$

Collecting the multiples of 514 and 24:

$$\begin{aligned}
10 &= 514 - 21 \cdot 24 && (1) \\
4 &= (-2) \cdot 514 + (1 + (-2)(-21)) \cdot 24 && (2') \\
2 &= (1) \cdot 514 + (-21) \cdot 24 - 2 \cdot 4 && (3')
\end{aligned}$$

and simplifying:

$$\begin{aligned} 10 &= 514 - 21 \cdot 24 && (1) \\ 4 &= (-2) \cdot 514 + (43) \cdot 24 && (2') \\ 2 &= (1) \cdot 514 + (-21) \cdot 24 - 2 \cdot 4 && (3') \end{aligned}$$

Substituting $(2')$ into $(3')$:

$$\begin{aligned} 10 &= 514 - 21 \cdot 24 && (1) \\ 4 &= (-2) \cdot 514 + (43) \cdot 24 && (2') \\ 2 &= (1) \cdot 514 + (-21) \cdot 24 - 2 \cdot ((-2) \cdot 514 + (43) \cdot 24) && (3') \end{aligned}$$

Tidying up:

$$\begin{aligned} 10 &= 514 - 21 \cdot 24 && (1) \\ 4 &= (-2) \cdot 514 + (43) \cdot 24 && (2') \\ 2 &= (1 - 2(-2)) \cdot 514 + (-21 - 2(43)) \cdot 24 && (3'') \end{aligned}$$

Simplifying:

$$\begin{aligned} 10 &= 514 - 21 \cdot 24 && (1) \\ 4 &= (-2) \cdot 514 + (43) \cdot 24 && (2') \\ 2 &= (5) \cdot 514 + (-107) \cdot 24 && (3'') \end{aligned}$$

(It's a good idea to check after each substitution that the equalities still hold. We all make mistakes, right?)

OK. That's great. It looks more organized now. So much so that you can now easily write a program to compute the above.

Now let's look at the general case. Suppose instead of 514 and 24, I write a and b. The computation will look like this:

$$\begin{aligned} a &= q_1 \cdot b + r_1 \\ b &= q_2 \cdot r_1 + r_2 \\ r_1 &= q_3 \cdot r_2 + r_3 \\ r_2 &= q_4 \cdot r_3 + 0 \end{aligned}$$

To make things even more regular and uniform, let me rewrite it this way:

$$\begin{aligned}
r_0 &= q_1 \cdot r_1 + r_2 \\
r_1 &= q_2 \cdot r_2 + r_3 \\
r_2 &= q_3 \cdot r_3 + r_4 \\
r_3 &= q_4 \cdot r_4 + 0
\end{aligned}$$

A lot nicer, right? Let me write it this way with the remainder term on the lefts:

$$\begin{aligned}
r_2 &= (1) \cdot r_0 + (-q_1) \cdot r_1 \\
r_3 &= (1) \cdot r_1 + (-q_2) \cdot r_2 \\
r_4 &= (1) \cdot r_2 + (-q_3) \cdot r_3
\end{aligned}$$

(Remember that r_4 is the gcd ... $r_0 = 514, r_1 = 24$ ... right?) Organized this way, I have the gcd on one side of the equation. Now if I substitute the first equation into the second I get

$$\begin{aligned}
r_2 &= (1) \cdot r_0 + (-q_1) \cdot r_1 \text{ ... USED} \\
r_3 &= (1) \cdot r_1 + (-q_2) \cdot ((1) \cdot r_0 + (-q_1) \cdot r_1) \\
r_4 &= (1) \cdot r_2 + (-q_3) \cdot r_3
\end{aligned}$$

i.e.,

$$\begin{aligned}
r_2 &= (1) \cdot r_0 + (-q_1) \cdot r_1 \text{ ... USED} \\
r_3 &= (-q_2) \cdot r_0 + (1 + q_1 q_2) \cdot r_1 \\
r_4 &= (1) \cdot r_2 + (-q_3) \cdot r_3
\end{aligned}$$

Note that I can't throw away the first equation yet! I need to keep r_2 around since it appears in the third equation! So when can I throw the first equation away? Look at the general case. Suppose we have

$$\begin{aligned}
r_2 &= (1) \cdot r_0 + (-q_1) \cdot r_1 \\
r_3 &= (1) \cdot r_1 + (-q_2) \cdot r_2 \\
r_4 &= (1) \cdot r_2 + (-q_3) \cdot r_3 \\
r_5 &= (1) \cdot r_3 + (-q_4) \cdot r_4 \\
r_6 &= (1) \cdot r_4 + (-q_5) \cdot r_5 \\
&\ldots
\end{aligned}$$

Aha! r_2 is used only in the next *two* equations.

Suppose we are at equation 3:

$$\begin{aligned} r_3 &= c_1 \cdot r_0 + d_1 \cdot r_1 \\ r_4 &= c_2 \cdot r_0 + d_2 \cdot r_1 \end{aligned}$$

We have to compute the next equation: This requires r_3, r_4. Then we have

$$r_5 = (1) \cdot r_3 + (-q_4) \cdot r_4$$

where

$$q_4 = \lfloor r_3 / r_4 \rfloor, \quad r_5 = r_3 - q_4 r_4$$

Altogether we have

$$\begin{aligned} r_3 &= c_1 \cdot r_0 + d_1 \cdot r_1 \\ r_4 &= c_2 \cdot r_0 + d_2 \cdot r_1 \\ r_5 &= (1) \cdot r_3 + (-q_4) \cdot r_4 \end{aligned}$$

The last equation becomes

$$r_5 = c_1 \cdot r_0 + d_1 \cdot r_1 + (-q_4) \cdot (c_2 \cdot r_0 + d_2 \cdot r_1)$$

i.e.

$$r_5 = (c_1 - q_4 c_2) \cdot r_0 + (d_1 - q_4 d_2) \cdot r_1$$

Let me repeat that in a slightly more general context. If we have

$$\begin{aligned} r_3 &= c_1 \cdot r_0 + d_1 \cdot r_1 \\ r_4 &= c_2 \cdot r_0 + d_2 \cdot r_1 \end{aligned}$$

then we get (throwing away the first equation):

$$\begin{aligned} r_4 &= c_2 \cdot r_0 + d_2 \cdot r_1 \\ r_5 &= (c_1 - q_4 c_2) \cdot r_0 + (d_1 - q_4 d_2) \cdot r_1 \end{aligned}$$

To put it in terms of numbers instead of equations this is what we get: If we have

$$c_1, d_1, c_2, d_2, r_3, r_4$$

then we get

$$c_2, d_2, c_1 - \lfloor r_3/r_4 \rfloor c_2, d_1 - \lfloor r_3/r_4 \rfloor d_2, r_4, r_3 - \lfloor r_3/r_4 \rfloor r_4$$

In general, if we have

$$c, d, c', d', r, r'$$

then we get

$$c', d', c - \lfloor r/r' \rfloor c', d - \lfloor r/r' \rfloor d', r', r - \lfloor r/r' \rfloor r'$$

Of course since we start off with r_0, r_1 (i.e. what we call a and b above), we have

$$\begin{aligned} r_0 &= 1 \cdot r_0 + 0 \cdot r_1 \\ r_1 &= 0 \cdot r_0 + 1 \cdot r_1 \end{aligned}$$

i.e., you would start off with

$$c = 1, d = 0, c' = 0, d' = 1, r = r_0, r' = r_1$$

Let's check this algorithm with our $r_0 = 514, r_1 = 24$.

STEP 1: The initial numbers are

$$c = 1, d = 0, c' = 0, d' = 1, r = 514, r' = 24$$

Again this corresponds to

$$\begin{aligned} r_3 &= 1 \cdot 514 + 0 \cdot 24 \\ r_4 &= 0 \cdot 514 + 1 \cdot 24 \end{aligned}$$

STEP 2: The new numbers (6 of them) are

$$\begin{aligned} c' &= 0 \\ d' &= 1 \\ c - \lfloor r/r' \rfloor c' &= 1 - \lfloor 514/24 \rfloor 0 = 1 \\ d - \lfloor r/r' \rfloor d' &= 0 - \lfloor 514/24 \rfloor 1 = 0 - 21 = -21 \\ r' &= 24 \\ r - \lfloor r/r' \rfloor r' &= 514 - \lfloor 514/24 \rfloor 24 = 514 - 504 = 10 \end{aligned}$$

So the new numbers (we reset the variables in the algorithm):

$$c = 0, d = 1, c' = 1, d' = -21, r = 24, r' = 10$$

These corresponds to the data on the second and third line of the following:

$$\begin{aligned} 514 &= 1 \cdot 514 + 0 \cdot 24 \\ 24 &= 0 \cdot 514 + 1 \cdot 24 \\ 10 &= 1 \cdot 514 + (-21) \cdot 24 \end{aligned}$$

STEP 3: From the 6 numbers from STEP 2 we get

$$\begin{aligned} c' &= 1 \\ d' &= -21 \\ c - \lfloor r/r' \rfloor c' &= 0 - \lfloor 24/10 \rfloor 1 = -2 \\ d - \lfloor r/r' \rfloor d' &= 1 - \lfloor 24/10 \rfloor (-21) = 1 + 42 = 43 \\ r' &= 10 \\ r - \lfloor r/r' \rfloor r' &= 24 - \lfloor 24/10 \rfloor 10 = 24 - 20 = 4 \end{aligned}$$

So the new numbers (we reset the variables in the algorithm):

$$c = 1, d = -21, c' = -2, d' = 43, r = 10, r' = 4$$

These corresponds to the data on the third and fourth line of the following:

$$\begin{aligned} 514 &= 1 \cdot 514 + 0 \cdot 24 \\ 24 &= 0 \cdot 514 + 1 \cdot 24 \\ 10 &= 1 \cdot 514 + (-21) \cdot 24 \\ 4 &= (-2) \cdot 514 + 43 \cdot 24 \end{aligned}$$

STEP 4: From the 6 numbers from STEP 3 we get

$$\begin{aligned} c' &= -2 \\ d' &= 43 \\ c - \lfloor r/r' \rfloor c' &= 1 - \lfloor 10/4 \rfloor (-2) = 1 + 4 = 5 \\ d - \lfloor r/r' \rfloor d' &= -21 - \lfloor 10/4 \rfloor (43) = -21 - 86 = -107 \\ r' &= 4 \\ r - \lfloor r/r' \rfloor r' &= 10 - \lfloor 10/4 \rfloor 4 = 10 - 8 = 2 \end{aligned}$$

So the new numbers (we reset the variables in the algorithm):

$$c = -2, d = 43, c' = 5, d' = -107, r = 4, r' = 2$$

These corresponds to the data on the fourth and fifth line of the following:

$$\begin{aligned} 514 &= 1 \cdot 514 + 0 \cdot 24 \\ 24 &= 0 \cdot 514 + 1 \cdot 24 \\ 10 &= 1 \cdot 514 + (-21) \cdot 24 \\ 4 &= (-2) \cdot 514 + 43 \cdot 24 \\ 2 &= 5 \cdot 514 + (-107) \cdot 24 \end{aligned}$$

STEP 5: From the 6 numbers from STEP 4 we get

$$\begin{aligned} c' &= 5 \\ d' &= -107 \\ c - \lfloor r/r' \rfloor c' &= -2 - \lfloor 4/2 \rfloor 5 = -2 - 10 = -12 \\ d - \lfloor r/r' \rfloor d' &= 43 - \lfloor 4/2 \rfloor (-107) = 43 + 214 = 257 \\ r' &= 2 \\ r - \lfloor r/r' \rfloor r' &= 4 - \lfloor 4/2 \rfloor 2 = 4 - 4 = 0 \end{aligned}$$

So the new numbers (we reset the variables in the algorithm):

$$c = 5, d = -107, c' = -12, d' = 257, r = 2, r' = 0$$

These corresponds to the data on the fifth and sixth line of the following:

$$\begin{aligned} 514 &= 1 \cdot 514 + 0 \cdot 24 \\ 24 &= 0 \cdot 514 + 1 \cdot 24 \\ 10 &= 1 \cdot 514 + (-21) \cdot 24 \\ 4 &= (-2) \cdot 514 + 43 \cdot 24 \\ 2 &= 5 \cdot 514 + (-107) \cdot 24 \\ 0 &= (-12) \cdot 514 + 257 \cdot 24 \end{aligned}$$

Of course (as before) at this point, you see that the $r' = 0$. Therefore

$$\gcd(514, 24) = 2$$

and furthermore from $c = 5, d = -107$, we get

$$5 \cdot 514 + (-107) \cdot 24 = \gcd(514, 24)$$

Here's a Python implementation with some test code:

```
ALGORITHM: EEA
INPUTS: a, b
OUTPUTS: r, c, d where r = gcd(a, b) = c*a + d*b

    a0, b0 = a, b
    d0, d = 0, 1
    c0, c = 1, 0
    q = a0 // b0
    r = a0 - q * b0

    while r > 0:
        d, d0 = d0 - q * d, d
        c, c0 = c0 - q * c, c

        a0, b0 = b0, r
        q = a0 // b0
        r = a0 - q * b0

    r = b0
    return r, c, d
```

You can pound real hard at the Extended Euclidean Algorithm with this:

By the way, this is somewhat similar to what we call *tail recursion* (CISS445) an extremely important technique in functional programming. All LISP hackers and people working in high performance computing and compilers swear by it. You don't see recursion in the above code, but you can replace the while-loop with recursion and if you have a compiler/interpreter that can perform true tail recursion, then it would run exactly like the above algorithm.

Exercise 202.7.4. Leetcode 365: `https://leetcode.com/problems/water-and-jug-problem/description/` and the Die Hard 3 problem `https://www.math.tamu.edu/~dallen/hollywood/diehard/diehard.htm`. You are given two jugs with capacities `jug1Capacity` and `jug2Capacity` liters. There is an infinite amount of water supply available. Determine whether it is possible to measure exactly `targetCapacity` liter using these two jugs.

If targetCapacity liters of water are measurable, you must have targetCapacity liters of water contained within one or both buckets by the end.

Operations allowed:

1. Fill any of the jugs with water.
2. Empty any of the jugs.
3. Pour water from one jug into another till the other jug is completely full, or the first jug itself is empty.

You'll see that there are times when you're only interested in the value of x and not y (or y and not x – the above is symmetric about x and y). Do you notice x comes from c? If you analyze the above algorithm, you see immediately that c is computed from c' and c' is computed from c, c', q, q is computed from r, r', r is computed from r', and finally (phew!) r' is computed from r, q, r'. Therefore if you're interested in c, you don't need to compute d or d'. So you can change the EEA to this:

```
ALGORTHM: EEA2 (sort of EEA ... without the d, d0)
INPUTS: a, b
OUTPUTS: r, c where r = gcd(a, b) = c*a + d*b for some d

    a0, b0 = a, b
    c0, c = 1, 0
    q = a0 // b0
    r = a0 - q * b0

    while r > 0:
        c, c0 = c0 - q * c, c

        a0, b0 = b0, r
        q = a0 // b0
        r = a0 - q * b0

    r = b0
    return r, c
```

Later you'll see why we compute only c. It's not that I have something against d.

Exercise 202.7.5. Compute the following gcd and the Bézout's coefficients of the given numbers by following the Extended Euclidean Algorithm.

1. $\gcd(0, 10)$
2. $\gcd(10, 0)$
3. $\gcd(10, 1)$
4. $\gcd(10, 10)$
5. $\gcd(107, 5)$
6. $\gcd(107, 26)$
7. $\gcd(84, 333)$
8. $\gcd(F_{10}, F_{11})$ where F_n is the n–th Fibonacci number. (Recall: $F_0 = 0, F_1 = 1, F_{n+2} = F_{n+1} + F_n$.)
9. $\gcd(ab, b)$
10. $\gcd(a, a + 1)$
11. $\gcd(ab + a, b)$ where $0 < a < b$. Go as far as you can.
12. $\gcd(a(a + 1) + a, (a + 1))$ where $0 < a < b$. Go as far as you can.

Exercise 202.7.6. Prove that if $a \mid c$, $b \mid c$, and $\gcd(a, b) = 1$, then $ab \mid c$.

Exercise 202.7.7. Prove that if $a \mid c$, $b \mid c$, then

$$\frac{ab}{\gcd(a, b)} \mid c$$

Exercise 202.7.8. Leetcode 920. https://leetcode.com/problems/number-of-music-playlists Your music player contains n different songs. You want to listen to goal songs (not necessarily different) during your trip. To avoid boredom, you will create a playlist so that:

Every song is played at least once.

A song can only be played again only if k other songs have been played.

Given `n`, `goal`, and `k`, return the number of possible playlists that you can create. Since the answer can be very large, return it modulo $109 + 7$.

File: computation-of-multiplicative-inverse-mod-N.tex

202.8 Computation of multiplicative inverse in $\mathbb{Z}/N$

Quick review: In your algebra classes, you spend lots of time solving equations. You usually start with something like: "Find the roots of $x + b$". This means finding a value for x such that

$$x + a = 0$$

Of course this is dead easy. As long as you have the concept of $-a$ you just add $-a$ to both sides and voila:

$$\begin{aligned}(x + a) + (-a) &= 0 + (-a)\\ x + (a + (-a)) &= 0 + (-a)\\ x + 0 &= 0 + (-a)\\ x &= 0 + (-a)\\ x &= -a\end{aligned}$$

using various ring properties/theorems/etc that we know. (Of course your algebra prof would be amazed that you actually show all the steps in such a precise manner ...)

Next up, you usually see this: "Find the roots of $ax + b$" which is the same as finding a value for x such that

$$ax + b = 0$$

Assuming you are working in $\mathbb{R}$, you would do something like this:

$$\begin{aligned}(ax + b) + (-b) &= 0 + (-b)\\ ax + (b + (-b)) &= 0 + (-b)\\ ax + 0 &= 0 + (-b)\\ ax &= 0 + (-b)\\ ax &= -b\end{aligned}$$

and at this point you would do this:

$$\begin{aligned} ax &= -b \\ a^{-1}(ax) &= a^{-1}(-b) \\ (a^{-1}a)x &= a^{-1}(-b) \\ 1x &= a^{-1}(-b) \\ x &= a^{-1}(-b) \end{aligned}$$

and vóila (again) and you're done.

In the case of real numbers (or even just fractions), if you're given a number a (which is not zero), you always have another number called a^{-1} such that

$$a \cdot a^{-1} = 1 = a^{-1}a$$

The reason why we like this is exactly because it allows us to solve the above equation.

Given a number a, the number $-a$ is called an *additive inverse* if it behaves like this:

$$a + (-a) = 0 = (-a) + a$$

The number a^{-1} is called a *multiplicative inverse* of a if it does this:

$$a \cdot a^{-1} = 1 = a^{-1}a$$

If a has a multiplicative inverse, we say that a is invertible.

Almost all values in $\mathbb{R}$ are invertible; the only exception being 0. On the other hand, note that most numbers in $\mathbb{Z}$ do not have multiplicative inverses. In fact the only numbers in $\mathbb{Z}$ that have inverses are 1 and -1. So the only units in $\mathbb{Z}$ are $1, -1$.

So the ability for a number (in some ring) to have a multiplicative inverse (in that ring) is special. We also say that a is a **unit** if a has a multiplicative inverse.

unit

On the other hand, $\mathbb{Z}/N$ $(N > 1)$ might contain lots of units. For instance look at $\mathbb{Z}/10$. Can you find a number x such that

$$3x \equiv 1 \pmod{10}$$

That x would be the multiplicative inverse of 3 in mod 10. In general, for modular arithmetic in $\mathbb{Z}/N$, $N > 1$. (You can talk about $\mathbb{Z}/N$ for $N = 1$, but it's not very interesting nor useful. $\mathbb{Z}/1$ has only one value, i.e., 0 which behaves like the additive identity element 0 and the multiplicative identity element 1.)

Exercise 202.8.1.

1. Find all the units and their multiplicative inverses in $\mathbb{Z}/10$.
2. Find all the units and their multiplicative inverses in $\mathbb{Z}/5$.
3. Find all the units and their multiplicative inverses in $\mathbb{Z}/6$.

□

... here endeth the review.

It turns out that in $\mathbb{Z}/N$, you can easily find all the elements with multiplicative inverses: $a \in \mathbb{Z}/N$ has a multiplicative inverse if $\gcd(a, N) = 1$. (Yes, it's that "useless-thing-from-high-school" ... the gcd.)

Proposition 202.8.1. *Let a, N be integers where $N > 0$. Then a is (multiplicatively) invertible in $\mathbb{Z}/N$ iff* $\gcd(a, N) = 1$.

If m, n are two integers such that $\gcd(m, n) = 1$, we say that m and n are **coprime**. coprime

Proof. ($\Longrightarrow$): Suppose a has an inverse in mod N. Then there is some b such that

$$ab \equiv 1 \pmod{N}$$

Hence $ab = Nk + 1$ for some integer k. If $\gcd(a, N) \neq 1$, then there is some $d > 1$ that divides a and N. This implies that d divides $ab - Nk$ (by linearity of divisibility). However $ab - Nk = 1$. This implies that $d > 1$ divides 1, which is a contradiction.

($\Longleftarrow$): Let $\gcd(a, N) = 1$. By Extended Euclidean Algorithm, there are integers x, y such that

$$1 = \gcd(a, N) = ax + Ny$$

Taking mod N, you get

$$1 \equiv ax + 0 \pmod{N}$$

i.e.,

$$ax \equiv 1 \pmod{N}$$

Hence a has a multiplicative inverse. □

Neat right? It's amazing how useful is the Extended Euclidean Algorithm.

The above immediately implies that

$$\begin{aligned}(\mathbb{Z}/N)^\times &= \{a \mid 0 \leq a \leq N-1,\ \ a \text{ is invertible mod } N\} \\ &= \{a \mid 0 \leq a \leq N-1,\ \ \gcd(a,N) = 1\}\end{aligned}$$

(Of course $\gcd(0, N) = N > 0$, so we can throw away 0 in the above if $N > 1$.) So the key is the computation of x, y such that

$$1 = \gcd(a, N) = ax + Ny$$

which is achieved by the Extended Euclidean Algorithm. But wait a minute ... I just need the x. So I'll just modify the Extended Euclidean Algorithm slightly.

Here's the (earlier) EEA algorithm, slightly modified, together with a function to compute the multiplicative inverse of $a \pmod{N}$:

```
ALGORITHM: EEA2 (sort of EEA ... without the d, d0)
INPUTS: a, b
OUTPUTS:  r, c where r = gcd(a, b) = c*a + d*b for some d

    a0, b0 = a, b
    c0, c = 1, 0
    q = a0 // b0
    r = a0 - q * b0

    while r > 0:
        c, c0 = c0 - q * c, c

        a0, b0 = b0, r
        q = a0 // b0
        r = a0 - q * b0

    r = b0
    return r, c

ALGORITHM: inverse
```

```
INPUTS: a, N
OUTPUT: x such that  (a * x) % N is 1

    g, x = EEA2(a, N)
    if g == 1:
        return x % N
    else:
        return None
```

Of course you can have a brute force (and inefficient) way to

```
ALGORITHM: brute-force-inverse
INPUT: a, N
OUTPUT: x such that a * x % N is 1

    for x = 1, 2, 3, ..., N - 1:
        if a * x % N == 1:
            return x
    return None
```

Example 202.8.1. Does 135 have an inverse mod 1673? If it does, find it using the Extended Euclidean Algorithm.

Solution.

1. c0, c, q, r: 1, 0, 0, 135
2. c0, c, q, r: 0, 1, 12, 53
3. c0, c, q, r: 1, -12, 2, 29
4. c0, c, q, r: -12, 25, 1, 24
5. c0, c, q, r: 25, -37, 1, 5
6. c0, c, q, r: -37, 62, 4 4
7. c0, c, q, r: 62, -285, 1, 1
8. c0, c, q, r: -285, 347, 4, 0
9. r, c: 1, 347

Therefore $135^{-1} \pmod{1673}$ is 347. And we check:

$$135 \cdot 347 = 34845 = 1 + 28 \cdot 1673 \equiv 1 \pmod{1673}$$

Exercise 202.8.2. Compute the $\gcd(16, 123)$. If it's 1, find x such that $16x \equiv$

1 (mod 123). Use the above version of Extended Euclidean Algorithm and compute by hand. When you're done, write a program implementing the above algorithm and check that it gives you the same result. Solve the equation

$$16x + 5 \equiv 0 \pmod{123}$$

i.e. find an integer x such that $0 \leq x < 123$ satisfying the above congruence.

File: prime.tex

202.9 Primes

A prime p is a positive integer greater than 1 that is divisible by only 1 and itself. In other words $p \in \mathbb{N}$ is a **prime** if $d \mid p$, then $d = 1$ or $d = p$. prime

Of course you know that.

Examples of primes are $2, 3, 5, 7, 11, 13, 17, 19, ...$

For integers at least zero, we can divide them into the following types:

- 0 – the zero element
- 1 – the unit element (i.e. the only invertible element ≥ 0)
- primes – $2, 3, 5, 7, 11, ...$
- composites – integers > 0 which are not primes

(It's also possible to define primes of $\mathbb{Z}$. A prime of $\mathbb{Z}$ is an integer not $-1, 0, 1$ such that if $d \mid p$, then $d = \pm 1$ or $d = \pm p$. In that case primes of $\mathbb{Z}$ are $\pm 2, \pm 3, \pm 5, \pm 7, \pm 11, ...$)

The follow lemma is extremely important and is used for instance in the fundamental theorem of arithmetic to prove the uniqueness of prime factorization in $\mathbb{N}$ (or $\mathbb{Z}$).

Theorem 202.9.1. *(Euclid's lemma) If p is a prime and $p \mid ab$, then either $p \mid a$ or $p \mid b$.*

Proof. We assume that $p \nmid a$ otherwise there's nothing to prove. We know that there exists integer x and y such that

$$\gcd(a, p) = ax + py$$

$\gcd(a, p)$ divides p and hence $\gcd(a, p)$ is 1 or p. $\gcd(a, p)$ cannot be p for otherwise $\gcd(a, p) = p$ would imply that p would divide a. Hence $\gcd(a, p)$ must be 1. We now have

$$1 = ax + py$$

Hence

$$b = abx + pby$$

Since p divides ab and p divides p, by linearity of divisibility, p must divide $abx + pby$. Therefore p divides b. □

The above generalizes easily to the following:

Theorem 202.9.2. *If p is a prime and $p \mid n_1 n_2 \cdots n_l$, then p divides one of the $n_1, \ldots, n_l$.*

File: `fundamental-theorem-of-arithmetic.tex`

202.10 Fundamental Theorem of Arithmetic

Theorem 202.10.1. *(Fundamental Theorem of Arithmetic) Every positive integer > 1 can be written as a unique product of primes up to permutation of the prime factors. This means*

(a) If $a > 1$ is an integer, then a can be written as a product of primes.

(b) If a is written as two products of primes:

$$a = p_1 p_2 \cdots p_m = q_1 q_2 \cdots q_n$$

where p_i and q_j are primes arranged in ascending order, i.e.,

$$\begin{aligned} p_1 \leq p_2 \leq \cdots \leq p_m \\ q_1 \leq q_2 \leq \cdots \leq q_n \end{aligned}$$

then $m = n$ and

$$p_1 = q_1, \quad p_2 = q_2, \quad \cdots, \quad p_m = q_m,$$

(a) This can be proven using mathematical induction. Let $P(n)$ be the statement:

$$P(n) : n \text{ is the product of primes}$$

for $n \geq 2$.

The base case is easy: $P(2)$ is true since 2 is a product of a single prime, i.e. itself.

Now suppose $P(2), P(3), \ldots, P(n)$ is true and consider $n + 1$. We know that either $n+1$ is a prime or a composite. ($n+1$ is not 0 or 1.) If $n+1$ is a prime, then $n+1$ is a product of itself; $P(n+1)$ is true. If $n+1$ is a composite, then n has a proper divisor, say d. This implies that

$$n + 1 = dm$$

We have

$$m < dm$$

since $1 < d$ and

$$1 < m$$

since $d < n$. Likewise $1 < d < n$. By inductive hypothesis, $P(d)$ and $P(m)$ are true. Therefore d and m is a product of primes, say

$$d = p_1 \cdots p_k m = q_1 \cdots q_l$$

Hence

$$n + 1 = dm = p_1 \cdots p_k q_1 \cdots q_l$$

which is also a product of primes.

Hence by mathematical induction, $P(n)$ is true for all $n > 1$. In other words if $n > 1$, then n is a product of primes.

(b) Now suppose

$$p_1 \cdots p_m = q_1 \cdots q_n$$

Note that p_1 divides $p_1 \cdots p_m$. Since $p_1 \cdots p_m = q_1 \cdots q_n$, p_1 also divide $q_1 \cdots q_n$. By Euclid's lemma, p_1 must divide one of $q_1, \ldots, q_n$. Say p_1 divides q_1. Since q_1 is a prime, p_1 must be 1 or q_1. But p_1 is a prime, therefore it cannot be 1. Hence $p_1 = q_1$. Since $p_1 \neq 0$, using the cancellation property of $\mathbb{Z}$ (because $\mathbb{Z}$ is an integral domain), we get

$$p_2 \cdots p_m = q_2 \cdots q_n$$

Continue this process, we see that primes in $p_1, \ldots, p_m$ must match exactly the primes in $q_1, \ldots, q_n$. □

Proposition 202.10.1. *Let $a = \prod_{p \in P} p^{a_p}$, $b = \prod_{p \in P} p^{b_p}$ and $c = \prod_{p \in P} p^{c_p}$. Then*

(a) $c = ab \implies c_p = a_p + b_p$. *What if $c = a + b$?*

(b) $a \mid b \implies a_p \leq b_p$ *for all $p \in P$.*

(c) $c = \gcd(a, b) \implies c_p = \min(a_p, b_p)$.

(d) $\gcd(a, b) = \prod_{p \in P} p^{\min(a_p, b_p)}$.

File: euler-totient-function.tex

202.11 Euler Totient Function

Definition 202.11.1. Let N be a positive integer. $\phi(N)$ is the number of positive integers from 0 to $N-1$ which are coprime to N, i.e.

$$\phi(N) = |\{a \,|\, 0 \leq a \leq N-1,\ \ \gcd(a,N)=1\}|$$

Note that you can also view ϕ in this way:

$$\begin{aligned}\phi(N) &= |\{a \mid 0 \leq a \leq N-1,\ \ \gcd(a,N)=1\}| \\ &= |\{a \mid 0 \leq a \leq N-1,\ \ a \text{ is invertible mod } N\}| \\ &= |(\mathbb{Z}/N^{\times})|\end{aligned}$$

Note that by definition $\phi(1) = 1$. Here are some important properties of ϕ.

Proposition 202.11.1.

(a) *If m, n are coprime, i.e.* $\gcd(m,n) = 1$, *then* $\phi(mn) = \phi(m)\phi(n)$.
(b) *If p is a prime and $k > 0$, then* $\phi(p^k) = p^{k-1}(p-1) = p^k - p^{k-1}$.
(c)
$$\phi(N) = N \prod_{p|N} (1 - 1/p)$$

Here "$\prod_{p|N}$" means "product over all primes p dividing N".

Proof. (a) Let a be an integer. Then

$$\gcd(a,mn) = 1 \iff \gcd(a,m) = 1 \text{ and } \gcd(a,n) = 1$$

since for a prime p,

$$\begin{aligned}p \mid \gcd(a,mn) &\iff p \mid a \text{ and } p \mid mn \\ &\iff p \mid a \text{ and } (p \mid m \text{ or } p \mid n) \\ &\iff (p \mid a \text{ and } p \mid m) \text{ or } (p \mid a \text{ and } p \mid n) \\ &\iff p \mid \gcd(a,m) \text{ or } p \mid \gcd(a,n)\end{aligned}$$

Hence if p is a prime,

$$p \nmid \gcd(a, mn) \iff p \nmid \gcd(a, m) \text{ and } p \nmid \gcd(a, n)$$

Therefore if $\gcd(a, mn) = 1$, then for any prime p, p does not divide $\gcd(a, m)$ nor $\gcd(a, n)$. Hence $\gcd(a, m) = 1 = \gcd(a, n)$. Likewise if $\gcd(a, m) = 1 = \gcd(a, n)$, then for every prime p, p does not divide $\gcd(a, mn)$, and hence $\gcd(a, mn) = 1$. (This can also be proven using prime factorization.)

(b) $\phi(p^k)$ is the number of integers a in $[0, p^k - 1]$ such that $\gcd(a, p^k) = 1$. Clearly $\gcd(a, p^k) = 1$ iff $\gcd(a, p)$. The number of integers in $[0, p-1]$ divisible by p is exactly 1 (only 0 is divisible by p). The number of integers in $[p, 2p-1]$ divisible by p is exactly 1 (only p is divisible by p). The number of integers in $[2p, 3p-1]$ divisible by p is exactly 1 (only $2p$ is divisible by p). Etc. Altogether the number of integers in $[0, \ell p - 1]$ (where $\ell > 0$) divisible by p is exactly ℓ. Hence the number of integers in $[0, p^{k-1}p - 1] = [0, p^k - 1]$ divisible by p is exactly p^{k-1}. Therefore the number of integers in $[0, p^k - 1]$ *not* divisible by p is $p^k - p^{k-1}$.

(c) Let n have prime factorization $\prod_{i=1}^g p_i^{e_i}$ where $e_i > 0$. Then

$$\begin{aligned}
\phi(n) &= \phi\left(\prod_{i=1}^g p_i^{e_i}\right) \\
&= \prod_{i=1}^g \phi\left(p_i^{e_i}\right) \\
&= \prod_{i=1}^g \left(p_i^{e_i} - p_i^{e_i-1}\right) \\
&= \prod_{i=1}^g p_i^{e_i}\left(1 - p_i^{-1}\right) \\
&= \prod_{i=1}^g p_i^{e_i} \cdot \prod_{i=1}^g (1 - 1/p_i) \\
&= n \prod_{i=1}^g (1 - 1/p_i)
\end{aligned}$$

□

If $n = p_1^{e_1} \cdots p_g^{e_g}$ where p_i's are distinct primes, then from the above, we have two different ways to write $\phi(n)$:

$$\phi(n) = \prod_{i=1}^{g} p_i^{e_i - 1}(p_i - 1) = n \prod_{i=1}^{g} \left(1 - \frac{1}{p_i}\right)$$

where the first expression uses (a) and (b) from the proposition.

Let's compute $\phi(10)$. Note that $10 = 2 \cdot 5$ and $\gcd(2, 5) = 1$. Therefore using (a) of the above theorem we get

$$\phi(10) = \phi(2^1 \cdot 5^1) = \phi(2^1) \cdot \phi(5^1)$$

since $\gcd(2^1, 5^1) = 1$. Using (b) of the above theorem I get

$$\phi(10) = \phi(2^1) \cdot \phi(5^1) = (2^1 - 2^{1-1}) \cdot (5^1 - 5^{1-1}) = 1 \cdot 4 = 4$$

Of course you can also use (c) above to get

$$\phi(10) = 10 \cdot (1 - 1/2) \cdot (1 - 1/5) = 10 \cdot \frac{1}{2} \cdot \frac{4}{5} = 4$$

You should also let a run through all the values $\{0, 1, 2, ..., 9\}$ and check if a has an inverse. You'll see that invertible a's are $1, 3, 7, 9$ with inverses $1, 7, 3, 9$ respectively. In other words, there are exactly 4 invertible elements in $\mathbb{Z}/10$ and $\mathbb{Z}/4^\times = \{1, 3, 7, 9\}$.

Exercise 202.11.1.

1. Compute $\phi(735)$.
2. Compute $\phi(900)$.
3. Compute $\phi(263891)$.

Exercise 202.11.2. 1. Let p be a prime. What is $\phi(2p)$ in terms of p?

2. How many solutions are there to $\phi(n) = 2n$?
3. How many solutions are there to $\phi(n) = n/2$?

Exercise 202.11.3. Easy: What is $\phi(pq)$ as an integer expression involving p and q? Can you write it as an expression involving the sum and product of p and q? (i.e., besides constants and operators, your expression contains only $p + q$ and pq).

Exercise 202.11.4.

(a) Solve $\phi(n) = 2$, i.e., find all positive integers n such that $\phi(n) = 2$. (Hint: Write down the prime factorization of $n = p_1^{e_1} \cdots p_g^{e_g}$ and use the equation $\phi(n) = 2$.)

(b) Solve $\phi(n) = 3$.

(b) Solve $\phi(n) = 6$.

Exercise 202.11.5. Prove that

$$\phi(mn) = \phi(m)\phi(n) \cdot \frac{g}{\phi(g)}$$

where $g = \gcd(m, n)$. (Note that the above does *not* assume m, n are coprime. What is the above if m, n are coprime?)

Exercise 202.11.6. * Plot a function of the graph $y = \phi(x)$ for integer values of x running through 1 to 10000. See any pattern? Note that if you want an approximation (for instance in the asymptotics) that's not too difficult.

But note this. Therefore two ways to compute $\phi(n)$:

1. $\phi(n) = |\{x \mid 0 \leq x < n-1, \ \gcd(x, n) = 1\}|$ which required Euclidean algorithm.
2. $\phi(n) = \phi(p_1^{\alpha_1} \cdots p_g^{\alpha_g}) = (p_1^{\alpha_1} - p_1^{\alpha_1 - 1}) \cdots (p_1^{\alpha_g} - p_1^{\alpha_g - 1})$

The first method is slow: you need to loop your x and for each x you need to execute the EEA which has a loop. The second method is fast only if you can find the prime factorization of n. If is possible to find $\phi(n)$ as a formula in n without finding the prime factorization of n?

Exercise 202.11.7. * Can you find an n such that $\phi(n)$ divides $n + 1$?

Exercise 202.11.8. *

1. If p is a prime, then $\phi(p) = p - 1$. Prove that if $\phi(n) = n - 1$, then n is a prime.
2. Can you find an $n > 1$ which is not a prime (i.e. composite) and such that $\phi(n)$ divides $n - 1$? If you can find one, let me know ASAP. Or if you can prove that such as n does not exist, let me know ASAP.

Exercise 202.11.9. *

1. Can you find some n such that
$$\phi(\phi(n)) = 1$$
2. Write $\phi^2(n) = \phi(\phi(n))$. Can you find *all* n such that $\phi^2(n) = 1$.
3. Write ϕ^k to the composition of k Euler ϕ. What about $\phi^3(n) = 1$? Can you find some n satisfying the above equation?
4. What about $\phi^k(2^n) = 1$? What is the smallest k for $\phi^k(2^n) = 1$?
5. What about $\phi^k(2^m \cdot 3^n) = 1$? What is the smallest k such that $\phi^k(2^m \cdot 3^n) = 1$?

As an aside, note that ϕ as a function has domain of $\mathbb{N}$. In this case, we say that ϕ is an **arithmetic function**. Furthermore, ϕ satisfies the property that if $\gcd(m, n) = 1$, then $\phi(mn) = \phi(m)\phi(n)$. A function $\mathbb{N} \to \mathbb{C}$ satisfying this property is said to be **multiplicative** . The Euler ϕ function is one of many multiplicative arithmetic functions. Multiplicative functions are extremely important in number theory.

arithmetic function

multiplicative

File: affine-cipher.tex

202.12 Affine Cipher

Recall the encryption and decryption of the affine cipher looks like

$$E_{a,b}(x) \equiv (ax + b) \pmod{26}, \qquad D_{a,b}(x) \equiv a^{-1}(x - b) \pmod{26}$$

Note that the key is (a, b). Note also that a must be invertible mod 26.

Therefore (by the multiplication principle in discrete mathematics), the total numbers of keys is

$$\phi(26) \cdot 26 = 312$$

This is not that big, but it's definitely bigger than the number of keys for the shift cipher (which is 26). This means that to carry out a brute force attack on an affine cipher, assume the attacker has the cipher, he/she must try 312 possible keys.

Exercise 202.12.1. This is easy: Show that $\phi(26) \cdot 26 = 312$.

File: fermat-little-theorem.tex

202.13 Fermat's Little Theorem

I've already talked about Fermat's Little Theorem. Let's do it again. Because the theorem is important and also because I want to give you a different proof that does not involve group theory.

Theorem 202.13.1. *(Fermat's Little Theorem). Let p is a prime number and a be a positive integer not divisible by p. Then*

$$a^{p-1} \equiv 1 \pmod{p}$$

Exercise 202.13.1. Compute r where r is the smallest positive integer satisfying

$$5^{642} \equiv r \pmod{641}$$

[Hint: 641 is prime.]

Corollary 202.13.1. *Let p be a prime. Then $a^p \equiv a \pmod{p}$.*

Note that the corollary does not require $p \nmid a$. The proof of the corollary is easy. If $p \mid a$, then both sides of the equation is 0 mod p, so the congruence is true. If $p \nmid a$, then Fermat's Little Theorem gives us

$$a^{p-1} \equiv a \quad (\text{mod } p)$$

on multiplying both sides by a, we get

$$a^p \equiv a \quad (\text{mod } p)$$

OK. Now let's prove Fermat's Little Theorem. Let $a > 0$ be a positive integer, p be a prime and suppose p does not divide a. Look at the set $X = \{1, 2, \ldots, p-1\}$ of nonzero remainders mod p. Let me define a function $f : X \to X$ by $f(x) = ax \pmod{n}$. Note ax (mod n is not zero and therefore is in X (X does not have 0 remember?).

What do I mean by this? Suppose $p = 5$ and $a = 2$. The function f will behave as following:

$$\begin{aligned}
f(1) &= 2 \cdot 1 = 2 \\
f(2) &= 2 \cdot 2 = 4 \\
f(3) &= 2 \cdot 3 = 6 \equiv 1 \pmod{5} \\
f(4) &= 2 \cdot 4 = 8 \equiv 3 \pmod{5}
\end{aligned}$$

Got it? You notice that for this case f is a bijective function and hence must have an inverse. Now back to the proof.

Now note that $f(0) = 0$. Remove 0 from the domain and range of f. I claim that f is a bijective function on X. Since p is a prime and and p does not divide a, $\gcd(p, a) = 1$. Right? Therefore a must be invertible mod p. Suppose b in the inverse of a mod p. Now define function g by $g(x) = bx$ in the same manner (reducing result mod p). You can show that this g is the inverse of f, i.e., $f(g(x)) = x$ and $g(f(x)) = x$.

So what? Look at $X = \{1, 2, \ldots, p-1\}$. Apply f to this collection to get $\{f(1), \ldots, f(p-1)\}$. Of course this is just $\{a, 2a, 3a, (p-1)a\}$. You also know from the above that this new set is just a permutation of the original. Of course we have actually taken mod p of some of the integers $a, 2a, 3a, \ldots, (p-1)a$. But anyway the two sets are the same integers mod p. So let's multiply them separate. The result numbers must be the same mod p:

$$1 \cdot 2 \cdot \cdots \cdot (p-1) \equiv a \cdot 2a \cdot \cdots \cdot (p-1)a \pmod{p}$$

which is the same as

$$(p-1)! \equiv (p-1)! \cdot a^{p-1} \pmod{p}$$

This is the formal proof of the above theorem.

Proof of Fermat's Little Theorem. Let $X = \{1, 2, \ldots, , p-1\}$. Define the $f(x)$ to the remainder of ax when divided by n. Note that

$$ax \not\equiv 0 \pmod{p}$$

otherwise $p|ax$ would imply $p|a$ or $p|x$. By our assumption, p does not divide a. Furthermore $1 \leq x \leq p-1$ and hence is not divisible by p since p is a prime. Therefore $f(x)$ is in X. Hence f is a well-defined function. Furthermore since

$\gcd(a, n) = 1$, a is invertible mod p. Suppose $1 \leq b \leq p - 1$ is an inverse of a. Define the function $g : X \to X$ so that $g(x)$ is the remainder of bx when divided by n. For the same reason g is a well-defined function. Furthermore

$$f(g(x)) \equiv abx \equiv x \pmod{n}$$
$$g(f(x)) \equiv bax \equiv x \pmod{n}$$

Therefore g is the inverse of f. In particular f is a bijective function. In other words f is a permutation on $1, 2, \ldots, p - 1$. Therefore $\{1, 2, \ldots, p - 1\}$ and $\{a \bmod p, 2a \bmod p, \ldots, (p-1)a \bmod p\}$ are the same sets. Multiplying the elements of each set give us the equation:

$$\prod_{x \in X} x \equiv \prod_{x \in X} (ax) \pmod{p}$$

Since $|X| = p - 1$, a appears $p - 1$ times on the left side of the congruence. Hence

$$(*) : \prod_{x \in X} x \equiv a^{p-1} \prod_{x \in X} x \pmod{p}$$

Now note that for each $x \in X$, since $1 \leq x \leq p - 1$, we have $\gcd(p, x) = 1$. Therefore x in invertible mod p. Hence x^{-1} mod p exists. Note that $\prod_{x \in X} x^{-1}$ is the inverse of $\prod_{x \in X} x$ mod p since

$$\prod_{x \in X} x^{-1} \cdot \prod_{x \in X} x \equiv \prod_{x \in X} (x^{-1}x) \equiv \prod_{x \in X} 1 \equiv 1 \pmod{p}$$

Therefore multiplying equation $(*)$ with $\prod_{x \in X} x^{-1}$ gives

$$1 \equiv a^{p-1} \pmod{p}$$

QED.

Exercise 202.13.2. What is the remainder of $3^{122436481}$ mod 13?

Exercise 202.13.3. Leetcode 372.
https://leetcode.com/problems/super-pow/
Your task is to calculate `ab mod 1337` where `a` is a positive integer and `b` is an extremely large positive integer given in the form of an array.

Exercise 202.13.4. Leetcode 1622 https://leetcode.com/problems/fancy-sequence/

Write an API that generates fancy sequences using the append, addAll, and multAll operations.

Implement the Fancy class:

- `Fancy()` Initializes the object with an empty sequence.
- `void append(val)` Appends an integer val to the end of the sequence.
- `void addAll(inc)` Increments all existing values in the sequence by an integer inc.
- `void multAll(m)` Multiplies all existing values in the sequence by an integer `m`.
- `int getIndex(idx)` Gets the current value at index `idx` (0-indexed) of the sequence modulo 109 + 7. If the index is greater or equal than the length of the sequence, return `-1`.

Exercise 202.13.5. Leetcode 1952
https://leetcode.com/problems/three-divisors/
Given an integer `n`, return `true` if `n` has exactly three positive divisors. Otherwise, return `false`.

An integer `m` is a divisor of `n` if there exists an integer `k` such that `n = k * m`.

File: euler-theorem.tex

202.14 Euler's Theorem

Note that Fermat's Little Theorem can be used to compute powers very rapidly if you work in mod p where p is a prime. What if you need to work in mod N where N is not a prime? There is a generalization of Fermat's Little Theorem due to Euler. Note that since $p - 1 = \phi(p)$, Fermat's Little Theorem

$$a^{p-1} \equiv 1 \pmod{p}$$

can be stated as

$$a^{\phi(p)} \equiv 1 \pmod{p}$$

This statement actually holds if p is replaced by any positive integer.

Theorem 202.14.1. *(Euler). Let a and n be positive integers. Then*

$$a^{\phi(n)} \equiv 1 \pmod{n}$$

The proof is just a generalization of the proof of Fermat's Little Theorem. Suppose X is the subset of $\{1, \ldots, n\}$ containing only the integers which are invertible mod n. (Of course $|X| = \phi(n)$.). You then consider $Y = \{f(x) \,|\, x \in X\}$ where $f(x) = ax \pmod{n}$. Since $\gcd(a, n) = 1$, again a is invertible mod n. Suppose b is the inverse of a mod n. Then $g(x) = bx \pmod{n}$ is the inverse function of f. So f is a permutation on X. Multiplying the integers in X and then the integers in Y you get

$$\prod_{x \in X} x \equiv \prod_{y \in Y} y \quad (\text{mod } n)$$

Each integer in Y is ax of some x in X. Therefore

$$\prod_{y \in Y} \equiv a^{|X|} \prod_{x \in X} x \pmod{n}$$

AHA! ... note that $|X| = \phi(n)$ ... ! Altogether we get

$$\prod_{x \in X} x \equiv a^{\phi(n)} \prod_{x \in X} \pmod{n}$$

Now note that every integer in X is invertible mod n. Therefore $\prod_{x \in X} x$ is also invertible. Right? So now hitting the above equation with $\left(\prod_{x \in X} x \right)^{-1}$ we finally get

$$1 \equiv a^{\phi(n)} \pmod{n}$$

DONE!

Make sure you step back and take in the whole proof and make your head one size larger. Look at the place where you used $\gcd(a, n) = 1$.

Exercise 202.14.1. Compute r where r is the smallest positive integer satisfying

$$5^{642} \equiv r \pmod{640}$$

[Hint: Use Euler's Theorem (duh).]

Exercise 202.14.2. What is the remainder of $3^{123456789}$ mod 100?

Exercise 202.14.3. What is the hundreds digit of $3^{123456789}$?

Exercise 202.14.4. Leetcode 372.
https://leetcode.com/problems/super-pow/
Your task is to calculate $a^b \pmod{1337}$ where a is a positive integer and b is an extremely large positive integer given in the form of an array. For instance for $a = 2, b = [1, 0]$, the output is 1024.

Exercise 202.14.5. Leetcode 1015.
https://leetcode.com/problems/smallest-integer-divisible-by-k/
Given a positive integer k, you need to find the length of the smallest positive integer n such that n is divisible by k, and n only contains the digit 1. Return the length of n. If there is no such n, return -1.

Exercise 202.14.6. Leetcode 204.
https://leetcode.com/problems/count-primes/description/

Given an integer n, return the number of prime numbers that are strictly less than n.

Exercise 202.14.7. https://acm.timus.ru/problem.aspx?space=1&num=1673 At the end of the previous semester the students of the Department of Mathematics and Mechanics of the Yekaterinozavodsk State University had to take an exam in network technologies. N professors discussed the curriculum and decided that there would be exactly N^2 labs, the first professor would hold labs with numbers $1, N+1, 2N+1, ..., N2-N+1$, the second one — labs with numbers $2, N+2, 2N+2, ..., N2-N+2$, etc. N-th professor would hold labs with numbers $N, 2N, 3N, ..., N2$. The professors remembered that during the last years lazy students didn't attend labs and as a result got bad marks at the exam. So they decided that a student would be admitted to the exam only if he would attend at least one lab of each professor. N roommates didn't know the number of labs and professors in this semester. These students had different diligence: the first student attended all labs, the second one — only labs which numbers were a multiple of two, the third one — only labs which numbers were a multiple of three, etc… At the end of the semester it turned out that only K of these students were admitted to the exam. Find the minimal N which makes that possible.

Input: An integer K ($1 \leq K \leq 2 \cdot 10^9$).

Output: Output the minimal possible N which satisfies the problem statement. If there is no N for which exactly K students would be admitted to the exam, output 0.

Example: Input:8, output:15. Input:3, output:0.

Index