Abstract Algebra Lecture Notes

David M. McClendon

Department of Mathematics
Ferris State University

Fall 2018 edition

Contents

Chapter 0

Review of Math 324

Math 324 is an introduction to mathematical language: that means sets, relations and functions. It is also an introduction to proof. In the first four sections of this chapter, we review the language (and general theorems about that language you should know) that was developed in Math 324; in the last section, we discuss general methods to prove statements.

Before we get started, here are some symbols and an abbreviation I use often:

$\forall$ means "for all"
$\exists$ means "there exists"
$\wedge$ means "and"
$\vee$ means "or"
$\sim$ means "not"
s.t. is short for "such that"

0.1 Sets

The fundamental objects of mathematics are called **sets**. A set is really just a list of objects (in math, the objects are usually numbers, or vectors, or functions).

Definition 0.1 *A* **set** *is a definable collection of objects. The objects which comprise a set are called the set's* **elements**. *If x is an element of set E, we write $x \in E$; if x is not an element of set E, we write $x \notin E$.*

Examples of sets (observe that sets are usually denoted by capital letters):

$$A = \{3, 5, 7, 9, 11\}$$
$$B = \{1, 2, 3, 4, 5, 6\}$$
$$C = \{3, 5, 7\}$$

The elements of set C described above are 3, 5 and 7. For the set A above, $3 \in A$ and $5 \in A$ but $8 \notin A$.

We often define a set without listing the elements (using English language). For example, the sets A, B and C given above could be described, respectively, by saying

"let A be the set of odd numbers from 3 to 11";
"let B be the set of integers from 1 to 6";
"let C be the set of odd numbers from 3 to 7".

We also describe sets by using what is called **set-builder notation**: to describe the same sets A, B, C as above using set-builder notation, we would write (or say)

$$A = \{x : 3 \leq x \leq 11 \text{ and } x \text{ is odd}\}$$
$$B = \{x : 1 \leq x \leq 6 \text{ and } x \text{ is an integer}\}$$
$$C = \{x : 3 \leq x \leq 7 \text{ and } x \text{ is odd}\}.$$

The first statement above is interpreted as follows: it says that set A is equal to the set of numbers x such that (the colon means "such that" in mathematics) $3 \leq x \leq 11$ and x is odd. Notice that this is exactly the set $\{3, 5, 7, 9, 11\}$.

To show you a different kind of example: if you were defining some set of functions (instead of a set of numbers), then instead of x you'd write f, and then after the colon you'd describe what has to be true about f for the function f to be in the set. For example, the set D of functions whose derivative at $x = 2$ is positive could be described by writing

$$D = \{f : f'(2) > 0\}.$$

For this set D, it would be valid to say that if $g(x) = x^3$, then $g \in D$ (because $g'(2) = 3(2^2) = 12 > 0$) but if $h(x) = 3 - 4x$, then $h \notin D$ (because $h'(2) = -4 \leq 0$).

Definition 0.2 *The* **empty set**, *denoted* $\emptyset$, *is the set with no elements.*

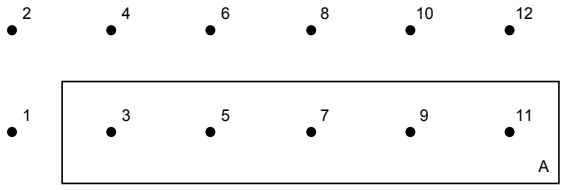
Definition 0.3 *Let E be a set. The* **power set** *of E, denoted 2^E or $\mathcal{P}(E)$, is the set of all subsets of E.*

The elements of a power set are themselves sets.

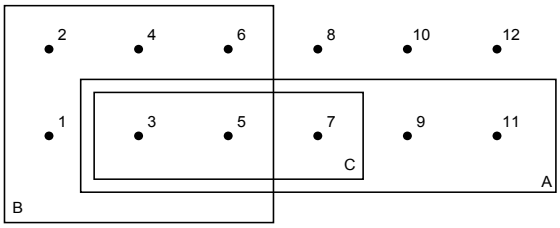
Example: if $E = \{1, 2\}$, then $2^E = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.

Venn diagrams

A useful way to think about sets is to draw pictures called **Venn diagrams**. To draw a Venn diagram, represent each set you're thinking about by a circle (or an oval, or a square, or a rectangle, or some other shape); think of an object as being an element of the set if and only if it is inside the shape corresponding to the set. For example, a Venn diagram for the set A described above (recall that $A = \{3, 5, 7, 9, 11\}$) would be given by something like



because the box describing A contains exactly the elements of A (nothing more and nothing less). Similarly, a Venn diagram representing the sets A, B and C from above at the same time would be something like



Subsets and set equality

Definition 0.4 *Let E and F be two sets.*

- *We say E is a* **subset** *of F, and write $E \subseteq F$, if $\forall\, x, x \in E \Rightarrow x \in F$.*
  If E is not a subset of F, we write $E \not\subseteq F$.
  If $E \subseteq F$, we also write $F \supseteq E$ and say that F is a **superset** *of E.*
- *We say E and F are* **equal**, *and write $E = F$, if $E \subseteq F$ and $F \subseteq E$.*
  If E and F are not equal, we write $E \neq F$.

Example: $\{0, 1, 2\} \subseteq \{0, 1, 2, 4, 8\}$ but $\{0, 1, 2\} \not\subseteq \{0, 2, 4\}$.

Note the difference between the symbols $\in$ and $\subseteq$: the first symbol should be preceded by an element, but the second symbol should be preceded by a set.

Generally speaking, to say $E \subseteq F$ means "everything in E also is in F" or "E is inside F". If you draw a Venn diagram, to say $S \subseteq T$ means that the shape corresponding to set S is completely inside the shape corresponding to set T.

For example, for the sets A and C given in the preceding section, $C \subseteq A$ since every element of C is also in A.

To say two sets are equal means that they contain exactly the same elements.

Example: $\{n \in \mathbb{N} : n \text{ is an even prime}\} = \{n \in \mathbb{N} : n - 2 = 0\}$ because the only element in each set is 2.

Operations on sets

Definition 0.5 *Let E and F be sets.*

- *The* **union** *of E and F, denoted $E \cup F$, is defined as follows:*

$$E \cup F = \{x : x \in E \text{ or } x \in F\}.$$

- *The* **intersection** *of E and F, denoted $E \cap F$, is defined as follows:*

$$E \cap F = \{x : x \in E \text{ and } x \in F\}.$$

Definition 0.6 *Let E and F be sets.*

- *We say E and F are* **disjoint** *if $E \cap F = \emptyset$.*
- *The* **complement** *of E, denoted E^C, is the set*

$$E^C = \{x : x \notin E\}.$$

- *The* **difference** *of E and F, denoted $E - F$ and read "E minus F", is the set $E - F = E \cap F^C$.*
- *The* **symmetric difference** *of E and F, denoted $E \triangle F$, is the set $E \triangle F = (E - F) \cup (F - E)$.*
- *The* **(Cartesian) product** *of E and F, denoted $E \times F$ and read "E cross F", is the set*

$$E \times F = \{(x, y) : x \in E \text{ and } y \in F\}.$$

  The elements of a Cartesian product are called **ordered pairs**. *We denote $E \times E$ by E^2.*

- *Let E be a set and let n be a positive integer. The n^{th}* **Cartesian power** *of E, denoted E^n, is the set of* **ordered n-tuples** *of elements from E:*

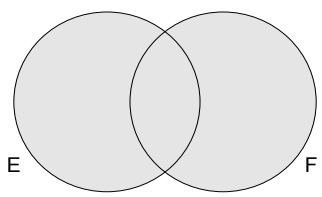
$$E^n = \{(x_1, x_2, ..., x_n) : \forall j, x_j \in E\}.$$

Concepts:

- $\cup$ is set language for "or"–the union of a bunch of sets is the set consisting of elements belonging to at least one of the sets. For example, using the sets described earlier,

$$A \cup B = \{1, 2, 3, 4, 5, 6, 7, 9, 11\}$$

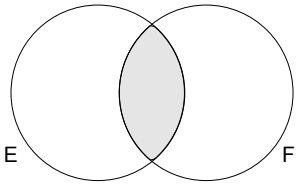
  because all of the numbers listed there either are in A, or in B, or both. In terms of Venn diagrams, the union is usually thought of as a "MasterCard-symbol" shaped region that encompasses the sets whose union you are taking. For example, the union of E and F in the figure below is exactly the shaded region (throughout these Venn diagrams, the set E is just the circle on the left; F is the circle on the right):



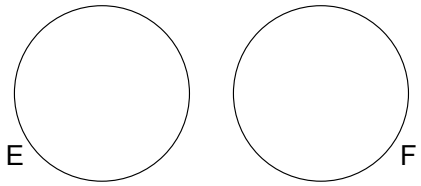
- $\cap$ is set language for "and"–the intersection of a bunch of sets is the set consisting of elements which belong to all of the given sets. For example, using the sets described A and B above,

$$A \cap B = \{3, 5\}$$

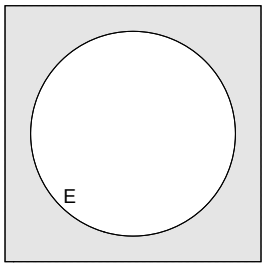
because the only numbers lying in both A and B are 3 and 5. In terms of Venn diagrams, the intersection of two sets is the overlap of the shapes representing the sets (below, the shaded region is $E \cap F$):



- Sets are disjoint if there are no objects which are both elements of S and elements of T. If you drew a Venn diagram with sets S and T where S and T are disjoint, then the shapes corresponding to S and T should not overlap: the Venn diagram would look like this:

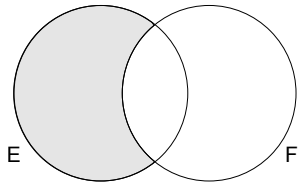


- Complement is set language for "not". In terms of Venn diagrams, the complement of E is the set of things outside the shape representing E (the shaded region in the picture below):

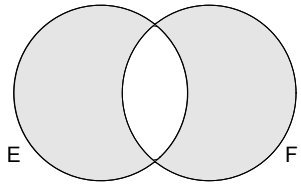


Complements have to be taken with respect to a universe of discourse U, which is usually understood without being stated.

- Difference means "in the first, but not the second". In this Venn diagram, $E - F$ is shaded:

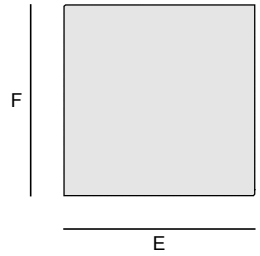


- Symmetric difference means "in one or the other, but not both". In this Venn diagram, $E \triangle F$ is shaded:



- The Cartesian product of two sets is the set of ordered pairs, where the first element comes from the first set and the second element comes from the second set. If (a, b) and (x, y) be ordered pairs. To say $(a, b) = (x, y)$ means $a = x$ and $b = y$. This means in particular that (unless $a = b$), $(a, b) \neq (b, a)$. Similarly, unless $E = F$, $E \times F$ is not equal to $F \times E$.

  Cartesian products are usually pictured like this (E is the line across the bottom, F is the line up the left-hand side, and $E \times F$ is the set of points in the square):



Sums of sets; multiples of sets

Suppose you have two sets (say S and T) where the elements of these objects are objects that can be added to one another and multiplied by constants, like numbers or vectors.

Define a new set, called the **sum** of S and T and denoted $S + T$, to be the set of all objects which are the sum of some element of S and some element of T. In set-builder notation, this means

$$S + T = \{s + t : s \in S, t \in T\}.$$

Also define the set $S - T$ to be the set of all objects which can be written as some element of S minus some element of T. In set-builder notation, this means

$$S - T = \{s - t : s \in S, t \in T\}.$$

If x is an element, we write $S + x$ and $S - x$ when we mean $S + \{x\}$ and $S - \{x\}$, respectively.

Given set S and real number c, define the set cS to be the set of all objects which are c times something in S.

Here are some examples, where E is given as $\{0, 2, 4\}$ and F is given as $\{0, 15\}$:

$5E$	$= \{0, 10, 20\}$	(obtained by multiplying each element in E by 5)
$E + F$	$= \{0, 2, 4, 15, 17, 19\}$	(obtained by adding each element in E to each element of F)
$-F$	$= \{0, -15\}$	(obtained by multiplying each element in F by -1)
$E + 3$	$= \{3, 5, 7\}$	(obtained by adding 3 to each element in E).

As a further example, we denote the integers $\{..., -3, -2, -1, 0, 1, 2, 3, ...\}$ by $\mathbb{Z}$. Therefore $7\mathbb{Z} = \{7n : n \in \mathbb{Z}\} = \{..., -21, -14, -7, 0, 7, 14, 21, ...\}$.

0.2 Relations

A *relation* is a subset of a Cartesian product space:

Definition 0.7 *Let E be a set. A* **relation on** *E is a subset R of E^2. If $(x, y) \in E^2$ is such that $(x, y) \in R$, we write $x\,R\,y$ (and say "x is related to y").*

There are two fundamentally important examples of relations: $\leq$ and $=$. Formally, the relation $\leq$ is the set $\{(x, y) : x \leq y\}$ (which can be thought of as a subset of $\mathbb{N}^2$ or $\mathbb{R}^2$, etc.), and the relation $=$ is the set $\{(x, x)\}$ (which can be thought of as a subset of E^2 for any set E). These two relations generalize, respectively, to *order relations* and *equivalence relations*.

Order relations

Definition 0.8 *Let R be a relation on a set E. We say R is a* **partial order** *on E if R has the following three properties:*

Reflexivity: $\forall\, x \in E$, $(x, x) \in R$.

Antisymmetry: *if $x, y \in E$ are such that $(x, y) \in R$ and $(y, x) \in R$, then $x = y$.*

Transitivity: *if $x, y \in E$ are such that $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$.*

A pair (E, R), where R is a partial order on E, is called a **partially ordered set**, *or* **poset** *for short.*

If, in addition to being a partial ordering, the relation R has the property that for any $x, y \in E$, either xRy or yRx, then we say R is a **total ordering** *on E.*

The prototype example of a total ordering is $\leq$ (on the real numbers), which is reflexive because $x \leq x$; it is antisymmetric because $x \leq y$ and $y \leq x$ certainly implies that $x = y$; it is transitive because $x \leq y$ and $y \leq z$ certainly implies $x \leq z$; and for any two real numbers x and y, either $x \leq y$ or $y \leq x$.

Definition 0.9 *Let R be a total ordering on E. We say R is a* **well ordering** *(or that E is* **well ordered** *(by R)) if every nonempty subset $A \subseteq E$ contains a smallest element, i.e.*

$$\exists\, x \in A \text{ s.t. } (y \in A \Rightarrow xRy)$$

Equivalence relations

Definition 0.10 *Let R be a relation on E.*

- *R is called* **reflexive** *if $\forall\, x \in E$, $x\, R\, x$.*
- *R is called* **symmetric** *if $\forall\, x, y \in E$, $x\, R\, y$ implies $y\, R\, x$.*
- *R is called* **transitive** *if $\forall\, x, y, z \in E$, ($x\, R\, y$ and $y\, R\, z$) implies $x\, R\, z$.*

If R is reflexive, symmetric and transitive, then R is called an **equivalence relation (on E)**.

Here is a fundamental axiom of mathematics (recall that an axiom is a statement that we accept without proof):

Axiom 0.11 (Axiom of equality) *Let E be any set. Then equality on E is an equivalence relation.*

Definition 0.12 *Suppose R is an equivalence relation on E. For each $x \in E$, define the set*

$$[x] = [x]_R = \{y \in E : x\, R\, y\}.$$

*This set is called the R-***equivalence class** *of x (or just the* **equivalence class** *of x). Any subset of E which is of the form $[x]_R$ for some $x \in E$ is called an* **equivalence class (of R)**.

As another example of an equivalence relation, let $E = \mathbb{R}^2$ and decree two points in E to be equivalent if they have the same x-coordinate. Then the equivalence class of a point $(x_0, y_0) \in E$ is the vertical line $x = x_0$ (this line is the set of all points which are equivalent to (x_0, y_0).

Theorem 0.13 *If R is an equivalence relation on E, then the R-equivalence classes form a partition of E (i.e. the union of the equivalence classes is all of E, and any two different equivalence classes are disjoint).*

0.3 Functions

In Math 324, we learn the following technical definition of a function which makes precise the idea of a "function" that you first encounter in high-school algebra or precalculus (generally speaking, this technical definition isn't useful):

Definition 0.14 *Let A and B be sets. A* **function**, *a.k.a.* **map** *f from A to B is a subset of $A \times B$ with the following property:*

$$\text{if } (x,y) \in f \text{ and } (x,z) \in f, \text{ then } y = z.$$

If $x \in A$ is such that $\exists\, y \in B$ with $(x,y) \in f$, then by the above hypothesis, this is the only y such that $(x,y) \in f$. In this situation, we write $y = f(x)$ (or just $y = f\,x$) and we say that y is the **value of** *f* **at** *x, or the* **image of** *x* **under** *f. The notation $f : A \to B$ means that f is a function from A to B.*

If $y = f(x)$, we think of x as being an input and y the corresponding output of the function, and f being a "procedure" that produces the y from the x. Thus if $f : A \to B$, A is the set of possible inputs to f, and B is the set of possible outputs to f. Values of a function are actual outputs of the function.

Definition 0.15 *Let $f : A \to B$.*

- *The* **domain** *of f, defined $Dom(f)$, is the set of inputs at which f has a value:*

$$Dom(f) = \{x \in A : \exists y \in B \text{ s.t. } (x,y) \in f\}.$$

- *The* **codomain** *of f is B.*
- *The* **range** *of f, a.k.a.* **image** *of f, denoted $Range(f)$ or $Im(f)$, is the set of the function's values:*

$$Range(f) = Im(f) = \{y \in B : \exists x \in A \text{ s.t. } f(x) = y\}.$$

- *A* **rule** *for f is a procedure or a formula which specifies how to determine $f(x)$ from each $x \in Dom(f)$.*

Example: Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x^2$. Then:
- the domain of f is $\mathbb{R}$;
- the codomain of f is $\mathbb{R}$;
- the range of f is $[0, \infty)$.

Example: Let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = \frac{1}{x}$. Then:
- the domain of f is $\mathbb{R} - \{0\}$;
- the codomain of f is $\mathbb{R}$;
- the range of f is $\mathbb{R} - \{0\}$.

Definition 0.16 (Equality of functions) *To say two functions $f : A \to B$ and $g : C \to D$ are* **equal** *(denoted $f = g$) means that $Dom(f) = Dom(g)$ and for all $x \in Dom(f)$, $f(x) = g(x)$.*

Definition 0.17 *Let $f : A \to B$.*

- *Given $E \subseteq A$, the* **image of E under f**, *denoted $f(E)$, is the set*

$$f(E) = \{y \in B : \exists x \in E \text{ s.t. } y = f(x).\}$$

- *Given $E \subseteq B$, the* **preimage (of E under f)**, *also called the* **inverse image (of E under f)**, *denoted $f^{-1}(E)$, is the set*

$$f^{-1}(E) = \{x \in A : f(x) \in E\}.$$

- *Given $y \in B$, the* **preimage (of y under f)**, *also called the* **inverse image (of y under f)**, *denoted $f^{-1}(y)$, is the set defined by*

$$f^{-1}(y) = f^{-1}(\{y\}) = \{x \in A : f(x) = y\}.$$

To emphasize, the preimage of a point is a set. As an example, let $f : \mathbb{R} \to \mathbb{R}$ be $f(x) = x^2$. Then $f^{-1}(25) = \{-5, 5\}$ since both 5 and -5 map to 25 under f.

Theorem 0.18 *Let $f : A \to B$ be a function. Then:*

- *For any set $E \subseteq A$, $f^{-1}(f(E)) \supseteq E$.*
  ***WARNING:** in general, $f^{-1}(f(E)) \neq E$.*
- *For any set $E \subseteq B$, $f(f^{-1}(E)) = E \cap Im(f)$.*
  ***WARNING:** in general, $f(f^{-1}(E)) \neq E$..*

Definition 0.19 *Given any set E, the* **identity function** *$I_E : E \to E$ is the function defined by $I_E(x) = x$.*

Compositions

Definition 0.20 *Let $g : A \to B$ and let $f : B \to C$. Define the* **composition** *of f with g, denoted $f \circ g$, to be the function from A to C defined by the rule*

$$(f \circ g)(x) = f(g(x)).$$

Example: If $f : \mathbb{R}^2 \to \mathbb{R}$ is $f(x, y) = x^2 - y$ and $g : \mathbb{R} \to \mathbb{R}^2$ is $g(t) = (t-2, 4t+3)$, then $f \circ g : \mathbb{R} \to \mathbb{R}$ has rule

$$(f \circ g)(t) = f(g(t)) = f(t-2, 4t+3) = (t-2)^2 - (4t+3).$$

Theorem 0.21 (Properties of compositions) *Let $h : A \to B$, $g : B \to C$ and $f : C \to D$. Then:*

1. Domain of a composition: $Dom(f \circ g) = g^{-1}(Dom(f))$;

2. Preimages under composition: *for any $E \subseteq D$,* $(f \circ g)^{-1}(E) = g^{-1}(f^{-1}(E))$.

3. Composition is associative: $(f \circ g) \circ h = f \circ (g \circ h)$.

4. Composition with identity function: $f \circ I_C = f$ *and* $I_D \circ f = f$.

Injectivity

An *injection* (a.k.a. a $1-1$ *function*) is a function which takes different inputs to different outputs. More precisely:

Definition 0.22 *A function $f : A \to B$ is called* **injective**, *a.k.a.* **one-to-one**, *a.k.a. $1-1$, if for every $x, y \in A$, $f(x) = f(y)$ implies $x = y$. If $f : A \to B$ is injective, we write $f : A \hookrightarrow B$.*

Equivalent characterizations of injectivity:
1. $f(x) = f(y)$ implies $x = y$.
2. $x \neq y$ implies $f(x) \neq f(y)$.
3. Different inputs go to different outputs.
4. f passes the Horizontal Line Test (in the situation where $f : \mathbb{R} \to \mathbb{R}$).

Example: $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^2$ is <u>not</u> injective because $f(1) = f(-1) = 1$.

Example: $f : [0, \infty) \to \mathbb{R}$ given by $f(x) = x^2$ <u>is</u> injective because for any $y \in \mathbb{R}$, there is at most one x in $[0, \infty)$ such that $f(x) = x^2 = y$.

Theorem 0.23 *Let $f : B \to C$ and $g : A \to B$ be functions.*

- *If f and g are both injective, then $f \circ g$ is injective.*
- *If $f \circ g$ is injective, then g is injective.*

Surjectivity

An *surjection* (a.k.a. an *onto function*) is a function which "hits" every point in its codomain. More precisely:

Definition 0.24 *A function $f : A \to B$ is called* **surjective**, *a.k.a.* **onto**, *if $f(A) = B$. If f is surjective, we write $f : A \twoheadrightarrow B$.*

The noun form of "surjective" is **surjection**.

Equivalent characterizations of surjectivity:
1. $Im(f) = f(A) = B$.
2. $B \subseteq Im(f)$.
3. Every potential output of f is an actual output.
4. For every $y \in B$, there is an $x \in A$ such that $f(x) = y$.

Example: $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^2$ is not onto, since $-1 \notin f(\mathbb{R})$.

Example: $f : \mathbb{R} \to [0, \infty)$ where $f(x) = x^2$ is onto, because for every $y \in [0, \infty)$, we can let $x = \sqrt{y}$. Then $f(x) = y$.

Theorem 0.25 *Let $f : B \to C$ and $g : A \to B$ be functions.*

- *If f and g are both surjective, then $f \circ g$ is surjective.*
- *If $f \circ g$ is surjective, then f is surjective.*

Bijectivity and inverse functions

A function which is both $1-1$ and onto is called a *bijection*:

Definition 0.26 *A function $f : A \to B$ is called* **bijective** *if f is both injective and surjective. In this situation, you can write $f : A \leftrightarrow B$, but keep in mind that this notation means that f is a bijective function from the left-hand set to the right-hand set.*

Equivalent characterizations of bijectivity:
1. f is both surjective and injective.
2. For every $y \in B$, there is one and only one $x \in A$ such that $f(x) = y$.
3. Every point in the codomain has a unique preimage.

Theorem 0.27 *If $f : B \to C$ and $g : A \to B$ are bijections, then $f \circ g$ is a bijection.*

The main reason we care about bijections is that bijections are exactly the functions that have inverses which are also functions:

Definition 0.28 *Let $f : A \to B$ be a function (with $Dom(f) = A$). If there is another function $f^{-1} : B \to A$ (with $Dom(f^{-1}) = B$) such that*

$$\forall x \in A, f^{-1}(f(x)) = x \quad \text{and} \quad \forall y \in B, f(f^{-1}(y)) = y$$

then we say f is **invertible** *and that f^{-1} is an* **inverse (function)** *of f.*

Example: Let $f : \mathbb{R} \to (0, \infty)$ be $f(x) = e^x$. Then $f^{-1} : (0, \infty) \to \mathbb{R}$ is $f^{-1}(x) = \ln x$. These are inverses because

$$f^{-1}(f(x)) = \ln e^x = x \quad \text{and } f(f^{-1}(x)) = e^{\ln x} = x.$$

Theorem 0.29 (Properties of inverse functions) *Let $f : A \to B$ and $g : B \to C$.*

1. *f is invertible if and only if f is bijective.*
2. *If f is invertible, then f has only one inverse function.*
3. *If f is invertible, then f^{-1} is invertible, and $(f^{-1})^{-1} = f$.*
4. *If f and g are invertible, then $f \circ g$ is invertible, and $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.*

Warnings on the notation "f^{-1}": the symbol f^{-1} is used for preimage and inverse function. Unless you know (or have proved) that the function f is invertible, f^{-1} means preimage, and is not actually referring to a function named "f^{-1}".

0.4 Cardinality

Loosely speaking, the *cardinality* of a set is its size. For example, the cardinality of $E = \{7, 17, 25, 32, 58\}$ is 5 (the count of the number of elements in the set). We write this as "$\#(E) = 5$". To define this more rigorously, we need some technical ideas:

Theorem 0.30 *Two sets A and B are said to be* **equinumerous** *(denoted $A \overset{1-1}{\longleftrightarrow} B$ or $A \leftrightarrow B$ or $A \approx B$) if there is a bijection $f : A \to B$. $\overset{1-1}{\longleftrightarrow}$ is an equivalence relation on the set of definable sets, and the equivalence classes under this relation are called* **cardinalities**. *The cardinality of a set E is denoted $\#(E)$.*

First, sets can be divided into two types: finite and infinite.

Definition 0.31 (Finiteness vs. infiniteness) *Let E be a set.*

1. *E is called* **finite** *if either:*
   - *E is empty, in which we write $\#(E) = 0$, or*
   - *$E \overset{1-1}{\longleftrightarrow} \{1, 2, 3, ..., n\}$ for some positive integer n, in which case we write $\#(E) = n$.*
2. *E is called* **infinite** *if E is not finite.*

The prototypical example of an infinite set is the set $\mathbb{N}$ of natural numbers.

Theorem 0.32 (Characterization of finite sets) *Let E be a set. TFAE (this means "the following are equivalent"):*

1. *E is finite;*
2. *for some finite set A, there is an injection $\hat{f} : E \hookrightarrow A$;*
3. *for some finite set A, there is a surjection $\hat{g} : A \hookrightarrow E$;*

Lemma 0.33 (Finite unions and Cartesian products of finite sets are finite) *Let $\{A_j\}_{j=1}^n$ be a finite collection of finite sets. Then, the following sets are all finite:*

1. *$A_1 \cup A_2$ (and in this case, $\#(A_1 \cup A_2) = \#(A_1) + \#(A_2) - \#(A_1 \cap A_2)$);*
2. *$\bigcup_{j=1}^{n} A_j$;*
3. *$A_1 \times A_2$;*
4. *A_1^n for any $n \in N$.*

Lemma 0.34 (Characterization of infinite sets) *Let E be a nonempty set. TFAE:*
1. *E is infinite;*
2. *for any $e \in E$, $E - \{e\}$ is infinite;*
3. *there is an infinite set A and an injection $\hat{f} : A \hookrightarrow E$;*
4. *there is an infinite set A and a a surjection $\hat{g} : E \twoheadrightarrow A$.*

We give two results which are central to the theory of cardinality here; I'm not sure if we will need these in Math 420:

Theorem 0.35 (Pigeonhole principle) *Let A and B be finite sets with $\#(A) < \#(B)$. Then:*
1. *there is no injection $g : B \hookrightarrow A$; and*
2. *there is no surjection $f : A \twoheadrightarrow B$.*

Theorem 0.36 (Cantor-Bernstein Theorem) *Let A and B be two sets. Suppose there are two injections $f : A \hookrightarrow B$ and $g : B \hookrightarrow A$. Then there is a bijection $h : A \leftrightarrow B$.*

Countability

Definition 0.37 *A set E is called* **countable** *if either E is empty, or one of these three equivalent conditions hold:*
1. *there is a surjective function $g : \mathbb{N} \twoheadrightarrow E$;*
2. *there is an injective function $f : E \hookrightarrow \mathbb{N}$;*
3. *either E is finite or in $1-1$ correspondence with $\mathbb{N}$.*

A set E is called **countably infinite** *if it is countable, but not finite (equivalently, if $E \overset{1-1}{\longleftrightarrow} \mathbb{N}$). An* **uncountable** *set is one which is not countable.*

NOTE: In order to list the elements of a set (as $\{x_1, x_2, x_3, ...\}$, for example), the set must be countable.

Theorem 0.38 (Classes of countable sets) *Let E be a countable set. Then:*
- *All of the following sets are countable:*
  1. *any set A with $E \overset{1-1}{\longleftrightarrow} A$;*
  2. *any set $f(E)$, where f is any function;*
  3. *any subset of E.*
- *Any union of finitely or countably many countable sets is countable.*
- *If A and B are countable, then $A \times B$ is countable.*

The natural numbers, the integers and the rational numbers are all countable. But any interval of real numbers is uncountable (including $\mathbb{R}$ itself).

0.5 Summary of proof techniques

Having zoomed through mathematical language in the previous four sections, now we give some guidelines on how to write typical classes of proofs. First, some general suggestions:

McCLENDON'S LAWS OF WRITING PROOFS

THE FIRST LAW: Work on scratch paper before writing your proof.

THE SECOND LAW: When in doubt, start by classifying the type of statement you are being asked to prove, and breaking down the logical structure of the statement to be proved.

This means identifying quantifiers, variables, conditionals, etc. The logical structure of the statement often suggests what your proof should "look like".

THE THIRD LAW: Keep in mind what you are always allowed to do in a proof.

- At any time, you can always state a hypothesis of the result you are to prove, state or use an axiom, state or apply a definition, and/or state or apply a previously proved result.
- At any time, you can state a sentence whose symbolic translation is a tautology.
- At any time, you can apply a rule of inference.
  - Most importantly, you can apply modus ponens. If you know P is true and you know $P \Rightarrow Q$, then you can state Q is true.
- At any time, you can change a statement into a logically equivalent form.

THE FOURTH LAW: Be willing to try something, even if you aren't sure it will work.

If you go through a line of reasoning that doesn't end where you want, that doesn't mean you are dumb. Try something else, and learn from what didn't work.

THE FIFTH LAW: The beginning and end of a proof should be easy to find. Start a proof by writing PROOF, and end a proof by writing $\square$ or Q.E.D. or $\#$ or some other symbol you use consistently.

General techniques for each type of logical structure

- To prove "If P, then Q":

  Direct proof: Suppose P. ... ... Therefore Q. $\square$
  Proof by contraposition: Suppose $\sim Q$. ... ... Therefore $\sim P$. $\square$
  Proof by contradiction: Suppose P and $\sim Q$. ... ... Therefore R. ... ... Therefore $\sim R$. Contradiction! $\square$
  Proof by cases: Suppose P. *Case 1:* ... ... Therefore Q. *Case 2:* ... ... Therefore Q. ... ... In all cases, Q. $\square$

- To prove "P if and only if Q":

  Biconditional proof: $(\Rightarrow)$ Suppose P. ... ... Therefore Q.
  $(\Leftarrow)$ Suppose Q. ... ... Therefore P. $\square$
  Shortcut biconditional proof: P iff ... iff ... iff ... ... iff Q. $\square$.

- To prove "TFAE: 1; 2; 3; ..."

  Circle of implications proof: Prove $(1 \Rightarrow 2)$, then $(2 \Rightarrow 3)$, then $(3 \Rightarrow 1)$, etc.

- To prove "$\forall x \in U, P(x)$":

  Generic particular argument: Let $x \in U$. ... ... Therefore $P(x)$. $\square$

- To disprove "$\forall x \in U, P(x)$":

  Disproof by counterexample: Let $x =$. ... ... Therefore $\sim P(x)$. $\square$

- To prove "$\exists x \in U : P(x)$":

  Constructive proof: Let $x =$. ... ... Therefore $P(x)$. $\square$
  Non-constructive proof: ... ... Therefore x exists by (some theorem). ... ... Therefore $P(x)$. $\square$

- To disprove "$\exists x \in U, P(x)$", prove the denial "$\forall x \in U, \sim P(x)$".

- To prove "$\exists! x \in U : P(x)$".

  Existence/uniqueness proof: First, prove $\exists x \in U : P(x)$.
  Then, suppose $P(x)$ and $P(y)$. ... ... Therefore $x = y$. $\square$

- To disprove "$\exists! x \in U : P(x)$", do one of two things:
  1. Disprove "$\exists x \in U, P(x)$" by proving the denial "$\forall x \in U, \sim P(x)$", or
  2. Disprove uniqueness by writing down specific x and y (with $x \neq y$) such that $P(x)$ and $P(y)$.

0.6 Outlines for common types of proofs

Proofs of conditionals

Most theorems that you are to prove are conditionals (i.e. "if [hypothesis], then [conclusion]"). The first approach one usually tries to prove such a statement is called a **direct proof**. Such a proof has the following structure:

DIRECT PROOF of $P \Rightarrow Q$:

Assume P.
.....
(some logical argument)
.....

Therefore, Q. $\square$

Sometimes a statement that you want to prove is of the form $(P_1 \text{ or } P_2) \Rightarrow Q$ (or is equivalent to this form). To do this, we use the following rule of inference (called "proof by cases" in the previous chapter):

$$[(P_1 \Rightarrow Q) \text{ and } (P_2 \Rightarrow Q)] \Rightarrow (P_1 \text{ or } P_2) \Rightarrow Q.$$

A proof that uses this rule of inference is called a **proof by exhaustion** or a **proof by cases**:

PROOF BY CASES of $(P_1 \text{ or } P_2 \text{ or } ... \text{ or } P_n) \Rightarrow Q$:

We consider n cases:

Case 1: Assume P_1 ..... (some logical argument) ..... Thus, Q.

Case 2: Assume P_2 ..... (some logical argument) ..... Thus, Q.

..... (more cases if necessary) .....

Case n: Assume P_n ..... (some logical argument) ..... Thus, Q.

In all cases, Q. $\square$

The phrase "WLOG" (which stands for "without loss of generality") is used when you are doing a proof by cases, and all the cases have basically the same proof. What "WLOG" signals is that you are going to prove one of several possible cases, and the other cases are identical (or their structure can be easily discerned from the proof of the case you give). For its use to be appropriate, either

(a) it must be totally obvious that the argument you give can be applied to any other cases; or

(b) any other cases should have some other totally obvious proof; or

(b) directions must be given to "reduce" other cases to the case you prove.

A typical application is when proving results about two integers (or real numbers) a and b. Assuming the roles of a and b are the same in the statement to be proven, you can assume WLOG that $a \leq b$. **Do not overuse the phrase WLOG. It is not a "catch-all" that allows you to assume anything under the sun.**

Any conditional $P \Rightarrow Q$ is logically equivalent to its contrapositive $\sim Q \Rightarrow \sim P$ ($\sim$ means "not", so to assume $\sim Q$ means to assume that Q is false). This means that we can prove conditional statements by giving a direct proof of their contrapositive:

PROOF BY CONTRAPOSITION of $P \Rightarrow Q$:

Assume $\sim Q$ ..... (some logical argument) ..... Thus, $\sim P$.

By contraposition, we are done. □

Recall also the rule of inference **reductio ad absurdum**, which says

$$[\sim P \Rightarrow (Q \text{ and } \sim Q)] \Rightarrow P.$$

A proof of statement P that applies this rule of inference is called a **proof by contradiction**:

PROOF BY CONTRADICTION of P:

Suppose not. (This is "math lingo" for "Assume $\sim P$.")

..... (some logical argument) .....

Therefore, Q.

..... (more logical argument) .....

Therefore, $\sim Q$.

Contradiction! Therefore, P. □

Sometimes, the contradiction needs to be explained (if it's not clear what the contradiction is).

Conditionals can be proven by contradiction with the following template:

PROOF BY CONTRADICTION of $P \Rightarrow Q$:

Assume P and $\sim Q$. *(because this is logically equivalent to* $\sim (P \Rightarrow Q)$*)*

..... (some logical argument) .....

Therefore, R.

..... (more logical argument) .....

Therefore, $\sim R$.

Contradiction! Therefore, $P \Rightarrow Q$. $\square$

Proofs of biconditionals

Recall that a biconditional is a statement of the form "$P \Leftrightarrow Q$", i.e. "P if and only if Q".

To prove a biconditional, there are two methods. The standard method is to rely on the fact that $P \Leftrightarrow Q$ is logically equivalent to $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$:

BICONDITIONAL PROOF of $P \Leftrightarrow Q$:

($\Rightarrow$) Prove $P \Rightarrow Q$ by some argument.

($\Leftarrow$) Prove $Q \Rightarrow P$ by some argument. $\square$

Sometimes you don't actually have to prove each direction of a biconditional:

SHORTCUT BICONDITIONAL PROOF of $P \Leftrightarrow Q$:

P iff .....
iff .....
iff .....

iff Q. $\square$

The acronym **TFAE** stands for "the following are (logically) equivalent". When you see a proposition of the form

"TFAE:
1. something
2. something else
3. another something else
4. one last thing"

that is the same thing as the proposition "$1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 4$". To prove such a proposition, you need to prove a "circle" of conditionals. For example, to prove

$$1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 4,$$

it is sufficient to prove $1 \Rightarrow 2$, $2 \Rightarrow 3$, $3 \Rightarrow 4$ and $4 \Rightarrow 1$.

Proofs of quantified statements

Mathematics contains many *quantified statements*. These statements are one of these three types:

1. $\forall x,\ P(x)$ (which means "for all x, the statement $P(x)$ is true");
2. $\exists x : P(x)$ (which means "there exists x such that $P(x)$ is true");
3. $\exists! x : P(x)$ (which means "exists a unique x such that $P(x)$ is true") (i.e. that there is exactly one x which makes $P(x)$ a true statement).

Here is how you prove (or disprove) each of these types of statements:

PROVING $\forall x, P(x)$ via **GENERIC PARTICULAR ARGUMENT**:
Write down a generic particular x and verify that $P(x)$ is true. □

The word "generic" is used because we assume nothing about x other than what is specified in the hypothesis of the result. The word "particular" is used because x is a particular element for which the open sentence $P(x)$ is checked.

Disproving $\forall x, P(x)$ via **COUNTEREXAMPLE**:
Write down a specific x, and show that for that x, $P(x)$ is false. □

CONSTRUCTIVE PROOF of $\exists x : P(x)$:

Write down a specific x, and show that for that x, $P(x)$ is true. □

NON-CONSTRUCTIVE PROOF of $\exists x : P(x)$:

Start with the assumptions of the proposition, and somehow, someway, show that there has to be an x for which $P(x)$ is true (usually by appealing to some other existence theorem which you already know). □

To disprove an existentially quantified sentence like $\exists x : P(x)$, there are two usual methods.

1. First, you can prove the denial $\forall x, \sim P(x)$ (usually by generic particular argument).
2. Disprove the statement by contradiction.

EXISTENCE/UNIQUENESS PROOF of $\exists! x : P(x)$:

First, give an existence proof of $\exists x : P(x)$ (as above)

Second, suppose $P(x)$ and $P(y)$ are true.

..... (logical argument) .....

Therefore $x = y$.

Therefore, $\exists! x : P(x)$. □

Subset and set equality proofs

Recall that the subset relationship $E \subseteq F$ can be restated as the conditional $\forall x, x \in E \Rightarrow x \in F$. This suggests a direct method for proving one set is a subset of another, called the **generic particular argument**:

GENERIC PARTICULAR ARGUMENT TO PROVE $E \subseteq F$:

Suppose $x \in E$ ..... (some logical argument) ..... Thus, $x \in F$.

Thus $E \subseteq F$.

Recall that two sets E and F are equal iff $E \subseteq F$ and $F \subseteq E$. This gives us a method of proving two sets are equal: you perform the generic particular argument twice, once to prove $E \subseteq F$ and again to prove $F \subseteq E$. This method should remind you of the method of proving a biconditional:

SET EQUALITY PROOF of $E = F$:

($\subseteq$) Suppose $x \in E$ ..... (some logical argument) ..... Thus, $x \in F$.

($\supseteq$) Suppose $x \in F$ ..... (some logical argument) ..... Thus, $x \in E$.

Since E and F are subsets of each other, $E = F$. $\square$

As with biconditionals, there is a shortcut method which is sometimes available:

SHORTCUT SET EQUALITY PROOF of $E = F$:

$x \in E$ iff ..... iff ..... iff ..... iff $x \in F$.

Thus $E = F$. $\square$

A shorthand version of this might be $E = ... = ... = ... = F$.

Proving a function surjective, injective, or bijective

PROVING that $f : A \to B$ is surjective:

Let $y \in B$.
Write a formula for some $x \in A$ (that comes from some scratch work).
Show that for the x you wrote down, $f(x) = y$.

Conclude that f is onto. $\square$

DISPROVING that $f : A \to B$ is surjective:

Find a specific $y \in B$.
Prove that $\sim \exists\, x \in A$ s.t. $f(x) = y$.

Conclude that f is not onto. $\square$

PROVING that $f : A \to B$ is injective:

Suppose $x, y \in A$ are such that $f(x) = f(y)$.
.....
Therefore, $x = y$.

Therefore f is $1 - 1$. $\square$

DISPROVING that $f : A \to B$ is injective:

Let $x =$ and $y =$ (choose specific $x, y \in A$).
.....
Therefore, $f(x) = f(y)$.

Therefore f is not $1 - 1$. $\square$

PROVING that $f : A \to B$ is a bijection:

1. Prove f is surjective.
2. Prove f is injective.

Therefore, f is a bijection. $\square$

PROVING that $f : A \to B$ is a bijection
(by constructing an inverse function of f)

Write down a formula for $f^{-1} : B \to A$
Show that for any $x \in A$, $f^{-1}(f(x)) = x$.
Show that for any $y \in B$, $f(f^{-1}(y)) = y$.

Conclude that f is invertible, hence f is a bijection. $\square$

DISPROVING that $f : A \to B$ is a bijection:

Either prove f is not surjective, or prove that f is not injective.

Therefore, f is not a bijection. $\square$

Proofs by induction

Mathematical induction can be used to prove statements which are universally quantified over the natural numbers, i.e. statements that look like

"$\forall\, n \in \mathbb{N}$, blah blah blah."

PROOF BY INDUCTION of $\forall n \in \mathbb{N}, P(n)$:

We proceed by induction (on n).

Base case: Verify $P(0)$.

Inductive step: Assume $P(k)$

..... (logical argument) ..... Therefore, $P(k+1)$.

Therefore, by the PMI, $P(n)$ is true for all $n \in \mathbb{N}$. $\square$

PROOF BY STRONG INDUCTION of $\forall n \in \mathbb{N}, P(n)$:

We proceed by induction (on n).

Base case(s): Verify $P(0)$ (sometimes it is necessary to verify $P(1)$ or $P(2)$ (maybe more), depending on how the inductive step works).

Inductive step: Assume that for all $j \leq k$, $P(j)$ is true.

..... (logical argument) ..... Therefore, $P(k+1)$.

Therefore, by the strong form of PMI, $P(n)$ is true for all $n \in \mathbb{N}$. $\square$

Proofs involving cardinality

Methods of proving a set E is finite

1. Show E is a subset of a finite set.
2. Show E is the union of finitely many finite sets.
3. Show E is the product of finitely many finite sets.
4. Show there is an injection from E to another finite set (such as $\{1, ..., n\}$).
5. Show E is the image of a finite set under some function.

Methods of proving a set E is infinite

1. Show E is in $1-1$ correspondence with one of its proper subsets.
2. Show E has a subset which is infinite.
3. Show E is the Cartesian product of some sets, of which at least one is infinite.
4. Show there is an injection from an infinite set (such as $\mathbb{N}$) to E.
5. Show there is a surjection from E to an infinite set (such as $\mathbb{N}$).

Methods of proving a set E is countable

1. Show E is finite.
2. Show E is a subset of a countable set.
3. Show E is the image of a countable set under some function.
4. Show there is an injective function from E to some other countable set.
5. Show E is a finite or countable union of countable sets.
6. Show E is the product of finitely many countable sets.

Methods of proving a set E is uncountable

1. Use a proof by contradiction.
2. Show there is an surjective function from E to some other uncountable set.
3. Show there is an injective function from some uncountable set to E.
4. Show E has a subset which is uncountable.
5. Show E is the product of some sets, at least one of which is uncountable.

Chapter 1

Major problems of abstract algebra

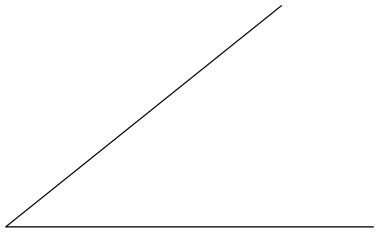
1.1 Straightedge and compass constructions

Geometric constructions with a straightedge (not a ruler) and a compass go back to the ancient Greeks.

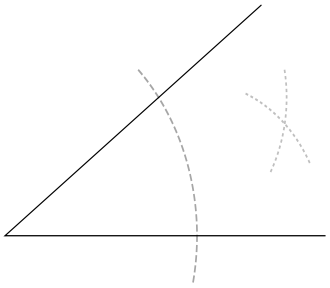
Things you can do with a straightedge and compass

1. Construct congruent segments

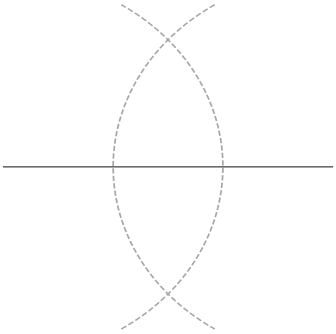
2. Construct congruent angles



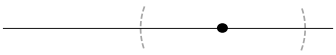
3. Bisect angles



4. Construct perpendicular bisectors



5. Construct perpendiculars through a point



6. Construct parallels

Wait a minute: This seems like geometry. Isn't this course "abstract **algebra**"?

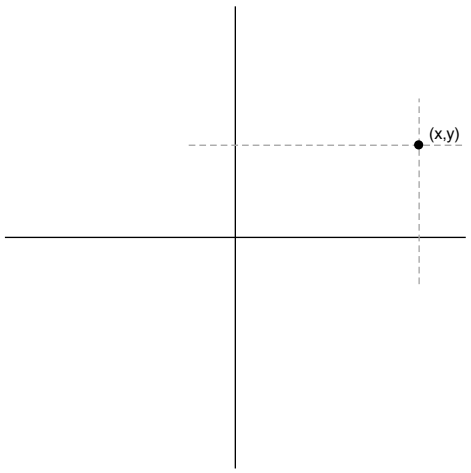
Turning geometry into algebra

Definition 1.1 *To* **construct** *a point $(x, y) \in \mathbb{R}^2$ means to obtain that point as the intersection of markings made with a straightedge and/or compass (i.e. lines and/or circles), given the locations of points $(0, 0)$ and $(1, 0)$. A point (x, y) is called* **constructible** *if you can construct it.*

Lemma 1.2 *Let x and y be real numbers. If (x, y) is constructible, then so is (y, x).*

PROOF Suppose (x, y) is constructible. Construct a parallel to the y-axis through (x, y); this gives you length x which you can mark on the y-axis. Similarly, construct a parallel to the x-axis through (x, y); this gives you length y which you can mark on the x-axis. Construct appropriate perpendiculars; where they meet is (y, x). □

Picture to explain:



Definition 1.3 *A real number is called* **constructible** *if it is either the x-coordinate or the y-coordinate of a constructible point.*

Question: What real numbers are constructible?

Activity: What points (and therefore what numbers) can you construct? **DO NOT LOOK AHEAD**.

Lemma 1 *If a and b are constructible, then $a + b$ and $a - b$ are constructible. (As a consequence, this means every integer is constructible.)*

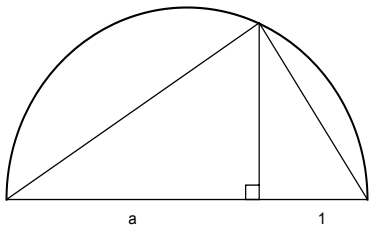
PROOF Suppose a and b are constructible. We construct $a \pm b$ as follows:

Lemma 2 *If a and b are constructible, then ab is constructible, and so long as $b \neq 0$, $\frac{a}{b}$ is constructible. (As a consequence, this means every rational number is constructible.)*

PROOF Suppose a and b are constructible, with $b \neq 0$. Here is how to construct ab and $\frac{a}{b}$:

Lemma 3 *Suppose $a \geq 0$ is constructible. Then $\sqrt{a}$ is constructible.*

PROOF Suppose a is constructible. Construct this diagram:



Definition 1.4 *A real number is called* **surd** *if it can be obtained from the integers by operations $+$, $-$, $\cdot$, $\div$ and $\sqrt{}$ (each of these operations may have to be done several times, but only finitely many operations are allowed). The set of surd numbers is called the* **real quadratic closure of** $\mathbb{Q}$.

Example: $\frac{5}{2} - 6\sqrt{3 + \sqrt{\frac{7}{13} + 5\sqrt{\frac{3+\sqrt{7}}{2}}}}$ is surd.

Example: $\sqrt[4]{15}$ is surd. (Why?)

Example: $\sqrt[3]{2}$ is not surd. (Or is it?)

Example: π

Together, the three lemmas we proved on the previous page proves the following theorem:

Theorem 1.5 *Every surd number is constructible.*

PROOF Let x be a surd number. x can be obtained from integers by a finite number of operations $+$, $-$, $\cdot$, $\div$ and $\sqrt{}$. By the lemmas on the previous page, any sum, difference, product, quotient and/or square root of constructible numbers is constructible, so x is constructible. □

Question: Are there any constructible numbers which are not surd?

Theorem 1.6 *Every constructible number is surd.*

PROOF First, some setup work. Suppose the two points (x_0, y_0) and (x_1, y_1) have surd coefficients. Then:

- The line passing through these points is

  $$y = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0);$$

  this equation can be rewritten as

  $$(x_1 - x_0)y + (y_1 - y_0)x = y_0(x_1 - x_0) - x_0(y_1 - y_0)$$

  In particular, this is a line of the form $Ax + By = C$ where A, B and C are surd numbers.

- The circle with center (x_0, y_0) passing through (x_1, y_1) has a radius

  $$r = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

  which is surd, so its equation is

  $$(x - x_0)^2 + (y - y_0)^2 = r^2$$

  where x_0, y_0 and r are surd.

Now for the main part of the proof. Every constructible number is obtained as the coordinate of a point constructed from $(0, 0)$ and $(1, 0)$ by a finite sequence of constructions. We prove the theorem by induction on the number of constructions it takes to actually construct the number.

Base case: Suppose it takes zero constructions to construct the number. Then the number is either 0 or 1; both of these are surd by definition.

Induction step: Suppose that every number/point which can be built from n constructions is surd. Let x be a constructible number built form $n + 1$ constructions. Since the only things you can make with a straightedge and compass are lines and circles, that means that x is a coordinate of one (or more) of these three things:

1. The intersection of two non-parallel lines passing through points with surd coefficients;

2. The intersection of a line passing through two points with surd coefficients and a circle whose center has surd coordinates and whose radius is surd;

3. The intersection of two circles, both of whom have center with surd coordinates and surd radius.

We will prove the theorem by showing that in each of these cases, the intersection has surd coordinates:

Case 1: Consider two lines with surd coefficients; call them

$$Ax + By = C \qquad \text{and } Dx + Ey = F$$

where $A, ...F$ are constructible. Solving these equations together (I'm leaving out the algebra) gives

$$x = \frac{CE - FB}{AE - BD} \text{ and } y = \frac{AF - BD}{AE - BD}$$

so x and y are surd.

Case 2: Consider a line $Ax + By = C$ and a circle $(x - x_0)^2 + (y - y_0)^2 = r^2$, where A, B, x_0, y_0 and r are surd. Solving for y in the first equation and plugging into the second gives

$$(x - x_0)^2 + \left(\frac{C - Ax}{B} - y_0\right)^2 = r^2,$$

a quadratic equation in x. By applying the quadratic formula (or doing other algebra), x can be obtained by $+$, $-$, $\cdot$, $\div$ and $\sqrt{}$ from surd numbers, so x is surd.

Case 3: Consider two circles where $(x - x_0)^2 + (y - y_0)^2 = r^2$ and $(x - x_1)^2 + (y - y_1)^2 = s^2$ where x_0, y_0, x_1, y_1, r and s are surd. If you solve these two equations for x (HW), you end up with a formula in terms of surd numbers with only $+$, $-$, $\cdot$, $\div$ and $\sqrt{}$ in it, so x is surd.

In every case, every constructible number is surd as wanted. □

Example: $\frac{5}{2} - 6\sqrt{3 + \sqrt{\frac{7}{13} + 5\sqrt{\frac{3+\sqrt{7}}{2}}}}$ is constructible.

Example: $\sqrt[8]{13}$ is constructible.

Example: $\sqrt[3]{2}$ is not constructible. (Or is it?)

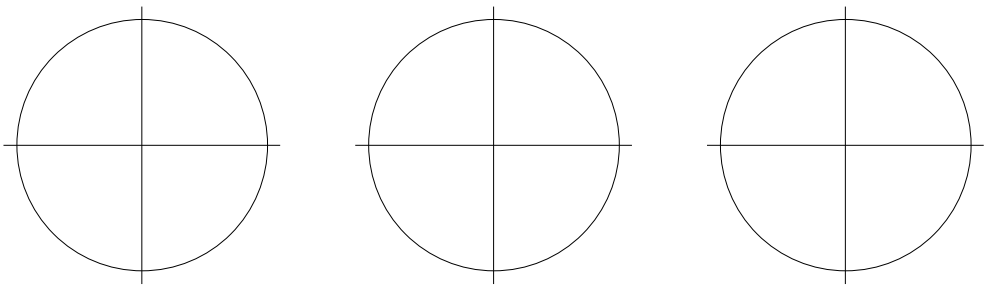
Other construction problems

I. Construction of regular polygons

Definition 1.7 *A* **regular polygon** *is a polygon where all the sides have the same length and all the angles have the same measure.*

Examples: equilateral triangles, squares, etc.

Question: for what n can a regular polygon with n sides be constructed?



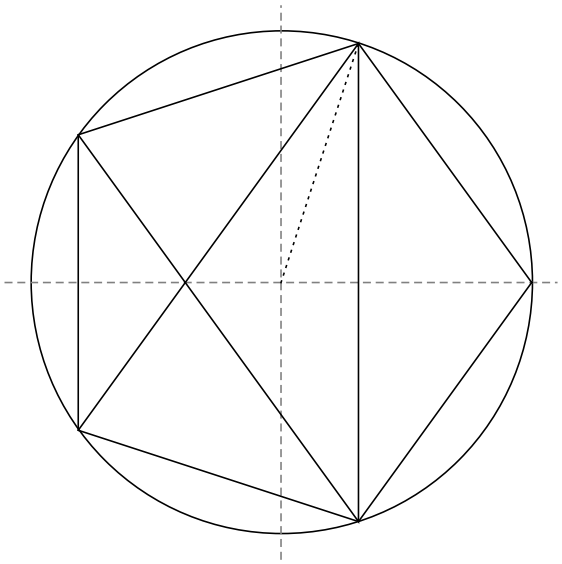
Equivalent formulation of construction problem I: Given $n \in \mathbb{N}$, is $\cos\frac{2\pi}{n}$ a constructible number?

Theorem 1.8 (Construction of regular polygons (preliminary cases)) *Let $n \in \mathbb{N}$. Then:*

- *if the regular polygon with n sides is constructible, so is the regular polygon with $2n$ sides.*
- *if $n \in \{3, 4, 5\}$, then the regular polygon with n sides is constructible.*

PROOF For the first statement, assume the regular n-gon is constructible. That means an angle of $\frac{2\pi}{n}$ can be constructed. Bisecting this angle gives an angle of measure $\frac{2\pi}{2n}$, which can be used to construct a regular $2n-$gon.

We verified the second statement above for $n = 3$ and $n = 4$. For $n = 5$, consider the picture on the next page:



II. Doubling the cube

Given a cube of volume 1, construct a cube of volume 2.

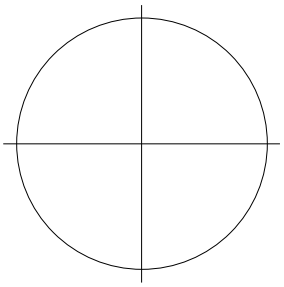
Equivalent formulation of II: Is $\sqrt[3]{2}$ a constructible number?

III. Trisection of a generic angle

Given an angle which you can construct, divide the angle into three equal parts.

Definition 1.9 *An angle θ is called* **constructible** *if $\cos\theta$ is a constructible number.*

Remark: An angle θ is constructible if and only if $\sin\theta$ is constructible, since $\cos^2\theta + \sin^2\theta = 1$.



Equivalent formulation of III: If $\cos\theta$ is a constructible number, must $\cos\frac{\theta}{3}$ also be a constructible number?

Remark: At least some of the time, this answer is yes: if $\theta = \frac{\pi}{2}$, then $\cos\frac{\theta}{3} = \cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}$ which is surd, hence constructible.

IV. Squaring the circle

Given a circle of radius 1, construct a square of the same area.

Equivalent formulation of IV: Is π a constructible number?

These construction problems baffled the ancient Greeks (for good reason, as we will see). One reason the Greeks got stuck is that they didn't have good language for turning geometry problems into algebra (they had no symbol for $\sqrt{}$ and no concept of variables or equations as we know them).

1.2 Polynomial equations

Definition 1.10 *A* **polynomial (in the one variable x)** *is any expression p of the form*

$$p(x) = a_0 + a_1x + a_2x^2 + ... + a_nx^n$$

where $a_0, a_1, ..., a_n \in \mathbb{R}$ and $a_n \neq 0$. n is called the **degree** *of the polynomial (we write $n = deg(p)$ for this) ; the numbers $a_0, a_1, ..., a_n$ are called the* **coefficients** *of the polynomial. a_n, the coefficient on the highest power of x, is called the* **leading coefficient** *of the polynomial. A polynomial is called* **monic** *if its leading coefficient is 1.*

Example: $p(x) = 3x^2 + \frac{5}{3}x - \sqrt{17\pi}$ is a polynomial of degree 2 whose leading coefficient is 3.

Example: $p(x) = x^9 - 17x + 4$ is a monic polynomial of degree 9.

Example: $p(x) = \sin x$ is not a polynomial (or is it?)

Theorem 1.11 *Let $p : \mathbb{R} \to \mathbb{R}$. p is a polynomial if and only if there exists an $n \in \mathbb{N}$ such that $p^{(n)}(x)$, the n^{th} derivative of p, is everywhere 0.*

PROOF HW

Theorem 1.12 *Let p and q be polynomials. Then pq is also a polynomial and $deg(pq) = deg(p) + deg(q)$.*

PROOF Let $m = deg(p)$ and $n = deg(q)$ so that

$$p(x) = a_0 + ... + a_mx^m \text{ and } q(x) = b_0 + ... + b_nx^n$$

for $a_0, ..., a_m, b_0, ..., b_n \in \mathbb{R}$ with $a_m \neq 0$, $b_n \neq 0$. Then

$$\begin{aligned} pq(x) &= (a_0 + a_1x + ... + a_mx^m)(b_0 + b_1x + ... + b_nx^n) \\ &= a_0b_0 + (a_1b_0 + a_0b_1)x + (something)x^2 + ... + a_mb_nx^n \end{aligned}$$

so pq is a polynomial, and since $a_mb_n \neq 0$, $deg(pq) = m + n$ as wanted. □

Classical problem: Solve polynomial equations $p(x) = 0$ in terms of the coefficients of p.

Remark: Every equation involving only polynomial stuff can be rewritten as $p(x) = 0$ by moving all the terms to one side.

We will study this classical problem (at first) by divvying up polynomial equations $p(x) = 0$ according to their degree.

Linear equations ($deg(p) = 1$)

Definition 1.13 *A* **linear equation** *is an equation of the form $p(x) = 0$ where p is a polynomial of degree 1.*

Every linear equation is therefore of the form

where $a \neq 0$. The solution of such an equation has been known to mankind for at least 4000 years (evidence from ancient Babylonian societies):

Remarks on linear equations:

- Every linear equation $ax + b = 0$ has exactly one real solution.
- To solve a linear equation, you need only the operations $+$, $-$, $\times$ and $\div$.

Quadratic equations ($deg(p) = 2$)

Definition 1.14 *A* **quadratic equation** *is an equation of the form $p(x) = 0$ where p is a polynomial of degree 2.*

Every quadratic equation is therefore of the form

$$ax^2 + bx + c = 0$$

where $a \neq 0$. If you are in this course, I hope you know the formula which gives the roots of this equation (which has been known to humanity since at least 400 BC). What you might not know (before today) is how to derive this formula:

Theorem 1.15 (Quadratic Formula) *The solutions of the quadratic equation*

$$ax^2 + bx + c = 0$$

(where $a \neq 0$) are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

PROOF Divide through the equation $ax^2 + bx + c = 0$ by a to get

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0$$

$$\text{i.e.} \qquad x^2 + \frac{b}{a}x = \frac{-c}{a}.$$

Now, "complete the square" by thinking of the problem geometrically:

Therefore, if we add $\left(\frac{b}{2a}\right)^2$ to the left-hand side, we will end up with $\left(x + \frac{b}{2a}\right)^2$. So let's add $\left(\frac{b}{2a}\right)^2 = \frac{b^2}{4a^2}$ to both sides to get:

$$\begin{aligned}
x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} &= \frac{b^2}{4a^2} - \frac{c}{a} \\
\left(x + \frac{b}{2a}\right)^2 &= \frac{b^2 - 4ac}{4a^2} \\
x + \frac{b}{2a} &= \pm\sqrt{\frac{b^2 - 4ac}{4a^2}} \\
x + \frac{b}{2a} &= \frac{\pm\sqrt{b^2 - 4ac}}{2a} \\
x &= \frac{-b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
\end{aligned}$$

as wanted. □

Historical footnote: The word "algebra" comes from the Arabic "al-jabr" which means "the reunion of broken parts". This nomenclature comes from the idea of completing the square (since you are "reuniting" the corner of the square that was missing. The Persian mathematician al-Khwarizmi was the first to write down the quadratic formula in algebraic language (i.e. variables, arithmetic symbols, etc.)

SECOND VERSION OF THE SAME PROOF As before, divide through the equation $ax^2 + bx + c = 0$ by a to get

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0 \qquad \textbf{Note:} \text{ this contains } x^2 \text{ and } x \text{ terms} \tag{1.1}$$

Now, instead of completing the square, think of the substitution $y = x + \frac{b}{2a}$. This means that

$$y^2 = x^2 + \frac{b}{a}x + \frac{b^2}{4ac}$$

so by substituting into (1.1), we get

$$\begin{aligned} y^2 - \frac{b^2}{4a^2} + \frac{c}{a} &= 0 \\ y^2 - \frac{b^2 - 4ac}{4a^2} &= 0 \qquad \textbf{Note:} \text{ this has } y^2, \text{ but no } y \text{ term} \\ y^2 &= \frac{b^2 - 4ac}{4a^2} \\ y &= \frac{\pm\sqrt{b^2 - 4ac}}{2a} \end{aligned}$$

Last, since $y = x + \frac{b}{2a}$, we conclude that

$$x = \frac{-b}{2a} + y = \frac{-b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a}$$

which is the usual quadratic formula. □

Question: Geometrically, what happened when we did the substitution $y = x + \frac{b}{2a}$?

Definition 1.16 *The* **discriminant** *of a quadratic equation is the number*

$$\Delta = b^2 - 4ac.$$

Remarks on quadratic equations:

- A quadratic equation $ax^2 + bx + c = 0$ could have 0, 1 or 2 real solutions.
  - If $\Delta = b^2 - 4ac > 0$, the equation has two real solutions.
  - If $\Delta = b^2 - 4ac = 0$, the equation has one real solution.
  - If $\Delta = b^2 - 4ac < 0$, the equation has no real solution.
- To solve a quadratic equation, you need only the operations $+$, $-$, $\times$, $\div$ and $\sqrt{\ }$. So every real solution of a quadratic equation is a surd number.

Question: Do you really need $\sqrt{\ }$?

Thinking about the number of solutions of a quadratic using calculus:

Let $p(x) = ax^2+bx+c$. Then $p'(x) = 2ax+b$, so the vertex of the parabola which is the graph of p is where $p'(x) = 0$, i.e. $x = \frac{-b}{2a}$. The corresponding y-coordinate of the vertex is

$$p\left(\frac{-b}{2a}\right) = a\left(\frac{-b}{2a}\right)^2 + b\left(\frac{-b}{2a}\right) + c = \frac{4ac - b^2}{4a} = \frac{-\Delta}{4a}.$$

	$\Delta > 0$	$\Delta < 0$	$\Delta = 0$
$a > 0$	p(vertex)	p(vertex)	p(vertex)
$a < 0$	p(vertex)	p(vertex)	p(vertex)

Cubic equations ($deg(p) = 3$)

Definition 1.17 *A* **cubic equation** *is an equation of the form $p(x) = 0$ where p is a polynomial of degree 3, i.e. an equation of the form*

$$ax^3 + bx^2 + cx + d = 0$$

where $a, b, c, d \in \mathbb{R}$ and $a \neq 0$.

To solve this, let's try "completing the cube" to get rid of the x^2 term. Let $y = x + \frac{b}{3a}$, so that $x = y - \frac{b}{3a}$. Then

$$ax^3 + bx^2 + cx + d = 0$$

$$a\left(y - \frac{b}{3a}\right)^3 + b\left(y - \frac{b}{3a}\right)^2 + c\left(y - \frac{b}{3a}\right) + d = 0$$

$$a\left(y^3 - 3\left(\frac{b}{3a}\right)y^2 + 3y\left(\frac{-b}{3a}\right) - \frac{b^3}{27a^3}\right) + b\left(y^2 - 2\frac{b}{3a}y + \frac{b^2}{9a^2}\right) + cy - \frac{bc}{3a} + d = 0$$

$$ay^3 - by^2 + -by - \frac{b^3}{27a^2} + by^2 - \frac{2b^2}{3a}y + \frac{b^3}{9a^2} + cy - \frac{bc}{3a} + d = 0$$

$$ay^3 + \left[-b - \frac{2b^2}{3a}\right]y + \left[\frac{-b^3}{27a^2} + \frac{b^3}{9a^2} - \frac{bc}{3a} + d\right] = 0$$

Now, divide through this last equation by a to get an equation of the form

$$y^3 + py + q = 0$$

where p and q are constants. Here is the punchline:

Lemma 1.18 *Every cubic equation $ax^3 + bx^2 + cx + d = 0$ can be transformed into an equation of the form*

$$y^3 + py + q = 0$$

by first performing the substitution $y = x + \frac{b}{3a}$ and then dividing through by a.

Question: How do you solve $y^3 + py + q = 0$?

Italian mathematicians del Ferro and Tartaglia (published in a book written by Cardano) figured out how to solve this in the 16th century. Here is their method:

Method of del Ferro and Tartaglia

Step 1: To solve $y^3 + py + q = 0$, let $y = u + v$. Then,

$$y^3 = (u+v)^3 = u^3 + 3u^2v + 3uv^2 + v^3 = u^3 + v^3 + 3uv(u+v).$$

Substitute into the equation you want to solve:

$$y^3 + py + q = 0$$

Step 2: Observe that $y = u + v$ is a solution of the equation if u and v satisfy the system of equations

Solve this system for u^3 and v^3 by substitution:

Step 3: Having obtained

$$\{u^3, v^3\} = \left\{ \frac{-q \pm \sqrt{q^2 + \frac{4p^3}{27}}}{2} \right\} = \left\{ \frac{-q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} \right\},$$

we see that $y = u + v = \sqrt[3]{u^3} + \sqrt[3]{v^3}$, i.e.

$$y = \sqrt[3]{\frac{-q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{\frac{-q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Potential problem: this method "fails" if the number $\frac{q^2}{4} + \frac{p^3}{27}$ under the square roots in the del Ferro / Tartaglia formula is negative.

Question: What is the significance of $\frac{q^2}{4} + \frac{p^3}{27}$?

Thinking about the number of solutions of a cubic using calculus:

Let $f(x) = x^3 + px + q$. Then $f'(x) = 3x^2 + p$ so the critical points of p are

Notice also that $f''(x) = 6x$ so by the Second Derivative Test, f has a local maximum at $x = -\sqrt{\frac{-p}{3}}$ and a local minimum at $x = \sqrt{\frac{-p}{3}}$:

Now, suppose f has three real roots. Then the quantity

$$\frac{1}{4} f\left(-\sqrt{\frac{-p}{3}}\right) f\left(\sqrt{\frac{-p}{3}}\right)$$

will be negative. If you work this out (HW), you will find that this quantity is exactly

$$\frac{q^2}{4} + \frac{p^3}{27}.$$

In other words, the cubic formula of del Ferro and Tartaglia will fail exactly when the cubic has three roots.

If f has one root, then

$$\frac{q^2}{4} + \frac{p^3}{27} = \frac{1}{4} f\left(-\sqrt{\frac{-p}{3}}\right) f\left(\sqrt{\frac{-p}{3}}\right)$$

will be positive, so the del Ferro / Tartaglia works to produce this one root.

If f has two roots, then the graph of p must look like

In this situation, the del Ferro / Tartaglia method works and produces one of these roots. In the HW, I ask you to figure out which one you get.

Definition 1.19 *The* **discriminant** *of the cubic equation $x^3 + px + q = 0$ is the quantity*

$$\Delta = -4p^3 - 27q^2.$$

(This quantity has the opposite sign as $\frac{q^2}{4} + \frac{p^3}{27}$.)

Remarks on cubic equations:

- After substitution and division by the leading coefficient, all cubic equations can be rewritten as

$$x^3 + px + q = 0.$$

- Such a cubic equation could have 1, 2 or 3 real solutions.
  - If $\Delta = -4p^3 - 27q^2 > 0$, the equation has three real solutions, and the method of del Ferro and Tartaglia fails.
  - If $\Delta = -4p^3 - 27q^2 = 0$, the equation has either one or two real solutions, and the method of del Ferro and Tartaglia gives one solution as
$$x = \sqrt[3]{\frac{-q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{\frac{-q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$
  - If $\Delta = -4p^3 - 27q^2 < 0$, the equation has one real solution, and the method of del Ferro and Tartaglia gives that solution by the same formula as above.
- To solve a cubic equation, you need the operations $+$, $-$, $\times$, $\div$, $\sqrt{\ }$ and $\sqrt[3]{\ }$.

Quartic equations ($deg(p) = 4$)

In the 1500s Ferrari showed that the ideas of del Ferro and Tartaglia could be used to solve quartic equations.

Definition 1.20 *A* **quartic equation** *is an equation of the form $p(x) = 0$ where p is a polynomial of degree 4, i.e. an equation of the form*

$$a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0$$

where $a_0, ..., a_4 \in \mathbb{R}$ and $a_4 \neq 0$.

This time, start with the substitution $y = x + \frac{a_3}{4a_4}$ and then divide by the leading coefficient. This gets rid of the x^3 term and leaves an equation of the form

$$y^4 + py^2 + qy + r = 0.$$

This can be rewritten using methods outlined in Section 1.7 of Stilwell's textbook, and solved for y.

Remarks on quartic equations:

- Quartic equations have 0, 1, 2, 3, or 4 real solutions.
- After substitution and division by the leading coefficient, all quartic equations can be rewritten as

  $$x^4 + px^2 + qx + r = 0.$$

- To solve a quartic equation, you need the operations $+$, $-$, $\times$, $\div$, $\sqrt{}$ and $\sqrt[3]{}$.

  (Why no $\sqrt[4]{}$?)

Quintic equations ($deg(p) = 5$)

The (hi)story so far

- Linear equations: Babylonians (4000 BC or perhaps earlier)
- Quadratic equations: Babylonians and Egyptians (2000 BC)
  - algorithms known to solve equations: Babylonians and Chinese (400 BC)
  - completing the square: al-Khwarizmi (800 AD)
  - quadratic formula first written down: Savasorda (1145 AD)
- Cubic equations: del Ferro and Tartaglia (1530)
- Quartic equations: Ferrari (1540)
- Quintic equations:

Definition 1.21 *A* **quintic equation** *is an equation of the form $p(x) = 0$ where p is a polynomial of degree 5. Every quintic equation is therefore*

$$a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0$$

where $a_0, ..., a_5 \in \mathbb{R}$.

What works: The substitution $y = x + \frac{b}{5a}$ gets rid of the fourth-power term. After dividing through by the leading coefficient, you are left with

$$x^5 + px^3 + qx^2 + rx + s = 0.$$

What doesn't work: Everything else.

Question: What *should* be true about quintic equations?

Definition 1.22 *A polynomial equation*

$$0 = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ... + a_n x^n$$

(with $a_n \neq 0$) is called **solvable by radicals** *if the solutions x are expressible in terms of $a_0, a_1, ..., a_n$ by means of the operations $+$, $-$, $\times$, $\div$, and the roots $\sqrt{}$, $\sqrt[3]{}$, $\sqrt[4]{}$, $\sqrt[5]{}$, etc.*

In this section (together with the exposition from Section 1.7 of Stillwell covering quartics), we have proven:

Theorem 1.23 *If p is a polynomial of degree less than or equal to 4, then the equation $p(x) = 0$ is solvable by radicals.*

Question: Is an arbitrary quintic equation solvable by radicals?

1.3 Summary: the major questions we want to tackle

Abstract algebra, at its heart, is about what you can and cannot do. Here are some questions we've seen that we will return to frequently in this course (recall that a number is **surd** if it can be obtained from integers by a finite string of the operations $+, -, \cdot, \div$ and $\sqrt{}$):

1. Which regular polygons are constructible using a straightedge and compass?

   What we know:

   - The regular n-gon is constructible if $n = 3, 4$ or 5.
   - If the regular n-gon is constructible, then the regular $2n$-gon is also constructible.
   - The regular n-gon is constructible if and only if $\cos \frac{2\pi}{n}$ is surd.

2. Can you double the cube using a straightedge and compass?

   Same problem, restated: Is $\sqrt[3]{2}$ surd?

   More generally: We know that an arbitrary cubic equation can be solve by radicals (you need $\sqrt{}$ and $\sqrt[3]{}$ at worst). Is an arbitrary cubic equation somehow solvable without using $\sqrt[3]{}$?

3. Can you trisect an arbitrary angle using a straightedge and compass?

   Same problem, restated: If $\cos\theta$ is surd, must $\cos\frac{\theta}{3}$ be surd?

4. Can you square the circle using a straightedge and compass?

   Same problem, restated: Is π constructible?

5. Is an arbitrary quintic equation (or higher-degree polynomial equation) solvable by radicals?

   If so, what is the quintic formula? Exactly what radicals are needed?

Chapter 2

Integers and rational numbers

2.1 $\mathbb{N}$: the natural numbers

Informally, the natural numbers are:

Formally speaking, the natural numbers are defined by some axioms that we just assume:

Axiom 2.1 (Peano's axioms) *The set of natural numbers is denoted $\mathbb{N}$. This is a set which satisfies these five axioms:*

Peano's first axiom: *0 is a natural number.*

Peano's second axiom: *Every natural number n has a* **successor** *$s(n)$ which is a natural number.*

Peano's third axiom: *0 is not the successor of any natural number.*

Peano's fourth axiom: *The successor function $s : \mathbb{N} \to \mathbb{N}$ is injective.*

Peano's fifth axiom: *The natural numbers are well ordered by $\leq$. This means that every nonempty subset of $\mathbb{N}$ has a least element.*

Remark: In the second axiom, "successor" just means the "next" natural number as you count. So the successor of 3 is 4, the successor of 6 is 7, etc.

Induction proofs

The reason the fifth axiom is so important is that it is equivalent to the following:

Axiom 2.2 (Principle of Mathematical Induction (PMI)) *Suppose $E \subseteq \mathbb{N}$ is a set with two properties:*

1. *$0 \in E$; and*
2. *if $n \in E$, then $n + 1 \in E$.*

Then $E = \mathbb{N}$.

Shorthand version: $[P(0)$ and $(P(n) \Rightarrow P(n+1))] \implies \forall n,\ P(n)$.

Application: Prove that for all $n \in \mathbb{N}$, $0 + 1 + 2 + ... + n = \frac{1}{2}n(n+1)$.

PROOF (Long version)

We usually shorthand the proof we just wrote as follows:

PROOF *Base case:* Let $n = 0$. Then $0 = \frac{1}{2}(0)(0+1)$.

Inductive step: Suppose $0 + 1 + ... + n = \frac{1}{2}n(n+1)$. Then

$$\begin{aligned} 0 + 1 + ... + n + (n+1) &= [0 + 1 + ... + n] + (n+1) \\ &= \frac{1}{2}n(n+1) + (n+1) \quad \text{(by the IH)} \\ &= \frac{1}{2}[n(n+1) + 2(n+1)] \\ &= \frac{1}{2}[n^2 + 3n + 2] \\ &= \frac{1}{2}(n+1)(n+2) \end{aligned}$$

as wanted. By induction, we are done. □

We may need this stronger version of induction:

Theorem 2.3 (Strong form of PMI) *Suppose $E \subseteq \mathbb{N}$ is a set with two properties:*

1. *$0 \in E$; and*
2. *if $\{0, 1, ..., n\} \subseteq E$, then $n + 1 \in E$.*

Then $E = \mathbb{N}$.

Shorthand version: $[P(0)$ and $(\forall k \leq n, P(k) \Rightarrow P(n+1))] \implies \forall n,\ P(n)$.

Application: Let F_n be the n^{th} Fibonacci number (i.e. $F_0 = F_1 = 1$ and for $n \geq 1$, $F_{n+1} = F_n + F_{n-1}$). Prove that for all $n \in \mathbb{N}$, $F_n \leq 2^n$.

PROOF *Base cases:* When $n = 0$, $F_0 = 1 = 2^0$ and when $n = 1$, $F_1 = 1 < 2^1$.

Induction step: Suppose that for all $k \leq n$, $F_k < 2^k$. Then

$$\begin{aligned} F_{n+1} &= F_n + F_{n-1} \\ &\leq 2^n + 2^{n-1} \quad \text{(by the IH)} \\ &< 2^n + 2^n \\ &= 2(2^n) = 2^{n+1} \end{aligned}$$

as wanted. By (strong) induction, we are done. □

Binary operations

Definition 2.4 *Let S be a set. A* **binary operation** *on S is a function $\odot : S \times S \to S$. An* **algebraic system**, *denoted $(S, \odot)$ or $(S, \odot_1, \odot_2, ...)$, is a set S together with one or more binary operations on S.*

Example 1: $(\mathbb{N}, +)$: addition on the natural numbers

Example 2: $(\mathbb{N}, \cdot)$: multiplication on the natural numbers

Binary operations have two inputs, and one output. We usually denote binary operations by symbols like $+$, $\times$, $-$, $\circ$, etc. and put the symbol in between the two coordinates of the input, so when we write something like

$$5 + 7 = 12,$$

we are technically thinking of a function which could (should?) written like

$$+(5, 7) = 12 \quad \text{or} \quad (5, 7) \overset{+}{\longmapsto} 12.$$

The binary operation presented in the above example (addition on $\mathbb{N}$) is symbolized by $+$. So technically, $+$ is a function from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$.

For now, we'll refer to an abstract binary operation (for now) by $\odot$. Think of this as a stand-in for something like $+$ or $\times$, etc.

Definition 2.5 ((Possible) properties of binary operations) *Let S be a set, and let $\odot$ be a binary operation on S.*

1. *$\odot$ is called* **associative** *if, for every $x, y, z \in S$, $x \odot (y \odot z) = (x \odot y) \odot z$.*
2. *$\odot$ is called* **commutative** *if, for every $x, y \in S$, $x \odot y = y \odot x$.*
3. *An element $e \in S$ is called an* **identity element (for $\odot$)** *if $x \odot e = x$ and $e \odot x = x$ for every $x \in S$.*
4. *Suppose e is an identity element for $\odot$. We say $x \in S$ has an* **inverse element (under $\odot$)** *if there is an element $y \in S$ such that $x \odot y = e$ and $y \odot x = e$.*

Of these properties, the most useful and important is associativity. Here's why:

Exercise: Complete the following chart with Yes or No answers (also, teach each other about the notation used for these sets):

Set S	Binary operation	Is this operation associative?	Is this operation commutative?	Is there an identity element? If so, what is it?	What members of S have inverses?
$\mathbb{N}$	$+$				
$\mathbb{N}$	$\times$, a.k.a. $\cdot$				
$M_{2\times 2}(\mathbb{R})$	$+$				
$M_{2\times 2}(\mathbb{R})$	matrix multiplication				
$\mathbb{N}$	exponentiation $a \odot b = a^b$				
$\{0\}$	trivial $0 \odot 0 \stackrel{\text{def}}{=} 0$				
2^E	$\bigcup$				
2^E	$\bigcap$				
2^E	$\triangle$				
$\mathbb{R}^3$	cross product				

ENRICHMENT: With regard to exponentiation, there is a natural number (namely 1) such that

$$a \odot 1 = a^1 = a$$

for all $a \in \mathbb{N}$. But 1 is **not** an identity element for this operation, because

$$1 \odot a = 1^a$$

is not generally equal to a. We say in this setting that 1 is a **right identity element for** $\odot$ (a **left identity element** would be some e such that $e \odot x = x$ for all $x \in S$; an identity element is an element that is both a left identity and a right identity).

What's good about $\mathbb{N}$?

One good thing about the natural numbers is that they are well-ordered, meaning that we can do induction proofs on statements indexed by natural numbers.

Other good things:

Theorem 2.6 *The operations $+$ and $\cdot$ on $\mathbb{N}$ are associative and commutative.*
There exists a unique identity element for $+$ (namely 0).
There exists a unique identity element for $\cdot$ (namely 1).

What's bad about $\mathbb{N}$?

Theorem 2.7 *The only element of $\mathbb{N}$ with an additive inverse is 0.*
The only element of $\mathbb{N}$ with a multiplicative inverse is 1.

Consequence: $(\mathbb{N}, +, \cdot)$ can be used to produce "unsolvable" equations:

ENRICHMENT: If you took Math 324 from me, you know that good mathematicians are skeptics and do not believe things that haven't been proven. On the other hand, I didn't prove Theorem 2.6 or 2.7. How do you prove these things?

Well, first you need formal definitions of addition and multiplication on $\mathbb{N}$. Here they are:

- Recall that by Peano's second axiom, there is a "successor" function $s : \mathbb{N} \to \mathbb{N}$. We define addition via iteration of this successor function, i.e.

  $$a + b \stackrel{\text{def}}{=} s(s(s(s(...s(s(a))))))$$

  where the successor function s is applied b times to a. Based on this definition, you can prove that addition is associative and commutative, and that 0 is the unique identity element under this operation, and that only 0 has an additive inverse (unfortunately, to make these proofs rigorous, you have to use induction and the definition of a function as a relation... that makes these proofs messy and technical and not worth doing, in my opinion).

- Once addition is defined and known to be associative, you can define multiplication by repeated addition:

  $$a \cdot b \stackrel{\text{def}}{=} a + a + ... + a$$

  where there are b as being added. You can prove (with messy induction proofs) that this multiplication operation is associative and commutative and that 1 is the unique identity element for this operation, and that 1 is the only natural number with a multiplicative inverse.

If you still don't believe Theorems 2.6 or 2.7, you can come to my office and I'll walk through these proofs with you.

2.2 $\mathbb{Z}$: the integers

First, an abstract definition of a set on which we can add, subtract and multiply:

Definition 2.8 (Definition of ring) *Let $(R, +, \cdot)$ be an algebraic system. $(R, +, \cdot)$ (or just R) is called a* **commutative ring with unit** *(or just a* **ring***) if*

1. *both $+$ and $\cdot$ are associative and commutative;*
2. *both $+$ and $\cdot$ have identity elements (usually denoted 0 and 1, respectively);*
3. *every member x of R has an inverse element under $+$ (which is denoted $-x$); and*
4. *$\cdot$ distributes over $+$, i.e. for all $x, y, z \in R$, $x(y + z) = xy + xz$.*

$\mathbb{N}$ is **not** a ring (Theorem 2.7 tells us there are no inverses under $+$ for nonzero elements of $\mathbb{N}$)... essentially this is what's "bad" about $\mathbb{N}$.

ENRICHMENT: Technically, for an algebraic system to be a a *ring*, one does not require that the multiplication is commutative, nor that there is an identity element for the multiplication. As an example, consider the set $M_{2\times 2}(\mathbb{R})$ of 2×2 matrices with real entries. The binary operations of matrix addition and matrix multiplication make this set into a (non-commutative) ring. We won't encounter non-commutative rings in this course, however, and in many (not all) advanced math textbooks, *ring* is used the way I use it in these lecture notes. Always look carefully at how an author defines the word *ring*!

Definition 2.9 *Let R be a ring. Define* **subtraction** *in R by*

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

where $+$ is the addition operation of R.

Consequence: Given $a, b \in R$ where R is a ring, we can always solve an equation of the form

$$x + a = b \qquad \text{or} \qquad x - a = b.$$

Theorem 2.10 (Cancellation law for rings) *Let R be a ring and suppose $a, b, c \in R$. If $a + b = a + c$, then $b = c$.*

PROOF

Theorem 2.11 (Properties of rings) *Let R be a ring, and denote its additive identity by 0 and its multiplicative identity by 1. Let $x \in R$. Then:*

Uniqueness of identity elements *0 is the only additive identity of R, and 1 is the only multiplicative identity of R.*

Uniqueness of additive inverses *x has exactly one additive inverse.*

Multiplication by zero $0x = 0$.

Additive inverse of additive inverse is original element $-(-x) = x$.

Multiplication by -1 gives additive inverse $-1(x) = -x$.

Additive inverse of multiplicative identity is itself $-0 = 0$.

PROOF First, we prove that 0 is the only additive identity of R. Suppose a is an additive identity of R. Then consider

$$a + 0 = \left\{ \right.$$

By transitivity, $a = 0$, so 0 is the only additive identity of R.

Next, let's prove uniqueness of additive inverses. Suppose y and z are both additive inverses of x. Then

$$x + y = 0 = x + z$$

and by the cancellation law, $y = z$. Thus additive inverses are unique.

The other parts of this are HW problems. □

Prototype ring: the integers

Informally, the integers are

i.e. they are the natural numbers, together with additive inverses of all the natural numbers (sort of... see the enrichment at the end of this section).

What's good about $\mathbb{Z}$:

1. the integers "contain" the natural numbers (sort of... see the enrichment at the end of this section).

2. the integers are totally ordered under $\leq$. Furthermore, inequalities are preserved under addition, subtraction and multiplication by positive integers (and reversed under multiplication by a negative integer). To prove this, however, you need messy, technical definitions of $\leq$ and induction proofs which I will omit.

3. $\mathbb{Z}$ is a ring:

   Theorem 2.12 *$\mathbb{Z}$ is a ring, under addition and multiplication.*

   PROOF See the enrichment at the end of this section.

4. $\mathbb{Z}$ is better than a ring:

   Definition 2.13 *A ring R is called an* **integral domain** *if "0 has no nontrivial divisors", i.e. if $a, b \in R$ are such that $ab = 0$, then at least one of a and/or b must be 0.*

   Theorem 2.14 *$\mathbb{Z}$ is an integral domain.*

   PROOF Suppose $ab = 0$ but $a \neq 0, b \neq 0$.

   Case 1: $b > 0$. Then

   $$ab = a + a + a + ... + a.$$

If $a > 0$, then $ab > 0$ since $+$ preserves $>$. If $a < 0$, then $ab < 0$ since $+$ preserves $<$. Either way, we have contradicted $ab = 0$.

Case 2: $b < 0$. Then by ring properties, $ab = 1(ab) = (-1)(-1)(ab) = (-1)a(-1)b = (-a)(-b)$. Applying Case 1 (which holds since $-b > 0$), the result follows. □

5. in $\mathbb{Z}$, you can perform "division with remainder" (more on this in a future section).

What's not so great about $\mathbb{Z}$?

1. $\mathbb{Z}$ is not well-ordered under $\leq$ (for instance, there is no least integer). So you can't do induction proofs directly on statements indexed by integers (although you can often do induction under the assumption the integer is at least zero, then prove a second case where the integer is negative).

2. $\mathbb{Z}$ does not contain reciprocals:

Theorem 2.15 *The only members of $\mathbb{Z}$ which have multiplicative inverses are ± 1.*

PROOF Clearly, 1 and -1 are their own multiplicative inverses.

Suppose $x \in \mathbb{Z}$ has multiplicative inverse y. That means $xy = 1$. Taking absolute values, this means $|xy| = 1$, i.e. $|x||y| = 1$. Therefore $|x|$ is a natural number with a multiplicative inverse, which by Theorem 2.7 means $|x| = 1$. Therefore $x = \pm 1$. □

Consequence: $(\mathbb{Z}, +, \cdot)$ can be used to produce "unsolvable" equations:

P.S. What other rings are there?

ENRICHMENT: In Math 324, we learned exactly what the integers are. Formally speaking, an integer is an equivalence class of a pair of natural numbers, where $(a, b) \in \mathbb{N}^2$ is equivalent to $(c, d) \in \mathbb{N}^2$ if

$$a + d = b + c.$$

Denoting the equivalence class of (a, b) by $[(a, b)]$, the integer n is the equivalence class $[(n, 0)]$ and the integer $-n$ is the equivalence class $[(0, n)]$.

Then, addition and multiplication on integers can be defined formally by

$$[(a, b)] + [(c, d)] = [(a + c, b + d)]$$
$$[(a, b)] \cdot [(c, d)] = [(ac + bd, ad + bc)]$$

These operations need to be shown to be "well-defined" (see my Math 324 notes). From these definitions, we can verify that $\mathbb{Z}$ is a ring (i.e. that addition and multiplication are associative and commutative and have identity elements (the additive identity is $[(0, 0)]$ and the multiplicative identity is $[(1, 0)]$), that multiplication distributes over addition, and that every class has an additive inverse (the additive inverse of $[(a, b)]$ is $[(b, a)]$).

This having been done, what do we mean when we say the integers "contain" the natural numbers? Really, this means there is an injective function $i : \mathbb{N} \hookrightarrow \mathbb{Z}$ (defined by $i(n) = [(n, 0)]$) such that i "preserves addition and multiplication", i.e.

$$i(a + b) = i(a) + i(b) \quad \text{and} \quad i(ab) = i(a)i(b)$$

for all $a, b \in \mathbb{N}$. The image $i(\mathbb{N})$ is essentially a copy of the natural numbers inside the integers. So when we write something like "$\mathbb{N} \subseteq \mathbb{Z}$", we really mean $i(\mathbb{N})$ that is a subset of $\mathbb{Z}$.

2.3 $\mathbb{Q}$: the rational numbers

First, an abstract definition of a set on which we can add, subtract, multiply and divide (but not by 0):

Definition 2.16 (Definition of field) *Let $(F, +, \cdot)$ be an algebraic system, where F contains at least two elements. F is called a* **field** *if*

1. *$(F, +, \cdot)$ is a ring; and*
2. *every element of F other than the additive identity element (i.e. 0) has an inverse element under multiplication (i.e. a* **reciprocal***). The reciprocal of $x \in F$ is denoted x^{-1} or $\frac{1}{x}$.*

Definition 2.17 *Let F be a field. For any $x \in F$, $y \in F - \{0\}$, we can define* **division** *in F by*

$$x \div y = \frac{x}{y} = x \cdot y^{-1}$$

where $\cdot$ is the multiplication operation of F.

Consequence: Given $a, b, c \in F$ where F is a field, if $a \neq 0$ we can always solve an equation of the form

$$ax + b = c.$$

Theorem 2.18 (Cancellation law for fields) *Let F be a field and suppose $a, b, c \in F$. If $ab = ac$ and $a \neq 0$, then $b = c$.*

Note: Since every field is a ring, the cancellation law for rings ($a + b = a + c$ implies $b = c$) also holds for fields.

PROOF

Just as $\mathbb{Z}$ is the prototype ring, $\mathbb{Q}$ is the prototype field. Informally, the rational numbers are fractions, where the numerator and denominator of the fraction are each an integer (and the denominator is nonzero):

What's good about $\mathbb{Q}$?

1. Like $\mathbb{Z}$, $\mathbb{Q}$ is totally ordered under $\leq$. Furthermore, inequalities are preserved under addition, subtraction and multiplication by positive rationals (and reversed under multiplication by negative rationals).
2. Like $\mathbb{Z}$, $\mathbb{Q}$ is an integral domain.
3. The rational numbers "contain" the integers (see the enrichment at the end of this section).
4. Unlike $\mathbb{Z}$, $\mathbb{Q}$ is a field (so you can take reciprocals, etc.):

   Theorem 2.19 *$\mathbb{Q}$ is a field, under addition and multiplication.*

   PROOF See the enrichment at the end of this section.

What's bad about $\mathbb{Q}$?

1. Like $\mathbb{Z}$, $\mathbb{Q}$ is not well-ordered under $\leq$ (so induction proofs are no good).
2. Unlike $\mathbb{Z}$, $\mathbb{Q}$ doesn't contain any "primes", which means that you can't really "factor" rational numbers in a formulaic way (we will see that in $\mathbb{Z}$, factorization works better).
3. There are some equations with coefficients in $\mathbb{Q}$ that arise naturally whose solutions can't be easily identified as rational numbers:

P.S. What other fields are there?

Lots of subsets of fields are themselves fields. The key concept is that if you perform the operations of addition, subtraction (i.e. additive inverse), multiplication, and/or division (i.e. reciprocal) on objects in the subset, then the answer must still be in the subset.

Theorem 2.20 *Let $(F, +, \cdot)$ be a field, with additive identity 0 and multiplicative identity 1. If $F_0 \subseteq F$ is a nonempty set with the following properties:*

1. *F_0 is "closed under addition", i.e. $x + y \in F_0$ whenever $x \in F_0$ and $y \in F_0$;*
2. *F_0 is "closed under multiplication", i.e. $xy \in F_0$ whenever $x \in F_0$ and $y \in F_0$;*
3. *F_0 is "closed under additive inverses", i.e. $-x \in F_0$ whenever $x \in F_0$; and*
4. *F_0 is "closed under reciprocals", i.e. $x^{-1} \in F_0$ whenever $x \in F_0$,*

then F_0 is also a field under operations $+$ and $\cdot$. In this setting we say F_0 is a **subfield** *of F and that F is an* **extension** *of F_0.*

PROOF This follows immediately from the definition of field. □

Theorem 2.21 *The set $\mathbb{S}$ of surd numbers (i.e. constructible numbers) is a field (under the usual operations of $+$ and $\cdot$).*

PROOF In Section 1.1, we showed that the sum and product of two surd numbers is surd, that the additive inverse of a surd number is surd, and that the reciprocal of a surd number is surd. Therefore $\mathbb{S}$ is a subfield of the field $\mathbb{R}$.

As a non-example, the integers fail to be a subfield of $\mathbb{Q}$ because they are not closed under reciprocals:

ENRICHMENT: In Math 324, we learned exactly what the rational numbers are. Formally speaking, an integer is an equivalence class of an element in $S = \mathbb{Z} \times (\mathbb{Z} - \{0\})$, where $(a, b) \in S$ is equivalent to $(c, d) \in S$ if $ad = bc$; denote the equivalence class of (a, b) by $\frac{a}{b}$.

Then, addition and multiplication on rational numbers can be defined formally by

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$$
$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$$

These operations need to be shown to be "well-defined" (see my Math 324 notes). From these definitions, we can verify that $\mathbb{Q}$ is a field (in particular, the additive identity is $\frac{0}{1}$, the multiplicative identity is $\frac{1}{1}$, the additive inverse of $\frac{a}{b}$ is $\frac{-a}{b}$, and the multiplicative inverse of $\frac{a}{b}$ exists if $a \neq 0$ and is $\frac{b}{a}$).

This having been done, we can see that there is an injective function $i : \mathbb{Z} \hookrightarrow \mathbb{Q}$ (defined by $i(n) = \frac{n}{1}$) such that i "preserves addition and multiplication", i.e.

$$i(a + b) = i(a) + i(b) \quad \text{and} \quad i(ab) = i(a)i(b)$$

for all $a, b \in \mathbb{Z}$. The image $i(\mathbb{Z})$ is essentially a copy of the integers inside the rational numbers. So when we write something like "$\mathbb{Z} \subseteq \mathbb{Q}$", we really mean $i(\mathbb{Z})$ that is a subset of $\mathbb{Q}$, and when we think of an integer (like 5) as a rational number, we are really thinking about $i(5) = \frac{5}{1}$.

2.4 Divisibility

Recall: what's bad about $\mathbb{Q}$ is that there aren't primes, so factoring doesn't make sense. But you can factor elements of $\mathbb{Z}$ in a reasonable way:

Definition 2.22 *Let $a, b \in \mathbb{Z}$. We say a* **divides** *b, or a is a* **factor** *of b, or b is a* **multiple** *of a, and write $a \mid b$, if there is an integer q such that $b = qa$. If a does not divide b, we write $a \nmid b$.*

Definition 2.23 *Let $S \subseteq \mathbb{Z}$. A* **(finite) linear combination** *of elements of S is any integer which can be written as*

$$\sum_{j=1}^{n} c_j s_j$$

for integers $c_1, ..., c_n$ and elements $s_1, ..., s_n \in S$.

For example, a linear combination of a and b is any integer which can be written as $ka + lb$ for $k, l \in \mathbb{Z}$.

Example: $17 \mid 221$ since $221 = 17(13)$.

Example: $5 \nmid 23$

Example: 5 is a linear combination of 10 and 25 because $5 = 3(10) + (-2)(25)$.

Example: 5 is not a linear combination of 24 and 28

Theorem 2.24 (Basic properties of divisibility) *Let $a, b, c \in \mathbb{Z}$.*

1. *$a \mid a$.*
2. *If $a \mid b$ and $b \neq 0$, then $|a| \leq |b|$.*
3. *If $a \mid b$ and $a \mid c$, then a divides any linear combination of b and c.*
   In particular, this implies
   - *$a \mid b \Rightarrow a \mid kb$ for any $k \in \mathbb{Z}$, and*
   - *$(a \mid b$ and $a \mid c) \implies$ both $a \mid (b + c)$ and $a \mid (b - c)$.*
4. *If $a \mid b$ and $b \mid c$, then $a \mid c$.*

PROOF Statement (1) is obvious, since $a = 1a$.
For statement (2), suppose $a \mid b$ and $b \neq 0$.

The proofs of (3) and (4) are left as HW. □

Application: Let's prove $5 \nmid 23$.

Definition 2.25 (Classification of ring elements) *Let R be a ring (with multiplicative identity 1).*

- *An element $u \in R$ is called a* **unit** *if $u \mid 1$.*
- *An element $p \in R$ is called* **prime** *if whenever $p = ab$ for $a, b \in R$, either a or b is a unit.*
- *An element $c \in R$ is called* **composite** *if it is nonzero, not a prime, and not a unit.*

Example: in $\mathbb{Z}$, the units are

in $\mathbb{Z}$, primes include

in $\mathbb{Z}$, composites include

Lemma 2.26 *Let $p \in \mathbb{Z}$. p is prime if and only if the only natural numbers which divide p are 1 and p.*
Let $c \in \mathbb{Z}$. c is composite if and only if $\exists\, a \in \mathbb{Z}$ with $1 < |a| < |c|$ such that $a \mid c$.

Question: Generalizing this definition, what would the units/primes/composites be in the rational numbers?

Theorem 2.27 *A ring is a field if and only if every nonzero element of the ring is a unit.*

Lemma 2.28 (Euclid's Lemma) *Let $n \in \mathbb{Z}$. If n is not a unit, then there is a prime p such that $p \mid n$.*

PROOF Let $n \in \mathbb{Z}$ be a non-unit.

Case 1: Suppose $n = 0$. Since $2 \mid 0$, the theorem holds.

Case 2: We will show that all positive non-units have prime divisors. To do this, let E be the set of integers which are at least 2 and which have no prime divisor. Assume $E \neq \emptyset$. Since $E \subseteq \mathbb{N}$, by the WOP E has a least element x.

Case 2(a): x is prime. Therefore x has prime divisor x, so $x \notin E$, a contradiction.

Case 2(b): x is composite. By definition, there is some integer y (WLOG y is positive), which is neither 1 nor x, which divides x. Since $y < x$, $y \notin E$ so y has a prime divisor p. But $p \mid y$ and $y \mid x$ implies $p \mid x$, so x has prime divisor p. This contradicts x being in E.

Either way, we have a contradiction, so E must be empty, proving the theorem whenever n is a natural number.

Case 3: Last, suppose n is a negative non-unit, i.e. $n \leq -2$. If n has no prime divisor, then neither does $-n$. Since $-n \geq 2$, this would contradict what we have already proven in Case 2. □

Theorem 2.29 (Euclid's Theorem) *There are infinitely many primes.*

PROOF Suppose not, i.e. that $\mathbb{P}$ is finite. That means we can list the positive prime numbers as $\{p_1, p_2, ..., p_n\}$.

The concept of a prime numbers has been around for thousands of years, but there are still unsolved problems involving primes that are simple to phrase:

Conjecture 2.30 (Twin Prime Conjecture) *There are infinitely many numbers p such that both p and $p+2$ are prime.*

Examples of twin primes: 3 and 5; 29 and 31; 71 and 73

Conjecture 2.31 (Goldbach Conjecture) *Every even integer greater than 2 can be written as the sum of two primes.*

Examples: $4 = 2 + 2$; $100 = 47 + 53$

The Goldbach conjecture has been verified for all even integers up to about 10^{18} via computer, but no one knows if it is true or not.

Division with remainder

One of the most important things about the integers is that you can perform division with remainder:

Example: $27 \div 2 =$

Example: $13 \div 5 =$

Example: $3 \div 7 =$

Example: $-22 \div 6 =$

Observations about these examples:

Theorem 2.32 (Division Theorem) *Let $a, b \in \mathbb{Z}$ with $a > 0$. Then there exist unique integers q and r such that:*

- $b = aq + r$, *and*
- $r \in \{0, 1, 2, ..., a - 1\}$.

In this theorem, q is referred to as the **quotient** *and r is referred to as the* **remainder**.

PROOF Let $a, b \in \mathbb{Z}$ with $a > 0$. The first part of our proof will establish the existence of q and r.

Case 1: $b \geq 0$. Since b is therefore a natural number, we can prove this by induction on b.

Base case: $0 = a(0) + 0$, so we can set $q = r = 0$ to prove the theorem.

Induction step: Suppose the theorem is true for b, i.e. there are integers q and r with $r \in \{0, ..., a - 1\}$ such that

$$b = aq + r.$$

Case 1(a): $r = a - 1$

Case 1(b): $r < a - 1$

In either case, by induction on b, we are done with Case 1.

Case 2: $b < 0$. Therefore $-b > 0$, so by Case 1, we can divide $-b$ by a to get

$$-b = aq + r$$

where $q, r \in \mathbb{Z}$ and $0 \leq r \leq a - 1$. That means

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

Case 2(a): $r > 0$. In this case, we have written

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

so the theorem holds.

Case 2(b): $r = 0$. In this case, $-b = aq$ so $b = a(-q) + 0$, and the theorem holds.

Between Cases 1 and 2, we have established the existence of the q and the r. Now for the uniqueness: to show this, suppose there are $q_1, q_2, r_1, r_2 \in \mathbb{Z}$ with $0 \leq r_1, r_2 \leq a - 1$ such that

$$b = aq_1 + r_1 \quad \text{and} \quad b = aq_2 + r_2.$$

WLOG $r_2 \leq r_1$, otherwise reverse the subscripts. Now subtract the two equations above to get

$$0 = b - b = aq_1 + r_1 - (aq_2 + r_2) = a(q_1 - q_2) + (r_1 - r_2),$$

i.e.

$$a(q_2 - q_1) = r_1 - r_2.$$

Since a divides the left-hand side, a divides the right-hand side, but the right-hand side is nonnegative and at most $(a - 1) - 0 < a$, a contradiction unless the right-hand side is 0. Therefore $r_1 = r_2$, so

$$a(q_2 - q_1) = 0$$

so since $a \neq 0$, it follows that $q_2 = q_1$. Thus the q and r of this theorem are unique. □

2.5 The Euclidean algorithm and applications

Definition 2.33 *Let $a, b \in \mathbb{Z}$ be nonzero. The* **greatest common divisor** *of a and b, denoted $\gcd(a, b)$, is the integer d with the following three properties:*

- *$d \mid a$;*
- *$d \mid b$; and*
- *for any $c \in \mathbb{Z}$ such that $c \mid a$ and $c \mid b$, $c \leq d$.*

a and b are called **relatively prime** *(a.k.a.* **coprime***) if $\gcd(a, b) = 1$.*

Definition 2.34 *Let $S \subseteq \mathbb{Z}$ be a set not containing 0. The* **greatest common divisor** *of S, denoted $\gcd(S)$, is the integer d with the following two properties:*

- *for every $a \in S$, $d \mid a$; and*
- *for any $c \in \mathbb{Z}$ such that $c \mid a$ and $c \mid b$, $c \leq d$ (i.e. d is the greatest common divisor).*

Remark: The $\gcd$ of two numbers is always positive: $\gcd(-5, -10) = 5$, for instance.

Example: Find these:

- $\gcd(49, 63)$
- $\gcd(\{18, 24, 10, 72\})$
- $\gcd(19, 418)$
- $\gcd(7458, 5082)$

Why do we say "the" $\gcd$ rather than "a" $\gcd$?

HOW TO PROVE $d = \gcd(a, b)$**:**

1. Show $d \mid a$.
2. Show $d \mid b$.
3. Suppose $c \mid a$ and $c \mid b$. Use a logical argument to deduce $c \leq d$ (often it is easier to show $c \mid d$ and $d \geq 0$).

Lemma 2.35 *Let $a, b \in \mathbb{Z}$ be nonzero.* $\gcd(a, b) = \gcd(|a|, |b|)$.

PROOF Let $d = \gcd(a, b)$. We will show $d = \gcd(|a|, |b|)$;

- $d = \gcd(a, b)$, so $d \mid a$. That means $\exists\, n \in \mathbb{Z}$ s.t. $dn = a$. If $a > 0$, then $dn = |a|$; if $a < 0$, then $d(-n) = |a|$. Either way, $d \mid |a|$.
- By the same argument as above applied to b instead of a, $d \mid |b|$.
- Last, suppose $c \mid |a|$ and $c \mid |b|$. Then $\exists\, n \in \mathbb{Z}$ s.t. $cn = |a|$ so either $cn = a$ or $c(-n) = a$, meaning $c \mid a$. Similarly, $c \mid b$. Since c is a common divisor of a and b, $c \leq d = \gcd(a, b)$ as wanted. □

P.S. Inside this proof we showed

Lemma 2.36 *Let $a, d \in \mathbb{Z}$ be nonzero. $d \mid a$ if and only if $d \mid |a|$.*

and in fact these statements are also equivalent to $|d| \mid a$ and $|d| \mid |a|$. So when you are messing with divisibility, you can often (not always) assume WLOG that the integers under consideration are positive.

Lemma 2.37 *Let $a, b \in \mathbb{Z}$ be nonzero. If $a \mid b$, then* $\gcd(a, b) = |a|$.

PROOF HW

Back to this example: $\gcd(7458, 5082) = ?$

One way to do this is to factor each of these numbers and identify the common factors. That way sucks, and I'm not going to spend any time doing it. A better way to compute this is called the *Euclidean algorithm*, and is based on this important principle:

Theorem 2.38 *Let $a, b \in \mathbb{Z}$ be nonzero. If $q, r \in \mathbb{Z}$ are such that $b = aq + r$, then*

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

PROOF Let $d = \gcd(a, b)$. We need to show $d = \gcd(a, r)$:

The **Euclidean algorithm** is a procedure to find the $\gcd$ of two numbers by repeatedly dividing the larger one by the smaller one (using the Division Theorem) and applying the preceding theorem, until you get a remainder of zero.

Example: Find $\gcd(7458, 5082)$.

Example: Find $\gcd(945, 672)$.

Theorem 2.39 (Bezout's Theorem) *Let $a, b \in \mathbb{Z}$ be nonzero, and let $d = \gcd(a, b)$.*

1. *d is a linear combination of a and b.*
2. *d the smallest positive integer which is a linear combination of a and b.*
3. *The set of integers which can be written as a linear combination of a and b coincides with the set of multiples of d. In set language, this statement can be phrased as*

$$a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z}.$$

Before proving this, let's see how this works in an example:

Example: Write $\gcd(945, 672)$ as a linear combination of 945 and 672.

Solution: From above (but in reverse order),

$$126 = 6 \cdot 21 + 0$$

$$273 = 2 \cdot 126 + 21$$

$$672 = 2 \cdot 273 + 126$$

$$945 = 1 \cdot 672 + 273$$

PROOF We begin by proving that d is a linear combination of a and b. First, perform the Euclidean algorithm on a and b (WLOG $0 < a < b$). Notice that the sequence $r_1, r_2, r_3, ...$ obtained below is a decreasing sequence of nonnegative integers, so this sequence must be finite by the WOP, i.e. the Euclidean algorithm terminates after some number (say N) of steps:

$$b = aq_1 + r_1$$

$$a = r_1q_2 + r_2$$

$$r_1 = r_2q_3 + r_3$$

$$\vdots$$

$$r_{N-2} = r_{N-1}q_N + r_N$$

$$r_{N-1} = r_Nq_{N+1} + d$$

$$r_N = dq_{N+1} + 0$$

To prove (2), note that from (1), d is a linear combination of a and b. Suppose $c > 0$ is a linear combination of a and b. But since d divides both a and b, d divides c by Theorem 2.24, so $d \,|\, c$. Since d and c are positive, $d \leq c$. This makes d the smallest positive linear combination of a and b, as wanted.

(3) is a set equality proof:

($\subseteq$) Let $x \in a\mathbb{Z} + b\mathbb{Z}$. That means $\exists\, m, n \in \mathbb{Z}$ such that $x = am + bn$, i.e. x is a linear combination of a and b. That means $d \,|\, x$ by Theorem 2.24, so $x = dl$ for some $l \in \mathbb{Z}$, so $x \in d\mathbb{Z}$ as wanted.

($\supseteq$) Let $x \in d\mathbb{Z}$. That means $\exists\, l \in \mathbb{Z}$ such that $x = dl$. By (1), this means $x = (am + bn)l = a(ml) + b(nl)$ for integers m and n. Therefore $x \in a\mathbb{Z} + b\mathbb{Z}$ as wanted. $\square$

Factorization of integers into primes

Lemma 2.40 *Let $a, p \in \mathbb{Z}$. If p is prime and $p \nmid a$, then $\gcd(a, p) = 1$.*

PROOF HW

Lemma 2.41 (Prime Divisor Lemma) *Let $p \in \mathbb{Z}$ be prime. If $p \mid a_1 a_2 a_3 \cdots a_n$, then there is some j such that $p \mid a_j$.*

PROOF The proof is by induction on n.

Base case: Suppose $p \mid a_1 a_2$. If $p \mid a_1$, we are done, so we suppose p does not divide a_1. By Lemma 2.40, that means $\gcd(a_1, p) = 1$.

Induction step: Suppose the result is true when $n = k$ and suppose

$$p \mid a_1 a_2 \cdots a_k a_{k+1}$$

From the base case, $p \mid a_1 \cdots a_k$ or $p \mid a_{k+1}$. In the first situation, p divides some a_j where $j \leq k$ by the IH; in the second situation, p divides a_{k+1}. Either way, by induction we are done. □

A more general version of the Prime Divisor Lemma is this statement:

Corollary 2.42 *Let $a, b, c \in \mathbb{Z}$. If $a \mid bc$ and $\gcd(a, b) = 1$, then $a \mid c$.*

PROOF HW

Theorem 2.43 (Fundamental Theorem of Arithmetic (FTAm)) *Let n be an integer which is greater than 1. Then there exist (finitely many) positive prime numbers $p_1, ..., p_k$ such that $n = p_1 p_2 \cdots p_k$. Furthermore, the p_j are unique (except perhaps for the order in which they are written).*

PROOF WLOG $n \geq 0$. First, we show the existence of the p_js. Let E be the set of natural numbers greater than 1 which do not have a factorization into positive primes. Suppose $E \neq \emptyset$.

Second, we show the uniqueness of the factorization. Suppose there are two different prime factorizations, i.e.

$$n = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s$$

where all the p_j and q_j are positive primes. Observe that

Remark: Why isn't 1 a prime number?

2.6 Congruence classes modulo n

Arguments related to division with remainder are made easier by means of what is called **modular arithmetic**. Essentially, this means taking the integers, divvying them up into sets based on what their remainder is when divided by n, and performing arithmetic on these remainders in a natural way. Here are the formal ideas:

Definition 2.44 *Let $n \in \mathbb{Z}$ be nonzero. Define the relation of* **equivalence modulo** *n, temporarily denoted $\equiv_n$, on $\mathbb{Z}$ by*

$$a \equiv_n b \quad \Leftrightarrow \quad n \,|\, (b-a).$$

The phrase "$a \equiv b \mod n$" is the typical way we write $a \equiv_n b$.

Example: $273 \equiv 48 \mod 5$ because $273 - 48 = 225$ is a multiple of 5.

Example: $17 \not\equiv 26 \mod 2$ because $2 \not| (17-26)$.

Theorem 2.45 *For any nonzero $n \in \mathbb{Z}$, $\equiv_n$ is an equivalence relation.*

PROOF Fix $n \neq 0$. First, we show $\equiv_n$ is reflexive:

Next, we show $\equiv_n$ is symmetric:

Last, we show $\equiv_n$ is transitive:

Question: Every equivalence relation on a space E partitions E into equivalence classes (recall that the equivalence class of $x \in E$ is the set of things in E that are related to x under the equivalence relation). With that in mind, a natural question is to ask what the equivalence classes are under the relation $\equiv_n$?

Example: Describe the equivalence classes under $\equiv_2$:

Example: Describe the equivalence classes under $\equiv_5$:

Theorem 2.46 *Let $n \neq 0$. The equivalence class of $a \in \mathbb{Z}$ under congruence modulo n is the set*

$$a + n\mathbb{Z} = \{a + nk : k \in \mathbb{Z}\}.$$

Any such set $a+n\mathbb{Z}$ is called a **coset (modulo n)**. *The set of equivalence classes under congruence modulo n is called a* **quotient space (modulo n)** *and denoted $\mathbb{Z}/n\mathbb{Z}$ (this is pronounced "$\mathbb{Z}$ mod $n\mathbb{Z}$").*

Some people denote the set $\mathbb{Z}/n\mathbb{Z}$ by $\mathbb{Z}_n$. **DO NOT DO THIS.**

Note: A coset $a + n\mathbb{Z}$ is a set of integers.

Example: $5 + 3\mathbb{Z} =$

Example: $7 + 4\mathbb{Z} = 3 + 4\mathbb{Z} = 115 + 4\mathbb{Z}$.

Example: $\mathbb{Z}/3\mathbb{Z} = \{0 + 3\mathbb{Z}, 1 + 3\mathbb{Z}, 2 + 3\mathbb{Z}\}$.

Here is what is interesting: you can define addition and multiplication in a natural way on $\mathbb{Z}/n\mathbb{Z}$, and these operations make $\mathbb{Z}/n\mathbb{Z}$ into a ring!

Definition 2.47 *Define the following binary operations on $\mathbb{Z}/n\mathbb{Z}$:*

Addition: $(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a + b) + n\mathbb{Z}$

Multiplication: $(a + n\mathbb{Z})(b + n\mathbb{Z}) = (ab) + n\mathbb{Z}$

WARNING: There is possibly a major problem with this. Let's take the operation of addition as defined above. Allegedly, I'm telling you how to add two cosets together and get another coset. For example, in $\mathbb{Z}/9\mathbb{Z}$, this definition means

$$(3 + 9\mathbb{Z}) + (4 + 9\mathbb{Z}) =$$

BUT! $3 + 9\mathbb{Z}$ is the same set as

and $4 + 9\mathbb{Z}$ is the same set as

so $(3 + 9\mathbb{Z}) + (4 + 9\mathbb{Z})$ had better be the same thing as $(21 + 9\mathbb{Z})$ + $(-5 + 9\mathbb{Z})$, otherwise this addition doesn't make sense.

Definition 2.48 *A function or operation is called* **well-defined** *if equal inputs produce equal outputs. Otherwise, it is called* **ill-defined**, *and it really isn't a function or operation at all.*

Theorem 2.49 *Addition and multiplication are well-defined on $\mathbb{Z}/n\mathbb{Z}$.*

PROOF First, we'll show addition is well-defined. Suppose we have equal inputs to addition: let $a_1 + n\mathbb{Z} = a_2 + n\mathbb{Z}$ and $b_1 + n\mathbb{Z} = b_2 + n\mathbb{Z}$. To show addition is well-defined, we need to show that the outputs are equal, i.e.

$$(a_1 + b_1) + n\mathbb{Z} = (a_2 + b_2) + n\mathbb{Z}.$$

From the definition of congruence, $n \mid (a_2 - a_1)$ and $n \mid (b_2 - b_1)$, i.e. there are integers k and l such that

$$nk = a_2 - a_1 \quad \text{and} \quad nl = b_2 - b_1.$$

Therefore

$$(a_2 + b_2) - (a_1 + b_1) = (a_2 - a_1) + (b_2 - b_1) = nk - nl = n(k - 1)$$

so

$$n \mid [(a_2 + b_2) - (a_1 + b_1)] .$$

Therefore $(a_1 + b_1) + n\mathbb{Z} = (a_2 + b_2) + n\mathbb{Z}$, so addition on $\mathbb{Z}/n\mathbb{Z}$ is well-defined.

Next, we show multiplication is well-defined. Suppose a_1, a_2, b_1, b_2, n, k and l are as above. Then

$a_2b_2 - a_1b_1 =$

Therefore $a_1b_1 + n\mathbb{Z} = a_2b_2 + n\mathbb{Z}$, so multiplication on $\mathbb{Z}/n\mathbb{Z}$ is well-defined. □

Theorem 2.50 *$(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ is a ring. In particular, the additive identity is $0 + n\mathbb{Z}$, the multiplicative identity is $1 + n\mathbb{Z}$ and the additive inverse of $a + n\mathbb{Z}$ is $-a + n\mathbb{Z}$.*

PROOF To prove this, you need to check all the properties in the definition of ring (Definition 2.8). Essentially, all this follows from the fact that $\mathbb{Z}$ is a ring and that the operations on $\mathbb{Z}/n\mathbb{Z}$ are inherited from those on $\mathbb{Z}$. I won't go through this in class because it's long and dull. □

Some example computations	Shorthand notation for the same computation
$(18 + 25\mathbb{Z}) + (13 + 25\mathbb{Z}) = 6 + 25\mathbb{Z}$	
$(10 + 8\mathbb{Z})(3 + 8\mathbb{Z}) = 6 + 8\mathbb{Z}$	
$(2 + 5\mathbb{Z}) - (4 + 5\mathbb{Z}) = 3 + 5\mathbb{Z}$	

You can use the shorthand notation on the right-hand side of the chart on the previous page, but keep in mind that what you really mean is the statement on the left involving the addition/multiplication of cosets. The advantage of the shorthand notation is that we can write things like

Another result that is easier to formulate with " $\mod$ " notation is the following fact. Many Math 130 and Math 220 students believe that every function (not just multiplication) distributes, i.e. that

$$(x+y)^2 = x^2 + y^2 \qquad \sin(a-b) = \sin a - \sin b \qquad \sqrt{x+4} = \sqrt{x} + 2 \qquad \text{etc.}$$

This common mistake is known in math professor circles as the *Freshman's Dream*.

If only these students were taking Math 420: in $\mathbb{Z}/p\mathbb{Z}$, the Freshman's Dream comes true!

Lemma 2.51 (Freshman's Dream) *Let p be prime. Then $(a+b)^p \equiv a^p + b^p \mod p$.*

PROOF HW (to prove this, you will prove lots of other stuff which is as important, if not more important, than the Freshman's Dream itself)

Addition and multiplication tables in $\mathbb{Z}/n\mathbb{Z}$

To get some intuition for what follows, let's construct addition and multiplication tables for $\mathbb{Z}/6\mathbb{Z}$ and $\mathbb{Z}/7\mathbb{Z}$ (in these tables, we represent the coset $a + n\mathbb{Z}$ with its unique element in $\{0, ..., n-1\}$):

$(\mathbb{Z}/6\mathbb{Z}, +)$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1					
2	2					
3	3					
4	4					
5	5					

$(\mathbb{Z}/6\mathbb{Z}, \times)$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2				
3	0	3				
4	0	4				
5	0	5				

$(\mathbb{Z}/7\mathbb{Z}, +)$	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	2	3	4	5	6	0
2	2	3	4	5	6	0	1
3	3	4	5	6	0	1	2
4	4	5	6	0	1	2	3
5	5	6	0	1	2	3	4
6	6	0	1	2	3	4	5

$(\mathbb{Z}/7\mathbb{Z}, \times)$	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2					
3	0	3					
4	0	4					
5	0	5					
6	0	6					

Observations:

Zero divisors in $\mathbb{Z}/n\mathbb{Z}$; integral domains

Definition 2.52 *Let R be a ring. $a \in R$ is called a* **zero divisor** *if $a \neq 0$, and there exists $b \neq 0$ in R such that $ab = 0$. R is called an* **integral domain** *if it has no zero divisors.*

Example: $(3 + 6\mathbb{Z})(2 + 6\mathbb{Z}) =$

Theorem 2.53 *A coset $a + n\mathbb{Z}$ is a zero divisor in $\mathbb{Z}/n\mathbb{Z}$ if and only if $\gcd(a, n) > 1$.*

PROOF ($\Rightarrow$) Suppose $a + n\mathbb{Z}$ is a zero divisor. That means there is nonzero $b + n\mathbb{Z}$ such that

$$(a + n\mathbb{Z})(b + n\mathbb{Z}) = 0 + n\mathbb{Z}, \text{ i.e. } n \mid ab.$$

Now suppose not, i.e. $\gcd(a, n) = 1$. By Bezout's Theorem, there are $k, l \in \mathbb{Z}$ such that

$$1 = ka + ln.$$

Multiply through by b to get

$$b = kab + lnb.$$

Notice that n divides the right-hand side of this, so $n \mid b$, i.e. $b + n\mathbb{Z}$ isn't nonzero, a contradiction.

($\Leftarrow$) Suppose $d = \gcd(a, n) > 1$. Since $d \mid a$, write $a = kd$ for $k \in \mathbb{Z}$. Since $d \mid n$, $\frac{n}{d}$ is an integer, so we can compute

$$(a + n\mathbb{Z})(\frac{n}{d} + n\mathbb{Z}) = \frac{an}{d} + n\mathbb{Z} = kn + n\mathbb{Z} = 0 + n\mathbb{Z}.$$

Since $d > 1$, $\frac{n}{d} + n\mathbb{Z}$ is not the zero coset, so $a + n\mathbb{Z}$ is a zero divisor as wanted. □

Units in $\mathbb{Z}/n\mathbb{Z}$; cancellation laws

Recall that a unit of a ring is an element which divides 1 (in the setting of $\mathbb{Z}/n\mathbb{Z}$, "1" means $1 + n\mathbb{Z}$).

Lemma 2.54 (Group properties of the set of units) *Let R be a ring with multiplicative identity 1. Let $x, y \in R$.*

1. *1 is a unit of R.*
2. *If x and y are units, then xy is also a unit.*
3. *If x is a unit, then x^{-1} is a unit.*

PROOF HW

Theorem 2.55 *A coset $a + n\mathbb{Z}$ is a unit in $\mathbb{Z}/n\mathbb{Z}$ if and only if* $\gcd(a, n) = 1$.

PROOF We use a shortcut biconditional proof.

$$\begin{aligned}\gcd(a, n) = 1 &\Leftrightarrow \exists\, k, l \in \mathbb{Z} \text{ s.t. } 1 = ka + ln \qquad \text{(by Bezout's Theorem)}\\ &\Leftrightarrow \exists k \in \mathbb{Z} \text{ s.t. } (k + n\mathbb{Z})(a + n\mathbb{Z}) = 1 + n\mathbb{Z}\\ &\Leftrightarrow a + n\mathbb{Z} \text{ divides } 1 + n\mathbb{Z}\\ &\Leftrightarrow a \text{ is a unit. } \square\end{aligned}$$

Application: Consider the equation

$$(a + n\mathbb{Z})(x + n\mathbb{Z}) = (ab + n\mathbb{Z}), \text{ i.e. } ax \equiv ab \mod n,$$

where the object is to solve for $x \mod n$. When can you "cancel" the as?

Corollary 2.56 (Cancellation law in $\mathbb{Z}/n\mathbb{Z}$) *If a is a unit in $\mathbb{Z}/n\mathbb{Z}$ (i.e. $\gcd(a, n) = 1$), then*

$$ax \equiv ab \mod n \quad \Rightarrow \quad x \equiv b \mod n$$

Definition 2.57 Euler's totient function, *a.k.a. the* **Euler phi function** ϕ *counts the number of units in $\mathbb{Z}/n\mathbb{Z}$. More precisely, $\phi : \{2, 3, 4, ...\} \to \mathbb{N}$ is defined by*

$$\phi(n) = \#\{x \in \mathbb{Z}/n\mathbb{Z} : x \textit{ is a unit}\} = \#\{x \in \{1, ..., n-1\} : \gcd(x, n) = 1\}.$$

Example: $\phi(7) =$

Example: $\phi(12) =$

$$1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11$$

Example: $\phi(7200) =$

To compute $\phi(n)$ for larger n, use the following properties:

Theorem 2.58 (Properties of the Euler phi function) *Let $\phi : \{2, 3, 4, ...\} \to \mathbb{N}$ be the Euler phi function. Then:*

1. *If p is prime, then $\phi(p) = p - 1$.*
2. *If p is prime, then for any $n \geq 1$, $\phi(p^n) = p^n - p^{n-1} = p^{n-1}(p - 1)$.*
3. *If m and n are relatively prime, the $\phi(mn) = \phi(m)\phi(n)$.*

PROOF We begin with the second statement. Notice that $x \in \{1, ..., p^n - 1\}$ is relatively prime to p^n if and only if $p \not| x$. So the only elements of $\{1, ..., p^n - 1\}$ not relatively prime to p^n are the multiples of p. These are every p^{th} number, and there are $\frac{1}{p}(p^n) = p^{n-1}$ of them. So

$$\phi(p^n) = p^n - p^{n-1} = p^{n-1}(p - 1).$$

The first statement is a special case of the second (when $n = 1$), so it is left to prove statement (3). We will postpone this until Chapter 5. □

Back to the example:

$$\phi(7200) = \phi(5^2 3^2 2^5) = \phi(5^2)\phi(3^2)\phi(2^5) = 5(5-1)\,3(3-1)\,2^4(2-1) = 20(6)(16) = 1920.$$

Corollary 2.59 *Let $n \in \mathbb{Z}$ be nonzero. The ring $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is prime (if and only if $\phi(n) = n - 1$).*

PROOF We know $\mathbb{Z}/n\mathbb{Z}$ is a ring; it is a field if and only if all of its nonzero elements have reciprocals if and only if all its nonzero elements are units. By the previous theorem, this holds if $\gcd(a, n) = 1$ for all $a \in \{1, 2, ..., n - 1\}$, which holds exactly when n is prime. □

With this in mind, if p is prime we denote $\mathbb{Z}/p\mathbb{Z}$ by $\mathbb{F}_p$ to emphasize that this set is a field. (Use this notation only if you are messing with both the addition and multiplication operations on the set; if you only need addition, call it $\mathbb{Z}/p\mathbb{Z}$).

Applications of modular arithmetic

Calendar / time problems

Example: Suppose it is $10:00$ AM right now. What time will it be in 205 hours?

Example: August 29, 2018 falls on a Wednesday. What day of the week will August 29, 2019 fall on?

Bar code problems

Universal Product Code (UPC) symbols are now found on most products in grocery and department stores. The UPC symbol is a 12-digit code identifying the product and its manufacturer. The first 11 digits contain information about the product, but the last digit, called the *check digit*, is sued for error detection.

If $d_1 d_2 \cdots d_{12}$ is a valid UPC, then

$$3d_1 + d_2 + 3d_3 + d_4 + ... + 3d_{11} + d_{12} \equiv 0 \mod 10.$$

Determine the check digit for a UPC which begins

$$0 - 49000 - 05014 - \_\_$$

Divisibility tests in $\mathbb{Z}$

Question: How can you tell if an integer is divisible by 3?

For example, does 3 divide 725163471?

Theorem 2.60 *Let $n \in \mathbb{Z}$.*
- *$3 \mid n$ if and only if the sum of the digits of n is divisible by 3.*
- *$9 \mid n$ if and only if the sum of the digits of n is divisible by 9.*
- *$11 \mid n$ if and only if the alternating sum of the digits of n is divisible by 11.*

PROOF Let the digits in the base 10 representation of n be $a_k a_{k-1} a_{k-1} \cdots a_3 a_2 a_1 a_0$. Denote the sum of these digits by $S(n) = a_k + ... + a_0$. Using what we mean by "base 10 representation", we see

$$n = 10^k a_k + 10^{k-1} a_{k-1} + ... + 100a^2 + 10a^1 + a_0 = \sum_{j=0}^{k} 10^k a^k.$$

Note that $10 \equiv 1 \mod 3$, so

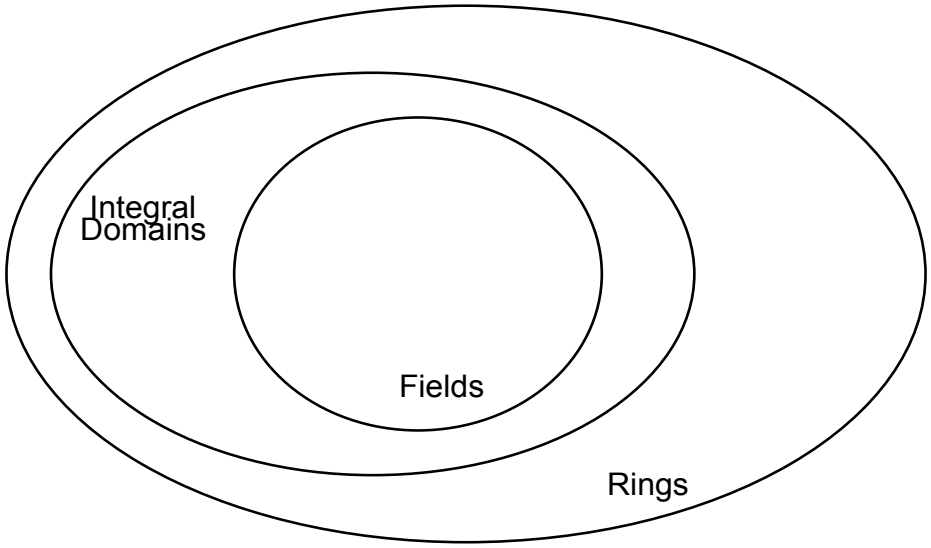
$$n = \sum_{j=0}^{k} 10^j a^j \equiv \sum_{j=0}^{k} 1^j a^j \equiv S(n) \mod 3.$$

Therefore 3 divides n if and only if 3 divides $S(n)$, as wanted. (The same proof as above works for divisibility by 9, since $10 \equiv 1 \mod 9$.)

The last statement is left as a HW problem. □

Objects we have studied so far

So far, we have seen a variety of types of sets with binary operations of addition and multiplication on them:



Question: What other examples of fields / integral domains / rings are there? Might those other examples be useful for something (like say, figuring out whether or not a regular n-gon is constructible or whether or not an equation is solvable by radicals)?

Question: What other categories of algebraic objects are there? How might those be useful?

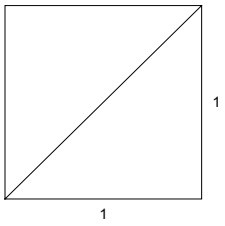
Question: We now know that for every prime p, there is a field with cardinality p (namely $\mathbb{Z}/p\mathbb{Z}$). If n isn't prime, is there a field with cardinality n? If so, what is it? (It can't be $\mathbb{Z}/n\mathbb{Z}$, because that isn't a field.) If not, why not?

Chapter 3

Real and complex numbers

3.1 Theorems of Hippasus and Theatitus

The ancient Greeks (at least some of them) believed that all numbers were rational. This is in part because they thought of numbers as ratios between lengths. They ran into a problem when they tried to find the length of a diagonal of a square of length 1:



A Greek mathematician named Hippasus discovered the following theorem:

Theorem 3.1 (Hippasus' Theorem) *There does not exist any rational number x such that $x^2 = 2$.*

PROOF Suppose not, i.e. that $x \in \mathbb{Q}$ is such that $x^2 = 2$. Write $x = \frac{a}{b}$ where $a, b \in \mathbb{Z}$ and $b \neq 0$. Rearranging the equation $2 = x^2 = \left(\frac{a}{b}\right)^2$, we obtain

$$a^2 = 2b^2$$

This contradicts the (uniqueness part of the) FTAm. □

Hippasus showed this proof to his Greek mathematician friends onboard a ship. His friends were so upset at him for proving this that they threw him overboard (he drowned).

Theorem 3.2 (Theatitus' Theorem) *Let $n \in \mathbb{N}$. If $\sqrt{n} \in \mathbb{Q}$, then $\sqrt{n} \in \mathbb{N}$.*

PROOF Suppose $\sqrt{n} \in \mathbb{Q}$. That means $\sqrt{n} = \frac{p}{q}$ for $p, q \in \mathbb{Z}$ with $\gcd(p, q) = 1$. Rearranging $\sqrt{n} = \frac{p}{q}$, we get

Now by Bezout's Theorem, there are integers k, l such that

$$1 = kp + lq$$

Multiply through by $\sqrt{n}$ to get

$$\sqrt{n} = r\sqrt{n}\,p + l\sqrt{n}\,q$$

$$=$$

Application: $\sqrt{73}$ is irrational by Theatitus' Theorem, because $8 < \sqrt{73} < 9$.

Definition 3.3 *Let f be a function taking values in ring R. A* **root** *of f is an element $x \in Dom(f)$ such that $f(x) = 0$.*

Theorem 3.4 (Rational Roots Theorem) *Let*

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0$$

be a polynomial with integer coefficients. If $x_0 \in \mathbb{Q}$ is a root of f, then $x_0 = \frac{p}{q}$ in lowest terms only if $p \mid a_0$ and $q \mid a_n$.

PROOF Suppose x_0 is a rational root of f, written in lowest terms as $\frac{p}{q}$. Then

$$f\left(\frac{p}{q}\right) = a_n \left(\frac{p}{q}\right)^n + a_{n-1} \left(\frac{p}{q}\right)^{n-1} + ... + a_1 \left(\frac{p}{q}\right) + a_0 = 0$$

$$\Rightarrow \qquad a_n p^n + a_{n-1} p^{n-1} q + ... + a_1 p q^{n-1} + a_0 q^n = 0$$

Corollary 3.5 *Let f be a monic polynomial (i.e. its leading coefficient is 1). Any rational root of f must be an integer.*

The Rational Roots Theorem allows us to generalize Theatitus' Theorem to arbitrary roots:

Corollary 3.6 *Let $n \in \mathbb{N}$, and let $d \geq 2$. If $\sqrt[d]{n} \in \mathbb{Q}$, then $\sqrt[d]{n} \in \mathbb{N}$.*

PROOF Apply the previous corollary to $f(x) = x^d - n$.

3.2 $\mathbb{R}$: the real numbers

Notice the way I phrased Hippasus' Theorem in the previous section. Suppose you rephrased it as

"$\sqrt{2}$ is not a rational number."

There is something that is (sort-of) "wrong" with this phrasing. What is it?

This leads to a question: is there such a thing as $\sqrt{2}$? What about $\sqrt{-1}$?

More generally, what exactly is a real number? Why is $\sqrt{2}$ real, but $\sqrt{-1}$ isn't?

A medium-sized discussion of what a real number is

If you draw a picture of the rational numbers, you will see something like this:

In particular, the rational numbers of larger and larger denominator get closer together, but never actually "fill in" the space between the rationals. Thus the rational numbers have "gaps" everywhere. The concept behind the real numbers is that they are the rationals, together with new objects (irrationals) designed to "fill in" the gaps to make a continuum of objects.

Construction # 1: limits of Cauchy sequences

We have seen there is no rational number x such that $x^2 = 2$. On the other hand, consider this *sequence* of rational numbers (which are better and better approximations to $\sqrt{2}$):

$$\{1, 1.4, 1.41, 1.414, 1.4142, 1.41421, 1.414213, ...\}$$

This sequence of numbers is getting closer and closer together. More precisely, if you give me any positive number ϵ, no matter how small ϵ is, I can delete a finite number of terms from the front of this sequence, such that any two of the numbers in the sequence I leave are less than ϵ apart.

(For example, if you give me $\epsilon = .01$, I can delete $1, 1.4$ and 1.41; everything that's left is at most $.001$ apart which is less than ϵ).

Definition 3.7 *A sequence $\{x_n\}$ of rational numbers is called a* **Cauchy sequence** *if, for any $\epsilon > 0$, there is $N \in \mathbb{N}$ such that whenever $m, n \geq N$, $|x_m - x_n| < \epsilon$.*

Essentially, the set of real numbers is the set of things which are limits of Cauchy sequences of rational numbers. For example, $\sqrt{2}$ is the real number corresponding to the Cauchy sequence

$$\{1, 1.4, 1.41, 1.414, 1.4142, 1.41421, 1.414213, ...\}$$

ENRICHMENT: To make this precise, you define an equivalence relation on the set of all Cauchy sequences of rational numbers:

$$\{x_n\} \sim \{y_n\} \Leftrightarrow \forall \epsilon > 0, \exists N \in \mathbb{N} \text{ s.t. } n \geq N \Rightarrow |x_n - y_n| < \epsilon$$

Essentially, if two sequences are "converging to the same thing", they will be equivalent under the relation $\sim$. Formally, we define a *real number* to be a $\sim$ −equivalence class.

Construction # 2: Dedekind cuts

A way of thinking about an irrational, but real number, is by means of a number line. Take the rational numbers, put them on a number line, and think about "inserting" an irrational number λ somewhere on this line:

You can describe this "inserting" procedure by saying that you have partitioned $\mathbb{Q}$ into two sets L_λ and R_λ with these properties:

1. Every member of L_λ is less than every member of R_λ;
2. L_λ has no greatest member; and
3. R_λ has no least member.

If you drop the third criterion above, you can account for rational numbers as well, by associating the rational number $\frac{p}{q}$ to the partition $L_{p/q} = (-\infty, \frac{p}{q}) \cap \mathbb{Q}$ and $R_{p/q} = [\frac{p}{q}, \infty) \cap \mathbb{Q}$. Thus a *real number* λ can be defined as a partition of $\mathbb{Q}$ into nonempty-sets L_λ and R_λ such that

1. Every member of L_λ is less than every member of R_λ; and
2. L_λ has no greatest member.

What's good about $\mathbb{R}$?

Irrespective of which construction you use, you get the "same object": the set of real numbers, denoted $\mathbb{R}$. This is a set with lots of useful properties.

Algebraic properties of $\mathbb{R}$

Theorem 3.8 *$\mathbb{R}$, with the usual operations of $+$ and $\cdot$, is a field.*

ENRICHMENT: The operations $+$ and $\cdot$ need formal definitions. If you define a real number using equivalence classes of Cauchy sequences of rationals, then the sum of the real number corresponding to the Cauchy sequence $\{x_n\}$ and the real number corresponding to $\{y_n\}$ is the real number corresponding to $\{x_n + y_n\}$ (multiplication is similar). You have to show these operations are well-defined, and that they obey the field laws, but they do.

If you define a real number using Dedekind cuts, then the sum of the real number corresponding to the partition $\{L_\lambda, R_\lambda\}$ and the real number corresponding to the partition $\{L_\mu, R_\mu\}$ is the partition $\{L_\lambda + L\mu, R_\lambda + R_\mu\}$. Here, you have to show that the sum of two real numbers is actually a real number, and that the field laws are obeyed (Stillwell does some of this in his textbook).

Theorem 3.9 *$\mathbb{Q}$ is a subfield of $\mathbb{R}$.*

ENRICHMENT: Technically, what we mean here is that there is a $1-1$ map $i : \mathbb{Q} \to \mathbb{R}$ preserving $+$ and $\cdot$, and it is really $i(\mathbb{Q})$ that is a subfield of $\mathbb{R}$. This is analogous to how we say $\mathbb{Z}$ is a subring of $\mathbb{Q}$.

Definition 3.10 *An* **irrational number** *is a real number which is not rational.*

Theorem 3.11 *$\mathbb{R}$ is totally ordered under $\leq$.*

Furthermore, the ordering $\leq$ is respected by the field operations (i.e. if $a \leq b$, $a + c \leq b + c$ and if $c \geq 0$, $a \leq b$ implies $ac \leq bc$, etc.).

Topological properties of $\mathbb{R}$

First, a statement that says that both rational numbers and irrational numbers are "virtually everywhere" in the real numbers:

Theorem 3.12 (Density Theorem) *For every $a, b \in \mathbb{R}$ with $a < b$, there exists $x \in \mathbb{Q}$ such that $a < x < b$ and $y \in \mathbb{R} - \mathbb{Q}$ such that $a < y < b$.*

PROOF Take Math 430.

Theorem 3.13 (Properties of absolute value) *Let $x, y \in \mathbb{R}$. The absolute value function $|\ | : \mathbb{R} \to \mathbb{R}$, defined by*

$$|x| = \sqrt{x^2} = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases},$$

has the following properties:

Positivity: $|x| \geq 0$.
Definiteness: $|x| = 0$ *only if* $x = 0$.
Symmetry: $|-x| = |x|$.
Triangle Inequality: $|x + y| \leq |x| + |y|$.

PROOF All of these are obvious except the triangle inequality, which is left as a homework problem (in Math 324, we did this with six cases, but there is an easier method which takes advantage of the fact that $|x| = \sqrt{x^2}$).

Thus if we define the **distance** between x and y to be $|x - y|$, this notion of distance makes $\mathbb{R}$ into what is called a **metric space**. Among other things, this allows us to define open and closed subsets of $\mathbb{R}$, gives us a mechanism to define limits, and lets us say what it means for a function $f : \mathbb{R} \to \mathbb{R}$ to be **continuous**. All these concepts are within the branch of mathematics called **topology**.

For our purposes (this isn't topology class), a continuous function $f : \mathbb{R} \to \mathbb{R}$ is one whose graph has no "gaps", or one that satisfies

$$\lim_{x \to c} f(x) = f(c)$$

for every c in the domain of f, just as it was in Math 220. All we need in Math 420 is the following fact, which is proven in Math 430 or in a topology course:

Theorem 3.14 *Every polynomial function $f : \mathbb{R} \to \mathbb{R}$ is continuous. Every rational function $f : \mathbb{R} \to \mathbb{R}$ is continuous, except at points which make its denominator zero.*

Now for an important theorem, sometimes introduced in Calculus 1 (but not by me):

Theorem 3.15 (Intermediate Value Theorem (IVT)) *If $f : \mathbb{R} \to \mathbb{R}$ is continuous on $[a, b]$, then for every z between $f(a)$ and $f(b)$, there is an $c \in [a, b]$ such that $f(c) = z$.*

PROOF Take Math 430. (This theorem will seem obvious after I draw the picture, but it is deep, and relies upon the rigorous construction of real numbers as Dedekind cuts or equivalence classes of Cauchy sequences of rationals.)

Picture:

Here is an important application of the IVT. Among other things, after proving this we could rephrase Hippasus' Theorem as "$\sqrt{2} \notin \mathbb{Q}$".

Corollary 3.16 (Existence of n^{th} roots) *Let $\lambda > 0$ be a real number. Then, for every nonzero $n \in \mathbb{N}$, there is a real number $c > 0$ s.t. $c^n = \lambda$ (i.e. $c = \sqrt[n]{\lambda}$ exists).*

PROOF Let $f(x) = x^n - \lambda$. f is a polynomial, hence continuous. Now

$$f(0) = -\lambda < 0$$

and

$$f(1 + \lambda) = (1 + \lambda)^n - \lambda > 0 \quad \text{(HW)}$$

so by the IVT, there is $c \in [a, b]$ s.t. $f(c) = 0$, i.e. $c^n - \lambda = 0$, i.e. $c^n = \lambda$. $\square$

Consequence: these are all real numbers:

$$\sqrt{17},\ \sqrt[5]{2},\ \sqrt[9]{23857} + \sqrt[3]{234},\ 6\left(\frac{\sqrt{7}}{\sqrt[3]{5} + \sqrt[8]{435}} + 34\sqrt[17]{35}\right)^{7/4}.$$

Question: Are expressions that can be made from $+$, $-$, $\times$, $\div$ and radicals <u>all</u> the real numbers? Can π be written this way, for example?

Theorem 3.17 (Extreme Value Theorem (EVT)) *If $f : \mathbb{R} \to \mathbb{R}$ is continuous on $[a, b]$, then f has an absolute maximum and/or absolute minimum on $[a, b]$.*

PROOF Take Math 430.

Picture:

What's bad about $\mathbb{R}$?

You can make relatively simple (i.e. polynomial) equations out of real numbers that have no solution:

Lemma 3.18 *The equation $x^2 + 1 = 0$ has no real solutions.*

PROOF (that uses calculus) Let $f(x) = x^2 + 1$. $f'(x) = 2x$, so f decreases when $x < 0$, increases when $x > 0$, and has absolute minimum when $x = 0$. But $f(0) = 1$, so $f(x) \geq 1$ for all x. $\square$

3.3 $\mathbb{C}$: complex numbers

Definition 3.19 *A* **complex number** *is an object of the form $z = x + iy$ where $x, y \in \mathbb{R}$ (for now, i is just a symbol... this symbol will be given meaning later). The set of complex numbers is denoted $\mathbb{C}$.*

Definition 3.20 *Given a complex number $z = x + iy$, the* **real part** *of z, denoted $\Re(z)$ or $Re(z)$, is x, and the* **imaginary part** *of z, denoted $\Im(z)$ or $Im(z)$, is y.*

A complex number z is called **pure imaginary** *if $\Re(z) = 0$; a complex number z is* **real** *if $\Im(z) = 0$.*

Note: for $z \in \mathbb{C}$, $\Re(z)$ and $\Im(z)$ are real numbers. For example:

$$\Re(2 + 5i) = 2 \qquad \Im(-1 - 4i) = -4$$

Examples: $\Re(7 - 4i) = 7$; $\Im(-6 + 3i) = 3$; $-3i$ is pure imaginary.

Denoting a complex number: Usually a complex number is denoted by $z, w,$ or a Greek letter like ζ (zeta) or ξ (xi) or ω (omega). Try not to use s, t, u, v, x and/or y to denote a complex numbers; these letters connote real numbers. In particular it is always understood with complex numbers that "z" means the complex number $z = x + iy$.

Remark: in general you want to avoid immediately thinking of a complex number as $x + iy$. Just think of it as z.

WARNING: $\mathbb{C}$ is not ordered in any meaningful sense (this is what is bad about $\mathbb{C}$). Don't ever talk about one complex number being $\leq$ another.

Arithmetic in $\mathbb{C}$

Addition and subtraction in $\mathbb{C}$ are defined by combining like terms. For example,

$$(2-3i)+(1+i)=3-2i \quad \text{and} \quad (-3+i)-(2+7i)=-5-6i.$$

Multiplication is defined by distributing terms, together with the law that $i^2=-1$ (this is the first time we need the idea that $i=\sqrt{-1}$). For example:

$$(2+5i)(-1-2i)=-2-5i-4i-10i^2=-2-9i+10=8-9i.$$

Division is trickier; to divide one complex number by a nonzero complex number, what you do is multiply through the numerator and denominator of the fraction by the conjugate of the denominator. An example:

$$(1+i)\div(3-4i)=\frac{1+i}{3-4i}=\frac{(1+i)(3+4i)}{(3-4i)(3+4i)}=\frac{-1+7i}{25}=\frac{-1}{25}+\frac{7}{25}i.$$

$$\frac{2-3i}{7+i}=$$

$$\frac{1}{1-4i}=$$

Theorem 3.21 *$\mathbb{C}$, with the addition and multiplication defined above, is a field. In particular:*

- *the additive identity element is $0=0+0i$;*
- *the multiplicative identity element is $1=1+0i$;*
- *the additive inverse of $z=x+iy$ is $-z=-x-iy$;*
- *the reciprocal of $z=x+iy$ is $\frac{1}{z}=\frac{1}{x+iy}=\frac{x-iy}{x^2+y^2}=\frac{x}{x^2+y^2}-i\frac{y}{x^2+y^2}$.*

Definition 3.22 *Any subfield of $\mathbb{C}$ is called a* **number field**.

Fields we have seen that are number fields:

Fields we have seen that are not number fields:

The complex numbers have an extra artithmetic operation that is important:

Definition 3.23 *The* **(complex) conjugate** *of $z = x + iy \in \mathbb{C}$ is $\overline{z} = x - iy$.*

Example: If $z = 2 - 7i$, then $\overline{z} = 2 + 7i$.

Lemma 3.24 (Properties of conjugation) *Let $z, w \in \mathbb{C}$.*

1. *Conjugation preserves addition, i.e. $\overline{z + w} = \overline{z} + \overline{w}$.*
2. *Conjugation preserves multiplication, i.e. $\overline{zw} = \overline{z}\,\overline{w}$.*
3. $\Re(z) = \frac{z+\overline{z}}{2}$.
4. $\Im(z) = \frac{z-\overline{z}}{2i}$.

PROOF HW (in each of these, you are write $z = x+iy$ and $w = u+iv$ for $x, y, u, v \in \mathbb{R}$, and work out both sides of the equations you need to verify in terms of x, y, u and v. You will see that they are equal.)

Geometric interpretation of $\mathbb{C}$

There is a natural bijection $\mathbb{C} \to \mathbb{R}^2$ defined by $x + iy \mapsto (x, y)$. So, we can think of complex numbers as being points in a plane, in the same way we think of real numbers as points on a line.

The "$x-$axis" of this plane contains the real numbers, and is called the **real axis**. The "$y-$axis" of this plane is called the **imaginary axis**, and contains the pure imaginary numbers.

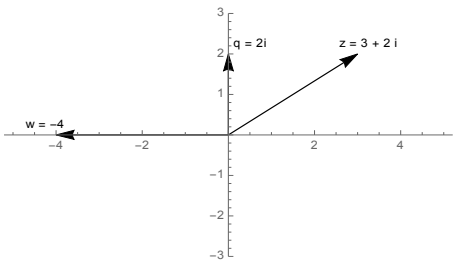
Addition of complex numbers corresponds to "head-to-tail" or "parallelogram" addition of vectors, i.e.

$$(3 + 2i) + (1 - 3i) = 4 - i$$

is essentially the same as the vector addition

$$(3, 2) + (1, -3) = (4, -1).$$

Observe that if we think of z as a vector, then $\overline{z}$ is the vector obtained by reflecting z through the real axis (note that $\overline{z} = z$ if and only if z is real):



To interpret multiplication of complex numbers geometrically, we use polar coordinates:

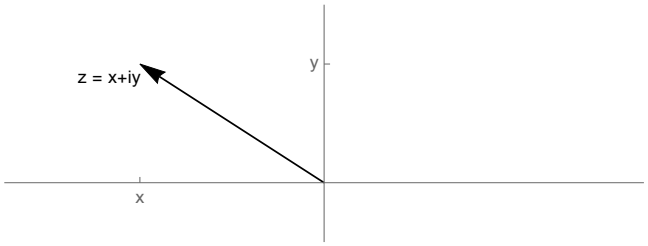
Lemma 3.25 *Let $z \in \mathbb{C}$. Then $z\overline{z} \in \mathbb{R}$, and $z\overline{z} \geq 0$.*

PROOF Write $z = x + iy$ for $x, y \in \mathbb{R}$. Then

$$z\overline{z} = (x + iy)(x - iy) = x^2 + iyx - iyx - i^2y^2 = x^2 + y^2 \geq 0$$

as wanted. □

Definition 3.26 *The* **absolute value** *a.k.a.* **norm** *a.k.a.* **modulus** *of a complex number $z = x + iy$ is $|z| = \sqrt{z\overline{z}} = \sqrt{x^2 + y^2}$.*



The norm of a complex number is its distance from zero, so the "norm of a complex number" generalizes the notion of "absolute value of a real number".

Another view of division in $\mathbb{C}$: given $z_1, z_2 \in \mathbb{C}$, we have

$$z_1 \div z_2 = \frac{z_1}{z_2} = \frac{z_1\overline{z_2}}{z_2\overline{z_2}} = \frac{z_1\overline{z_2}}{|z_2|^2}.$$

Special case (reciprocals): If $z \neq 0$, then

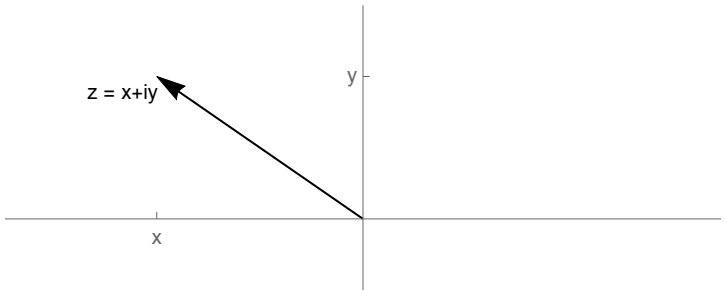
$$z^{-1} = \frac{1}{z} = \frac{\overline{z}}{z\overline{z}} = \frac{\overline{z}}{|z|^2}.$$

In particular, if $|z| = 1$, then $z^{-1} = \overline{z}$ (useful special case: $\frac{1}{i} = i^{-1} = \overline{i} = -i$).

Lemma 3.27 *Let $z_1, z_2 \in \mathbb{C}$. Then $|z_1 z_2| = |z_1||z_2|$.*

PROOF HW

Definition 3.28 *Let $z = x + iy \in \mathbb{C}$. The* **argument** *of z, denoted $\arg(z)$, is any angle θ (in radians) such that $x = |z| \cos\theta$ and $y = |z| \sin\theta$.*

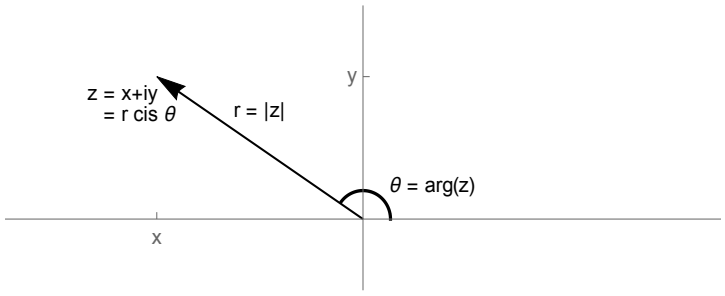


Given z, you can solve for $\theta = \arg z$ by setting $\theta = \arctan\left(\frac{y}{x}\right)$ if $x > 0$; if $x = 0$ then $\theta = \pi/2$ if $y > 0$ and $\theta = -\pi/2$ if $y < 0$. Notice that arguments are only defined up to multiples of 2π.

Definition 3.29 *The* **polar coordinates** *of a complex number z are (r, θ) where $r = |z|$ and $\theta = \arg z$. If the polar coordinates of z are (r, θ), we (temporarily) write*

$$z = r\cos\theta + ir\sin\theta = r(\cos\theta + i\sin\theta)$$

or $z = r\,\mathrm{cis}\,\theta$.



Example: Find the modulus and argument of $z = -5 + 5\sqrt{3}i$, and write z in $r\,\mathrm{cis}\,\theta$ form.

Euler's formula

To define functions like exponentials and trig functions of complex numbers, we use power series (because power series are made up only of addition, subtraction, multiplication and division, and all these operations are already defined for complex numbers).

There is an issue regarding what it means for a series of complex numbers to converge, but it turns out that any power series which converges for all real numbers also converges for all complex numbers.

Definition 3.30 *For any complex number $z \in \mathbb{C}$, define*

$$\begin{aligned} e^z = \exp(z) &= \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2} + \frac{z^3}{3!} + \frac{z^4}{4!} + ... \\ \cos z &= \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!} = 1 - \frac{z^2}{2} + \frac{z^4}{4!} - \frac{z^6}{6!} + ... \\ \sin z &= \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{(2n+1)!} = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \frac{z^7}{7!} + ... \end{aligned}$$

From this, you can show that all the usual trigonometric and exponential identities that hold for real numbers also hold for complex numbers. Amazingly, we have the following amazing identity which links exponential and trigonometric functions:

Theorem 3.31 (Euler's formula) *For any $z \in \mathbb{C}$, $e^{iz} = \cos z + i \sin z$.*

PROOF

$$\begin{aligned} e^{iz} &= \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} \\ &= 1 + iz + \frac{(iz)^2}{2!} + \frac{(iz)^3}{3!} + \frac{(iz)^4}{4!} + ... \\ &= 1 + iz + i^2\frac{z^2}{2!} + i^3\frac{z^3}{3!} + i^4\frac{z}{4!} + ... \\ &= 1 + iz - \frac{z^2}{2!} - i\frac{z^3}{3!} + \frac{z^4}{4!} + i\frac{z^5}{5!} - \frac{z^6}{6!} - i\frac{z^7}{7!} + ... \\ &= \left[1 - \frac{z^2}{2!} + \frac{z^4}{4!} - ...\right] + i\left[z - \frac{z^3}{3!} + \frac{z^5}{5!} - ...\right] \\ &= \cos z + i \sin z. \ \square \end{aligned}$$

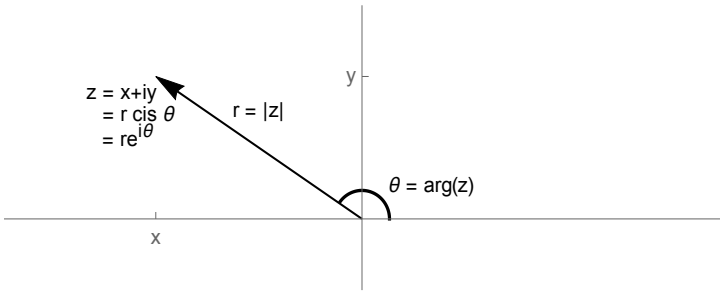
As an important consequence, we see that if z has polar coordinates (r, θ), then

$$z = r \text{ cis } \theta = r(\cos\theta + i\sin\theta) = r\cos\theta + ir\sin\theta = re^{i\theta}.$$

For such a z,

$$\overline{z} = r\cos\theta - ir\sin\theta = r\cos(-\theta) + i\sin(-\theta) = re^{-i\theta}.$$

In particular, if we write $z = re^{i\theta}$ where $r \geq 0$ and $\theta \in \mathbb{R}$, this means (r, θ) are the polar coordinates of z, so $r = |z|$ and $\theta = \arg z$.



Theorem 3.32 *Suppose $z_1 = r_1$ cis θ_1 and $z_2 = r_2$ cis θ_2 (this means $r_1 = |z_1|, r_2 = |z_2|$). Then $z_1 z_2 = r_1 r_2$ cis $(\theta_1 + \theta_2)$.*

PROOF Let $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$, then by elementary properties of exponentials, $z_1 z_2 = r_1 r_2 e^{i(\theta_1+\theta_2)}$. □

This theorem tells us how to interpret multiplication geometrically in $\mathbb{C}$. Given two complex numbers, if those numbers are multiplied, then the "moduli multiply" (since $|z_1 z_2| = |z_1||z_2|$) and the "arguments add" (since this theorem implies $\arg(z_1 z_2) = \arg z_1 + \arg z_2$).

Example: Suppose $z = 6e^{\pi i/3}$ and $w = 2e^{\pi i/4}$. Find zw and $\frac{z}{w}$.

Theorem 3.33 (de Moivre's Theorem) *Suppose $z \in \mathbb{C}$ (assume $z \neq 0$) has polar form $re^{i\theta} = r \operatorname{cis} \theta$. Then, for any $n \in \mathbb{Z}$,*

$$z^n = r^n e^{in\theta} = r^n \operatorname{cis} n\theta.$$

PROOF We start by proving the theorem when $n \in \mathbb{N}$. To do this, we use induction on n. The base case $n = 0$ is obvious since $z^0 = 1 = r^n e^{0i}$. Now, assume the result is true when $n = k$. Then,

$$\begin{aligned} z^{k+1} &= z^k z \\ &= \left(r^k e^{ik\theta}\right)(re^{i\theta}) \qquad \text{(by the IH)} \\ &= r^{k+1} e^{ik\theta + i\theta} \qquad \text{(by Theorem 3.32)} \\ &= r^{k+1} e^{(k+1)i\theta}. \end{aligned}$$

By induction, the theorem holds for $n \in \mathbb{N}$.

Now suppose n is negative.

$$\begin{aligned} z^n z^{-n} &= 1 \\ \left(r^n e^{in\theta}\right) z^{-n} &= 1e^{0i} \end{aligned}$$

By Theorem 3.32, the modulus of z^{-n}, say s, must satisfy $r^n s = 1$. Thus $s = r^{-n}$ as wanted. The argument of z^{-n} must satisfy $n\theta + \arg(z^{-n}) = 0$, so $\arg(z^{-n}) = -n\theta$ as wanted. □

Example: Compute $(3 + 3i)^{13}$, writing your answer in $a + ib$ form.

3.4 Fundamental Theorem of Algebra

Remember that what was bad about $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ was...

You might ask if the same thing is bad about $\mathbb{C}$. This theorem says that with regard to polynomial equations, $\mathbb{C}$ is good:

Theorem 3.34 (Fundamental Theorem of Algebra (FTAl)) *Let f be a polynomial of degree at least 1, whose coefficients are complex numbers. Then f has a root $z_0 \in \mathbb{C}$.*

Interestingly, all known proofs of the FTAl require at least a moderate amount of topology. I have written out a somewhat accessible proof for you here, just so you have seen it, but I won't test you on this, because the material is more closely related to what you would study in topology or real or complex analysis.

PROOF This proof has several steps. Start by letting $f : \mathbb{C} \to \mathbb{C}$ be the non-constant polynomial

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + ... + a_1 z + a_0.$$

Now for two lemmas. The first says that if you are in $\mathbb{C}$ but aren't at a root of a polynomial f, then you can move a little bit in some direction to make the value of f smaller (in norm):

Lemma 1 (d'Alembert's Lemma): If $f(z_0) \neq 0$ for $z_0 \in \mathbb{C}$, then there is $z_1 \in \mathbb{C}$, arbitrarily close to z_0, such that $|f(z_1)| < |f(z_0)|$.

Proof of d'Alembert's Lemma: Let $\Delta z \in \mathbb{C}$. Then

$$\begin{aligned} f(z) = f(z_0 + \Delta z) &= a_n(z_0 + \Delta z)^n + a_{n-1}(z_0 + \Delta z)^{n-1} + ... + a_1(z_0 + \Delta z) + a_0 \\ &= a_n z_0^n + a_{n-1} z_0^{n-1} + ... + a_1 z_0 + a_0 + A_1 \Delta z + A_2 (\Delta z)^2 + ... \end{aligned}$$

$$= f(z_0) + A_1(\Delta z) + \epsilon.$$

By choosing the argument of Δz to be $-\arg f(z_0)$, and choosing the norm of Δz small enough so that $|\epsilon|$, which contains powers of $|\Delta z|$, is far smaller than $|A_1(\Delta z)|$, and by setting $z_1 = z_0 + \Delta z$, we get $f(z_1) = f(z_0 + \Delta z)$ which has smaller norm than $f(z_0)$, as wanted.

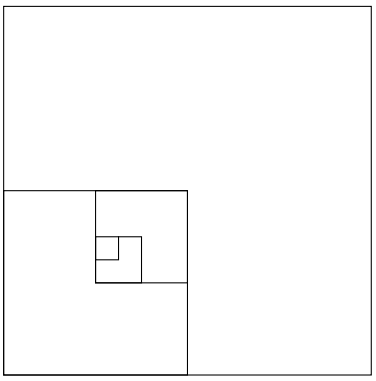
The second lemma says that any continuous function on a closed square achieves its absolute minimum value:

Lemma 2 (Extreme Value Theorem for $\mathbb{C}$): Fix $R > 0$. If $g : \mathbb{C} \to \mathbb{R}$ is continuous, then g has an absolute minimum value on $S = \{z = x + iy : |x| \leq R, |y| \leq R\}$.

Sketch of proof of Extreme Value Theorem for $\mathbb{C}$: Suppose not, i.e. that g is unbounded on the square S. Cut S into four smaller squares; g must be unbounded on at least one of these smaller squares. Let S_1 be a square on which g is unbounded. Repeating this argument, we obtain a sequence of squares

$$S \supseteq S_1 \supseteq S_2 \supseteq S_3 \supseteq \cdots$$

which intersect on a single point $\lambda \in S$. Thus $\{|g(z)|\}$ is unbounded on a small square containing λ, which is absurd since $g(\lambda)$ exists and is continuous at λ.



Back to the proof of the FTAl. For large z, say for $|z| \geq R$,

$$|f(z)| = |a_n z^n + ... + a_0|$$

is dominated by $|a_n z^n|$. hence cannot be small. Now, let z_0 be the location of the minimum value of $|f(z)|$ on $\{z : |z| \leq R\}$ (this exists by the Extreme Value Theorem for $\mathbb{C}$). Clearly, $|f(z_0)| \geq 0$.

However, if $|f(z_0)| > 0$, then by d'Alembert's Lemma, $|f(z_0)|$ isn't actually the minimum, a contradiction. That means $f(z_0) = 0$, as wanted. $\square$

3.5 Complex roots, cubic equations and regular polygons

Recall from an earlier section: de Moivre's Theorem, which says

$$z = r \operatorname{cis} \theta = re^{i\theta} \quad \Rightarrow \quad z^n = r^n \operatorname{cis} n\theta = r^n e^{in\theta}.$$

Earlier, we used this formula to compute powers; now, we use it to compute roots.

Example: Find all complex numbers z such that $z^3 = -4 + 4\sqrt{3}\, i$.

Example: Find all complex numbers z such that $z^6 = 1$.

These examples generalize:

Theorem 3.35 (Roots of complex numbers) *Let $w \in \mathbb{C}$ be nonzero and let $n \in \mathbb{N}$. Then there are exactly n complex numbers z such that $z^n = w$. All of these z have modulus equal to $|w|^{1/n}$, and their arguments come from the set*

$$\left\{ \frac{1}{n} \arg w + \frac{2\pi j}{n} : j \in \{0, 1, ..., n-1\} \right\}.$$

PROOF For $j \in \{0, ..., n-1\}$, let

$$z_j = |w|^{1/n} e^{\left(\frac{1}{n} \arg w + \frac{2\pi}{n} j\right)}.$$

By de Moivre's Theorem, $z_j^n = w$. That there are no other n^{th} roots of w follows from the fact that a polynomial of degree n (such as $z^n - w$) has at most n roots; this will be proved in the next chapter (Corollary 4.12). □

The generic picture that goes with this theorem:

Remember cubic equations?

In Chapter 1, we learned the del Ferro / Tartaglia method of solving a cubic equation, and saw that (as far as we knew then) that said method worked so long as the cubic did not have three real roots. Now, suppose you have a cubic equation $x^3 + px + q = 0$ with three real roots, i.e. the discriminant

$$\Delta = -4p^3 - 27q^2$$

is positive. Recall that the del Ferro / Tartaglia formula gives, as a solution to $x^3 + px + q = 0$,

$$\begin{aligned} x &= \sqrt[3]{\frac{-q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{\frac{-q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ &= \sqrt[3]{\frac{-q}{2} + \sqrt{\frac{-\Delta}{108}}} + \sqrt[3]{\frac{-q}{2} - \sqrt{\frac{-\Delta}{108}}} \\ &= \sqrt[3]{\frac{-q}{2} + i\sqrt{\frac{\Delta}{108}}} + \sqrt[3]{\frac{-q}{2} - i\sqrt{\frac{\Delta}{108}}} \end{aligned}$$

In other words, the del Ferro / Tartaglia method works even if there are three real roots, so long as you are willing to introduce complex numbers along the way!

Example: $x^3 - 3x + \sqrt{2} = 0$

Solution: Here, $p = -3$, $q = \sqrt{2}$ so $\Delta = -4(-27) - 27(2) = 54 > 0$. Thus this equation has three roots. From del Ferro / Tartaglia, we get

$$\begin{aligned} x &= \sqrt[3]{\frac{-q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{\frac{-q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \\ &= \sqrt[3]{\frac{-\sqrt{2}}{2} + \sqrt{\frac{-1}{2}}} + \sqrt[3]{\frac{-\sqrt{2}}{2} - \sqrt{\frac{-1}{2}}} \\ &= \sqrt[3]{\frac{-\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}} + \sqrt[3]{\frac{-\sqrt{2}}{2} - i\frac{\sqrt{2}}{2}} \end{aligned}$$

Examples like this one illustrate why mathematicians initially cared about complex numbers. del Ferro, Cardano, et al. didn't care about finding *imaginary* solutions to *quadratics* like $x^2 + 1 = 0$, but they cared a lot about finding *real* solutions to *cubics* like the example above. Complex numbers were a means to an end (as they are today in differential equations).

Another mathematician named Bombelli in 1572 called these square roots of negative numbers "imaginary numbers". He decided to see what kinds of arithmetic one could do with these "imaginary" numbers, and showed that you could make sense of addition, multiplication, division, powers and roots of them. Imaginary numbers were controversial until 1742, when Euler proved the Fundamental Theorem of Algebra.

Today, we have discovered that complex numbers are not really "imaginary". They describe physical quantities in fluid dynamics, electromagnetism, signal and image processing, quantum mechanics, and special and general relativity.

Back to regular polygons

Recall from Chapter 1:

The regular n-gon is constructible $\Leftrightarrow \cos\frac{2\pi}{n}$ and/or $\sin\frac{2\pi}{n}$ are constructible numbers

$\Leftrightarrow \cos\frac{2\pi}{n}$ and/or $\sin\frac{2\pi}{n}$ are surd (i.e. they are real numbers that can be obtained from integers using only $+, -, \times, \div, \sqrt{}$)

We can rephrase this material in the language of complex numbers.

Definition 3.36 *A complex number $z = x + iy$ is called* **constructible (surd)** *if x and y are surd real numbers (in the sense of Chapter 1). Call the set of surd complex numbers $\mathbb{S}$.*

Since $i = \sqrt{-1}$, a complex number is surd if and only if it can be obtained from the integers using $+, -, \times, \div$ and $\sqrt{}$.

Putting this together, we have:

Theorem 3.37 *The regular n-gon is constructible if and only if $\zeta_n = e^{2\pi i/n}$ is surd.*

Now for some alternate language and notation:

Definition 3.38 *A complex number z is called an n^{th}* **root of unity** *if it satisfies $z^n = 1$. z is called a* **primitive** *n^{th}* **root of unity** *if $z^n = 1$ and $z^k \neq 1$ for any $k < n$.*

Note: All roots of unity lie on the unit circle $|z| = 1$.

Example: The fourth roots of unity are ± 1 and $\pm i$.

By applying de Moivre's Theorem, the n^{th} roots of unity are

$$\left\{ e^{i(2\pi j/n)} = \cos\frac{2\pi j}{n} + i\sin\frac{2\pi j}{n} : 0 \leq j \leq n-1 \right\}.$$

The n^{th} root of unity corresponding to $j = 1$, $e^{2\pi i/n}$, is denoted by by ζ_n.

Theorem 3.39 *The n^{th} roots of unity are $\zeta_n, \zeta_n^2, \zeta_n^3, ..., \zeta_n^{n-1}$ and $\zeta_n^n = 1$.*
An n^{th} root of unity ζ_n^k is primitive if and only if $\gcd(n,k) = 1$.

PROOF The first statement follows directly from de Moivre's theorem.

For the second statement:

$$\begin{aligned}
\zeta_n^k \text{ is a primitive root of unity } &\Leftrightarrow \left(\zeta_n^k\right)^j \neq 1 \text{ for all } j \leq n \\
&\Leftrightarrow \left(e^{2\pi ik/n}\right)^j \neq 1 \text{ for all } j \leq n \\
&\Leftrightarrow jk \not\equiv 0 \mod n \text{ for all } j \leq n \\
&\Leftrightarrow k \text{ is not a zero divisor in } \mathbb{Z}/n\mathbb{Z} \\
&\Leftrightarrow \gcd(n,k) = 1 \ \square.
\end{aligned}$$

Let's suppose we were trying to think of an equation we would solve to find ζ_n. Since $\zeta_n^n = 1$, we know ζ_n is a root of

$$z^n - 1 = 0$$

Recall:

Definition 3.40 *Let p be prime. The* **cyclotomic polynomial** Φ_p *is the polynomial*

$$\Phi_p(z) = z^{p-1} + z^{p-2} + ... + z^2 + z + 1.$$

(The p^{th} roots of unity other than 1 are the roots of this polynomial.)

More generally, for any positive $n \in \mathbb{N}$, define the **cyclotomic polynomial** Φ_n *to be*

$$\Phi_n(z) = \prod_{\{k:\gcd(n,k)=1\}} (z - \zeta_n^k).$$

(The primitive n^{th} roots of unity are therefore the roots of this polynomial.)

It turns out that for any n, Φ_n has integer coefficients.

Conclusion, so far: the following are equivalent:

1. The regular n-gon is constructible.
2. $\cos \frac{2\pi}{n}$ is surd.
3. ζ_n is surd.
4. The roots of Φ_n are surd numbers.

Theorem 3.41 *Let $m, n \in \mathbb{N}$ be such that $\gcd(m, n) = 1$. If the regular m-gon and regular n-gon are constructible, then the regular mn-gon is constructible.*

PROOF HW (Here is the idea: if the regular m- and n-gons are constructible, then ζ_m and ζ_n are surd. These numbers, respectively, satisfy $\Phi_m(z) = 0$ and $\Phi_n(z) = 0$. Now ζ_{mn} satisfies $\Phi_{mn} = 0$; show that Φ_{mn} factors in a particular way and deduce that ζ_{mn} is surd from that factorization.)

Example: $n = 5$ (regular pentagon)

We seek to construct $\zeta_5 = e^{2\pi i/5} = \cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}$, which is a root of

$$\Phi_5(z) = z^4 + z^3 + z^2 + z + 1 = 0.$$

In this example, a trick is useful: divide through Φ_5 by z^2 to get

$$z^2 + z + 1 + \frac{1}{z} + \frac{1}{z^2} = 0$$

and observe that since $|z| = 1$, $\overline{z} = z^{-1}$. This yields the following equation (of which ζ_5 is a solution):

$$z^2 + z + 1 + \overline{z} + \overline{z}^2 = 0. \tag{3.1}$$

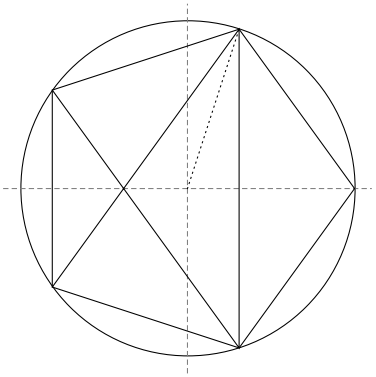
Now let

$$x = 2\Re(z) = z + \overline{z}.$$

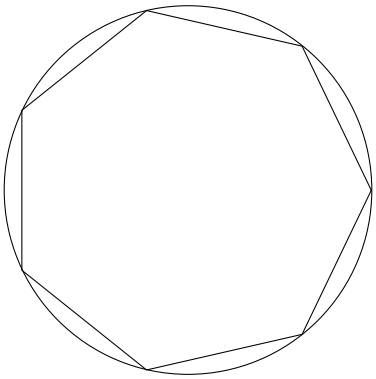
Notice

$$x^2 = (z + \overline{z}^2 = z^2 + 2z\overline{z} + \overline{z}^2 = z^2 + 2|z| + \overline{z}^2 = z^2 + 2 + \overline{z}^2$$

so equation (3.1) becomes



Example: $n = 7$ (regular septagon)



We seek to construct $\zeta_7 = e^{2\pi i/7} = \cos\frac{2\pi}{7} + i\sin\frac{2\pi}{7}$, which is a root of

$$\Phi_7(z) = z^6 + z^5 + z^4 + z^3 + z^2 + z + 1 = 0.$$

Divide through by z^3 to get

$$z^3 + z^2 + z + 1 + \overline{z} + \overline{z}^2 + \overline{z}^3 = 0. \tag{3.2}$$

Again, let $x = z + \overline{z}$; as before $x^2 = z^2 + 2 + \overline{z}^2$ and also,

$$x^3 = z^3 + 3(z + \overline{z}) + \overline{z}^3.$$

Substituting all this into equation (3.2), we get

$$x^3 + x^2 - 2x - 1 = 0. \tag{3.3}$$

Punchline: the regular septagon is constructible if and only if the equation

$$x^3 + x^2 - 2x - 1 = 0$$

has roots which are surd. (In this setting, $x = 2\cos\frac{2\pi}{7}$.)

Chapter 4

Polynomial rings

4.1 Definition and basic properties

Earlier, we asked what rings exist, other than fields, the ring $\mathbb{Z}$ of integers, and $\mathbb{Z}/n\mathbb{Z}$. In this chapter we discuss a new class of rings:

Definition 4.1 *Let F be a field. A* **polynomial** *with coefficients in F is an expression*

$$f = f(x) = a_0 + a_1x + a_2x^2 + a_3x_3 + ... + a_nx^n$$

where $n \in \mathbb{N}$, $a_0, a_1, ..., a_n \in F$, and $a_n \neq 0$. In this setting n is called the **degree** *of f; we write $n = \deg f$ to denote this. The x is called an* **indeterminate** *and an object that is being used only in a formal, place-holding role. The set of polynomials with coefficients in field F is denoted $F[x]$.*

Examples of the F in this definition include $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{F}_p$.

Convention: The degree of the zero polynomial is $-\infty$. To say that a polynomial is **nonzero** means it is not the constant zero polynomial. Thus any nonzero polynomial has degree at least 0, and any non-constant polynomial has degree at least 1.

Definition 4.2 *Let F be a field and let $f, g \in F[x]$ be*

$$f(x) = a_0 + a_1x + a_2x^2 + ... \quad \text{and} \quad g(x) = b_0 + b_1x + b_2x^2 + ...$$

Define addition on $F[x]$ as follows:

$$f + g = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + ...$$

Define multiplication on $F[x]$ as follows:

$$\begin{aligned} fg &= (a_0 + a_1x + a_2x^2 + ...)(b_0 + b_1x + b_2x^2 + ...) \\ &= a_0b_0 + (a_1b_0 + a_0b_1)x + (a_2b_0 + a_1b_1 + a_0b_2)x^2 + ... \end{aligned}$$

In particular, the x^k term of the product fg is

$$\left(\sum_{j=0}^{k} a_{k-j}b_j\right) x^k = (a_kb_0 + a_{k-1}b_1 + a_{k-2}b_2 + ... + a_1b_{k-1} + a_0b_k)x^k.$$

Theorem 4.3 *The addition and multiplication defined above make $F[x]$ into a ring, called a* **polynomial ring**. *In particular:*

- *the additive identity is the constant polynomial 0;*
- *the multiplicative identity is the constant polynomial 1; and*
- *the additive inverse of $f = a_0 + ... + a_nx^n$ is $-f = -a_0 - ... - a_nx^n$.*

PROOF To actually prove this, you have to check all the properties that a ring is supposed to have (addition and multiplication are commutative and associative, etc.). These properties are essentially inherited from the fact that any field F is also a ring, but writing the proofs is is tedious and you won't learn anything from it, so I won't bother doing it. □

Note: $F[x]$ is **not** a field (HW).

Theorem 4.4 *Let F be a field and let $f, g \in F[x]$. Then*

$$\deg(fg) = \deg(f) + \deg(g).$$

PROOF We did this in Chapter 1. □

Main concept: $F[x]$ is a lot like $\mathbb{Z}$

CONCEPT	STATEMENT IN $\mathbb{Z}$	STATEMENT IN $F[x]$
Divisibility	Let $a, b \in \mathbb{Z}$. We say $a \mid b$ if there exists $c \in \mathbb{Z}$ such that $b = ac$.	
Divisibility condition	If $a \mid b$, then $\lvert a\rvert \leq \lvert b\rvert$.	
Division Theorem	Let $a, b \in \mathbb{Z}$. Then $\exists! q, r \in \mathbb{Z}$ such that $b = aq + r$ and $0 \leq r < a$.	
Units	The units in $\mathbb{Z}$ are 1 and -1.	
Primes	$p \in \mathbb{Z}$ is prime if $p = ab$ implies that a or b is a unit.	
Unique Factorization	$\forall n \in \mathbb{N}$ with $n \geq 2$, $\exists$ primes $\{p_j\}_{j=1}^{k}$ (unique except for their order) s.t. $n = p_1 p_2 \cdots p_k$.	
Congruence	Let $n \in \mathbb{Z}$ be nonzero. We say $a, b \in \mathbb{Z}$ are **congruent (modulo n)** if $n \mid (b - a)$.	
Cosets and " $\mod$ " notation	The equivalence class of an integer a modulo n is called a **coset** and is denoted $a + n\mathbb{Z}$. If cosets $a + n\mathbb{Z}$ and $b + n\mathbb{Z}$, coincide, then we write $a \equiv b \mod n$.	
Quotient Space	The set $\mathbb{Z}/n\mathbb{Z}$ of cosets modulo n forms a ring. This ring is a field iff n is prime.	

Divisibility in $F[x]$

Definition 4.5 *Let F be a field and let $f, g \in F[x]$. We say f* **divides** *g and write $f \mid g$ if there exists $q \in F[x]$ such that $g = fq$. If f does not divide g, we write $f \nmid g$.*

Lemma 4.6 *Let F be a field and let $f, g \in F[x]$. If $f \mid g$ and g is not the zero polynomial, then $\deg(f) \leq \deg(g)$.*

PROOF Suppose $f \mid g$, i.e. $g = fq$ for $q \in F[x]$. Since $g \neq 0$, $q \neq 0$, so $\deg(q) \geq 0$. Then $\deg g = \deg(fq) = \deg f + \deg q \geq \deg f$. □

Lemma 4.7 *Let F be a field and let $f, g, h \in F[x]$. Then:*

- *if $f \mid g$ and $f \mid h$, then f divides any linear combination of g and h;*
- *if $f \mid g$ and $g \mid h$, then $f \mid h$.*

PROOF Same as in $\mathbb{Z}$ (except that the letters in the proof represent polynomials rather than integers). □

One catch: "linear combination" in $F[x]$ includes things like

$$(x^2 + 1)(x^3 - 3) + (x^3 - 3x^2 + 2x - 3)(x^2 + 4)$$

being a linear combination of $x^2 - 3$ and $x^2 + 4$, etc.

Theorem 4.8 (Division Theorem in *F*[*x*]) *Let F be a field and let $f, g \in F[x]$. There exist $q, r \in F[x]$ such that*

$$f(x) = g(x)q(x) + r(x)$$

and $\deg r < \deg g$.

PROOF Fix $f, g \in F[x]$.

By the WOP, D has a least element, say k. Thus

$$f(x) = g(x)q(x) + r(x)$$

where $r(x) = a_k x^k + ... + a_1 x + a_0$. It remains to show that $k < \deg g$.

Suppose not, i.e. $k \geq \deg g = m$. Write $g(x) = b_m x^m + ... + b_1 x + b_0$. But then,

$$f - gq - \frac{a_k}{b_m} x^{k-m} g = f - \left(q - \frac{a_n}{b_m} x^{k-m} \right) g$$

has degree less than k (because the x^k terms cancel), contradicting the definition of k. Thus $k = \deg r < \deg g$ as desired. □

Note: The reason the coefficients in our polynomials have to be in a field (as opposed to a ring) is because of the division necessary in the above proof.

Example: In $\mathbb{R}[x]$, take $g(x) = x^4 - 1$ and $f(x) = x^2 + x - 2$:

Units in $F[x]$

Theorem 4.9 *Let F be a field. The units in $F[x]$, i.e. the divisors of 1, are the nonzero constant polynomials.*

PROOF HW (use the fact that $\deg(fg) = \deg f + \deg g$).

Irreducible polynomials

Definition 4.10 *Let F be a field. A polynomial $p \in F[x]$ is called* **irreducible (over F)** *(a.k.a.* **prime***), if it is not a unit, not zero and whenever $p = fg$ for $f, g \in F[x]$, either f or g must be constant. A non-unit, non-irreducible polynomial in $F[x]$ is called* **reducible (over F)***.*

Example: $x^2 + 1$ is irreducible over $\mathbb{R}$, but $x^2 + 1 = (x + i)(x - i)$ so $x^2 + 1$ is reducible over $\mathbb{C}$.

Example: x^2+2 is irreducible over $\mathbb{R}$, but in $\mathbb{F}_3[x]$, $x^2+2 \equiv x^2-1 = (x+1)(x-1)$ so $x^2 + 2$ is reducible over $\mathbb{F}_3$.

Theorem 4.11 (Factor Theorem of Descartes) *Let F be a field and let $f \in F[x]$. If $c \in F$ is such that $f(c) = 0$, then*

$$f(x) = (x - c)q(x)$$

for some $q \in F[x]$.

PROOF Divide $f(x)$ by $(x - c)$ using the Division Theorem to get

$$f(x) = (x - c)q(x) + r(x)$$

where $\deg r < \deg(x - c) = 1$. That means $\deg r = 0$, i.e. r is constant, i.e.

$$f(x) = (x - c)q(x) + r. \tag{4.1}$$

Now plug in $x = c$ to both sides of (4.1) to get

$$f(c) = 0 + r.$$

Hence $r = 0$ and $f(x) = (x - c)q(x)$ as wanted. □

Corollary 4.12 (Number of roots of a polynomial) *Let F be a field. If $p \in F[x]$ has degree $d \geq 1$, then p has at most d roots in F.*

PROOF Induction on $\deg p$. When $\deg p = 1$, $p(x) = ax + b$ (where $a \neq 0$) has only the one root $x = \frac{-b}{a}$. Now assume any polynomial of degree k has at most k roots. Let p be a polynomial of degree $k + 1$. If p has no roots, we are done. Otherwise, let c be a root of p; by applying the Factor Theorem we have $p(x) = (x - c)q(x)$ where q has degree k. By the IH, q has at most k roots, so p has at most $k + 1$ roots. By induction, we are done. □

Corollary 4.13 *Suppose F is an infinite field. Let $f, g \in F[x]$. f and g are equal as polynomials (i.e. they have the same degree and all the same coefficients on the same powers of x) if and only if they are equal as functions $F \to F$ (i.e. $f(x) = g(x)$ for all $x \in F$).*

PROOF $(\Rightarrow)$ is obvious, even if F is finite.

$(\Leftarrow)$ Suppose $f = g$ as functions. Consider $h = f - g$. Every element of F is therefore a root of h, so h has infinitely many roots. By the previous corollary, that means h is the zero polynomial, so f and g agree as polynomials. □

Lemma 4.14 *Let F be a field and suppose h is an irreducible polynomial with a root $c \in F$. Then h is unique up to a constant factor.*

PROOF Suppose $h, h^* \in F[x]$ are both irreducible over F and both have root c. WLOG h^* has the smallest degree of any nonzero polynomial in $F[x]$ which has c as a root.

Divide h by h^* to get $h(x) = h^*(x)q^*(x)+r^*(x)$ and plug in $x = c$ to get $0 = r^*(c)$. But $\deg r^* < \deg h^*$, so r^* must be the zero polynomial. Thus $h = h^*q^*$. But h is irreducible, so q^* must be constant, i.e. $h = h^*$ up to a constant as wanted. □

Theorem 4.15 (General Factor Theorem) *Let F be a field and let $p \in F[x]$. If $c \in F$ is such that $p(c) = 0$, and if $h \in F[x]$ is an irreducible polynomial s.t. $h(c) = 0$, then $p(x) = h(x)q(x)$ for some $q \in F[x]$.*

PROOF Divide p by h to get

$$p(x) = h(x)q(x) + r(x).$$

Plug in $x = c$ to see that $r(c) = 0$. But $\deg r < \deg h$, contradicting the preceding lemma unless r is the zero polynomial. □

Theorem 4.16 (Irreducibles in $\mathbb{C}[x]$) *A polynomial $f \in \mathbb{C}[x]$ is irreducible if and only if it is linear.*

PROOF Suppose $f \in \mathbb{C}[x]$ is irreducible. By the FTAl, f has root z_0, and by the Factor Theorem that means $f(z) = (z - z_0)q(x)$ for some $q \in \mathbb{C}[x]$. But since f is irreducible, either $z - z_0$ or q is a unit, meaning q is a unit (i.e. a constant), meaning f is linear as wanted. □

Remark: In any $F[x]$, linear polynomials are always irreducible. The converse is what is interesting in the preceding statement.

Theorem 4.17 (Irreducibles in $\mathbb{R}[x]$) *A polynomial $f \in \mathbb{R}[x]$ is irreducible if and only if it is either linear, or quadratic with negative discriminant.*

PROOF HW

Unique factorization into irreducibles

Theorem 4.18 (Unique Factorization in $F[x]$) *Let F be a field and let $f \in F[x]$. Then there exist irreducible polynomials $p_1, p_2, ..., p_k \in F[x]$, unique up to their order and constant factors, such that*

$$f = p_1 p_2 \cdots p_k.$$

PROOF First, we have to show that every $f \in F[x]$ can be factored into irreducibles. To do this, let

$$D = \{\deg(f) : f \in F[x] \text{ cannot be factored into irreducibles}\}.$$

Suppose $D \neq \emptyset$. By the WOP, D has a least member, say n. Thus there is a degree n polynomial f which cannot be factored into irreducibles.

If f is irreducible, this is a contradiction.

But if f is reducible, we have $f = gh$ where $\deg g, \deg h < n$. This means $\deg g, \deg h \notin D$ so g and h both factor into irreducibles. Thus f factors into irreducibles, again a contradiction.

Either way, D must be empty, meaning every polynomial factors into irreducibles.

Second, we have to show the uniqueness of the factorization. This proof is the same as the one for $\mathbb{N}$, except that the letters represent (irreducible) polynomials instead of (prime) numbers. □

Congruence classes and modular arithmetic in $F[x]$

Definition 4.19 (Congruence in $F[x]$) *Let F be a field and let $l \in F[x]$ be nonzero. We say $f, g \in F[x]$ are* **congruent (modulo l)** *if $l \mid (f - g)$. The equivalence class of f under this relation is called a* **coset (modulo l)** *and is denoted $f + lF[x]$; if the equivalence classes of f and g coincide we write $f \equiv g \mod l$. The set of cosets modulo l is denoted $F[x]/lF[x]$.*

Note: One has to prove that this relation is an equivalence relation (the proof is essentially the same as the one for equivalence modulo n on the integers).

Definition 4.20 *Let F be a field, and let $l \in F[x]$ be nonzero. The set $F[x]/lF[x]$ is a ring under the addition and multiplication defined by*

$$(f + lF[x]) + (g + lF[x]) = (f + g) + lF[x];$$

$$(f + lF[x])(g + lF[x]) = (fg) + lF[x].$$

In particular, the additive identity is $0 + lF[x]$; the multiplicative identity is $1 + lF[x]$; the additive inverse of $f + lF[x]$ is $-f + lF[x]$.

Example:

Example: From the division we did earlier,

$$(x^4 - 1) \equiv (-4x + 5) \mod (x^2 + x - 2).$$

More generally, every polynomial $p \in F[x]$ is congruent modulo l to a polynomial of degree less than $\deg l$.

Theorem 4.21 *Let F be a field and let $l \in F[x]$ be nonzero. The ring $F[x]/lF[x]$ is a field if and only if l is irreducible.*

PROOF HW

4.2 Irreducibility tests

In light of the previous theorem, it seems like a good idea to have some idea of how to show whether or not a polynomial is irreducible.

- If a polynomial has a root in F, then it is reducible over F by the Factor Theorem.

  Application: Show $f(x) = x^4 - 3x + 2$ is reducible over $\mathbb{Q}$:

- The Rational Roots Theorem says that if a polynomial with integer coefficients has a root in $\mathbb{Q}$, then that root must be $\frac{p}{q}$ in lowest terms, where p divides the constant term and q divides the leading coefficient.

  Application: Show $f(x) = x^2 - x - 3$ is irreducible over $\mathbb{Q}$:

Now, for some new results. The first says that if a polynomial with integer coefficients factors over the rationals, then it factors over the integers:

Theorem 4.22 (Gauss' Lemma) *Suppose $g, h \in \mathbb{Q}[x]$. If $f = gh$ has integer coefficients, then $gh = g_0 h_0$ where both g_0 and h_0 must have integer coefficients.*

PROOF Let $f = gh$ where

$$\begin{aligned} f(x) &= a_n x^n + ... + a_1 x_0 + a_0; \\ g(x) &= \frac{b_m}{c_m} x^m + ... + \frac{b_1}{c_1} x + \frac{b_0}{c_0}; \\ h(x) &= \frac{r_k}{s_k} x^k + ... + \frac{r_1}{s_1} x + \frac{r_0}{s_0} \end{aligned}$$

where all the a_j, b_j, c_j, r_j and s_j are integers. Multiply through by the least common denominators in g and h to obtain

$$\hat{g}(x) = u_m x^m + ... + u_1 x + u_0$$
$$\hat{h}(x) = v_k x^k + ... + v_1 x + v_0$$

where $\hat{g}(x) = cg(x)$ and $\hat{h}(x) = sh(x)$ for integers c, h (the u_js and v_j are, of course, integers). WLOG $c > 0$ and $s > 0$. Thus

$$f(x) = g(x)h(x) = \frac{1}{cs}\hat{g}(x)\hat{h}(x) \quad \Rightarrow \quad csf(x) = \hat{g}(x)\hat{h}(x).$$

If $cs = 1$, we are done (let $g_0 = \hat{g}$ and $h_0 = \hat{h}$). Otherwise, write

$$cs = p_1 p_2 \cdots p_l$$

for primes $p_1, ..., p_l$ and observe that for any $t \in \{1, ..., l\}$, p_t divides every coefficient of $csf(x)$, i.e. every coefficient of $\hat{g}(x)\hat{h}(x)$.

Now, suppose that there is at least one coefficient of $\hat{g}(x)$, and at least one coefficient of $\hat{h}(x)$, both of which are <u>not</u> multiples of p_t. Let r and s be the smallest integers such that u_r and v_s are not multiples of p_t. That means

$$p_t \mid u_j \text{ for } j < r, \text{ and } p_t \mid v_j \text{ for } j < s.$$

Now, the coefficient of x^{r+s} in $\hat{g}(x)\hat{h}(x)$, which is a multiple of p_t since it is a coefficient of $csf(x)$, is

$$u_{r+s}v_0 + u_{r+s-1}v_1 + ... + u_{r+1}v_{s-1} + u_r v_s + u_{r-1}v_{s+1} + ... + u_1 v_{r+s-1} + u_0 v_{r+s}$$

Consequently $p_t \mid u_r v_s$. By the Prime Divisor Lemma, $p_t \mid u_r$ or $p_t \mid v_s$, a contradiction. Therefore p_t divides all the coefficients of either $\hat{g}$ or $\hat{h}$. WLOG p_t divides all the coefficients of $\hat{g}$. Then factor out p_t from both sides of the equation

$$csf(x) = \hat{g}(x)\hat{h}(x)$$

to get

$$\frac{cs}{p_t}f(x) = \hat{\hat{g}}(x)\hat{h}(x).$$

If $\frac{cs}{p_t} = 1$, we are done; otherwise, repeat the procedure above to factor out more primes in the factorization of cs. Eventually we will have

$$f(t) = (\text{some number of hats of } g)(x)\,(\text{some number of hats of } h)(x) = g_0(x)h_0(x)$$

as wanted. □

Application: Determine if $f(x) = x^4 + 2x^3 + 5x^2 + 4x + 3$ is irreducible over $\mathbb{Q}$.

We have ended up with the system (where $b, c, e \in \mathbb{Z}$)

$$\begin{cases} b + e &= 2 \\ 4c + be &= 5 \\ 3bc + ce &= 4 \end{cases}$$

Theorem 4.23 (Irreducibility Test mod *p*) *Let $f \in \mathbb{Q}[x]$ have integer coefficients. For any prime p which does not divide the leading coefficient of f, if $f \mod p$ is irreducible over $\mathbb{F}_p$, then f is irreducible over $\mathbb{Q}$.*

PROOF We prove the contrapositive. Suppose $f = gh$ for $g, h \in \mathbb{Q}[x]$ where $\deg g, \deg h \geq 1$. By Gauss' Lemma, we can assume WLOG that g and h have integer coefficients. Let b_m be the leading coefficient of g and let c_k be the leading coefficient of h; thus $b_m c_k$ is the leading coefficient of f. If $p \nmid b_m c_k$, then $p \nmid b_m$ and $p \nmid c_k$, so

$$\deg(g \mod p) = \deg g = m \geq 1 \quad \text{and} \quad \deg(h \mod p) = \deg h = k \geq 1.$$

That means

$$f \mod p = (g \mod p)(h \mod p)$$

with $\deg(g \mod p) \geq 1$ and $\deg(h \mod p) \geq 1$, i.e. $f \mod p$ is reducible over $\mathbb{F}_p$. □

Application: Determine if $f(x) = x^3 + 2x + 2$ is irreducible over $\mathbb{Q}$.

Lemma 4.24 (Substitution Trick) *Let $f(x) \in \mathbb{Q}[x]$. Let $x = ay + b$ for $a, b \in \mathbb{Q}$ with $a \neq 0$. If $g(y) = f(ax + b)$ is irreducible in $\mathbb{Q}[y]$, then f is irreducible in $\mathbb{Q}[x]$.*

PROOF If g factors into two polynomials in the variable y, then f factors in the same way by replacing each of the ys in the g-factorization with $\frac{1}{a}(x - b)$. □

Application: Determine if $f(x) = (x + 2)^3 + 2(x + 2) + 2$ is irreducible over $\mathbb{Q}$.

Theorem 4.25 (Eisenstein Criterion) *Suppose*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0$$

is a polynomial with integer coefficients. If p is a prime such that:

- $p \mid a_j$ *for all* $j \in \{1, ..., n-1\}$*;*
- $p \nmid a_n$*; and*
- $p^2 \nmid a_0$.

Then f is irreducible over $\mathbb{Q}$.

PROOF Suppose all the hypotheses of this criterion are satisfied, but f is reducible. Write $f(x) = g(x)h(x)$; by Gauss' Lemma there are integers $b_0, ..., b_m, c_0, ..., c_k$ such that

$$g(x) = b_m x^m + ... + b_1 x + b_0;$$
$$h(x) = c_k x^k + ... + c_1 x + c_0.$$

Observe:

- since $a_0 = b_0 c_0$, p divides either b_0 or c_0, but cannot divide both because $p^2 \nmid a_0$. WLOG assume $p \nmid b_0$;
- since $p \nmid a_n$ and $a_n = b_m c_k$, p cannot divide either b_m or c_k.

Now let l be the smallest integer so that $p \nmid c_l$. By hypothesis, $p \mid a_l$. Now

$$a_l = b_l c_0 + b_{l-1}c_1 + ... + b_1 c_{l-1} + b_0 c_l$$

so by subtraction,

$$b_0 c_l = a_l - [b_l c_0 + b_{l-1}c_1 + ... + b_1 c_{l-1}].$$

p divides every term on the right, so $p \mid b_0 c_l$, so $p \mid b_0$ by the Prime Divisor Lemma, contradicting the first bullet point above. Thus f must be irreducible. □

Application: Determine if $2x^5 + 36x^2 + 15x + 21$ is irreducible over $\mathbb{Q}$.

Corollary 4.26 *Let p be a prime. Then the p^{th} cyclotomic polynomial Φ_p is irreducible over $\mathbb{Q}$.*

PROOF For every pair of natural numbers n and k with $k \leq n$, define

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

The Binomial Theorem says that for any two complex numbers z and w, and any natural number n,

$$(z+w)^n = \sum_{k=0}^{n} \binom{n}{k} z^n w^{n-k}.$$

We will apply this to prove the corollary. Let

$$\Phi_p(z) = z^{p-1} + z^{p-2} + ... + z^2 + z + 1 = \frac{z^p - 1}{z-1}.$$

Let $z = x + 1$, so that

$$\begin{aligned}
\Phi_p(x+1) = \frac{(x+1)^p - 1}{x} & \\
&= \frac{1}{x}\left[\sum_{k=0}^{p} \binom{p}{k} x^k 1^{p-k} - 1\right] \\
&= \frac{1}{x} \sum_{k=1}^{p} \frac{p!}{k!(p-k)!} x^k \\
&= \sum_{k=1}^{p} \frac{p!}{k!(p-k)!} x^{k-1} \\
&= x^{p-1} + \binom{p}{1} x^{p-2} + \binom{p}{2} x^{p-3} + ... + \binom{p}{p-1} x + \binom{p}{p-1} \\
&= x^{p-1} + px^{p-2} + \binom{p}{2} x^{p-3} + ... + \binom{p}{p-1} x + p
\end{aligned}$$

Notice $p \mid p!$ but when $k \in \{1, ..., p-1\}$, since k and $p-k$ are less than p, $p \not| k!(p-k)!$. Therefore $p \mid \binom{p}{k}$.

But p does not divide the leading coefficient 1, nor does p^2 divide the constant term p. So by Eisenstein, $\Phi_p(x+1)$ is irreducible over $\mathbb{Q}$, meaning by the substitution trick that $\Phi_p(z)$ is irreducible over $\mathbb{Q}$ as well. □

Chapter 5

Fields

Recall that a **field** is an algebraic system $(F, +, \cdot)$ where $+$ and $\cdot$ are associative and commutative, have identity elements and inverses, and where the distributive law works. (More simply, a field is a ring where every nonzero element has a reciprocal.)

In this course, we've encountered the following examples of fields:

Number fields (subfields of $\mathbb{C}$):

Fields that aren't number fields:

Quotients of polynomial rings:

What other fields are there? How do various fields relate to one another? What do fields have to do with construction problems and/or quintic equations? That is the subject of this chapter.

5.1 Field extensions

One way to create a new field is to take an old field (especially $\mathbb{Q}$) and "adjoin" one or more elements to it. This means making a bigger field, which contains the old field and contains the elements being adjoined. Here are the details:

Definition 5.1 *Let F be a number field and suppose $\alpha \in \mathbb{C}$. Define $F(\alpha)$ to be the smallest subfield of $\mathbb{C}$ containing F and α.*

Note: Since $F \subseteq F(\alpha)$, we say F is a **subfield** of $F(\alpha)$ and that $F(\alpha)$ is an **extension** of F.

The worst-case scenario when you adjoin α to F to make $F(\alpha)$ is that you get a field of rational functions in α:

Lemma 5.2 *Let F be a number field and suppose $\alpha \in \mathbb{C}$. Then*

$$F(\alpha) = \left\{ \frac{f(\alpha)}{g(\alpha)} : f, g \in F[x] \text{ and } g(\alpha) \neq 0 \right\}.$$

PROOF First, the $F(\alpha)$ described above is a field (it is closed under $+, \cdot$, additive inverses and reciprocals of nonzero elements).

Now, let G be any field containing F and α. Since G is closed under $+, -, \cdot$ and $\div$, G must contain $F(\alpha)$ as described above, so $F(\alpha)$ is the smallest subfield containing F and α, as wanted. □

Often, the description of $F(\alpha)$ simplifies (see the next section).

You can adjoin more than one element to a field by adjoining one element at a time:

Definition 5.3 *Let F be a number field and let $\alpha_1, ..., \alpha_n \in \mathbb{C}$. Define*

$$F(\alpha_1, \alpha_2, ..., \alpha_n) = [\cdots[[[F(\alpha_1)](\alpha_2)](\alpha_3)]\cdots(\alpha_n)].$$

Theorem 5.4 *Let F be a number field and let $\alpha_1, ..., \alpha_n \in \mathbb{C}$. Then $F(\alpha_1, ..., \alpha_n)$ is the closure of $F \cup \{\alpha_1, ..., \alpha_n\}$ under $+, -\cdot, \div$. As a consequence, the order in which the elements are adjoined to create $F(\alpha_1, ..., \alpha_n)$ doesn't matter.*

PROOF Induction on n. The base case $n = 1$ is obvious. Now suppose the result is true when $n = k$. Thus

$$\begin{aligned} F(\alpha_1, ..., \alpha_k, \alpha_{k+1}) &= F(\alpha_1, ..., \alpha_k)(\alpha_{k+1}) \\ &= \text{closure under } +, -, \cdot, \div \text{ of } F(\alpha_1, ..., \alpha_k) \cup \{\alpha_{k+1}\} \\ &= \text{closure under } +, -, \cdot, \div \text{ of } F \cup \{\alpha_1, ..., \alpha_k\} \cup \{\alpha_{k+1}\} \\ &= \text{closure under } +, -, \cdot, \div \text{ of } F \cup \{\alpha_1, ..., \alpha_{k+1}\}. \end{aligned}$$

By induction, we are done. □

5.2 Algebraic extensions

Recall that $F(\alpha)$ can be thought of as the set of rational functions in α:

$$F(\alpha) = \left\{ \frac{f(\alpha)}{g(\alpha)} : f, g \in F[x] \text{ and } g(\alpha) \neq 0 \right\}.$$

Does a simpler description of $F(\alpha)$ exist?

Definition 5.5 *Let F be a number field and let $\alpha \in \mathbb{C}$. α is called* **algebraic (over F)** *if there is a nonzero polynomial $f \in F[x]$ such that $f(\alpha) = 0$. If α is not algebraic over F, α is called* **transcendental (over F)**. *The* **algebraic closure** *of F, denoted $\widehat{F}$, is the set of all algebraic numbers over F.*

Example: $\sqrt{2}$ is algebraic over $\mathbb{Q}$, because $\sqrt{2}$ is a root of $x^2 - 2$.

Example: i is algebraic over $\mathbb{Q}$, because i is a root of $x^2 + 1$.

Example: for all $n \geq 2$, $\zeta_n = e^{2\pi i/n}$ is algebraic over $\mathbb{Q}$, because ζ_n is a root of Φ_n.

Fact: π is transcendental over $\mathbb{Q}$ (proof is beyond the scope of this class).

The FTAl can be restated as:

Theorem 5.6 *$\widehat{\mathbb{Q}}$ is countable (therefore $\widehat{\mathbb{Q}}$ cannot be all of $\mathbb{C}$ nor can it contain all of $\mathbb{R}$).*

PROOF For each $n \geq 0$, let $\mathcal{P}_n$ be the set of all polynomials in $\mathbb{Q}[x]$ whose degree is n. There is an injection from $\mathcal{P}_n$ to $\mathbb{Q}^{n+1}$ given by

$$f = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0 \mapsto (a_n, a_{n-1}, ..., a_0)$$

and consequently, since $\mathbb{Q}^{n+1}$ is countable, so is $\mathcal{P}_n$. Therefore

$$\mathbb{Q}[x] - \{0\} = \bigcup_{n=0}^{\infty} \mathcal{P}_n$$

is the union of countably many countable sets, hence countable. Therefore,

$$\widehat{\mathbb{Q}} = \bigcup_{f \in \mathbb{Q}[x] - \{0\}} \{\alpha : f(\alpha) = 0\}$$

is the union of countably many finite sets (each set in the union is finite because a polynomial of degree n has at most n roots), hence is countable. □

Lemma 5.7 *Let F be a number field and suppose α is algebraic over F. Then:*

1. *there is an irreducible polynomial $h \in F[x]$ such that $h(\alpha) = 0$;*
2. *any two irreducible polynomials which have α as a root must have the same degree.*

PROOF HW

Definition 5.8 *Let F be a number field and suppose α is algebraic over F. A* **minimal polynomial for α (over F)** *is an irreducible polynomial $h \in F[x]$ such that $h(\alpha) = 0$. Define the* **degree** *of α over F, denoted $\deg(\alpha/F)$, to be the degree of any minimal polynomial for α.*

Example: $\deg(\sqrt{2}/\mathbb{Q}) = \deg(i/\mathbb{Q}) = 2$.

Example: $\deg(\zeta_n/\mathbb{Q}) = \deg \Phi_n =$

Example: $F = \mathbb{Q}$; $\alpha = 2 + \sqrt{3}$.

Theorem 5.9 *Let F be a number field and let α be algebraic over F with $\deg(\alpha/F) = n$. Then*

$$F(\alpha) = \{a_0 + a_1\alpha + a_2\alpha^2 + ... + a_{n-1}\alpha^{n-1} : a_0, a_1, ..., a_{n-1} \in F\}.$$

PROOF Let $G = \{a_0 + a_1\alpha + a_2\alpha^2 + ... + a_{n-1}\alpha^{n-1} : a_0, a_1, ..., a_{n-1} \in F\}$. It is clear that $G \subseteq F(\alpha)$.

To show $F(\alpha) \subseteq G$, first let h be a minimal polynomial for α over F. Since h is irreducible, $F[x]/hF[x]$ is a field. That means that for any $g \in F[x]$ with $g(\alpha) \neq 0$, g is not a multiple of h, so $g + hF[x]$ is a unit, i.e. there is $g^{-1} \in F[x]$ such that

$$(g + hF[x])(g^{-1} + hF[x]) = 1 + hF[x].$$

In other words, by plugging in α to this equation, we get

$$g(\alpha)g^{-1}(\alpha) = 1.$$

Now let $p \in F(\alpha)$. Thus $p = \frac{f(\alpha)}{g(\alpha)} = f(\alpha)g^{-1}(\alpha)$. Thus p can be written as a polynomial in α.

Last, since h has degree n and $h(\alpha) = 0$, for any $k \geq n$, α^k can be rewritten as a polynomial in terms of $\alpha, \alpha^2, ..., \alpha^{n-1}$, so p is a polynomial in α of degree at most $n - 1$, i.e. $p \in G$ as wanted. □

Consequence: for example, $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$.

Question: Is $\sqrt[4]{2}$ an element of $\mathbb{Q}(\sqrt{2})$?

5.3 Linear algebra and field extensions

Review of linear algebra

Definition 5.10 *A* **vector space** *is a collection of objects called* **vectors**, *together with two operations (addition and multiplication by scalars) satisfying a bunch of laws:*

- *addition is commutative, associative, there is an additive identity and there are additive inverses;*
- *scalar multiplication is associative and* 1 *is an identity element; and*
- *scalar multiplication distributes over addition.*

WARNING: vector spaces are not usually rings: the scalar multiplication is not a binary operation on a vector space because it is a *scalar* times a vector (not a *vector* times a *vector*.

Definition 5.11 *Let* V *be a vector space, and let* $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n \in V$.

- $\mathbf{v}_1, ..., \mathbf{v}_n$ *are called* **linearly independent** *if for any* $c_1, ..., c_n \in \mathbb{R}$,

$$c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n = \mathbf{0} \text{ implies } c_1 = c_2 = ... = c_n = 0.$$

- $\mathbf{v}_1, ..., \mathbf{v}_n$ *are said to* **span** V *if for any* $\mathbf{w} \in V$,

$$\mathbf{w} = c_1\mathbf{v}_1 + ... + c_n\mathbf{v}_n$$

  for scalars $c_1, ..., c_n$.
- $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ *is called a* **basis** *of* V *if the vectors* $\mathbf{v}_1, ..., \mathbf{v}_n$ *are linearly independent and span* V.

Theorem 5.12 *Any two bases of vector space* V *must have the same number of elements.*

Definition 5.13 *Let* V *be a vector space. The* **dimension** *of* V *is the number of elements in any basis of* V.

Dimension of a field extension

Big idea: Let field E be an extension of field F (i.e. $F \subseteq E$). Then:

Example: $\mathbb{R} \subseteq \mathbb{C}$.

Definition 5.14 *Let E be an extension of field F. We say $e_1, ..., e_n \in E$ are* **(linearly) independent (over F)** *if for any $f_1, ..., f_n \in F$,*

$$f_1e_1 + f_2e_2 + ... + f_ne_n = 0 \text{ implies } f_1 = f_2 = ... = f_n = 0.$$

Example: $\{1, i\} \subseteq \mathbb{C}$ (viewed as an extension of $\mathbb{R}$)

Example: $\{1, i, 3 - 2i\} \subseteq \mathbb{C}$ (viewed as an extension of $\mathbb{R}$)

Example: $\{1, i, i^2\} \subseteq \mathbb{C}$ (viewed as an extension of $\mathbb{R}$)

More general example: if $\deg(\alpha/\mathbb{Q}) = n$, then $1, \alpha, \alpha^2, ..., \alpha^n$ are not linearly independent over $\mathbb{Q}$.

Definition 5.15 *Let E be an extension of field F. We say $e_1, ..., e_n \in E$* **span** *E if for any $e \in E$, there exist $f_1, ..., f_n \in F$ such that*

$$e = f_1 e_1 + f_2 e_2 + ... + f_n e_n.$$

Example: $\{1, i\} \subseteq \mathbb{C}$ (viewed as an extension of $\mathbb{R}$)

Theorem 5.9 restated: if $\deg(\alpha/F) = n$, then $1, \alpha, \alpha^2, ..., \alpha^{n-1}$ span $F(\alpha)$.

Definition 5.16 *Let E be an extension of field F. We say $e_1, ..., e_n \in E$ is a* **basis for** *E* **(over** *F***)** *if $e_1, ..., e_n$ are linearly independent over F and span E.*

Theorem 5.17 *Let E be an extension of field F. Any two bases of E over F have the same number of elements.*

PROOF Essentially the same proof as the proof from Math 322 that any two bases of a vector space have the same number of elements (i.e. the Exchange Lemma). □

Theorem 5.18 *Let F be a number field and let $\alpha \in \mathbb{C}$. If $\deg(\alpha/F) = n$, then $\{1, \alpha, \alpha^2, ..., \alpha^{n-1}\}$ is a basis of $F(\alpha)$ over F.*

PROOF HW (the set spans by Theorem 5.9; you have to show the set is independent).

Definition 5.19 *Let E be an extension of field F. The* **dimension of** *E* **over** *F, a.k.a. the* **degree of** *E* **over** *F, denoted $\dim(E/F)$ or $\dim(E : F)$ or $(E : F)$, is the number of elements in any basis of E over F.*

Example: $\dim(\mathbb{C}/\mathbb{R}) = 2$

Example: $\dim(\mathbb{Q}(\sqrt{2})/\mathbb{Q}) = 2$

Corollary 5.20 *Let E be an extension of field F with $\dim(E/F) = n$. Then every $\alpha \in E$ is algebraic over F, and $\deg(\alpha/F) \leq n$.*

PROOF Consider the set $\{1, \alpha, \alpha^2, ..., \alpha^n\} \subseteq E$. This set has $n+1$ elements, which is more than $\dim(E/F)$ elements, so this set cannot be independent. Therefore there is an equation

$$a_0 + a_1\alpha + ... + a_n\alpha^n = 0$$

where $a_j \in F$ for all j and not all a_j are zero. In other words, there is a polynomial $f = a_0 + ... + a_n x^n$ of degree $\leq n$ such that $f(\alpha) = 0$, proving the corollary. □

Iterated extensions

Recall that by definition, $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \left[\mathbb{Q}(\sqrt{2})\right](\sqrt{3})$.

Thus $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ is an extension of $\mathbb{Q}$ and an extension of $\mathbb{Q}(\sqrt{2})$. So

$$\mathbb{Q}(\sqrt{2}, \sqrt{3}) \supseteq \mathbb{Q}(\sqrt{2}) \supseteq \mathbb{Q}.$$

How do the dimensions of these extensions relate to one another?

Remark: In this context, $1, \sqrt{2}$ (thought of as elements of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$) are independent over $\mathbb{Q}$, but not independent over $\mathbb{Q}(\sqrt{2})$:

Question: What is $\dim(\mathbb{Q}(\sqrt{2}, \sqrt{3})/\mathbb{Q}(\sqrt{2}))$?

Theorem 5.21 (Dedekind Product Theorem) *Suppose $E \supseteq B \supseteq F$ are fields with $\{e_1, ..., e_m\}$ a basis for E over B and $\{b_1, ..., b_n\}$ a basis for B over F. Then*

$$\{e_1b_1, e_1b_2, e_1b_3, ...e_1b_n, e_2b_1, e_2b_2, ..., e_mb_n\}$$

is a basis for E over F. In particular, this means

$$\dim(E/F) = \dim(E/B)\dim(B/F).$$

PROOF Let $\epsilon \in E$. Then since $\{e_1, ..., e_m\}$ is spans E over B,

$$\epsilon = \sum_{j=1}^{m} \beta_j e_j = \beta_1 e_1 + \beta_2 e_2 + ... + \beta_m e_m$$

for $\beta_1, ..., \beta_m \in B$. Now since $b_1, ..., b_n$ spans B over F, we can write, for each j,

$$\beta_j = b_1 f_{j,1} + b_2 f_{j,2} + ... + b_n f_{j,n}$$

for $f_{j,1}, ..., f_{j,n} \in F$. By substitution, we see that

$$\epsilon = \sum_{j=1}^{m} \beta_j e_j = \sum_{j=1}^{m} \left(\sum_{k=1}^{n} b_k f_{j,k} \right) e_j = \sum_{(j,k)=(1,1)}^{(m,n)} e_j b_k f_{j,k}$$

meaning that the products $\{e_j b_k : 1 \leq j \leq m, 1 \leq k \leq n\}$ span E over F as wanted.

To show independence, suppose there are constants $a_{j,k} \in F$ such that

$$0 = \sum_{j=1}^{m} \sum_{k=1}^{n} a_{j,k} e_j b_k = \sum_{j=1}^{m} \left(\sum_{k=1}^{n} a_{j,k} b_k \right) e_j.$$

By the independence of $\{e_1, ..., e_m\}$ over B, it follows that for all j,

$$0 = \sum_{k=1}^{n} a_{j,k} b_k.$$

But by the independence of $\{b_1, ..., b_n\}$ over F, it follows that for all j, k that $a_{j,k} = 0$. Thus the products $\{e_j b_k : 1 \leq j \leq m, 1 \leq k \leq n\}$ are independent over F, meaning they form a basis of E over F. □

Back to the example: We saw on the previous page that

$$\dim(\mathbb{Q}(\sqrt{2},\sqrt{3})/\mathbb{Q}(\sqrt{2}) = 2 \qquad \text{and} \qquad \dim(\mathbb{Q}(\sqrt{2})/\mathbb{Q}) = 2.$$

By the Dedekind Product Theorem,

$$\dim(\mathbb{Q}(\sqrt{2},\sqrt{3})/\mathbb{Q}) =$$

and a basis of $\mathbb{Q}(\sqrt{2},\sqrt{3})$ over $\mathbb{Q}$ is

Therefore, we can describe $\mathbb{Q}(\sqrt{2},\sqrt{3})$ as

Corollary 5.22 *Let F be a number field. Then its algebraic closure $\widehat{F}$ is a field.*

PROOF Suppose $\alpha, \beta \in \widehat{F}$. Then $\dim(F(\alpha)/F) = m$ and $\dim(F(\beta)/F) = n$ for positive integers m and n. That means $\deg(\alpha/F) = m$, so α is the root of an irreducible $h \in F[x] \subseteq F(\beta)[x]$. Therefore α has degree at most m over $F(\beta)$, so $\dim(F(\alpha,\beta)/F(\beta)) \leq m$.

Therefore by the Dedekind Product Theorem,

$$\begin{aligned}\dim(F(\alpha,\beta)/F) &= \dim(F(\alpha,\beta)/F(\beta))\ \dim(F(\beta)/F) \\ &\leq mn < infty\end{aligned}$$

so all members of $F(\alpha,\beta)$ (including $\alpha \pm \beta$, $\alpha\beta$ and $\alpha \div \beta$) are algebraic over F (they have degree $\leq mn$). Thus $\widehat{F}$ is closed under $+, -, \cdot$ and $\div$, making it a field. $\square$

5.4 Classical construction problems, revisited

What we know at this point

In order to perform a geometric construction with straightedge and compass, the coordinates of any points so obtained must be **surd** numbers. For each of the classical construction problems encountered in Chapter 1, we found a number $\alpha \in \mathbb{C}$ so that the construction problem is doable exactly if α is surd:

CLASSICAL CONSTRUCTION PROBLEM	CORRESPONDING NUMBER α THAT HAS TO BE SURD	CORRESPONDING POLYNOMIAL FOR WHICH α IS A ROOT
Squaring the circle		
Doubling the cube		
Trisecting a 60° angle		
Constructing a regular p-gon (p prime)		

The key result

Theorem 5.23 (Characterization of surd numbers) *If $\alpha \in \mathbb{C}$ is a surd number, then $\dim(\mathbb{Q}(\alpha)/\mathbb{Q}) = 2^m$ for some $m \in \mathbb{N}$.*

PROOF Let $\alpha \in \mathbb{C}$ be surd. Recall that any such α must be obtained from rational numbers by $+, -, \cdot, \div, \sqrt{}$. This means α lies in a field extension

$$\mathbb{Q}(\alpha_1, \alpha_2, ..., \alpha_n)$$

where $\alpha_1^2 \in \mathbb{Q}$ and for all $j > 1$, $\alpha_j^2 \in \mathbb{Q}(\alpha_1, ..., \alpha_{j-1})$.

Since each α_j satisfies a quadratic equation over $\mathbb{Q}(\alpha_1, ..., \alpha_{j-1})$, we see

So by the Dedekind Product Theorem,

$$\dim(\mathbb{Q}(\alpha_1, ..., \alpha_n)/\mathbb{Q}) = \dim(\mathbb{Q}(\alpha_1)/\mathbb{Q}) \prod_{j=2}^{n} \dim(\mathbb{Q}(\alpha_1, ..., \alpha_j)/\mathbb{Q}(\alpha_1, ..., \alpha_{j-1}))$$

Since $\alpha \in \mathbb{Q}(\alpha_1, \alpha_2, ..., \alpha_n)$, $\dim(\alpha/\mathbb{Q})$ divides 2^n, so it must be 2^m for some $m \in \mathbb{N}$. □

Impossibility results

Theorem 5.24 *Squaring the circle with straightedge and compass is impossible.*

PROOF π is transcendental over $\mathbb{Q}$, so $\dim(\mathbb{Q}(\pi)/\mathbb{Q}) = \infty$, which is not a power of 2. Thus π is not surd. □

Theorem 5.25 *Doubling the cube with straightedge and compass is impossible.*

PROOF $\sqrt[3]{2}$ has minimal polynomial $x^3 - 2$ (irreducible by either Eisenstein ($p = 2$) or Rational Roots Theorem), so $\dim(\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}) = 3$. This dimension is not a power of 2, so $\sqrt[3]{2}$ is not surd. □

More generally, this argument shows that the equation $x^3 - 2 = 0$ ***cannot*** be solved with $+, -, \cdot, \div, \sqrt{}$ alone (if it was possible to do this, the solution would be surd). So the solution of a cubic equation really does require cube roots, in general.

Theorem 5.26 *Trisecting a $60°$ angle with straightedge and compass is impossible.*

PROOF From earlier HW, $x = \cos 20°$ satisfies the polynomial equation

$$8x^3 - 6x - 1 = 0.$$

Let $y = 2x - 1$ so that the above polynomial becomes

$$y^3 + 3y^2 - 3 = 0.$$

By Eisenstein ($p = 3$), $y^3 + 3y^2 - 3$ is irreducible, so $\dim(\mathbb{Q}(y)/\mathbb{Q}) = 3$, which is not a power of 2. Thus y is not a constructible number, so neither is $x = \frac{1}{2}(y + 1)$. Since $\cos 20°$ is not a constructible number, a $20°$ angle cannot be constructed from a straightedge and compass. □

Theorem 5.27 *If p is a prime, then constructing a regular p-gon with straightedge and compass is impossible, unless $p - 1$ is a power of 2.*

PROOF Let $\zeta_p = e^{2\pi i/p}$; since ζ_p is a root of the irreducible cyclotomic polynomial Φ_p (which has degree $p - 1$), we see that $\deg(\mathbb{Q}(\zeta_p)/\mathbb{Q}) = p - 1$ from whence the result follows. □

Question: For which primes p is $p - 1$ a power of 2?

Definition 5.28 *A prime $p \in \mathbb{Z}$ is called a* **Fermat prime** *if $p - 1$ is a power of 2.*

Lemma 5.29 *Every Fermat prime is of the form $2^{2^n} + 1$ for $n \in \mathbb{N}$.*

WARNING: the converse of this is false: $641 \mid (2^{2^5} + 1)$, for example.

Remark: the only known Fermat primes are $2^{2^n} + 1$ when $n \in \{0, 1, 2, 3, 4\}$:

$$3, 5, 17, 257, 65537$$

No one knows if there are any other Fermat primes.

Theorem 5.30 *A regular n-gon can be constructed with a straightedge and compass only if*

$$n = 2^m p_1 p_2 \cdots p_k$$

where $p_1, p_2, ..., p_k$ are distinct Fermat primes.

PROOF Suppose the regular n-gon is constructible. Since ζ_n is a root of Φ_n, which is irreducible over $\mathbb{Q}$ (I proved irreducibility if n is prime; if n is not prime, consult sections 4.8 and 4.9 of Stillwell for a proof that Φ_n is irreducible if n is composite). Therefore

$$\deg(\zeta_n/\mathbb{Q}) = \phi(n),$$

so $\phi(n)$ has to be a power of 2.

Now, write the prime factorization of n as

$$n = p_0^{j_0} p_1^{j_1} \cdots p_k^{j_k}.$$

By properties of the Euler phi function,

$$\phi(n) = p_0^{j_0-1}(p_0-1)p_1^{j_1-1}(p_1-1)\cdots p_k^{j_k-1}(p_k-1).$$

This is a power of 2 only if

- $p_0 = 2$, and
- $j_1 - 1 = j_2 - 1 = ... = j_k - 1 = 0$, and
- $p_1 - 1$, $p_2 - 1$, ..., $p_k - 1$ are all powers of 2 (i.e. $p_1, ..., p_k$ are all Fermat primes).

This completes the proof. □

Question: What about the converse: if $n = 2^m p_1 p_2 \cdots p_k$ where $p_1, p_2, ..., p_k$ are distinct Fermat primes, is the regular n-gon constructible with a straightedge and compass?

Chapter 6

Morphisms

6.1 What is a homomorphism?

A *morphism* is math lingo for a function between two sets, each with some kind of structure, so that the "structures" on the two sets are preserved. Every branch of mathematics has its own concept(s) of "structure", so it has its own version of morphisms:

BRANCH OF MATHEMATICS	MORPHISM
Linear algebra	Linear transformation
Topology	Continuous function / homeomorphism
Set theory	Function / bijection
Abstract algebra	?

This class is about abstract algebra, so we will study morphisms that preserve algebraic structures.

Definition 6.1 *Let $(S, \odot)$ and $(S', \odot')$ be two algebraic systems. We say that a function $\sigma : S \to S'$ is a* **homomorphism** *if*

$$\sigma(x \odot y) = \sigma(x) \odot' \sigma(y)$$

for every $x, y \in S$.

In English: $\sigma : S \to S'$ is a homomorphism if, whenever you take two elements of S, if you first $\odot$ them and then do σ, you get the same thing as if you applied σ to both of them (sending them into S') and then $\odot'$ing them.

A "commutative diagram" to explain:

$$\begin{array}{ccc} S \times S & \xrightarrow{\sigma \otimes \sigma} & S' \times S' \\ {\scriptstyle \odot}\downarrow & & \downarrow{\scriptstyle \odot'} \\ S & \xrightarrow{\sigma} & S' \end{array}$$

Some examples

Example A: logarithm (with arbitrary base), thought of as a function

$$\log : ((0, \infty), \cdot) \to (\mathbb{R}, +).$$

Why is this a homomorphism?

Example B: "quotient map" $\sigma : (\mathbb{Z}, +) \to (\mathbb{Z}/n\mathbb{Z}, +)$ defined by $\sigma(x) = x + n\mathbb{Z}$

Why is this a homomorphism?

Example C: $\sigma : (\mathbb{Z}/6\mathbb{Z}, +, \cdot) \to (\mathbb{Z}/2\mathbb{Z}, +, \cdot)$ defined by $\sigma(x + 6\mathbb{Z}) = x + 2\mathbb{Z}$

Why is this well-defined?

Suppose $x + 6\mathbb{Z} = y + 6\mathbb{Z}$. That means $x - y = 6k = 2(3k)$ for $k \in \mathbb{Z}$. Therefore $2 \mid (x - y)$, so $x + 2\mathbb{Z} = y + 2\mathbb{Z}$, i.e. $\sigma(x + 6\mathbb{Z}) = \sigma(y + 6\mathbb{Z})$.

Why is this a homomorphism?

Example D: conjugation $\sigma : (\mathbb{C}, +, \cdot) \to (\mathbb{C}, +, \cdot)$ defined by $\sigma(z) = \overline{z}$.

Why is this a homomorphism?

Example E: $\sigma : (\mathbb{Z}/4\mathbb{Z}, +) \to (\{\pm 1, \pm i\}, \cdot)$ defined by

$$\sigma(0 + 4\mathbb{Z}) = 1 \quad \sigma(1 + 4\mathbb{Z}) = i \quad \sigma(2 + 4\mathbb{Z}) = -1 \quad \sigma(3 + 4\mathbb{Z}) = -i$$

Why is this a homomorphism?

HOW TO PROVE $\sigma : (S, \odot) \to (S', \odot')$ is a homomorphism:

Let $x, y \in S$.
$\sigma(x \odot y) = ... =$ something.
$\sigma(x) \odot' \sigma(y) = ... =$ the same something as above.

Therefore σ is a homomorphism. $\square$

HOW TO PROVE $\sigma : (S, \odot) \to (S', \odot')$ is **NOT** a homomorphism:

Write down two specific elements $x, y \in S$, and show that for those two elements, $\sigma(x \odot y) \neq \sigma(x) \odot' \sigma(y)$. $\square$

(If you took linear algebra from me, this should remind you of how we prove whether or not a function between vector spaces is a linear transformation. That's because linear transformations are "homomorphisms of vector spaces".)

Example: Prove that $\sigma : (M_{2\times 2}(\mathbb{R}), \text{matrix multiplication}) \to (\mathbb{R}, \cdot)$ defined by

$$\sigma \begin{pmatrix} a & b \\ c & d \end{pmatrix} = a$$

is not a homomorphism.

6.2 Isomorphisms and invariants

Here are three examples of algebraic structures (they are all fields, in fact):

SET	DEFINITION OF ADDITION	DEFINITION OF MULTIPLICATION
$\mathbb{Z}/2\mathbb{Z}$	$+ \mod 2$	$\cdot \mod 2$
$\{1, -1\}$	$\cdot$	$\max$
$\{\emptyset, E\}$	$\triangle$	$\cap$

I claim these three fields are really the "same" field. Why?

Addition tables

$(\mathbb{Z}, 2\mathbb{Z}, +)$	$0 + 2\mathbb{Z}$	$1 + 2\mathbb{Z}$
$0 + 2\mathbb{Z}$		
$1 + 2\mathbb{Z}$		

$(\{1, -1\}, \cdot)$	1	-1
1		
-1		

$(\{\emptyset, E\}, \triangle)$	$\emptyset$	E
$\emptyset$		
E		

Multiplication tables

$(\mathbb{Z}, 2\mathbb{Z}, \cdot)$	$0 + 2\mathbb{Z}$	$1 + 2\mathbb{Z}$
$0 + 2\mathbb{Z}$		
$1 + 2\mathbb{Z}$		

$(\{1, -1\}, \max)$	1	-1
1		
-1		

$(\{\emptyset, E\}, \cap)$	$\emptyset$	E
$\emptyset$		
E		

Observations:

Definition 6.2 *Let $(R, +, \cdot)$ and $(R', +, \cdot)$ be rings. A function $\sigma : R \to R'$ is a* **ring isomorphism** *if*

1. *σ is a homomorphism from $(R, +)$ to $(R', +)$;*
2. *σ is a homomorphism from $(R, \cdot)$ to $(R, \cdot)$; and*
3. *σ is a bijection (i.e. σ is injective and surjective).*

If there is a ring isomorphism $\sigma : R \to R'$, we say R and R' are **isomorphic** *and write $(R, +, \cdot) \cong (R', +, \cdot)$ (or just $R \cong R'$).*

If R and R' are fields, we can say that σ is a **field isomorphism**.

Concept: Two isomorphic rings are essentially the same object (their operations are performing the same algebra, but in different symbols and/or language). An isomorphism is a function which acts as a "translator", converting the language of one ring to the other.

Example: Let $\sigma : (\mathbb{Z}/2\mathbb{Z}, + \mod 2, \cdot \mod 2) \to (\{1, -1\}, \cdot, \max)$ be defined by

$$\sigma(0 + 2\mathbb{Z}) = 1 \qquad \sigma(1 + 2\mathbb{Z}) = -1.$$

This is a field isomorphism (from the previous page, σ preserves the addition and multiplication tables, and σ is clearly $1 - 1$ and onto).

Example: Let $\sigma : (\mathbb{C}, +, \cdot) \to (\mathbb{C}+, \cdot)$ be $\sigma(z) = \overline{z}$.

- Let $z, w \in \mathbb{C}$. $\sigma(z + w) = \overline{z + w} = \overline{z} + \overline{w} = \sigma(z) + \sigma(w)$.
- Let $z, w \in \mathbb{C}$. $\sigma(zw) = \overline{zw} = \overline{z}\,\overline{w} = \sigma(z)\sigma(w)$.
- Let $z, w \in \mathbb{C}$ be such that $\sigma(z) = \sigma(w)$. Thus $\overline{z} = \overline{w}$. Conjugating both sides again, we see $\overline{\overline{z}} = \overline{\overline{w}}$, i.e. $z = w$. Thus σ is $1 - 1$.
- Let $z \in \mathbb{C}$. Notice that $\sigma(\overline{z}) = \overline{\overline{z}} = z$ so σ is onto.

Therefore σ is a field isomorphism.

HOW TO PROVE two rings (fields) are isomorphic:
Write down a specific σ mapping one ring (field) to the other. Prove that σ is a ring (field) isomorphism.

How do you prove two rings aren't isomorphic?

Theorem 6.3 *$(\mathbb{C}, +, \cdot)$ and $(\mathbb{R}, +, \cdot)$ are not isomorphic.*

PROOF Suppose not, i.e. that $\sigma : \mathbb{C} \to \mathbb{R}$ is a field isomorphism.

Claim 1: $\sigma(0) = 0$.

Claim 2: $\sigma(1) = 1$.

Now consider the following equation in $\mathbb{C}$:

$$i^2 + 1 = 0$$

The crux of the above proof is that $\mathbb{C}$ has an element i with $i^2 = -1$, whereas $\mathbb{R}$ has no such element.

Definition 6.4 *An* **invariant** *is a property that is preserved by isomorphism. In other words, if property P is an invariant and rings R and R' are isomorphic, then either both R and R' have property P, or neither R nor R' have property P.*

Examples:
- Cardinality (# of elements) (obvious since any isomorphism σ is a bijection)
- Whether or not the ring is an integral domain (HW)
- Whether or not the ring is a field (HW)

To prove two rings are not isomorphic, it is often easiest to exhibit an invariant which one has but the other doesn't.

Chinese Remainder Theorem

Theorem 6.5 (Chinese Remainder Theorem) *Suppose $m, n \in \mathbb{Z}$ are nonzero, non-units such that $\gcd(m, n) = 1$. Then*

$$\mathbb{Z}/(mn)\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}.$$

PROOF We start by defining a map $\sigma : \mathbb{Z}/(mn)\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ by setting

$$\sigma(x + mn\mathbb{Z}) = (x + m\mathbb{Z}, x + n\mathbb{Z}).$$

Claim 1: σ is well-defined.
Suppose $x + mn\mathbb{Z} = y + mn\mathbb{Z}$. Thus $mn \mid (y - x)$ so $y - x = kmn$ for $k \in \mathbb{Z}$. That means

$$y - x = m(kn) \Rightarrow y \equiv x \mod m \text{ and } y - x = n(km) \Rightarrow y \equiv x \mod n,$$

meaning σ is well-defined.

Claim 2: σ is an additive homomorphism. To show this, let $x, y \in \mathbb{Z}$. Then

$$\begin{aligned}\sigma((x + mn\mathbb{Z}) + (y + mn\mathbb{Z})) &= \sigma(x + y + mn\mathbb{Z}) \\ &= (x + y + m\mathbb{Z}, x + y + n\mathbb{Z}) \\ &= (x + m\mathbb{Z}, x + n\mathbb{Z}) + (y + m\mathbb{Z}, y + n\mathbb{Z}) \\ &= \sigma(x + mn\mathbb{Z}) + \sigma(y + mn\mathbb{Z}).\end{aligned}$$

Claim 3: σ is a multiplicative homomorphism.
To show this, let $x, y \in \mathbb{Z}$. Then

$$\begin{aligned}\sigma((x + mn\mathbb{Z})(y + mn\mathbb{Z})) &= \sigma(xy + mn\mathbb{Z}) \\ &= (xy + m\mathbb{Z}, xy + n\mathbb{Z}) \\ &= (x + m\mathbb{Z}, x + n\mathbb{Z})(y + m\mathbb{Z}, y + n\mathbb{Z}) \\ &= \sigma(x + mn\mathbb{Z})\sigma(y + mn\mathbb{Z}).\end{aligned}$$

Claim 4: σ is onto.
To show this, let $(a + m\mathbb{Z}, b + n\mathbb{Z})$ be an arbitrary element of $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. By Bezout's Theorem, there are integers k, l such that

$$km + ln = 1.$$

In particular, this means $km + n\mathbb{Z} = 1 + n\mathbb{Z}$ and $ln + m\mathbb{Z} = 1 + m\mathbb{Z}$. Now, consider $x = bkm + aln + mn\mathbb{Z} \in \mathbb{Z}/(mn)\mathbb{Z}$.

$$\begin{aligned}\sigma(x) &= (bkm + aln + m\mathbb{Z}, bkm + aln + n\mathbb{Z}) \\ &= (aln + m\mathbb{Z}, bkm + n\mathbb{Z}) \\ &= (a1 + m\mathbb{Z}, b1 + n\mathbb{Z}) \\ &= (a + m\mathbb{Z}, b + n\mathbb{Z}).\end{aligned}$$

Claim 5: σ is $1-$ by the Pigeonhole Principle (since $\mathbb{Z}/(mn)\mathbb{Z}$ and $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ both have cardinality mn, any surjection between the sets is automatically a bijection).

Claims 1-5 show σ is a ring isomorphism. □

Application: Suppose we want to solve the system of congruences

$$\begin{cases} x \equiv 3 \mod 10 \\ x \equiv 5 \mod 7 \end{cases}.$$

The Chinese Remainder Theorem says that this system has a unique solution $\mod 70$, and to find the solution, you can use the method we used to prove Claim 4 in the Chinese Remainder Theorem: first, write $\gcd(10, 7)$ as a linear combination of 7 and 10:

$$\begin{aligned}10 &= 1 \cdot 7 + 3 \\ 7 &= 2 \cdot 3 + 1 \\ 3 &= 3 \cdot 1 + 0 \Rightarrow \gcd(10, 7) = 1\end{aligned}$$

$$1 = 7 - 2(3) = 7 - 2(10 - 1 \cdot 7) = 3 \cdot 7 - 2 \cdot 10 = -2(10) + 3(7)$$

In the context of the proof of Claim 4, that means $k = -2$ and $l = 3$. The solution to the system of congruences is therefore

$$x = bkm + aln + mn\mathbb{Z} = 5(-2)(10) + 3(3)(7) + 10(7)\mathbb{Z} = -37 + 70\mathbb{Z} = 33 + 70\mathbb{Z}.$$

Corollary 6.6 *Let ϕ be the Euler phi function. If $\gcd(m, n) = 1$, then $\phi(mn) = \phi(m)\phi(n)$.*

PROOF Recall that $\phi(mn)$ is the number of units in $\mathbb{Z}/(mn)\mathbb{Z}$.

Observe that $(u + m\mathbb{Z}, v + n\mathbb{Z})$ is a unit in $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ if and only if $u + m\mathbb{Z}$ is a unit in $\mathbb{Z}/m\mathbb{Z}$ and $v + n\mathbb{Z}$ is a unit in $\mathbb{Z}/n\mathbb{Z}$, so the number of units in $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ is $\phi(m)\phi(n)$.

Since $\mathbb{Z}/(mn)\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, these two rings have the same number of units (since units are preserved by ring homomorphisms). The corollary follows. □

6.3 Ring homomorphisms

It turns out that even if a would-be isomorphism isn't bijective, it still has value, so we give such a map another name:

Definition 6.7 *Let $(R, +, \cdot)$ and $(R', +, \cdot)$ be rings. A function $\sigma : R \to R'$ is a* **ring homomorphism** *if, for any $x, y \in R$, we have*

1. *σ preserves addition, i.e. $\sigma(x + y) = \sigma(x) + \sigma(y)$; and*
2. *σ preserves multiplication, i.e. $\sigma(xy) = \sigma(x)\sigma(y)$.*

If $\sigma : R \to R'$ preserves $+$ and $\cdot$, lots of other stuff comes for free:

Lemma 6.8 (Properties of ring homomorphisms) *Let $\sigma : R \to R'$ be a ring homomorphism. Let $x \in R$. Then:*

1. *σ preserves additive identities, i.e. $\sigma(0) = 0$;*
2. *σ preserves additive inverses, i.e. $\sigma(-x) = -\sigma(x)$;*
3. *If σ is nontrivial, then σ preserves multiplicative identities, i.e. $\sigma(1) = 1$;*
4. *If σ is nontrivial, then σ preserves units, i.e. if x is a unit in R, then $\sigma(x)$ is a unit in R'.*

PROOF (1): Let $x \in R$. Since σ is a ring homomorphism,

$$\sigma(x) = \sigma(0 + x) = \sigma(0) + \sigma(x).$$

By the cancellation law (in R'), $\sigma(0) = 0$.

(2): Let $x \in R$. Since σ is a ring homomorphism,

Therefore $\sigma(-x)$ must equal $-\sigma(x)$ by uniqueness of additive inverses (in R').

(3) is a HW problem.

(4): Suppose $x \in R$ is a unit. Thus there is $y \in R$ such that $xy = 1$. Then

$$1 = \sigma(1) = \sigma(xy) = \sigma(x)\sigma(y)$$

so $\sigma(x)$ divides 1 in R', i.e. $\sigma(x)$ is a unit in R'. □

Standard examples of ring homomorphisms

Zero homomorphism: Let R and R' be any two rings, and define $\sigma : R \to R'$ by $\sigma(x) = 0$ for all $x \in \mathbb{R}$. σ is a ring homomorphism.

Identity map: Let R be a ring. Then $I_R : R \to R$ defined by $I_R(x) = x$ is a ring homomorphism.

Evaluation maps: Let F be a field, and let $\alpha \in F$. Then $\sigma : F[x] \to F$ defined by $\sigma(f) = f(\alpha)$ is a ring homomorphism.

Quotient maps (integer setting): Let $n \in \mathbb{Z}$ be nonzero. Then $\sigma : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ defined by $\sigma(x) = x + n\mathbb{Z}$ is a ring homomorphism.

Quotient maps (polynomial setting): Let F be a field and let $l \in F[x]$ be nonzero. Then $\sigma : F[x] \to F[x]/lF[x]$ defined by $\sigma(f) = f + lF[x]$ is a ring homomorphism.

Compositions of homomorphisms: Let R, R' and R'' be rings. If $\sigma : R \to R'$ and $\tau : R' \to R''$ are ring homomorphisms, then so is $\tau \circ \sigma : R \to R''$ (HW).

Conjugation (on $\mathbb{C}$): $\sigma : \mathbb{C} \to \mathbb{C}$ defined by $\sigma(z) = \overline{z}$ is a ring homomorphism.

(This is Example D from earlier in this chapter.)

Kernels of homomorphisms

Definition 6.9 *Let $\sigma : R \to R'$ be a ring homomorphism, and let 0 denote the additive identity element of R'. The* **kernel** *of σ, denoted $\ker(\sigma)$, is the set*

$$\ker(\sigma) = \{x \in R : \sigma(x) = 0\}.$$

Example C (from earlier): $\sigma : \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ defined by $\sigma(x + 6\mathbb{Z}) = x + 2\mathbb{Z}$.

Example D (from earlier): $\sigma : \mathbb{C} \to \mathbb{C}$ defined by $\sigma(z) = \overline{z}$.

Example B (from earlier): quotient map $\sigma : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$

Since the identity of $\mathbb{Z}/n\mathbb{Z}$ is $0 + n\mathbb{Z}$,

$$\ker(\sigma) = \{x \in \mathbb{Z} : \sigma(x) = 0 + n\mathbb{Z}\} = n\mathbb{Z}.$$

Lemma 6.10 (Properties of kernels) *Let $\sigma : R \to R'$ be a ring homomorphism. Let 0 denote the additive identity element of R. Then:*

1. $0 \in \ker(\sigma)$*;*
2. $\ker(\sigma) = \{0\}$ *if and only if σ is injective;*
3. *if $x, y \in \ker(\sigma)$, then $x + y \in \ker(\sigma)$;*
4. *if $x \in \ker(\sigma)$, then $xy \in \ker(\sigma)$ for any $y \in R$.*

PROOF (1) follows from (1) of Lemma 6.8.

(2): ($\Rightarrow$) Suppose $\ker(\sigma) = \{0\}$. Now suppose $x, y \in R$ are such that $\sigma(x) = \sigma(y)$. Consider $\sigma(x - y) = \sigma(x) - \sigma(y) = 0$. This means $x - y \in \ker(\sigma)$, so $x - y = 0$, so $x = y$. That means σ is injective.

($\Leftarrow$) Suppose $\ker(\sigma) \neq \{0\}$; that means there is $k \neq 0$ belonging to $\ker(\sigma)$. Thus $\sigma(0) = \sigma(k)$ so σ is not injective. By contraposition, we are done with (2).

(3): Suppose $x, y \in \ker(\sigma)$. That means $\sigma(x) = 0$ and $\sigma(y) = 0$. Thus

$$\sigma(x + y) = \sigma(x) + \sigma(y) = 0 + 0 = 0$$

so $x + y \in \ker(\sigma)$ as wanted.

(4) is a HW problem (it is similar to (3)). □

Note: A surjective ring homomorphism is a ring isomorphism iff its kernel is $\{0\}$.

Quotient rings

Definition 6.11 *Let $\sigma : R \to \mathbb{R}'$ be a ring homomorphism. Define the relation $\equiv_\sigma$ on R by*

$$x \equiv_\sigma y \Leftrightarrow \sigma(x) = \sigma(y).$$

Prototype example: $\sigma : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ defined by $\sigma(x) = x + n\mathbb{Z}$. Then $\equiv_\sigma$ is the relation of equivalence modulo n discussed in detail in Chapter 2.

Lemma 6.12 *Let $\sigma : R \to R'$ be a ring homomorphism. Then $\equiv_\sigma$ is an equivalence relation, and the equivalence class of $x \in R$ is*

$$x + \ker(\sigma) = \{x + k : k \in \ker(\sigma)\}.$$

PROOF First, we prove $\equiv_\sigma$ is an equivalence relation:

- $\sigma(x) = \sigma(x)$ obviously, so $\equiv_\sigma$ is reflexive.
- If $\sigma(x) = \sigma(y)$, then $\sigma(y) = \sigma(x)$, so $\equiv_\sigma$ is symmetric.
- If $\sigma(x) = \sigma(y)$ and $\sigma(y) = \sigma(z)$, then clearly $\sigma(x) = \sigma(z)$, so $\equiv_\sigma$ is transitive.

Second, let $x, y \in R$. Then

$$\begin{aligned} y \equiv_\sigma x &\Leftrightarrow \sigma(x) = \sigma(y) \\ &\Leftrightarrow \sigma(y) - \sigma(x) = 0 \\ &\Leftrightarrow \sigma(y - x) = 0 \\ &\Leftrightarrow y - x \in \ker(\sigma) \\ &\Leftrightarrow y = x + k \text{ for } k \in \ker(\sigma) \\ &\Leftrightarrow y \in x + \ker(\sigma). \ \square \end{aligned}$$

Example B (from earlier): $\sigma : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ quotient map; $\ker(\sigma) = n\mathbb{Z}$

Example D (from earlier): $\sigma : \mathbb{C} \to \mathbb{C}$ defined by $\sigma(z) = \overline{z}$; $\ker(\sigma) = \{0\}$

Example C (from earlier): $\sigma : \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ defined by $\sigma(x + 6\mathbb{Z}) = x + 2\mathbb{Z}$; $\ker(\sigma) = \{0 + 6\mathbb{Z}, 2 + 6\mathbb{Z}, 4 + 6\mathbb{Z}\}$.

Theorem 6.13 *Let $\sigma : R \to R'$; denote the set of $\equiv_\sigma$-equivalence classes by $R/\ker(\sigma)$. This set is called a* **quotient ring (of R by $\ker(\sigma)$)**. *$R/\ker(\sigma)$ is a ring, where the addition and multiplication are defined as follows:*

$$(x + \ker(\sigma) + (y + \ker(\sigma)) = (x + y) + \ker(\sigma)$$

$$(x + \ker(\sigma)(y + \ker(\sigma)) = xy + \ker(\sigma)$$

In particular, the additive identity of $R/\ker(\sigma)$ is $0 + \ker(\sigma)$ and the multiplicative identity of $R/\ker(\sigma)$ is $1 + \ker(\sigma)$.

PROOF The main thing to prove is that this addition and multiplication are well-defined. These proofs are essentially the same as the proofs that $+$ and $\cdot$ are well-defined on $\mathbb{Z}/n\mathbb{Z}$ (Chapter 2) and are left as HW. The ring axioms need to be checked, but they all hold. $\square$.

So what?

Theorem 6.14 (First Isomorphism Theorem) *Let $\sigma : R \to R'$ be a surjective ring homomorphism. Then*

$$R/\ker(\sigma) \cong R'.$$

PROOF Define $\tau : R/ker(\sigma) \to R'$ by

$$\tau(x + \ker(\sigma)) = \sigma(x).$$

Claim 1: τ is well-defined.
To show this, suppose $x + \ker(\sigma) = y + \ker(\sigma)$. Then $x - y \in \ker(\sigma)$, so

$$0 = \sigma(x - y) = \sigma(x) - \sigma(y).$$

Thus $\tau(x + \ker(\sigma)) = \sigma(x) = \sigma(y) = \tau(y + \ker(\sigma))$, so τ is well-defined.

Claim 2: τ is an additive homomorphism.
To show this, let $x, y \in R$. Then

$$\begin{aligned}\tau((x + \ker(\sigma)) + (y + \ker(\sigma))) &= \tau(x + y + \ker(\sigma)) \\ &= \sigma(x + y) \\ &= \sigma(x) + \sigma(y) \\ &= \tau(x + \ker(\sigma)) + \tau(y + \ker(\sigma)).\end{aligned}$$

Claim 3: τ is a multiplicative homomorphism.
To show this, let $x, y \in R$. Then

$$\begin{aligned}\tau((x + \ker(\sigma))(y + \ker(\sigma))) &= \tau(xy + \ker(\sigma)) \\ &= \sigma(xy) \\ &= \sigma(x)\sigma(y) \\ &= \tau(x + \ker(\sigma))\tau(y + \ker(\sigma)).\end{aligned}$$

Claim 4: τ is onto. To show this, let $y \in R'$. Since σ is surjective, there is $x \in R$ such that $\sigma(x) = y$. Thus $\tau(x + \ker(\sigma)) = \sigma(x) = y$ so τ is onto.

Claim 5: τ is $1 - 1$. To show this, suppose $x + \ker(\sigma) \in \ker(\tau)$. Thus $\sigma(x) = 0$ so $x \in \ker(\sigma)$ so $x + \ker(\sigma) = 0 + \ker(\sigma)$. Thus $\ker(\tau) = \{0 + \ker(\sigma)\}$, so τ is $1 - 1$.

Claims 1-5 show that σ is a ring isomorphism. □

Corollary 6.15 *Let F be a number field and let α be algebraic over F. Let $h \in F[x]$ be a minimal polynomial for α (i.e. an irreducible polynomial such that $h(\alpha) = 0$). Then*

$$F[x]/hF[x] \cong F(\alpha).$$

PROOF Define $\sigma : F[x] \to F(\alpha)$ by $\sigma(f) = f(\alpha)$. σ is an evaluation map, hence a ring homomorphism. By Theorem 5.18, σ is onto (this theorem says that

$$\{1, \alpha, \alpha^2, ..., \alpha^{n-1}\}$$

spans $F(\alpha)$). Therefore, this result follows from the First Isomorphism Theorem if we can show $\ker(\sigma) = hF[x]$. We show this here:

$$\begin{aligned} f \in \ker(\sigma) &\Leftrightarrow f(\alpha) = 0 \\ &\Leftrightarrow f = hq \quad \text{(by the General Factor Theorem from Chapter 4)} \\ &\Leftrightarrow f \in hF[x]. \square \end{aligned}$$

Theorem 6.16 (Conjugation Theorem) *Suppose $\alpha, \beta \in \mathbb{C}$ are algebraic over number field F. TFAE:*

1. *α and β are roots of the same irreducible polynomial $h \in F[x]$.*
2. *There is a field isomorphism $\sigma : F(\alpha) \to F(\beta)$ with $\sigma(\alpha) = \beta$ and $\sigma(x) = x$ for all $x \in F$.*

PROOF ($\Rightarrow$) The preceding corollary says

$$F(\alpha) \;\cong\; F[x]/hF[x] \;\cong\; F(\beta)$$

via evaluation maps which send elements as follows:

$$\alpha \;\leftarrow\!\shortmid\; x + hF[x] \;\mapsto\; \beta$$

If $c \in F$, then the maps work like this:

$$c \;\leftarrow\!\shortmid\; c + hF[x] \;\mapsto\; c$$

By composition, we get the isomorphism σ with the appropriate properties.

($\Leftarrow$) Suppose there exists an isomorphism σ with the appropriate properties given in (2). Let h be a minimal polynomial for α over F; then $h(\alpha) = 0$. Applying σ to both sides of this equation, we get

$$0 = \sigma(h(\alpha)) = h(\sigma(\alpha)) = h(\beta)$$

proving the theorem. □

Application: The map $\sigma : \mathbb{Q}(\sqrt[3]{2}) \to \mathbb{Q}(\zeta_3\sqrt[3]{2})$ defined by $\sigma(x) = x$ for all $x \in \mathbb{Q}$ and $\sigma(\sqrt[3]{2}) = \zeta_3\sqrt[3]{2}$ is an isomorphism.

Application: The only isomorphism $\sigma : \mathbb{Q}(\sqrt[3]{2}) \to \mathbb{Q}(\sqrt[3]{2})$ is the identity map.

Theorem 6.17 (Isomorphism Extension Theorem) *Let F, F' be fields. Then any field isomorphism $\sigma : F \to F'$ extends to a (ring) isomorphism $\sigma : F[x] \to F'[x]$ by setting $\sigma(x) = x$.*

PROOF This involves checking isomorphism properties, which is straight-forward but tedious, so I won't do it here. See page 94 of Stillwell for the details.

6.4 Automorphisms

Definition 6.18 *Let $(R, +, \cdot)$ be a ring. A ring isomorphism $\sigma : R \to R$ is called an* **automorphism** *of R. The set of automorphisms of R is denoted $Aut(R)$.*

Example: $Aut(\mathbb{Q}) = \{I_{\mathbb{Q}}\}$ (earlier HW).

Example: $Aut(\mathbb{R}) = \{I_{\mathbb{R}}\}$ (essentially the same proof as for $\mathbb{Q}$).

Example: $Aut(\mathbb{C}) \ni \{I_{\mathbb{C}}, \sigma\}$ where $\sigma(z) = \overline{z}$. Are there other automorphisms?

Theorem 6.19 (Group properties of automorphisms) *Let R be a ring. Then:*

$Aut(R)$ **is closed under composition:** *If $\sigma, \tau \in Aut(R)$, then $\sigma \circ \tau \in Aut(R)$.*

$Aut(R)$ **contains an identity:** *The identity function $I_R \in Aut(R)$ (where $I_R(x) = x$ for all $x \in R$).*

$Aut(R)$ **is closed under inverses:** *If $\sigma \in Aut(R)$, then $\sigma^{-1} \in Aut(R)$.*

WARNING: There is no reason to believe that $Aut(F)$ is commutative under composition.

Theorem 6.20 (Automorphism Extension Theorem) *Let F be a number field and let α be algebraic over F. Then any $\sigma \in Aut(F)$ extends to an automorphism σ of $F(\alpha)$ by setting $\sigma(\alpha) = \alpha$.*

PROOF This involves checking isomorphism properties, which is straight-forward but tedious, so I won't do it here. See page 94-95 of Stillwell for the details.

Chapter 7

Groups

7.1 An update on the big picture

One of our motivating problems from the first week of the course was solving polynomial equations. More specifically, we want to describe the roots of polynomial p where $p \in \mathbb{Q}[x]$. Here's what we know now:

In general

Let F be any field, and let $p \in F[x]$. The equation $p(x) = 0$ has at most $\deg p$ solutions. By the Factor Theorem, $x_0 \in F$ is a root of $p \in F[x]$ if and only if

$$(x - x_0) \mid p(x).$$

If $p \in \mathbb{Q}[x]$, then every solution of such an equation belongs to the algebraic closure $\widehat{\mathbb{Q}}$ of $\mathbb{Q}$, which is a countable subfield of $\mathbb{C}$. The degree of any root of p over $\mathbb{Q}$ is at most $\deg p$.

The Fundamental Theorem of Algebra guarantees that if $p \in \mathbb{C}[x]$, then p has a root in $\mathbb{C}$ (so it has exactly p roots, counting multiplicities).

If $p \in \mathbb{R}[x]$, then for any complex root z of p, the conjugate $\overline{z}$ must also be a root.

π is transcendental over $\mathbb{Q}$, meaning it is not a root of any polynomial $p \in \mathbb{Q}[x]$.

Linear equations

Let F be any field. If $p \in F[x]$ has degree 1, then every equation $p(x) = 0$ has exactly one solution in F. This solution is obtainable with only the operations $+, -, \cdot, \div$.

Quadratic equations

Let F be any number field. Then for any $p \in F[x]$ which has degree 2, the roots of p either:

- belong to F, in which case p is reducible over F; or
- belong to the same 2-dimensional extension E of F, in which case p is irreducible over F and both roots of p have degree 2 over F. Furthermore, there is an automorphism of E which sends every element of F to itself and interchanges the two roots of p.

Any quadratic equation whose coefficients belong to $\mathbb{C}$ is solvable via the quadratic formula using $+, -, \cdot, \div, \sqrt{}$, and all these operations are actually necessary to solve an arbitrary quadratic equation (as an example, $x^2 - 2 = 0$ has no solution in the rationals).

More generally, any constructible (a.k.a. surd) number must have degree 2^m over $\mathbb{Q}$ for some natural number m. This means, for instance:

- $\sqrt[3]{2}$ is not surd, i.e. doubling the cube is not possible;
- π is not constructible, i.e. squaring the circle is not possible;
- $\cos 20^\circ$ is not surd, i.e. trisecting a 60° angle is not possible;
- if p is a prime such that $p - 1$ is not a power of 2, then $\cos \frac{2\pi}{n}$ is not surd, and therefore the regular p-gon is not constructible.

Cubic equations

Let F be any number field. Then for any $p \in F[x]$ with $\deg p = 3$, the roots of p have degree at most 3 over F. There are two disjoint possibilities:

- If there is a root of p whose degree is less than 3, then p is reducible over f, and p must have a root in F.
- Otherwise, p is irreducible over F, all roots of p have degree 3 over F, and each root belongs to a 3-dimensional extension of F (different roots might not lie in the same 3-dimensional extension, however).

Cubic equations whose coefficients belong to $\mathbb{C}$ can be solved using the cubic formula of del Ferro and Tartaglia using $+, -, \cdot, \div, \sqrt{}, \sqrt[3]{}$, and these operations are all actually necessary to solve arbitrary cubic equations (as an example, $x^3 - 2 = 0$ has root $\sqrt[3]{2}$ which is not surd).

Quartic equations

Let F be any number field. Then for any $p \in F[x]$ with $\deg p = 4$, the roots of p have degree at most 4 over F. There are two disjoint possibilities:

- p has a root in F, in which case p factors over F into a linear term and a cubic (the cubic can be studied by the methods above).
- p is reducible over F but has no root in F. In this case, p factors over F into two irreducible quadratics, and the four roots of p all have degree 2 over F.
- p is irreducible over F, all roots of p have degree 4 over F, and each belongs to a 4-dimensional extension of F.

Quartic equations whose coefficients belong to $\mathbb{C}$ can be solved using the "quartic formula" (a method similar to that of del Ferro and Tartaglia) using the operations $+, -, \cdot, \div, \sqrt{}, \sqrt[3]{}$ (you don't need $\sqrt[4]{}$ only because $\sqrt[4]{} = \sqrt{\sqrt{}}$), and these operations are all actually necessary to solve arbitrary quartic equations.

Quintic equations (and beyond)

Let F be any number field. Then for any $p \in F[x]$ with $\deg p = 5$, the roots of p have at most 5 over F. If p is reducible, then we can analyze each factor of p separately by the above cases. But if p is irreducible, we don't know very much. We are interested in the following questions:

- Is an arbitrary quintic solvable by radicals?
- Is there a "quintic formula"?
- If so, what is it and what operations does it use?
- If not, why not?

Strategy

Take polynomial $p \in \mathbb{Q}[x]$ and consider the equation $p(x) = 0$. This equation has at most $\deg p$ solutions in $\mathbb{C}$.

Then, make a new field by adjoining the roots of p to $\mathbb{Q}$; call this field F.

New idea: look at the set $Aut(F)$ of automorphisms of F. It turns out that this set has some algebraic structure which can be interpreted geometrically as the symmetries of some figure. This translates back into information about when and why you can (or can't) solve a polynomial equation by radicals.

7.2 What is a group?

We are aiming for an algebraic structure on $Aut(F)$, where F is some field. Unlike rings and fields (where there are two natural operations $+$ and $\cdot$, there's only one natural operation on $Aut(F)$.

The elements of $Aut(F)$ are

and the only thing you know can do with these types of objects to obtain another element of $Aut(F)$ is

Definition 7.1 (Definition of group) *An algebraic system $(G, \odot)$ is called a* **group** *if it has the following four properties:*

G **is closed under $\odot$:** *If $x, y \in G$, then $x \odot y \in G$.*

$\odot$ is associative: *If $x, y, z \in G$, then $(x \odot y) \odot z = x \odot (y \odot z)$.*

$\odot$ has an identity: *There is an element $e = e_G \in G$ such that for all $x \in G$, $e \odot x = x \odot e = x$.*

Elements of G have inverses under $\odot$: *For any $x \in G$, there is an element $x^{-1} \in G$ such that $x^{-1} \odot x = x \odot x^{-1} = e$.*

The operation $\odot$ is called the **composition rule** *for the group.*

Definition 7.2 *Let G be a group. The number of elements of G is called the* **order** *of G and is denoted $|G|$.*

Look back at Theorem 6.19 for some motivation as to where these properties come from. By that theorem, the set of automorphisms of any field form a group under composition.

With groups, we tend to write the composition rule $\odot$ without using any symbol. So instead of writing

$$g \odot h \quad \text{or} \quad g + h \quad \text{or} \quad g \cdot h \quad \text{or} \quad g \circ h,$$

we just write gh.

WARNING: Nothing in the definition of group requires that the composition law is commutative, so in general, in group G,

$$gh \neq hg.$$

Definition 7.3 *Let G be a group. If the operation $\odot$ is commutative (i.e $gh = hg$ for all $g, h \in G$), we say that $(G, \odot)$ is an* **abelian** *group.*

The word "abelian" is in honor of the Norwegian mathematician Niels Abel.

When performing group operations, we write g^2 for gg, and g^5 means $ggggg$, etc. What does g^{-3} mean?

So a group is abelian if $gh = hg$ for all $g, h \in G$. But in general, $gh \neq hg$ in a group.

The notation "gh" for the composition rule of a group can be confusing, because sometimes "composition" really means addition or multiplication. See the chart below:

IF THE BINARY OPERATION IN THE GROUP IS ...	WHEN WE WRITE "gh", WE MEAN...	WHEN WE WRITE "g^4", WE MEAN...	WHEN WE WRITE "g^{-1}", WE MEAN...	THE IDENTITY ELEMENT e MEANS...
addition $+$				
multiplication $\cdot$				
composition of functions $\circ$				

Remark: Remember that you compose functions from right to left. So if you write gh where $g, h \in G$, you should think of doing h first, then g.

Properties of groups

Theorem 7.4 (Properties of groups) *Let G be a group. Then:*

Identity is unique: *there is only one identity element of G.*
Inverses are unique: *for any $g \in G$, there is only one element $g^{-1} \in G$ which is an inverse of g.*

Inverse of the identity is the identity: $e^{-1} = e$.
Inverse of an inverse: *for any $g \in G$,* $(g^{-1})^{-1} = g$.
Inverse of a product: *for any $g, h \in G$,* $(gh)^{-1} = h^{-1}g^{-1}$.

Left cancellation law: *if $gh = gk$, then $h = k$.*
Right cancellation law: *if $hg = kg$, then $h = k$.*
Abelian cancellation law: *if G is abelian and $gh = kg$ or $hg = gk$, then $h = k$.*

Left or right inverse must be an inverse: *$gh = e$ if and only if $hg = e$ if and only if $h = g^{-1}$.*

Exponent rules: *if $g \in G$ and $m, n \in \mathbb{Z}$, then $g^m g^n = g^{m+n}$ and $(g^m)^n = g^{mn}$.*

PROOF For the first statement, suppose there are two identity elements e and e'. Then

$$ee' = \begin{cases} e & \text{since } e' \text{ is an identity} \\ e' & \text{since } e \text{ is an identity} \end{cases}$$

Thus $e = e'$, meaning there is only one identity element.

Uniqueness of inverses is a HW problem.

Third, $ee = e$ so $e = e^{-1}$.

Fourth, $gg^{-1} = g^{-1}g = e$ since g^{-1} is an inverse of g, so g is the inverse of g^{-1}.

Fifth, by applying associativity, $(gh)(h^{-1}g^{-1}) = ghh^{-1}g^{-1} = geg^{-1} = gg^{-1} = e$, so $h^{-1}g^{-1}$ is the inverse of gh as wanted.

For the left cancellation law, suppose $gh = gk$. Multiply both sides of this equation on the left by g^{-1} to get

$$g^{-1}gh = g^{-1}gk,$$

i.e. $h = k$ as wanted. For the right cancellation law, multiply both sides of the equation on the right by g^{-1}. If G is abelian, then either of the given equations

$gh = kg$ or $hg = gk$ are equivalent to $gh = gk$, from which the left cancellation law can be applied to give $h = k$.

Suppose $gh = e$. Then $gh = gg^{-1}$ so by the left cancellation law, $h = g^{-1}$. Similarly, if $hg = e = g^{-1}g$, by the right cancelllation law $h = g^{-1}$.

Last, the exponent rules are obvious by associativity. □

Application: in a group G, you can always solve $gx = h$ and $xg = h$ for x:

Example of a group: $(\mathbb{Z}, +)$. This is an abelian group, because addition on $\mathbb{Z}$ is commutative.

This is a terrible example to start with, because

(1) $(\mathbb{Z}, +)$ is abelian, whereas most groups aren't, and

(2) the elements of $\mathbb{Z}$ are *numbers* that are *added* under the group operation, and really, you should think of a group as being comprised of *functions* being *composed* under the group operation.

How can you think of $\mathbb{Z}$ as as being a collection of functions?

Group or not a group?

Exercise: Recall that a group is an algebraic system where:

1. the set is closed under the operation;
2. the operation is associative;
3. the operation has an identity; and
4. every element in the set has an inverse under the operation.

Decide whether each of the algebraic systems below comprises a group (no proofs needed). If the system isn't a group, explain which of the four items above fails:

1. $(\mathbb{Z}, \cdot)$
2. $(2\mathbb{Z}, +)$
3. $(\mathbb{Z}/8\mathbb{Z}, +)$
4. $(\mathbb{Z}, -)$
5. $(M_2(\mathbb{R})$, matrix addition$)$
6. $(M_2(\mathbb{R})$, matrix multiplication$)$
7. $((0, \infty), \cdot)$
8. $(\mathbb{Q}, +)$
9. $(\mathbb{Q}, \cdot)$
10. $(GL_2(\mathbb{R})$, matrix addition$)$
    Notation: $GL_2(\mathbb{R})$ is the set of 2×2 invertible matrices with real entries.
11. $(GL_2(\mathbb{R})$, matrix multiplication$)$
12. $(2^E, \cup)$
13. $(2^E, \triangle)$

Subgroups

Definition 7.5 *Let G be a group. A subset $H \subseteq G$ is called a* **subgroup** *of G if it is itself a group under the same operation as G. Since H is automatically associative, this means H has to satisfy three things:*

H **contains the identity:** $e \in H$.

H **is closed under the composition rule:** *For any $h_1, h_2 \in H$, $h_1 h_2$ must be in H.*

H **is closed under inverses:** *If $h \in H$, then $h^{-1} \in H$.*

If H is a subgroup of G, we write $H \leq G$. If H is a subgroup of G which is not equal to all of G, we write $H < G$ and call H a **proper subgroup** *of G.*

Example: Let $G = (\mathbb{Z}, +)$ and let H be the set of odd numbers. Is H a subgroup of G?

Example: Let $G = (\mathbb{Z}, +)$ and let H be the set of perfect squares. Is H a subgroup of G?

Example: Let $G = (\mathbb{Z}, +)$ and let H be the set of positive numbers. Is H a subgroup of G?

Example: Let $G = (\mathbb{Z}, +)$ and let H be the set of even numbers. Is H a subgroup of G?

Every group G has at least two subgroups: $\{e\}$ and G itself. These are called **trivial subgroups** of G; a nontrivial subgroup of G is a subgroup H with $\{e\} < H < G$.

HOW TO PROVE H is a subgroup of G:

0. If it's not obvious, prove $H \subseteq G$.
   (Usually this is obvious or given and can be omitted.)
1. ... (logical argument) ... $e \in H$.
   (this shows H contains the identity)
2. Let $h_1, h_2 \in H$. ... (logical argument) ... Therefore $h_1 h_2 \in H$.
   (this shows H is closed under the composition rule)
3. Let $h \in H$. ... (logical argument) ... Therefore $h^{-1} \in H$.
   (this shows H is closed under inverses)

Therefore $H \leq G$. $\square$

HOW TO PROVE H is **NOT** a subgroup of G:

Do any one of the following things:

0. Write down an element of H that isn't in G.
1. Prove that $e \notin H$.
2. Write down two explicit elements of H whose composition is not in H.
3. Write down an explicit element of H whose inverse is not in H.

Example: Determine, with proof, whether or not $7\mathbb{Z} = \{7n : n \in \mathbb{Z}\}$ is a subgroup of $(\mathbb{Z}, +)$.

Group homomorphisms and isomorphisms

Definition 7.6 *Let G and G' be groups. A map $\sigma : G \to G'$ is called a* **(group) homomorphism** *if for every $g_1, g_2 \in G$,*

$$\sigma(g_1 g_2) = \sigma(g_1)\sigma(g_2).$$

A group homomorphism that is also a bijection is called a **(group) isomorphism***; if there is a group isomorphism $\sigma : G \to G'$ we say G and G' are* **isomorphic** *and write $G \cong G'$. An* **invariant** *is a property of a group that is preserved under isomorphism.*

As with isomorphic rings and fields, two isomorphic groups should be thought of as being "the same group expressed in a different language".

Theorem 7.7 (Properties of group homomorphisms) *Let $\sigma : G \to G'$ be a group homomorphism. Then:*

1. *σ preserves identities, i.e. $\sigma(e_G) = e_{G'}$;*
2. *σ preserves inverses, i.e. $\sigma(g^{-1}) = [\sigma(g)]^{-1}$ for all $g \in G$;*
3. *σ preserves subgroups, i.e. if $H \leq G$, then $\sigma(H) \leq G'$.*

PROOF (1): Let $g \in G$. Then

$$\sigma(g) = \sigma(e_G g) = \sigma(e_G)\sigma(g).$$

By the right cancellation law, $e_{G'} = \sigma(e_G)$ as wanted.

For (2), let $g \in G$. Then:

$$e_{G'} = \sigma(e_G) = \sigma(gg^{-1}) = \sigma(g)\sigma(g^{-1}).$$

Left-multiply both sides of this equation by $[\sigma(g)]^{-1}$ to get $[\sigma(g)]^{-1} = \sigma(g^{-1})$.

The last statement is a HW problem (you have to prove $1 \in \sigma(H)$, that $\sigma(H)$ is closed under composition, and that $\sigma(H)$ is closed under inverses). □

Kernels, quotient groups and the First Isomorphism Theorem for groups

The stuff we did in the previous chapter for rings carries over to groups:

Definition 7.8 *Let $\sigma : G \to G'$ be a group homomorphism. The* **kernel** *of σ, denoted $\ker(\sigma)$, is the set of elements which map to the identity under σ:*

$$\ker(\sigma) = \{g \in G : \sigma(g) = e\}.$$

Theorem 7.9 *Let $\sigma : G \to G'$ be a group homomorphism. Then $\ker(\sigma)$ is a subgroup of G.*

PROOF First, $\sigma(e_G) = e_{G'}$ so $e_G \in \ker(\sigma)$.

Second, suppose $h_1, h_2 \in \ker(\sigma)$. That means $\sigma(h_1) = e$ and $\sigma(h_2) = e$. Then $\sigma(h_1 h_2) = \sigma(h_1)\sigma(h_2) = ee = e$ so $h_1 h_2 \in \ker(\sigma)$.

Last, suppose $h \in \ker(\sigma)$. $\sigma(h^{-1}) = (\sigma(h))^{-1} = e^{-1} = e$ so $h^{-1} \in \ker(\sigma)$. Thus $\ker(\sigma) \leq G$. □

Lemma 7.10 *Let $\sigma : G \to G'$ be a group homomorphism. σ is injective if and only if $\ker(\sigma) = \{e\}$.*

PROOF HW

Theorem 7.11 *Let $\sigma : G \to G'$ be a group homomorphism. Define a relation $\equiv_\sigma$ on G by setting $g_1 \equiv_\sigma g_2$ if $\sigma(g_1) = \sigma(g_2)$. This is an equivalence relation, and the equivalence classes are sets*

$$g\ \ker(\sigma) = \{gk : k \in \ker(\sigma)\}.$$

Denote the set of equivalence classes under $\equiv_\sigma$ by $G/\ker(\sigma)$, and define an operation on $G/\ker(\sigma)$ by

$$(g_1\ \ker(\sigma))(g_2\ \ker(\sigma)) = g_1 g_2\ \ker(\sigma).$$

Then $G/\ker(\sigma)$ is a group.

PROOF Essentially, this is the same proof as what was given in the context of rings in the preceding chapter. There is a catch, however, which we will dive more deeply into later when we discuss *normal subgroups* in Section 7.6.

Theorem 7.12 (First Isomorphism Theorem) *Let $\sigma : G \to G'$ be a surjective group homomorphism. Then*

$$G/\ker(\sigma) \cong G'.$$

PROOF As with the First Isomorphism Theorem for rings, you define $\tau : G/\ker(\sigma) \to G'$ by $\tau(g\ \ker(\sigma)) = \sigma(\tau)$. This map is well-defined, is a group homomorphism, and is a bijection, applying arguments that are essentially the same as those given for rings. □

HOW TO PROVE $\sigma : G \to G'$ is a group homomorphism:

Let $g_1, g_2 \in g$.
$\sigma(g_1 g_2) = ... =$ something.
$\sigma(g_1)\sigma(g_2) = ... =$ the same something as above.
Therefore σ is a group homomorphism.

(To show σ is an isomorphism, first prove that it is a homomorphism, then prove σ is surjective and injective.)

Example: Determine, with proof, whether or not the function $\sigma : (\mathbb{Z}, +) \to (7\mathbb{Z}, +)$ defined by $\sigma(x) = 14x$ is a group homomorphism. (Is σ injective? Is σ an isomorphism?)

7.3 Examples of groups

1. Automorphism groups

Let F be a field. Then $Aut(F)$ is a group under the operation of composition of functions.

Why are $Aut(F)$ important examples of groups? Two reasons:

1.

2. Cayley's Theorem, which says that every group is isomorphic to a set of functions under composition:

Theorem 7.13 (Cayley's Theorem) *Every group G is isomorphic to a set G' of bijections from G to itself, where the group operation on G' is composition of functions.*

PROOF Let $\psi_g : G \to G$ be the function $\psi_g(x) = xg$. This is a bijection of G, because its inverse is

$$(\psi_g)^{-1} = \psi_{g^{-1}}.$$

Now consider the set

$$G' = \{\psi_g : g \in G\};$$

this is a group under composition since

$$\psi_{gh} = \psi_g \circ \psi_h.$$

Now consider the function $\sigma : G \to G'$ given by $\sigma(g) = \psi_g$. First,

$$\sigma(gh) = \psi_{gh} = \psi_g\psi_h = \sigma(g)\sigma(h).$$

so σ is a group homomorphism. Now, suppose $g \in \ker(\sigma)$. That means $\sigma(g) = \psi_g$ is the identity function, so $xg = x$ for all $x \in G$. By the left cancellation law, that means $g = e$. This proves $\ker(\sigma) = \{e\}$, so σ is injective. Clearly σ is onto, because for any $\psi_g \in G'$, $\sigma(g) = \psi_g$. Therefore σ is a group isomorphism between G and G'. □

2. Additive groups of rings; cyclic groups

Let R be a ring. Then $(R, +)$ is an abelian group called the **additive group of** R, where the identity element is $e = 0$ and the inverse "x^{-1}" of x is $-x$.

Note: any ring homomorphism (isomorphism) $\sigma : R \to R'$ automatically is a group homomorphism (isomorphism) $\sigma : (R, +) \to (R', +)$.

Specific examples in this context: $(\mathbb{Z}, +)$; $(\mathbb{Q}, +)$; $(\mathbb{R}, +)$; $(\mathbb{C}, +)$; $(\mathbb{Z}/n\mathbb{Z}, +)$.

Some additive groups of rings have extra structure:

Definition 7.14 *A group G is called* **cyclic** *if there is an element $g \in G$ such that every element of G is of the form g^n for some $n \in \mathbb{Z}$. For any element g such that $G = \{g^n : n \in \mathbb{Z}\}$, we say g* **generates** *G, and write $G = <g>$.*

So a cyclic group must look like:

Theorem 7.15 *Every infinite cyclic group is isomorphic to $(\mathbb{Z}, +)$. Every finite cyclic group is isomorphic to $(\mathbb{Z}/n\mathbb{Z}, +)$.*

PROOF Let G be a cyclic group. Then $\exists\, g \in G$ such that $G = \{g^k : k \in \mathbb{Z}\}$. Now define $\sigma : \mathbb{Z} \to G$ by $\sigma(k) = g^k$.

$$\sigma(k + l) = g^{k+l} = g^k g^l = \sigma(k)\sigma(l)$$

so σ is a group homomorphism. Since G is cyclic, σ is surjective. Thus, by the First Isomorphism Theorem, $G \cong \mathbb{Z}/\ker(\sigma)$. If σ is injective, we have $G \cong \mathbb{Z}$. If σ is not injective, then $\ker(\sigma)$ is a subset of $\mathbb{Z}$, containing $\{0\}$, closed under addition. Thus $\ker(\sigma) = n\mathbb{Z}$ for some $n \in \mathbb{Z}$ (HW from Chapter 2), and by the First Isomorphism Theorem $G \cong \mathbb{Z}/n\mathbb{Z}$ as wanted. □

Corollary 7.16 *Every cyclic group is abelian.*

Application: Consider the set $\{1, \zeta_n, \zeta_n^2, ..., \zeta_n^{n-1}\}$ under multiplication. This is a cyclic group of order n, so it is isomorphic to $(\mathbb{Z}/n\mathbb{Z}, +)$.

3. Units of a ring, under multiplication

Let R be a ring. Let $R^\times$ be the set of elements of R that are units. Then $(R^\times, \cdot)$ is a group under the ring multiplication called the **multiplicative group of** R or **group of units of** R, where the identity element is $e = 1$ and the inverse of $x \in R^\times$ is its reciprocal.

Example: $(\mathbb{Z}/6\mathbb{Z})^\times$

4. The trivial group

Let $G = \{e\}$ with binary operation $ee = e$. This forms a group, called the **trivial group**. The order of this group is 1, and this is the only group of order 1.

P.S. the trivial group is the group of units in $\mathbb{Z}/2\mathbb{Z}$.

5. Products of groups

Theorem 7.17 *Let G_1 and G_2 be groups. Then $G_1 \times G_2$ is a group, where the group operation is defined coordinate-wise by*

$$(g_1, g_2)(h_1, h_2) = (g_1 h_1, g_2 h_2).$$

PROOF HW

Example: By the Chinese Remainder Theorem, if $\gcd(m, n) = 1$, then

$$(\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}, +) \cong (\mathbb{Z}/(mn)\mathbb{Z}, +).$$

So $\mathbb{Z}/12\mathbb{Z} \cong \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$, for example.

What if $\gcd(m, n) \neq 1$? For example, is $(\mathbb{Z}/4\mathbb{Z}, +) \cong (\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, +)$?

Definition 7.18 *The* **Klein** 4**-group**, *denoted* V, *is* $(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, +)$. V *is the smallest non-cyclic group.*

Lemma 7.19 *The only groups of order* 4 *(up to isomorphism) are* $(\mathbb{Z}/4\mathbb{Z}, +)$ *and the Klein* 4*-group* V.

PROOF HW (later on)

Definition 7.20 *Let* G *be a group and let* $g \in G$. *If there is a positive* $n \in \mathbb{N}$ *such that* $g^n = e$, *then we say* g **has finite order (in** G**)** *and we call the smallest positive* n *such that* $g^n = e$ *the* **order of** g **(in** G**)**.

The crux of the argument on the previous page is that the order of any element is preserved by isomorphism, so if for some k, one group has a different number of elements of order k than a second group has, the two groups cannot be isomorphic.

Lemma 7.21 *Let* G *be a group. The only element of order* 1 *in* G *is the identity* e.

Corollary 7.22 *Every group of order* 2 *is isomorphic to* $(\mathbb{Z}/2\mathbb{Z}, +)$.

PROOF HW

Example: $Aut(\mathbb{C}) \cong (\mathbb{Z}/2\mathbb{Z}, +)$.

Corollary 7.23 *Let* p *be prime. Every group of order* p *is isomorphic to* $(\mathbb{Z}/p\mathbb{Z}, +)$.

PROOF HW (later on)

Example: Find the order of every element in $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$.

ELEMENT	ORDER	ELEMENT	ORDER
$e = (0, 0)$		$(1, 0)$	
$(0 + 2\mathbb{Z}, 1 + 4\mathbb{Z}) = (0, 1)$		$(1, 1)$	
$(0, 2)$		$(1, 2)$	
$(0, 3)$		$(1, 3)$	

6. Dihedral groups

Imagine a shape lying flat in a box of the same shape. You can take the shape out of the box, and if the shape has some symmetry, you can flip over and/or rotate the shape before putting it back in the box.

The transformations of an object are called its *symmetries* the symmetries of an object form a group. More formally,

Definition 7.24 *Let S be a geometric figure. A* **symmetry** *of S is a function $f : S \to S$ which preserves the distance between any two points in S.*

Of particular importance are the symmetries of regular polygons:

Definition 7.25 *Let $n \geq 3$. The* **(n^{th}) dihedral group** *D_n is the set of symmetries of a regular n-gon, i.e. the set of linear transformations of a plane which send a regular polygon to itself.*

Example: Give notation for the elements of D_3 and construct a composition table for D_3.

Exercise: Give notation for the elements of D_4 (*Hint:* "r" and "f" may be useful), and construct a composition table for D_4. Then try to find some nontrivial subgroups of D_4.

Theorem 7.26 (Properties of dihedral groups) *Let D_n be the n^{th} dihedral group. Then:*

- $|D_n| = 2n$.
- $D_n = \{e, r, r^2, ..., r^{n-1}, f, rf, r^2f, ...r^{n-1}f\}$ *where $r \in D_n$ has order n and $f \in D_n$ has order 2 (if the regular polygon is centered at the origin and has a vertex on the x-axis, think of r as a counterclockwise rotation by $\frac{2\pi}{n}$ and think of f as a reflection across the x-axis).*
- *If $n \geq 3$, then D_n is not abelian (in particular $fr = r^{n-1}f \neq rf$).*
- *D_n has a cyclic subgroup $\{e, r, r^2, ..., r^{n-1}\}$ of order n.*
- *D_n has a cyclic subgroup $\{e, f\}$ of order 2.*

Application: Recall that $\zeta_3 = e^{2\pi i/3}$ and let $F = \mathbb{Q}(\sqrt[3]{2}, \zeta_3)$. Let's study F and $Aut(F)$:

- $\sqrt[3]{2}$ has degree 3 over $\mathbb{Q}$ because it is a root of the irreducible polynomial $x^3 - 2$. So
$$\{1, \sqrt[3]{2}, \left(\sqrt[3]{2}\right)^2\}$$
is a basis for $\mathbb{Q}(\sqrt[3]{2})$ over $\mathbb{Q}$.

- ζ_3 has degree 2 over $\mathbb{Q}$ because it is a root of the irreducible polynomial $\Phi_3(x) = x^2 + x + 1$. This polynomial is also irreducible over $\mathbb{Q}(\sqrt[3]{2})$, because it has no real roots and $\mathbb{Q}(\sqrt[3]{2}) \subseteq \mathbb{R}$. Therefore ζ_3 has degree 2 over $\mathbb{Q}(\sqrt[3]{2})$ and
$$\{1, \zeta_3\}$$
is a basis for $\mathbb{Q}(\sqrt[3]{2})(\zeta_3) = F$ over $\mathbb{Q}(\sqrt[3]{2})$.

- By the Dedekind Product Theorem, $\dim(F/\mathbb{Q}) = 3 \cdot 2 = 6$ and
$$\{1, \sqrt[3]{2}, \left(\sqrt[3]{2}\right)^2, \zeta_3, \sqrt[3]{2}\zeta_3, \left(\sqrt[3]{2}\right)^2 \zeta_3\}$$
is a basis for F over $\mathbb{Q}$.

- any automorphism of F must be the identity on $\mathbb{Q}$, since $Aut(\mathbb{Q}) = \{I_{\mathbb{Q}}\}$. Therefore any element of $Aut(F)$ is determined by its values at $\sqrt[3]{2}$ and ζ_3.

  What can those values be? First, if $\sigma \in Aut(F)$, what can $\sigma(\zeta_3)$ be?

Second, if $\sigma \in Aut(F)$, what can $\sigma(\sqrt[3]{2})$ be?

To summarize, these are the six possibilities:

NAME OF σ	$\sigma(\zeta_3)$	$\sigma(\sqrt[3]{2})$	$\sigma(\sqrt[3]{2})$	$\sigma(\sqrt[3]{2}\zeta_3)$	$\sigma(\sqrt[3]{2}\zeta_3^2)$
e	ζ_3	$\sqrt[3]{2}$	$\sqrt[3]{2}$	$\sqrt[3]{2}\zeta_3$	$\sqrt[3]{2}\zeta_3^2$
	ζ_3	$\sqrt[3]{2}\zeta_3$	$\sqrt[3]{2}\zeta_3$	$\sqrt[3]{2}\zeta_3^2$	$\sqrt[3]{2}$
	ζ_3	$\sqrt[3]{2}\zeta_3^2$	$\sqrt[3]{2}\zeta_3^2$	$\sqrt[3]{2}$	$\sqrt[3]{2}\zeta_3$
	ζ_3^2	$\sqrt[3]{2}$	$\sqrt[3]{2}$	$\sqrt[3]{2}\zeta_3^2$	$\sqrt[3]{2}\zeta_3$
	ζ_3^2	$\sqrt[3]{2}\zeta_3$	$\sqrt[3]{2}\zeta_3$	$\sqrt[3]{2}$	$\sqrt[3]{2}\zeta_3^2$
	ζ_3^2	$\sqrt[3]{2}\zeta_3^2$	$\sqrt[3]{2}\zeta_3^2$	$\sqrt[3]{2}\zeta_3$	$\sqrt[3]{2}$

Notice that each of these automorphisms permutes the set $\{\sqrt[3]{2}, \sqrt[3]{2}\zeta_3, \sqrt[3]{2}\zeta_3^2\}$. Let's graph this set in a complex plane:

This example suggests how we study $Aut(F)$ in general. We try to find some "common" group which is isomorphic to $Aut(F)$ that we know a lot about.

7.4 Permutation groups

Now we turn to the last class of groups. This class is important because by Cayley's Theorem, every finite group is isomorphic to a subgroup of one of these groups.

Definition 7.27 *Let N be a positive integer. The* **symmetric group on N letters**, *denoted $\mathcal{S}_N$ or $Sym(N)$ or $S(N)$, is the set of all bijections of the set $\{1, 2, ..., N\}$. (This set forms a group under composition of functions.) Elements of $\mathcal{S}_N$ are called* **permutations (of $\{1, ..., N\}$)**.

Exercise: List all the elements of $\mathcal{S}_2$:

Exercise: List all the elements of $\mathcal{S}_3$:

Exercise: Construct a composition table for $\mathcal{S}_3$:

Question: What is the order of S_N? We've seen $|S_2| = 2$ and $|S_3| = 6$.

Inconvenient notation for permutations that is sometimes used: Suppose $\sigma \in \mathcal{S}_5$ is $\sigma(1) = 2$; $\sigma(2) = 4$; $\sigma(3) = 3$; $\sigma(4) = 1$. Then we can write:

We'd like more efficient notation for permutations, and we turn to that issue next.

Cycles

Definition 7.28 *Let $\sigma \in \mathcal{S}_N$ be a permutation. If there is a subset $\{a_1, a_2, ..., a_k\}$ of $\{1, ..., N\}$ such that*

$$\sigma(x) = \begin{cases} a_{k+1} & \text{if } x \in \{a_1, a_2, ..., a_{k-1}\} \\ a_1 & \text{if } x = a_k \\ x & \text{if } x \notin \{a_1, ..., a_k\} \end{cases},$$

then σ is called a **cycle (of length k)** *(or a k-**cycle**) and we write*

$$\sigma = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_k \end{pmatrix}.$$

Two cycles $\sigma, \tau \in \mathcal{S}_N$ are called **disjoint** *if their corresponding sets $\{a_1, ..., a_k\}$ in this definition are disjoint.*

A 2-cycle $\begin{pmatrix} a_1 & a_2 \end{pmatrix}$ is also called a **transposition**.

Note: The cycles $(3\ 5\ 8)$ and $(5\ 8\ 3)$ are the same, because they both represent a permutation that does this:

But neither $(3\ 5\ 8)$ nor $(5\ 8\ 3)$ are equal to $(3\ 8\ 5)$:

Example: Let $\sigma \in \mathcal{S}_6$ be $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 6 & 1 & 5 & 4 \end{pmatrix}$. Write σ in cycle notation.

Example: Let $\tau \in \mathcal{S}_6$ be $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 5 & 4 & 3 & 6 \end{pmatrix}$. Write τ in cycle notation.

Theorem 7.29 *Let $\sigma \in \mathcal{S}_N$. Then σ is a product of disjoint cycles.*

PROOF Set $b_1 = 1$ and let $k_1 = \min\{n \geq 1 : \sigma^n(b_1) = b_1\}$. Then define

$$O_1 = \{b_1, \sigma(b_1), \sigma^2(b_1), ..., \sigma^{k_1-1}(b_1)\}$$

and let $\sigma_1 = \left(b_1\ \sigma(b_1)\ ...\ \sigma^{k_1-1}(b_1)\right)$.

If $O_1 = \{1, ..., N\}$, we are done, since $\sigma = \sigma_1$, a cycle. Otherwise, let $b_2 = \min\{n \in \{1, ..., N\} : n \notin O_1\}$ and notice $b_2 \geq 2$. Now define $k_2 = \min\{n \geq 1 : \sigma^n(b_2) = b_2\}$ and set

$$O_2 = \{b_2, \sigma(b_2), \sigma^2(b_2), ..., \sigma^{k_2-1}(b_2)\}$$

and

$$\sigma_2 = \left(b_2\ \sigma(b_2)\ ...\ \sigma^{k_2-1}(b_2)\right).$$

Notice $O_1 \cap O_2 = \emptyset$, because if not, $\sigma^j(b_1) = b_2$ for some j, meaning $b_2 \in O_1$, a contradiction. Therefore σ_1 and σ_2 are disjoint cycles.

At this point, if $O_1 \cup O_2 = \{1, ..., N\}$, we are done, since $\sigma = \sigma_1\sigma_2$, a product of two cycles. Otherwise, let $b_3 = \min\{n \in \{1, ..., N\} : n \notin O_1 \cup O_2\}$; notice $b_3 \geq 3$. Define $k_3 = \min\{n \geq 1 : \sigma^n(b_3) = b_3\}$ and set

$$O_3 = \{b_3, \sigma(b_3), \sigma^2(b_3), ..., \sigma^{k_2-1}(b_3)\}$$

and

$$\sigma_3 = \left(b_3\ \sigma(b_3)\ ...\ \sigma^{k_3-1}(b_3)\right).$$

As before, σ_3 is disjoint from both σ_1 and σ_2. Continuing in this fashion, eventually $O_1 \cup O_2 \cup \cdots \cup O_m = \{1, ..., N\}$, and we see that

$$\sigma = \sigma_1\sigma_2 \cdots \sigma_m$$

as wanted. □

Remark: Although $\mathcal{S}_N$ is not abelian, disjoint cycles always commute, so the order the cycles are written in the previous theorem doesn't matter, i.e.

$$(1\ 2\ 4)(3\ 7) = (3\ 7)(1\ 2\ 4).$$

Example: Let $\sigma \in \mathcal{S}_6$ be $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 6 & 5 & 2 & 3 \end{pmatrix}$. Write σ in cycle notation.

Henceforth, all permutations will be written in cycle notation.

Example computations with cycle notation

Example: Compute and simplify the following:

$(1\ 2\ 3)(1\ 2)$

$(1\ 2\ 5)(3\ 5)(1\ 2)$

the inverse of $(1\ 4\ 2\ 3)(5\ 8)$:

Example: Find the order of $(1\ 2\ 3)(4\ 5)(6\ 7\ 8\ 9)$.

Example: Suppose $\sigma = (1\ 3\ 4)(2\ 6)$ and $\tau = (1\ 5\ 2)(3\ 6)$. Compute $\sigma\tau$.

Example: List all the elements of $\mathcal{S}_3$ in cycle notation:

(Having done these examples, it is a good idea to return to the first page of this section and rewrite things in terms of cycle notation.)

Parity and alternating groups

Corollary 7.30 *Let $\sigma \in \mathcal{S}_N$. Then σ is a product of transpositions.*

PROOF In light of Theorem 7.29, it is sufficient to show that any cycle is a product of transpositions. But by direct calculation,

$$(a_1\ a_2\ \cdots\ a_{k-1}\ a_k) = (a_1\ a_k)(a_1\ a_{k-1})\cdots(a_1\ a_4)(a_1\ a_3)(a_1\ a_2).$$

So we are done. □

Note: This corollary shows explicitly that a $k-$cycle can be written as a product of $k-1$ transpositions.

Note: There is nothing unique about the transpositions in this corollary (there are lots of ways to write a permutation as a product of transpositions). However, there is more to this story. It turns out that permutations, like integers, come in two flavors: "odd" and "even". Here's how we distinguish them:

Definition 7.31 *The* **sign**, *or* **signature** *function $sgn : \mathcal{S}_N \to \{-1, 1\}$ is the function*

$$sgn(\sigma) = \frac{P(x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(N)})}{P(x_1, x_2, ..., x_N)}$$

where

$$P(x_1, ..., x_N) = \prod_{i<j}(x_i - x_j).$$

Examples: If $N = 3$, then

$$P(x_1, x_2, x_3) = (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$$

but

$$P(x_3, x_1, x_2) =$$

Furthermore, if $\sigma = (1\ 2)$, then

$$\text{sgn}(\sigma) =$$

and if $\sigma = (1\ 3\ 2)$, then

$$\text{sgn}(\sigma) = \frac{P(x_3, x_1, x_2)}{P(x_1, x_2, x_3)} = \frac{(x_3 - x_1)(x_3 - x_2)(x_1 - x_2)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} =$$

In general, P is the product of $(N-1)+(N-2)+...+3+2+1 = \frac{1}{2}N(N-1)$ differences.

Lemma 7.32 *If $\tau \in \mathcal{S}_N$ is a transposition, then $sgn(\tau) = -1$.*

PROOF Suppose $\tau = (a\ b)$ where $a < b$. Then

$$\begin{aligned}\text{sgn}(\tau) &= \frac{P(x_{\tau(1)}, ..., x_{\tau(N)})}{P(x_1, ..., x_N)} \\ &= \frac{P(x_1, x_2, ..., x_b, x_{a+1}, ..., x_{b-1}, x_a, ..., x_N)}{P(x_1, x_2, ..., x_a, x_{a+1}, ..., x_{b-1}, x_b, ..., x_N)}.\end{aligned}$$

In the above expression, all the differences $(x_i - x_j)$ where $i < a, i < b$, $j > a$ and $j > b$ cancel, and all the differences $(x_i - x_j)$ where neither i nor j are a or b cancel, leaving

$$\text{sgn}(\tau) = \frac{[(x_b - x_{a+1})(x_b - x_{a+2}) \cdots (x_b - x_a)]\,[(x_{a+1} - x_a)(x_{a+2} - x_a) \cdots (x_{b-1} - x_a)]}{[(x_a - x_{a+1})(x_a - x_{a+2}) \cdots (x_a - x_b)]\,[(x_{a+1} - x_b)(x_{a+2} - x_b) \cdots (x_{b-1} - x_b)]}$$

$$= (-1)^{[b-a+1]+[b-a]} = (-1)^{2(b-a)+1} = -1. \square$$

Theorem 7.33 *$sgn : \mathcal{S}_N \to \{-1, 1\}$ is a group homomorphism (where the group operation on $\{-1, 1\}$ is multiplication).*

PROOF Let $\sigma, \tau \in \mathcal{S}_N$. Then

$$\begin{aligned}\text{sgn}(\sigma\tau) &= \frac{P(x_{\sigma\tau(1)}, ..., x_{\sigma\tau(N)})}{P(x_1, ..., x_N)} \\ &= \frac{P(x_{\sigma\tau(1)}, ..., x_{\sigma\tau(N)})}{P(x_{\tau(1)}, ..., x_{\tau(N)})} \cdot \frac{P(x_{\tau(1)}, ..., x_{\tau(N)})}{P(x_1, ..., x_N)} \\ &= \text{sgn}(\sigma)\ \text{sgn}(\tau). \square\end{aligned}$$

Corollary 7.34 *If σ can be written as the product of m_1 transpositions and also written as the product of m_2 transpositions, then either (both m_1 and m_2 are even), or (both m_1 and m_2 are odd).*

PROOF By previous results, $\text{sgn}(\sigma) = (-1)^{m_1} = (-1)^{m_2}$. The result follows. $\square$

Definition 7.35 *Let $\sigma \in \mathcal{S}_N$. σ is called* **even** *if it can be written as the product of an even number of transpositions; σ is called* **odd** *if it can be written as the product of an odd number of transpositions. The evenness/oddness of a permutation is called its* **parity**.

Equivalently, σ is even iff $\text{sgn}(\sigma) = 1$ and σ is odd iff $\text{sgn}(\sigma) = -1$.

Since sgn is a homomorphism, the product of two permutations of the same parity is even, and the product of two permutations of opposite parity is odd.

Remember from Corollary that a k-cycle can be written as a product of $k-1$ transpositions. So 3-cycles, 5-cycles, etc. are even, and 2-cycles, 4-cycles, etc. are odd.

Example: Is $(3\ 5\ 8)(2\ 7)(4\ 6\ 11)$ even or odd?

Definition 7.36 *The set of even permutations in $\mathcal{S}_N$ is called the* **alternating group (on N letters)** *and is denoted $\mathcal{A}_N$.*

Since $\mathcal{A}_N = \ker(\text{sgn})$, it is clear that $\mathcal{A}_N$ is a subgroup of $\mathcal{S}_N$.

Lemma 7.37 $|\mathcal{A}_N| = \frac{1}{2}n!$.

PROOF Note that $\sigma \in \mathcal{A}_N$ if and only if $\sigma(1\ 2) \in \mathcal{S}_N - \mathcal{A}_N$. Thus the function $f : \mathcal{A}_N \to \mathcal{S}_N - \mathcal{A}_N$ defined by $f(\sigma) = \sigma(1\ 2)$ is a bijection (it is bijective because it is its own inverse), meaning

$$|\mathcal{A}_N| = \#(\mathcal{S}_N - \mathcal{A}_N) = |S_N| - |\mathcal{A}_N| = N! - |\mathcal{A}_N|.$$

Solving for $|\mathcal{A}_N|$ gives the result. □

Elements of $\mathcal{S}_N$ can be studied according to their cycle structure, and we can easily discern members of $\mathcal{A}_N$ by doing this. As an example, let's study all the members of $\mathcal{S}_3$ and $\mathcal{S}_4$:

DISJOINT CYCLE STRUCTURE	PERMUTATIONS IN $\mathcal{S}_3$ WITH THAT STRUCTURE	ORDER	NUMBER OF THESE IN $\mathcal{S}_3$	EVEN OR ODD?

DISJOINT CYCLE STRUCTURE	ORDER	NUMBER OF THESE IN $\mathcal{S}_4$	EVEN OR ODD?

ENRICHMENT: You may remember from linear algebra the concept of the *determinant* of an $N \times N$ square matrix. For example, the determinant of a 3×3 square matrix is

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + cdh + bfg - bdi - ceg - afh.$$

Notice that this formula for the determinant contains six terms, of which three are added and three are subtracted. Each of the terms being added or subtracted is the product of three numbers, such that the three numbers consist of one number from each row of the matrix and one number from each column of the matrix. Therefore, for each of the numbers being added or subtracted, there is a permutation $\sigma \in \mathcal{S}_3$ so that the term is

$$a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)}$$

where, as usual, a_{ij} represents the element of the matrix in the i^{th} row and j^{th} column. For example, the term bfg in the determinant comes from $\sigma = (1\ 2\ 3)$ since for this σ,

$$a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)} = a_{12}a_{23}a_{31} = bfg.$$

If you look carefully, you will see that the terms being added in the determinant (like bfg) come from the even permutations in $\mathcal{S}_3$, and the terms being subtracted come from the odd permutations in $\mathcal{S}_3$. So we can write the formula for the determinant of a 3×3 matrix as follows:

$$\det A = \sum_{\sigma \in \mathcal{S}_3} \text{sgn}(\sigma) a_{1\sigma(1)}a_{2\sigma(2)}a_{3\sigma(3)}.$$

This formula generalizes: for any $N \times N$ matrix A, the formal definition of the determinant of A is

$$\det A = \sum_{\sigma \in \mathcal{S}_N} \left(\text{sgn}(\sigma) \prod_{j=1}^{N} a_{j\sigma(j)} \right).$$

From this formula, you can prove all the properties of determinants you know and love. For example, if you swap two rows of a matrix, all the signatures of the corresponding σs in the determinant formula are multiplied by -1 (since the σ gets multiplied by a transposition coming from swapping the rows), meaning the determinant gets multiplied by -1 as well.

Symmetries of three-dimensional figures

Exercise: Describe the group of symmetries of a regular tetrahedron:

7.5 Subgroups and cosets

Examples of subgroups

1. Kernel of any homomorphism

Special case: for any N, $\mathcal{A}_N < \mathcal{S}_N$, since $\mathcal{A}_N = \ker(\text{ sgn})$.

2. Cyclic subgroups

Definition 7.38 *Let G be a group and let $g \in G$. The* **subgroup generated by** g, *denoted <g>, is the subgroup*

$$<g> = \{g^k : k \in \mathbb{Z}\}.$$

For any $g \in G$, <g> is a cyclic group whose order is equal to the order of element g. Thus we have proven that every group contains a cyclic subgroup.

Example: $G = \mathcal{S}_4$; $g = (1\ 2\ 3)$.

Example: $G = (\mathbb{Z}, +)$; $g = 7$

Cosets

Every subgroup H of G generates two equivalence relations on G. The equivalence classes under these relations are called *cosets*.

Definition 7.39 *Let G be a group and let $H \leq G$. For every $g \in G$, define:*

- *the* **left coset** *of H containing g, denoted gH, by $gH = \{gh : h \in H\}$, and*
- *the* **right coset** *of H containing g, denoted Hg, by $Hg = \{hg : h \in H\}$.*

Example: Let $G = \mathcal{S}_3$ and let $H = <(1\,2)> = \{e, (1\,2)\}$. Then:

ELEMENT OF G	LEFT COSET gH	RIGHT COSET Hg
e		
$(1\,2)$		
$(1\,3)$		
$(2\,3)$		
$(1\,2\,3)$	$\{(1\,2\,3), (1\,3)\}$	$\{(1\,2\,3), (2\,3)\}$
$(1\,3\,2)$	$\{(1\,3\,2), (2\,3)\}$	$\{(1\,3\,2), (1\,3)\}$

Observe:

1. Two left cosets are either the same, or disjoint.
2. Two right cosets are either the same, or disjoint.
3. The left cosets and the right cosets are not the same subsets of $\mathcal{S}_3$.

Theorem 7.40 *Let G be a group and let $H \leq G$. Two left (right) cosets of H in G are either equal or disjoint. Thus they form a partition of G.*

PROOF Suppose $x \in g_1H \cap g_2H$. Then $x = g_1h_1$ and $x = g_2h_2$ for $h_1, h_2 \in H$. Thus

$$g_1h_1 = g_2h_2 \Rightarrow g_1 = g_2h_2h_1^{-1}$$

and for any $h' \in H$,

$$g_1h' = g_2h_2h_1^{-1}h'$$

and since H is a subgroup, $h_2h_1^{-1}h' \in H$. We have proved $g_1H \subseteq g_2H$, and by symmetric argument, $g_2H \subseteq g_1H$. Thus any two left cosets which intersect in at least one point are equal, proving the theorem (the proof for right cosets is similar). □

Consequence: Since the left (right) cosets of H in G partition G, the relation of "being in the same left (right) coset of H" is an equivalence relation on G.

Theorem 7.41 (Lagrange's Theorem) *Let G be a finite group and let $H \leq G$. Then $|H|$ divides $|G|$.*

PROOF Let $g \in G$. The function $f : H \to gH$ defined by $f(x) = gx$ is a bijection (because it has inverse $f^{-1}(x) = g^{-1}x$), so the cardinalities of H and gH are the same.

That means every left coset of H in G has the same cardinality as H. Let the number of cosets be denoted by k; then

$$k\,|H| = |G|,$$

proving the theorem. □

Application: If $|G| = 30$, then G has no subgroup of order 7.

Corollary 7.42 *Let G be a finite group and let $g \in G$. Then the order of g divides $|G|$.*

PROOF The order of cyclic subgroup $<g> \leq G$ is the order of the element g. The result follows from the preceding theorem. □

Application: S_4 has no element of order 5 (since $5 \not| 24 = |S_4|$).

Note: The function $x \mapsto xg$ is also a bijection between H and Hg for any $g \in G$, so every right coset of H in G also has the same cardinality as H.

Definition 7.43 *If the number of left cosets of H in G is finite, we say H* **has finite index** *in G and we denote the number of left cosets by $[G : H]$. This number is called the* **index of H in G.**

Note: By Lagrange's Theorem, if G is finite, this means $[G : H] = \frac{|G|}{|H|}$.

Note: $[G : H]$ is also the number of right cosets, by the following lemma:

Lemma 7.44 *The set of left cosets of H in G has the same cardinality as the set of right cosets in G.*

PROOF The function $gH \mapsto Hg$ is well-defined and gives a bijection between the set of left cosets and the set of right cosets. □

Theorem 7.45 (Index Product Theorem) *Suppose $K \leq H \leq G$ are groups with G finite, then $[G : K] = [G : H][H : K]$.*

PROOF This follows from Lagrange's Theorem directly:

$$[G : K] = \frac{|G|}{|K|} = \frac{|G|}{|H|} \cdot \frac{|H|}{|K|} = [G : H][H : K]. \square$$

Fermat's Little Theorem and Euler's Theorem

Fermat's most famous "theorem" is his "last" theorem, which says:

If $n \geq 3$ is an integer, the equation $a^n + b^n = c^n$ has no integer solutions $(a, b, c) \in \mathbb{Z}^3$ other than when a and/or b are zero.

Fermat wrote this down and scribbled "I have discovered a truly marvelous proof of this!" next to it, and then died before he wrote more on it. In 1994 Andrew Wiles gave the first published proof of this theorem, so us math nerds now call the above fact Wiles' Theorem rather than Fermat's Last Theorem.

Fermat's most useful theorem is not his last theorem, but his "little" theorem:

Theorem 7.46 (Fermat's Little Theorem) *If p is prime and $\gcd(a, p) = 1$, then*

$$a^{p-1} \equiv 1 \mod p.$$

PROOF Consider the group of units in $\mathbb{Z}/p\mathbb{Z}$. This group has order $p-1$. Therefore, the order of any $a + p\mathbb{Z}$ divides $p - 1$, so $(a + p\mathbb{Z})^{p-1} = 1 + p\mathbb{Z}$ as wanted. □.

Application: What is the remainder when 2^{290} is divided by 73?

Remark: Fermat's Little Theorem implies $a^p \equiv a \mod p$ whenever $\gcd(a, p) = 1$. In fact, if $\gcd(a, p) = p$, then $a^p \equiv 0^p \equiv 0 \equiv a \mod p$ so Fermat's Little Theorem can also be stated this way:

Corollary 7.47 (Fermat's Little Theorem (restated)) *If p is prime, then for any a,*

$$a^p \equiv a \mod p.$$

Fermat's Little Theorem generalizes as follows:

Theorem 7.48 (Euler's Theorem) *If $\gcd(a, n) = 1$, then*

$$a^{\phi(n)} \equiv 1 \mod p$$

where ϕ is the Euler phi function.

PROOF Consider the group of units in $\mathbb{Z}/n\mathbb{Z}$. This group has order $\phi(n)$. Therefore, the order of any unit $a + n\mathbb{Z}$ divides $\phi(n)$, so $(a + n\mathbb{Z})^{\phi(n)} = 1 + n\mathbb{Z}$ as wanted. □.

7.6 Normal subgroups and quotient groups

Definition 7.49 *Let G be a group and let $H \leq G$. H is called a* **normal subgroup of** *G if for every $g \in G$ and $h \in H$, $ghg^{-1} \in H$. We denote a normal subgroup H of G by writing $H \triangleleft G$.*

HOW TO PROVE H is a normal subgroup of G:

1. Prove H is a subgroup of G.
2. Let $g \in G, h \in H$. ... (logical argument) ... Therefore $ghg^{-1} \in H$.
   Therefore $H \triangleleft G$. □

Lemma 7.50 *If G is an abelian group, then every subgroup of G is normal.*

PROOF Let $H \leq G$ and suppose $g \in G, h \in H$. Then

$$ghg^{-1} = gg^{-1}h = eh = h \in H$$

so H is normal by definition. □

Example 1: Every subgroup $n\mathbb{Z}$ of $\mathbb{Z}$ is normal, since $(\mathbb{Z}, +)$ is abelian.

Example 2: Let $G = \mathcal{S}_3$ and let $H_1 = <(1\,2)> = \{e, (1\,2)\}$.

Example 3: Let $G = \mathcal{S}_3$ and $H_2 = \mathcal{A}_3 = <(1\,2\,3)> = \{e, (1\,2\,3), (1\,3\,2)\}$.

The "real" reason $\mathcal{A}_3 \triangleleft \mathcal{S}_3$ is that $\mathcal{A}_3$ is a kernel (of the sgn homomorphism):

Theorem 7.51 *Let G and G' be groups and let $\sigma : G \to G'$ be a group homomorphism. Then $\ker(\sigma)$ is a normal subgroup of G.*

PROOF We have already proven that $\ker(\sigma) \leq G$. Let $h \in \ker(\sigma)$ and $g \in G$. Since $h \in \ker(\sigma)$, $\sigma(h) = e$.

Therefore $ghg^{-1} \in \ker(\sigma)$, so $\ker(\sigma) \triangleleft G$ as wanted. □

In fact, a converse of this theorem is also true: every normal subgroup of G is the kernel of some homomorphism $G \to G'$, where G' is some other group.

ALTERNATE METHOD TO PROVE H is a normal subgroup of G:

1. Write down a function $\sigma : G \to G'$ where G' is some group.
2. Prove σ is a group homomorphism.
3. Prove $H = \ker(\sigma)$.

   Therefore $H \triangleleft G$. □

Theorem 7.52 *Let G be a group and let $H \leq G$. TFAE:*

1. *$H \triangleleft G$.*
2. *For all $g \in G$, $gH = Hg$.*
3. *For all $g \in G$, $gHg^{-1} = H$.*
4. *For all $g \in G$, $gHg^{-1} \subseteq H$.*
5. *Every left coset of H in G is also a right coset.*
6. *Every right coset of H in G is also a left coset.*

PROOF (1 ⇒ 4): obvious by definition.

(4 ⇒ 3): $gHg^{-1} \subseteq H$ is given, so we need to show $H \subseteq gHg^{-1}$. Let $h \in H$. Then $g^{-1}h(g^{-1})^{-1} = h' \in H$ by (4), so $h = gh'g^{-1} \in gHg^{-1}$ as wanted.

(3 ⇒ 2): ($\subseteq$) Let $x \in gH$; then $x = gh$ for $h \in G$. Then $xg^{-1} = ghg^{-1} \in H$ by (3) so $xg^{-1} = h' \in H$. Thus $x = h'g$, so $x \in Hg$ as wanted.

($\supseteq$) Let Let $x \in Hg$; then $x = hg$ for $h \in G$. Then $g^{-1}x = g^{-1}hg^{-1} = g^{-1}h(g^{-1})^{-1} \in H$ by (3) so $g^{-1}x = h' \in H$. Thus $x = gh'$, so $x \in gH$ as wanted.

(2 ⇔ 5,6) is obvious (just a restatement of (2) in English).

(2 ⇒ 1): Let $h \in H$ and $g \in G$. Then $gh \in gH = Hg$ so there is $h' \in H$ such that $gh = h'g$, i.e. $ghg^{-1} = h' \in H$. □

As a consequence, if $H \triangleleft G$, we don't have to worry about "left" or "right" cosets, because they are the same thing. We can then denote the set of cosets by G/H and define a binary operation on G/H coming from the binary operation on G:

Definition 7.53 *Let G be a group and let $H \triangleleft G$. The set of cosets of H in G is denoted G/H (read "$G \mod H$"). We define an operation on G/H by setting, for any $g_1, g_2 \in G$,*

$$(g_1H)(g_2H) = (g_1g_2)H.$$

Theorem 7.54 *Let G be a group and let $H \triangleleft G$. Then, the set of cosets G/H forms a group called either the* **quotient group (of G by H)** *or "$G \mod H$".*

PROOF We need to show G/H is a group:

1. First, by definition, G/H is closed under the operation if it is well-defined, so we check that it is in fact well-defined. Suppose $g_1H = g_1'H$ and $g_2H = g_2'H$. Since H is normal, we also know by the previous theorem that $g_2'H = Hg_2'$.

$$\begin{aligned}(g_1H)(g_2H) = (g_1g_2)H &= g_1g_2H \\ &= g_1g_2'H \\ &= g_1Hg_2' \\ &= g_1'Hg_2' \\ &= g_1'g_2'H = (g_1'H)(g_2'H),\end{aligned}$$

   so the operation is well-defined.

2. We need to show that the operation is associative. Let $g_1, g_2, g_3 \in H$.

$$\begin{aligned}(g_1H)\left[(g_2H)(g_3H)\right] &= (g_1H)(g_2g_3H) \\ &= (g_1(g_2g_3))H \\ &= ((g_1g_2)g_3)H \quad \text{(by the associativity of } G\text{)} \\ &= (g_1g_2H)(g_3H) \\ &= \left[(g_1H)(g_2H)\right](g_3H)\end{aligned}$$

   as wanted.

3. We need to show $eH = H$ is the identity for G/H. To see this, let $g \in G$ and note

$$(gH)(eH) = (ge)H = gH \quad \text{and} \quad (eH)(gh) = (eg)H = gH$$

   as wanted.

4. Last, we need to check that for any $g \in G$, $g^{-1}H$ is an inverse element of gH. To see this, note

   $(g^{-1}H)(gH) = (g^{-1}g)H = eH = H \quad$ and $\quad (gH)(g^{-1}H) = (gg^{-1})H = eH = H$

   as wanted. $\square$

Example: $G = (\mathbb{Z}, +)$, $H = 5\mathbb{Z}$.

Example: $G = \mathcal{S}_3$, $H = \mathcal{A}_3$

Remarks: Restating previous theorems gives:

- If $H \triangleleft G$ has finite index in G, then $|G/H| = [G : H]$.
- If G is finite, then for any $H \triangleleft G$, $|G/H| = |G|/|H|$.
- If $K \triangleleft H \triangleleft G$, then $|G/K| = |G/H|\,|H/K|$.

Simple groups

Every group G has two normal subgroups: $\{e\}$ and G itself. Some groups have no other normal subgroups:

Definition 7.55 *If G is a group with no normal subgroups other than $\{e\}$ and G, then G is called* **simple**.

Being simple sounds good, but it is actually bad. A non-simple group can be studied by looking at its normal subgroups and examining how those normal subgroups "fit together" to form the whole group. As an example, the dihedral group D_n is not simple:

But you can think about a dihedral group by considering the "rotation part" (the normal subgroup) and the "flip" (what's not in the normal subgroup) and thinking about how the rotation and the flip interact.

If a group is simple, there's no analogous reasoning available. You have to look at the whole bleeping group altogether.

Example: $(\mathbb{Z}/p\mathbb{Z}, +)$ is simple if p is prime:

Example: if $N \geq 3$, the symmetric group $\mathcal{S}_N$ is not simple:

Here is the most important example of a simple group:

Theorem 7.56 *$\mathcal{A}_5$ is simple.*

PROOF Let $H \triangleleft \mathcal{A}_5$ and suppose $H \neq \{e\}$. We prove this theorem in three steps:

Step 1: Prove that ($H \triangleleft \mathcal{A}_5$ and $H \neq \{e\}$) implies that H contains a 3-cycle.
Step 2: Prove that ($H \triangleleft \mathcal{A}_N$, $N \geq 5$ and H contains a 3-cycle) implies that H contains all 3-cycles in $\mathcal{S}_N$.
Step 3: Prove that every element in $\mathcal{A}_N$ (for $N \geq 5$) can be written as a composition of 3-cycles.

Proof of Step 1: Since $H \neq \{e\}$, H must contain an even permutation $\sigma \neq e$. We will show by cases, based on the possible cycle structure of σ, that no matter what σ is, H must contain a 3-cycle:

Case 1: $\sigma = (a\,b\,c)$, a 3-cycle. In this situation, $H = \mathcal{A}_5$ by the previous lemma.

Case 2: $\sigma = (a\,b)(c\,d)$ where a, b, c and d are different. Here, since H is normal,

$$(a\,b\,e)\sigma(a\,b\,e)^{-1} = (b\,e)(c\,d) \in H$$

and since H is a subgroup, H is closed under composition so

$$[(a\,b)(c\,d)]\,[(b\,e)(c\,d)] = (a\,b\,e) \in H.$$

Case 3: $\sigma = (a\,b\,c\,d\,e)$, a 5-cycle. Again, since H is normal,

$$[(a\,b)(c\,d)]\,(a\,b\,c\,d\,e)\,[(a\,b)(c\,d)]^{-1} = (a\,d\,c\,e\,b)$$

and since H is a subgroup, H is closed under composition so

$$(a\,b\,c\,d\,e)(a\,d\,c\,e\,b) = (a\,e\,c) \in H.$$

In any case, H contains a 3-cycle. This completes Step 1.

Proof of Step 2: Consider the 3-cycle $(1\,2\,3)$. Given $(x\,y\,z) \in H$, let $\sigma = (3\,z)(2\,y)(1\,x)$. There are two cases:

Case 1: $\sigma \in \mathcal{A}_N$. In this case, consider

$$\sigma(1\,2\,3)\sigma^{-1} = (3\,z)(2\,y)(1\,x)(1\,2\,3)(x\,1)(y\,2)(z\,3) = (x\,y\,z).$$

Case 2: $\sigma \notin \mathcal{A}_N$. In this case, note $\sigma(4\,5) \in \mathcal{A}_N$ and consider

$$\begin{aligned}
[\sigma(4\,5)]\,(1\,2\,3)\,[\sigma(4\,5)]^{-1} &= (3\,z)(2\,y)(1\,x)(4\,5)(1\,2\,3)(4\,5)(x\,1)(y\,2)(z\,3) \\
&= (3\,z)(2\,y)(1\,x)\,[(4\,5)(1\,2\,3)(4\,5)]\,(x\,1)(y\,2)(z\,3) \\
&= (3\,z)(2\,y)(1\,x)(1\,2\,3)(x\,1)(y\,2)(z\,3) \\
&= (x\,y\,z).
\end{aligned}$$

Either way, $\tau(1\,2\,3)\tau^{-1} = (x\,y\,z)$ for some $\tau \in \mathcal{A}_N$. Equivalently, $[\tau^{-1}](x\,y\,z)[\tau^{-1}]^{-1} = (1\,2\,3)$ so the punchline is

$$(1\,2\,3) \in H \text{ if and only if } (x\,y\,z) \in H.$$

Therefore if H contains one 3-cycle, it contains $(1\,2\,3)$ and therefore contains all 3-cycles. This finishes Step 2.

Proof of Step 3: Let $N \geq 5$ and suppose $\sigma \in \mathcal{S}_N$ is the product of two transpositions. There are three cases:

Case 1: $\sigma = (a\,b)(a\,b)$. Here, $\sigma = e = (1\,2\,3)(1\,3\,2)$.

Case 2: $\sigma = (a\,b)(b\,c)$ where a, b and c are distinct. In this situation,

$$(a\,b)(b\,c) = (a\,b\,c).$$

Case 3: $\sigma = (a\,b)(c\,d)$ where a, b, c and d are all distinct. In this situation,

$$(a\,b)(c\,d) = (e\,a\,f)(a\,b\,e)(e\,c\,f)(c\,d\,e).$$

This proves that any product of two transpositions can be written as the product of 3-cycles, finishing Step 3.

From Steps 1-3, $H = \mathcal{A}_5$, so $\mathcal{A}_5$ is simple. □

Fact: if $N \geq 5$, then $\mathcal{A}_N$ is simple.

ENRICHMENT: A reasonable question to ask is whether or not there is a "catalog" of all groups. That's way too hard to do - the notion of "group" is just too broad.

A simpler question is to ask whether or not there is a "catalog" of all finite groups. That's also way too hard.

An even simpler question is to ask whether or not there is a "catalog" of all finite simple groups. It turns out that such a catalog exists, and the catalog is kind of weird. Here are four categories of groups which are known to be simple (the first two of which we have studied in this class):

1. cyclic groups of prime order (i.e. $(\mathbb{Z}/p\mathbb{Z}, +)$);
2. alternating groups $\mathcal{A}_N$ where $N \geq 5$;
3. groups of "Lie type";
4. "Tits" groups.

It turns out that if G is a finite simple group, then either G is in one of these categories, or G is one of 26 other groups, called **sporadic groups**. The largest sporadic group is called the **Monster** and has order

$$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$$
$$= 808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000.$$

That's 808 sexdecillion and change.

Chapter 8

Quintic equations

8.1 Solvability by radicals

Background: We are interested in determining whether or not polynomial equations (especially quintic equations) are solvable by radicals. In Chapter 1, I said that an equation was *solvable by radicals* if you can find its solutions using $+, -, \cdot, \div$ and roots.

The problem with this definition is that the phrase "you can find" is too vague. We need a definition of solvability by radicals that is more "mathematical", i.e. refers to the existence of some kind of well-defined *object*.

Definition 8.1 *Let $p \in \mathbb{C}[x]$ be the monic polynomial*

$$p(x) = x^n + a_{n-1}x^{n-1} + ... + a_1x + a_0.$$

A **coefficient field** *of p is any number field $\mathbb{Q}(a_0, a_1, ..., a_{n-1})$.*

POLYNOMIAL	COEFFICIENT FIELD(S)
$x^2 - 3x + 4$	
$x^3 + \sqrt{2}x - 4\sqrt{2} + 3$	
$x^5 - \zeta_3 x^2 + x + 4\zeta_3^2$	

Definition 8.2 *Let $p \in \mathbb{C}[x]$ be the monic polynomial*

$$p(x) = x^n + a_{n-1}x^{n-1} + ... + a_1x + a_0$$

and denote the roots of p as $x_1, x_2, ..., x_n$. The **root field** *of p is $\mathbb{Q}(x_1, x_2, ..., x_n)$.*

Definition 8.3 *Let E and F be number fields with $F \subseteq E$. We say E is a* **radical adjunction** *of F if $E = F(\alpha)$ for some $\alpha \in \mathbb{C}$ such that $\alpha^m \in F$ (i.e. if $E = F(\sqrt[m]{f})$ for some $f \in F$).*

Let E and F be number fields with $F \subseteq E$. We say E is a **radical extension** *of F if*

$$E = F(\alpha_1, ..., \alpha_k)$$

where for each $j \in \{1, ..., k-1\}$, $F(\alpha_1, ..., \alpha_{j+1}) = F(\alpha_1, ..., \alpha_j)(\alpha_{j+1})$ is a radical adjunction of $F(\alpha_1, ..., \alpha_{j+1})$.

A radical extension E of F is called **regular** *if $E = F(\alpha_1, ..., \alpha_k)$ is a radical extension where*

1. *each α_j is a p^{th} root of some element in $F(\alpha_1, ..., \alpha_{j-1})$ for some prime p, and*
2. *either α_j is a p^{th} root of unity, or $F(\alpha_1, ..., \alpha_{j-1})$ already contains all p^{th} roots of unity.*

Example 1: $\mathbb{Q}(\pi)$ is not a radical extension of $\mathbb{Q}$.

Example 2: $\mathbb{Q}(\sqrt{2})$ is a regular radical extension of $\mathbb{Q}$.

Example 3: $\mathbb{Q}(\sqrt[3]{2})$ is a radical extension of $\mathbb{Q}$ which is not regular.

Example 4: $\mathbb{Q}(\zeta_3, \sqrt[3]{2})$ is a regular radical extension of $\mathbb{Q}$.

Now for the definition of solvability by radicals which is useful:

Definition 8.4 *The polynomial equation*

$$p(x) = x^n + a_{n-1}x^{n-1} + ... + a_1x + a_0 = 0$$

is called **solvable by radicals** *(over a given coefficient field F) if there is a radical extension of F containing the root field of p. (WLOG this radical extension can be taken to be regular.)*

Example: quadratic equations

$$p(x) = x^2 + bx + c = 0$$

Example: cubic equations

$$f(x) = x^3 + px + q = 0$$

Example: a quartic equation

$$p(x) = x^4 - 16x^2 + 4 = 0$$

Theorem 8.5 *Every quadratic, cubic and quartic equation is solvable by radicals over* $\mathbb{Q}$.

8.2 Galois groups

Question: How might we show that a polynomial equation $p(x) = 0$ is **not** solvable by radicals?

Definition 8.6 *Let E be an extension of field F. The* **Galois group of E over F,** *denoted $Gal(E/F)$ or $Gal(E : F)$, is the set of automorphisms of E which fix every element of F, i.e.*

$$Gal(E/F) = \{\sigma \in Aut(E) : \sigma(x) = x \text{ for every } x \in F\}.$$

Lemma 8.7 *$Gal(E/F)$ is a group under composition of functions.*

Example 1: $E = \mathbb{Q}(\sqrt{2})$, $F = \mathbb{Q}$

Example 2: $E = \mathbb{Q}(\sqrt[3]{2})$, $F = \mathbb{Q}$

Example 3: $E = \mathbb{Q}(\sqrt{2}, \sqrt{3})$, $F = \mathbb{Q}$

Example 4: $E = \mathbb{Q}(\sqrt{2}, \sqrt{3})$, $F = \mathbb{Q}(\sqrt{2})$

Example 5: $E = \mathbb{Q}(\zeta_3, \sqrt[3]{2})$, $F = \mathbb{Q}$

Example 5: $E = \mathbb{Q}(\zeta_3, \sqrt[3]{2})$, $F = \mathbb{Q}(\zeta_3)$

Theorem 8.8 *Let E be the root field of polynomial $p \in F[x]$. Suppose $Gal(E/F)$ is isomorphic to a subgroup of $\mathcal{S}_n$, where n is the number of roots of E. Then, there is a radical extension $\overline{E}$ of E and a surjective group homomorphism $\pi : Gal(\overline{E}/F) \to Gal(E/F)$.*

PROOF Let the roots of p be $x_1, ..., x_n$. Write $E = F(\alpha_1, ..., \alpha_n)$ and for each α_j, write $\alpha_j = r_j(x_1, ..., x_n)$.

Now for each α_j and each $\sigma \in S_n$, adjoin $r_j(x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(n)})$ to E. Call this radical extension $\overline{E}$.

Now given $\sigma \in Gal(E/F)$, we can think of σ as an element of S_n by looking at how it permutes $x_1, ..., x_n$, and therefore view σ as an element of $Gal(\overline{E}/F)$, which when restricted to E is an automorphism of E. That means the map $\pi : Gal(\overline{E}/F) \to Gal(E/F)$ given by $\pi(\sigma) = \sigma|_E$ is surjective. It is left to show this is a group homomorphism (this is clear, since restriction of functions preserves

composition, but we'll show it rigorously): let $\sigma, \sigma' \in Gal(\overline{E}/F)$, and observe

$$\pi(\sigma\sigma') = (\sigma\sigma')|_E = \sigma|_E\sigma'|_E = \pi(\sigma)\pi(\sigma').$$

Thus π gives the desired group homomorphism. $\square$

Consequence: Every element of $Gal(E/F)$ (where E is the root field and F is a coefficient field of the polynomial p) comes from an element of $Gal(\overline{E}/F)$ for some regular radical extension $\overline{E}$ of F.

Example: $p(x) = x^3 - 2 \in \mathbb{Q}[x]$.

Solvability of groups

Here's an adjective which describes some, but not all groups. Its importance will be seen in a bit:

Definition 8.9 *Let G be a group. G is called* **solvable** *if there exist a sequence of subgroups*

$$\{e\} = G_k \leq G_{k-1} \leq G_{k-2} \leq ... \leq G_2 \leq G_1 \leq G_0 = G$$

such that:

- *each G_j is a normal subgroup of G_{j-1}; and*
- *each quotient group G_{j-1}/G_j is abelian.*

The sequence $G = G_0, G_1, G_2, ..., G_k = \{e\}$ is called a (not "the") **derived series** *for G.*

Remark: Derived series for a solvable group are not unique, but there is a standard way to find the derived series for a solvable group.

Example 1: $G =$ any abelian group.

Example 1a: $G = V$, the Klein 4-group.

Example 2: $G = \mathcal{S}_3 \cong D_3$.

Example 3: $G = \mathcal{S}_4$.

Example 4: $G = \mathcal{S}_5$.

Lemma 8.10 *If G is a solvable group and $\sigma : G \to H$ is a surjective group homomorphism, then H is solvable.*

PROOF HW

Why does solvability of a group matter?

Lemma 8.11 *If $B \subseteq B(\alpha) \subseteq E$ are number fields where $B(\alpha)$ is a regular radical extension of B, then $Gal(E/B(\alpha)) \triangleleft Gal(E/B)$ and $Gal(E/B)/Gal(E/B(\alpha)$ is abelian.*

PROOF Define the function $\pi : Gal(E/B) \to Aut(B(\alpha))$ by

$$\pi(\sigma) = \sigma|_{B(\alpha)},$$

i.e. $\pi(\sigma)$ restricts the domain of σ (which was originally E) to $B(\alpha) \subseteq E$.

First, we show π is a group homomorphism (this is clear since restriction of functions preserves composition):

$$\pi(\sigma\sigma') = (\sigma\sigma')|_{B(\alpha)} = \sigma|_{B(\alpha)}\sigma'|_{B(\alpha)} = \pi(\sigma)\pi(\sigma').$$

Second, we have to show that $\pi(\sigma)$ is in fact an automorphism of $B(\alpha)$. To do this, let $\sigma \in Gal(E/B)$. Since $\sigma(b) = b$ for all $b \in B$, $\sigma|_{B(\alpha)}$ is determined by its value at α. Since $B(\alpha)$ is a regular extension of B, there are two cases:

Case 1: α is a p^{th} root of unity. Then

$$[\sigma(\alpha)]^p = \sigma(\alpha^p) = \sigma(1) = 1$$

so $\sigma(\alpha)$ must also be a p^{th} root of unity, hence is in $B(\alpha)$.

Case 2: α is not a p^{th} root of unity, bur $\alpha^p \in B$ for some prime p and B contains all p^{th} roots of unity. In this situation,

$$[\sigma(\alpha)]^p = \sigma(\alpha^p) = \alpha^p \in B$$

so $\sigma(\alpha) = \zeta\alpha$ where ζ is a p^{th} root of unity (so $\zeta \in B$). Thus $\sigma(\alpha) \in B(\alpha)$.

Third, we show $\ker(\pi) = Gal(E/B(\alpha))$:

$$\begin{aligned}\sigma \in \ker(\pi) &\Leftrightarrow \pi(\sigma) = I_{B(\alpha)} \\ &\Leftrightarrow \sigma(b) = b \text{ for all } b \in B(\alpha) \\ &\Leftrightarrow \sigma \in Gal(E/B(\alpha)) \quad \text{(by definition of } Gal(E/B(\alpha))).\end{aligned}$$

This proves $Gal(E/B(\alpha)) \triangleleft Gal(E/B)$ since every kernel is a normal subgroup.

Last, we show $Gal(E/B)/Gal(E/B(\alpha))$ is abelian. Let $\sigma + \ker(\pi)$ and $\tau + \ker(\pi)$ be cosets in $Gal(E/B)/Gal(E/B(\alpha))$. Then:

Case 1: α is a p^{th} root of unity. Then from above, for any $\sigma \in Gal(E/B)$, $\pi(\sigma)(\alpha)$ is also a p^{th} root of unity, i.e. $\pi(\sigma) = \sigma_i$ where $\sigma_i(\alpha) = \alpha^i$ for some i. Therefore, for any $\sigma, \tau \in Gal(E/B)/Gal(E/B(\alpha))$ with $\pi(\sigma) = \sigma_i$ and $\pi(\tau) = \sigma_j$, we have

$$\pi(\sigma\tau)(\alpha) = \sigma(\alpha^j) = \alpha^{ij} = \alpha^{ji} = \tau(\alpha^i) = \pi(\tau\sigma)(\alpha).$$

Therefore $\pi(\sigma\tau) = \pi(\tau\sigma)$. Now let $\sigma[\ker(\pi)]$ and $\tau[\ker(\pi)]$ be elements of $Gal(E/B)/Gal(E/B(\alpha))$. Let $x \in \sigma\tau\ker(\pi)$; this means $\pi(x) = \pi(\sigma\tau) = \pi(\tau\sigma)$ so $x \in \tau\sigma\ker(\pi)$. Thus $Gal(E/B)/Gal(E/B(\alpha))$ is abelian.

Case 2: α is not a p^{th} root of unity, bur $\alpha^p \in B$ for some prime p and B contains all p^{th} roots of unity. In this situation, for $\sigma \in Gal(E/B)$, $\pi(\sigma)(\alpha) = \zeta^j\alpha$ where $\zeta \in B$ is some p^{th} root of unity, i.e. $\pi(\sigma) = \sigma_j$ where $\sigma_i(\alpha) = \zeta^j\alpha$ for some j. Therefore, for any $\sigma, \tau \in Gal(E/B)/Gal(E/B(\alpha))$ with $\pi(\sigma) = \sigma_i$ and $\pi(\tau) = \sigma_j$, we have

$$\pi(\sigma\tau)(\alpha) = \sigma_i\sigma_j(\alpha) = \zeta^{ij}\alpha = \zeta^{ji}\alpha = \sigma_j\sigma_i(\alpha) = \pi(\tau\sigma)(\alpha).$$

As with Case 1, it follows that $Gal(E/B)/Gal(E/B(\alpha))$ is abelian.

This completes the proof of the lemma. □

Theorem 8.12 *Suppose E is a regular radical extension of F. Then $Gal(E/F)$ is a solvable group.*

PROOF If E is a regular radical extension of F, then

$$F \subseteq F(\alpha_1) \subseteq F(\alpha_1, \alpha_2) \subseteq \cdots F(\alpha_1, ..., \alpha_{k-1}) \subseteq F(\alpha_1, ..., \alpha_k) = E$$

and after renaming, this chain can be written as

$$F = F_0 \subseteq F_1 \subseteq F_2 \subseteq \cdots \subseteq F_{k-1} \subseteq F_k = E$$

where each triple $F_j \subseteq F_{j+1} \subseteq F_{j+2}$ satisfies the hypotheses of the $B, B(\alpha)$ and E of Lemma 8.11. By applying the lemma, we see that

$$\{1\} = Gal(E/E) = Gal(E/F_k) \triangleleft Gal(E/F_{k-1}) \triangleleft \cdots \triangleleft Gal(E/F_1) \triangleleft Gal(E/F)$$

and $Gal(E/F_j)/Gal(E/F_{j+1})$ is abelian for each $j \in \{0, ..., k-1\}$. In other words, $Gal(E/F)$ is a solvable group. □

Remark: The converse of this theorem is also true (see Chapter 9 of Stillwell).

Corollary 8.13 *Let E be the root field of $p \in F[x]$. p is solvable by radicals only if $Gal(E/F)$ is a solvable group.*

PROOF If p is solvable by radicals, then there is a regular radical extension $\overline{E}$ of F containing E. By the preceding theorem, $Gal(\overline{E}/F)$ is a solvable group. By Theorem 8.8, there is a surjective homomorphism $Gal(\overline{E}/F) \to Gal(E/F)$, so by Lemma 8.10, $Gal(E/F)$ must be a solvable group. □

8.3 Quintic equations, revisited

What we know: If there is a polynomial p with root field E such that $Gal(E/\mathbb{Q}) \cong S_5$, then p cannot be solvable by radicals (since S_5 is not a solvable group).

What's left: Is there actually a polynomial whose root field has Galois group $\mathcal{S}_5$ over $\mathbb{Q}$?

Definition 8.14 *Let $H \leq \mathcal{S}_N$. H is called* **transitive** *if for every $a, b \in \{1, ..., N\}$, there is a permutation $\sigma \in H$ such that $\sigma(a) = b$.*

Example 1: $H = \mathcal{S}_N$

Example 2: $H = \mathcal{A}_N$ ($N \geq 3$)

Example 3: $H = <(1\,2\,\cdots\,N)>$

Example 4: $H = \{e, (1\,2), (1\,3), (2\,3), (1\,2\,3), (1\,3\,2)\}$

Theorem 8.15 *Let $p \in \mathbb{Q}[x]$ be an irreducible polynomial with root field E. Then $Gal(E/\mathbb{Q})$ is isomorphic to a transitive subgroup of $\mathcal{S}_N$, where N is the number of distinct roots of p in $\mathbb{C}$.*

PROOF First, let α be a root of p, and let $\sigma \in Gal(E/\mathbb{Q})$. Since σ is an automorphism, and since σ preserves $\mathbb{Q}$, we see $p(\sigma(\alpha)) = \sigma(p(\alpha)) = \sigma(0) = 0$, i.e. $\sigma(\alpha)$ must also be a root of p. Therefore any $\sigma \in Gal(E\mathbb{Q})$ permutes the roots of p, and is therefore isomorphic to a subgroup of $\mathcal{S}_N$.

Recall the Conjugation Theorem from Chapter 6, which says (in the language of this chapter) that for any two roots α and β of p, there is a field isomorphism $\sigma_{\alpha,\beta} : \mathbb{Q}(\alpha) \to \mathbb{Q}(\beta)$ with $\sigma(\alpha) = \beta$ but $\sigma(q) = q$ for all $q \in \mathbb{Q}$.

By the Automorphism Extension Theorem from Chapter 6, this σ extends to an automorphism of E fixing F (i.e. an element of $Gal(E/F)$) by setting $\sigma(x) = x$ for any root of p not in $\mathbb{Q}(\alpha)$. Thus $Gal(E/F)$ contains an element σ taking α to β as desired. □

Theorem 8.16 *Let $H \leq \mathcal{S}_5$ be a transitive subgroup containing a transposition. Then $H = \mathcal{S}_5$.*

PROOF Let $H \leq \mathcal{S}_5$ be any subgroup. Define a relation $\sim$ on $\{1, 2, 3, 4, 5\}$ by

$$i \sim j \Leftrightarrow (i\,j) \in H.$$

This is an equivalence relation (reflexivity follows since $(i\,i) = e \in H$; symmetry is obvious; for transitivity, if $i \sim j$ and $j \sim k$, then $(i\,j)$ and $(j\,k)$ are in H, so since H is a subgroup, $(i\,j)(j\,k) = (i\,k) \in H$ so $i \sim k$).

Now, suppose there is $k \in \{1, ..., 5\}$ such that $1 \not\sim k$. Since H is transitive, there is $\sigma \in H$ with $\sigma(1) = k$, meaning that for any j wth $1 \sim j$,

$$\sigma(i\,j)\sigma^{-1} = (k, \sigma(j)) \in H.$$

Therefore $1 \sim j$ implies $k \sim \sigma(j)$.

At the same time, for any j with $k \sim j$,

$$\sigma^{-1}(k\,j)\sigma = (1, \sigma^{-1}(j)) \in H$$

so $k \sim j$ implies $1 \sim \sigma^{-1}(j)$.

Therefore σ is a bijection between the equivalence class of 1 and the equivalence class of k. Since k is arbitrary, it follows that each equivalence class has the same

cardinality.

Thus each class must have size 1 or 5. It can't be 1, since H contains a transposition, meaning there is a class of size at least 2. Therefore there must only be one class of size 5, i.e. H contains every transposition. Since every element of $\mathcal{S}_5$ is a product of transpositions, H must be all of $\mathcal{S}_5$. $\square$

After 180+ pages, we finally come to a resolution of a problem first posed in Chapter 1:

Theorem 8.17 *Arbitary quintic equations with coefficients in $\mathbb{Q}$ are not solvable by radicals.*

In other words, there is no such thing as a "quintic formula" involving only $+, -, \cdot, \div, \sqrt{}, \sqrt[3]{}, \sqrt[5]{}, \sqrt[p]{}$, etc. which solves generic quintic equations.

PROOF We will show that $p(x) = x^5 - 4x + 2$ is not solvable by radicals. Let E be the root field of p.

Claim 1: p is irreducible.

Claim 2: p has five distinct roots, of which three are real.

Claim 3: $Gal(E/F)$ contains a transposition.

Claims 1-3 imply that $Gal(E/F) \cong \mathcal{S}_5$, which is not a solvable group. Thus p is not solvable by radicals. $\square$

Index