NOTES TO ACCOMPANY *TOOLS OF THE TRADE*

EUGENE LERMAN

Contents

typeset May 19, 2021.

Lecture 1: Sets, informally

Introduction. The goal of the course is to help you get better at writing and reading proofs. Since one learns by doing, we will do a lot of math, including construction of the rational numbers out of integers and real numbers out of rationals.

Warning: Sally sweeps some important details under the rug. I will try and point out when this happens.

An informal discussion of sets. We start by talking about sets. Informally a set is a collection of some (mathematical) objects, which are called elements of the set.

Notation 1.1. We write $x \in A$ to mean "x is an element of a set A."
We write $x \notin A$ to mean "x is not an element of a set A."

For example:

- $\mathbb{N}$ = the set of all natural numbers, $\mathbb{N} = \{1, 2, 3, 4, 5 \ldots\}$;
- $\mathbb{Z}$ = the set of all integers, $\mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}$;
- $\mathbb{Q}$ = the set of all rational numbers:

$$\mathbb{Q} = \left\{ \frac{p}{q} \mid p, q \in \mathbb{Z},\ q \neq 0 \right\}.$$

- $\mathbb{R}$ is the set of all real numbers.

(The notation above is currently standard in North America and, possibly, Western Europe.) Elements of $\mathbb{N}$ are natural numbers. For example, $10^{23} \in \mathbb{N}$, that is, 10^{23} is an element of $\mathbb{N}$. $-1 \notin \mathbb{N}$, that is, -1 is not an element of $\mathbb{N}$. We will show that there is a real positive number x so that $x^2 = 2$. This number is usually denoted by $\sqrt{2}$. We will show that $\sqrt{2} \notin \mathbb{Q}$, that is, $\sqrt{2}$ is not an element of $\mathbb{Q}$, or, less formally, $\sqrt{2}$ is not rational.

We take the point of view that sets are determined by their elements. That is two sets A and B are equal (we write $A = B$) if and only if[1] they have the same elements:

$$A = B \Leftrightarrow (a \in A \Leftrightarrow a \in B).$$

Definition 1.2 (Empty set)**.** The empty set, denoted $\varnothing$, is the set with no elements.

Definition 1.3 (Subsets)**.** A set A is a subset of a set B (we write $A \subset B$ or $A \subseteq B$) if (and only if) any element in A is also an element of B:

$$a \in A \Rightarrow a \in B.$$

(Here and elsewhere $\Rightarrow$ is a shorthand for "implies."

Example 1.4. $\mathbb{N} \subseteq \mathbb{Z}$, $\mathbb{Z} \subseteq \mathbb{Q}$, $\mathbb{Q} \subseteq \mathbb{R}$.

Remark 1.5. For any set A the empty set $\varnothing$ is a subset of A. Why?

Remark 1.6. $A = B \iff (A \subseteq B$ and $B \subseteq A)$. (This is Exercise 1.2.2 in Sally's book.)

Exercise 1.7 (Exercise 1.2.1 in Sally's book.)**.** For a set A,

$$A \subseteq A$$

[1] We may abbreviate "if and only if" as $\Leftrightarrow$ or iff

Proof. $A \subseteq A$ means: for all $a \in A$, $a \in A$, which is true. □

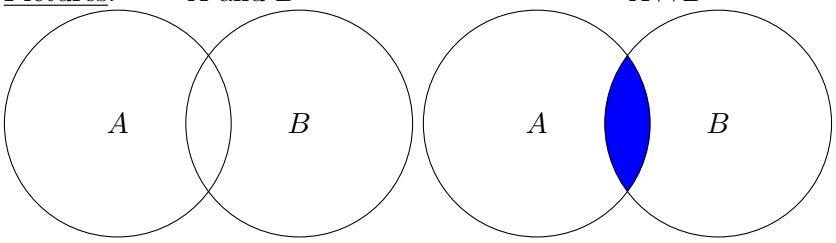
WARNING: In math, "X or Y" means:
X, Y, or *both* X and Y.

Definition 1.8 (union)**.** The union of two sets A and B is the set $A \cup B$ defined by
$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$

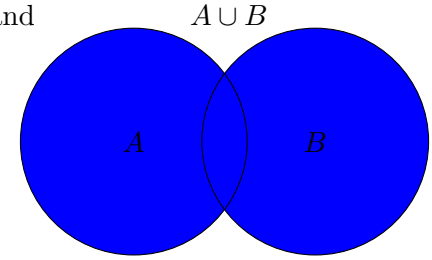
Definition 1.9 (intersection)**.** The intersection of two sets A and B is the set $A \cap B$ defined by

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

Pictures: A and B $A \cap B$



and $A \cup B$



Example 1.10. $A = \{1, 2, 3\}, B = \{1, 2, 4\}$
$A \cup B = \{1, 2, 3, 4\}$
$A \cap B = \{1, 2\}$

Definition 1.11. Two sets A and B are disjoint if $A \cap B = \varnothing$.

Example 1.12. $A = \{1, 2, 3\}, B = \{4, 5, 6, 23\}$.
$A \cap B = \varnothing$
$\implies$ A and B are disjoint. $\{1, 2, 3\}$ and $\{2, 3, 6\}$ are *not* disjoint.

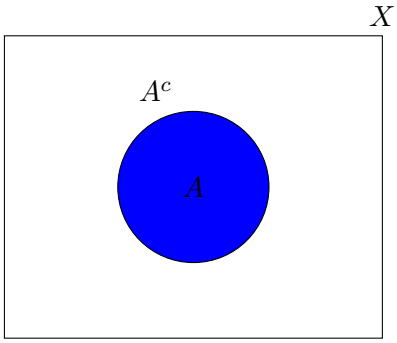
Example 1.13.

$$\{1, 2, 3\} \cap \{2, 4\} = \{2\}$$

$$\{1, 2, 3\} \cap \{4, 5\} = \varnothing$$

Definition 1.14 (Complements)**.** Let X be a set and A a subset of X. The *complement* of A in X is the set

$$A^c := \{x \in X \mid x \notin A\} :$$



Example 1.15. The complement of the set $\mathbb{N}$ of natural numbers inside the set $\mathbb{Z}$ of all integers is the set $\mathbb{N}^c = \{n \in \mathbb{Z} \mid n \notin \mathbb{N}\} = \{0, -1, -2, \ldots\}$.

Example 1.16. The complement of the set A of even integers inside the set X of a all integers is the set of all odd integers.

Example 1.17 (Example (1.3.7) in Sally)**.** For any three sets A, B and C

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

Proof. Suppose $x \in A \cap (B \cup C$. Then

$$x \in A \text{ and } x \in B \cup C.$$

Since $x \in B \cup C$ either $x \in B$ or $x \in C$ (or both). Now if $x \in B$ then $x \in A \cap B$ (since x is always in A). Similarly if $x \in C$ then $x \in A \cap C$. Hence $x \in (A \cap B) \cup (A \cap C)$. Since x is arbitrary we conclude that

$$A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C).$$

Conversely if $x \in (A \cap B) \cup (A \cap C)$ then either $x \in A \cap B$ (and then $x \in A$ and $x \in B$) or $x \in A \cap C$ (and then $x \in A$ and $x \in C$).

In either case $x \in A$ and ($x \in B$ or $x \in C$). Hence $x \in A$ and $x \in B \cup C$. Consequently $x \in A \cap (B \cup C)$. It follows that

$$(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C).$$

We now conclude that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ since they have the same elements. □

Example 1.18. Suppose A, B are two subsets of a set X. Then $(A \cup B)^c = A^c \cap B^c$ (here A^c, B^c mean complements with respect to the set X).

Proof.

$$\begin{aligned}
x \in (A \cup B)^c &\iff x \in X \text{ and } x \notin (A \cup B) \\
&\iff x \in X \text{ and } (x \notin A \text{ and } x \notin B) \\
&\iff (x \in X \text{ and } x \notin A) \text{ and } (x \in X \text{ and } x \notin B) \\
&\iff x \in A^c \text{ and } x \in B^c \\
&\iff x \in A^c \cap B^c
\end{aligned}$$

□

Lecture 2: Ordered pairs, functions

Last time: Informal notion of a set, being an element of a set, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\varnothing$, "$x \in A$", $A \cap B$, $A \subseteq B$, $A \cup B$, If $A \subseteq X$, $A^c := \{x \in X \mid x \notin A\}$.

More generally, given two sets A and B, we can define $A \smallsetminus B := \{a \in A \mid a \notin B\}$. Then if $A \subset X$, the complement of A in X is $A^c = X \smallsetminus A$.

Definition 2.1 (ordered pairs)**.** Let A, B be two sets, $a \in A$, $b \in B$. An *ordered pair* (a, b) is a set with the following property:
For all $a' \in A, b' \in B$

$$(a, b) = (a', b') \quad \iff \quad (a = a' \text{ and } b = b').$$

At this point we don't know if ordered pairs exist, but this issue is not hard to take care of. For example we could define

$$(a, b) := \{\{a\}, \{a, b\}\} \tag{2.1}$$

(See Wikipedia https://en.wikipedia.org/wiki/Ordered_pair for a proof that the set (a, b) defined by (2.1) has the desired property). There are other constructions. What matters is not the construction but the fact that ordered pairs with the desired properties exist.

I presume you have seen ordered pairs before. For example, the coordinate plane $\mathbb{R}^2$ is the set of all ordered pairs of real numbers. Note that the ordered pair (x, y) is *not* the same thing as the set $\{x, y\}$ since $\{x, y\} = \{y, x\}$ and $(x, y) \neq (y, x)$ unless $x = y$.

More generally, given two sets A and B, we define their Cartesian product $A \times B$ as follows:

Definition 2.2. The Cartesian product $A \times B$ of two sets A and B is the set of all ordered pairs (a, b) with the first element in A and the second in B:

$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$

Example 2.3. $A = \{a, b\}$, $B = \{1, 2, 3\}$.
$A \times B = \{(a, 1), (a, 2), (a, 3), (b, 1), (b, 2), (b, 3)\}$

Example 2.4. $A = \{0, 1\}$, $B = \{1, 2, 3\}$.
$A \times B = \{(0, 1), (0, 2), (0, 3), (1, 1), (1, 2), (1, 3)\}$
$B \times A = \{(1, 0), (1, 1), (2, 0), (2, 1), (3, 0), (3, 1)\}$
In particular,

$$A \times B \neq B \times A$$

Remark 2.5. For any set A the Cartesian product $\varnothing \times A$ is empty, since $\varnothing \times A$ has no elements. Similarly $A \times \varnothing = \emptyset$.

Definition 2.6. A relation R on a set X is a subset of $X \times X$.

Notation 2.7. Given a relation R on a set X it is common to write $a \sim b$ or $a \sim_R b$ if (and only if) $(a, b) \in R$.

Recall that an integer $k \in \mathbb{Z}$ is *even* if (and only if) $k = 2\ell$ for some $\ell \in \mathbb{Z}$. An integer $k \in \mathbb{Z}$ is *odd* if (and only if) $k = 2\ell + 1$ for some $\ell \in \mathbb{Z}$. We are now ready for our first example of a relation. We'll see shortly that the relation below is an equivalence relation.

Example 2.8 (Parity relation on the set $\mathbb{Z}$ of integers)**.**

$$\begin{aligned} R &= \{(n, m) \in \mathbb{Z} \times \mathbb{Z} \mid (n \text{ and } m \text{ are odd}) \text{ or } (n \text{ and } m \text{ are even})\} \\ &= \{(n, m) \in \mathbb{Z} \times \mathbb{Z} \mid n - m \text{ is even}\} \end{aligned}$$

Thus

$$n \sim_R m \quad \Leftrightarrow n - m = 2k \text{ for some } k \in \mathbb{Z}.$$

Example 2.9. X any set. The *equality relation* on X is the subset

$$R = \{(x, y) \in X \times X \mid x = y\}$$

Thus $x \sim_R y$ if and only if $x = y$.

Example 2.10. $X = \mathbb{R}$, "$<$" defines a relation

$$R = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x < y\}.$$

Thus, by construction $x \sim_R y$ if and only if $x < y$.

2.1. **Equivalence Relations.**

Definition 2.11. An *equivalence relation* on a set X is a relation $R \subseteq X \times X$ such that the following three conditions hold.

(ER1) (reflexivity) $a \sim a$ for all $a \in X$,
(ER2) (symmetry) If $a \sim b$ then $b \sim a$ (for all $a, b \in X$.
(ER3) (transitivity) if $a \sim b$ and $b \sim c$ then $a \sim c$ (for all $a, b, c \in X$).

(You should memorize this definition — the definition will be on the quiz on Friday, 01/29/2021.)

Example 2.12 (important!)**.** The parity relation R (Example 2.8) is an equivalence relation.

Proof. (1) Since $n - n = 0 = 2 \cdot 0$, $n \sim n$ for all $n \in \mathbb{Z}$. Hence R is reflexive.

(2) If $n - m = 2k$ for some $k \in Z$ then

$$m - n = -(n - m) = (-1)(2k) = 2(-k).$$

Hence R is symmetric.

(3) Suppose $(n, m), (m, r) \in R$. Then $n - m = 2k$ and $m - r = 2\ell$ for some $k, \ell \in \mathbb{Z}$. Hence

$$n - r = n - m + m - r = 2k + 2\ell = 2(k + \ell).$$

Hence $n \sim r$.

Since R is reflexive, symmetric and transitive, it's an equivalence relation. □

We will come back to equivalence relations in a few lectures and see why they are important. In particular we will use equivalence relations to construct the rationals out of integers and the reals out of the rationals.

2.2. **Functions.**

Informally, a function f from a set A to a set B is a rule that assigns to *each* element $a \in A$ *exactly one* element $f(a) \in B$. We write $f : A \longrightarrow B$ to indicate that f is a function from a set A to a set B.

The set A is called the domain of the function f. The set B is the called the range or the codomain or the target of the function f.

Example 2.13. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ is a function from the reals to the reals.

Example 2.14 (A nonexample)**.** $h : [0, \infty) \to \mathbb{R}$, $h(x) = \pm\sqrt{x}$, is not a function: it assigns to 4 *both* 2 and -2.

Formally, there are several ways to define functions. For example we can declare a "function" to be an undefined notion, just like "set" or "is an element of" are undefined notions.

Another approach defines functions in terms of their graphs. This has an advantage (?) of defining functions to be sets. Given a function $f : A \longrightarrow B$ we have a set

$$\text{Graph}(f) = \{(a, b) \in A \times B \mid b = f(a)\},$$

the graph of the function f. Not every subset R of $A \times B$ can be a graph of a function; R has to pass "the vertical line test":

$$(a, b), (a, b') \in R \implies b = b'$$

The set $R = \{(x, y) \in \mathbb{R}^2 \mid y\} = \pm\sqrt{x} = \{(x, y) \in \mathbb{R}^2 \mid y = x^2\}$ fails the vertical line est and so is *not* the graph of a function.

Definition 2.15 (compare Sally, 1.7.3)**.** Let A and B be two sets. A function from A to B is a subset R of the Cartesian product $A \times B$ such that

(1) For all $a \in A$, there is $b \in B$ with $(a, b) \in \mathbb{R}$.
(2) If (a, b) and (a, b') are in R then $b = b'$.

"Each element of A occurs exactly once as a first coordinate."

Remark 2.16. Recall that for any set B, $\varnothing \times B = \varnothing$. So there is *exactly one* function from $\varnothing$ to B: $\varnothing \subseteq \varnothing \times B = \varnothing$.

There are *no* functions from a *nonempty* set A to $\varnothing$. This is because given $a \in A$ there does not exists a $b \in \varnothing$ such that $(a, b) \in A \times \varnothing = \varnothing$.

There is exactly one function from $\varnothing$ to itself: it is the empty function.

Remark 2.17. Of course most mathematicians **don't** think of functions as subsets of Cartesian products. We think of them as rules that assign elements to elements.

Pairs of functions with the appropriate domain and range can be composed.

Definition 2.18 (Composition of functions)**.** Given $f : A \longrightarrow B$, $g : B \longrightarrow C$ we define their composite $g \circ f : A \longrightarrow C$ by

$$(g \circ f)(a) := g(f(a)) \text{ for all } a \in A$$

Thus (secretly ?) we think of functions in terms of what they do to elements of their domain and not as subsets of Cartesian products.

Remark 2.19. For every set A we have the identity function $id_A : A \to A$. It is defined by

$$id_A(a) = a \text{ for all } a \in A.$$

Theorem 2.20 (Sally, 1.7.13)**.** *Composition of functions is associative: for any three functions* $f : A \longrightarrow B$, $g : B \longrightarrow C$, $h : C \longrightarrow D$

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

Proof. For any element $a \in A$ we have

$$\begin{aligned}(h \circ (g \circ f))(a) &= h((g \circ f)(a)) \\ &= h(g(f(a))) \\ &= (h \circ g)(f(a)) \\ &= ((h \circ g) \circ f)(a).\end{aligned}$$

□

Lecture 3: Rules of arithmetic for integers

Sally states "Rules of arithmetic for $\mathbb{Z}$" as follows:

(A1) For all $a, b \in \mathbb{Z}$, the sum $a + b$ is in $\mathbb{Z}$.

(M1) For all $a, b \in \mathbb{Z}$, the product ab is in $\mathbb{Z}$.

(A2) (associativity of $+$) For all $a, b, c \in \mathbb{Z}$,

$$a + (b + c) = (a + b) + c.$$

(M2) (associativity of $\cdot$) For all $a, b, c \in \mathbb{Z}$,

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c$$

(A3) (commutativity of $+$) For all $a, b \in \mathbb{Z}$

$$a + b = b + a.$$

(M3) (commutativity of $\cdot$) For all $a, b \in \mathbb{Z}$

$$a \cdot b = b \cdot a$$

(A4) There is $0 \in \mathbb{Z}$ so that

$$a + 0 = a = 0 + a$$

for all $a \in \mathbb{Z}$.

(M4) There is $1 \in \mathbb{Z}$ so that

$$a \cdot 1 = a = 1 \cdot a$$

for all $a \in \mathbb{Z}$.

(A5) (additive inverses) For all $a \in \mathbb{Z}$ there is $(-a) \in \mathbb{Z}$ such that $a + (-a) = 0 = (-a) + a$.

(D) (distributive law) For all $a, b, c \in \mathbb{Z}$,

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c).$$

Remark 3.1. (A1) says that addition of integers is a function $+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$.

(M1) says that multiplication of integers is a function $\cdot : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$.

(A1), (A2), (A4), (A5) say: $\mathbb{Z}$ with $+$ and 0 is a group.

Additionally (A3) says that the group $(\mathbb{Z}, +, 0)$ is abelian.

(A1-A5), (M1), (M2), (M4) and (D) say: $(\mathbb{Z}, +, \cdot, 0, 1)$ is a ring with unity (that is, with 1).

Adding the condition (M3) says that $(\mathbb{Z}, +, \cdot, 0, 1)$ is a commutative ring.

In other words

Definition 3.2. A commutative ring is a set R together with two functions ("binary operations") $+ : R \times R \to R$, $\cdot : R \times R \to R$ and two elements $0, 1 \in R$ (with $0 \neq 1$) so that the conditions (A1-A5), (M1-M4) and (D) hold for all elements of R.

Facts In a commutative ring R

1) The additive identity is unique. That is, if there is $0' \in R$ so that $a + 0' = a$ for all $a \in R$ then $0' = 0$.

Proof. Suppose $0'$ is another additive identity. Then

$$\begin{aligned} 0' &= 0' + 0 && \text{(since 0 is an additive identity)} \\ &= 0 && \text{(since } 0' \text{ is an additive identity)} \end{aligned}$$

□

(2) Additive inverses are unique: if $a + a' = 0$ then $a' = -a$.

Proof. Suppose $a + a' = 0$. Then

$$\begin{aligned} (-a) &= (-a) + 0 \\ &= (-a) + (a + a') \\ &= ((-a) + a) + a' \\ &= 0 + a' \\ &= a' \end{aligned}$$

□

Remark 3.3. Since additive inverses are unique we have a well-defined *additive inverse function*

$$\text{inv} : R \to R, \quad \text{inv}(a) = -a.$$

(3) Cancellation for addition: For $a, b, c \in R$, if $a + b = a + c$ then $b = c$.

Proof.

$$\begin{aligned} a + b = a + c &\implies (-a) + (a + b) = (-a) + (a + c) \\ &\implies ((-a) + a) + b = ((-a) + a) + c \\ &\implies 0 + b = 0 + c \\ &\implies b = c \end{aligned}$$

□

(4) Multiplicative inverses are unique: if there exists $1' \in R$ so that $1' \cdot a = a$ for all $a \in R$ then $1' = 1$.
(Prove it!)

(5) For any $a \in R$, $a \cdot 0 = 0$:

Proof.

$$\begin{aligned} 0 + (a \cdot 0) &= a \cdot 0 \\ &= a \cdot (0 + 0) \\ &= (a \cdot 0) + (a \cdot 0). \end{aligned}$$

Now, as in the proof of (3) add $-(a \cdot 0)$ to both sides. We get $0 = a \cdot 0$. □

(6) For any $a \in \mathbb{Z}$, $(-1) \cdot a = -a$, that is $(-1)a$ is the additive inverse of a.

Proof.

$$\begin{aligned} a + (-1) \cdot a &= 1 \cdot a + (-1) \cdot a \\ &= (1 + (-1)) \cdot a \\ &= 0 \cdot a \\ &= 0 \end{aligned}$$

Since additive inverses are unique (see (2) above)

$$(-1) \cdot a = -a.$$

□

The ring of integers $\mathbb{Z}$ has an additional property that an arbitrary commutative ring may not have (we'll see examples later on): "$\mathbb{Z}$ has no zero divisors." That is:
for any $a, b \in \mathbb{Z}$, if $a \cdot b = 0$ then either $a = 0$ or $b = 0$.

Lemma 3.4. *Suppose R is a commutative ring with no zero divisors:*

$$a \cdot b = 0 \Rightarrow (a = 0 \text{ or } b = 0).$$

Then R has a cancellation law:

$$(a \cdot b = a \cdot c \text{ and } a \neq 0) \Rightarrow b = c.$$

To prove the Lemma we'll need

Exercise 3.5 (compare with Sally, 1.5.2)**.** For any two elements a, b in a ring R

$$(-a) \cdot b = -(a \cdot b) = a \cdot (-b).$$

Hint: additive inverses are unique.

Proof of Lemma 3.4.

$$\begin{aligned}
& a \cdot b = a \cdot c \\
\Longrightarrow \quad & a \cdot b + (-(a \cdot c)) = 0 \\
\Longrightarrow \quad & a \cdot b + (a \cdot (-c)) = 0 \qquad \text{by the exercise above} \\
\Longrightarrow \quad & a \cdot (b + (-c)) = 0 \qquad \text{distributive law}
\end{aligned}$$

Since $a \neq 0$ by assumption and since the ring R has no zero divisors, we must have $b + (-c) = 0$. This implies that $b = c$ (add c to both sides, use associativity of $+$ and the fact that $b + 0 = b$). □

Lecture 4: Order and the well-ordering principle

Given a commutative ring R there are two ways to define the notion of an order: as a relation $<$ ("less") which must have certain properties or by singling out a set P of positive elements (which also must have certain properties). The two definitions are equivalent. Here are the details of the two definitions:

Definition 4.1. An order on a commutative ring R is a relation $<$ so that

(O1) (trichotomy): For any $a, b \in R$, one and only one of the following holds:

$$a < b, \quad a = b, \quad b < a.$$

(O2) (transitivity): for all $a, b, c \in R$,

$$(a < b \text{ and } b < c) \implies a < c.$$

(O3) (addition) for all $a, b, c \in R$

$$a < b \implies a + c < b + c.$$

(O4) for all $a, b, c \in \mathbb{Z}$,

$$(a < b \text{ and } 0 < c) \implies a \cdot c < b \cdot c.$$

Definition 4.2. Let R be a commutative ring. A subset P of R is a positive cone if (and only if) the following two conditions hold:

(ORD1) For any $a \in R$ either $a \in P$ or $a = 0$ or $-a \in P$, and the three possibilities are *mutually exclusive.*

(ORD2) For any $a, b \in P$ we have: $a + b \in P$ and $ab \in P$.

Lemma 4.3. *The two definitions are equivalent:*

(1) *Given a ring R with a positive cone P the relation $<$ defined by*

$$a < b \quad \Leftrightarrow \quad b - a \in P$$

is an order.

(2) *Given a ring R with an order $<$ the set*

$$P := \{a \in R \mid 0 < a\}$$

is a positive cone.

Proof. Exercise. (This is exercise 1.5.9 in Sally in the case when $R = \mathbb{Z}$. The only thing the proof uses is the fact that $\mathbb{Z}$ is a ring.) □

Notation 4.4. Given an ordered ring R with an order $<$ we write $a \leq b$ iff ($a < b$ or $a = b$);
$a > b$ iff $b < a$
$a \geq b$ iff ($b < a$ or $b = a$)
and so on.

Example 4.5. Consider the set $\mathbb{Z}[x]$ of polynomials in one variable with integral coefficients:

$$\mathbb{Z}[x] = \{a_0 + a_1 x + \cdots + a_n x^n \mid n \geq 0, a_0, \ldots, a_n \in \mathbb{Z}\}.$$

It is a commutative ring under the usual operations of addition and multiplication of polynomials. We define

$$P := \{a_0 + a_1 x + \cdots + a_n x^n \mid a_n > 0\}$$

the set of polynomials with *positive* highest order term. It is easy to check that (ORD1) and (ORD2) hold for P. By Lemma 4.3 there is an order relation $<$ on the ring $\mathbb{Z}[x]$ of polynomials.

Remark 4.6. For any $a \in \mathbb{Z}$, the polynomial $p(x) = x - a$ (that is, $p(x) = 1 \cdot x + (-a)$ if we are being pedantic) is in P since $1 > 0$ in $\mathbb{Z}$. Hence $0 < x - a$. Consequently (using (O2) above) $a < x$ for any $a \in \mathbb{Z}$. We'll come back to this example.

Facts (compare with Facts 1.5.5 in Sally). Let $(R, <)$ be an ordered ring. Then

(1) $a > 0 \iff -a < 0$
(2) If $a > 0$ and $b > 0$ then $ab > 0$
(3) If $a < 0$ and $b > 0$ then $ab < 0$
(4) If $a < 0$ and $b < 0$ then $ab > 0$
(5) If $a \neq 0$ then $a^2 > 0$
(6) $1 > 0$
(7) If $a > b$ and $c < 0$ then $ac < bc$
(8) If $a > b$ then $-a < -b$

I will prove (1), (2), and (6). Read the rest!

Proof. (1) Suppose $0 < a$. Then by (O3), $0 + (-a) < a + (-a)$, i.e. $-a < 0$.
Suppose $-a < 0$ then $a + (-a) < 0 + a$, i.e. $0 < a$.
(2) If $a > 0$ and $b > 0$ then by (O4),

$$0 \cdot a < b \cdot a$$

i.e. $0 < a \cdot b$
(6) $1 = 1 \cdot 1$. If $1 > 0$ then $1 \cdot 1 > 0$ by (2).
If $1 < 0$ then $1 \cdot 1 > 0$ by (4). □

Lemma 4.7. *Suppose $(R, <)$ is an ordered commutative ring. Then R is an integral domain. That is, for any $a, b \in R$*

$$a \cdot b = 0 \text{ implies that } a = 0 \text{ or } b = 0.$$

Proof. Suppose not. Then there are $a, b \in R$ so that $a \cdot b = 0$ but both a and b are nonzero. Since $a \neq 0$ then by trichotomy either $a > 0$ or $a < 0$. Similarly since $b \neq 0$ either $b > 0$ or $b < 0$. We now have four cases to consider.

If $a > 0$ and $b < 0$ that $a \cdot b < 0$ (by Fact (3) above). Hence in particular $a \cdot b \neq 0$. The other three cases are similar. In each case $a \cdot b \neq 0$. But we assumed that $a \cdot b = 0$. Contradiction.

We conclude that $a \cdot b = 0$ implies that $a = 0$ or $b = 0$ (or both). □

Well ordering principle: Any nonempty subset A of *positive* integers has a least element, i.e., there exists $a_0 \in A$ such that $a_0 \leq a$ for all $a \in A$.

Exercise 4.8. Prove that if a, a' are two least elements of $A \subset \mathbb{Z}$, then $a = a'$. Conclude that the notation $\min A$ for a least element of A is justified.

Consequences of the well ordering principle. (1.5.11 in Sally)

(1) There are no integers $a \in \mathbb{Z}$ with $0 < a < 1$.

Proof. Let $A = \{a \in \mathbb{Z} \mid 0 < a < 1\}$. If $A \neq \varnothing$, then by the well-ordering principle A has a least element $a_0 \in A$. Since a_0 is in A, $0 < a_0$ and $a_0 < 1$. Since $0 < a_0 < 1$, $a_0^2 < a_0 \cdot 1 = a_0$. Since $a_0 < 1$, $a_0^2 < 1$.

Since $a_0 > 0$, $a_0^2 > 0$. Now since $0 < a_0^2$ and since $a_0^2 < 1$, $a_0^2 \in A$. But $a_0^2 < a_0$, as we have just seen. That contradicts the fact that a_0 is a least element of A. Hence $A = \varnothing$. □

(2) Well-ordering principle implies the principle of mathematical induction. Recall what this the principle says:

Principle of mathematical induction Let A be a set of positive integers such that
(1) $1 \in A$
(2) if $k \in A$ then $k + 1 \in A$.
Then $A = \mathbb{N}$, the set of *all* positive integers.

Proof that well-ordering ⇒ mathematical induction.
Suppose A is a subset of the set of natural numbers $\mathbb{N}$, A satisfies (1) and (2) above and $A \neq \mathbb{N}$.

Then $A' = N \setminus A := \{n \in \mathbb{N} \mid n \notin A\}$ is nonempty. By the well-ordering principle A' has a least element c. Since $1 \in A$, $c \in A'$ and $A \cap A' = \varnothing$, $c \neq 1$. By trichotomy, either $c < 1$ or $c > 1$ (but not both!). Since $c > 0$, and $c \in \mathbb{Z}$, $c < 1$ is impossible by (1) above. Therefore we must have $c > 1$. Hence $c - 1 := c + (-1) > 1 + (-1) = 0$. Since $c - 1 < c$ and since c is a least element of A' $c - 1 \notin A'$. Therefore $c - 1 \in A$. But then $c = (c - 1) + 1 \in A$ as well (by the assumption on A). Contradiction, since $c \in A'$ and $A \cap A' = \varnothing$.
We got into trouble by assuming that $A \neq \mathbb{N}$. Therefore $A = \mathbb{N}$. □

Lecture 5: Mathematical induction, division algorithm for integers, partitions

Recall what the principle of mathematical induction says:

Principle of mathematical induction Let A be a set of positive integers such that
(1) $1 \in A$
(2) if $k \in A$ then $k + 1 \in A$.
Then $A = \mathbb{N}$, the set of *all* positive integers.

Lemma 5.1. *Well-ordering principle implies principle of mathematical induction.*

Proof. Suppose A is a subset of the set of natural numbers $\mathbb{N}$, A satisfies (1) and (2) above and $A \neq \mathbb{N}$.

Then $A' = N \setminus A := \{n \in \mathbb{N} \mid n \notin A\}$ is nonempty. By the well-ordering principle A' has a least element c. Since $1 \in A$, $c \in A'$ and $A \cap A' = \varnothing$, $c \neq 1$. By trichotomy, either $c < 1$ or $c > 1$ (but not both!). Since $c > 0$, and $c \in \mathbb{Z}$, $c < 1$ is impossible by (1) above. Therefore we must have $c > 1$. Hence $c - 1 := c + (-1) > 1 + (-1) = 0$. Since $c - 1 < c$ and since c is a least element of A' $c - 1 \notin A'$. Therefore $c - 1 \in A$. But then $c = (c - 1) + 1 \in A$ as well (by the assumption on A). Contradiction, since $c \in A'$ and $A \cap A' = \varnothing$.
We got into trouble by assuming that $A \neq \mathbb{N}$. Therefore we musthave $A = \mathbb{N}$. □

Example 5.2. Prove by induction that

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

for any natural number n.
Solution: Let

$$A = \{n \in \mathbb{N} \mid 1 + 2 + \cdots + n = \frac{n(n+1)}{2}\}.$$

We would like to show that $A = \mathbb{N}$.

Note first that

$$1 = \frac{1(1+1)}{2}.$$

Hence $1 \in A$.

Now *suppose* that $k \in A$. Then

$$\begin{aligned} 1 + 2 + \cdots + k + (k+1) &= \tfrac{k(k+1)}{2} + (k+1) \\ &= (k+1)(\tfrac{k}{2} + 1) \\ &= \tfrac{(k+1)((k+1)+1)}{2}. \end{aligned}$$

Consequently $k+1 \in A$. Since $1 \in A$ and since $k \in A \Rightarrow k+1 \in A$, we conclude that $A = \mathbb{N}$ (by the principal of mathematical induction). That is,

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

for all $n \in \mathbb{N}$.

Remark 5.3. The well-ordering principle has the following useful variant:

Well ordering principle 2: Any nonempty subset A of *nonnegative* integers has a least element

This is because if $\varnothing \neq A \subset \mathbb{N} \cup \{0\}$ and if $0 \in A$ then $0 = \min A$. And if $0 \notin A$ then $A \subset \mathbb{N}$ and $\min A$ exists by the well-ordering princple.

Proposition 5.4 (Division algorithm for $\mathbb{Z}$)**.** *For all $a, d \in \mathbb{Z}$ with $d \geq 1$, there exists unique $q, r \in \mathbb{Z}$ such that*

(i) $a = q \cdot d + r$
(ii) $0 \leq r < d$

Proof. (Existence of q and r) We use the well-ordering principle 2.

If $a = 0$, take $q = 0$, $r = 0$. Then $0 = 0 \cdot d + 0$ and $0 < d$. If $d = 1$, let $q = a$, $r = 0$. Then $a = a \cdot 1 + 0$ and $0 < 1$.

Now assume $a \neq 0$, $d > 1$ and consider

$$X = \{a - td \mid t \in \mathbb{Z}, a - td \geq 0\}$$

Claim: $X \neq \varnothing$.

Proof of claim. If $a > 0$, $a - 0 \cdot d = a \in X$. Hence $X \neq \varnothing$.
If $a < 0$, $a - ad = a(1-d) > 0$ since $a < 0$ and $1 - d < 0$. Hence $a - ad \in X$ and $X \neq \varnothing$ in this case as well. □

By well-ordering principle 2, X has the least element, call it r. Then since $r \in X$, $r \geq 0$ and $r = a - q \cdot d$ for some $q \in \mathbb{Z}$. Hence

$$a = q \cdot d + r.$$

It remains to prove that

$$r < d.$$

Suppose not, $r \geq d$. Then

$$0 \leq r - d = a - q \cdot d - d = a - (q+1)d.$$

Therefore $r - d = a - (q+1)d \in X$. On the other hand, since $d > 0$, $r - d < r$. This contradicts the fact that $r \in X$ least. We conclude that $r < d$. This proves the existence of $r, q \in \mathbb{Z}$ with $a = q \cdot d + r$, $0 \leq r \leq d$.

(Uniqueness) Suppose there exist $q_1, q_2, r_1, r_2 \in \mathbb{Z}$ such that

$$a = q_1 d + r_1 = q_2 d + r_2,\ 0 \leq r_1, r_1 < d$$

Then either $r_1 \leq r_2$ or $r_2 \leq r_1$. Say $r_2 \leq r_1$. Then

$$0 \leq r_1 - r_2 = (a - q_1 d) - (a - q_2 d) = (q_2 - q_1)d.$$

Since $r_1, r_2 < d$ and $0 \leq r_2$,

$$d > r_1 - r_2 = (q_1 - q_2)d$$

Therefore

$$0 > (q_1 - q_2)d - d = (q_1 - q_2 - 1)d$$

Since $d > 0$ this can only happen if $q_1 - q_2 - 1 < 0$, i.e., $q_1 - q_2 < 1$. We have seen that there are no integers between 0 and 1. Hence $q_1 - q_2 = 0$. Consequently $r_1 - r_2 = (q_1 - q_2) \cdot d = 0$. We conclude that $q_2 = q_1$ and $r_2 = r_1$. □

Back to equivalence relations. A few lectures ago we defined equivalence relations, but we haven't done anything with them and I haven't told you what they are for. The equivalence relations turn out to be a convenient way to define partitions. To talk about partitions we need two definitions. These definitions generalize intersections and unions of pairs of sets.

Definition 5.5. Let $\mathcal{A}$ be a set of sets (a set whose elements are sets). We define

$$\bigcup \mathcal{A} := \{x \mid x \in A \text{ for some } A \in \mathcal{A}\}$$

Example 5.6. Let A and B be two sets, $\mathcal{A} = \{A, B\}$. Then

$$\bigcup \mathcal{A} = \{x \mid x \in A \text{ or } x \in B\} = A \cup B.$$

Example 5.7. Let $\mathcal{A}$ be the set of horizontal lines in the strip $\mathbb{R} \times [0, 1]$:

$$\mathcal{A} = \{\mathbb{R} \times \{y\} \mid 0 \leq y \leq 1\}.$$

Then

$$\bigcup \mathcal{A} = \mathbb{R} \times [0, 1].$$

Example 5.8. Let A be a set, $\mathcal{A} = \{\{a\} \mid a \in A\}$. Then

$$\bigcup \mathcal{A} = A.$$

Definition 5.9. Let $\mathcal{A}$ be a set of sets. We define

$$\bigcap \mathcal{A} := \{x \mid x \in A \text{ for all } A \in \mathcal{A}\}$$

Example 5.10. Suppose A, B, C are three sets and $\mathcal{A} = \{A, B, C\}$. Then

$$\bigcap \mathcal{A} = A \cap (B \cap C) = \{x \mid x \in A \text{ and } x \in B \text{ and } x \in C\}.$$

Example 5.11. Let $\mathcal{A}$ be the set of horizontal lines in the strip $\mathbb{R} \times [0, 1]$:

$$\mathcal{A} = \{\mathbb{R} \times \{y\} \mid 0 \leq y \leq 1\}.$$

Then

$$\bigcap \mathcal{A} = \varnothing.$$

Lecture 6: partitions and equivalence relations

Definition 6.1. A partition $\mathcal{P}$ of a set X is a set of (nonempty) subsets of X such that

(i) $\bigcup \mathcal{P} = X$ and
(ii) for all $A, B \in \mathcal{P}$, if $A \cap B \neq \varnothing$ then $A = B$. (Equivalently: for all $A, B \in \mathcal{P}$ either $A = B$ or $A \cap B = \varnothing$.)

Example 6.2. $\mathcal{P} = \{\mathbb{R} \times \{y\} \mid 0 \leq y \leq 1\}$ is a partition of the set $\mathbb{R} \times [0, 1]$.

Example 6.3. $\mathcal{P} = \{\text{the set of all even integers}, \text{the set of all odd integers}\}$ is a partition of $\mathbb{Z}$. Note: $\mathcal{P}$ has exactly two elements.

Example 6.4. Let $X = \{1, 2, 3, 4, 5\}$. Then $\mathcal{P} = \{\{1, 2\}, \{4, 3\}\}$ is **not** a partition of X since $\bigcup \mathcal{P} = \{1, 2, 3, 4\} \neq \{1, 2, 3, 4, 5\} = X$.

$\mathcal{P}' = \{\{1, 2\}, \{1, 4\}, \{3, 5\}\}$ is not a partition of X since $\{1, 2\} \cap \{1, 4\} \neq \varnothing$ and $\{1, 2\} \neq \{1, 4\}$.

Recall the definition of a level set of a function.

Definition 6.5. Let $f : X \to Y$ be a function between two sets. For $y \in Y$ the y-level set of f is the set

$$f^{-1}(y) := \{x \in X \mid f(x) = y\}$$

Example 6.6. Suppose $f : X \to Y$ is a function between two sets. The collection $\mathcal{P}$ of all non-empty level sets of f is a partition of X:

$$\mathcal{P} = \{f^{-1}(y) \mid y \in Y, f^{-1}(y) \neq \varnothing\}.$$

This is because:

(1) for any $x \in X$, $x \in f^{-1}(f(x))$ and
(2) if $f^{-1}(y_1) \cap f^{-1}(y_2) \neq \varnothing$, then there is $x_0 \in X$ so that $f(x_0) = y_1$ and $f(x_0) = y_2$. Therefore $y_1 = y_2$. Hence $f^{-1}(y_1) = f^{-1}(y_2)$.

Recall: An *equivalence relation* on a set X is a relation $R \subseteq X \times X$ such that the following three conditions hold.

(reflexivity) $a \sim a$ for all $a \in X$,
(symmetry) If $a \sim b$ then $b \sim a$ (for all $a, b \in X$).
(transitivity) if $a \sim b$ and $b \sim c$ then $a \sim c$ (for all $a, b, c \in X$).

Definition 6.7. Let R be an equivalence relation on a set X and $x \in X$. The equivalence class of x is the set

$$[x] := \{y \in X \mid x \sim y\} = \{y \in X \mid y \sim x\}.$$

Note that Sally denotes the equivalence class of x by $C(x)$.

Example 6.8. Let $X = \mathbb{Z}$. Recall that a relation $\sim$ on X defined by

$$n \sim m \quad \text{if and only if} \quad n - m \quad \text{is even}$$

is an equivalence relation. We have two equivalence classes:

$$[0] = \{n \in \mathbb{Z} \mid n - 0 \text{ is even}\} = \{2k \mid k \in \mathbb{Z}\}$$

and

$$[1] = \{n \in \mathbb{Z} \mid n - 1 \text{ is even}\} = \{2k + 1 \mid k \in \mathbb{Z}\}.$$

Example 6.9. Let $f : X \to Y$ be a function. The relation $\sim$ defined by

$$x \sim x' \quad \text{if and only if} \quad f(x) = f(x')$$

is an equivalence relation (check this!). The equivalence class of $x \in X$ is then

$$[x] = \{x' \in X \mid f(x') = f(x)\} = f^{-1}(f(x)), \quad \text{the } f(x)\text{-level set of the function } f.$$

Proposition 6.10. *Let R be an equivalence relation on a set X. The equivalence classes of R form a partition of X.*

Proof. We need to check that $\mathcal{P} := \{[x] \mid x \in X\}$ is a partition of X: $\bigcup \mathcal{P} = X$ and $[x] \cap [x'] \neq \varnothing \Rightarrow [x] = [x']$.

Since for any $x \in X$, $x \sim x$, $x \in [x]$, $X \subset \bigcup_{x \in X} [x] = \bigcup \mathcal{P}$. On the other hand since for any $x \in X$, $[x] \subset X$, $\bigcup_{x \in X} [x] \subset X$. Therefore $X = \bigcup \mathcal{P}$.

Suppose $[x] \cap [x'] \neq \varnothing$. Then there is $z \in X$ so that $z \in [x]$ and $z \in [x']$. Therefore $z \sim x$ and $z \sim x'$. By symmetry of $R = \sim$, $x \sim z$. By transitivity of $\sim$, $x \sim x'$.

For any $a \in [x]$, $a \sim x$. Transitivity of $\sim$v implies that $a \sim x'$. Therefore $a \in [x']$. We conclude that $[x] \subset [x']$. Similarly $[x'] \subset [x]$. Therefore $[x] = [x']$.

We conclude that $\mathcal{P}\{[x] \mid x \in X\}$ is a partition of X. □

Notation 6.11. Let $\sim$ be an equivalence relation on a set X. We denote the set of equivalence classes of X by $X/_\sim$. That is,

$$X/_\sim := \{[x] \mid x \in X\}$$

The set $X/_\sim$ is sometimes referred to as the quotient set.

Example 6.12. If $X = \mathbb{Z}$ and $n \sim n'$ if and only if $n - n'$ is even, then $\mathbb{Z}/_\sim$ is the two element set $\{\text{even integers}, \text{odd integers}\}$.

Note well (NOTA BENE) Given an equivalence relation $\sim$ on a set X we have a function $\pi : X \to X/_\sim$:

$$\pi(x) = [x]$$

for all $x \in X$.

To have more examples of equivalence relations, equivalence classes and quotient sets, we introduce divisibility of integers.

Definition 6.13. Let $a, b \in \mathbb{Z}$. a divides b if $b = ac$ for some $c \in \mathbb{Z}$.
We write $a|b$ if a divides b.

Lemma 6.14.

(i) $a|0$ *for all* $a \in \mathbb{Z}$.
(ii) $a|a$ *for all* $a \in \mathbb{Z}$.
(iii) *If* $a|b$ *then* $a|(-b)$.
(iv) *If* $a|b$ *and* $b|c$ *then* $a|c$.

Proof. (i) $0 = 0 \cdot a$.
(ii) $a = a \cdot 1$.
(iii) $a|b \Rightarrow b = qa$ for some $q \in \mathbb{Z} \Rightarrow (-b) = (-q)a \Rightarrow a|(-b)$.
(iv) $a|b \Rightarrow b = qa$ for some $q \in \mathbb{Z}$. $b|c \Rightarrow c = q'b$ for some $q' \in \mathbb{Z}$. Hence $c = q'(qa) = (q'q)\, a. \Rightarrow a|c$. □

Lemma 6.15. *Fix an integer* $n \geq 1$. *The relation* $\sim_n$ *on* $\mathbb{Z}$ *defined by*

$$a \sim_n b \quad \Leftrightarrow n|(b - a)$$

is an equivalence relation. The corresponding equivalence classes are

$$[a] = \{a + nk \mid k \in \mathbb{Z}\}.$$

Proof. We check reflexivity, symmetry and transitivity.
(i) For all $a \in \mathbb{Z}$, $a - a = 0 = 0 \cdot n. \implies n|(a - a) \implies a \sim a$. Hence $\sim_n$ is reflexive.
(ii) If $a \sim_n b$ then $n|(b - a$. Hence $n|(-(b - a)) = a - b. \implies b \sim_n a$. Therefore $\sim_n$ is symmetric.
(iii) If $a \sim_n b$ and $b \sim_n c$ then there are $k, \ell \in \mathbb{Z}$ so that

$$a - b = kn \qquad b - c = \ell n.$$

Then

$$a - c = (a - b) + (b - c) = kn + \ell n = (k + \ell)n \implies n|(a - c) \implies a \sim_n c.$$

Therefore $\sim_n$ is transitive. We conclude that $\sim_n$ is an equivalence relation.
$[a] = \{b \in \mathbb{Z} \mid n|b - a\} = \{b \in \mathbb{Z} \mid b - a = nk \text{ for some } k \in \mathbb{Z}\}$
$= \{b \in \mathbb{Z} \mid b = a + nk \text{ for some } k \in \mathbb{Z}\} = \{a + nk \mid k \in \mathbb{Z}\}$. □

Notation 6.16. $a + n\mathbb{Z} := \{a + nk \mid k \in \mathbb{Z}\}$.

Remark 6.17. It's common to write

$$a \equiv b \mod n$$

whenever $n|(a-b)$.

Remark 6.18. The partition of $\mathbb{Z}$ defined by the relation $\sim_n$ has exactly n elements: $[0]$, $[1]$, ..., $[n-1]$.
Reason: By the division algorithm for any $a \in \mathbb{Z}$ there exist unique $q, r \in \mathbb{Z}$ so that $a = qn + r$ and $0 \leq r < n$. Therefore

$$a - r = qn \iff a \sim_n r \iff [a] = [r].$$

Notation 6.19.

$$\mathbb{Z}_n = \mathbb{Z}/_{\sim_n} = \{[0], [1], \ldots, [n-1]\}.$$

It's pronounced "Z mod n" or "Z modulo n."

It turns out that $\mathbb{Z}_n$ is naturally a commutative ring with the addition and multiplication defined by

$$[a] + [b] := [a + b] \quad \text{and} \quad [a][b] := [ab].$$

We'll prove this later.

Lecture 7: Construction of the rationals from the integers; $\mathbb{Z}_n$

Recall: we defined the set of rational numbers to be

$$\mathbb{Q} := \{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\}.$$

This is somewhat imprecise: what exactly is a/b when a, b are two integers? Note that

$$\frac{1}{2} = \frac{3}{6}$$

so the ordered pair $(1, 2) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ represents the same rational number as the pair $(3, 6)$. This suggests that $\mathbb{Q}$ is the quotient of $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ by some equivalence relation $\sim$. The question are:

- what exactly should the relation $\sim$ be? and
- given two equivalence classes $[(a, b]]$ and $[(c, d)]$ in $\mathbb{Q} = (\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}))/_\sim$ how exactly should we define their addition and multiplication so that $\mathbb{Q}$ is a commutative ring?

We define a relation $\sim$ on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ by

$$(a, b) \sim (c, d) \iff ad = bc. \tag{7.2}$$

(This is the familiar fact that $\frac{a}{b} = \frac{c}{d} \iff ad = bc$.)
Claim The relation $\sim$ defined by (7.2) is an equivalence relation.

Proof. (1) $(a, b) \sim (a, b)$ since $ab = ab$.
(2) if $(a, b) \sim (c, d)$ then $ad = bc$. $\implies$ $bc = ad$ $\implies$ $(c, d) \sim (a, b)$.
(3) Suppose $(a, b) \sim (c, d)$, $(c, d) \sim (e, f)$. Then

$$ad = bc \qquad \text{and} \qquad cf = ed.$$

Hence

$$adf = bcf \qquad \text{and} \qquad bcf = bed.$$

$\implies$

$$0 \cdot d = (af - be) \cdot d.$$

Since $d \neq 0$, we have $af - be$. Therefore $(a, b) \sim (e, f)$. □

Notation 7.1. Write $\frac{p}{q}$ for the equivalence class $[(p,q)]$ of $(p,q) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$.

We now turn the set of equivalence classes $\mathbb{Q} = \{[(a,b)] \mid a,b \in \mathbb{Z}, b \neq 0\} = \{\frac{a}{b} \mid a,b \in \mathbb{Z}, b \neq 0\}$ into a commutative ring with 1. In order to do that we need to define two operations $+ : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ and $\cdot : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$. We'd like to define $+$ by

$$\frac{a}{b} + \frac{c}{d} := \frac{ad+bc}{bd}.$$

Issue/potential problem:

If $\frac{a'}{b'} = \frac{a}{b}$ and $\frac{c'}{d'} = \frac{c}{d}$, is $\frac{ad+bc}{bd} = \frac{a'd'+b'c'}{b'd'}$?

In other words we need to prove: if $a'b = ab'$ and $c'd = cd'$ then $(ad+bc)(b'd') = (a'd'+b'c')(bd)$.

Proof.

$$\begin{aligned}(ad+bc)\cdot b'd' - (a'd'+b'c')\cdot bd &= adb'd' + bcb'd' - a'd'bd - b'c'bd \\ &= (ab'-a'b)dd' + (cd'-c'd)bb' = 0+0 = 0\end{aligned}$$

□

Similarly, define $\cdot : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ by

$$\frac{a}{b} \cdot \frac{c}{d} := \frac{ac}{bd}.$$

We need to deal with a similar issue: if $\frac{a}{b} = \frac{a'}{b'}$, $\frac{c}{d} = \frac{c'}{d'}$, does $\frac{ac}{bd} = \frac{a'c'}{b'd'}$?
This is easier. Here is the computation:

$$acb'd' - a'c'bd = acb'd' - a'bcd' + a'bcd' - a'c'bd = (ab'-a'b)cd' + a'b(cd'-c'd) = 0cd' + a'b0 = 0.$$

Therefore $\cdot : \mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$ is well-defined.

- By construction $+$ and $\cdot$ are commutative.
- For any $\frac{a}{b} \in \mathbb{Q}$, $\frac{0}{1} + \frac{a}{b} = \frac{0\cdot b + 1\cdot a}{1\cdot b} = \frac{a}{b}$, so $\frac{0}{1}$ acts like 0. Or, more precisely, $\frac{0}{1}$ is the identity for $+$.
- $\frac{1}{1}$ is the identy for $\cdot$

A few more calculations show: $(\mathbb{Q}, +, \cdot, \frac{0}{1}, \frac{1}{1})$ is a commutative ring.

Unlike the ring $\mathbb{Z}$ of integers, nonzero elements of the ring $\mathbb{Q}$ have multiplicative inverses: if $\frac{a}{b} \neq \frac{0}{1}$ then $a = a \cdot 1 \neq 0 \cdot b = 0$ and then

$$\frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ab} = \frac{1}{1}.$$

In other words: if $x \in \mathbb{Q}$ and $x \neq 0$ there is $y \in \mathbb{Q}$ so that $x \cdot y = 1$. Such commutative rings have a name.

Definition 7.2. A commutative ring $(R, +, \cdot, 0, 1)$ is a field if and only if for every $x \in R$ with $x \neq 0$ there is $y \in R$ so that $xy = 1$.

Example 7.3. The ring $\mathbb{Q}$ of rational numbers is a field. The ring $\mathbb{Z}$ of integers is not a field: the only integers with multiplicative inverses are ± 1. The ring $\mathbb{R}$ of real numbers is a field. We'll spend a fair amount of time constructing it out of rationals.

$\mathbb{Z}_n$, the integers modulo n. Recall that $\mathbb{Z}_n$ is the set of equivalence classes of the relation $\sim_n$ on $\mathbb{Z}$ defined by $a \sim_n b \iff n|(a-b)$. We have seen that

$$\mathbb{Z}_n = \{[0], [1], \ldots, [n-1]\}.$$

This is a consequence of the division algorithm. I promised to construct $+$ and $\cdot$ on $\mathbb{Z}_n$ and make $\mathbb{Z}_n$ into a ring.

Lemma 7.4 (Exercise 1.6.28 in Sally)**.** *The operations $+, \cdot : \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$ defined respectively by*

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

are well-defined.

Proof. We need to check that if $[a] = [a']$ and $[b] = [b']$ then $[a+b] = [a'+b']$ and $[ab] = [a'b']$.

Since $[a] = [a']$, $a \sim_n a' \implies n|(a-a') \implies a - a' = kn$ for some $k \in \mathbb{Z}$. Similarly, since $[b] = [b']$, $b - b' = \ell n$ for some $\ell \in \mathbb{Z}$. Now

$$a+b = (a'+kn)+(b'+\ell n) = (a'+b')+(k+\ell)n. \implies n|\big((a+b)-(a'+b')\big) \implies [a+b] = [a'+b'].$$

Therefore $+ : \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$ is well-defined.

Similarly

$$ab = (a'+kn)(b'+\ell n) = a'b' + knb' + a'\ell n + kn\ell n = a'b' + n(kb' + a'\ell + kn\ell).$$
$$\implies n|(ab - a'b') \implies [ab] = [a'b'].$$

Hence $\cdot : \mathbb{Z}_n \times \mathbb{Z}_n \to \mathbb{Z}_n$ is well-defined as well. □

It is not hard (but it's a bit tedious) to check that $(\mathbb{Z}_n, +, \cdot, [0], [1])$ is a commutative ring.

Example 7.5 (A number is divisible by 3 iff the sum of its digits is a multiple of 3)**.** In $\mathbb{Z}_3$ $[10] = [1]$ since $10 - 1 = 9 = 3 \cdot 3$. Hence (and this should really be proved by induction) for any $n \geq 1$

$$[10^n] = ([10])^n = [1]^n = [1^n] = [1]$$

Therefore $3|(a_0 + a_1 10 + a_2 10^2 + \cdots + a_n 10^n)$ where $a_0, \ldots, a_n \in \{0, 1, \ldots, 9\}$ iff in $\mathbb{Z}_3$

$$[0] = [a_0 + a_1 10 + a_2 10^2 + \cdots + a_n 10^n]$$

Since $\mathbb{Z}_3$ is a ring and since $[10]^k = [1]$ for any $k \geq 1$,

$$\begin{aligned} [a_0 + a_1 10 + a_2 10^2 + \cdots + a_n 10^n] &= [a_0] + [a_1][10] + [a_2][10^2] + \cdots + [a_n][10^n] \\ &= [a_0] + [a_1][1] + \cdots + [a_n][1] = [a_0 + \cdots + a_n]. \end{aligned}$$

Therefore

$$[0] = [a_0 + a_1 10 + a_2 10^2 + \cdots + a_n 10^n] \quad \iff \quad [0] = [a_0 + \cdots + a_n].$$

Lecture 8: Injective, surjective, bijective and invertible functions

Last time:

- Defined the rationals $\mathbb{Q} := \big(\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})\big)/\sim$ where $\sim$ is an appropriate equivalence relation and constructed operations $+$ and $\cdot$ on $\mathbb{Q}$ that make $\mathbb{Q}$ into a field.
- Constructed operations $+$, $\cdot$ on $\mathbb{Z}_n$ that make $\mathbb{Z}_n$ into a commutative ring.

It is important to understand that just because you have a set X with some binary operation $m : X \times X \to X$ and an equivalence relation $\sim$ on X it does not follow that m will give rise to a binary operation on $X/\sim$.

Here is an example of things going wrong.

Example 8.1. Let $X = \mathbb{R}^2$. The level sets of a function $f(x, y) = x^2 + y^2$ partition $\mathbb{R}^2$ into circles centered at $(0, 0)$. The elements of the partition are equivalence classes of the relation $\sim$ defined by

$$(x, y) \sim (x', y') \iff x^2 + y^2 = (x')^2 + (y')^2.$$

Suppose we try to define addition on the equivalence classes (i.e., addition on the set of circles) by setting

$$[(a, b)] + [(c, d)] := [(a + c, b + d)]$$

(i.e., we are treating points of $\mathbb{R}^2$ as vectors and are using the vector addition). Is + well-defined? $[(0, 1)] = [(0, -1)]$ but

$$[(0, 1) + (0, 1)] = [(0, 2)]$$

while

$$[(0, 1) + (0, -1)] = [(0, 0)]$$

and $[(0, 2)] \neq [(0, 0)]$. So + is not well-defined.

Definition 8.2 (1-1 (injective))**.** A function $f : A \to B$ is 1-1 (injective) if and only if

$$f(a) = f(a') \implies a = a'$$

for all $a, a' \in A$.

Equivalently a function $f : A \to B$ is injective if and only if for any $b \in B$ the equation $f(x) = b$ has *at most* one solution (it may have none for a given b but it may not have more than one).

Example 8.3. $f : [0, \infty) \to [0, \infty)$, $f(x) = x^2$ is injective since $x^2 = a$ has at most one *nonnegative* solution.

$f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ is not $1-1$ since the equation $x^2 = 1$ has two solutions: $x = 1$ and $x = -1$.

$f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^x$ is $1 - 1$ since $f(x)$ is strictly increasing. Note that the equation $e^x = -1$ has no solutions. This is why the definition of injective says "at most one."

Definition 8.4 (Onto (surjective))**.** A function $f : A \longrightarrow B$ is *surjective (onto)* iff for any $b \in B$ there is $a \in A$ such that $f(a) = b$ (the equation $f(x) = b$ has a solution for *any* $b \in B$)

Example 8.5. $f : \mathbb{R} \to [0, \infty)$, $f(x) = x^2$ is onto: any nonegative real number has a square root.

$h : \mathbb{R} \to \mathbb{R}$, $h(x) = x^2$ is not onto: negative real numbers have no real square roots.

$g : [-1, 1] \to \mathbb{R}$, $g(x) = x^3$ is not onto. Why?

Definition 8.6 (Bijection)**.** A function $f : A \longrightarrow B$ is a *bijection* if f is $1 - 1$ and onto:
for any $b \in B$ the equation $f(x) = b$ has a *unique* solution.

Example 8.7. Given a set X there is a function $\mathrm{id}_X : X \to X$ that does nothing: $\mathrm{id}_X(x) = x$ for all $x \in X$. id_X is a bijection: the solution of $b = \mathrm{id}_X(x)$ is b.

Example 8.8. The function $f : \mathbb{R} \to (0, \infty)$, $f(x) = e^x$ is a bijection: the equation $b = e^x$ has a unique solution: $x = \ln b$.

The function $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x^3$ is a bijection. The equation $b = x^3$ has a unique solution: $x = \sqrt[3]{(b)}$.

Definition 8.9 (Invertibility)**.**
A function $f : A \to B$ is *invertible* if there exists a function $h : B \longrightarrow A$ such that

$$f \circ h = \mathrm{id}_B \qquad \text{and} \qquad h \circ f = \mathrm{id}_A$$

Equivalently we require $f(h(b)) = b$ for all $b \in B$ and $h(f(a)) = a$ for all $a \in A$.

Lemma 8.10. *Suppose $f : A \to B$ and $h : B \to A$ are two functions so that*

$$h \circ f = \mathrm{id}_A .$$

Then f is 1-1 and h is onto.

Proof. Since $h \circ f = \mathrm{id}_A$

$$h(f(a)) = a$$

for all $a \in A$. Suppose $f(a_1) = f(a_2)$. Then

$$a_1 = h(f(a_1)) = h(f(a_2)) = a_2.$$

Hence f is one-to-one.

Given $a \in A$ the equation $h(x) = a$ has $x = f(a)$ as a solution since $h(f(a)) = a$. □

Corollary 8.11. *If $k : A \to B$ is invertible then k is 1-1 and onto.*

Proof. Since k is invertible, there is $h : B \to A$ (an inverse of k) so that $h \circ k = \mathrm{id}_A$. By Lemma 8.10 k is 1-1.

Also, $k \circ h = \mathrm{id}_B$ since h is an inverse of k. By Lemma 8.10 k is onto. □

Exercise 8.12. Prove a converse of Corollary 8.11: if $f : A \to B$ is 1-1 and onto then f is invertible. Hint: define the inverse $g : B \to A$ by setting $g(b)$ to be the unique solution a of the equation $f(x) = b$.

Exercise 8.12 together with Corollary 8.11 imply that

A function $f : A \to B$ is invertible $\iff$ f is 1-1 and onto.

Lemma 8.13. *Inverses of invertible functions are unique: suppose $f : A \to B$ is a function, $h_1, h_2 : B \to A$ are two of its inverses:*

$$f \circ h_i = \mathrm{id}_B, \qquad \text{and} \qquad h_i \circ f = \mathrm{id}_A \qquad \text{for } i = 1, 2.$$

Then $h_1 = h_2$.

Proof. $h_2 = \mathrm{id}_A \circ h_2 = (h_1 \circ f) \circ h_2 = h_1 \circ (f \circ h_2) = h_1 \circ \mathrm{id}_B = h_1$. □

It now makes sense to talk about the inverse of an invertible map $f : A \to B$. We denote the inverse of f by f^{-1}.

Lecture 9: Pigeonhole principle, size of a finite set

Last time:

- Injective, surjective, bijective and invertible maps.
- $f : A \to B$ is invertible $\iff$ f is a bijection.

Example 9.1. Let X be a set, $\sim$ an equivalence relation on X, $X/\sim$ the set of equivalence classes of $\sim$. Then the quotient map $\pi : X \to X/\sim$, $\pi(x) = [x]$ is always surjective: an element of $X/\sim$ is an equivalence class and it has to be an equivalence class of something.

Example 9.2. There is a map $i : \mathbb{Z} \to \mathbb{Q}$ which is defined by $i(n) = [(n, 1)] \equiv \frac{n}{1}$. This map is 1-1.
Reason:

$$\begin{aligned} i(n) = i(m) &\iff [(n,1)] = [(m,1)] \\ &\iff (n,1) \sim (m,1) \\ &\iff n \cdot 1 = m \cdot 1 \\ &\iff n = m. \end{aligned}$$

Hence i is 1-1.

Definition 9.3. A set A has $n \in \mathbb{N}$ elements if there is a bijection $g : \{1, \ldots, n\} \longrightarrow A$.
A set A is *finite* if either $A = \varnothing$ or if A has $n > 0$ elements. (If $A = \varnothing$ we say that A has 0 elements.)

Question: Suppose we have two bijection

$$f : \{1, \ldots, n\} \longrightarrow A \qquad \text{and} \qquad g : \{1, \ldots, m\} \longrightarrow A$$

for some $n, m \in \mathbb{N}$. For Definition 9.3 to make sense we must have $n = m$. How do we know that? To address this question we first prove:

Lemma 9.4. *The composite of two bijections is a bijection: suppose $f : A \longrightarrow B$, $h : B \longrightarrow C$ are bijections. Then $h \circ f : A \longrightarrow C$ is also a bijection.*

Proof. Since $h : B \longrightarrow C$ is a bijection, for any $c \in C$, there exists a unique $b \in B$ such that $h(b) = c$. Since $f : A \longrightarrow B$ is a bijection, there is a unique $a \in A$ such that $f(a) = b$. And then $h(f(a)) = h(b) = c$.

Hence the equation $h(f(x)) = c$ has a unique solution for any $c \in C$, i.e., $h \circ f$ is a bijection. □

How does Lemma 9.4 help us answer the question? Suppose $f : \{1, \ldots, n\} \longrightarrow A$ and $g : \{1, \ldots, m\} \longrightarrow A$ are two bijections. Then the inverse of g, $g^{-1} : A \to \{1, \ldots, m\}$ is a also a bijection (why?). Next, by Lemma 9.4, $g^{-1} \circ f : \{1, \ldots, n\} \longrightarrow \{1, \ldots, m\}$ is a bijection. We now argue in several steps that n must equal m.

Theorem 9.5 (Pigeonhole Principle)**.** *Suppose $m, n \in \mathbb{N}$ with $m > n$ and $f : \{1, \ldots, m\} \to \{1, \ldots, n\}$ is a function. Then f is* not *injective: there exist $a, b \in \{1, \ldots, n\}$ such that $f(a) = f(b)$.*

Proof. (Induction on n):
Base Case: $n = 1$, $m > 1$. Then a functions $f : \{1, \ldots, m\} \to \{1\}$ can take only one value, namely 1. In particular we have $f(1) = 1 = f(2)$. Hence f is not $1 - 1$.

Inductive step. Suppose the Pigeonhole Principle holds for $n = k$: any function $f : \{1, \ldots, m\} \longrightarrow \{1, \ldots, k\}$ with $m > k$ is not injective. Consider a function $h : \{1, \ldots, l\} \longrightarrow \{1, \ldots, k, k+1\}$, $l > k + 1$.

If there are $x_1, x_2 \in \{1, \ldots, l\}$ with $h(x_1) = k + 1 = h(x_2)$ and $x_1 \neq x_2$, then h is not injective and we are done.

If $k + 1 \neq h(x)$ for any $x \in \{1, \ldots, l\}$, then h is a map from $\{1, \ldots, l\}$ to $\{1, \ldots, k\}$. Since $k < k + 1 < l$, h is *not* injective by the inductive hypothesis. So we are done again.

Finally suppose there is a unique $x \in \{1, \ldots, l\}$ with $h(x) = k + 1$. Define a function $g : \{1, \ldots, l - 1\} \to \{1, \ldots, k\}$ by:

$$g(i) = \begin{cases} h(i) & \text{if } i < x \\ h(i+1) & \text{if } i \geq x \end{cases}.$$

Since $l > k+1$, $l-1 > k$. Hence g is *not* injective: there are $i < j$ with $g(i) = g(j)$. This implies that the function h is not injective either.

Conclusion: by the principle of mathematical induction the pigeonhole principle holds. □

Corollary 9.6. *If a function $f : \{1, \ldots, m\} \longrightarrow \{1, \ldots, n\}$ is injective, then $m \leq n$*

Proof. If $m > n$ then the Pigeonhole Principles implies that the function f is *not* injective. Therefore, since f *is* injective, we must have $m \leq n$. □

Corollary 9.7. *If $h : \{1, \ldots, m\} \longrightarrow \{1, \ldots, n\}$ is a bijection then $m = n$.*

Proof. Since h is injective, Corollary 9.6 implies that $m \leq n$. Since h^{-1} is also injective, Corollary 9.6 implies that $n \leq m$. Hence $m = n$. □

Conclusion: If $f : \{1, \ldots, n\} \to A$, $g : \{1, \ldots, m\} \to A$ are two bijections for some $n, m \in \mathbb{N}$, then $g^{-1} \circ f : \{1, \ldots, n\} \to \{1, \ldots, m\}$ is a bijection hence by Corollary 9.7 $n = m$. Therefore if A is finite we *can* define the size (or cardinality) $|A|$ of A as follows:

$$|A| := \begin{cases} 0 & \text{if } A = \varnothing \\ n & \text{if there is a bijection } f :]\{1, \ldots, n\} \to A. \end{cases}$$

Definition 9.8. Two sets A and B (not necessarily finite) have the same cardinality (i.e., same size) iff there is a bijection $f : A \to B$.

Notation 9.9. We write $|A| = |B|$ if two sets A and B have the same cardinality.

Comparison of sizes/cardinalities

- If there is an injective function $f : A \to B$ between two sets we write $|A| \leq |B|$ (i.e., the cardinality of A is less than or equal the cardinality of B).
- If there is a surjective map $g : A \to B$ we write $|A| \geq |B|$ (i.e., A is no smaller than B).

However we should be careful: it's not obvious that if $g : B \to A$ is surjective then there is an injective map $f : A \to B$ (i.e., $|A| \geq |B| \implies |B| \leq |A|$). Nor is it clear that if $|A| \leq |B|$ and $|B| \leq |A|$ then $|A| = |B|$. We'll come back to these issues.

Lecture 10: Powerset of a set, images and premiamges

Definition 10.1 (Power set $\mathcal{P}(A)$ of a set A)**.** The power set $\mathcal{P}(A)$ of a set A is the collection of all subsets of A.

Example 10.2. $A = \{1, 2, 3\}$. $\mathcal{P}(A) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\} = A\}$. Note that $|\mathcal{P}(A)| = 8 = 2^3$. This is not a coincidence.

Notation 10.3. Let A be a set.

$$2^A := \text{ the set of all functions from } A \text{ to } \{0, 1\}.$$

For any set A there is a bijection between the set 2^A of indicator functions and the power set of A; see Proposition 10.15 below.

Exercise 10.4. For any finite set A, $|2^A| = 2^{|A|}$.
Hint: induction on $n = |A|$.

Remark 10.5. For any set $A \neq \varnothing$ there is an injective map $f : A \to \mathcal{P}(A)$; it is defined by

$$f(a) = \{a\},$$

where $\{a\}$ is the set with one element — a. So our intuition should tell us that $|A| \leq |\mathcal{P}(A)|$. (If $A = \varnothing$, we have the empty map $\varnothing : \varnothing \to \mathcal{P}(\varnothing) = \{\varnothing\}$ and $|\varnothing| = 0 < 1 = |\{\varnothing\}|$, so in this case too $|A| \leq |\mathcal{P}(A)|$. I suppose we can consider the empty map $\varnothing$ injective.) The point of the next theorem is that for any set A the cardinality of the power set $\mathcal{P}(A)$ is strictly bigger than the cardinality of A.

Theorem 10.6 (Cantor, 1891)**.** *For any set A there is no surjective map from A to $\mathcal{P}(A)$.*

Proof. If $A = \varnothing$ then $\mathcal{P}(A) = \{\varnothing\}$. There is only one map from $\varnothing$ to $\mathcal{P}(\varnothing)$ — the empty map $\varnothing$ — and the map $\varnothing : \varnothing \to \{\varnothing\}$ is not onto.

Now suppose $A \neq \varnothing$ and suppose there is a surjective map $f : A \to \mathcal{P}(A)$. Consider the set

$$S = \{a \in A \mid a \notin f(a)\}.$$

Then S is a subset of A hence an element of $\mathcal{P}(A)$. Since f is onto there must exist $x \in A$ so that $f(x) = S$.

If $x \in S = f(x)$ then $x \notin f(x) = S$. So we cannot have $x \in S$ and we must have $x \notin S$.

But if $x \notin S = f(x)$ then $x \in S$. Contradiction.

We conclude that no function $f : A \to \mathcal{P}(A)$ can be surjective. □

Images and preimages of sets.

Definition 10.7. Let A and B be two sets and $f : A \to B$ a function. For any subset $X \subset A$ of A we define the image of X to be the set

$$f(X) = \{f(x) \mid x \in X\}.$$

The set $f(X)$ is a subset of B.

For any subset $Y \subset B$ we define the preimage of Y to be the set

$$f^{-1}(Y) := \{a \in A \mid f(a) \in Y\}.$$

The set $f^{-1}(Y)$ is a subset of A.

Example 10.8. Consider a function $f : \mathbb{R} \longrightarrow \mathbb{R}$, $f(x) = x^2$.
$f([-1, 2]) = \{x^2 \mid x \in [-1, 2]\} = [0, 4]$.
$f^{-1}([0, 4]) = \{x \mid x^2 \in [0, 4]\} = \{x \mid 0 \leq x^2 \leq 4\} = [-2, 2]$.

Remark 10.9.

(1) $f^{-1}(Y)$ is ambiguous if $f : A \to B$ is invertible: is this the preimage of $Y \subset B$ under f or the image of Y under f^{-1}? Fortunately the two sets are equal:

$$f^{-1}(Y) = \{a \in A \mid f(a) \in Y\} = \{f^{-1}(y) \mid y \in Y\} = (f^{-1})(Y).$$

(2) Given $f : A \to B$ and $b \in B$, $f^{-1}(\{b\}) = f^{-1}(b)$, the b-level set of f.

(3) The notation is abused further: a function $f : A \to B$ induces a function

$$f : \mathcal{P}(A) \to \mathcal{P}(B), \quad X \mapsto f(X)$$

and a function

$$f^{-1} : \mathcal{P}(B) \to \mathcal{P}(A) \quad Y \mapsto f^{-1}(Y).$$

(4) $f : A \to B$ is onto if and only if $f(A) = B$.

(5) $f : A \to B$ is injective if and only if for any $b \in B$, $|f^{-1}(\{b\})| \leq 1$.

Families of sets (Sally, p. 27).

Definition 10.10. A family $\{A_i\}_{i\in I}$ of sets indexed by a set I is a function that assigns to each element $i \in I$ a set A_i.

Definition 10.11. Let $\{A_i\}_{i\in I}$ be a family of sets indexed by a set I. The union of the family of sets is the set

$$\bigcup_{i\in I} A_i := \{a \mid a \in A_i \text{ for some } i \in I\},$$

The intersection of the family of sets is the set

$$\bigcap_{i\in I} A_i := \{a \mid a \in A_i \text{ for all } i \in I\},$$

Exercise 10.12 (Exercise 1.7.27 (i) in Sally)**.** For a function $f : A \longrightarrow B$ and a family $\{A_i\}_{i\in I}$ of subsets of A,

$$f(\bigcup_{i\in I} A_i) = \bigcup_{i\in I} f(A_i)$$

Solution.

$$\begin{aligned} b \in f(\bigcup_{i\in I} A_i) &\iff b = f(a) \text{ for some } a \in \bigcup_{i\in I} A_i \\ &\iff b = f(a) \text{ for some } a \in A_j \text{ and some } j \in I \\ &\iff b \in f(A_j) \text{ for some } j \in I \\ &\iff b \in \bigcup_{i\in I} f(A_i). \end{aligned}$$

□

Exercise 10.13 (Exercise 1.7.27 (iv) in Sally)**.** For any map $f : A \longrightarrow B$ and for any family $\{B_i\}_{i\in I}$ of subsets of B,

$$f^{-1}(\bigcap_{i\in I} B_i) = \bigcap_{i\in I} f^{-1}(B_i)$$

Solution.

$$\begin{aligned} a \in f^{-1}(\bigcap_{i\in I} B_i) &\iff f(a) \in \bigcap_{i\in I} B_i \\ &\iff f(a) \in B_i \text{ for all } i \in I \\ &\iff a \in f^{-1}(B_i) \text{ for all } i \in I \\ &\iff a \in \bigcap_{i\in I} f^{-1}(B_i). \end{aligned}$$

□

Remark 10.14. In general $f(\bigcap_{i\in I} A_i) \subseteq \bigcap_{i\in I} f(A_i)$ but $f(\bigcap_{i\in I} A_i)$ may be stricktly smaller than $\bigcap_{i\in I} f(A_i)$. Here is an example. Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. Let $I = \{1, 2\}$, $A_1 = (-1, 0)$, $A_2 = (0, 1)$. Then $\bigcap_{i\in I} A_i = (-1, 0) \cap (0, 1) = \varnothing$ and $f(\bigcap_{i\in I} A_i) = f(\varnothing) = \varnothing$. On the other hand $\bigcap_{i\in I} f(A_i) = f((-1, 0)) \cap f((0, 1)) = (0, 1) \cap (0, 1) = (0, 1)$.

Back to the power sets.

Proposition 10.15. *For a set A there is a bijection*

$$2^A \longleftrightarrow \mathcal{P}(A)$$

where 2^A is the set of all functions from A to $\{0,1\}$ (the so called indicator functions) and $\mathcal{P}(A)$ is the power set of A.

Proof. Given a function $f : A \to \{0,1\}$ define the corresponding subset of A to be $f^{-1}(1) = \{a \in A \mid f(a) = 1\}$. This gives us a function $\varphi : 2^A \to \mathcal{P}(A)$, $f \mapsto f^{-1}(1)$.

Next we construct an inverse $\mathcal{X} : \mathcal{P}(A) \to 2^A$ of φ as follows: Given a subset $X \subset A$ define $\mathcal{X}_X : A \to \{0,1\}$ to be the indicator function:

$$\mathcal{X}_X(a) = \begin{cases} 1 & \text{if } a \in X \\ 0 & \text{if } a \notin X. \end{cases}$$

This gives us a map $\mathcal{X} : \mathcal{P}(A) \to 2^A$, $X \mapsto \mathcal{X}_X$.

We still need to check that $\mathcal{X}$ and φ are inverses of each other. For any $X \subset A$,

$$\varphi(\mathcal{X}_X) = (\mathcal{X}_X)^{-1}(1) = X.$$

Hence $(\varphi \circ \mathcal{X})(X) = X$ for all $X \subset A$. Conversely given a function $f : A \to \{0,1\}$ the indicator function $\mathcal{X}_{f^{-1}(1)}$ takes value 1 exactly on the set $f^{-1}(1)$ and is zero elsewhere. Hence $\mathcal{X}_{f^{-1}(1)} = f$. Therefore $(\mathcal{X} \circ \varphi)(f) = f$ for all $f \in 2^A$ and we are done. □

Remark 10.16. Proposition 10.15 together with Exercise 10.4 show that $|\mathcal{P}(A)| = 2^{|A|}$ for any finite set A.

Lecture 11: Schröder-Bernstein theorem

Last time: for any set A there is an injective map from A to its power set $\mathcal{P}(A)$ but there is no surjective map from A to $\mathcal{P}(A)$. Think: the size of the power set $\mathcal{P}(A)$ is strictly bigger than the size of A.

Recall: we think $|A| \leq |B|$ (the size of a set A is no bigger than the size of a set B) if there is an injective map $f : A \to B$. Now suppose $|A| \leq |B|$ and $|B| \leq |A|$. Are the sets A and B then of the same size? The answer is "yes":

Theorem 11.1 (Schröder-Bernstein)**.** *Suppose A, B are sets and $f : A \to B$, $g : B \to A$ are two injective maps. Then there is a bijection $h : A \longrightarrow B$.*

Example 11.2. The function $f : [-\pi/2, \pi/2] \to \mathbb{R}$, $f(x) = x$ is injective. On the other hand, $g : \mathbb{R} \to [-\pi/2, \pi/2]$, $g(y) = \arctan(y)$ is also injective. Neither f nor g are bijections. Nonetheless by Schröder-Bernstein theorem there is a bijection $h : [-\pi/2, \pi/2] \to \mathbb{R}$

To prove Schröder-Bernstein theorem, we need a bit of preparation.

Definition 11.3 (Restriction of a function to a subset)**.** . Suppose $f : X \to Y$ is a function, $Z \subset X$ a subset. The restriction of f to Z is the function $f|_Z : Z \to Y$ defined by

$$(f|_Z)(z) = f(z) \qquad \text{for all } z \in Z.$$

Example 11.4. The function $f : [-\pi/2, \pi/2] \to \mathbb{R}$, $f(x) = x$ is the restriction of the identity function $\mathrm{id}_{\mathbb{R}} : \mathbb{R} \to \mathbb{R}$ to the closed interval $[-\pi/2, \pi/2]$.

Exercise 11.5. Suppose $f : X \to Y$ is 1-1 and $Z \subset X$ is a subset. Then the restriction $f|_Z$ is also 1-1.

Definition 11.6 (Set difference)**.** Let A and B be two sets. We define

$$A \smallsetminus B := \{a \in A \mid a \notin B\}.$$

Definition 11.7 (Disjoint union)**.** A set Z is a disjoint union of sets A and B iff

- $Z = A \cup B$ and
- $A \cap B = \varnothing$.

We denote the disjoint union of A and B by $A \sqcup B$.

Example 11.8. For any set Z and any subset $X \subset Z$ of Z,

$$Z = X \sqcup (Z \smallsetminus X).$$

Exercise 11.9. Suppose $f : A \to A'$, $g : B \to B'$ are two bijections. Then $h : A \sqcup B \to A' \sqcup B'$ defined by

$$h(x) = \begin{cases} f(x) & x \in A \\ g(x) & x \in B \end{cases}$$

is a bijection.

Proof of Schröder-Bernstein theorem. Since g is injective, for any $a \in A$ either $g^{-1}(a) = \varnothing$ or there is a unique $b \in B$ such that $g(b) = a$. If such b exists we call it the predecessor of a. Similarly we call $a \in A$ the predecessor of $b \in B$ if $f(a) = b$.

Lineage. For $a_0 \in A$ the linage of a_0 is the sequence (possibly infinite, possibly finite):

$$a_0, b_1, a_1, b_2, a_2, \ldots$$

so that $g(b_n) = a_{n-1}$ and $f(a_n) = b_n$ for all $n > 0$.
Similarly every $b_0 \in B$ has a lineage:

$$b_0, a_1, b_1, a_2, b_2, \ldots$$

with $f(a_n) = b_{n-1}$, $g(b_n) = a_n$ for all $n > 0$.

We say that the lineage of $a \in A$ ends in A if it is of the form

$$a, b_1, a_1, b_2, \ldots, b_n, a_n$$

for some $n \geq 0$ (so $g^{-1}(a_n) = \varnothing$). Similarly the lineage of $b \in B$ ends in A if it is of the form

$$b, a_1, b_1, a_2, b_2, \ldots, b_n, a_n.$$

Let

$$A_A = \{a \in A \mid \text{ lineage of } a \text{ ends in } A\}$$
$$B_A = \{b \in B \mid \text{ lineage of } b \text{ ends in } A\}$$

Claim 1: $f(A_A) = B_A$.

Proof. If $b \in f(A_A)$, $b = f(a)$ and a has lineage

$$a, b_1, a_1, b_2, \ldots, b_n, a_n$$

(and $g^{-1}(a_n) = \varnothing$). Then b has the lineage

$$b, a, b_1, a_1, b_2, \ldots, b_n, a_n.$$

Hence $b \in B_A$. This implies that $f(A_A) \subseteq B_A$.

Conversely if $b \in B_A$ then the lineage of b must be of the form

$$b, a_1, b_1, \ldots, a_n \quad \text{for some } n \geq 1.$$

Hence $b = f(a_1)$ for some $a_1 \in A$ and $a_1 \in A_A$. Therefore $B_A \subseteq f(A_A)$. □

Since $f : A \longrightarrow B$ is injective, $f|_{A_A} : A_A \longrightarrow B$ is injective. Since $f(A_A) = B_A$,

$$f|_{A_A} : A_A \longrightarrow B_A$$

is a *bijection*.

Claim 2: $g(B \setminus B_A) = A \setminus A_A$.

Proof. If $b \in B$ has lineage $b, a_1, b_1, a_2, b_2 \ldots$ then $g(b) \in A$ has lineage

$$g(b), b, a_1, b_1, a_2, b_2, \ldots$$

Hence if $b \in B \setminus B_A$ then $g(b) \in A \setminus A_A$. Therefore

$$g(B \setminus B_A) \subseteq A \setminus A_A.$$

If $a \in A \setminus A_A$, then $g^{-1}(a) \neq \varnothing$ for otherwise the lineage of a is just $a \in A$. Hence $a = g(b)$ for some $b \in B$. Moreover, since $a \notin A_A$, the lineage of a does not end in A. Consequently the lineage of b doesn't end in A either, that is, $b \in B \setminus B_A$. Hence $a \in g(B \setminus B_A)$ and therefore

$$A \setminus A_A \subseteq g(B \setminus B_A).$$

□

Claim 2 implies that $g|_{B \setminus B_A} : B \setminus B_A \to A \setminus A_A$ is a bijection (same argument as right after the proof of claim 1). Since $g|_{B \setminus B_A}$ is a bijection, it's invertible and the inverse

$$(g|_{B \setminus B_A})^{-1} : A \setminus A_A \to B \setminus B_A, \qquad a \mapsto g^{-1}(a)$$

is a bijection. Now define $h : A \to B$ by

$$h(a) = \begin{cases} f(a) & a \in A_A \\ g^{-1}(a) & a \notin A_A. \end{cases}$$

By Exercise 11.9 h a bijection. □

Lecture 12: Axiom of choice, countable sets

Last time we proved Schröder-Bernstein theorem:

Given two injective maps $f : A \to B$ and $g : B \to A$ there exists a bijection $h : A \to B$.

Here is an "application" of Schröder-Bernstein theorem:

Theorem 12.1. *The collection of all sets is* not *a set.*

Proof. Suppose the collection of all sets *is* a set; call it A. Then the power set $\mathcal{P}(A)$ of A (the set of all subsets of A) is also a set. Every element of $x \in \mathcal{P}(A)$ is also a set, hence is an *element* of A. This implies that $\mathcal{P}(A)$ is a *subset* of A. Consider the inclusion map $i : \mathcal{P}(A) \longrightarrow A$, $i(x) = x$; the map i is injective. On the other hand we have an injective map $j : A \longrightarrow \mathcal{P}(A)$, $j(x) = \{x\}$ for all $x \in A$. By Schröder-Bernstein Theorem (Theorem 11.1) there exists a bijection $h : A \longrightarrow \mathcal{P}(A)$, but this contradicts Cantor's Theorem (Theorem 10.6). Therefore the collection of all sets cannot be a set. □

We next discuss the axiom of choice. This axiom exists in many equivalent versions that often look quite different. Here is one popular version:

Axiom of choice Let $\{A_i\}_{i\in I}$ be a family of nonempty sets index by some (nonempty) set I. Then there is a function $f : I \to \bigcup_{i\in I} A_i$ with $f(i) \in A_i$ for all $i \in I$. In other words
"given a family of sets $\{A_i\}_{i\in I}$ we can choose an element in each set A_i."

The axiom of choice looks innocuous, but it has many consequences. Some are very useful, other are paradoxical (look up Banach-Tarski paradox if you are curious). We will not touch on paradoxes in this lecture.

Theorem 12.2. *The axiom of choice is equivalent to:*

Given a surjective function $g : B \to A$ there is a function $h : A \to B$ so that

$$g(h(a)) = a$$

for all $a \in A$.

In particular the axiom of choice implies that for any two sets A and B if there is a surjective function $g : B \to A$ then there exists an injective function $h : A \to B$.

Proof. Assume the axiom of choice. Suppose the map $g : B \to A$ is onto. The for each $a \in A$ the set $g^{-1}(a) = \{b \in B \mid g(b) = a\}$ is *nonempty*. Hence by the axiom of choice, for each $a \in A$ we can choose an element in $g^{-1}(a)$. Call this choice $h(a)$. These choices define a function $h : A \to B$ with the property that $g(h(a)) = a$ for all $a \in A$.

Conversely suppose that for any sets B and A and for any surjective function $g : B \to A$ there is an function $h : A \to B$ so that $g(h(a)) = a$ for all $a \in A$. Suppose next we have a family of nonempty sets $\{A_i\}_{i\in I}$ indexed by some set I. Let $B = \bigcup_{i\in I} A_i \times \{i\}$. Then each $b \in B$ is of the form $b = (a, i)$ for some $i \in I$ and $a \in A_i$. Moreover if $(a, i) = (a', i')$ then $i = i'$. Therefore we have a well-defined function $g : B \to I$ which sends $b = (a, i)$ to $i \in I$. Clearly g is onto. Therefore by our assumption there exists a function $h : I \to B$ with $g(h(i)) = i$ for all $i \in I$. Each element $h(i)$ is of the form (a, i) with $a \in A_i$. Define $f(i) \in A_i$ to be this a. In other words we define $f : I \to \bigcup_{i\in I} A_i$ by setting $f(i)$ to be the first coordinate of the function $h(i)$. Therefore the axiom of choice holds.

□

Corollary 12.3. *Suppose $k : A \to B$ and $l : B \to A$ are two surjective maps. Then there exists a bijection $h : A \to B$.*

Proof. Since the map k is onto, there exists by Theorem 12.2 an injection $g : B \to A$. Similarly there exists an injection $f : A \to B$. Schröder-Bernstein Theorem implies that there exists a bijection $h : A \to B$. □

Lemma 12.4. *Suppose D is an infinite subset of the set $\mathbb{N}$ of natural numbers. Then there is a bijection $h : D \to \mathbb{N}$.*

Proof. The inclusion map $i : D \to \mathbb{N}$ is injective. By Schröder-Bernstein Theorem it is enough to construct an injective map $f : \mathbb{N} \to D$. By well-ordering, D has a smallest element, d_1. Set $f(1) = d_1$. Since D is infinite, $D \setminus \{f(1)\}$ is nonempty. Consequently the set $D \setminus \{f(1)\}$ has a smallest element. Set

$$f(2) = \min(D \setminus \{f(1)\}.$$

More generally for any $k \in \mathbb{N}$ define $f(k)$ recursively by

$$f(k) = \min(D \setminus \{f(1), f(2), \ldots, f(k-1)\}).$$

By construction,

$$f(1) < f(2) < \cdots < f(k).$$

Hence for any $k, l \in \mathbb{N}$ with $l < k$

$$f(l) < f(k).$$

It follows that the map f is injective and we are done. (One can also prove that the map f constructed above is in fact a bijection.) □

Definition 12.5. A set A is countably infinite if there is a bijection $f : \mathbb{N} \to A$. A set A is countable if it is finite or countably infinite. The set A is uncountable if it is *not* countable.

For example the power set $\mathcal{P}(\mathbb{N})$, the set of all subsets of the set of natural numbers is uncountable (why?).

Lecture 13: Countable and uncountable sets

Last time :

- Discussed the axiom of choice. Proved (Theorem 12.2) that the axiom of choice is equivalent to:

  "Any surjective map $g : B \to A$ has a right inverse. That is, if $g : B \to A$ is surjective then there is a map $f : A \to B$ with $g \circ f = id_A$."
- Defined a set A to be countable if either A is finite or if there is a bijection $h : \mathbb{N} \to A$.
- Proved (Lemma 12.4) that if $D \subset \mathbb{N}$ is infinite then there is a bijection $h : D \to \mathbb{N}$.

The following lemma is very useful.

Lemma 13.1. *A nonempty set A is countable if and only if there is a surjective map $\mathbb{N} \to A$.*

Proof. ($\Longrightarrow$) Suppose a set A is countable. If A is finite, there exists a bijection $f : \{1, \ldots, n\} \to A$ for some $n \in \mathbb{N}$. Define $g : \mathbb{N} \to A$ by

$$g(i) = \begin{cases} f(i) & \text{if } 1 \leq i \leq n \\ f(1) & \text{if } i > n \end{cases}.$$

If A is infinite, there is a bijection $g : \mathbb{N} \to A$, which is onto.

($\Longleftarrow$). Suppose $g : \mathbb{N} \to A$ is onto. If the set A is finite, we're done. Suppose A is infinite. We need to construct a bijective map $\mathbb{N} \to A$, or, equivalently a bijective map $A \to \mathbb{N}$.

By Theorem 12.2, we have an injection $f : A \to \mathbb{N}$. Let $D = f(A)$. Then $f : A \to D$ is a bijection. By Lemma 12.4, we have a bijection $h : D \to \mathbb{N}$. The map $h \circ f : A \to \mathbb{N}$ is the desired bijection. □

Corollary 13.2. *If A is countable and $h : A \to B$ is onto, then B is countable.*

Proof. By Lemma 13.1 there is a surjective map $f : \mathbb{N} \to A$. The composite $h \circ f : \mathbb{N} \to B$ is surjective since it is the composite of two surjective maps. By Lemma 13.1 the existence of a surjective map from $\mathbb{N}$ to B implies that B is countable. □

Uncountable sets are easy to come up with.

Example 13.3. By Cantor's theorem there is no surjective map from the set $\mathbb{N}$ of natural numbers to its power set $\mathcal{P}(\mathbb{N})$. Therefore $\mathcal{P}(\mathbb{N})$ is uncountable.

Theorem 13.4 (also due to Cantor)**.** *The set of real numbers $\mathbb{R}$ is uncountable.*

Remark 13.5. A positive real number x has a unique decimal expansion

$$x = a_n a_{n-1} \dots a_1 . a_{-1} a_{-2} \dots \qquad a_i \in \{0, \dots, 9\}$$

as long as as we don't allow the expansion to end in an infinite sequence of 9's. Recall why this is the case. First of all for any $q \in \mathbb{R}$ with $|q| < 1$,

$$\sum_{n=1}^{\infty} q^n = \frac{q}{1-q}.$$

Therefore

$$0.999999\dots = 9 \cdot \sum_{n=1}^{\infty} \left(\frac{1}{10}\right)^n = 9 \cdot \frac{\frac{1}{10}}{1 - \frac{1}{10}} = 1.$$

Consequently $a_n \dots a_1 . a_{-1} a_{-2} \dots a_{-k} 9999999 \dots = a_n \dots a_1 . a_{-1} a_{-2} \dots a_{-k} 000000 \dots + 10^{-k}$.

Proof. We will argue that the set $(0, \infty)$ of positive real numbers is uncountable. This is enough since the exponential map

$$\exp : \mathbb{R} \to (0, \infty), \qquad \exp(x) = e^x$$

is a bijection.

Suppose the set $(0, \infty)$ of positive real numbers is countable. That is, suppose there is a surjective function $x : \mathbb{N} \to (0, \infty)$, $n \mapsto x_n$. Consider the decimal expansions of x_n's:

$$\begin{aligned} x_1 &= a_{n_1}^{(1)} \dots a_1^{(1)} . a_{-1}^{(1)} a_{-2}^{(1)} \dots \\ x_2 &= a_{n_2}^{(2)} \dots a_1^{(2)} . a_{-1}^{(2)} a_{-2}^{(2)} \dots \\ &\vdots \end{aligned}$$

We produce a positive real number r that's not on the list above as follows. Choose digits $r_{-1}, r_{-2}, \dots, r_{-n}, \dots$, i.e., elements of the set $\{1, \dots, 9\}$, such that

$$\begin{aligned} r_{-1} &\neq a_{-1}^{(1)} \\ r_{-2} &\neq a_{-2}^{(2)} \\ &\vdots \\ r_{-n} &\neq a_{-n}^{(n)} \\ &\vdots \end{aligned}$$

(and such that we don't end with an infinite sequence of 9s). Let $r = 0.r_{-1} r_{-2} \dots r_{-n} \dots$. For example if our list starts with

$$\begin{aligned} a_1 &= 102.\underline{1}30\dots \\ a_2 &= 10,029.1\underline{4}60\dots \\ a_3 &= 0.13\underline{0}0\dots \\ &\vdots \end{aligned}$$

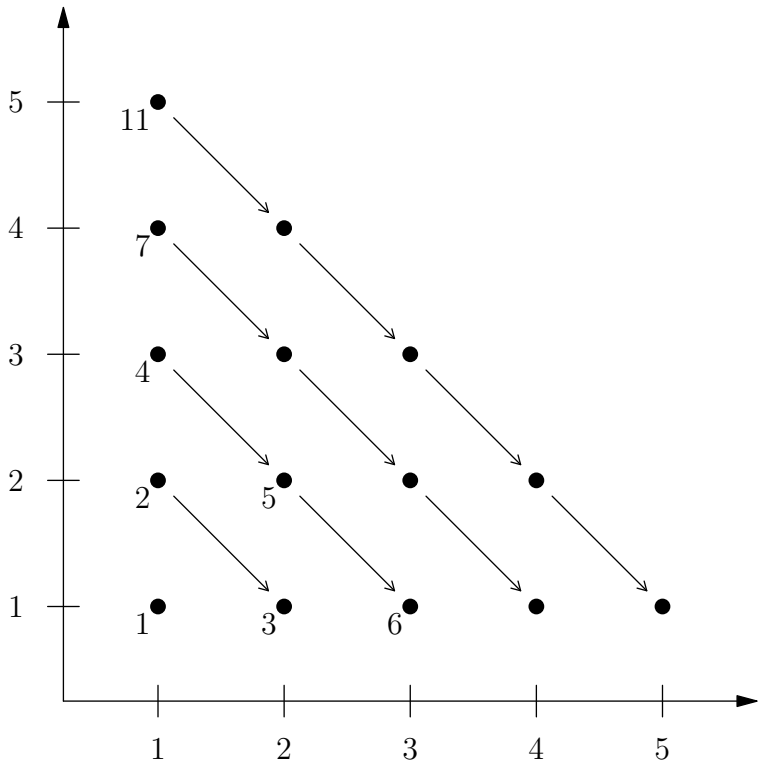
we may choose $r = 0.251\dots$.

Note that $r \neq x_n$ for any n. This is because $r_{-n} \neq a_{-n}^{(n)} = n^{\text{th}}$ digit of x_n past the decimal point. Contradiction!. Hence there is no surjective map $\mathbb{N} \to (0, \infty)$. It follows that the set $\mathbb{R}$ of real numbers is uncountable. $\square$

Remark 13.6. We will prove that the set $\mathbb{Q}$ of rational numbers is countable. We first need to build some tools.

Lemma 13.7. *The product* $\mathbb{N} \times \mathbb{N}$ *is countable.*

Sketch of proof. The easiest thing to do is to draw a picture:



A counting of elements of $\mathbb{N} \times \mathbb{N}$.

But pictures may be deceiving, so here is a formular for a map $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$:

$$f(k, n-k+1) = 1 + 2 + \cdots + (n-1) + k$$

for all $n \in \mathbb{N}$, and all $k \in \mathbb{N}$ with $1 \leq k \leq n$.
Then
$f(1,1) = f(1, 1-1+1) = (1-1)+1 = 1$
$f(1,2) = f(1, 2-1+1) = (2-1)+1 = 2$
$f(2,1) = f(2, 2-2+1) = (2-1)+2 = 3$
$f(1,3) = f(1, 3-1+1) = 1+(3-1)+1 = 4$
$\vdots$
Check that f is a bijection. □

Lemma 13.8. *If two sets* A *and* B *are countable then so is their product* $A \times B$.

Proof. Since A and B are countable, we have surjections

$$f : \mathbb{N} \to A, \quad g : \mathbb{N} \to B.$$

Then

$$(f, g) : \mathbb{N} \times \mathbb{N} \to A \times B, \qquad (f, g)(i, j) = (f(i), g(j))$$

is onto (check this!). Since $\mathbb{N} \times \mathbb{N}$ is countable, the product $A \times B$ is countable by Corollary 13.2. □

Lemma 13.9. *Let $\{A_n\}_{n\in\mathbb{N}}$ be a countable family of countable sets. Then the union $\bigcup_{n\in\mathbb{N}} A_n$ is countable.*

Proof. It is no loss of generality to assume that all sets A_n are nonempty. Since each A_n is countable (and nonempty) we have a surjection $f_n : \mathbb{N} \longrightarrow A_n$. Now define $\varphi : \mathbb{N} \times \mathbb{N} \longrightarrow \cup_{n\in\mathbb{N}} A_n$ as follows:

$$\varphi(k, n) := f_n(k).$$

I claim that the map φ is onto. Let's check that. If $a \in \bigcup_{n\in\mathbb{N}} A_n$, $a \in A_n$ for some n. Since $f_n : \mathbb{N} \to A_n$ is onto there is $k \in \mathbb{N}$ such that $a = f_n(k)$. Hence $a = \varphi(k, n)$. Since $\mathbb{N} \times \mathbb{N}$ is countable, the union $\bigcup_{n\in\mathbb{N}} A_n$ is countable by Corollary 13.2. □

Exercise 13.10.

(1) Prove that the union of three countable sets is countable.
(2) Prove that the set $\mathbb{Z}$ of natural numbers is countable. Hint: $\mathbb{Z}$ is the union of the sets $\mathbb{N}$, $\{0\}$, $-\mathbb{N}$.
(3) Prove that the set $\mathbb{Q}$ of rational numbers is countable. Hint: prove that the set $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ is countable.

Lecture 14: Least upper bounds in ordered rings

Recall the definition of a commutative ring with 1 from lecture 3 (Definition 3.2).

Definition 14.1. A *commutative ring* R with unity consists of the following data:

- a set R,
- two maps $+ : R \times R \longrightarrow R$ ("addition"), $\cdot : R \times R \longrightarrow R$ ("multiplication"),
- two special elements $0_R = 0, 1_R = 1 \in R$, $0 \neq 1$ such that

(1) the addition $+$ is commutative and associative
(2) the multiplication $\cdot$ is commutative and associative
(3) distributive law holds: $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ for all $a, b, c \in R$.
(4) 0 is an additive identity.
(5) 1 is a multiplicative identity
(6) for all $a \in R$ there exists $b \in R$ such that $a + b = 0$ (we call this b the additive inverse of a and denote it by $-a$).

Example 14.2. The integers $\mathbb{Z}$, the rationals $\mathbb{Q}$, the reals $\mathbb{R}$, and $\mathbb{Z}_n$ (the integers modulo n) are all commutative rings with 1. The set $M_2(\mathbb{R})$ of 2×2 matrices with the usual matrix addition and multiplication is a *noncommutative* ring with 1 (where "1" is the identity matrix). The set $2\mathbb{Z}$ of even integers with the usual addition and multiplication is a commutative ring without 1.

Next we introduce the notion of an integral domain. In fact we have seen integral domains in lecture 3, where they were called "commutative rings without zero divisors" (see Lemma 3.4).

Definition 14.3. A commutative ring R (with 1) is an integral domain if

$$a \cdot b = 0 \implies a = 0 \text{ or } b = 0$$

for all $a, b \in R$. (This is equivalent to: $a \cdot b = a \cdot c$ and $a \neq 0 \implies b = c$.)

Example 14.4. The rings $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are integral domains, but $\mathbb{Z}_6$ is not. For example in $\mathbb{Z}_6$

$$[2] \cdot [3] = [6] = [0]$$

but $[2], [3] \neq [0]$.

Next recall the definition of a field.

Definition 14.5. A commutative ring R (with 1) is a field if every nonzero element of R has a multiplicative inverse: for any $a \in R$, $a \neq 0$, there exists $b \in R$ such that $a \cdot b = 1$.

Lemma 14.6. *Any field is an integral domain.*

Proof. Suppose R is a field, $a, b \in R$ and $a \cdot b = 0$. If $a = 0$, we are done. If $a \neq 0$, then there exists $a^{-1} \in R$ such that $a \cdot a^{-1} = 1$. Therefore

$$a^{-1} \cdot a \cdot b = a^{-1} \cdot 0.$$

But $a^{-1} \cdot a \cdot b = 1 \cdot b = b$. Hence $b = 0$. □

Next recall that there are two equivalent ways to defined an ordered (commutative) ring.

Definition 14.7. An *ordered commutative ring* is a pair (R, P) where R is a commutative ring (with 1), and P a positive cone, that is, a subset of $R \setminus \{0\}$ such that

(1) for all $a, b \in P$, $a + b, a \cdot b \in P$.
(2) for all $a \in R \setminus \{0\}$, either $a \in P$ or $(-a) \in P$ (but not both! i.e., an exclusive "or"). We write $R \setminus \{0\} = P \sqcup (-P)$.

We refer to the elements of the positive cone P as positive elements.

Definition 14.8. An *ordered ring* is a pair $(R, <)$ where R is a commutative ring (with 1) and $<$ isa relation $<$ on R so that the following four conditions hold:

(O1) (trichotomy): For any $a, b \in R$, one and only one of the following holds:

$$a < b, \quad a = b, \quad b < a.$$

(O2) (transitivity): for all $a, b, c \in R$,

$$(a < b \text{ and } b < c) \implies a < c.$$

(O3) (addition) for all $a, b, c \in R$

$$a < b \implies a + c < b + c.$$

(O4) for all $a, b, c \in R$,

$$(a < b \text{ and } 0 < c) \implies a \cdot c < b \cdot c.$$

In Homework 2 (problem 1.5.9 of Sally) you proved that the two definitions are equivalent in the following sense (strictly speaking the problem was to prove it for the ring of integers $\mathbb{Z}$ but the proof only used the fact that $\mathbb{Z}$ is a commutative ring).

Lemma 4.3. *The two definitions are equivalent:*

(1) *Given a ring R with a positive cone P the relation $<$ on R defined by*

$$a < b \quad \Leftrightarrow b - a \in P$$

is an order (in the sense of Definition 14.8) .

(2) *Given a ring R with an order $<$ the set*

$$P := \{a \in R \mid 0 < a\}$$

is a positive cone (in the sense of Definition 14.7).

Remark 14.9. It will be important for us later (when, for example, we argue hat the field $\mathbb{C}$ of complex numbers cannot be ordered) that in an ordered commutative ring (R, P) the multiplicative identity 1_R has to be positive, i.e., $1_R \in P$. The reason is simple. We proved that in an ordered ring (R, P) the squares of non-zero elements are positive, and the multiplicative identity $1_R \in R$ is the square $1_R \cdot 1_R$.

Definition 14.10 (bounded above, bounded below)**.** Let $(R, <)$ be an ordered ring, and A a nonempty subset of R.

The set A is bounded above if and only if there exists $M \in R$ with $a \leq M$ for all $a \in A$.

The set A is bounded below if and only if there exists $m \in R$ with $m \leq a$ for all $a \in A$.

Example 14.11. Let R be the rationals $\mathbb{Q}$ with the usual order. and consider the subset

$$A = \{x \in \mathbb{Q} \mid x^2 \leq 2\}.$$

The set A is bounded above by 3 and below by -3:

$$x^2 \leq 2 \implies x^2 < 9 \implies -3 < x < 3.$$

We now come to an important definition.

Definition 14.12. Let R be an ordered ring and $A \subseteq R$ a subset bounded above. An element $L \in R$ is a *least upper bound* of A if and only if

(1) L is an upper bound of A.
(2) L is a *least* upper bound: if M is any upper bound of A then $L \leq M$.

Least upper bounds may or may not exists. For example the set $A = \{x \in \mathbb{Q} \mid x^2 \leq 2\}$ has no least upper bound (in $\mathbb{Q}$!). Morally the argument is that the least upper bound of A should be $\sqrt{2}$ but $\sqrt{2}$ is not an element of the rationals $\mathbb{Q}$.

Exercise 14.13. Prove that $\sqrt{2}$ is not rational. Hints.
Prove first: if $a, b \in \mathbb{N}$ and the product ab is even, then either a or b is even.

Next suppose $\sqrt{2}$ is rational: $\sqrt{2} = p/q$ for some pair of natural numbers p, q. It is no loss of generality that p and q have no common factors (otherwise cancel the factor). Then $2q^2 = p^2$. Hence (why?) p is even. What does it tell you about q?

Lemma 14.14. *Least upper bounds of a given set are unique (provided they exist).*

Proof. Suppose L_1, L_2 are two least upper bound of a subset A of an ordered ring R. Then $L_1 \leq L_2$ since L_1 is a least upper bound. Similarly $L_2 \leq L_1$ since L_2 is least. Hence $L_1 = L_2$. □

Notation 14.15. Let A be a subset of an ordered ring R which is bounded above. Suppose the least upper bound of A exists. Then we denote the least upper bound by $\sup A$ (Sally prefers lub A.)

Definition 14.16. An ordered commutative ring R has the least upper bound property iff every nonempty subset of R that is bounded above has a least upper bound (which is necessarily unique).

Sally defines real numbers as follows.

Definition 14.17. The *real numbers* are an ordered field with the least upper bound property: every nonempty subset that is bounded above has the least upper bound.

The definition begs two question: (1) do real numbers exist? (2) are they unique in some sense? To answer the second question we need to be able to say precisely what we mean by two ordered fields "being the same." Answering the first question will take several lectures — we will construct an ordered field in which every nonempty subset has the least upper bound out of the rationals. What we are going to do next is to *assume* that real numbers exist and study some of their properties. This will motivate a number of ideas that we will need in order to construct them.

Lecture 15: More on least upper bounds

Last time: reviewed rings, ordered rings, fields. Defined sets in ordered rings bounded above and below. Defined least upper bounds and proved that they are unique (if they exist). The set $\{q \in \mathbb{Q} \mid q^2 \leq 2\}$ has upper and lower bounds but no least upper bounds.

We defined real numbers as an ordered fields with a least upper bound property: any set bounded above has a least upper bound.

Notation 15.1. $\sup A =$ the least upper bound of a set A bounded above (if it exists).

Sally denotes the least upper bound of a set A by lub$A$.

Definition 15.2. Let R be an ordered ring, $\varnothing \neq A \subseteq R$ a subset bounded below. An element $l \in R$ is the *greatest lower bound* of the set A if

(1) l is a lower bound of A: $l \leq x$ for all $x \in A$ and
(2) if m is a lower bound of A then $m \leq l$.

Again it is easy to see that greatest lower bounds are unique (if they exist).

Notation 15.3. $\inf A =$ greatest lower bound of a set A. (The ordered ring itself is not explicitly mentioned in this notation.)

Notation 15.4. Let A be a subset of a ring R. We define

$$-A := \{-x \mid x \in A\}.$$

Exercise 15.5. let A be a subset of an ordered ring R which is bounded below. If $\sup(-A)$ exists then $\inf A$ exists and

$$\inf A = -\sup(-A)$$

Consequently any nonempty subset of the reals bounded below has the greatest lower bound.

Remark 15.6. For the next few lectures we will assume that $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$. Given the way we constructed the rationals out of integers, the ring $\mathbb{Z}$ of integers is not literally a subring of the ring $\mathbb{Q}$ of the rationals. However there is a bijection $\mathbb{Z} \longleftrightarrow \{\frac{n}{1} \in \mathbb{Q} \mid n \in \mathbb{Z}\}$ so there is no harm in assuming that $\mathbb{Z} \subset \mathbb{Q}$ (and keeps notation simpler).

Later in the course we will prove that if F is any ordered field then F contains a copy of $\mathbb{Z}$ and, more generally, a copy of $\mathbb{Q}$.

Theorem 15.7 (Archimedian property of $\mathbb{R}$)**.** *For all real numbers $a, b \in \mathbb{R}$ with $a, b > 0$ there exists $n \in \mathbb{N}$ such that $b < na$.*

In particular for any $c \in \mathbb{R}$ there is $n \in \mathbb{N}$ with $c < n$.

Proof. If $b \leq a$, then $b < a + a = 2a$ (since $0 < a$), so take $n = 2$.

Suppose now that $a < b$ and there exists no $n \in \mathbb{N}$ so that $b < na$. Then $na \leq b$ for all $n \in \mathbb{N}$. Hence the set

$$S = \{na \mid n \in N\}$$

is bounded above by b. By the least upper bound property of the reals $\mathbb{R}$ there exists the least upper bound L of S. Since $a > 0$, $L - a < L$. Hence $L - a$ is *not* an upper bound of S. Therefore there exists $n_0 \in \mathbb{N}$ such that $L - a < n_0 a$. But then

$$L < n_0 a + a = (n_0 + 1)a$$

and $(n_0 + 1)a \in S$ (since $n_0 + 1 \in \mathbb{N}$). This contradicts that L is an upper bound of S. We conclude that there exists $n \in \mathbb{N}$ such that $b < na$.

Now suppose $c \in \mathbb{R}$. If $c \leq 0$ then $c < 1 \in \mathbb{N}$. If $c > 0$, let $b = c$ and $a = 1$. By the first part of the theorem there is $n \in \mathbb{N}$ such that $c = b < a \cdot n = 1 \cdot n$. □

We note a corollary which will be useful later when we talk about convergence of sequences. It says, in effect, that the sequence $\{\frac{1}{n}\}$ converges to 0. We will say more about convergence of sequences later.

Corollary 15.8. *For any positive number ε there exists $n \in \mathbb{N}$ such that $\frac{1}{n} < \varepsilon$.*

Proof. By Theorem 15.7 there exists $n \in \mathbb{N}$ such that $1 < n\varepsilon$. □

Recall that for the time being we are assuming without proof that any ordered field F contains (a copy of) the integers $\mathbb{Z}$ (Remark 15.6).

Definition 15.9. An ordered field F is an *Archimedian ordered field* if for all $x \in F$, there exists $n \in \mathbb{Z}$ such that $x < n$.

Example 15.10. Theorem 15.7 implies that the field $\mathbb{R}$ of the reals with its usual order is an Archimedian ordered field.

Remark 15.11. Non-Archemidian ordered fields do exist. Non-Archemidian ordered *rings* are even easier to come by. Recall the ring $\mathbb{Z}[x]$ of polynomials with integer coefficients from Example 4.5:

$$\mathbb{Z}[x] = \{a_0 + a_1x + \cdots + a_nx^n \mid n \geq 0, a_0, \ldots, a_n \in \mathbb{Z}\}.$$

It is a commutative ring under the usual operations of addition and multiplication of polynomials. We defined the positive cone P in $\mathbb{Z}[x]$ to be

$$P := \{a_0 + a_1x + \cdots + a_nx^n \mid a_n > 0\}$$

the set of polynomials with *positive* highest order term. In this ring the polynomail $x - n$ is positive for any natural number n. Hence in $\mathbb{Z}[x]$

$$n < x$$

for all $n \in \mathbb{N}$.

Our next goal is to prove that the rationals are *dense* inside the set $\mathbb{R}$ of real numbers. "Dense" means that given a real number x and any $\varepsilon > 0$ there is a rational number $q \in (x - \varepsilon, x + \varepsilon)$. More generally "dense" means the following.

Definition 15.12. A set $A \subseteq \mathbb{R}$ is dense in $\mathbb{R}$ if for every $x \in \mathbb{R}$ and for every $\varepsilon > 0$

$$(x - \varepsilon, x + \varepsilon) \cap A \neq \varnothing.$$

We prove the density of $\mathbb{Q}$ in $\mathbb{R}$ in several steps.

Theorem 15.13. *For any real number $a \in \mathbb{R}$ there exists $N \in \mathbb{Z}$ such that*

$$N - 1 \leq a < N.$$

Proof. By the Archemidean property of the reals there exists $n \in \mathbb{Z}$ such that $a < n$. Hence the set $S = \{n \in \mathbb{Z} \mid n > a\}$ is nonempty.

We now argue that S has the least element. Note that S may contain zero or negative integers, so we can't apply the Well-Ordering principle right away. We proceed as follows. By the Archemidean property of $\mathbb{R}$ there exists $M \in \mathbb{Z}$ such that $(-a) < M$. Consequently $a > (-M)$.

Now for any $n \in S$, $n > a > -M$. Hence $n + M > 0$. Therefore the set

$$S' = \{n + M \mid n \in S\}$$

is a nonempty subset of natural numbers. By the Well-Ordering Principle, S' has least element. This least element has to be of the form $N + M$ for some $N \in \mathbb{Z}$. By construction of S' N is the least element of S. Since $N - 1 < N$, $N - 1$ is not in S. Therefore $N - 1 \leq a$. On the other hand, since $N \in S$, $a < N$.

We conclude that there exists $N \in \mathbb{Z}$ such that

$$N - 1 \leq a < N.$$

□

Lecture 16: Density of rationals, limits of sequences

Last time:
Proved the Archemidian property of $\mathbb{R}$ — for all real numbers $a, b \in \mathbb{R}$ with $a, b > 0$ there exists $n \in \mathbb{N}$ such that $b < na$. In particular for any $c \in \mathbb{R}$ there is $n \in \mathbb{N}$ with $c < n$.

We also proved: for any $a \in \mathbb{R}$ there is $N \in \mathbb{Z}$ so that $N - 1 \leq a < N$.

Our first result of the day:

Theorem 16.1. *For all $a, b \in \mathbb{R}$ with $a < b$, there exists $r \in \mathbb{Q}$ such that $a < r < b$. Consequently $\mathbb{Q}$ is dense in $\mathbb{R}$.*

Proof. By Corollary 15.8, there exists $q \in \mathbb{N}$ such that $\frac{1}{q} < b - a$. Hence

$$a + \frac{1}{q} < b. \tag{16.3}$$

By Theorem 15.13, there exists $p \in \mathbb{Z}$ such that $p - 1 \leq q \cdot a < p$. Since $\frac{1}{q} > 0$, this implies that

$$\frac{p}{q} - \frac{1}{q} = \frac{p-1}{q} \leq a < \frac{p}{q}.$$

In particular

$$a < \frac{p}{q} \quad \text{and} \quad \frac{p}{q} \leq a + \frac{1}{q}. \tag{16.4}$$

$$a < \frac{p}{q} \overset{(16.4)}{\leq} a + \frac{1}{q} \overset{(16.3)}{<} b.$$

Now let $r = \frac{p}{q}$. Then $a < r < b$. □

Corollary 16.2. *For all $a, b \in \mathbb{R}$ with $a < b$, there exists $x \in \mathbb{R} \setminus \mathbb{Q}$ with $a < x < b$.*

Proof. Recall that $\sqrt{2}$ is not rational. That is, $\sqrt{2} \in \mathbb{R} \setminus \mathbb{Q}$. Hence for any rational number $r \neq 0$ the product $\sqrt{2}r$ could not be rational either. For if $\sqrt{2}r = s \in \mathbb{Q}$, then $\sqrt{2} = sr^{-1} \in \mathbb{Q}$, which is not true.

Now, $a < b$ implies that $\frac{a}{\sqrt{2}} < \frac{b}{\sqrt{2}}$ (since $\sqrt{2}$ is positive). By Theorem 16.1, there exists $r \in \mathbb{Q}$ with $\frac{a}{\sqrt{2}} < r < \frac{b}{\sqrt{2}}$. If $r = 0$, there exists $r' \in \mathbb{Q}$ such that $\frac{a}{\sqrt{2}} < r' < 0$ (Theorem 16.1 again).

Therefor we may assume that $r \neq 0$. Then $a < \sqrt{2}r < b$ **and** $\sqrt{2}r \notin \mathbb{Q}$. Let $x = \sqrt{2}r$. □

Absolute Value.

Definition 16.3. For $a \in \mathbb{R}$ we define the absolute value of a to be

$$|a| = \begin{cases} a & a \geq 0 \\ -a & a < 0 \end{cases}.$$

We think of the absolute value $|a|$ of $a \in \mathbb{R}$ as the distance from 0 to a. Note that $|a| = |-a|$. More generally we define:

Definition 16.4. For $a, b \in \mathbb{R}$, the distance from a to b is

$$d(a,b) := |a-b| (= |b-a|).$$

Theorem 16.5. (1) *For any $a \in \mathbb{R}$, $|a| \geq 0$. Moreover $|a| = 0$ if and only if $a = 0$.*
(2) *For any $a, b \in \mathbb{R}$, $|ab| = |a| \cdot |b|$.*
(3) *(Triangle Inequality) For any $a, b \in \mathbb{R}$, $|a+b| \leq |a| + |b|$.*
(4) *$|a| \leq x \iff -x \leq a \leq x$ for all $a, x \in \mathbb{R}$.*

Proof. (1) holds by definition of $|a|$.
(2) follows by considering cases:

- If a or b is zero, (2) holds.
- If $a, b > 0$, (2) holds as well.
- If $a, b < 0$, then $ab > 0$ hence we have $|ab| = ab = (-a) \cdot (-b) = |a||b|$.
- The case $a > 0$, $b < 0$ is similar.

(4) ($\Longrightarrow$) If $|a| \leq x$ then $x \geq 0$ (hence $-x \leq 0$). If $a \geq 0$, then $-x \leq 0 \leq a = |a| \leq x$.
If $a \leq 0$, then $x \geq |a| = -a > 0$. Hence $-x \leq a \leq 0 \leq x$.
($\Longleftarrow$) Suppose $-x \leq a \leq x$. Then, since $x \geq -x$, $2x \geq 0$ and therefore $0 \leq x$.
If $a > 0$, $a = |a| \leq x$.
If $a < 0$, $|a| = -a$. Now $-x \leq a$ implies that $x \geq -a = |a|$.
(3) $-|a| \leq a \leq |a|$, $-|b| \leq b \leq |b|$. Therefore

$$-(|a| + |b| \leq a + b \leq |a| + |b|.$$

It follows from (4) that

$$|a+b| \leq |a| + |b|.$$

□

Corollary 16.6 (Triangle inequality for distances)**.** $\forall a, b, c \in \mathbb{R}$,

$$d(a,c) \leq d(a,b) + d(b,c)$$

Proof.

$$\begin{aligned} d(a,c) &= |a-c| \\ &= |(a-b)+(b-c)| \\ &\leq |a-b| + |b-c| \qquad \text{(by Theorem 16.5 (3))} \\ &= d(a,b) + d(a,c) \end{aligned}$$

□

Sequences.

Definition 16.7. A sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers is a function $\mathbb{N} \to \mathbb{R}$, $n \mapsto a_n$.

You should memorize the following definition.

Definition 16.8. A sequence $\{a_n\}_{n=1}^{\infty}$ converges to $L \in \mathbb{R}$ if for any $\varepsilon > 0$ there exists $N \in \mathbb{R}$ such that if $n \geq N$ then $|a_n - L| < \varepsilon$.

Example 16.9. The sequence $a_n = \frac{1}{n}$ converges to 0.

Proof. By Corollary 15.8 for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $\frac{1}{N} < \varepsilon$.
Then for any integer $n \geq N$ we have $\varepsilon > \frac{1}{N} \geq \frac{1}{n} = |\frac{1}{n} - 0|$. □

Notation 16.10. If a sequence $\{a_n\}$ converges to a number L we write

$$\lim_{n\to\infty} a_n = L \qquad \text{or, equivalently,} \qquad a_n \xrightarrow[n\to\infty]{} L.$$

Example 16.11. $\lim_{n\to\infty} \frac{1}{n^2} = 0$.

Proof. Given $\varepsilon > 0$ we need to find $N > 0$ such that if $n \geq N$ then $|\frac{1}{n^2} - 0| < \varepsilon$.

$$\begin{aligned} \left|\frac{1}{n^2} - 0\right| < \varepsilon &\iff \frac{1}{n^2} < \varepsilon \\ &\iff \frac{1}{n} < \sqrt{\varepsilon} \\ &\iff n > \frac{1}{\sqrt{\varepsilon}} \end{aligned}$$

Take $N = \frac{1}{\sqrt{\varepsilon}} + 1$. If $n \geq N$ then $n \geq \frac{1}{\sqrt{\varepsilon}} + 1 > \frac{1}{\sqrt{\varepsilon}}$. And then $n^2 > \frac{1}{\varepsilon}$. Consequently if $n > N$

$$\varepsilon > \frac{1}{n^2} = \left|\frac{1}{n^2} - 0\right|.$$

We conclude that if $n \geq \frac{1}{\sqrt{\varepsilon}} + 1$ then $\left|\frac{1}{n^2} - 0\right| < \varepsilon$. That is, $\lim_{n\to\infty} \frac{1}{n^2} = 0$ as we wanted to show. □

Lemma 16.12. *Suppose $x \in \mathbb{R}$ is a real number such that $x \geq 0$ and $x < \varepsilon$ for all $\varepsilon > 0$. Then $x = 0$.*

Proof. Suppose not. Then x is strictly positive: $x > 0$. Consider $\varepsilon_0 = x/2$. Then $0 < \varepsilon_0 < x$, which contradicts the assumption that $x < \varepsilon$ for all $\varepsilon > 0$. □

Lecture 17: Cauchy sequences

Last time:

• Defined absolute value $|a|$ of a real number a. Note that this definition makes sense in any ordered field.

• Defined distance between two real numbers $d(a, b) = |a - b|$ and prove the triangle inequality.

• Defined sequences and what it means for a sequence to converge:

> A sequence $\{a_n\}_{n=1}^{\infty}$ of real numbers converges to $L \in \mathbb{R}$ if for any $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, there exists $N \in \mathbb{R}$ such that if $n \geq N$ then $|a_n - L| < \varepsilon$.

(The same definition makes sense in any ordered field F.)

• A sequence $\{a_n\}_{n=1}^{\infty}$ converges if it converges to some $L \in \mathbb{R}$.

We should have proved last time (but I forgot to do it):

Lemma 17.1. *Suppose $x \in \mathbb{R}$ is a real number such that $x \geq 0$ and $x < \varepsilon$ for all $\varepsilon > 0$. Then $x = 0$.*

Proof. Suppose not. Then x is strictly positive: $x > 0$. Consider $\varepsilon_0 = x/2$. Then $0 < \varepsilon_0 < x$, which contradicts the assumption that $x < \varepsilon$ for all $\varepsilon > 0$. □

Lemma 17.2. *Limits are unique: if $a_n \to L_1$ and $a_n \to L_2$ then $L_1 = L_2$.*

Proof. We argue that $|L_1 - L_2| < \varepsilon$ for any $\varepsilon > 0$. Note that this implies what we want to show: $|L_1 - L_2| = 0$, hence $L_1 - L_2 = 0$.

Fix $\varepsilon > 0$. Since $a_n \to L_1$, there exists N_1 such that for $n \geq N_1$ we have $|a_n - L_1| < \frac{\varepsilon}{2}$. Since $a_n \to L_2$, there exists N_2 such that for $n \geq N_2$ we have $|a_n - L_2| < \frac{\varepsilon}{2}$. If $n \geq N = \max\{N_1, N_2\}$ then

$$|L_1 - L_2| = |L_1 - a_n + a_n - L_2 \leq |L_1 - a_n| + |a_n - L_2| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

By Lemma 17.1 we are done. □

Aside. What does it mean for a sequence $\{a_n\}$ **not** to converge to $L \in \mathbb{R}$?

It means that there is $\varepsilon > 0$ so that the statement "there exists $N \in \mathbb{R}$ such that if $n \geq N$ then $|a_n - L| < \varepsilon$" is false.

Equivalently there is $\varepsilon > 0$ so that for any $N \in \mathbb{R}$ the statement "if $n \geq N$ then $|a_n - L| < \varepsilon$" is false.

Equivalently there is $\varepsilon > 0$ so that for any $N \in \mathbb{R}$ there is $n \in \mathbb{N}$ (which depends on N) with $n \geq N$ and $|a_n - L| \geq \varepsilon$.

We conclude that a sequence a sequence $\{a_n\}$ **does not converge** if and only if

> for all $L \in \mathbb{R}$ there exists $\varepsilon = \varepsilon(L) \in \mathbb{R}$, $\varepsilon > 0$, so that for all $N \in \mathbb{R}$ there exists $n = n(N) \in \mathbb{N}$ with $n \geq N$ and $|a_n - L| \geq \varepsilon$.

Here are some examples of sequences that don't converge:

$a_n = n$, $\quad a_n = (-1)^n$, $\quad a_n = \sum_{k=1}^{n} \frac{1}{k}$.

Proving directly from the definition that these sequences don't converge requires work. Try it! There is an easier way.

Definition 17.3. A sequence $\{a_n\}$ is Cauchy if for all $\varepsilon > 0$, there exists N such that if $n, m \geq N$ then $|a_n - a_m| < \varepsilon$.

Exercise 17.4 (Sally, exercise 3.5.4)**.** If a sequence $\{a_n\}$ converges then $\{a_n\}$ is Cauchy. That is, being Cauchy is a necessary condition for a sequence to be convergent.

Solution. Let L denote the limit of the sequence $\{a_n\}$. Since $a_n \to L$, for all $\varepsilon > 0$, there exists N such that if $n \geq N$ then $|a_n - L| < \varepsilon/2$. Therefore, if $n, m \geq N$ then

$$|a_n - a_m| < |a_n - L + L - a_m| \leq |a_n - L| + |L - a_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

We conclude that $\{a_n\}$ is Cauchy. □

Example 17.5. $a_n = n$ is *not* Cauchy hence does not converge to any $L \in \mathbb{R}$:

Proof. Let $\varepsilon = 1/2$. Then for all $N \in \mathbb{R}$, if $n \geq N$, then $n + 1 > n \geq N$ as well but

$$|a_{n+1} - a_n| = |n + 1 - n| = 1 > 1/2 = \varepsilon.$$

□

Example 17.6. The sequence $a_n = \sum_{k=1}^{n} \frac{1}{k}$ does not converge (i.e., the series $\sum_{k=1}^{\infty} \frac{1}{k}$ does not converge).

Proof. For any natural number n

$$|a_{2n} - a_n| = \left|\sum_{k=1}^{2n} \frac{1}{k} - \sum_{k=1}^{n} \frac{1}{k}\right| = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{n+n}$$
$$\geq \underbrace{\frac{1}{2n} + \cdots + \frac{1}{2n}}_{n \text{ terms}} = \frac{n}{2n} = \frac{1}{2}.$$

Let $\varepsilon = 1/3$. Then for any $N \in \mathbb{R}$ and for any $n > N$

$$|a_{2n} - a_n| \geq \frac{1}{2} > \frac{1}{3} = \varepsilon.$$

Therefore the sequence $a_n = \sum_{k=1}^{n} \frac{1}{k}$ is not Cauchy hence does not converge. □

Question: suppose $\{a_n\}$ is a Cauchy sequence in an ordered field F (i.e., $a_n \in F$ for all n and for any $\varepsilon \in F$ with $\varepsilon > 0$, there exists $N \in F$ such that if $n, m \geq N$ then $|a_n - a_m| < \varepsilon$). Does the sequence converge?

Answer: yes if $F = \mathbb{R}$, the reals (this is Theorem 3.6.14 in Sally; we'll prove it). No if $F = \mathbb{Q}$. Here is a reason:

Lemma 17.7. *For any real number x there is a sequence $\{a_n\}$ so that $a_n \in \mathbb{Q}$ for all n and $a_n \to x$. In particular if $x \in \mathbb{R} \setminus \mathbb{Q}$ the sequence $\{a_n\}$ above does not converge to any rational number r (because $a_n \to x \notin \mathbb{Q}$ and limits are unique).*

Proof. Fix $x \in \mathbb{R}$. Since the rational are dense in $\mathbb{R}$ (Theorem 16.1) for any $n \in \mathbb{N}$ there is $a_n \in \mathbb{Q}$ with

$$x < a_n < x + \frac{1}{n}.$$

Then

$$0 < a_n - x < \frac{1}{n} \qquad \text{hence} \qquad |a_n - x| < \frac{1}{n}.$$

Now given $\varepsilon > 0$ we can choose $N \in \mathbb{N}$ with $\frac{1}{N} < \varepsilon$ (Corollary 15.8). Then for any $n \geq N$ we have

$$\varepsilon > \frac{1}{N} \geq \frac{1}{n} > |a_n - x|.$$

We conclude that $a_n \to x$. □

Remark 17.8. Given $x \in \mathbb{R}$ there are many different sequences of rational numbers that converge to x. For example if $x = 0$ then both $\{1/n\}$ and $\{1/n^2\}$ converge to x and there are many more.

The key idea of section 3.5 in Sally's book is:

> The field $\mathbb{R}$ of real numbers is the set of equivalence classes of Cauchy sequence of rational numbers for an appropriate equivalence relation.

To understand what equivalence relation is "appropriate" we observe:

Lemma 17.9. *Suppose the two sequence $\{a_n\}$ and $\{b_n\}$ of real numbers both converge to $x \in \mathbb{R}$. Then $a_n - b_n \to 0$.*

Proof. We need to show: for all $\varepsilon > 0$, there exists $N \in \mathbb{R}$ such that if $n \geq N$ then

$$|(a_n - b_n) - 0| < \varepsilon.$$

Fix $\varepsilon > 0$. Since $a_n \to x$, there exists N_1 such that if $n \geq N_1$ then $|a_n - x| \leq \frac{\varepsilon}{2}$.

Since $b_n \to x$, there exists N_2 such that if $n \geq N_1$ then $|b_n - x| \leq \frac{\varepsilon}{2}$. Therefore for $n \geq N = \max(N_1, N_2)$,

$$\begin{aligned} |a_n - b_n| &= |a_n - x + x - b_n| \\ &\leq |a_n - x| + |x - b_n| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

□

Lemma 17.9 suggests that we should define a relation $\sim$ on the set $\mathcal{C}$ of all Cauchy sequences of rational numbers by

$$\{a_n\} \sim \{b_n\} \quad \iff \quad a_n - b_n \to 0.$$

Unfortunately it's far from clear that $\sim$ is an equivalence relation let alone that the quotient $\mathcal{C}/_\sim$, the set of equivalence classes of $\sim$, is an ordered field with the least upper bound property.

To construct the field $\mathbb{R}$ of real numbers as a set of equivalence classes of Cauchy sequences we'll need a number of ideas/notions from abstract algebra:

- homomorphisms of rings
- subrings and ideals
- quotient rings
- first isomorphism theorem.

Lecture 18: Subrings, homomorphisms and isomorphisms of rings

Last time:

- Defined Cauchy sequences: $\{a_n\}$ is Cauchy if for any $\varepsilon > 0$ there is N so that for all $n, m \geq N$ we have $|a_n - a_m| < \varepsilon$.
- Proved that if a sequence converges then it is Cauchy.
- Proved that for any $x \in \mathbb{R}$ there is a sequence $\{a_n\}$ of *rational* numbers that converges to x
- Consequently in $\mathbb{Q}$ there are Cauchy sequences with no limit in $\mathbb{Q}$
- We'll prove: *any* Cauchy sequence in $\mathbb{R}$ has a limit in $\mathbb{R}$ but we are far from doing that.
- We proved: if $\{a_n\}$, $\{b_n\}$ are two sequence in $\mathbb{R}$ with $\lim a_n = \lim b_n$ then $a_n - b_n \to 0$.
- We plan to construct the reals $\mathbb{R}$ as the quotient $\mathcal{C}/_\sim$ of the set $\mathcal{C}$ of all Cauchy sequences of rational numbers by the relation $\sim$ which is defined by

$$\{a_n\} \sim \{b_n\} \quad \iff \quad a_n - b_n \to 0.$$

Our first step is the following lemma:

Lemma 18.1. *Let X be a set and R a commutative ring (with 1 as usual). Then R^X, the set of all functions from X to R is a commutative ring with 1 with the ring operations defined pointwise. In particular for any commutative ring R the set of sequences $R^{\mathbb{N}}$ is a ring.*

Sketch of proof. To turn R^X into a ring, we define addition of functions pointwise:

$$(f + g)(x) := f(x) + g(x) \quad \text{for all } x \in X;$$

here on the right $f(x), g(x) \in R$ so we use the $+$ operation R to add $f(x)$ and $g(x)$. It's easy to check that the addition of functions defined above is commutative and associative. Similarly we define multiplication of functions from X to R pointwise:

$$(f \cdot g)(x) := f(x) \cdot g(x) \quad \text{for all } x \in X.$$

Again $\cdot$ is commutative, associative and distributes over addition. The zero in R^X is the zero function 0 defined by

$$0(x) = 0_R$$

for all $x \in X$ (here 0_R denotes the zero in the ring R). Similarly we have a function $1 : X \to R$ defined by

$$1(x) = 1_R$$

for all $x \in X$ (where 1_R is the one in the ring R). It's a fairly easy exercise at this point to check that the data $(R^X, +, \cdot, 0, 1)$ form a commutative ring with 1. □

Exercise 18.2. Check that the additive inverse of a function $f : X \to R$ in R^X is the function $-f$ defined by

$$(-f)(x) = -f(x)$$

for all $x \in X$.

Definition 18.3. A *nonempty* subset S of a ring R is a subring iff

(i) for all $s_1, s_2 \in S$ both $s_1 + s_2$ and $s_1 s_2$ are in S;
(ii) $1_R, 0_R \in S$
(iii) $(S, +, \cdot, 0_R, 1_R)$ is a ring.

Remark 18.4. Since S is a ring, for any $s \in S$, $-s$ has to be in S as well.

Example 18.5. For any ring R and any set X the set of constant functions S is a subring of R^X.

Example 18.6. The set of 2×2 diagonal matrices $D = \left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \middle| a, b \in \mathbb{R} \right\}$ is a subring of the ring of 2×2 real matrices $M_{2,2}(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| a, b, c, d \in \mathbb{R} \right\}$.

Warning 18.7. $\mathbb{Z}^{\geq 0} := \{n \in Z \mid n \geq 0\}$ is a nonempty subset of $\mathbb{Z}$ satisfies the conditions (i) and (ii) but it's not a ring — it doesn't have additive inverses. For example there is no $n \in \mathbb{Z}^{\geq 0}$ so that $n + 2 = 0$. So conditions (i) and (ii) on a subset S of a ring R *by themselves* are not enough to guarantee that S is a subring.

It will be convenient to have the following lemma at our disposal.

Lemma 18.8. *Suppose R is a ring, $\varnothing \neq S \subset R$ a subset with $1_R \in S$ so that for all $s, t \in S$ we have*

$$st \in S \quad \text{and } s - t \in S.$$

Then S is a subring of R

Proof. Sketch of proof (i) Since $S \neq \varnothing$, there is $s \in S$. Then $0 = 0_R = s - s \in S$.

(ii) For any $s \in S$, $S \ni 0 - s = -s$.

(iii) For any $s, t \in S$, $-t \in S$ by (ii) above. Hence $S \ni s - (-t) = s + t$.

It is not hard to check that $(S, +, \cdot, 0_R, 1_R)$ is a ring and therefore S is a subring of R. □

Example 18.9. Consider the subset S of the ring $\mathbb{Q}$ of rational numbers defined by

$$S := \left\{ \frac{n}{1} \in \mathbb{Q} \;\middle|\; n \in \mathbb{Z} \right\},$$

Lemma 18.8 makes it easy to check that S is a subring of $\mathbb{Q}$: for any $n, m \in \mathbb{Z}$

$$\frac{n}{1} - \frac{m}{1} = \frac{n-m}{1} \in S$$
$$\frac{n}{1}\frac{m}{1} = \frac{nm}{1} \in S$$
$$\frac{0}{1}, \frac{1}{1} \in S.$$

Next we'd like to have the right language to express the fact that the subring S "is" the ring of integers $\mathbb{Z}$. We start with a definition.

Definition 18.10. Let R, S be two rings. A map $\varphi : R \to S$ is a homomorphism if it preserves $+$ and $\cdot$:

- $\varphi(a+b) = \varphi(a) + \varphi(b)$
- $\phi(a \cdot b) = \varphi(a) \cdot \varphi(b)$

for all $a, b \in R$.

Example 18.11. Fix $n \in \mathbb{N}$. The quotient map $\pi : \mathbb{Z} \to \mathbb{Z}_n$ is a ring homomorphism because

$$\pi(a+b) = [a+b] = [a] + [b] = \pi(a) + \pi(b) \qquad \text{and}$$
$$\pi(ab) = [ab] = [a][b] = \pi(a)\pi(b).$$

Example 18.12. Let X be a set and R a commutative ring. Then the map $R \to R^X$ defined by

$$R \ni r \mapsto \text{ the constant function } \varphi_r : X \to R, \quad \varphi_r(x) = r \text{ for all } x \in X$$

is a homomorphism (check it!).

Example 18.13. The map

$$f : \mathbb{Z} \to S = \left\{ \frac{n}{1} \in \mathbb{Q} \,\middle|\, n \in \mathbb{Z} \right\}, \quad f(n) = \frac{n}{1}$$

is a ring homomorphism.

Lemma 18.14. *Let K and L be two rings and $f : K \to L$ a ring homomorphism. Then*

(1) $f(0_K) = 0_L$ *and*
(2) $f(-a) = -(f(a))$ *for all $a \in K$.*

Proof. (1) $f(0_K) = f(0_K + 0_K) = f(0_K) + f(0_K)$. Now add $-f(0_K)$ to both sides. We get

$$0_L = f(0_K).$$

(2) For any $a \in K$

$$0_L = f(0_K) = f(a + (-a)) = f(a) + f(-a).$$

Add $-f(a)$ to both sides. We get

$$-f(a) = (-f(a)) + f(a) + f(-a) = 0_L + f(-a) = f(-a).$$

□

Warning 18.15. A ring homomorphism $f : K \to L$ need not send 1_K to 1_L. Here is an example.

Example 18.16. Let $K = \mathbb{R}$, the ring of real numbers. Let $L = \left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} a, b \in \mathbb{R} \right\}$, the ring of diagonal 2×2 matrices. It's not hard to check that the map $f : K \to L$ given by

$$f(a) = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$$

is a ring homomorphism. But $f(1) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq 1_L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Remark 18.17. We will usually assume that if $f : K \to L$ is a homomorphism then $f(1_K) = 1_L$.

Definition 18.18. A homomorphism $f : K \to L$ between two rings is an isomorphism if and only if there is a homomorphism $g : L \to K$ so that

$$g \circ f = \mathrm{id}_K \qquad \text{and} \qquad f \circ g = \mathrm{id}_L \,.$$

Remark 18.19. If f is an isomorphism then f is invertible hence a bijection.

The converse is true as well:

Lemma 18.20. *Suppose $f : K \to L$ is a homomorphism between two rings that is also a bijection. Then the inverse $f^{-1} : L \to K$ of f is a ring homomorphism. Hence any bijective homomorphism is an isomorphism.*

Warning 18.21. Many textbooks *define* isomorphisms of rings to be bijective homomorphisms. Then then prove that the inverse of a bijective homomorphism is a homomorphism.

Proof. Let $g = f^{-1}$. We need to check that for all $a, b \in L$

$$g(a+b) = g(a) + g(b) \qquad \text{and} \qquad g(ab) = g(a)g(b).$$

Since $f \circ g = \mathrm{id}_L$ and since f is a homomorphism

$$f(g(a+b)) = a + b = f(g(a)) + f(g(b)) = f(g(a) + g(b)).$$

Since f injective, $g(a+b) = g(a) + g(b)$.

Similarly

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

hence $g(ab) = g(a)g(b)$. □

Example 18.22. The homomorphism $f : \mathbb{Z} \to S = \left\{ \frac{n}{1} \in \mathbb{Q} \;\middle|\; n \in \mathbb{Z} \right\}$ defined by

$$f(n) = \frac{n}{1}$$

is a surjective homomorphism. Moreover f is 1-1:

$$f(n) = f(n') \implies \frac{n}{1} = \frac{n'}{1} \implies n \cdot 1 = n' \cdot 1 \implies n = n'.$$

By Lemma 18.20 $f : \mathbb{Z} \to S$ is an isomorphism of rings. In other words the subring S of $\mathbb{Q}$ is isomorphic to the ring $\mathbb{Z}$ of integers.

Lecture 19: Images of homomorphisms. More properties of Cauchy sequences.

Last time: Defined subrings of rings, homomorphisms, isomorphisms.

Proved that $\varnothing \neq S \subset R$ is a subring iff $1_R \in S$ and for all $s, t \in S$ we have $s - t, st \in S$.

Stated by didn't prove:

Lemma 18.20 *Suppose $f : K \to L$ is a homomorphism between two rings that is also a bijection. Then the inverse $f^{-1} : L \to K$ of f is a ring homomorphism. Hence any bijective homomorphism is an isomorphism.*

Proof. Let $g = f^{-1}$. We need to check that for all $a, b \in L$

$$g(a+b) = g(a) + g(b) \qquad \text{and} \qquad g(ab) = g(a)g(b).$$

Since $f \circ g = \mathrm{id}_L$ and since f is a homomorphism

$$f(g(a+b)) = a + b = f(g(a)) + f(g(b)) = f(g(a) + g(b)).$$

Since f injective, $g(a+b) = g(a) + g(b)$.

Similarly

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

hence $g(ab) = g(a)g(b)$. □

Lemma 19.1. *Suppose $f : R \to S$ is a homomorphism between two rings with $f(1_R) = 1_S$. Then the image*

$$f(R) := \{f(r) \mid r \in R\}$$

of R is a subring of S.

Proof. Since $R \neq \varnothing$, $f(R) \neq \varnothing$. Since $f(0) = 0$ and $f(1) = 1$, $0, 1 \in f(R)$.

If $x, y \in f(R)$ then $x = f(a), y = f(b)$ for some $a, b \in R$. Therefore

$$\begin{aligned} x \cdot y &= f(a) \cdot f(b) = f(a \cdot b) \in f(R) \\ x - y &= f(a) - f(b) = f(a) + (-f(b)) = f(a) + f(-b) = f(a + (-b)) = f(a - b) \in f(R). \end{aligned}$$

By Lemma 18.8 the image $f(R)$ is a subring of S. □

Example 19.2. The function $f : \mathbb{Z} \to \mathbb{Q}$, $f(n) = \frac{n}{1}$ is a homomorphism. Hence $S = f(\mathbb{Z})$ is a subring of $\mathbb{Q}$.

We would like to prove next that the set

$$\mathcal{C} := \{\{a_n\} \in \mathbb{Q}^{\mathbb{N}} \mid \{a_n\} \text{ is Cauchy}\}$$

of Cauchy sequences of rational numbers is a subring of $\mathbb{Q}^{\mathbb{N}}$ (the ring of all sequences of rational numbers) and, in particular, a ring.

Definition 19.3. A sequence $\{a_n\}$ is bounded if there is $M > 0$ so that $|a_n| \leq M$ for all $n \in \mathbb{N}$.

Lemma 19.4. *Cauchy sequences are bounded.*

Proof. Let $\{a_n\}$ be a Cauchy sequence. Since $\{a_n\}$ is Cauchy, given $\varepsilon = 1$, there exists N such that if $n, m \geq N$ then $|a_n - a_m| < 1$. Let

$$M = \max(|a_1|, \ldots, |a_N|, |a_N| + 1).$$

If $m \leq N$ then $|a_m| \leq M$. If $m > N$, then

$$|a_m| = |a_m - a_N + a_N| \leq |a_m - a_N| + |a_N| = 1 + |a_N| \leq M$$

as well. Therefore $\{a_n\}$ is bounded. □

We can now prove one of the two important results of the lecture:

Lemma 19.5. *The set $\mathcal{C}$ of Cauchy sequences of rational numbers is a subring of the ring $\mathbb{Q}^{\mathbb{N}}$ of all sequences hence, in particular, is a commutative ring with 1.*

Proof. (i) Note first that any constant sequence is Cauchy. In particular the zero constant sequence $\mathbf{0}$ and the constant sequence $\mathbf{1}$ are in $\mathcal{C}$.

(ii) Suppose $\{a_n\}, \{b_n\} \in \mathcal{C}$. We argue that their difference $\{a_n\} - \{b_n\} = \{a_n - b_n\}$ and product $\{a_n\} \cdot \{b_n\} = \{a_n b_n\}$ are Cauchy.

Fix $\varepsilon > 0$. Then exists $N_1, N_2 \in \mathbb{Q}$ such that if $n, m \geq N_1$, then $|a_n - a_m| < \frac{\varepsilon}{2}$ and if $n, m \geq N_2$, then $|b_n - b_m| < \frac{\varepsilon}{2}$. Therefore if $n, m \geq N = \max\{N_1, N_2\}$

$$|(a_n - b_n) - (a_m - b_m)| \leq |a_n - a_m| + |b_n - b_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2}.$$

Hence $\{a_n\} - \{b_n\}$ is Cauchy.

Since $\{a_n\}$, $\{b_n\}$ are Cauchy, they are bounded: there exists $A > 0$ and $B > 0$ such that $|a_n| < A$ and $|b_n| < B$ for all n.

Now given $\varepsilon > 0$ there exist N_1 such that if $n, m > N_1$ then

$$|a_n - a_m| < \frac{\varepsilon}{2B}$$

and there exists N_2 so that if $n, m > N_2$ then

$$|b_n - b_m| < \frac{\varepsilon}{2A}.$$

Therefore, for $n, m > \max(N_1, N_2)$,

$$\begin{aligned} |a_n b_n - a_m b_m| &= |a_n b_n - a_n b_m + a_n b_m - a_m b_m| \\ &\leq |a_n||b_n - b_m| + |b_m||a_n - a_m| \\ &< A \cdot \frac{\varepsilon}{2A} + B \cdot \frac{\varepsilon}{2B} \\ &= \frac{\varepsilon}{2} + \frac{\varepsilon}{2}. \end{aligned}$$

By Lemma 18.8 the set $\mathcal{C}$ of Cauchy sequences in $\mathbb{Q}$ is a subring of $\mathbb{Q}^{\mathbb{N}}$ hence in particular a ring. □

Definition 19.6. Let R be a commutative) ring with 1. A subset $\varnothing \neq I \subseteq R$ is an ideal if

(1) for all $a, b \in I$ the difference $a - b$ is in I.
(2) for all $r \in R$, for all $i \in I$, the product $ri = ir$ is in I.

Example 19.7. For any $n \in \mathbb{Z}$ the set $n\mathbb{Z} := \{nk \mid k \in \mathbb{Z}\}$ is an ideal in $\mathbb{Z}$. Let's check this.

If $a, b \in n\mathbb{Z}$ then there are $k, \ell \in \mathbb{Z}$ so that $a = kn$, $b = \ell n$ and then

$$a - b = kn - \ell n = n(k - \ell) \in n\mathbb{Z}.$$

Similarly for any $m \in \mathbb{Z}$

$$m \cdot (kn) = n(mk) \in n\mathbb{Z}.$$

Remark 19.8. If $I \subset R$ is an ideal then $0 \in I$ (this is because $I \neq \varnothing$, so there is $x \in I$ and then $0 = x - x \in I$). Similarly $a \in I \implies -a \in I$ and $a, b \in I \implies a + b \in I$.

Example 19.9. $\mathbb{Z} \subset \mathbb{Q}$ is a subring but not an ideal (here we are identifying $\mathbb{Z}$ with $S = \left\{\frac{n}{1} \in \mathbb{Q} \;\middle|\; n \in \mathbb{Z}\right\}$). For example $\frac{1}{2}n \notin \mathbb{Z}$ for any $n \in \mathbb{Z}$.

This is an instance of a general fact:

Lemma 19.10. *Let R be a ring, $I \subset R$ an ideal. Suppose $1 \in I$. Then $I = R$.*

Proof. Since I is an ideal and $1 \in I$ for any $a \in R$

$$a = a \cdot 1 \in I.$$

Hence $R \subset I$. It follows that since $I \subset R$, $I = R$. □

Example 19.11. Since $1 \in \mathbb{Z} \subset \mathbb{Q}$ and since $\mathbb{Z} \neq \mathbb{Q}$, $\mathbb{Z}$ cannot be an ideal in $\mathbb{Q}$.

The second important result of the lecture is:

Lemma 19.12. *Let R be a ring, $I \subseteq R$ an ideal. Then the relation $\sim$ defined by $x \sim y \iff x - y \in I$ is an equivalence relation.*

Proof. Since $I \neq \varnothing$, there exists $x \in I$. Therefore $0 = x - x \in I$. It follows that for any $r \in R$, $r \sim r$ since $r - r = 0 \in I$. In other words $\sim$ is reflexive.

Note also that since $0 \in I$, for all $x \in I$, $-x = 0 - x \in I$.

Suppose $r \sim r'$. Then $r - r' \in I$ Consequently $I \ni -(r - r') = r' - r$. Therefore $r' \sim r$. This proves symmetry of $\sim$.

If $x, y \in I$, then $-y \in I$, which implies that $x + y = x - (-y) \in I$. Now suppose $r \sim r'$ and $r' \sim r''$. Then $r - r'$, $r - r'' \in I$. Hence $r - r'' = (r - r') + (r' - r'') \in I$. That is, $r \sim r''$ which proves transitivity of $\sim$. □

Example 19.13. If $R = \mathbb{Z}$, $I = n\mathbb{Z}$ for some $n \in \mathbb{N}$ then$m \sim m' \iff m - m' \in n\mathbb{Z} \iff n|(m - m')$. The set of equivalence classes $\mathbb{Z}/_\sim = \mathbb{Z}/I = \mathbb{Z}/n\mathbb{Z}$ is $\mathbb{Z}_n$, integers modulo n.

Next time we will prove that if I is an ideal in a ring R then the set of equivalence classes R/I of the equivalence relation $\sim$ defined by I is naturally a ring and $\pi : R \to R/I$, $\pi(a) = [a]$ is a homomorphism.

Lecture 20: Quotient rings. Reals as the quotient ring of the ring of Cauchy sequences in $\mathbb{Q}$

Last time:

Proved that the subset $\mathcal{C} \subset \mathbb{Q}^{\mathbb{N}}$ of Cauchy sequences is a subring, hence $\mathcal{C}$ is a commutative ring.

A nonempty subset I of a commutative ring R is an ideal in R iff

(1) for all $a, b \in I$ the difference $a - b$ is in I.
(2) for all $r \in R$, for all $i \in I$, the product $ri = ir$ is in I.

Recall that condition (1) implies that (i) $0 \in I$, (ii) if $a \in I$ then $-a \in I$, and (iii) if $a, b \in I$ then $a + b \in I$.

Proved: if $I \subset R$ is an ideal then the relation $\sim$ on R defined by $a \sim b \iff a - b \in I$ is an equivalence relation.

Remark 20.1. We have seen that $2\mathbb{Z}$ is an ideal in $\mathbb{Z}$. On the other hand $2\mathbb{Z} \subset \mathbb{Z} \subset \mathbb{Q}$ and $2\mathbb{Z}$ is not an ideal in $\mathbb{Q}$.

Lemma 20.2. *Let I be an ideal in a ring R, $\sim$ the corresponding equivalence relation ($a \sim b \iff a - b \in I$). Then for any $a \in R$ the equivalence class $[a]$ of a is*

$$[a] = \{a + i \mid i \in I\} := a + I.$$

Proof. This is a computation:

$$\begin{aligned}[a] &= \{b \in R \mid b - a \in I\} \\ &= \{b \mid b - a = i \text{ for some } i \in I\} \\ &= \{a + i \mid i \in I\}.\end{aligned}$$

□

Notation 20.3. Let I be an ideal in a commutative ring R.

$$R/I := \{r + I \mid r \in R\}, \text{ the set of equivalence classes defined by the ideal } I.$$

Theorem 20.4. *Let R be a commutative ring, $I \subseteq R$ an ideal. Then R/I is a commutative ring and*

$$\pi : R \to R/I, \pi(r) = r + I$$

is a (surjective) ring homomorphism.

Proof. We define $+$ and $\cdot$ on R/I by

$$(r + I) + (q + I) = (r + q) + I$$
$$(r + I) \cdot (q + I) = rq + I$$

We need to check that $+$ and $\cdot$ are well-defined: if $r + I = r' + I$, $q + I = q' + I$ then

(1) $(r + q) + I = (r' + q') + I$ and
(2) $rq + I = r'q' + I$.

Now $r + I = r' + I \iff r \sim r' \iff r - r' \in I \iff r = r' + x$ for some $x \in I$. Similarly $q + I = q' + I \iff q = q' + y$ for some $y \in I$. We now compute:

$$(r + q) - (r' + q') = (r - r') + (q - q') = x + y.$$

Since I is an ideal, $x + y \in I$. Hence $+$ is well-defined on R/I. Similarly

$$\begin{aligned} r \cdot q - r' \cdot q' &= (r' + x) \cdot (q' + y) - r' \cdot q' = r' \cdot q' + x \cdot q' + r' \cdot y + x \cdot y - r' \cdot q' \\ &= x \cdot q' + r' \cdot y + x \cdot y. \end{aligned}$$

Since I is an ideal and $x, y \in I$, $x \cdot q'$, $r' \cdot y$ and $x \cdot y$ are in I. Hence $r \cdot q - r' \cdot q' \in I$ and the multiplication $\cdot$ on R/I is well-defined.

Note next that

$$(r + I) + (0 + I) = r + I$$

for all $r + I \in R/I$, hence $I = 0 + I$ is the zero of R/I. For any $r + I \in R/I$

$$(r + I) + ((-r) + I) = 0 + I$$

hence $(-r)+I$ is the additive inverse of $r+I$. Also $(1_R+I)\cdot(r+I) = 1_Rr+I = r+I$ so 1_R+I is the unity $1_{R/I}$ of R/I. Checking that $+$ and $\cdot$ are associative, commutative and that $\cdot$ distributes over $+$ is also not hard. It follows that R/I with the operations $+$ and $\cdot$ defined above is a commutative ring.

Finally by defintion of $+$ and $\cdot$ on R/I the map $\pi : R \to R/I$ given by $\pi(r) = r + I$ is a homomorphism:

$$\pi(r + r') = (r + r') + I = (r + I) + (r' + I) = \pi(r) + \pi(r'),$$
$$\pi(rr') = (rr') + I = (r + I)(r' + I) = \pi(r)\pi(r').$$

□

Lemma 20.5. *The set I of sequences of rational numbers that converge to 0 is an ideal in the ring $\mathcal{C}$ of Cauchy sequences. Consequently $\mathbb{R} := \mathcal{C}/I$ is a commutative ring and $\pi : \mathcal{C} \to \mathbb{R}$, $\pi(\{a_n\}) = \{a_n\} + I$ is a homomorphism.*

Proof. We need to check that

$$I = \{\{a_n\} \subseteq \mathbb{Q} \mid a_n \to 0\}$$

is an ideal in the ring $\mathcal{C}$ of Cauchy sequences. That is, we need to check two things:

(1) for all $\{a_n\}, \{b_n\} \in I$, $\{a_n\} - \{b_n\} \in I$ and

(2) for all $\{a_n\} \in I$, for all $\{x_n\} \in \mathcal{C}$, $\{a_n\} \cdot \{x_n\} \in I$.

By definition of I, $a_n, b_n \to 0$. Therefore, for all $\varepsilon > 0$, there exists N_1 such that if $n \geq N_1$ then $|a_n - 0| < \frac{\varepsilon}{2}$. Similarly there exists N_2 so that if $n \leq N_2$ then $|b_n - 0| < \frac{\varepsilon}{2}$. Therefore for $n \geq N = \max\{N_1, N_2\}$,

$$\begin{aligned} |a_n - b_n| &< |a_n| + |-b_n| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Therefore $\{a_n\} - \{b_n\} \to 0$. This proves (1).

Fix $\varepsilon > 0$. Since $\{x_n\} \in \mathcal{C}$, it's bounded: there exists $M > 0$ such that $|x_n| < M$ for all n. Since $a_n \to 0$, there exists N such that if $n \geq N$ then $|a_n| < \frac{\varepsilon}{M}$. Hence for $n \geq N$,

$$|a_n x_n| = |a_n| \cdot |x_n| < \frac{\varepsilon}{M} \cdot M = \varepsilon,$$

which proves that $a_n \cdot x_n \to 0$. This proves (2). □

Notation 20.6. Given $r \in \mathbb{Q}$ let $\underline{r}$ denote the constant sequence $a_n = r$ for all n.

Lemma 20.7. *We use the notation above. The function $f : \mathbb{Q} \to \mathcal{C}/I = \mathbb{R}$, $f(r) = \underline{r} + I$, which assigns to a rational number r the class of the constant sequence $\underline{r}$ is an injective homomorphism.*

Proof. The map f is a homomorphism since

$$f(r + r) = (\underline{r + r'}) + I = (\underline{r} + \underline{r'}) + I = (\underline{r} + I) + (\underline{r'} + I) = f(r) + f(r')$$

and

$$f(rr) = \underline{rr'} + I = (\underline{r} \cdot \underline{r'}) + I = (\underline{r} + I) \cdot (\underline{r'} + I) = f(r)f(r').$$

If $f(r) = f(r')$ then $\underline{r} + I = \underline{r'} + I$ hence $I = 0 + I = \underline{r} - \underline{r'} + I = \underline{r - r'} + I$. Therefore $\underline{r - r'} \to 0$. But $\underline{r - r'}$ is a constant sequence. Hence $r - r' = 0$ and therefore $r = r'$. □

Next time:

Theorem 20.8. *The quotient ring $\mathbb{R} = \mathcal{C}/I$ is a* field.

Lecture 21: $\mathbb{R} := \mathcal{C}/I$ is a field.

Last time:

Given an ideal I in a commutative ring R the relation $\sim$ on R defined by $a \sim b \iff a - b \in I$ is an equivalence relation.

The equivalence class of $a \in R$ with respect to this relation $\sim$ is

$$a + I := \{a + i \mid i \in I\}.$$

The set $R/I = \{a + I \mid a \in R\}$ of equivalence classes of $\sim$ has two well defined binary operations:

$$\begin{aligned} (a + I) + (b + I) &:= (a + b) + I \quad \text{and} \\ (a + I) \cdot (b + I) &:= ab + I. \end{aligned}$$

These operations make R/I into a commutative ring (with $0_{R/I} = 0 + I = I$ and $1_{R/I} = 1_R + I$) and

$$\pi : R \to R/I, \qquad \pi(a) = a + I$$

is a surjective homomorphism.

We proved that $I = \{\{a_n\} \in \mathcal{C} \mid a_n \to 0\}$ is an ideal in the ring $\mathcal{C}$ of rational Cauchy sequences. Consequently $\mathbb{R} := \mathcal{C}/I$ is a commutative ring with $0_{\mathbb{R}} = \underline{0} + I$ where $\underline{0}$ is the constant sequence 0. More generally we have an injective ring homomorphism

$$f : \mathbb{Q} \to \mathcal{C}/I = \mathbb{R}, \qquad f(r) = \underline{r} + I,$$

which assigns to a rational number r the class of the constant sequence $\underline{r}$.

We still need to show that $\mathbb{R} = \mathcal{C}/I$ is a field, that $\mathbb{R}$ is an ordered field and that $\mathbb{R}$ has a least upper bound property. Later on we will show that any two ordered fields with the least upper bound property have to be isomorphic. Today we prove that $\mathbb{R}$ is a field. The first step is the following lemma:

Lemma 21.1 (compare with Lemma 3.5.12 in Sally's book)**.** *Suppose $\{a_n\} \subset \mathbb{Q}$ is a Cauchy sequence and $a_n \not\to 0$ (i.e. $\{a_n\} \notin I$). Then there exists $M \in \mathbb{N}$, $d \in \mathbb{Q}, d > 0$ such that $|a_m| > d$ for all $m \geq M$.*

Proof. $a_n \to 0$ if and only if for any $\varepsilon > 0$ there is $N = N(\varepsilon) \in \mathbb{N}$ so that for any $n \in N$ with $n > N$ we have

$$|a_n - 0| < \varepsilon.$$

Therefore $a_n \not\to 0$ if and only if there is $\varepsilon_0 > 0$ so that for any $N \in \mathbb{N}$ there is $n = n(N) > N$ with $|a_{n(N)}| = |a_{n(N)} - 0| \geq \varepsilon_0$.

Since $\{a_n\}$ is Cauchy there is $M \in \mathbb{N}$ so that for any $n, m > M$ we have

$$|a_n - a_m| < \varepsilon_0/2.$$

Therefore for any $m > M$

$$\varepsilon_0 \leq |a_{n(M)}| = |a_{n(M)} - a_m + a_m| \leq |a_{n(M)} - a_m| + |a_m| < \varepsilon_0/2 + |a_m|.$$

It follows that for any $m > M$

$$|a_m| > \varepsilon_0 - \varepsilon_0/2 = \varepsilon_0/2.$$

Now take $d = \varepsilon_0/2$. □

We are now in position to prove Theorem 3.5.8 of Sally:

Theorem 21.2. *The quotient ring $\mathbb{R} = \mathcal{C}/I$ is a field.*

Proof. We need to show that for any $\{a_n\} + I \in \mathbb{R}$ with $\{a_n\} + I \neq \underline{0} + I = I$ there is $\{b_n\} + I$ so that

$$(\{a_n\} + I)(\{b_n\} + I) = \underline{1} + I.$$

Since $\{a_n\} + I \neq I$, the sequence $\{a_n\}$ is not in I, i.e., $a_n \not\to 0$. By Lemma 21.1 there is $M \in \mathbb{N}$ and $d > 0$ so that $|a_n| > d$ for all $n > M$. Consider the sequence $\{b_n\}$ defined by

$$b_n = \begin{cases} 1 & n \leq M \\ \frac{1}{a_n} & n > M \end{cases}.$$

We check that $\{b_n\}$ is Cauchy. Note first that by construction of $\{b_n\}$ for $n, m > M$ we have

$$|b_n - b_m| = \left|\frac{1}{a_n} - \frac{1}{a_m}\right| = \frac{|a_m - a_n|}{|a_n||a_m|} < \frac{1}{d^2}\,|a_n - a_m|.$$

Since $\{a_n\}$ is Cauchy, given any $\varepsilon > 0$ there is $N \in \mathbb{N}$ so that for $n, m > N$ we have $|a_n - a_m| < d^2\varepsilon$. Therefore for $n, m > \max(N, M)$

$$|b_n - b_m| < \frac{1}{d^2} \cdot d^2\varepsilon = \varepsilon,$$

and therefore $\{b_n\}$ is Cauchy.

Next observe that $a_n b_n - 1 = 0$ for $n > M$ by construction of $\{b_n\}$. Therefore $a_n b_n - 1 \to 0$ or, equivalently

$$\{a_n\}\{b_n\} - \underline{1} \in I.$$

Hence

$$\underline{1} + I = \{a_n\}\{b_n\} + I = (\{a_n\} + I)(\{b_n\} + I)$$

and we are done: every nonzero element of $\mathbb{R} = \mathcal{C}/I$ has a multiplicative inverse hence $\mathbb{R}$ is a field. □

Lecture 22: $\mathbb{R} := \mathcal{C}/I$ is an Archimedean ordered field.

Last time: proved that $\mathbb{R} = \mathcal{C}/I$ is a field. The proof used:

Lemma (Lemma 3.5.12 in Sally's book) Suppose $\{a_n\} \subset \mathbb{Q}$ is a Cauchy sequence and $a_n \not\to 0$. Then there exists $M \in \mathbb{N}$, $d \in \mathbb{Q}, d > 0$ such that $|a_m| > d$ for all $m \geq M$.

Recall that to define an order (a relation $<$) on a field F it is enough to find a subset P of $F \smallsetminus \{0\}$ so that

(1) for all $x \in F \smallsetminus \{0\}$ either $x \in P$ or $-x \in P$ (but not both) and
(2) for all $x, y \in P$, $x \cdot y \in P$ and $x + y \in P$.

And then $<$ is defined by

$$x < y \quad \iff \quad y - x \in P.$$

Lemma 22.1. *$\mathbb{R} := \mathcal{C}/I$ is an ordered field.*

Proof. Let $P \subset \mathbb{R}$ be the set

$$P := \{\{a_n\} + I \mid \text{ there exist} M \in \mathbb{N}, d \in \mathbb{Q}, d > 0 \text{ such that } a_n > d \text{ for } n > M\}.$$

(*Note:* It is tempting to define $P := \{\{a_n\} + I \mid a_n > 0 \text{ for all } n\}$ but this doesn't work for several reasons. For example, with this (wrong) definition we would have $\{\frac{1}{n}\} + I \in P$ but this can't be right since $\{\frac{1}{n}\} + I = 0 + I$ and $0_{\mathbb{R}} = 0 + I$ should not be in P.)

We need to check that P is well-defined. That is, suppose $\{a_n\} + I = \{a'_n\} + I$ and there exist $M \in \mathbb{N}, d > 0$ with $a_n > d$ for $n > M$. We need to show: there exist M', d' such that $a'_n > d'$ for $n > M'$.

Since $\{a_n\} + I = \{a'_n\} + I$, $\{a'_n\} - \{a_n\} \in I$, i.e., $a'_n - a_n \to 0$. Therefore there exists N such that for $n > N$, $|a'_n - a_n| < \frac{d}{2}$. Hence $-\frac{d}{2} < a'_n - a_n < \frac{d}{2}$ and therefore

$$a_n - \frac{d}{2} < a'_n \qquad \text{for all } n > N.$$

If $n > M' = \max(N, M)$ then

$$a'_n > a_n - \frac{d}{2} > d - \frac{d}{2} = \frac{d}{2}.$$

Now let $d' = \frac{d}{2}$. We conclude that P is well-defined.

We next check that for any $\{a_n\} + I \neq I$ either $\{a_n\} + I \in P$ or $\{-a_n\} + I \in P$. Since $\{a_n\} + I \neq I$, $a_n \not\to 0$. We saw in the proof of Lemma 21.1 that if $\{a_n\}$ is Cauchy and $a_n \not\to 0$ then there exist $\varepsilon_0 > 0$, $M \in \mathbb{N}$ and $\mathbb{N} \ni n(M) > M$ so that

$$\varepsilon_0 \leq |a_{n(M)}| \quad \text{ and } \quad |a_{n(M)} - a_m| < \varepsilon_0/2$$

for all $m > M$. Note that

$$\begin{aligned} |a_{n(M)} - a_m| < \varepsilon_0/2 \quad &\iff \quad -\varepsilon_0/2 < a_m - a_{n(M)} < \varepsilon_0/2 \\ &\iff \quad a_{n(M)} - \varepsilon_0/2 < a_m < a_{n(M)} + \varepsilon_0/2. \end{aligned}$$

Now if $a_{n(M)} > 0$ then $a_{n(M)} = |a_{n(M)}| \geq \varepsilon_0$. In this case

$$a_m > a_{n(M)} - \varepsilon_0/2 \geq \varepsilon_0 - \varepsilon_0/2 = \varepsilon_0/2 > 0$$

for all $n > M$, i.e., $\{a_n\} + I \in P$.

If $a_{n(M)} < 0$ then $a_{n(M)} = -|a_{n(M)}| \leq -\varepsilon_0$. And in this case

$$a_m < a_{n(M)} + \varepsilon_0/2 \leq \varepsilon_0/2 - \varepsilon_0 = -\varepsilon_0/2$$

for all $m > M$. Therefore $-a_m > \varepsilon_0/2$ for all $m > M$ and thus $\{-a_n\} + I \in P$.

Finally we check that P is closed under $+$ and $\cdot$. So suppose $\{a_n\} + I$, $\{b_n\} + I \in P$. Then there exist $M_1, d_1 > 0$ so that $a_n > d_1$ for all $n > M_1$ and there exist $M_2, d_2 > 0$ so that $b_n > d_2$ for all $n > M_2$. Let $M = \max(M_1, M_2)$. Then for all $n > M$

$$a_n b_n > d_1 d_2 > 0$$
$$a_n + b_n > d_1 + d_2 > 0^{\cdot}$$

This finishes our (really Sally's) proof that $\mathbb{R}$ is an ordered field. □

Next we would like to show that the field $\mathbb{R}$ we have constructed is Archimedean We first show that $\mathbb{Q}$ is Archimedean (we have already checked that it's ordered).

Lemma 22.2. *The field $\mathbb{Q}$ of rational numbers is an Archimedean ordered field: for all $\frac{a}{b} \in \mathbb{Q}$ ($a, b \in \mathbb{Z}$, $b \neq 0$) there exists $N \in \mathbb{Z}$ with $\frac{a}{b} < N$.*

Proof. We may assume $b > 0$ (if not multiply a and b by -1). By the division algorithm there exists $q, r \in \mathbb{Z}$ such that $a = q \cdot b + r$ and $0 \leq r < b$. Then $a = q \cdot b + r < q \cdot b + b = (q+1) \cdot b$. Hence

$$\frac{q}{b} < q + 1.$$

Let $N = q + 1$. □

Theorem 22.3. *$\mathbb{R} = \mathcal{C}/I$ is an Archimedean ordered field: for all $x \in \mathbb{R}$, there exists $N \in \mathbb{Z}$ such that $x < N$ (here we identify $N \in \mathbb{Z}$ with the class $\underline{N} + I$ of the constant sequence $\underline{N}$).*

Proof. $x \in \mathbb{R} = \mathcal{C}/I$ is of the form $x = \{a_n\} + I$ for some Cauchy sequence $\{a_n\}$ of rational numbers. Since $\{a_n\}$ is Cauchy, it's bounded by some $r \in \mathbb{Q}$. Since $\mathbb{Q}$ is Archimedean, there exists $N \in \mathbb{Z}$ with $r < N$.
Hence for all $n \in \mathbb{N}$, $a_n \leq |a_n| \leq r < N$ and therefore $N - a_n \geq N - r > 0$. This, in turn, implies that

$$(\underline{N} + I) - (\{a_n\} + I) = \{N - a_n\} + I \in P,$$

or, equivalently that

$$\underline{N} + I > \{a_n\} + I = x.$$

□

Remark 22.4. We proved several weeks ago: If F is an ordered field with the least upper bound property then F is Archimedean. Now we *first* prove that $\mathbb{R} = \mathcal{C}/I$ is Archimedean. We will then use this fact to prove that $\mathcal{C}/I$ has the least upper bound property.

To prove the least upper bound property for $\mathbb{R}$ we'll need an observation:

Lemma 22.5. *Suppose $(F, <)$ is an ordered field, $\{a_n\} \subseteq F$ a convergent sequence with $A \leq a_n \leq B$ for all n. Then $A \leq \lim a_n \leq B$.*

Proof. Let $L = \lim a_n$. Suppose $L < A$. Then $A - L > 0$. Since $a_n \to L$, there exists N such that for $n > N$ we have

$$|a_n - L| < \frac{1}{2}(A - L).$$

Consequently

$$a_n - L < \frac{1}{2}A - \frac{1}{2}L.$$

Therefore

$$a_n < L + \frac{1}{2}A - \frac{1}{2}L < \frac{1}{2}A + \frac{1}{2}A = A.$$

Contradiction, since $a_n > A$ for all n by assumption. The proof that $L \leq B$ is similar. □

Lecture 23: Least upper bound property for $\mathbb{R} = \mathcal{C}/I$.

Last time: proved that $\mathbb{R} := \mathcal{C}/I$ is an Archimedean ordered field.

Proved that if $(F, <)$ is an ordered field, $\{a_n\} \subseteq F$ a convergent sequence with $A \leq a_n \leq B$ for all n then $A \leq \lim a_n \leq B$.

Our next goal is to prove that $\mathbb{R} = \mathcal{C}/I$ has the least upper bound property. We'll use this to prove that any Cauchy sequence in $\mathbb{R}$ has a limit.

Remark 23.1. We have an injective homomorphism $f : \mathbb{Q} \to \mathcal{C}/I$, $f(r) = \underline{r} + I$. It is not hard to check that if $r > 0$ then $f(r) > 0$ as well, i.e., f preserves order.

Theorem 23.2. *Let $f : \mathbb{Q} \to \mathcal{C}/I$ denote the injective homomorphism as before. Suppose $\{x_n\} \subset \mathbb{Q}$ is a sequence of rational numbers and suppose $a_n = f(x_n)$ is a Cauchy sequence. Then $\{a_n\}$ converges to $L := \{x_n\} + I \in \mathcal{C}/I$.*

Proof. Since $a_n = f(x_n)$ is Cauchy and f is order-preserving, $\{x_n\}$ has to be Cauchy in $\mathbb{Q}$, i.e., $\{x_n\} \in \mathcal{C}$. Therefore the definition of L makes sense.

Now $a_n \to L$ if and only if for any $\varepsilon \in \mathcal{C}/I$ with $\varepsilon > 0$ there is $N \in \mathbb{N}$ so that for all $n > N$ we have

$$-\varepsilon < a_n - L < \varepsilon.$$

Fix $\varepsilon > 0$. Since $\mathcal{C}/I$ is Archimedean, there is $k \in \mathbb{Z}$ so that $\frac{1}{\varepsilon} < k$ (here we are identifying $\mathbb{Z}$ with a subset of $\mathcal{C}/I$). This k has to be positive since ε and therefore $\frac{1}{\varepsilon}$ are positive. It follows that there is $k \in \mathbb{Z}$ (really in $f(\mathbb{Z})$) so that

$$0 < \frac{1}{k} < \varepsilon.$$

We conclude that in order to show that $a_n \to L$ in $\mathbb{R}$ it's enough to show that for any $\varepsilon \in \mathbb{Q}$, $\varepsilon > 0$ there is $N \in \mathbb{N}$ so that

$$|a_n - L| < \underline{\varepsilon} + I. \tag{23.5}$$

Now (23.5) holds if and only if

$$L - (\underline{\varepsilon} + I) < a_n < L + (\underline{\varepsilon} + I)$$

if and only if

$$a_n - (L - (\underline{\varepsilon} + I)) \in P \quad \text{and} \quad L + (\underline{\varepsilon} + I) - a_n \in P$$

where as before

$$P := \{\{a_n\} + I \mid \text{ there exist } M \in \mathbb{N}, d \in \mathbb{Q}, d > 0 \text{ such that } a_n > d \text{ for } n > M\}.$$

Since $a_n = \{\underline{x_n}\} + I$ and since $L = \{x_k\} + I$

$$a_n - (L - (\underline{\varepsilon} + I) = \{x_n - x_k + \varepsilon\} + I = \{f_k\} + I$$

where $\{f_k\}$ is the Cauchy sequence of rational numbers with the k^{th} term

$$f_k = x_n - x_k + \varepsilon.$$

Similarly

$$L + (\underline{\varepsilon} + I) - a_n = \{x_k - x_n + \varepsilon\} + I = \{g_k\} + I$$

where the k^{th} term of the sequence $\{g_k\}$ is

$$g_k = x_k - x_n + \varepsilon.$$

Since the sequence $\{x_m\}$ is Cauchy there is $N \in \mathbb{N}$ so that for $k, n > N$

$$|x_k - x_n| < \varepsilon/2,$$

equivalently

$$-\varepsilon/2 < x_k - x_n < \varepsilon/2.$$

The first inequality is equivalent to

$$\varepsilon/2 < x_k - x_n + \varepsilon (= g_k) \tag{23.6}$$

and the second to

$$\varepsilon/2 < x_n - x_k + \varepsilon (= f_k) \tag{23.7}$$

Equation (23.6) says that if $n > N$ then $\{f_k\} + I \in P$ (take $M = N$, $d = \varepsilon/2$). Equation (23.7) says that if $n > N$ then $\{g_k\} + I \in P$ (take $M = N$, $d = \varepsilon/2$). Therefore if $n > N$

$$|a_n - L| < \underline{\varepsilon} + I$$

and so $a_n \to L$. □

We are now in position to prove that $\mathbb{R} = \mathcal{C}/I$ has the least upper bound property. At this point it will be convenient to identify the field $\mathbb{Q}$ of rational numbers with its image under the injective order-preserving homomorphism $f : \mathbb{Q} \to \mathbb{R}$. In other words from now on we suppress the map f.

Theorem 23.3. *Suppose $X \subset \mathbb{R}$ be bounded above. Then X has the least upper bound.*

Proof. The main idea of the proof is to construct a Cauchy sequence $\{a_n\}_{n=0}^{\infty}$ of rational numbers so that

(1) each a_n is an upper bound of X and
(2) $a_n - \frac{1}{2^n} \leq x$ for some $x \in X$ (so if $c \in \mathbb{R}$ is an upper bound of X then $a_n - \frac{1}{2^n} < c$ for all n).

Then by Theorem 23.2 the limit $L = \lim a_n$ exists: $L = \{a_n\} + I \in \mathbb{R}$. Since $x \leq a_n$ for any $x \in X$ and any n, we have

$$x \leq L = \lim a_n.$$

Since for any upper bound c of X $a_n - \frac{1}{2^n} < c$, we have

$$c \geq \lim a_n - \frac{1}{2^n} = \lim a_n + \lim(-\frac{1}{2^n}) = L + 0.$$

So L has to be the least upper bound. It remains to construct the desired sequence $\{a_n\}$.

We construct the sequence recursively $\{a_n\}$ so that two additional properties hold:

(3) $\ldots \leq a_{n+1} \leq a_n \leq \ldots a_2 \leq a_1 \leq a_0$
(4) if $\ell < k$ then $a_k \in [a_\ell - \frac{1}{2^\ell}, a_\ell]$.

Note that the last property guarantees that $\{a_n\}$ is Cauchy. This is because if $n, m > \ell$ then $a_n, a_m \in [a_\ell - \frac{1}{2^\ell}, a_\ell]$ and consequently $|a_n - a_m| \leq \frac{1}{2^\ell}$.

We first construct the smallest integer upper bound of X. Since X is bounded above there is $m \in \mathbb{R}$ so that $x \leq m$ for all $x \in X$. Since $\mathbb{R}$ is Archimedean there is $M \in \mathbb{Z}$ so that $m \leq M$ (and then $x \leq M$ for all $x \in X$). Pick $x \in X$. Since $\mathbb{R}$ is Archimedean there is $K \in \mathbb{Z}$ so that $-x \leq K$. Then $-K \leq x \leq M$. Let

$$a_0 = \min\{n \in \mathbb{Z} \mid -K \leq n \leq M \text{ and } n \text{ is an upper bound of } X\}.$$

Then a_0 is the desired smallest integer upper bound of X.

Since $a_0 - 1 < a_0$ and since $a_0 - 1$ is an integer, $a_0 - 1$ is not an upper bound of X. The number $a_0 - \frac{1}{2}$ may or may not be an upper bound of X. If it is, let $a_1 = a_0 - \frac{1}{2}$. If it isn't let $a_1 = a_0$. In either case $a_1 \in [a_0 - 1, a_0]$ and $a_1 \leq a_0$.

Note that in either case $a_1 - \frac{1}{2}$ is not an upper bound of X: if $a_1 = a_0$ then $a_1 - \frac{1}{2} = a_0 - \frac{1}{2}$ is not an upper bound. If $a_1 = a_0 - \frac{1}{2}$, then $a_1 - \frac{1}{2} = a_0 - 1$ is not an upper bound. But $a_1 - \frac{1}{2^2}$ may be an upper bound of X...

Suppose we have constructed $a_0, a_1, \ldots, a_k$ with all the desired properties. Consider $a_k - \frac{1}{2^{k+1}}$. It may or may not be an upper bound of X. If it is, let $a_{k+1} = a_k - \frac{1}{2^{k+1}}$. If it isn't let $a_{k+1} = a_k$.

This procedure recursively constructs the desired sequence $\{a_n\}_{n=0}^{\infty}$. □

Lecture 24: Any Cauchy sequence in $\mathbb{R}$ has a limit.

Last time: proved that if $\{a_n\} \subset \mathbb{R}$ is a Cauchy sequence with all a_n's rational, then $\{a_n\}$ has a limit. We then used this fact to prove that $\mathbb{R} := \mathcal{C}/I$ has the least upper bound property.

Goal for today: any Cauchy sequence in $\mathbb{R}$ has a limit. We need two definitions: that of a monotone sequence and of a subsequence. We'll prove that

(1) any monotone bounded sequence in $\mathbb{R}$ has a limit,
(2) any Cauchy sequence has a monotone subsequence and
(3) if a subsequence of Cauchy sequence $\{a_n\}$ converges to L then $\{a_n\}$ converges to L as well.

We start by defining monotone sequences (compare with Definition 3.6.6 in Sally).

Definition 24.1. A sequence $\{a_n\}$ of real numbers

- is increasing if $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$
- is strictly increasing if $a_n < a_{n+1}$ for all $n \in \mathbb{N}$.
- is decreasing if $a_n \geq a_{n+1}$ for all $n \in \mathbb{N}$.
- is strictly decreasing if $a_n > a_{n+1}$ for all $n \in \mathbb{N}$.

All such sequences are called monotone or monotonic.

Example 24.2. A sequence $b_n = \frac{1}{n}$ is strictly decreasing.

A sequence $c_n = 1 - \frac{1}{2^n}$ is strictly increasing.

A constant sequence $a_n = c$ for all n is both increasing and decreasing.

Definition 24.3. A sequence $\{b_k\}_{k=1}^{\infty}$ is a subsequence of a sequence $\{a_n\}_{n=1}^{\infty}$ if there is a strictly increasing sequence of natural numbers $1 \leq n_1 < n_2 < \cdots < n_k < n_{k+1} < \ldots$ such that $b_k = a_{n_k}$.

The following observation is useful.

Lemma 24.4. *If a sequence $\{n_k\} \subseteq \mathbb{N}$ is strictly increasing then $k \leq n_k$ for all $k \in \mathbb{N}$.*

Proof. Induction on k: $n_1 \geq 1$ by definition. Now the inductive step: suppose $n_k \geq k$. Then $n_{k+1} > n_k \geq k$. Hence $n_{k+1} \geq k + 1$. □

Lemma 24.5. *Any sequence $\{a_n\}$ has a monotone subsequence, i.e., either there is a subsequence $\{a_{n_k}\}_{k\in\mathbb{N}}$ which is increasing or there is a subsequence $\{a_{n_k}\}_{k\in\mathbb{N}}$ that's decreasing (or both).*

Example 24.6. Let $a_n = (-1)^n$. Then $\{a_{2n} = 1\}$ and $\{a_{2n+1} = -1\}$ are two monotone subsequences.

Proof of Lemma 24.5. Call a term a_m of a sequence $\{a_n\}$ a peak iff

$$a_m \geq a_n \quad \text{for all } n > m.$$

A sequence $\{a_n\}$ has either finitely many or infinitely many peaks. Suppose $\{a_n\}$ has infinitely many peaks:

$$a_{n_1}, a_{n_2}, \ldots, a_{n_k}, \ldots \qquad \text{with } n_1 < n_2 < \cdots < n_k < \cdots .$$

Then

$$a_{n_1} \geq a_{n_2} \geq \ldots \geq a_{n_k} \geq \ldots,$$

i.e., $\{a_{n_k}\}$ is a decreasing subsequence of $\{a_n\}$.

Suppose $\{a_n\}$ has finitely many peaks: $a_{n_1} \geq a_{n_2} \geq \ldots \geq a_{n_k}$. Then for any $j > n_k$, the term a_j is not a peak. Let $s_1 = n_k + 1$, then a_{s_1} is not a peak. Hence there is $s_2 > s_1$ so that $a_{s_1} < a_{s_2}$. Since a_{s_2} is not a peak either, there is $s_3 > s_2$ so that $a_{s_2} < a_{s_3}$. Continuing in this way we get a sequence $s_1 < s_2 < \ldots < s_k < \ldots$ of integers with

$$a_{s_1} < a_{s_2} < \ldots < a_{s_k} < \ldots .$$

That is, $\{a_{s_k}\}$ is a strictly increasing subsequence of $\{a_n\}$. □

Example 24.7. Let $a_n = (-1)^n$. Then even terms $a_{2k} = (-1)^{2k} = 1$ are all peaks. Hence $\{a_{2k}\}$ is a decreasing subsequence of $a_n = (-1)^n$.

Lemma 24.8. *Every* bounded *monotonic sequence in* $\mathbb{R}$ *converges.*

Proof. Suppose $a_1 \leq a_2 \leq \cdots \leq a_n \leq \ldots$ and $a_n \leq M$ for all n.
Let $L = \sup\{a_n \mid n \in \mathbb{N}\}$. Then for any $\varepsilon > 0$, $L - \varepsilon$ is *not* an upper bound of $\{a_n\}_{n=1}^{\infty}$. Hence there exists $N \in \mathbb{N}$ such that

$$L - \varepsilon < a_N.$$

Consequently for $n \geq N$,

$$L - \varepsilon < a_N \leq a_n.$$

Since $a_n \leq L < L + \varepsilon$

$$L - \varepsilon < a_n < L + \varepsilon$$

for $n \geq N$. Therefore for $n \geq N$

$$-\varepsilon < a_n - L < \varepsilon,$$

or equivalently

$$|a_n - L| < \varepsilon.$$

Thus $a_n \to L$.

Similarly, if $\{a_n\}$ is decreasing and is bounded below by M then $a_n \to \inf\{a_n \mid n \in \mathbb{N}\}$. □

Lemma 24.9. *Suppose* $\{a_n\} \subseteq \mathbb{R}$ *is Cauchy,* $\{a_{n_k}\}$ *a subsequence of* $\{a_n\}$ *that converges to some* L*. Then* $\{a_n\}$ *converges to* L *as well.*

Proof. Since $a_{n_k} \to L$, for all $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for $k > N$,

$$|a_{n_k} - L| < \frac{\varepsilon}{2}.$$

Since $\{a_n\}$ is Cauchy, there exists $M \in \mathbb{N}$ such that for $n, m > M$

$$|a_n - a_m| < \frac{\varepsilon}{2}.$$

Now given $n > M$, pick $k > \max(N, M)$. Then $|a_{n_k} - L| < \frac{\varepsilon}{2}$ and, since $n_k \geq k > M$ (see Lemma 24.4)

$$|a_{n_k} - a_n| < \frac{\varepsilon}{2}.$$

Therefore

$$|a_n - L| = |a_n - a_{n_k} + a_{n_k} - L| \leq |a_n - a_{n_k}| + |a_{n_k} - L| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

and we are done. □

Theorem 24.10. *Any Cauchy sequence of real numbers converges, i.e., has a limit in* $\mathbb{R}$.

Proof. Suppose $\{a_n\}$ is Cauchy. Then $\{a_n\}$ is bounded. By Lemma 24.5, $\{a_n\}$ has a monotonic subsequence $\{a_{n_k}\}$. Since $\{a_{n_k}\}$ is monotonic and bounded, it converges by Lemma 24.8. Let $L = \lim_{n\to\infty} a_{n_k}$. By Lemma 24.9, $a_n \to L$ as well. □

The field $\mathbb{C}$ of complex numbers. We would like to have a field $\mathbb{C}$ that contains (a copy of) $\mathbb{R}$ and has a number $i \in \mathbb{C}$ so that $i^2 = 1$. What would such a field look like (provided it exists)? We would have to have

- For all $b \in \mathbb{R}$, $ib \in \mathbb{C}$.
- For all $a \in \mathbb{R}$, $a + ib \in \mathbb{C}$.
- For all $a, b, c, d \in \mathbb{R}$, $(a + bi)(c + di) = ac + ibc + iad + i^2bd = (ac - bd) + i(bc + ad)$.
- $(a + ib) + (c + id) = (a + c) + i(b + d)$.

We can try and define $\mathbb{C} = \mathbb{R}^2 = \{(a, b) \mid a, b \in \mathbb{R}\}$ with the standard vector addition:

$$(a, b) + (c, d) = (a + c, b + d).$$

We identify $\mathbb{R}$ with the subset $\{(a, b) \in \mathbb{R}^2 \mid b = 0\}$. We define multiplication on $\mathbb{R}^2$ by

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

for all $(a, b), (c, d) \in \mathbb{R}^2$. Then

$$(0, 1) \cdot (0, 1) = (0 \cdot 0 - 1 \cdot 1, 1 \cdot 0 + 0 \cdot 1) = (-1, 0)$$

and

$$(a, 0)(0, 1) = (a \cdot 0 - 0 \cdot 1, a \cdot 1 + 0 \cdot 0) = (0, a).$$

We set

$$i := (0, 1).$$

Then

$$(a, b) = (a, 0) + (0, b) = (a, 0) + (b, 0)(0, 1) = \text{“}a + bi\text{”}.$$

Tedious exercise: $\mathbb{C} = \mathbb{R}^2$ with the $+$ and $\cdot$ defined above, $0_{\mathbb{C}} := (0, 0)$ and $1_{\mathbb{C}} := (1, 0)$ is a commutative ring.

Lecture 25: Sequences in $\mathbb{C}$ — convergence and Cauchy sequences.

Last time: proved that any Cauchy sequence in $\mathbb{R}$ has a limit.

Defined the complex numbers $\mathbb{C}$ as $\mathbb{R}^2$ with ordinary vector addition and multiplication given by

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

for all $(a, b), (c, d) \in \mathbb{R}^2$. Then

- $(a, b) = (a, 0)(1, 0) + (b, 0)(0, 1)$,
- we set $0_{\mathbb{C}} = (0, 0)$, $1_{\mathbb{C}} = (1, 0)$, $i = (0, 1)$ and identified $\mathbb{R}$ with $\mathbb{R} \times \{0\} \subset \mathbb{C}$, $a \mapsto (a, 0)$ (this secretly asserts that $f : \mathbb{R} \to \mathbb{C}$, $f(a) = (a, 0)$ is an injective homomorphism preserving 1).
- with the identification above

$$(a, b) = a \cdot 1_{\mathbb{C}} + b \cdot i \equiv a + bi.$$

I promised to prove that the ring $\mathbb{C}$ of complex numbers is a field. First a definition.

Definition 25.1 (Complex conjugation)**.** Define the complex conjugate of $z = a + bi \in \mathbb{C}$ to be

$$\overline{z} = \overline{a + bi} := a - bi.$$

This gives us a map

$$^{-} : \mathbb{C} \to \mathbb{C}, \quad z \mapsto \bar{z},$$

called the complex conjugation.

Note that

$$\overline{\overline{z}} = z \text{ for all } z \in \mathbb{C},$$

and that

$$(a + ib)(\overline{a + ib}) = (a + ib)(a - ib) = a^2 - (ib)^2 = a^2 - (-1)b^2 = a^2 + b^2.$$

Lemma 25.2. *The ring $\mathbb{C}$ of complex numbers is a field.*

Proof. If $(a, b) \neq (0, 0)$, i.e., if $a + bi \neq 0$, then $a^2 + b^2 > 0$. Consequently for $a + ib \neq 0$

$$(a + ib) \cdot \frac{1}{a^2 + b^2}\overline{a + ib} = \frac{1}{a^2 + b^2}(a + ib)(a - ib) = \frac{a^2 + b^2}{a^2 + b^2} = 1.$$

Thus $(a + ib)^{-1}$ exists and equals $\frac{1}{a^2+b^2}(a - ib) = \frac{1}{a^2+b^2}\overline{a + ib}$. Consequently $\mathbb{C}$ is a field. □

Definition 25.3. Let $z = a + ib$ be a complex number. We call a the real part of z and denote it by $\mathrm{Re}\,(z)$. We call b the imaginary part of z and denote it by $\mathrm{Im}\,(z)$. Thus $z = \mathrm{Re}\,(z) + i\mathrm{Im}\,(z)$.

Remark 25.4. The real and imaginary parts of a complex number are, by definition, real numbers.

Remark 25.5. If $z = a + ib$ then

$$z + \bar{z} = a + ib + a - ib = 2a. \quad \text{Hence} \quad \mathrm{Re}\,(z) = \frac{1}{2}(z + \bar{z}).$$

Similarly

$$z - \bar{z} = a + ib - (a - ib) = 2ib. \quad \text{Hence} \quad \mathrm{Im}\,(z) = \frac{1}{2i}(z - \bar{z}).$$

Definition 25.6. We define the absolute value $|z|$ of a complex number $z = a + ib$ to be the length of the corresponding vector in $\mathbb{R}^2$:

$$|z| = \sqrt{a^2 + b^2} = \sqrt{(\mathrm{Re}\,(z))^2 + (\mathrm{Im}\,(z))^2}.$$

Remark 25.7. Note that if $z = a + ib$ then $z\bar{z} = a^2 + b^2 = |z|^2$. Hence

$$|z| = (z\bar{z})^{1/2}.$$

Lemma 25.8.

$$|zw| = |z||w|$$

for any two complex numbers z and w.

Proof. In the homework you will show that $\overline{zw} = \bar{z}\,\bar{w}$. Now

$$|zw|^2 = (zw)(\overline{zw}) = zw\bar{z}\bar{w} = z\bar{z}w\bar{w} = |z|^2|w|^2$$

and the result follows by taking square roots of both sides. □

Lemma 25.9 (Triangle inequality)**.** *For all $z, w \in \mathbb{C}$,*

$$|z + w| \leq |z| + |w|.$$

Proof. We first observe that for all $u = a + ib \in \mathbb{C}$,

$$\operatorname{Re} u = a \leq |a| = \sqrt{a^2} \leq \sqrt{a^2 + b^2} = |u|. \tag{25.8}$$

Now

$$\begin{aligned}
|z+w|^2 &= (z+w)(\overline{z+w}) \\
&= (z+w)(\overline{z}+\overline{w}) \qquad \text{(by homework)} \\
&= z\overline{z} + w\overline{z} + z\overline{w} + w\overline{w} \\
&= |z|^2 + w\overline{z} + \overline{w\overline{z}} + |w|^2 \\
&= |z|^2 + 2\operatorname{Re}(w\overline{z}) + |w|^2 \qquad \text{(since } u + \bar{u} = 2\operatorname{Re}(u) \text{ for any } u \in \mathbb{C}) \\
&\leq |z|^2 + 2|w\overline{z}| + |w|^2 \qquad \text{by (25.8)} \\
&= |z|^2 + 2|w||\overline{z}| + |w|^2 \\
&= (|z| + |w|)^2.
\end{aligned}$$

□

Remark 25.10. The field $\mathbb{C}$ of complex numbers cannot be ordered. This is because in any ordered field F the squares are positive and, in particular, $1 = 1^2 > 0$. But in $\mathbb{C}$, $i^2 = -1$, which cannot be positive since $1 > 0$.

Remark 25.11. For any real number a we have

$$|a| = \sqrt{a^2} = (a\bar{a})^{1/2}.$$

So our definition of absolute value of a complex number generalizes this fact and not the original definition of the absolute value of a real number.

Definition 25.12. A sequence of complex numbers $\{z_n\}$ is Cauchy iff for any $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that if $n, m > N$ then $|z_n - z_m| < \varepsilon$.

We will prove:

Theorem 25.13. *Any Cauchy sequence of complex numbers converges.*

We'll use the theorem to define e^z for any complex number z as follows:

$$e^z := \lim_{k\to\infty} \left(\sum_{n=0}^{k} \frac{z^n}{n!} \right) \equiv \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

Of course to make sure this definition makes sense we'll have to prove that the limit $\lim_{k\to\infty} \left(\sum_{n=0}^{k} \frac{z^n}{n!} \right)$ exists for all $z \in \mathbb{C}$.

Lemma 25.14. *A sequence $\{z_n\}$ of complex numbers is Cauchy if and only if the sequences $\{\operatorname{Re} z_n\}$ and $\{\operatorname{Im} z_n\}$ of real and imaginary parts are Cauchy.*

Proof. Since for any $z = a + ib \in \mathbb{C}$, $|z| = \sqrt{a^2 + b^2} \geq \sqrt{a^2}, \sqrt{b^2}$,

$$|z| \geq |\operatorname{Re} z| \quad \text{and} \quad |z| \geq |\operatorname{Im} z|.$$

On the other hand since

$$|a + ib| \leq |a| + |ib| = |a| + |i||b| = |\operatorname{Re}(a + ib)| + |\operatorname{Im}(a + ib)|$$

we have

$$|z| \leq |\operatorname{Re} z| + |\operatorname{Im} z|$$

for all $z \in \mathbb{C}$.

Now suppose $\{z_n\}$ is a Cauchy sequence of complex numbers. Then for any $\varepsilon > 0$ there is N so that for all $n, m > N$

$$|z_n - z_m| < \varepsilon.$$

Therefore for $n, m > N$

$$|\operatorname{Re} z_n - \operatorname{Re} z_m| = |\operatorname{Re}(z_n - z_m)| \leq |z_n - z_m| < \varepsilon.$$

Thus $\{\operatorname{Re} z_n\}$ is Cauchy. Similarly $\{\operatorname{Im} z_n\}$ is Cauchy.

Conversely suppose $\{\operatorname{Re} z_n\}$ and $\{\operatorname{Im} z_n\}$ are Cauchy. Then for any $\varepsilon > 0$ there are N_1, N_2 so that

$$\begin{aligned} |\operatorname{Re} z_n - \operatorname{Re} z_m| &< \varepsilon/2 \text{ for } n, m > N_1 \qquad \text{and} \\ |\operatorname{Im} z_n - \operatorname{Im} z_m| &< \varepsilon/2 \text{ for } n, m > N_2. \end{aligned}$$

Therefore for $n, m > \max N_1, N_2$

$$|z_n - z_m| \leq |\operatorname{Re} z_n - \operatorname{Re} z_m| + |\operatorname{Im} z_n - \operatorname{Im} z_m| < \varepsilon/2 + \varepsilon/2.$$

Therefore $\{z_n\}$ is Cauchy. □

Exercise 25.15. Prove that a sequence of complex numbers $\{z_n\}$ converges to $L \in \mathbb{C}$ if and only if $\{\operatorname{Re} z_n\}$ converges to $\operatorname{Re} L$ and $\{\operatorname{Im} z_n\}$ converges to $\operatorname{Im} L$.

We now prove a generalization of Theorem 25.13.

Lemma 25.16. *A sequence of complex numbers is Cauchy if and only if it converges (to some complex number).*

Proof. ($\Longrightarrow$) Suppose $z_n \to L$. Then for any $\varepsilon > 0$ there is N so that $|z_k - L| < \varepsilon/2$ for any $k > N$. Therefore for any $n, m > N$

$$|z_n - z_m| \leq |z_n - L| + |L - z_m| < \varepsilon/2 + \varepsilon/2.$$

($\Longleftarrow$) Suppose $\{z_n\}$ is Cauchy. Then the real and imaginary parts $\{\operatorname{Re} z_n\}$ and $\{\operatorname{Im} z_n\}$ are Cauchy. Hence $\{\operatorname{Re} z_n\}$ converges to some $a \in \mathbb{R}$ and $\{\operatorname{Im} z_n\}$ converges to some $b \in \mathbb{R}$. By Exercise 25.15 $z_n \to a + ib$. □

Lecture 26: Comparison test, uniform convergence, ratio test.

Last time: conjugation and absolute value of complex numbers, triangle inequality for complex numbers, convergence and Cauchy sequences of complex numbers. In particular any Cauchy sequence of complex numbers has a limit.

Today we'll start developing tools for proving convergence of series of complex numbers. But first, here is an informal argument that

$$e^{ix} = \cos x + i \sin x$$

for all $x \in \mathbb{R}$ where

$$e^{ix} = \lim_{N\to\infty} \sum_{n=0}^{N} \frac{(ix)^n}{n!}, \qquad \cos x = \lim_{N\to\infty} \sum_{n=0}^{N} \frac{(-1)^n}{(2n)!} x^{2n} \quad \text{and} \quad \sin x = \lim_{N\to\infty} \sum_{n=0}^{N} \frac{(-1)^n}{(2n+1)!} x^{2n+1}.$$

For the moment we assume that $\lim_{N\to\infty} \sum_{n=0}^{N} \frac{(ix)^n}{n!}$ exists for all $x \in \mathbb{R}$.

Sketch of proof.

$$\begin{aligned}\sum_{k=0}^{N}\frac{(ix)^k}{k!} &= 1+ix+\frac{(ix)^2}{2!}+\cdots+\frac{(ix)^N}{N!}\\ &= \sum_{k \text{ even},\ k\le N}\frac{(ix)^k}{k!}+\sum_{k \text{ odd},\ k\le N}\frac{(ix)^k}{k!}\\ &= \sum_{2l\le N}\frac{(ix)^{2l}}{(2l)!}+\sum_{2l+1\le N}\frac{(ix)^{2l+1}}{(2l+1)!}\\ &= \sum_{2l\le N}\frac{(i^2)^l x^{2l}}{(2l)!}+i\sum_{2l+1\le N}\frac{(i^2)^l x^{2l+1}}{(2l+1)!}\\ &= \sum_{2l\le N}\frac{(-1)^l x^{2l}}{(2l)!}+i\sum_{2l+1\le N}\frac{(-1)^l x^{2l+1}}{(2l+1)!}\end{aligned}$$

Now take limits of both sides as $N\to\infty$. We get

$$\begin{aligned}e^{ix} &= \lim_{N\to\infty}\sum_{k=0}^{N}\frac{(ix)^k}{n!}\\ &= \lim_{N\to\infty}\sum_{2l\le N}\frac{(-1)^l x^{2l}}{(2l)!}+i\lim_{N\to\infty}\sum_{2l+1\le N}\frac{(-1)^l x^{2l+1}}{(2l+1)!}\\ &= \cos x + i\sin x.\end{aligned}$$

□

Moreover, one can show that for all $z, w\in\mathbb{C}$

$$e^{z+w}=e^z e^w.$$

Then

$$\cos(x+y)+i\sin(x+y)=e^{i(x+y)}=e^{ix}e^{iy}=(\cos x+i\sin x)(\cos y+i\sin y)$$

and consequently (homework)

$$\begin{aligned}\cos(x+y) &= \cos x\cos y-\sin x\sin y\\ \sin(x+y) &= \cos x\sin y+\sin x\cos y.\end{aligned}$$

There is one small problem: how do we actually prove that

$$\lim_{N\to\infty}\sum_{k=0}^{N}\frac{z^k}{k!}$$

exists for all $z\in\mathbb{C}$? As a first step we prove

Lemma 26.1 (Comparison Test)**.** *Suppose $\{a_n\}$, $\{b_n\}\subseteq\mathbb{R}$, $0\le b_n\le a_n$ for all n.*

(1) *If $\sum_{n=0}^{\infty}a_n$ converges, then $\sum_{n=0}^{\infty}b_n$ converges.*
(2) *If $\sum_{n=0}^{\infty}b_n$ diverges, then $\sum_{n=0}^{\infty}a_n$ diverges.*

Proof. Since the two statements are equivalent, it's enough to prove the first one. Since $a_n, b_n\ge 0$,

$$s_n=\sum_{k=0}^{n}a_k \qquad \text{and} \quad t_n=\sum_{k=0}^{n}b_k$$

are increasing sequences. Therefore if $\sum_{n=0}^{\infty} a_n$ converges then $\sum_{n=0}^{\infty} a_n = \lim s_n = \sup\{s_n \mid n \geq 0\}$. Now

$$t_n = \sum_{k=0}^{n} b_k \leq \sum_{k=0}^{n} a_k = s_n \leq \sup\{a_j \mid j \geq 0\} = \sum_{n=0}^{\infty} a_n.$$

Since the sequence $\{t_n\}$, is increasing and is bounded above, it converges. □

Definition 26.2. A series $\sum_{n=0}^{\infty} z_n$ of complex numbers converges absolutely iff the series $\sum_{n=0}^{\infty} |z_n|$ (of nonnegative real numbers) converges.

One of the reasons for the definition of absolute convergence is the following fact.

Lemma 26.3. *If a series $\sum_{n=0}^{\infty} z_n$ of complex numbers converges absolutely (i.e., if $\sum_{n=0}^{\infty} |z_n|$ convereges) then $\sum_{n=0}^{\infty} z_n$ converges.*

Proof. We argue that the sequence $s_n = \sum_{k=0}^{n} z_k$ of partial sums is Cauchy, which is enough to prove convergence of the sequence $\{s_n\}$.

Since $\sum_{n=0}^{\infty} |z_n|$ converges, the sequence $t_n = \sum_{k=0}^{n} |z_k|$ is Cauchy. Hence for any $\varepsilon > 0$ there is N so that for $m > n > N$

$$\begin{aligned}
\varepsilon > |t_m - t_n| &= \left| \sum_{k=0}^{m} |z_k| - \sum_{k=0}^{n} |z_k| \right| = ||z_m| + |z_{m-1}| + \cdots + |z_{n+1}|| \\
&\geq |z_m + z_{m-1} + \cdots + z_{n+1}| \qquad (\text{ triangle inequality and induction on } m-n) \\
&= \left| \sum_{k=0}^{m} z_k - \sum_{k=0}^{n} z_k \right| = |s_m - s_n|
\end{aligned}$$

and we are done. □

Remark 26.4. In general a series may converge but not converge absolutely. Here is an example. One can show (we won't) that the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ converges. Since $\sum_{n=1}^{\infty} \left| \frac{(-1)^n}{n} \right| = \sum_{n=1}^{\infty} \frac{1}{n}$ and since $\sum_{n=1}^{\infty} \frac{1}{n}$ does not converge (we did prove that!), $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ doesn't converge absolutely.

A necessary condition for a series to converge is for the terms to go to zero. More precisely:

Lemma 26.5. *If $\sum_{n=0}^{\infty} z_n$ converges, then $\{z_n\}$ converges to 0.*

Proof. If the series $\sum_{n=0}^{\infty} z_n$ converges then, by definition, the sequence of partial sums $s_n = \sum_{k=0}^{n} z_k$ converges. This implies that $\{s_n\}$ is Cauchy.

Therefore for all $\varepsilon > 0$, there exists N such that if $n, m > N$ then $|s_m - s_n| < \varepsilon$. Hence if $n > N$ then $|z_{n+1}| = |s_{n+1} - s_n| < \varepsilon$. Therefore for all $k > N+1$, $|z_k - 0| < \varepsilon$, i.e, $z_k \to 0$. □

Lemma 26.6 (Ratio test)**.** *Suppose $\{z_n\}$ is a sequence of nonzero complex numbers and $|\frac{z_{n+1}}{z_n}| \to \rho \in \mathbb{R}$ (so necessarily $\rho \geq 0$).*

- *If $\rho < 1$, then $\sum_{n=0}^{\infty} z_n$ converges (in fact converges absolutely).*
- *If $\rho > 1$, then $\sum_{n=0}^{\infty} z_n$ diverges.*

We'll prove the lemma next time. The proof will use the completeness of complex numbers: any Cauchy sequence of complex numbers converges. Here are some examples of what ratio test is good for.

Example 26.7. $\sum_{n=0}^{\infty} \frac{z_n}{n!}$ converges for all $z \in \mathbb{C}$.

Proof. If $z = 0$, $\sum_{n=0}^{\infty} \frac{0^n}{n!} = 1 + \frac{0^1}{1!} + \frac{0^2}{2!} \cdots$ converges (here $0! = 1$ and $0^0 = 1$ as well so $\frac{0^0}{0!} = 1$).

If $z \neq 0$,

$$\left| \frac{z^{n+1}}{(n+1)!} \Big/ \frac{z^n}{n!} \right| = \left| \frac{z^{n+1} n!}{z^n (n+1)!} \right| = \frac{|z|}{n+1} \to 0.$$

Since $0 < 1$, ratio test tells us that the series $\sum_{n=0}^{\infty} \frac{z^n}{n!}$ converges for all $z \neq 0$. □

One defines

$$e^z := \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

for all $z \in \mathbb{C}$.

Example 26.8. $\cos z := \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n}}{(2n)!}$ converges for all $z \in \mathbb{C}$.

Proof. If $z = 0$, then just as in the case of e^z, $\cos z = 1$. Now suppose $z \neq 0$. Then

$$\left| \frac{(-1)^{n+1} z^{2(n+1)}}{((2(n+1))!} \Big/ \frac{(-1)^n z^{2n}}{(2n)!} \right| = \left| \frac{z^{2n+2}}{(2n+2)!} \cdot \frac{(2n)!}{z^{2n}} \right| = \frac{|z^2|}{(2n+2)(2n+1)}.$$

For each $z \in \mathbb{C}$,

$$\lim_{n \to \infty} \frac{|z^2|}{(2n+2)(2n+1)} = 0.$$

Ratio test tells us that the series converges. □

The next two examples show that if $|\frac{z_{n+1}}{z_n}| \to 1$ then ratio test tells us nothing. We know that $\sum_{n=0}^{\infty} \frac{1}{n}$ diverges (since the sequence of partial sums is not Cauchy) and

$$\left| \frac{1}{n+1} \Big/ \frac{1}{n} \right| = \frac{n}{n+1} \to 1.$$

On the other hand one can show that the series $\sum_{n=0}^{\infty} \frac{1}{n^2}$ converges. Again

$$\left| \frac{1}{(n+1)^2} \Big/ \frac{1}{n^2} \right| = \frac{n^2}{(n+1)^2} \to 1.$$

Lecture 27: Proof of ratio test, uniqueness of real numbers

Last time: Proved the comparison test. Proved that if a series converges then the terms have to go to 0. Defined absolute convergence of series and proved that the series that converge absolutely converge. Stated but didn't prove the ratio test:

Ratio test Suppose $\{z_n\}$ is a sequence of nonzero complex numbers and $|\frac{z_{n+1}}{z_n}| \to \rho \in \mathbb{R}$ (so necessarily $\rho \geq 0$).

- If $\rho < 1$, then $\sum_{n=0}^{\infty} z_n$ converges (in fact converges absolutely).
- If $\rho > 1$, then $\sum_{n=0}^{\infty} z_n$ diverges.
- If $\rho = 1$ then the series $\sum_{n=0}^{\infty} z_n$ may or may not converge.

Proof. Suppose $\rho > 1$. Then for any $\varepsilon > 0$ there is N so that for $n \geq N$

$$\left| \left| \frac{z_{n+1}}{z_n} \right| - \rho \right| < \varepsilon, \qquad \text{that is} \qquad \rho - \varepsilon < \left| \frac{z_{n+1}}{z_n} \right| < \rho + \varepsilon$$

Choose ε so that $\rho - \varepsilon > 1$ (for example we may take $\varepsilon = \frac{\rho - 1}{2}$). Then for any n with $n \geq N$,

$$1 < \left| \frac{z_{n+1}}{z_n} \right| \qquad \text{hence } |z_n| < |z_{n+1}|.$$

Consequently

$$|z_N| < |z_{N+1}| < |z_{N+2}| < \cdots < |z_{N+k}| < \cdots$$

for all $k \geq 1$. It follows that the sequence $\{z_n\}$ does not converge to 0. Hence by Lemma 26.5 the series $\sum_{n=0}^{\infty} z_n$ does not converge.

Now suppose $\rho < 1$. Choose $\varepsilon > 0$ so that $r := \rho + \varepsilon < 1$ (for example $\varepsilon = \frac{1-\rho}{2}$ works). Then for $n \geq N$

$$\left|\frac{z_{n+1}}{z_n}\right| < \rho + \epsilon = r < 1.$$

Hence $|z_{N+1}| < r|z_N|$. Moreover $\quad |z_{N+2}| < r|z_{N+1}| < r^2|z_N|$

$|z_{N+3}| < r|z_{N+2}| < r^3|z_N|$

$\vdots$

$|z_{N+k}| < r^k|z_N|$ for all $k \geq 1$.

It follows that

$$\sum_{m=0}^{\ell} |z_{N+m}| < \sum_{m=0}^{\ell} r^m|z_N| = |z_N| \sum_{m=0}^{\ell} r^m$$

In the next homework you'll show that for all $0 \leq r < 1$ the geometric series $\sum_0^{\infty} r^k$ converges and equals $\frac{1}{1-r}$. Hence

$$\sum_{m=0}^{\ell} |z_{N+m}| < |z_N| \frac{1}{1-r}.$$

It follows that

$$\sum_{n=0}^{N+\ell} |z_n| = \sum_{n=0}^{N-1} |z_n| + \sum_{n=N}^{N+\ell} |z_n| = \sum_{n=0}^{N-1} |z_n| + \sum_{m=0}^{\ell} |z_{N+m}| < \sum_{n=0}^{N-1} |z_n| + |z_N| \frac{1}{1-r}.$$

Therefore the increasing sequence of real numbers $s_k = \sum_{n=0}^{k} |z_n|$ is bounded above. Hence $\{s_k\}$ converges, i.e. the series $\sum_{k=0}^{\infty} z_k$ converges absolutely. But absolute convergence implies convergence (Lemma 26.3) so we are done. □

We have some unfinished business to take care of: uniqueness of the reals. It's left as an exercise in Sally's book. We want to show the following:

> Let F be an ordered field with the least upper bound property. Then F is isomorphic to $\mathbb{R} = \mathcal{C}/I$ (where $\mathcal{C}$ is the ring of Cauchy sequences in $\mathbb{Q}$ and I is the ideal of sequences converging to 0). That is, there is a bijection $\varphi : \mathcal{C}/I \to F$ so that
>
> - φ is an homomorphism (i.e., φ preserves addition and multiplication)
> - $\varphi(\underline{1} + I) = 1_F$ and
> - φ preserves order: if $x < y$ then $\varphi(x) < \varphi(y)$.

To prove the uniqueness of reals we need some preparation.

Characteristic of a ring. Suppose R is a commutative ring. For each natural number n define $n \cdot 1_R$ by

$$n1_R := \overbrace{1_R + \cdots + 1_R}^{n}.$$

Then either there is a natural number $m > 1$ so that $m1_R = 0_R$ (for example if $R = \mathbb{Z}_m$) or $n1_R \neq 0$ for any $n \in \mathbb{N}$.

In the first case we define the characteristic of R to be

$$\operatorname{char} R := \min\{m \in \mathbb{N} \mid m1_R = 0\}.$$

In the second case ($n1_R \neq 0$ for any $n \in \mathbb{N}$) we define the characteristic of R to be 0: $\operatorname{char} R := 0$.

Example 27.1. $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{R}$ and $\mathbb{C}$ all have characteristic 0 while $\operatorname{char} \mathbb{Z}_n = n$.

Lemma 27.2. *If R is an ordered ring then the characteristic of R is 0.*

Proof. In an ordered ring R, $1_R = 1_R \cdot 1_R > 0$ and for any $x, y > 0$, $x + y > 0$. Now induction on n shows that $n1_R > 0$ for all $n \in \mathbb{N}$. In particular $n1_R \neq 0$ for any $n \in \mathbb{N}$. □

Lemma 27.3. *Suppose R is a commutative ring with $\operatorname{char} R = 0$. Then the function $f : \mathbb{Z} \to R$ defined by*

$$f(n) = \begin{cases} \overbrace{1_R + \cdots + 1_R}^{n} & n > 1 \\ 0 & n = 0 \\ \underbrace{(-1_R) + \cdots + (-1_R)}_{|n|} & n < 0 \end{cases}$$

is an injective homomorphism.

Sketch of proof. Checking that f preserves $+$ requires looking at various cases. For example if $n, m > 0$

$$f(n+m) = \overbrace{1_R + \cdots + 1_R}^{n+m} = (\overbrace{1_R + \cdots + 1_R}^{n}) + (\overbrace{1_R + \cdots + 1_R}^{m}) = f(n) + f(m).$$

Similarly if $n, m > 0$

$$\begin{aligned} f(n \cdot m) &= \overbrace{1_R + \cdots + 1_R}^{n \cdot m} = \overbrace{(\underbrace{1_R + \cdots + 1_R}_{n}) + \cdots + (\underbrace{1_R + \cdots + 1_R}_{n})}^{m} \\ &= \overbrace{1_R \cdot n1_R + \cdots + 1_R \cdot n1_R}^{m} \\ &= (\overbrace{1_R + \cdots + 1_R}^{m}) \cdot n1_R = (m1_R) \cdot (n1_R) = f(m) \cdot f(n). \end{aligned}$$

It remains to check that f is injective. In Homework 8 problem 6 you proved that a homomorphism $\varphi : R \to S$ is 1-1 if and only if $\ker \varphi = \{0\}$. Therefore if f is not 1-1 there is $n \in \mathbb{Z}$, $n \neq 0$ so that $0 = f(n) = n \cdot 1_R$. If $n > 0$ this cannot happen since $\operatorname{char} R = 0$. If $n < 0$ then $-n > 0$ and $f(-n) = -f(n) = -0 = 0$, which again cannot happen. Therefore f is 1-1. □

Lemma 27.4. *Let F be a field, $f : \mathbb{Z} \to F$ an injective ring homomorphism (with $f(1) = 1_F$). Then*

$$\tilde{f} : \mathbb{Q} \to F, \quad \tilde{f}(p/q) := f(p)f(q)^{-1}$$

is a well-defined injective ring homomorphism. That is, F contains a copy of the rationals $\mathbb{Q}$. Thus all fields of characteristic 0, in particular all ordered fields, contain a copy of the rationals.

Proof. Suppose $\frac{p}{q} = \frac{p'}{q'} \in \mathbb{Q}$. Then $pq' = p'q$ and consequently

$$f(p)f(q') = f(pq') = f(p'q) = f(p')f(q) \quad \Longrightarrow \quad f(p)f(q)^{-1} = f(p')f(q')^{-1}.$$

Thus $\tilde{f}$ is well-defined. Checking that $\tilde{f}$ is a homomorphism is not hard. For example

$$\tilde{f}\left(\frac{p}{q}\frac{r}{s}\right) = \tilde{f}\left(\frac{rp}{qs}\right) = f(rp)f(qs)^{-1} = f(r)f(p)(f(q)f(s))^{-1} = f(r)f(p)f(s)^{-1}f(q)^{-1}$$
$$= f(r)f(q)^{-1}f(r)f(s)^{-1} = \tilde{f}\left(\frac{p}{q}\right)\tilde{f}\left(\frac{r}{s}\right).$$

You should check that $\tilde{f}$ preserves $+$; it's a similar computation. Since $\ker f$ is an ideal and $\mathbb{Q}$ is a field either $\ker f = \mathbb{Q}$ or $\ker f = \{0\}$. But $\ker f$ cannot be all of $\mathbb{Q}$ since $f(1) = 1_F \neq 0$. Hence $\ker f = \{0\}$ and therefore f is injective. □

Lecture 28: Uniqueness of reals (up to isomorphism), first isomorphism theorem

Last time: Proved the ratio test. Defined characteristic of a ring. Proved that any ordered ring has characteristic 0. Proved that if F is a field of characteristic 0 then there is an injective ring homomorphism $f : \mathbb{Q} \to F$ so that for $p, q \in \mathbb{N}$

$$f\left(\frac{p}{q}\right) = \overbrace{1_F + \cdots + 1_F}^{p} \cdot \left(\overbrace{1_F + \cdots + 1_F}^{q}\right)^{-1}.$$

We repeatedly used a fact from the homework: a homomorphism $\varphi : R \to S$ is injective if and only if $\ker \varphi = \{0\}$.

Recall our goal: any two ordered fields with the least upper bound property are isomorphic.

The following abstract algebra result is extremely useful:

Theorem 28.1 (First Isomorphism Theorem)**.** *Let $f : R \to S$ be a surjective ring homomorphism, $I = \ker f$. Then the map $\bar{f} : R/I \to S$ given by*

$$\bar{f}(r + I) = f(r)$$

is a well defined isomorphism.

Proof. Recall first that the kernel I of f is an ideal, so the quotient ring R/I makes sense. To prove that $\bar{f}$ is well-defined, we need to check that if $a + I = a' + I$ then $f(a) = f(a')$.

So suppose $a + I = a' + I$. Then $a - a' \in I = \ker f$. Hence $0 = f(a - a')$. Therefore

$$0 = f(a) - f(a').$$

We conclude that if $a + I = a' + I$, then $f(a) = f(a')$. Therefore $\bar{f}$ is well-defined.

We next check that $\bar{f}$ preserves $+$ and $\cdot$. For any $a + I, b + I \in R/I$,

$$\bar{f}((a + I) + (b + I)) = \bar{f}((a + b) + I) = f(a + b) = f(a) + f(b) = \bar{f}(a + I) + \bar{f}(b + I).$$

Similarly

$$\bar{f}((a + I) \cdot (b + I)) = \bar{f}(ab + I) = f(ab) = f(a)f(b) = \bar{f}(a + I) \cdot \bar{f}(b + I).$$

Therefore $\bar{f}$ is a homomorphism.

Since f is onto, for any $s \in S$ there is $a \in R$ so that $s = f(a)$. Since $\bar{f}(a + I) = f(a)$ it follows that $s = \bar{f}(a + I)$. Therefore $\bar{f}$ is onto.

Finally we check that $\bar{f}$ is injective. For this it's enough to show that $\ker(\bar{f}) = 0_{R/I} = 0 + I = I$. Now

$$a + I \in \ker(\bar{f}) \iff 0 = \bar{f}(a + I) = f(a) \iff a \in \ker f = I.$$

Therefore $\ker(\bar{f}) = \{a + I \mid a \in I\} = I$. Thus $\bar{f}$ is injective.

We conclude that $\bar{f}$ is a well-defined isomorphism of rings. □

Remark 28.2. Note that if additionally $f(1_R) = 1_S$ then $\bar{f}(1_R + I) = 1_S$.

We are now in position to prove that the real numbers are the uniqe ordered field with the least upper bound property.

Theorem 28.3. *Let F be an ordered field with the least upper bound property (equivalently, any Cauchy sequence in F has a limit). Then F is isomorphic, as an ordered field, to $\mathcal{C}/I$ where as before $\mathcal{C}$ is the ring of all Cauchy sequence in the rationals $\mathbb{Q}$ and I is the ideal of sequences that converge to 0.*

Proof. Since F is an ordered field, it has characteristic 0, hence contains a copy of $\mathbb{Q}$. It will be convenient for us to think of $\mathbb{Q}$ as a subset of F.

Since F has the least upper bound property any Cauchy sequence in $\mathbb{Q}$ contains to something in F. (We proved this by showing that any sequence has a monotone subsequence, that monotone bounded sequences converge, and that for a Cauchy sequence $\{a_n\}$ if any subsequence $\{a_{n_k}\}$ converges to L then $\{a_n\}$ also converges to L.)

Consider the function $f : \mathcal{C} \to F$ defined by

$$f(\{a_n\}) := \lim a_n.$$

Since $\lim a_n b_n = \lim a_n \cdot \lim b_n$ and since $\lim(a_n + b_n) = \lim a_n + \lim b_n$, the function f preserves $+$ and $\cdot$. Note also that $1_{\mathcal{C}}$ is the constant sequence $\underline{1}$ and that $\lim \underline{1} = 1$. Hence f preserves 1.

The homomorphism f is onto since for any $x \in F$ there is a sequence $\{a_n\}$ of rational numbers with $a_n \to x$.

Since $I = \ker f = \{\{a_n\} \mid \lim a_n = 0\} \neq \{0\}$, the homomorphism f is not injective. By the first isomorphism theorem the homomorphism

$$\bar{f} : \mathcal{C}/I \to F, \qquad \bar{f}(\{a_n\} + I) = f(\{a_n\}) = \lim a_n$$

is an isomorphism, i.e., the field F is isomorphic to $\mathcal{C}/I = \mathbb{R}$.

It remains to check that $\bar{f}$ preserves order. Suppose $\mathbb{R} \ni \{a_n\} + I > 0$. By definition of $>$ on $\mathbb{R}$ this means that there is $M \in \mathbb{N}$ and $d \in \mathbb{Q}$ with $d > 0$ so that $a_n > d$ for all $n > M$. But this implies that

$$0 < d \leq \lim a_n = \bar{f}(\{a_n\} + I).$$

Therefore $\bar{f}$ is order preserving and we are done. □

Remark 28.4. In the first isomorphism theorem the assumption that $f : R \to S$ is surjective is not essential.

If f is not surjective, we still get an well-defined injective homomorphism $\bar{f} : R/\ker f \to S$. The proof is the same as in the proof of the first isomorphism theorem. The image of $\bar{f}$ is the image $f(R)$ of f. Note that $f(R)$ is a subring of S. So even if f is not surjective, $\bar{f} : R/I \to f(R)$ is an isomorphism of rings.

Remark 28.5. For any ring R and any ideal $I \subset R$, we have the canonical surjective homomorphism

$$\pi : R \to R/I, \qquad \pi(a) = a + I.$$

Therefore if $f : R \to S$ is a homomorphism, $I = \ker f$ and $\bar{f} : R/I \to S$ the induced homomorphism, then the equation

$$\bar{f}(a + I) = f(a)$$

can also be written as

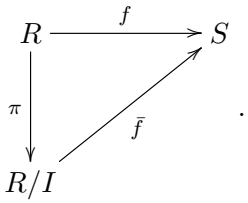
$$\bar{f}(\pi(a)) = f(a).$$

or, equivalently, as

$$\bar{f} \circ \pi = f.$$

The last equation can be visualized as a commuting diagram

(28.9)



.

Since $\pi : R \to R/I$ is surjective there is only one homomorphism from R/I to S making the diagram (28.9) commute. This is because if $h : R/I \to S$ is another homomorphism so that

$$h \circ \pi = f,$$

then for any $a + I \in R/I$,

$$h(a + I) = h(\pi(a)) = (h \circ \pi) = f(a) = \bar{f}(a + I),$$

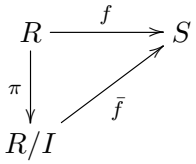
and therefore h has to equal $\bar{f}$.

Lecture 29: Characteristics of integral domains, polynomial rings.

Last time: Proved that any ordered field with the least upper bound property is isomorphic to $\mathcal{C}/I$, hence any two ordered fields with the least upper bound property are isomorphic.

The proof used

First Isomorphism theorem Let $f : R \to S$ be a homomorphism between two rings, $I = \ker f$. There exists a unique injective homomorphism $\bar{f} : R/I \to S$ with $\bar{f}(a + I) = f(a)$ for all $a \in R$. That is, the diagram



commutes. Here $\pi : R \to R/I$, $\pi(a) = a + I$ is the canonical projection.

Moreover $\bar{f} : R/I \to f(R)$ is an isomorphism. Here $f(R) = \{f(a) \mid a \in R\}$ is the image of R.

Characteristic of integral domains. Recall that a nonzero element a of a ring R is a zero divisor if there is $b \in R$ so that $b \neq 0$ and $ab = 0$.

For example $[2], [3] \in \mathbb{Z}_6$ are zero divisors since $[2], [3] \neq [0]$ and

$$[2][3] = [6] = [0].$$

Recall that a commutative ring R is an integral domain iff R has no zero divisors. We saw in lecture 3 that integral domains have a cancellation property: if R is an integral domain, $a, b, c \in R$, $a \neq 0$ then

$$ab = ac \implies b = c.$$

We have also seen that any field is an integral domain.

We have shown in lecture 27 that if R is a commutative ring and $\operatorname{char} R = 0$ then R contains a copy of $\mathbb{Z}$.

Lemma 29.1. *Let R be a commutative ring. Suppose* $\operatorname{char} R = n \neq 0$. *Then R contains a copy of $\mathbb{Z}_n$. Moreover if R is an integral domain then n has to be* prime.

Proof. We still have a homomorphism $f : \mathbb{Z} \to R$ so that

$$f(m) = \overbrace{1_R + \cdots + 1_R}^{m} \equiv m \cdot 1_R$$

for any $m > 0$ (and a similar formula if $m \leq 0$). The homomorphism f no longer injective since $f(n) = 0$ where

$$n = \min\{k \in \mathbb{N} \mid k \cdot 1_R = 0\} = \operatorname{char} R > 0.$$

We know that $I = \ker f$ is an ideal in $\mathbb{Z}$. You have proved in the homework that any ideal J of $\mathbb{Z}$ has to be of the form $d\mathbb{Z}$ where

$$d = \min\{k \in \mathbb{N} \mid k \in I\}.$$

Therefore

$$\ker f = n\mathbb{Z}, \qquad n = \operatorname{char} R.$$

The first isomorphism theorem tells us that $f : \mathbb{Z} \to R$ induces an isomorphism $\bar{f} : \mathbb{Z}/\ker f = \mathbb{Z}/n\mathbb{Z} \to f(Z)$. Therefore if $\operatorname{char} R = n$ then R contains a copy of $\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n$.

If R is an integral domain, then any subring of R is an integral domain as well. Therefore $\mathbb{Z}_n$ has to be an integral domain. Suppose n is not prime. Then n can be factored as $n = m\ell$, $1 < m, \ell < n$. Since $[0] = [n] = [m][\ell]$ in $\mathbb{Z}_n$, $[m]$, $[\ell]$ are zero divisors, which is impossible. Therefore $n = \operatorname{char} R$ has to be prime if R is an integral domain. □

Our next goal is to construct the field $\mathbb{C}$ of complex numbers as a quotient of the ring $\mathbb{R}[x]$ of polynomials with real coefficients. More specifically we'll prove that

$$\mathbb{C} = \mathbb{R}[x]/(x^2+1)$$

where (x^2+1) is the ideal in $\mathbb{R}[x]$ consisting of all multiples of x^2+1:

$$(x^2+1) := \{(x^2+1)q(x) \mid q(x) \in \mathbb{R}[x]\} = (x^2+1)\mathbb{R}[x].$$

In order to carry this out we need to study polynomial rings more systematically.

Let S be a commutative ring with 1. For example $S = \mathbb{Z}, \mathbb{Z}_n, \mathbb{Q}, \mathbb{R}, \mathbb{C}, \ldots...$ We define

$$S[x] = \{a_0 + a_1 x + \cdots + a_n x^n \mid n \geq 0, a_0, \ldots, a_n \in S\}.$$

The set $S[x]$ of polynomials with the coefficients in the ring S can be made into a commutative ring (with 1) as follows. We define addition on $S[x]$ by

$$\big(\sum a_k x^x\big) + \big(\sum b_k x^k\big) = \sum (a_k + b_k) x^k$$

and multiplication by

$$\big(\sum_{k=0}^{n} a_k x^k\big) \cdot \big(\sum_{\ell=0}^{m} b_\ell x^\ell\big) = \sum_{k=0}^{n+m} \left(\sum_{i+j=k} a_i b_j \right) x^k.$$

Caution: Every polynomial $p(x) = a_0 + a_1 + \cdots + a_n x^n \in S[x]$ defines a function

$$\text{“}p\text{”} : S \to S, \qquad \text{“}p\text{”}(s) = a_0 + a_1 s + \cdots + a_n s^n.$$

However *different* polynomials may define the *same* function.

Example 29.2. Let $S = \mathbb{Z}_2$. $p(x) = [1]x^2 + [1]x + [1]$, $q(x) = [1]$. Then

$$\text{“}p\text{”}([0]) = [1][0]^2 + [1][0] + [1] = [1] = \text{“}q\text{”}([0]),$$
$$\text{“}p\text{”}([1]) = [1][1]^1 + [1][1] + [1] = [1] = \text{“}q\text{”}([1]).$$

So “p” = “q” as functions. But they are different as polynomials, since they have different coefficients.

It is easy to see that there are a lot more polynomials with coefficients in $\mathbb{Z}_2$ than there are functions from $\mathbb{Z}_2$ to $\mathbb{Z}_2$. The number of functions from $\mathbb{Z}_2$ to $\mathbb{Z}_2$ is $2 \cdot 2 = 4$, while the ring of polynomials $\mathbb{Z}_2[x]$ is an infinite set: for any $n \in \mathbb{N}$, $[1]\cdot x^n$ is a polynomial in $\mathbb{Z}_2[x]$, and $[1]\cdot x^n \neq [1]x^m$ unless $n = m$!

Definition 29.3 (degree of a polynomial)**.** Given a polynomial $p(x) = a_0 + a_1 x + \cdots + a_n x^n \in S[x]$, S commutative ring, we define its degree $\deg p$ to be the largest integer k such that $a_k \neq 0$:

$$\deg p = \max\{k \in \mathbb{Z} \mid a_k \neq 0\}.$$

By convention the degree of the zero polynomial is $-\infty$. We'll see that there is a good reason for this convention.

Notation 29.4. We will abbreviate $1_S x^k$ as x^k for all $k \geq 1$.

Example 29.5. Consider $p(x) = 1 + 2x + 0x^2 + 1x^3 + 0x^4 = 1 + 2x + x^3 \in \mathbb{R}[x]$. The degree of $p(x)$ is 3.

Remark 29.6. For a polynomial $p(x) \in S[x]$, the degree of p is zero if and only if p is a **nonzero** constant polynomial: $p(x) = a_0$ for some $a_0 \in S$, $a_0 \neq 0$.

Lecture 30: Division algorithm for polynomial rings and applications

Lemma 30.1. *Let S be a commutative ring, $p, q \in S[x]$ two polynomials. Then*

(1) $\deg(p+q) \leq \max\{\deg p, \deg q\}$

(2) *If S is an integral domain (i.e., S has no zero divisors) then* $\deg(p \cdot q) = \deg(p) + \deg(q)$. *In general* $\deg(p \cdot q) \leq \deg(p) + \deg(q)$.

Sketch of proof. Suppose $\deg p = n$, $\deg q = m$. Then $p(x) = a_0 + \cdots + a_n x^n$, $q(x) = b_0 + \cdots + b_m x^m$ with $a_n \neq 0$, $b_m \neq 0$. We have

$$p(x) \cdot q(x) = a_n b_m x^{n+m} + \text{ lower order terms.}$$

If S is an integral domain then $a_n b_m \neq 0$ (since $a_n, b_m \neq 0$). Therefore

$$\deg(p \cdot q) = n + m = \deg p + \deg q.$$

Of course if $a_n b_m = 0$, $\deg(p \cdot q) < n + m$. Next note that

$$p(x) + q(x) = \sum_{i=0}^{\max(m,n)} (a_i + b_i) x^i$$

If $n = m$ it may happen that $a_n + b_m = a_n + b_n = 0$. In this case $deg(p+q) < n = \deg p = \deg q = \max(\deg p, \deg q)$. Otherwise $\deg(p+q) = \max(m, n) = \max(\deg p, \deg q)$. □

Recall that in order to prove that all ideals in the ring $\mathbb{Z}$ of integers are of the form $n\mathbb{Z}$ for some $n \in \mathbb{Z}$ we needed a division algorithm (for $\mathbb{Z}$). To prove the analogous result for polynomial rings we need the analogue of the division algorithm.

Theorem 30.2. *[Division Algorithm] Let F be a field, $f, g \in F[x]$ two polynomials with $g \neq 0$. Then there exist unique polynomials $q(x), r(x)$ with*

(1) $f = q \cdot g + r$
(2) $\deg r < \deg g$.

Existence. If $\deg f < \deg g$, $f = 0 \cdot g + f$. Next suppose $n = \deg f \geq \deg g$. We use induction on $n = \deg f$. Base case: $n = 0$. Then f, g are constant polynomials: $f = a_0 \in F$, $g = b_0 \in F$ and $b_0 \neq 0$ (otherwise $g = 0$). Then $a_0 = (a_0 {b_0}^{-1}) \cdot b_0 + 0$. Let $q = a_0 {b_0}^{-1}$, $r = 0$. We have

$$\deg r = \deg 0 = -\infty < \deg g = 0.$$

Now suppose theorem holds for all polynomials $l(x), s(x)$ with $n > \deg l \geq \deg s$ and suppose $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ with $a_n \neq 0$, $g(x) = b_0 + b_1 x + \cdots + b_m x^m$, $b_m \neq 0$. Consider

$$\begin{aligned} h(x) :=& f(x) - a_n {b_m}^{-1} x^{n-m} g(x) = (a_n x^n + \cdots + a_0) - a_n {b_m}^{-1} x^{n-m}(b_m x^m + \cdots + b_0) \\ =& (a_{n-1} - a_n b_m^{-1} b_{m-1}) x^{n-1} + \text{ lower order terms.} \end{aligned}$$

Then $\deg h \leq n - 1$. By inductive assumption, there exist $q_1(x), r_1(x) \in F[x]$ such that

(1) $h(x) = q_1(x) g(x) + r_1(x)$
(2) $\deg r_1 < \deg g$.

(1) implies that $f(x) - a_n b_m{}^{-1} x^{n-m} g(x) = q_1(x) \cdot g(x) + r_1(x)$. Hence $f(x) = (a_n b_m{}^{-1} x^{n-m} + q_1(x))g(x) + r_1(x)$.

(Uniqueness) Suppose $f = q_1 g + r_1 = q_2 g + r_2$ and $\deg r_1, \deg r_2 < \deg g$. Then $(q_1 - q_2)g = r_2 - r_1$.
If $q_1 - q_2 \neq 0$, then $\deg(q_1 - q_2) \geq 0$ and

$$\deg(q_1 - q_2)g = \deg(q_1 - q_2) + \deg g \geq \deg g,$$

while

$$\deg(r_2 - r_1) \leq \max(\deg r_2, \deg r_1) < \deg g.$$

Contradiction, since $\deg(r_2 - r_1) = \deg((q_1 - q_2)g)$. Therefore $q_1 - q_2 = 0$. Hence

$$r_2 - r_1 = (q_1 - q_2)g = 0 \cdot q = 0.$$

We conclude that $q_1 = q_2$ and $r_2 = r_1$. □

Here is is our first application of the division algorithm.

Lemma 30.3. *Let F be a field, $0 \neq g \in F[x]$ a polynomial of degree $n \geq 1$. Then the map*

$$\varphi : F^n \to F[x]/(g), \qquad \varphi(a_0, a_1, \ldots, a_{n-1}) := a_0 + a_1 x + \cdots + a_{n-1} x^{n-1} + (g)$$

is a bijection. Here and elsewhere (g) denotes the ideal consisting of multiples of g: $(g) = \{p(x)g(x) \mid p(x) \in F[x]\}$.

We'll not prove this lemma now and prove a better version next time.

Example 30.4. Let $F = \mathbb{R}$, the field of real numbers. Let $g(x) = x^2 + 1$. Then by Lemma 30.3 the map

$$\varphi : \mathbb{R}^2 \to \mathbb{R}[x]/(x^2 + 1), \qquad \varphi(a, b) = a + bx + (x^2 + 1)$$

is a bijection. The bijection φ turns $\mathbb{R}^2$ into a ring: for all $(a, b), (c, d) \in \mathbb{R}^2$ we have

$$(a + bx + (x^2 + 1)) + (c + dx + (x^2 + 1)) = (a + c) + (b + d)x + (x^2 + 1).$$

So we define $+$ on $\mathbb{R}^2$ so that φ would preserve it:

$$(a, b) + (c, d) := (a + c, b + d).$$

On the other hand

$$\begin{aligned}(a + bx + (x^2 + 1))(c + dx + (x^2 + 1)) =& (a + bx)(c + dx) + (x^2 + 1) = ac + (bc + ad)x + bdx^2 + (x^2 + 1) \\ =& ac + (bc + ad)x + bd(x^2 + 1) - bd + (x^2 + 1) \\ =& (ac - bd) + (bc + ad)x + (x^2 + 1).\end{aligned}$$

So if we want φ to preserve multiplication we should define a multiplication $\cdot$ on $\mathbb{R}^2$ by

$$(a, b) \cdot (c, d) := (ac - bd, bc + ad).$$

This is exactly what we did when we defined the ring of complex numbers $\mathbb{C}$. But now we know we are going to get a ring since $\varphi : \mathbb{R}^2 \to \mathbb{R}[x]/(x^2 + 1)$ is a bijection, $\mathbb{R}[x]/(x^2 + 1)$ is a ring, and φ preserves $+$ and $\cdot$.

Lecture 31: Vector spaces, linear maps and quotients of polynomial rings

Last time: proved the devision algorithm for polynomials with coefficients in a field F (the fact that F is a field is important): for any two polynomials $f, g \in F[x]$ with $g \neq 0$ there exists two unique polynomials q and r so that

(1) $f = q \cdot g + r$
and
(2) $\deg r < \deg g$.

Stated without proof: Let F be a field, $g \in F[x]$ a polynomial of degree $n \geq 1$. Then the map

$$\varphi : F^n \to F[x]/\langle g \rangle, \qquad \varphi(a_0, a_1, \ldots, a_{n-1}) := a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + \langle g \rangle$$

is a bijection. Here and from now on:

$$\langle g \rangle := gF[x] = \{gp \mid p \in F[x]\}.$$

We can do better because F^n is not just a set of n-tuples — it's a vector space over the field F. So is the quotient ring $F[x]/\langle g \rangle$. So what's a vector space over a field? We start by defining groups (which were mentioned briefly awhile back).

Definition 31.1. A group is a set G together with an associative binary operation ("multiplication")

$$\bullet : G \times G \to G, \qquad (a, b) \mapsto a \bullet b,$$

an element $e \in G$ (the "identity") so that

$$a \bullet e = a = e \bullet a$$

for all $a \in G$ and such that for all $a \in G$ there is $b \in G$ with

$$a \bullet b = e = b \bullet a.$$

Such b is called an inverse of a.

A group G is abelian if the operation $\bullet$ is commutative:

$$a \bullet c = c \bullet a$$

for all $a, c \in G$. If G is abelian the operation $\bullet$ is usually written as $+$, e is written as 0 and the inverse of a as $-a$.

Remark 31.2. (1) It's not hard to show that in a group inverses are unique.
(2) An abelian group G is a triple $(G, +, 0)$.

Example 31.3 (Really an "example")**.** If we start with a ring $(R, +, \cdot, 0, 1)$ and forget the multiplication $\cdot$ we get an abelian group $(R, +, 0)$.

Definition 31.4. Let F be a field. A vector space over F is an abelian group $(V, +, 0 = 0_V)$ together with a map

$$F \times V \to V, \qquad (\lambda, v) \mapsto \lambda v$$

called multiplication by scalars ("scalars" are elements of F) such that for all $\lambda, \mu \in F$, $v, w \in V$

- $\lambda(v + w) = (\lambda v) + (\lambda w)$
- $(\lambda \cdot \mu)v = \lambda(\mu v)$ (here $\cdot$ is the multiplication in F)
- $(\lambda + \mu)v = (\lambda v) + (\mu v)$ and
- $1_F v = v$.

Exercise 31.5. It's not hard to show that for all vectors v in a vector space V,

$$0_F v = 0_V$$

Hint: $0_F + 0_F = 0_F$. Therefore $0_F v = (0_F + 0_F)v = \ldots$

Example 31.6. The ring $F[x]$ of polynomials with coefficients in a field F is a vector space over F. The scalar multiplication is given by

$$\lambda \cdot (a_0 + a_1 x + \cdots + a_n x^n) = \lambda a_0 + \lambda a_1 x + \cdots + \lambda a_n x^n.$$

Example 31.7. For any $n \geq 1$ the set of n-tuples $F^n = \{(a_0, \ldots, a_{n-1}) \mid a_i \in F\}$ is a vector space over F: the addition and scalar multiplication are defined coordinate-wise:

$$\begin{aligned}(a_0, \ldots, a_{n-1}) + (b_0, \ldots, b_{n-1}) :=& (a_0 + b_0, \ldots, a_{n-1} + b_{n-1}) \\ \lambda \cdot (a_0, \ldots, a_{n-1}) :=& (\lambda a_0, \ldots, \lambda a_{n-1})\end{aligned}.$$

Example 31.8. For any field F and any ideal $I \subset F[x]$ the quotient ring $F[x]/I$ is a vector space over F. The vector addition comes from the ring addition of $F[x]/I$. The scalar multiplication comes from the ring multiplication too:

$$\lambda \cdot (p + I) = (\lambda + I)(p + I) = \lambda p + I$$

for all $\lambda \in F \subset F[x]$ and all $p + I \in F[x]/I$.

Definition 31.9. Let V and W be two vector spaces over a field F. A function $T : V \to W$ is linear if T preserves $+$ and scalar multiplication:

$$T(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 T(v_1) + \lambda_2 T(v_2)$$

for all $\lambda_1, \lambda_2 \in F$, $v_1, v_2 \in V$.

Example 31.10. Let V be the set of all differentiable functions on the real line and W the space of all function on $\mathbb{R}$. Both a vector spaces over $\mathbb{R}$ since one can add function and multiply them by scalars. The derivative

$$\frac{d}{dx} : V \to W$$

is a linear map.

Remark 31.11. Just as in the case of homomorphisms, a linear map $T : V \to W$ always sends $0 \in V$ to $0 \in W$. Reason:

$$T(0) = T(0 + 0) = T(0) + T(0).$$

Now add $-T(0)$ to both sides. We get $0 = T(0)$.

Remark 31.12. Let V be a vector space, $v \in V$. Then

$$(-1)v + v = (-1)v + 1v = ((-1) + 1)v = 0v = 0_V.$$

Hence $(-1)v = -v$.

Definition 31.13. A linear map $T : V \to W$ between two vector spaces over a field F is an isomorphism if there is a linear map $S : W \to V$ so that $S \circ T = \mathrm{id}_V$, $T \circ S = \mathrm{id}_W$.

Exercise 31.14. A linear map $T : V \to W$ is an isomorphism if and only if T is a bijection.

Definition 31.15. The kernel (also called the null space) of a linear map $T : V \to W$ is the set

$$\ker T := \{v \in V \mid T(v) = 0\}.$$

Lemma 31.16. *A linear map $T : V \to W$ is 1-1 if and only if $\ker T = \{0\}$.*

Proof. Now suppose $T : V \to W$ is 1-1 and $v \in \ker T$. Then $T(v) = 0 = T(0)$. Since T is 1-1, $v = 0$. Therefore $\ker T = \{0\}$.

Suppose $\ker T = \{0\}$ and $T(v) = T(v')$ for some $v, v' \in V$. Then

$$0 = T(v) - T(v') = T(v) + (-1)T(v') = T(v + (-1)v').$$

Hence $\ker T = \{0\} \ni v + (-1)v' = v - v'$ and consequently $v = v'$. Therefore T is 1-1. □

Lemma 31.17. *Let F be a field, $g \in F[x]$ a polynomial of degree $n \geq 1$. Then the map*

$$\varphi : F^n \to F[x]/\langle g\rangle, \qquad \varphi(a_0, a_1, \ldots, a_{n-1}) := a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + \langle g\rangle$$

is an isomorphism of vector spaces over F.

Proof. We need to check that φ is 1-1 and onto. Given $p + \langle g\rangle \in F[x]/\langle g\rangle$ there are, by the division algorithm, two unique polynomials $q, r \in F[x]$ so that

$$p = qg + r \qquad \text{and} \qquad \deg r < \deg g = n.$$

Then $r = a_0 + a_1 x + \cdots a_{n-1}x^{n-1}$ for some $(a_0, \ldots, a_{n-1}) \in F^n$ and

$$p - r = qg \in \langle g\rangle.$$

Therefore

$$p + \langle g\rangle = r + \langle g\rangle = \varphi(a_0, \ldots, a_{n-1}).$$

Hence φ is onto.

To check that φ is 1-1 we compute its kernel.

$$\begin{aligned}\ker\varphi &= \{(a_0, \ldots, a_{n-1}) \in F^n \mid \varphi(a_0, \ldots, a_{n-1}) = 0 + \langle g\rangle\} \\ &= \{(a_0, \ldots, a_{n-1}) \in F^n \mid a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} \in \langle g\rangle.\end{aligned}$$

Now $a_0 + a_1 x + \cdots a_{n-1}x^{n-1} \in \langle g\rangle$ if and only if $a_0 + a_1 x + \cdots a_{n-1}x^{n-1} = q(x)g(x)$ for some $q(x) \in F[x]$ which can only happen if

$$\begin{aligned} n - 1 \geq \deg(a_0 + a_1 x + \cdots a_{n-1}x^{n-1}) &= \deg q + \deg g = \deg q + n \\ \iff \quad -1 &\geq \deg q \\ \iff \quad \deg q &= -\infty \quad \text{i.e., } q = 0 \\ \iff a_0 + a_1 x + \cdots a_{n-1}x^{n-1} &= 0g \quad \text{ is the zero polynomial} \\ \iff (a_0, \ldots, a_{n-1}) &= (0, \ldots, 0).\end{aligned}$$

Hence $\ker\varphi = \{(0, \ldots, 0)\}$ and φ is injective. □

Lecture 32: Metrics, norms, Cauchy-Schwarz

We have seen that the absolute value function from $\mathbb{C}$ to $[0, \infty)$ defines a distance function $d : \mathbb{C} \times \mathbb{C} \to [0, \infty)$ by

$$d(z, w) = |z - w| = \sqrt{\mathrm{Re}\,(z - w)^2 + \mathrm{Im}\,(z - w)^2}$$

for all $a, b \in \mathbb{C}$. We now turn the properties of d into a definition.

Definition 32.1. Let S be a set. A function $d : S \times S \to [0, \infty)$ is a metric (a distance) iff for all $x, y, z \in S$

(1) $d(x, y) = d(y, x)$ (symmety)
(2) $d(x, y) = 0$ if and only if $x = y$ (nondegenracy) and
(3) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inquality).

Example 32.2. $S = \mathbb{R}$, $d(x, y) = |x - y|$.
$S = \mathbb{C} = \mathbb{R}^2$, $d(z, w) = |z - w| = \sqrt{\mathrm{Re}\,(z - w)^2 + \mathrm{Im}\,(z - w)^2}$.

Example 32.3. $S = \mathbb{R}^2$, $d((x_1, x_2), (y_1, y_2)) := |x_1 - y_1| + |x_2 - y_2|$. We need to check that $d : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty)$ is a metric.

Symmetry is easy since $|a - b| = |b - a|$ for any $a, b \in \mathbb{R}$.

d is nondegenerate because

$$\begin{aligned} 0 = |x_1 - y_1| + |x_2 - y_2| &\iff 0 = |x_1 - y_1| = |x_2 - y_2| \\ &\iff (x_1, x_2) = (y_1, y_2). \end{aligned}$$

Finally we check the triangle inequality: for all $(x_1, x_2), (y_1, y_2), (z_1, z_2) \in \mathbb{R}^2$

$$\begin{aligned} d((x_1, x_2), (z_1, z_2)) = |x_1 - z_1| + |x_2 - z_2| \leq &(|x_1 - y_1| + |y_1 - z_1|) + (|x_2 - y_2| + |y_2 - z_2|) \\ = &(|x_1 - y_1| + |x_2 - y_2|) + (|y_1 - z_1| + |y_2 - z_2|) \\ = &d((x_1, x_2), (y_1, y_2)) + d(y_1, y_2), (z_1, z_2)). \end{aligned}$$

Example 32.4. Any (nonempty) set S has a "trivial" metric d. It is defined by

$$d(x, y) = \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}.$$

Check that this d is in fact a metric.

Definition 32.5. A metric space is a set S together with a metric $d : S \times S \to [0, \infty)$. In other words a metric space is a pair (S, d) where S is a set and d is a metric on S.

Remark 32.6. The same set S may have many different metrics. For example if $S = \mathbb{R}^2$ then

$$\begin{aligned} d_2((x_1, x_2), (y_1, y_2)) &:= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \\ d_1((x_1, x_2), (y_1, y_2)) &:= |x_1 - y_1| + |x_2 - y_2| \qquad \text{and} \\ d_{\text{triv}}((x_1, x_2), (y_1, y_2)) &:= \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases} \end{aligned}$$

are three different metrics.

We use metrics to define convergence of sequences.

Definition 32.7. A sequence in a set S is a function $a : \mathbb{N} \to S$, $n \mapsto a_n$. We write $\{a_n\} \subset S$ when $\{a_n\}$ is a sequence in S.

Definition 32.8. A sequence $\{a_n\}$ is a metric space (S, d) converges to $L \in S$ if and only if for any $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$ there is $N \in \mathbb{N}$ so that for any $n > N$

$$d(a_n, L) < \varepsilon.$$

Definition 32.9. A sequence $\{a_n\}$ is a metric space (S, d) is Cauchy if and only if for any $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$ there is $N \in \mathbb{N}$ so that for any $n, m > N$

$$d(a_n, a_m) < \varepsilon.$$

The definitions above include three familiar cases:

(1) $S = \mathbb{Q}$, $d(x, y) = |x - y|$ for all $x, y \in \mathbb{Q}$
(2) $S = \mathbb{R}$, $d(x, y) = |x - y|$ for all $x, y \in \mathbb{R}$.
(3) $S = \mathbb{C}$, $d(x, y) = |x - y|$ for all $x, y \in \mathbb{C}$.

Many metrics on vector spaces over $\mathbb{R}$ come from norms.

Definition 32.10. Let V be a vector space over $\mathbb{R}$ (for example $V = \mathbb{R}^n$). A norm on V is a function

$$|| \cdot || : V \to [0, \infty)$$

so that

(1) for all $v \in V$, $||v|| = 0$ if and only if $v = 0$;

(2) for all $v \in V$, all $\lambda \in \mathbb{R}$

$$||\lambda v|| = |\lambda|||v||;$$

(3) (triangle inequality) $||v + w|| \leq ||v|| + ||w||$.

Exercise 32.11. If $|| \cdot || : V \to [0, \infty)$ is a norm, then $d(v, w) := ||v - w||$ is a metric on V.

Recall that $\mathbb{R}^n$ comes with the standard inner product $\langle \cdot \rangle$ which is defined by

$$\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n$$

for all vectors $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$. Recall also $\langle x, x \rangle \geq 0$ for all x, that $\langle \cdot \rangle$ is symmetric:

$$\langle x, y \rangle = \langle y, x \rangle \qquad \text{for all } x, y \in \mathbb{R}^n.$$

The inner product is also linear in each slot. That is, for all $\lambda_1, \lambda_2 \in \mathbb{R}$, $v_1, v_2, w \in \mathbb{R}^n$

$$\langle \lambda_1 v_1 + \lambda_2 v_2, w \rangle = \lambda_1 \langle v_1, w \rangle + \lambda_2 \langle v_2, w \rangle$$

and

$$\langle v, \mu_1 w_1 + \mu_2 w_2 \rangle = \mu_1 \langle v, w_1 \rangle + \mu_2 \langle v, w_2 \rangle$$

for all $v, w_1, w_2 \in \mathbb{R}^n$, $\mu_1, \mu_2 \in \mathbb{R}$.

One uses the standard inner product to define a norm on $\mathbb{R}^n$ by

$$||v|| = (\langle v, v \rangle)^{1/2} = \left(\sum v_i^2 \right)^{1/2}$$

for all $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$. To prove that $|| \, ||$ is a norm, we need to prove an inequality.

Lemma 32.12 (Cauchy-Schwarz inequality)**.** *For all $x, y \in \mathbb{R}^n$*

$$|\langle x, y \rangle| \leq ||x|| \cdot ||y||. \tag{32.10}$$

Proof. If x or y are 0, (32.10) reduces to $0 \leq 0$, which is true. We now assume that x, y are both nonzero vectors. Consider the function $f : \mathbb{R} \to [0, \infty)$ defined by

$$f(t) = ||x + ty||^2 = \langle x + ty, x + ty \rangle.$$

We have

$$\begin{aligned} 0 \leq f(t) &= \langle x + ty, x + ty \rangle = \langle x, x + ty \rangle + t\langle y, x + ty \rangle \\ &= \langle x, x \rangle + t\langle x, y \rangle + t\langle y, x \rangle + t^2 \langle y, y \rangle = ||x||^2 + 2t\langle x, y \rangle + t^2 ||y||^2. \end{aligned}$$

The function $f(t)$ is nonnegative and is quadratic in t. Hence $f(t)$ has a unique minumum at $t_{\min}$. To compute $t_{\min}$ we set the derivative $f"(t) = 2\langle x, y \rangle + 2t||y||^2$ to 0 and solve for t. We get

$$t_{\min} = -\frac{\langle x, y \rangle}{||y||^2}$$

Since $f(t) \geq 0$ for all t,

$$\begin{aligned} 0 \leq f(t_{\min}) &= ||x||^2 - 2\frac{\langle x, y \rangle}{||y||^2}\langle x, y \rangle + \left(\frac{\langle x, y \rangle}{||y||^2} \right)^2 ||y||^2 \\ &= ||x||^2 - \frac{(\langle x, y \rangle)^2}{||y||^2}. \end{aligned}$$

Hence

$$||x||^2 \geq \frac{(\langle x, y \rangle)^2}{||y||^2}$$

and therefore

$$||x||||y|| \geq |\langle x, y \rangle|.$$

And we are done. □

Corollary 32.13. *The function* $||\cdot|| : \mathbb{R}^n \to [0,\infty)$, $||x|| = \langle x,x\rangle^{1/2} = \left(\sum x_i^2\right)^{1/2}$ *is a norm and the function*

$$d(x,y) = ||x-y|| = \left(\sum (x_i - y_i)^2\right)^{1/2}$$

is a metric on $\mathbb{R}^n$.

Proof. The only hard part is checking the triangle inequality for $||\cdot||$, which we'll now do: for all $x, y \in \mathbb{R}^n$

$$\begin{aligned} ||x+y||^2 &= \langle x+y, x+y\rangle = \langle x,x\rangle + \langle y,x\rangle + \langle x,y\rangle + \langle y,y\rangle \\ &= ||x||^2 + 2\langle x,y\rangle + ||y||^2 \le ||x||^2 + 2|\langle x,y\rangle| + ||y||^2 \\ &\le ||x||^2 + 2||x||||y|| + ||y||^2 \qquad \text{by Cauchy-Schwarz} \\ &= (||x|| + ||y||)^2 \end{aligned}$$

and consequently

$$||x+y|| \le ||x|| + ||y||.$$

□

Remark 32.14. The norm $||x|| = \left(\sum x_i^2\right)^{1/2}$ on $\mathbb{R}^n$ is called an L^2-norm or 2-norm for short. We'll denote it by $||x||_2$ to distinguish it from other norms.

Remark 32.15. There are two more very useful norms on $\mathbb{R}^n$. The L^∞ norm $||x||_\infty$ is defined by

$$||x||_\infty := \max\{|x_1|, |x_2|, \ldots, |x_n|\} \equiv \max_{1\le j\le n} |x_j|.$$

To make sure it's a norm we check the triangle inequality (the other two conditions are easy to check and are left as an exercise). For all $x, y \in \mathbb{R}^n$

$$||x||_\infty + ||y||_\infty = \max_{1\le j\le n} |x_j| + \max_{1\le j\le n} |y_j| \ge |x_i| + |y_i| \ge |x_i + y_i| \qquad \text{for all } i.$$

Hence

$$||x||_\infty + ||y||_\infty \ge \max_i |x_+ y_i| = ||x+y||_\infty.$$

The third useful norm on $\mathbb{R}^n$ is the L^1 norm defined by

$$||x||_1 := \sum_{i=1}^n |x_i|.$$

We check the triangle inequality:

$||x+y||_1 = \sum i = 1^n |x_i + y_i| \le \sum i = 1^n (|x_i| + |y_i|) = \sum i = 1^n |x_i| + \sum i = 1^n |y_i| = ||x||_1 + ||y||_1.$

Lecture 33: Equivalence of the three norms on $\mathbb{R}^n$, completeness and equivalent metrics

Last time: defined metrics and norms, proved Cauchy-Schwarz inequality. Noted that any norm $||\cdot||$ on a vector space V gives rise to a metric d:

$$d(x,y) = ||x-y||$$

for all $x, y \in V$.

Defined 3 norms on $\mathbb{R}^n$ and proved that they are norms:

$$||x||_2 := (\sum x_i^2)^{1/2}, \qquad ||x||_1 := (\sum |x_i|)^1 = \sum |x_i|, \qquad ||x||_\infty := \max_i |x_i|.$$

These three norms give rise to three metrics, which, in turn allow us to talk about convergence of sequences and Cauchy sequences.

Definition 33.1. Two norms $||\cdot||$ and $||\cdot||'$ on a real vector space V are equivalent if there are two *positive* constants $c_1, c_2 \in \mathbb{R}$ so that

$$c_1||x|| \leq ||x||' \leq c_2||x||$$

for all $x \in V$.

Two metrics d, d' on a set S are equivalent if there are two *positive* constants $c_1, c_2 \in \mathbb{R}$ so that

$$c_1 d(x,y) \leq d'(x,y) \leq c_2 d(x,y)$$

for all $x, y \in S$.

Exercise 33.2. Two equivalent norms define equivalent metrics.

Exercise 33.3. The relation of being equivalent (for norms and for metrics) is symmetric and transitive.

Theorem 33.4. *The three norms $||\cdot||_2$, $||\cdot||_1$ and $||\cdot||_\infty$ on $\mathbb{R}^n$ are equivalent: for any $x \in \mathbb{R}^n$*

$$\frac{1}{n}||x||_1 \leq ||x||_\infty \leq ||x||_2 \leq ||x||_1.$$

Proof. Fix $x \in \mathbb{R}^n$. By definition of $||x||_\infty$ there is an index j so that $||x||_\infty = |x_j|$. Then

$$||x||_\infty = |x_j| = \sqrt{x_j^2} \leq \sqrt{\sum x_i^2} = ||x||_2.$$

Next

$$(||x||_2)^2 = \sum |x_i|^2 \leq \left(\sum |x_i|\right)^2 = (||x||_1)^2.$$

Hence $||x||_2 \leq ||x||_1$. Finally

$$\frac{1}{n}||x||_1 = \frac{1}{n}\sum_i |x_i| \leq \frac{1}{n}\overbrace{(|x_j| + \cdots + |x_j|)}^{n} = \frac{1}{n} n||x||_\infty = ||x||_\infty.$$

□

Lemma 33.5. *Suppose $d, d' : S \times S \to [0, \infty)$ are two equivalent metrics on a set S. Then*

(a) *A sequence $\{a_n\} \subset S$ is Cauchy with respect to d $\iff$ it is Cauchy with respect to d'.*
(b) *A sequence $\{a_n\} \subset S$ converges to $L \in S$ with respect to d $\iff$ it converges to L with respect to d'.*

Proof. Since the metrics d, d' are equivalent there are positive constants c_1, c_2 so that

$$c_1 d(x,y) \leq d'(x,y) \leq c_2 d(x,y)$$

for all $x, y \in S$.

(a) Suppose $\{a_n\}$ is Cauchy with respect to d'. We want to show: for any $\varepsilon > 0$ there is $N \in \mathbb{N}$ so that $d(a_n, a_m) < \varepsilon$ for all $n, m > N$. Fix $\varepsilon > 0$. Since $\{a_n\}$ is Cauchy with respect to d' there is $N \in \mathbb{N}$ so that $d(a_n, a_m) < c_1\varepsilon$ for all $n, m > N$. But then

$$d(a_n, a_m) \leq \frac{1}{c_1} d'(a_n, a_m) \leq \frac{1}{c_1} c_1 \varepsilon = \varepsilon$$

for all $n, m > N$.

Similarly if $\{a_n\}$ is Cauchy with respect to d, it is Cauchy with respect to d'.

(b) Suppose $a_n \to L$ with respect to d'. Then for any $\varepsilon > 0$ there is $N \in \mathbb{N}$ so that

$$d'(a_n, L) < c_1\varepsilon.$$

for all $n > N$. Therefore

$$d(a_n, L) \leq \frac{1}{c_1} d'(a_n, L) \leq \frac{1}{c_1} c_1\varepsilon = \varepsilon$$

for all $n > N$. Hence $a_n \to L$ with respect to d. Similarly if $a_n \to L$ with respect to d then $a_n \to L$ with respect to d'. □

Definition 33.6. A metric space (S, d) is complete if any Cauchy sequence in (S, d) converges to some $L \in S$ (again with respect to the metric d).

We would like to show that $\mathbb{R}^n$ is complete with respect to the metrics corresponding to the three compatible norms $||\cdot||_2, ||\cdot||_1$ and $||\cdot||_\infty$. Our strategy is to show first that $\mathbb{R}^n$ is complete with respect to the metric d_1 that corresponds to $||\cdot||_1$. We'll then argue that if two metrics d, d' on a set S are equivalent and (S, d) is compelete then (S, d') is also complete.

Lemma 33.7. *For any $n > 0$ the set $\mathbb{R}^n$ is complete with respect to the metric d_1 defined by the 1-norm $||\cdot||_1$.*

Proof. Suppose $\{x^{(k)} = (x_1^{(k)}, \ldots, x_n^{(k)})\}$ is a Cauchy sequence in $(\mathbb{R}^n, d_1)$. We first argue that for each i, $1 \leq i \leq n$, the sequences $\{x_i^{(k)})\} \subset \mathbb{R}$ are Cauchy the respect to the standard metric $d(a, b) = |a - b|$.

Since $\{x^{(k)}\}$ is Cauchy, for any $\varepsilon > 0$ there is $N \in \mathbb{N}$ so that for $r, s > N$

$$\varepsilon > d_1(x^{(r)}, x^{(s)}) = \sum_{j=1}^{n} |x_j^{(r)} - x_j^{(s)}| \geq |x_i^{(r)} - x_i^{(s)}|.$$

Hence the sequences $\{x_i^{(k)})\} \subset \mathbb{R}$, $1 \leq i \leq n$, are Cauchy.

Since $\mathbb{R}$ is complete, for each i the sequence $\{x_i^{(k)}\}$ converges to some $L_i \in \mathbb{R}$. Let $L = (L_1, \ldots, L_n)$. Then given $\varepsilon > 0$ there exist $N_1, \ldots, N_n \in \mathbb{N}$ so that for $k > N_i$

$$|x_i^{(k)} - L_i| < \varepsilon/n.$$

Then for all $k > \max\{N_1, \ldots, N_n\}$

$$||x^{(k)} - L||_1 = \sum_i |x_i^{(k)} - L_i| < \overbrace{\varepsilon/n + \cdots + \varepsilon/n}^{n} = \varepsilon.$$

□

Lemma 33.8. *Let d, d' be two equivalent metrics on a set S. Then (S, d) is complete $\iff$ (S, d') is complete.*

Consequently $\mathbb{R}^n$ is complete with respect to d_2 and d_∞ (where $d_2(x, y) = ||x-y||_2$ and $d_\infty(x, y) = ||x - y||_\infty$).

Proof. Suppose (S, d) is complete and $\{a_n\}$ is a Cauchy sequence in (S, d'). Then by Lemma 33.5 (a) the sequence $\{a_n\}$ is a Cauchy sequence in (S, d). Since (S, d) is complete, $\{a_n\}$ converges to some $L \in S$ with respect to d. By Lemma 33.5 (b) $\{a_n\}$ converges to L with respect to d'. Hence (S, d') is complete with respect to d'.

The converse argument is the same. □

Definition 33.9. Let (S, d) be a metric space. An ε-ball centered at $x \in S$ also called an open ball centered at x of radius ε is the set

$$B_\varepsilon(x) := \{y \in S \mid d(x, y) < \varepsilon\}$$

consisting of points of distance strictly less than ε from x.

Open balls play an important in defining continuity, convergence and other so called "topological" property.

Example 33.10. If $S = \mathbb{R}^3$ and $d = d_2$ then $B_r(x)$ is the standard round ball in 3-space centered at x.

Exercise 33.11. Let $S = \mathbb{R}^2$. Draw $B_r(0)$ with respect to the metrics d_1 and d_∞. Hint: they are squares.

Example 33.12. Let $d_{\text{triv}} : S \times S \to [0, \infty)$ is the trivial metric,

$$d_{\text{triv}}(x, y) := \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}$$

for all $x, y \in S$. Then $B_\varepsilon(x) = \{x\}$ if $\varepsilon \leq 1$. On the other hand $B_\varepsilon(x) = S$ if $\varepsilon > 1$.

Lecture 34: Open sets in a metric space, topology

Last time:

- equivalent norms and metrics
- $\mathbb{R}^n$ is complete with respect to d_1, d_2, d_∞.
- Given a metrics space (S, d) an open ball of radius r centered at $x \in S$ is

$$B_r(x) := \{y \in S \mid d(x, y) < r\}.$$

Today we define and study an idea that is more fundamental than a metric space: a topological space.

Definition 34.1. Let (S, d) be a metric spaces. A subset $U \subset S$ is open if for any $x \in U$ there is $r > 0$ so that $B_r(x) \subset U$.

Remark 34.2. By definition the whole space S and the empty set $\varnothing$ are always open, no matter the metric.

Example 34.3. Consider $\mathbb{R}$ with the standard metric: $d(x, y) = |x - y|$. Then an open interval $(a, b) \subset \mathbb{R}$ is open: given $x \in (a, b)$ let $r = \min\{x - a, b - x\}$. Then $a = x - (x - a) < x - r$ and $x + r < x + (b - x) = b$ so

$$B_r(x) = (x - r, x + r) \subset (a, b).$$

Lemma 34.4. *For any metric space (S, d) open balls are open sets.*

Proof. Let (S, d) be a metric space, $x \in S$, $r > 0$. Suppose $y \in B_r(x)$. Then $0 < r - d(x, y)$. Set $\rho := r - d(x, y)$. Now suppose $z \in B_\rho(y)$. Then

$$d(x, z) \leq d(x, y) + d(y, z) < d(x, y) + \rho = d(x, y) + (r - d(x, y)) = r.$$

Hence $z \in B_r(x)$ and consequently $B_\rho(y) \subset B_r(x)$. Since y is arbitrary, $B_r(x)$ is open. □

Theorem 34.5. *Let (S, d) be a metric space.*

(i) *For any two open sets $O_1, O_2 \subset S$, the intersection $O_1 \cap O_2$ is open.*
(ii) *For any collection $\mathcal{A}$ of open subsets of S the union $\bigcup_{O \in \mathcal{A}} O$ is open.*

Proof. (i) Suppose $x \in O_1 \cap O_2$. Then there are $r_1, r_2 \geq 0$ so that $B_{r_1}(x) \subseteq O_1$, $B_{r_2}(x) \subseteq O_2$. Let $r = \min(r_1, r_2)$. Then $B_r(x) \subseteq O_1$ *and* $B_r(x) \subseteq O_2$. Hence $B_r(x) \subseteq O_1 \cap O_2$. Since x is arbitrary this shows that the intersection $O_1 \cap O_2$ is open

(ii) Suppose $x \in \bigcup_{O\in\mathcal{A}} O$. Then $x \in U$ for some $U \in \mathcal{A}$. Since U is open, there is $r > 0$ so that $B_r(x) \subseteq U (\subseteq \bigcup_{O\in\mathcal{A}} O)$. Therefore $\bigcup_{O\in\mathcal{A}} O$ is open. □

Corollary 34.6. *The intersection of any* finite *collection of open sets is open.*

Proof. Induction on the number of n of sets in the collection. Nothing to prove for $n = 1$. Suppose the intersection of any k open sets is open and $O_1, \ldots, O_{k+1}$ are open. Then $O_1 \cap \ldots \cap O_k$ is open (by the inductive assumption) and then

$$O_1 \cap \ldots \cap O_k \cap O_{k+1} = (O_1 \cap \ldots \cap O_k) \cap O_{k+1}$$

is open by Theorem 34.5(i). □

WARNING In general the intersection of infinitely many open sets need not be open. Here is an example: let $\{O_n = (-\frac{1}{n}, \frac{1}{n})\}_{n\in\mathbb{N}}$. It's a countable collection of open intervals in $\mathbb{R}$. But

$$\bigcap_n (-\frac{1}{n}, \frac{1}{n}) = \{0\},$$

and the singleton $\{0\}$ is not open in $\mathbb{R}$.

We now turn Theorem 34.5 into a definition.

Definition 34.7. A topology $\mathcal{T}$ on a set S is a collection of subsets of S (i.e., $\mathcal{T} \subseteq \mathcal{P}(S)$) so that

(1) $\varnothing, S \in \mathcal{T}$
(2) for any $O_1, O_2 \in \mathcal{T}$ we have $O_1 \cap O_2 \in \mathcal{T}$
(3) for any subset $\mathcal{A} \subset \mathcal{T}$ the union $\bigcup_{O\in\mathcal{A}} O$ is in $\mathcal{T}$.

Elements of $\mathcal{T}$ are called open sets.

With this defintion Theorem 34.5 can be restated as follows.

Theorem 34.8. *Let (S, d) be a metric space. Then the collection $\mathcal{T}_d \subseteq \mathcal{P}(S)$ of subsets of S given by*

$$O \in \mathcal{T}_d \iff \text{for any } x \in O \text{ there is } r > 0 \text{ so that } B_r(x) \subseteq O$$

is a topology on S, the topology induced by the metric *d.*

Definition 34.9. A topological space is a pair $(S, \mathcal{T})$ where S is a set and $\mathcal{T}$ a topology on S.

Example 34.10. Let S be a set. Then $\mathcal{T} = \{\varnothing, S\}$ is a topology on S called the trivial topology on S.

Example 34.11. Let S be a set and d the trivial metric ($d(x, y) = 1 \iff x \neq y$). Then for any $x \in S$

$$B_1(x) = \{x\}.$$

Hence singletons are open in the metric topology $\mathcal{T}_d$. But then any subset $Y \subseteq S$ is open, ie, is an element of $\mathcal{T}_d$. This is because

$$Y = \bigcup_{y\in Y} \{y\}$$

and singletons $\{y\}$ are open. Hence in this case $\mathcal{T}_d = \mathcal{P}(S)$, the set of all subsets of S.

Exercise 34.12. Let X be an infinite set. For example $X = \mathbb{R}$. Define $\mathcal{T} \subset \mathcal{P}(X)$ by

$$\mathcal{T} = \{\varnothing\} \cup \{Y \subset X \mid X \smallsetminus Y \text{ is finite } \}.$$

Show that $\mathcal{T}$ is a topology. That is, in this topology O is open iff $O = \varnothing$ or if its complement $X \smallsetminus O$ is finite. This topology $\mathcal{T}$ is called the cofinite topology on X.

Having a topology on a set is enough to define convergence of sequences.

Definition 34.13. Let $(S, \mathcal{T})$ be a topological space. A sequence $\{a_n\}$ converges to $L \in S$ if for any open set O containing L there is $N \in \mathbb{N}$ so that $a_n \in O$ for all $n \geq N$.

If the topology on S comes from a metric d then the convergence with respect to the topology $\mathcal{T}_d$ is equivalent to the convergence with respect to the metric d.

Lemma 34.14. *Let (S, d) be a metric space, $\mathcal{T}_d$ the corresponding topology $\{a_n\}$ a sequence in S and $L \in S$. Then*

$$a_n \to L \text{ with respect to the metric } d \quad \iff \quad a_n \to L \text{ with respect to the topology } \mathcal{T}_d.$$

Proof. ($\implies$) Suppose that $a_n \to L$ with respect to the metric d and O is an open set containing L. Since O is open, there is $\varepsilon > 0$ so that $B_\varepsilon(L) \subseteq O$. Since $a_n \to L$ with respect to d there is N so that for $n > N$, $d(a_n, L) < \varepsilon$. Then for $n > N$

$$a_n \in B_\varepsilon(L) \subset O.$$

($\impliedby$) Suppose $a_n \to L$ with respect to the topology $\mathcal{T}_d$. Consider $\varepsilon > 0$. Since the open ball $B_\varepsilon(L)$ is open, since $L \in B_\varepsilon(L)$ and since $a_n \to L$ with respect to $\mathcal{T}_d$ there is N so that for all $n > N$ we have $a_n \in B_\varepsilon(L)$. But then for $n > N$

$$d(a_n, L) < \varepsilon.$$

Hence $a_n \to L$ with respect to d. □

We now revisit the fact that if two metrics d, d' on a set S are equivalent then they have the same convergent sequences (with the same limits).

Lemma 34.15. *Let d, d' be two equivalent metrics on the set S: there are $c_1, c_2 > 0$ so that*

$$c_1 d(x, y) \leq d'(x, y) \leq c_2 d(x, y)$$

Then $\mathcal{T}_d = \mathcal{T}_{d'}$. That is, the two metrics give rise to the same open sets.

Proof. We first argue that for any $x \in S$ and any $r > 0$

$$B^{d'}_{c_1 r}(x) \subseteq B^d_r(x) \quad \text{and} \quad B^d_{r/c_2}(x) \subseteq B^{d'}_r(x).$$

Let $y \in B^{d'}_{c_1 r}(c)$. Then $d'(x, y) < c_1 r$. Since $c_1 d(x, y) \leq d'(x, y)$,

$$c_1 d(x, y) < c_1 r. \quad \implies d(x, y) < r \implies y \in B^d_r(x).$$

Hence $B^{d'}_{c_1 r}(c) \subseteq B^d_r(x)$.

Now suppose $y \in B^d_{r/c_2}(x)$. Then $d(x, y) < r/c_2$. Since $d'(x, y) < c_2 d(x, y)$ we have

$$d'(x, y) \leq c_2 d(x, y) < c_2 r/c_2 = r. \implies y \in B^{d'}_r(x).$$

Thus $B^d_{r/c_2}(x) \subseteq B^{d'}_r(x)$.

Now suppose $O \in \mathcal{T}_d$ and $x \in O$. Then there is $r > 0$ so that $B^d_r(x) \subseteq O$. Since $B^{d'}_{c_1 r}(x) \subseteq B^d_r(x)$, $B^{d'}_{c_1 r}(x) \subseteq O$. Since $x \in O$ is arbitrary, $O \in \mathcal{T}_{d'}$.

Suppose $O \in \mathcal{T}_{d'}$. Let $x \in O$. Then there is $r > 0$ so that $B^{d'}_r(x) \subseteq O$. Since $B^d_{r/c_2}(x) \subseteq B^{d'}_r(x)$, $B^d_{r/c_2}(x) \subseteq O$. Since $x \in O$ is arbitrary, $O \in \mathcal{T}_d$. □

Lecture 35: Continuity

Last time: open sets in metric spaces and the notion of a topology.

Equivalent metrics define the same topology.

Today: continuity; first for metric spaces then for general topological spaces.

Definition 35.1. Let (S,d), (S',d') be two metric spaces. A function $f:S\to S'$ is continuous at $x_0\in S$ if for any sequence $\{a_n\}$ in S with $a_n\to x_0$, $f(a_n)\to f(x_0)$ in S'.

Example 35.2. The function $f:\mathbb{R}\to\mathbb{R}$, $f(x)=\begin{cases} x\sin\frac{1}{x} & x\neq 0\\ 0 & x=0\end{cases}$ is continuous at 0 (where $\mathbb{R}$ is given the standard metric; $d(x,y)=|x-y|$).

Proof. Note that if $x\neq 0$ then $|f(x)|=|x\sin\frac{1}{x}|\leq|x|$. If $x=0$ then $|f(x)|=0\leq|x|$ as well. Now suppose $a_n\to 0$. Then for any $\varepsilon>0$ there is $N\in\mathbb{N}$ so that for $n>N$ we have $|a_n-0|<\varepsilon$. Therefore for $n>N$

$$|f(a_n)-f(0)|=|f(a_n)-0|\leq|a_n|\leq\varepsilon.$$

Hence $f(a_n)\to 0$ and therefore f is continuous at 0. □

Example 35.3. The function $f:\mathbb{R}\to\mathbb{R}$, $f(x)=\begin{cases}\sin\frac{1}{x} & x\neq 0\\ 0 & x=0\end{cases}$ is not continuous at 0.

Proof. Consider the sequence $a_n=\frac{1}{\pi/2+2\pi n}$. Since $0\leq a_n\leq\frac{1}{2\pi n}$ and $\frac{1}{2\pi n}\to 0$, $a_n\to 0$. On the other hand

$$f(a_n)=\sin\pi/2+2\pi n=1\not\to 0.$$

So f is not continuous at 0. □

Checking a function on all the possible sequences in order to check continuity can get pretty old. Fortunately there is another criterion.

Theorem 35.4. *A function $f:(S,d)\to(S',d')$ is continuous at $x_0\in S$ if and only if for any $\varepsilon>0$ there is $\delta>0$ so that*

$$d(x,x_0)<\delta \quad\Longrightarrow\quad d'(f(x),f(x_0))<\varepsilon.$$

Proof. ($\Longleftarrow$) Suppose for any $\varepsilon>0$ there is $\delta>0$ so that

$$d(x,x_0)<\delta \quad\Longrightarrow\quad d'(f(x),f(x_0))<\varepsilon. \tag{35.11}$$

Fix $\varepsilon>0$. Then there is $\delta>0$ so that (35.11) holds. Let $\{a_n\}$ be a sequence in S so that $a_n\to x_0$. Then there is $N\in\mathbb{N}$ so that for $n>N$ we have

$$d(a_n,x_0)<\delta.$$

Since (35.11) holds, $d'(f(a_n),f(x_0))<\varepsilon$ for all $n>N$. It follows that $f(a_n)\to f(x_0)$.

($\Longrightarrow$) Suppose for any sequence $\{a_n\}\subseteq S$ with $a_n\to x_0$ we have $f(a_n)\to f(x_0)$ but (35.11) fails: there is $\varepsilon_0>0$ so that for any $\delta>0$ there is $x(\delta)\in S$ with

$$d(x(\delta),x_0)<\delta \quad\text{and}\quad d'(f(x(\delta)),f(x_0))\geq\varepsilon_0.$$

Define a sequence $\{a_n\}$ in S by $a_n:=x(\frac{1}{n})$. Then

$$d(a_n,x_0)=d(x(\frac{1}{n}),x_0)<\frac{1}{n}$$

for all n and consequently $a_n\to x_0$. On the other hand, by definition of $x(\delta)$,

$$d'(f(a_n),f(x_0))=d'(f(x(\frac{1}{n})),f(x_0))\geq\varepsilon_0$$

and consequently $f(a_n) \not\to f(x_0)$. Contradiction. Therefore for any ε there is δ so that (35.11) holds. □

Theorem 35.4 shows that we could equivalently define continuity of a function $f : (S, d) \to (S', d')$ as follows (this is the so called ε-δ definition of continuity).

Definition 35.5. A function $f : (S, d) \to (S', d')$ is continuous at a point $x_0 \in S$ if for every $\varepsilon > 0$ there is $\delta > 0$ so that

$$d(x, x_0) < \delta \quad \Longrightarrow \quad d'(f(x), f(x_0)) < \varepsilon.$$

Remark 35.6. In one variable calculus books and in beginning analysis books one often sees the following definition of continuity of a function $f : \mathbb{R} \to \mathbb{R}$ at a point $x_0 \in \mathbb{R}$: for every $\varepsilon > 0$ there is $\delta > 0$ so that

$$|x - x_0| < \delta \quad \Longrightarrow \quad |f(x) - f(x_0)| < \varepsilon.$$

This is a special case of Definition 35.5 with $S = S' = \mathbb{R}$ and $d = d' =$ the standard metric: $d(x, y) = |x - y|$.

Example 35.7. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$. It doesn't look continuous at any point $x_0 \in \mathbb{R}$. Let's check it using the second definition (Definition 35.5).

Proof. Choose $\varepsilon = 1/2$. Then for any $\delta > 0$, there is $x_1 \in \mathbb{Q}$, $x_2 \notin \mathbb{Q}$ so that $|x_1 - x_0| < \delta$ and $|x_2 - x_0| < \delta$. If x_0 is rational then $f(x_0) = 1$. But then $|f(x_0 - f(x_2)| = |1 - 0| > \frac{1}{2}$. If x_0 is not rational, then $f(x_0) = 0$ and then $|f(x_0) - f(x_1)| = |0 - 1| > \frac{1}{2}$ again. So f is not continuous at x_0. □

Definition 35.8. A function $f : (S, d) \to (S', d')$ between two metric spaces is continuous if and only f is continuous at every point $x_0 \in S$.

We would like to reformulate the definition of continuity in terms of topologies. The following theorem is then essential:

Theorem 35.9. *A function $f : (S, d) \to (S', d')$ is continuous $\iff$ for any open subset U of S' the preimage $f^{-1}(U)$ is open in S.*

Proof. ($\Longrightarrow$) Suppose $f : (S, d) \to (S', d')$ is continuous in the sense of Definition 35.5 and $U \subset S'$ is open. We'd like to show that $f^{-1}(U)$ is open in S. Pick $x_0 \in f^{-1}(U)$. Then $f(x_0) \in U$. Since U is open there is $\varepsilon > 0$ so that $B^{d'}_{\varepsilon}(f(x_0)) \subset U$. Since f is continuous at x_0 there is $\delta > 0$ so that

$$d(x, x_0) < \delta \quad \Longrightarrow \quad d'(f(x), f(x_0)) < \varepsilon. \tag{35.12}$$

Equation (35.12) is equivalent to

$$x \in B^{d}_{\delta}(x_0) \quad \Longrightarrow \quad f(x) \in B^{d'}_{\varepsilon}(f(x_0)), \tag{35.13}$$

that is,

$$B^{d}_{\delta}(x_0) \subseteq f^{-1}\left(B^{d'}_{\varepsilon}(f(x_0))\right). \tag{35.14}$$

Since $B^{d'}_{\varepsilon}(f(x_0)) \subset U$, $f^{-1}\left(B^{d'}_{\varepsilon}(f(x_0))\right) \subseteq f^{-1}(U)$. Hence

$$B_{\delta}(x_0) \subseteq f^{-1}\left(B^{d'}_{\varepsilon}(f(x_0))\right) \subseteq f^{-1}(U)$$

Since x_0 is arbitrary the set $f^{-1}(U)$ is open.

($\Longleftarrow$) Suppose for any open set $U \subset U'$ the preimage $f^{-1}(U)$ is open in S, $x_0 \in S$ and $\varepsilon > 0$.

Then the open ball $B_\varepsilon(f(x_0))$ is open in S'. By assumption on f the preimage $f^{-1}(B_\varepsilon(f(x_0)))$ is open in S. Since $x_0 \in f^{-1}(B_\varepsilon(f(x_0)))$ there is $\delta > 0$ so that

$$B_\delta(x_0) \subset f^{-1}\left(B_\varepsilon(f(x_0))\right).$$

But we have seen that the last inclusion is equivalent to

$$d(x, x_0) < \delta \quad \Longrightarrow \quad d'(f(x), f(x_0)) < \varepsilon,$$

which is continuity of f at x_0. Since x_0 is arbitrary, f is continuous in the sense of Definition 35.5 . □

We record two consequences of Theorem 35.9.

(1) The property of a function $f : (S, d) \to (S', d')$ being continuous doesn't really depend on the metrics: all that matters is that the preimages of open sets are open.
(2) It makes sense to define a function $f : (X, \mathcal{T}) \to (X', \mathcal{T}')$ between two topological spaces to be continuous as follows:

Definition 35.10. A function $f : (X, \mathcal{T}) \to (X', \mathcal{T}')$ between two topological spaces is continuous if and only if for every open set $U \subset X'$ the preimage $f^{-1}(U)$ is open in X.

Lecture 36: Compactness

Last time: if $f : (S, d) \to (S', d')$ is a function then we can define its continuity at $x_0 \in S$ in two equivalent ways:

(1) f is continuous at x_0 if for any sequence $\{a_n\}$ in S with $a_n \to x_0$ we have $f(a_n) \to f(x_0)$.
(2) f is continuous at x_0 if for any $\varepsilon > 0$ there is $\delta > 0$ so that if $d(x, x_0) < \delta$ then $d'(f(x), f(x_0)) < \varepsilon$.

Recall also that $f : (S, d) \to (S', d')$ is continuous if it is continuous at every $x \in S$. We proved: **Theorem** $f : (S, d) \to (S', d')$ is continuous if and only if for every $U \subseteq S'$ open, the preimage $f^{-1}(U)$ is open in S.

Consequently for a function between two general topological spaces (with topologies not necessarily coming from metrics) the following definition makes sense:

> A function $f : (X, \mathcal{T}) \to (X', \mathcal{T}')$ between two topological spaces is continuous if and only if for every open set $U \subset X'$ the preimage $f^{-1}(U)$ is open in X.

You may remember the following fact from calculus: if $f : [a, b] \to \mathbb{R}$ is continuous then f is bounded (i.e., the image $f([a, b])$ is bounded in $\mathbb{R}$) and f achieves its maximum and minimum on $[a, b]$: there are $x_1, x_2 \in [a, b]$ so that

$$f(x_1) \leq f(x) \leq f(x_2)$$

for all $x \in [a, b]$. The result generalizes to continuous functions on certain subsets of $\mathbb{R}^n$ and to more functions on more general topological spaces. To state the result we need a number of definitions.

Definition 36.1. A subset C of a topological space $(X, \mathcal{T})$ is closed if its complement $X \setminus C$ is open.

Example 36.2. A closed interval $[a, b] \subseteq \mathbb{R}$ ($\mathbb{R}$ has the standard topology) is closed.

Proof. We need to check that $U = \mathbb{R} \setminus [a,b]$ is open. That is, we need to check that for any $x \notin [a,b]$ (i.e., for $x \in U$) there is $r > 0$ so that $B_r(x) = (x-r, x+r) \subseteq U$, or, equivalently $(x-r, x+r) \cap [a,b] = \varnothing$. Now take $r = \min(|x-a|, |x-b|)$. This r works. □

Example 36.3. Let X be any set, $\mathcal{T} = \mathcal{P}(X)$, the largest possible topology on X. Then for any subset $Y \subseteq X$ the complement $X \setminus Y \in \mathcal{P}(X) = \mathcal{T}$. So any subset of X is both open and closed in this topology.

Example 36.4. Let $X = \mathbb{R}$ with the standard topology. Then an interval $[a,b)$ is neither open nor closed. (Why?)

Moral: subsets of topological spaces are not like doors, that are either open or closed. Subsets can be open and closed at the same time or they can be neither open nor closed.

Lemma 36.5. *Let (S,d) be a metric space and $C \subset S$ a closed subset (with respect to the metric topology). Let $\{a_n\}$ be a convergent sequence with $a_n \in C$ for all n. Then $L = \lim a_n \in C$.*

Proof. Suppose not. Then $L \in S \setminus C$. Since C is closed its complement $S \setminus C$ is open. Therefore there is $\varepsilon > 0$ so that $B_\varepsilon(L) \subseteq S \setminus C$. On the other hand, since $a_n \to L$ there is N so that for all $n > N$, $a_n \in B_\varepsilon(L)$. But $a_n \in C$ for all n hence $B_\varepsilon(L) \cap C \neq \varnothing$. This contradicts $B_\varepsilon(L) \subseteq S \setminus C$. □

Definition 36.6. A subset C of a metric space (S,d) is bounded if there is $x \in S$ and $r > 0$ so that $C \subset B_r(x)$.

Example 36.7. Suppose $S = \mathbb{R}$ with the standard metric. Then $C \subset \mathbb{R}$ is bounded if there is $x \in \mathbb{R}$ and $r > 0$ so that $C \subset (x-r, x+r)$, i.e., there are $K, L \in \mathbb{R}$ so that $K < c < L$ for all $c \in C$.

Example 36.8. Let S be a nonempty set and d the trivial metric ($d(x,y) = 1$ for all $x \neq y$). Then S is bounded: Fix $x \in S$. Then for any $y \in S$, $d(x,y) < 2$ hence $S \subset B_2(x)$.

Note that any subset of S is bounded as well.

Remark 36.9. The fact that a continuous function on a closed interval achieves its maximum and minimum on the interval generalizes to the vector calculus setting as follows:

Theorem Let $C \subset \mathbb{R}^n$ be closed and bounded and $f : \mathbb{R}^n \to \mathbb{R}$ continuous. Then f achieves its maximum and minumum on C: there are $a, b \in C$ so that for any $c \in C$

$$f(a) \leq f(c) \leq f(b).$$

We'll prove this theorem and its generalization. To state the generalization we need more definitions.

Definition 36.10. Let $(X, \mathcal{T})$ be a topological space, $E \subset X$ a subset. An open cover of E is a collection $\{U_\alpha\}_{\alpha \in A}$ of open subsets of X so that $E \subseteq \bigcup_{\alpha \in A} U_\alpha$.

Example 36.11. Let $X = \mathbb{R}^n$ with the standard topology (coming from the Euclidian metric d_2). Then the collection of open balls $\{B_n(0)\}_{n \in \mathbb{N}}$ is an open cover of $\mathbb{R}^n$: for any $x \in \mathbb{R}^n$ there is $n \in \mathbb{N}$ so that $d(0,x) < n$ and so $x \in B_n(x)$.

Example 36.12. Let S be a set, $\mathcal{T} = \mathcal{P}(S)$ so singletons $\{x\}$ are open for any $x \in S$. For any subset E of S the set $\{\{x\}\}_{x \in E}$ is an open cover of E since $E = \bigcup_{x \in E} \{x\}$.

Now an important definition:

Definition 36.13. A subset C of a topological space X is compact if for any open cover $\{U_\alpha\}_{\alpha\in A}$ of C there is a finite subcover. That is there are finitely many indices $\alpha_1,\dots,\alpha_n$ so that

$$C\subseteq U_{\alpha_1}\cup U_{\alpha_2}\cup\ldots\cup U_{\alpha_n}.$$

Example 36.14. Let X be a topological space. Then any finite subset of X is compact. Why?

An important result in undergraduate real analysis is the following theorem.

Theorem 36.15 (Heine-Borel)**.** *A subset C of $(\mathbb{R}^n,d_2)$ is compact if and only if C is closed and bounded with respect to d_2.*

We first prove:

Lemma 36.16. *Suppose (S,d) is a metric space and $C\subset S$ is compact (with respect to the metric topology $\mathcal{T}_d$). Then C is closed and bounded.*

Proof. Proving that C is bounded is not tricky. Fix a point $x\in S$. Consider the countable set $\{B_n(x)\}_{n\in\mathbb{N}}$ of open balls centered at x of integral radius. For any $c\in\mathbb{C}$ there is $n\in\mathbb{N}$ so that $d(c,x)<n$ (this is the Archimedean property of the reals). Hence $C\subset\bigcup_n B_n(x)$, i.e., $\{B_n(x)\}_{n\in\mathbb{N}}$ is an open cover of C. Since C is compact there are $n_1,n_2,\dots n_k$ so that $C\subset B_{n_1}\cup\ldots\cup B_{n_k}$. Let $M=\max(n_1,\dots,n_k)$. Then $C\subset B_M(x)$, and so C is bounded.

Proving that a compact subset of a metric space has to be closed is a little trickier. We'll argue that $S\smallsetminus C$ is open. So let $x\in S\smallsetminus C$. For any $c\in C$ let $r(c)=\frac{1}{2}d(x,c)$. Then $B_{r(c)}(x)\cap B_{r(c)}(c)=\varnothing$. This is because if $z\in B_{r(c)}(x)\cap B_{r(c)}(c)$, then

$$d(x,c)\le d(x,z)+d(z,c)\lneqq r(c)+r(c)=d(x,c),$$

which is impossible. The collection $\{B_{r(c)}(c)\}_{c\in C}$ is an open cover of C. Since C is compact, there is a finite subcover $\{B_{r(c_1)}(c_1),\dots,B_{r(c_k)}(c_k)\}$. Let $r=\min(r(c_1),\dots,r(c_k))$. By construction

$$B_r(x)\cap B_{r(c_i)}(c_i)=\varnothing$$

for $i=1,\dots,k$. Therefore

$$B_r(x)\cap C\subseteq B_r(x)\cap\big(B_{r(c_1)}(c_1)\cup\ldots\cup B_{r(c_k)}(c_k)\big)=(B_r(x)\cap B_{r(c_1)}(c_1))\cup\ldots\cup(B_r(x)\cap B_{r(c_k)}(c_k))=\varnothing.$$

Hence $B_r(x)\subset S\smallsetminus C$. Since x is arbitrary $S\smallsetminus C$ is open and therefore C is closed. □

Lecture 37: Closed and bounded subsets of $\mathbb{R}^n$ are compact

Last time: A subset C of a topological space $(X,\mathcal{T})$ is closed if its complement $X\smallsetminus C$ is open. Note: the whole space X is closed since $X\smallsetminus X=\varnothing$, which is open.

A subset C of a metric space (S,d) is bounded if there is $x\in S$ and $r>0$ so that $C\subset B_r(x)$.

An open cover of a subset E of a topological space $(X,\mathcal{T})$ is a collection $\{U_\alpha\}_{\alpha\in A}$ of open subsets of X so that $E\subseteq\bigcup_{\alpha\in A}U_\alpha$. That is, for every $x\in E$ there is an index α so that $x\in U_\alpha$.

A subset C of a topological space X is compact if for any open cover $\{U_\alpha\}_{\alpha\in A}$ of C there is a finite subcover. That is, there is $n>0$ and indices $\alpha_1,\dots,\alpha_n$ so that

$$C\subseteq U_{\alpha_1}\cup U_{\alpha_2}\cup\ldots\cup U_{\alpha_n}.$$

We proved: (Lemma 36.16) if a subset C of a metric space (S,d) is compact then C is closed and bounded.

Note that the converse is false! Take any infinite set S and give it a trivial metric d (so that $d(x,y)=1$ for all $x,y\in S$). Then S is closed (it's the whole space) and bounded but it's not compact since the collection of singletons $\{\{x\}\}_{x\in S}$ is an open cover with not finite subcover: you can't leave anything out of the cover and the whole collection of points in S is infinite.

The main result of today's lecture is the following theorem (Sally has a version for $\mathbb{R}$ and for $\mathbb{C} = \mathbb{R}^2$).

Theorem 37.1 (Heine-Borel)**.** *Consider $\mathbb{R}^n$ with the Euclidean metric d_2. Let $C \subset \mathbb{R}^n$ be a subset. The following conditions are equivalent:*

(1) *Every sequence in C has a subsequence that converges to a point in C (with respect to d_2).*
(2) *C is closed and bounded (with respect to d_2).*
(3) *C is compact (with respect to the standard topology on $\mathbb{R}^n$, the one defined by d_2).*

Proof. We have seen that (3) $\implies$ (2) in any metric space, not just in $(\mathbb{R}^n, d_2)$. That's Lemma 36.16.

(2) $\implies$ (1) Suppose $C \subset \mathbb{R}^n$ is closed and bounded and $\{a_k = ((a_k)_1, \dots (a_k)_n)\}_{k\in\mathbb{N}}$ a sequence in C. Since C is bounded, $\{a_k\} \subseteq C$ is bounded in $\mathbb{R}^n$ and therefore each of the sequence $\{(a_k)_i\}_{k\in\mathbb{N}}$, $1 \le i \le n$ is bounded in $\mathbb{R}$. Every bounded sequence in $\mathbb{R}$ has a convergent subsequence. So $\{(a_k)_1\}_{j\in\mathbb{N}}$ has a convergent subsequence $\{(a_{k_j})_1\}_{j\in\mathbb{N}}$. Since $\{(a_{k_j})_2\}_{j\in\mathbb{N}}$ is bounded, it has a convergent subsequence $\{(a_{k_{j_\ell}})_2\}_{\ell\in\mathbb{N}}$. After n steps we get a subsequence $\{a_{n_r}\}_{r\in\mathbb{N}}$ so that the coordinate sequences $\{(a_{n_r})_i\}_{r\in\mathbb{N}}$, $1 \le i \le n$, all converge. Say $(a_{n_r})_i \to L_i$ in $\mathbb{R}$ for $i = 1, \dots, n$. Then $a_{n_r} \underset{r\to\infty}{\longrightarrow} L = (L_1, \dots, L_n)$. Since C is closed, $L \in C$ (Lemma 36.5).

(1) $\implies$ (3) This proof is the most complicated one. We'll do it in several steps.

First of all recall that the set $\mathbb{Q}$ of rational numbers is countable and that the product of two countable sets is countable. Consequently the n-fold product $\mathbb{Q}^n$ is countable.

Claim 0 If $B_R(x)$ is a ball in a metric space (S, d), $y \in S$ and $d(x, y) < R/2$ then for any $r \le R/2$ the ball $B_r(y)$ is contained in $B_R(x)$.

Proof of Claim 0. For any $z \in B_r(y)$

$$d(z, x) \le d(z, y) + d(y, x) < r + R/2 \le R/2 + R/2 = R.$$

Hence $B_r(y) \subseteq B_R(x)$. □

Claim 1 For any $x \in \mathbb{R}^n$ and for any $r > 0$ there is $q \in \mathbb{Q}^n$ so that $d(x, q) < r$ or, equivalently, $q \in B_r(x)$. In other words $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$.

Proof of Claim 1. A point $x \in \mathbb{R}^n$ is an n-tuple of reals numbers: $x = (x_1, \dots, x_n)$. For each i there is $q_i \in Q$ so that $|x_i - q_i| < \frac{r}{\sqrt{n}}$. Then

$$d_2(x, q) = \left(\sum_i |x_i - q_i|^2\right)^{1/2} < \left(\frac{r^2}{n} + \cdots + \frac{r^2}{n}\right)^{1/2} = r.$$

□

Now suppose that $C \subset \mathbb{R}^n$ has the property that any sequence in C has a subsequence that converges to a point in C. Let $\{U_\alpha\}_{\alpha\in A}$ be an open cover of C. We'd like to show that $\{U_\alpha\}_{\alpha\in A}$ has a finite subcover.

Claim 2 There is a countable collection $\mathcal{B}$ of open balls so that

(1) $C \subset \bigcup_{B\in\mathcal{B}} B$, i.e., $\mathcal{B}$ is an (open) cover of C, and
(2) for each $B \in \mathcal{B}$ there is $\alpha \in A$ so that $B \subset U_\alpha$.

Proof of Claim 2. Let

$$\mathcal{B} = \{B_r(q) \mid q \in \mathbb{Q}^n, r \in (0, \infty) \cap \mathbb{Q} \text{ and there is } \alpha \in A \text{ so that } B_r(q) \subseteq U_\alpha\}.$$

The set $\mathcal{B}$ is countable since it is a subset of the countable set of balls

$$\{B_r(q) \mid q \in \mathbb{Q}^n, r \in (0,\infty) \cap \mathbb{Q}\}.$$

We need to show that $\mathcal{B}$ is an open cover of C.

If $x \in C$ then $x \in U_\alpha$ for some α. Since U_α is open, there is $R > 0$ so that $B_R(x) \subseteq U_\alpha$. Choose $r \in \mathbb{Q}$ with $0 < r < R/2$. By Claim 1 there is $q \in \mathbb{Q}^n$ so that $d(q,x) < r$. Then $q \in B_r(x)$ and, by Claim 0,

$$B_r(q) \subseteq B_R(x) \subseteq U_\alpha.$$

Therefore $B_r(q) \in \mathcal{B}$. Since $x \in B_r(q)$, $x \in \bigcup_{B\in\mathcal{B}} B$. Since x is an arbitrary point of C, $C \subseteq \bigcup_{B\in\mathcal{B}} B$. □

Since $\mathcal{B}$ is countable, there is a bijection $f : \mathbb{N} \to \mathcal{B}$. Let $B_i = f(i)$. Then $\mathcal{B} = \{B_i\}_{i\in\mathbb{N}}$.

Claim 3 There is $n \in \mathbb{N}$ and indicies $i_1, \ldots, i_n \in \mathbb{N}$ so that $C \subseteq B_{i_1} \cup \ldots \cup B_{i_n}$.

Proof of Claim 3. Suppose not. Then $C \not\subseteq B_1$ for otherwise we may take $n = 1$ and $i_1 = 1$. Therefore there is $x_1 \in C$ with $x_1 \notin B_1$.

By the same reasoning $C \not\subseteq B_1 \cup B_2$ (for otherwise we may take $n = 2$ and $i_1 = 1, i_2 = 2$). Therefore there is $x_2 \in C$ with $x_2 \notin B_1 \cup B_2$.

Proceeding recursively we see that for each $k \in \mathbb{N}$ there is $x_k \in C$ with $x_k \notin B_1 \cup \ldots B_k$. In particular for any $\ell \leq k$, $x_k \notin B_\ell$.

Now by assumption on C there is a convergent subsequence $\{x_{k_i}\}$ of $\{x_k\}$ with $L = \lim_{i\to\infty} x_{k_i} \in C$. Since $L \in C$ there is $m \in \mathbb{N}$ so that $L \in B_m$. For any index $r \geq m$, we have $k_r \geq r \geq m$. Hence $x_{k_r} \notin B_m$. On the other hand since $L \in B_m$ and B_m is open there is $\varepsilon > 0$ so that $B_\varepsilon(L) \subset B_m$. Since $x_{k_i} \to L$, there is N so that for $i > N$ $x_{k_i} \in B_\varepsilon(L) \subset B_m$. So for $r > \max(N, m)$ we have both that $x_{k_r} \in B_m$ and $x_{k_r} \notin B_m$, which is impossible. □

By construction of the collection $\mathcal{B} = \{B_i\}_{i\in\mathbb{N}}$ of balls for each index i_j there is $\alpha = \alpha(i_j)$ so that $B_{i_j} \subseteq U_{\alpha(i_j)}$. Therefore

$$C \subseteq B_{i_1} \cup \ldots \cup B_{i_n} \subseteq U_{\alpha(i_1)} \cup \ldots \cup U_{\alpha(i_n)}$$

and we are done. □

Lecture 38: Continuous functions on compact sets achieve max and min

Last time: Proved Heine-Borel theorem: for a subset C of $\mathbb{R}^n$ the following conditions are equivalent:

(1) Every sequence in C has a subsequence that converges to a point in C.
(2) C is closed and bounded.
(3) C is compact.

Remark 38.1. If $C \subset \mathbb{R}$ is closed and bounded then $\inf C$, $\sup C$ exist and moreover $\inf C, \sup C \in C$. This is because there are sequences $\{a_n\}, \{b_n\}$ in C with $a_n \to \inf C$, $b_n \to \sup C$. Since C is closed $\inf C = \lim a_n \in C$ and $\sup C = \lim b_n \in C$ as well.

The main result of today's lecture is

Theorem 38.2. *Let $f : (X, \mathcal{T}_X) \to (Y, \mathcal{T}_Y)$ be a continuous function, $C \subset X$ a compact set. Then the set $f(C)$ is compact in Y.*

Proof. Let $\{U_\alpha\}_{\alpha\in A}$ be an open cover of $f(C)$. We need to show that it has a finite subcover. Since f is continuous and each U_α is open, the preimages $f^{-1}(U_\alpha)$ are open in X. Since $f(C) \subseteq \bigcup_{\alpha\in A} U_\alpha$ we have

$$C \subseteq f^{-1}\left(\bigcup_{\alpha\in A} U_\alpha\right) = \bigcup_{\alpha\in A} f^{-1}(U_\alpha).$$

Since C is compact, there is n and indicies $\alpha_1, \ldots, \alpha_n$ so that

$$C \subseteq f^{-1}(U_{\alpha_1}) \cup \ldots \cup f^{-1}(U_{\alpha_n}) = f^{-1}\left(U_{\alpha_1} \cup \ldots \cup U_{\alpha_n}\right)$$

Consequently

$$f(C) \subseteq U_{\alpha_1} \cup \ldots \cup U_{\alpha_n}.$$

Therefore $\{U_{\alpha_1}, \ldots, U_{\alpha_n}\}$ is a finite subcover of $\{U_\alpha\}_{\alpha\in A}$ and thus $f(C)$ is compact. □

Corollary 38.3. *Let $f : (X, \mathcal{T}) \to \mathbb{R}$ be a continuous function (with $\mathbb{R}$ being given the standard topology) and $C \subset X$ be compact. Then there exist $x_1, x_2 \in C$ (not necessarily unique) so that*

$$f(x_1) \leq f(x) \leq f(x_2)$$

for all $x \in C$. That is, the restriction of f to C is bounded and achieves maximum and minimum.

Proof. Since C is compact and f is continuous $f(C)$ is compact by Theorem 38.2. By Heine-Borel (really by Lemma 36.16), $f(C)$ is closed and bounded. By Remark 38.1 $y_1 := \inf f(C)$ and $y_2 := \sup f(C)$ are in $f(C)$. Since $y_1, y_2 \in f(C)$, there are $x_1, x_2 \in C$ with $y_1 = f(x_1)$, $y_2 = f(x_2)$. Therefore, for any $x \in C$

$$f(x_1) = \inf f(C) \leq f(x) \leq \sup f(C) = f(x_2).$$

□

Corollary 38.4. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is continuous (standard topologies) and $C \subset \mathbb{R}^n$ is closed and bounded. Then there exist $x_1, x_2 \in C$ (not necessarily unique) so that*

$$f(x_1) \leq f(x) \leq f(x_2)$$

for all $x \in C$.

Proof. Closed and bounded subsets of $\mathbb{R}^n$ are compact. □

Remark 38.5. It is not essential for f to be defined on all of $\mathbb{R}^n$. It only needs to be defined (and be continuous) on a subset X of $\mathbb{R}^n$ that contains C. Note that any subset of a metric space is a metric space again, so any subset X of $\mathbb{R}^n$ is a metric space. For example we can define $d : X \times X \to [0, \infty)$ by $d(x, y) := \left(sum(x_i - y_i)^2 \right)^{1/2}$. Consequently continuity of $f : X \to \mathbb{R}$ makes sense.