Notebook

https://github.com/v--/notebook

Ianis Vasilev, ianis@ivasilev.net

This ever-expanding document started as a set of study notes and exercises and gradually outgrew itself to become slightly encyclopedic. Having all these notes in one place is quite helpful for both expressing my own thoughts clearly and for later reference.

Contents

1. Analysis

1.1. Geometry of vector spaces

Definition 1. Let X be a real vector space[82] and let $t_1, \dots, t_n \in \mathbb{R}$ and $x_1, \dots, x_n \in X$. We say that their linear combination[77] $x := \sum_{k=1}^{n} t_k x_k$ is

a) an affine combination if $\sum_{k=1}^{n} t_k = 1$.

b) a conic combination if all of the coefficients are nonnegative.

c) a convex combination if it is both affine and conic. A convex combination of two elements $x, y \in X$ is usually written as $tx + (1-t)y$ for some scalar $t \in [0,1]$.

Definition 2. Let A be a subset of the real vector space X. We define the convex hull $\operatorname{conv} A$ (resp. affine hull or conic hull) of A by the set of all convex (resp. affine or conic) combinations of finite subsets of A.

The convex hull $\operatorname{conv} A$ of two elements $x, y \in X$ is often called the segment between x and y and is denoted as

$$[x, y] := \{tx + (1-t)y \colon t \in [0,1]\}.$$

If the set A is equal to its convex (resp. affine or conic) hull, we say that it is a convex set (resp. affine set or cone)

Definition 3. Let $v_1, \dots, v_n$ be linearly independent vectors and v_0 be any other vector. The convex hull S of the vectors $v_0 + v_1, \dots, v_0 + v_n$ is called an n-simplex.

The convex hull of any nonempty subset of $\{v_0, \dots, v_n\}$ of cardinality $m+1$ is an m-simplex and is called an m-face of S.

1.2. Topological vector spaces

Definition 4. Let X be any vector space and let $\mathcal{T}$ be a topology on X. The space $(X, +, \cdot, \mathcal{T})$ is called a topological vector space if the linear and topological structure agree, that is, the operations $+ : X \times X \to X$ and $\cdot : X \times \mathbb{R} \to X$ are continuous with respect to $\mathcal{T}$.

Proposition 5. *Let X and Y be real Hausdorff topological vector spaces, let $D \subseteq X$ and let $f : D \to Y$ be a continuous function.*

Then f is locally bounded.

Proof. Let $x_0 \in D$. The set $f(x_0) + B_Y \subseteq Y$ is obviously bounded and open. Since f is continuous, $f^{-1}(f(x_0) + B_Y)$ is also open. Even though f may not be surjective, $f^{-1}(f(x_0) + B_Y)$ is nonempty, since it contains x_0.

This implies that $f^{-1}(f(x_0) + B_Y)$ is a neighborhood of x_0 with a bounded image. Hence f is locally bounded. □

1.3. Differentiability

Let X and Y be Hausdorff topological vector spaces[4], let $D \subseteq X$ be open and let $D : X \to Y$ be any function.

Definition 6. We fix a point $x \in D$ and a direction $h \in S_X$. We introduce a few definitions of derivatives. In all cases we say that the derivative (of the corresponding type) exists for f at x in the direction h. The quotient under the limit sign is called a difference quotient (of the corresponding type).

a) The classical two-sided derivative is defined as

$$f'(x)(h) := \lim_{t \to 0} \frac{f(x+th) - f(x)}{t}$$

This definition is often too strict and so there exist a few generalizations.

b) [Phe93, lemma 1.2] The one-sided (or right-hand) directional derivative is defined as

$$f'_+(x)(h) := \lim_{t \downarrow 0} \frac{f(x+th) - f(x)}{t}$$

It is also denoted as $\partial^+ f(x)(h)$ in [Phe93, lemma 1.2]. We do not need to define a left-hand directional derivative because it would equal $-f'_+(x)(-h)$.

c) [Cla13, definition 11.18] The upper (resp. lower) Dini derivative is defined as

$$\overline{f}'(x)(h) := \limsup_{t \downarrow 0} \frac{f(x+th) - f(x)}{t}$$
$$\underline{f}(x)(h) := \liminf_{t \downarrow 0} \frac{f(x+th) - f(x)}{t}$$

Dini derivatives are useful when the difference quotients are bounded but do not have a limit.

d) [Cla13, Section 10.1] The generalized Clarke derivative is defined as

$$f^\circ(x)(h) := \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+th) - f(y)}{t} = \lim_{\delta \to 0} \sup_{\substack{y \in B(x,\delta) \\ t \in (0,\delta)}} \frac{f(y+th) - f(y)}{t}.$$

Refer to section 1.9 for their usefulness.

Definition 7. We fix a point $x \in D$. We will now introduce several types of differentiability. Each is implied by the next one.

a) We do not introduce a special name for functions that have a one-sided derivative at x in the direction h, although some authors call these functions Gateaux differentiable at x in the direction h.

b) [ЙT74, p. 0.2.1] If the two-sided directional derivative $f'(x)(h)$ exists for all directions $h \in S_X$, we call it the first variation at x and denote the corresponding linear operator $\delta f(x) : X \to Y$.

c) [Phe93, definition 1.12] Let X be a Banach space. We say that f is Gateaux-differentiable at x if there exists a continuous linear operator $f'_G(x) : X \to Y$, called the Gateaux derivative of f at x, such that

$$f'_G(x)(h) = \lim_{t \to 0} \frac{f(x + th) - f(x)}{t}.$$

The Gateaux derivative $f'_G(x)$ exists precisely when the first variation $\delta f(x)$ operator exists and is continuous. They are obviously equal.

The Gateaux derivative is also denoted by $df(x)$ in [Phe93, definition 1.12].

d) [Phe93, definition 1.12] Let X be a Banach space. We say that f is Frechet-differentiable at x if there exists a continuous linear operator $f'(x) : X \to Y$, called the Frechet derivative of f at x, such that for each $\varepsilon > 0$ there exists a radius $\delta > 0$ such that and for every direction $h \in S_X$ we have

$$t \in (0, \delta) \implies \left\| \frac{f(x + th) - f(x)}{t} - f'(x)(h) \right\| < \varepsilon.$$

Note that for each $\varepsilon > 0$, Gateaux differentiability gives us a radius $\delta_h > 0$ such that

$$t \in (0, \delta_h) \implies \left\| \frac{f(x + th) - f(x)}{t} - f'_G(x)(h) \right\| < \varepsilon.$$

If the limit is uniform over $h \in S_X$, i.e. if $\sup_{h \in S_X} \delta_h < \infty$, then f is Frechet differentiable at x and $f'(x) = f'_G(x)$.

e) [DR14, p. 33] We say that f is strictly differentiable at x if there exists a continuous linear operator $f'(x) : X \to Y$ such that

$$\lim_{\substack{y \to x \\ z \to x}} \frac{f(y) - f(z) - f'(x)(y - z)}{\|y - z\|} = 0.$$

1.4. Subdifferentials

Let X be a Hausdorff topological vector space[4], let $D \subseteq X$ be an open set and $f : D \to \mathbb{R}$ be any function.

Definition 8. We fix a point $x \in D$. We define different types of subgradients and subdifferentials. Subgradients are linear functionals $x^* \in X^*$ that approximate f at the point x in a certain way, and a subdifferential is the set of all subgradients of a given type.

a) [Cla13, p. 59] We say that $x^* \in X^*$ is a subgradient of f at x if for every $y \in D$ we have

$$f(y) - f(x) \geqslant \langle x^*, y - x \rangle .$$

The subdifferential of f at x is denoted by $\partial f(x)$ and is also sometimes called the convex subdifferential because of proposition 29.

b) [Cla13, definition 10.3] We say that $x^* \in X^*$ is a Clarke (generalized) subgradient of f at x if for every direction $h \in X$ we have

$$f^\circ(x)(h) \geqslant \langle x^*, h \rangle ,$$

where $f^\circ(x)(h)$ is the generalized Clarke derivative[6 (d)].

The subdifferential of f at x is denoted by $\partial_C f(x)$. Confusingly, the Clarke subdifferential is called the "generalized gradient" by Clarke himself with no special name for the Clarke subgradients.

See section 1.9 for properties of these subgradients.

c) [Cla13, p. 227] We say that $x^* \in X^*$ is a proximal subgradient of f at x if there exist $\sigma > 0$ and a neighborhood $V \subseteq X$ of x such that for every $y \in D \cap V$ we have

$$f(y) - f(x) + \sigma \|y - x\|^2 \geqslant \langle x^*, y - x \rangle .$$

The proximal subdifferential of f at x is denoted by $\partial_P f(x)$.

d) [Cla13, definition 11.10] Suppose the following are satisfied:

1. $\{x_n\}_n \subseteq D$ is a sequence of points converging to x
2. $f(x_n) \to f(x)$ (redundant if f is continuous)
3. x_n^* is a proximal subgradient for f at x_n for every $n \in \mathbb{Z}^{>0}$.

If the limit $x^* := \lim_n x_n^*$ exists and is a continuous linear functional, we call x^* a limiting subgradient of f at x.

The limiting subdifferential of f at x is denoted by $\partial_P f(x)$.

1.5. Norms

Definition 9. [Phe93, Example 2.26] We define the duality mapping

$$D : E \rightrightarrows X^*,$$
$$D(x) := \{x^* \in X^* \colon \; \|x\| = \|x^*\| \text{ and } \langle x^*, x\rangle = \|x^*\| \, \|x\|\}.$$

Note 10. We will usually use this mapping for unit vectors, so we may as well consider its restriction to the unit spheres, where

$$D' : S_X \rightrightarrows S_{X^*},$$
$$D'(x) := \{x^* \in S_{X^*} \colon \; \langle x^*, x\rangle = 1\}.$$

Lemma 11. *For every point $x \in X$, the set $D(x)$ is nonempty.*

Proof. Fix $x \in X$ and consider the one-dimensional subspace

$$\operatorname{span}\{x\} = \{\lambda x \colon \lambda \in \mathbb{R}\}.$$

Define $\xi : \operatorname{span}\{x\} \to \mathbb{R}$ by $\xi(\lambda x) = \lambda \|x\|^2$.

The functional ξ is linear and, since it acts on a finite-dimensional space, it is also continuous. The norm of ξ is

$$\|\xi\| = \max\left\{\left\langle \xi, \frac{x}{\|x\|}\right\rangle, \left\langle \xi, -\frac{x}{\|x\|}\right\rangle\right\} = \max\{\|x\|, -\|x\|\} = \|x\|.$$

The Hahn-Banach theorem allows us to extend ξ to a continuous linear functional $x^* \in X^*$ such that $\|x^*\| = \|\xi\| = \|x\|$ and $\langle x^*, x\rangle = \langle \xi, x\rangle = \|x\|^2 = \|x\| \, \|x^*\|$. Thus $x^* \in D(x)$ and $D(x)$ is nonempty. ☐

Definition 12. [Phe93, Definition 2.36] The norm $\|\cdot\|$ on X is called smooth if any of if for each $x \in S_X$ the duality mapping is single-valued.

Definition 13. [Phe93, Definition 2.36] The norm $\|\cdot\|$ on X is called rotund or strictly convex if any of the following equivalent conditions hold:

a) There are no line segments in the unit sphere S_X.

b) Every convex subset of X has at most one point of least norm.

c)

$$\|x + y\| = \|x\| + \|y\| \implies x \text{ and } y \text{ are linearly dependent.} \tag{1.1}$$

Proof. (13 (a) $\implies$ 13 (b)) Let the norm in E be rotund and let $C \subseteq E$ be a (potentially empty) convex set. We will prove that C contains at most one point of least norm.

If C is empty or otherwise contains no element of least norm, trivially contains at most one point of least norm.

Now let C contain at least one element $x \in C$ of least norm. Assume that $y \in C$ is another element of least norm. Necessarily $\|x\| = \|y\|$.

Fix $t \in (0,1)$ and define $z := tx + (1-t)y$. Since C is convex, it contains z. Since x and y are elements of least norm, we have $\|z\| \geqslant \|x\|$. By the triangle inequality,

$$\|z\| = \|tx + (1-t)y\| \leqslant t\|x\| + (1-t)\|y\| = \|x\|,$$

thus $\|z\| = \|x\|$.

This implies that the entire segment $[x, y]$ are elements of least norm in C. Hence the segment $[x, y]$ is contained in the sphere $\|x\| S_E$, which contradicts the rotundity of the norm $\|\cdot\|$.

Hence C contains at most one element of least norm.

(13 (b) $\implies$ 13 (a)) Let every convex set $C \subseteq E$ have at most one element of least norm.

Assume$^{\mathrm{LEM}}$ that the norm $\|\cdot\|$ is not rotund. Then the unit sphere S_E contains a line segment $[x, y], x \neq y$. The set $[x, y]$ is compact and, by the Weierstrass extreme value theorem, the norm attains its minimum on the segment in a point $z \in [x, y]$. Since the segment is also convex and we assumed that convex sets have at most one element of least norm, it follows that this element z is unique.

Then for any point $s \in [x, y], s \neq z$, we have $\|s\| > \|z\| = 1$, thus s cannot be an element of the unit sphere. The obtained contradiction shows that the norm $\|\cdot\|$ is rotund.

(13 (a) $\implies$ 13 (c)) Let E be rotund let $x, y \in E$ be distinct vectors such that

$$\|x + y\| = \|x\| + \|y\|. \tag{1.2}$$

If either of them is the zero vector, then they are trivially linearly dependent.

Assume that both x and y are nonzero and define

$$\xi := \frac{x}{\|x\|} \qquad \eta := \frac{y}{\|y\|} \qquad t := \frac{\|x\|}{\|x + y\|}$$

Equation (1.2) implies that

$$1 - t = 1 - \frac{\|x\|}{\|x + y\|} = \frac{\|x + y\| - \|x\|}{\|x + y\|} = \frac{\|y\|}{\|x + y\|}.$$

Since both ξ and η are in S_E, by rotundity, their convex combination

$$\nu := t\xi + (1-t)\eta$$

should not be contained in S_E unless $\xi = \eta$.

Calculating the norm, we obtain

$$\begin{aligned}\|\nu\| &= \|t\xi + (1-t)\eta\| = \\ &= \left\| \frac{\|x\|\,\xi}{\|x + y\|} + \frac{\|y\|\,\eta}{\|x + y\|} \right\| = \\ &= \left\| \frac{x + y}{\|x + y\|} \right\| = 1,\end{aligned}$$

hence $\nu \in S_E$. Thus $\xi = \eta$ and $x = \frac{\|x\|}{\|y\|} y$, so x and y are linearly dependent.

(13 (c) $\implies$ 13 (a)) Let eq. (1.1) hold and fix $x, y \in S_E, t \in (0,1)$. Define $z := tx + (1-t)y$. First, assume that the vectors tx and $(1-t)y$ satisfy the left part of eq. (1.1), i.e.

$$\|z\| = \|tx + (1-t)y\| = t\,\|x\| + (1-t)\,\|y\| = 1.$$

This does not refute rotundity since x and y are not necessarily distinct. It follows from eq. (1.1) that tx and $(1-t)y$ are linearly dependent, hence x and y are also linearly dependent. Since x and y both have unit norm, either $y = x$ or $y = -x$.

If we assume that $y = -x$, then

$$\|z\| = \|tx + (1-t)y\| = (2t-1)\,\|x\| = 2t - 1,$$

which is only possible if $t = 1$ since $\|z\| = 1$. But t is strictly less than 1.

Hence $y \neq -x$ and the only remaining possibility is that $y = x$.

Now assume that the vectors tx and $(1-t)y$ do not satisfy the left part of eq. (1.1). This implies $\|z\| < 1$. Thus x and y are necessarily distinct, but z is not contained in the unit sphere and the segment $[x, y]$ is not contained in S_E.

We have shown that $x, y \in S_E$ implies that either $y = x$ or that the segment $[x, y]$ is not contained in S_E, thus the norm in E is rotund. □

Theorem 14. *[Phe93, exercise 2.37(a)] If the norm in X is such that its dual norm in X^* is rotund (resp. smooth), then it is itself smooth (resp. rotund).*

Proof. 1. First, let the dual norm $\|\cdot\|^*$ be rotund and assume that $\|\cdot\|$ is not smooth.

Fix $x \in S_X$. Since $D(x)$ is nonempty (by lemma 11) and since $\|\cdot\|$ is not smooth, then there exist two different functionals $x^*, y^* \in D(x)$, such that

$$\langle x^*, x\rangle = \langle y^*, x\rangle = 1.$$

We will show that the segment $[x^*, y^*]$ is contained in S_{X^*}, i.e. that the dual norm is not rotund.

Fix any $t \in (0,1)$ and define $z^* := tx^* + (1-t)y^*$. We only need to show that $\|z^*\| = 1$.

By the triangle inequality, we have

$$\|z^*\| = \|tx^* + (1-t)y^*\| \leqslant t\,\|x^*\| + (1-t)\,\|y^*\| = t + (1-t) = 1.$$

For the reverse inequality, note that

$$\|z^*\| \geqslant \langle z^*, x\rangle = t\,\langle x^*, x\rangle + (1-t)\,\langle y^*, x\rangle = t + (1-t) = 1,$$

thus $\|z^*\| = 1$. Hence $[x^*, y^*]$ is contained in S_{X^*} and the dual space is not smooth. The obtained contradiction proves that the norm in X is rotund.

2. Now let the dual norm $\|\cdot\|^*$ be smooth and assume that $\|\cdot\|$ is not rotund. Then there exist points $x, y \in S_X$ such that the while segment $[x, y]$ is contained in S_X. Fix $t \in (0,1)$ and define $z := tx + (1-t)y \in S_X$. Denote by $J : X \to X^{**}$ the canonical embedding into the double-dual. By lemma 11, there exists a functional $z^* \in X^*$, such that

$$\langle J(z), z^* \rangle = \langle z^*, z \rangle = 1.$$

Because the dual norm $\|\cdot\|^*$ is smooth, we cannot have $\langle J(x), z^* \rangle = \langle z^*, x \rangle = 1$ or $\langle J(y), z^* \rangle = \langle z^*, y \rangle = 1$ and since $\|z^*\| = 1$, necessarily

$$\langle z^*, x \rangle < 1 \text{ and } \langle z^*, y \rangle < 1.$$

If follows that

$$1 = \langle z^*, z \rangle = t \langle z^*, x \rangle + (1-t) \langle z^*, y \rangle < t + (1-t) = 1,$$

which is a contradiction. Hence $\|\cdot\|$ is rotund.

□

Proposition 15. *[Phe93, exercise 2.37(c]) Norms in Hilbert spaces are both smooth and rotund.*

Proof. Let X be a Hilbert space, i.e. the norm is generated by an inner product and, due to Riesz's theorem, we identify the space X with its continuous dual X^*.

To prove that X is rotund, choose $x, y \in S_X, x \neq y$. We will show that the segment $[x, y]$ is not contained in S_X.

If x and y are linearly dependent, necessarily $y = -x$ and all non-trivial convex combinations of x and y are contained in the open unit ball, hence $[x, y] \nsubseteq S_X$.

Not let x and y be linearly independent. By the Cauchy-Bunyakovsky-Schwarz inequality, we have

$$\langle x, y \rangle \leqslant |\langle x, y \rangle| < \|x\| \, \|y\| = 1. \tag{1.3}$$

Fix $t \in (0,1)$ and let $z := tx + (1-t)y$. We will show that $z \notin S_X$. Indeed,

$$\begin{aligned} \|z\|^2 = \langle z, z \rangle &= t^2 \|x\|^2 + t(1-t) \langle x, y \rangle + (1-t)t \langle y, x \rangle + (1-t)^2 \|y\|^2 = \\ &= t^2 + (1-t)^2 + 2t(1-t) \langle x, y \rangle < \\ &\overset{(1.3)}{<} t^2 + (1-t)^2 + 2t(1-t) = \\ &= t^2 + 1 - 2t + t^2 + 2t - t^2 = 1. \end{aligned}$$

Thus $\|z\|^2 < 1$ and $\|z\| < 1$ and $z \notin S_X$.

In both cases, no interior point of the segment $[x, y]$ is contained in S_X, hence the norm in X is rotund.

Since we identify X with its dual, the norm in X^* is also rotund and by theorem 14, the norm in X is also smooth. □

Example 16. [Phe93, exercise 2.37(c]) The norms in c_0 and l^1 are neither smooth nor rotund.

Proof. Consider the space c_0 of all real sequences that converge to zero equipped with the uniform norm

$$\|x\|_{c_0} := \sup_i |x_i| \,.$$

Note that the dual space of c_0 is (isometrically isomorphic to) the space l^1 of absolutely summable sequences with norm

$$\|x\|_{l^1} := \sum_i |x_i| \,.$$

Let $\{e_n\}_{n=1}^{\infty}$ be the canonical basis of c_0, i.e. the coordinates $e_n^{(i)}$ of e_n are given by the Dirac delta function, $e_n^{(i)} := \delta_{i,n}$.

For every natural $n \geqslant 1$, define x_n to be the same as e_n except that the first coordinate of x_n is always 1.

The corresponding norms of e_n are all equal to 1 and the norms of x_n are

$$\|x_n\|_{c_0} = 1 \qquad\qquad \|x_n\|_{l^1} = 2.$$

For every n we have

$$\langle e_1, x_n \rangle = \langle e_n, x_n \rangle = 1,$$

hence $J_{c_0}(x_n)$ has at least two elements e_1 and e_n and the norm in c_0 is not smooth.

Given that $\{x_1, x_2, \ldots\} \subseteq S_{c_0}$, consider the convex combinations of x_2 and x_3:

$$tx_2 + (1-t)x_3 = (1, t, (1-t), 0, 0, \ldots).$$

Evidently $tx_2 + (1-t)x_3 \in S_{c_0}$ for every $t \in (0,1)$, hence the norm in c_0 is not rotund.

The contrapositions$^{\mathrm{LEM}}$ to the statements in theorem 14 say that if X is not rotund (resp. smooth), then the dual space X^* is not smooth (resp. rotund). Thus l^1 is neither smooth or rotund as the dual of c_0. □

1.6. Dentable sets

Definition 17. [Phe93, Example 3.2(a]) Let X be a Banach space.
We define the support function σ_{A^*} for the set of functionals $A^* \subseteq X^*$ by

$$\sigma_{A^*} : X \to \mathbb{R} \cup \{\infty\}$$
$$\sigma_{A^*}(x) := \sup\{\langle x^*, x\rangle : x^* \in A^*\}$$

and the weak* support function σ_A^* for the set of points $A \subseteq X$ by

$$\sigma_A^* : X^* \to \mathbb{R} \cup \{\infty\}$$
$$\sigma_A^*(x^*) := \sup\{\langle x^*, x\rangle : x \in A\}.$$

Definition 18. [Phe93, definition 2.17] Given a linear functional x^*, a nonempty subset A of X and a diameter $\alpha > 0$, the value $S(x^*, A, \alpha)$ is called a slice of A, where

$$S : X^* \times \mathcal{P}(X) \times \mathbb{R}^{>0} \mapsto \mathcal{P}(A)$$
$$S(x^*, A, \alpha) := \{x \in A \colon \langle x^*, x\rangle > \sigma_A^*(x^*) - \alpha\}.$$

We define a weak* slice of $A^* \subseteq X^*$ as $S^*(x, A^*, \alpha)$, where

$$S^* : X \times \mathcal{P}(X) \times \mathbb{R}^{>0} \mapsto \mathcal{P}(A)$$
$$S^*(x, A^*, \alpha) := \{x^* \in A^* \colon \langle x^*, x\rangle > \sigma_{A^*}(x) - \alpha\}.$$

If we need to make the underlying space explicit, we will use $S_X(x^*, A, \alpha)$ and $S_X^*(x, A^*, \alpha)$.

Definition 19. [Phe93, definition 5.1] A subset A of a Banach space X is called dentable if it admits slices of arbitrarily small diameter, i.e. for every $\varepsilon > 0$ there exist a functional $x^* \in X^*$ and a diameter $\alpha > 0$, such that $\operatorname{diam} S(x^*, A, \alpha) < \varepsilon$.
Weak* dentability is defined in an obvious way.

Definition 20. [Phe93, definition 5.2] The space X is said to have the Radon-Nikodym property (RNP) if every nonempty bounded set A of X is dentable.

Proposition 21. *Let X be a Banach space and $A^* \subseteq X^*$ be a weak*-dentable set. Then A^* is dentable in X^*.*

Proof. Let $\varepsilon > 0$ and let $x \in X$ and $\alpha > 0$ be such that $\operatorname{diam} S^*(x, A^*, \alpha) < \varepsilon$. We denote by $J(x)$ the embedding of $x \in X$ into the double-dual X^{**} and by $T(J(x), A^*, \alpha)$ the slice of A^* in X^*. We have that

$$\begin{aligned} S^*(x, A^*, \alpha) &= \{x^* \in A^* \colon \langle x^*, x\rangle > \sigma_{A^*}(x) - \alpha\} = \\ &= \{x^* \in A^* \colon \langle x^*, x\rangle > \sup\{\langle y^*, x\rangle : y^* \in A^*\} - \alpha\} = \\ &= \{x^* \in A^* \colon \langle J(x), x^*\rangle > \sup\{\langle J(x), y^*\rangle : y^* \in A^*\} - \alpha\} = T(J(x), A^*, \alpha), \end{aligned}$$

Since J is an isometry, this equality implies that

$$\operatorname{diam} T(J(x), A^*, \alpha) = \operatorname{diam} S(x, A^*, \alpha) < \varepsilon.$$

Hence A^* admits arbitrarily small slices in X^*, i.e. it is dentable in X^*. ☐

1.7. Asplund spaces

Definition 22. The Banach space X is called an Asplund (resp. weak Asplund) space if any of the following equivalent conditions hold:

a) [Phe93, theorem 2.14] Every continuous convex function on a convex open subset D of X is Frechet (resp. Gateaux) differentiable at a dense G_δ subset of D.

b) [Phe93, definition 5.2] The dual space X^* has the Radon-Nikodym property.

c) [Phe93, theorem 5.12] Every nonempty weak* compact convex subset of X^* is the weak* closed convex hull of its weak* strongly exposed points.

1.8. Convex functions

Let X be a Hausdorff topological vector space[4] and D be a convex subset[2] of X.

Definition 23. A function $f : D \to \mathbb{R}$ is called convex if any of the following equivalent conditions hold:

a) For any two points $x, y \in D$ and any $t \in [0,1]$ we have

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y).$$

b) The epigraph[170]

$$\operatorname{epi} f := \{(x,a) \in X \times \mathbb{R} \colon f(x) \leqslant a\}$$

is convex.

Note that definitions do not require any topological structure on X. Most of their properties, however, require a topology.

Proof. Let $x, y \in D$ and let $t \in [0,1]$.
(23 (a) $\implies$ 23 (b)) Let $\operatorname{epi} f$ be a convex set. Obviously $(x, f(x)) \in D$ and $(y, f(y)) \in D$. By the convexity of $\operatorname{epi} f$, we have

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y).$$

Thus f is a convex function.
(23 (b) $\implies$ 23 (a)) Let f be convex. Let $a \geqslant f(x)$ and $b \geqslant f(y)$ so that $(x,a) \in \operatorname{epi} f$ and $(y,b) \in \operatorname{epi} f$. Hence

$$f(tx + (1-t)y) \leqslant tf(x) + (1-t)f(y) \leqslant ta + (1-t)b,$$

which implies that

$$(tx + (1-t)y, ta + (1-t)b) \in \operatorname{epi} f.$$

Thus $\operatorname{epi} f$ is a convex set. □

Proposition 24. *[Phe93, exercise 1.10] For any convex function f and any $x \in D$, the set $\partial f(x)$ is convex and weak* closed.*

Proof. Fix $x \in D$. If $\partial f(x)$ is empty, then the theorem is trivially true.
Suppose it is nonempty and $y^*, z^* \in \partial f(x)$. For any $x \in D$ we then have

$$\begin{cases} \langle y^*, x - x\rangle \leqslant f(x) - f(x), \\ \langle z^*, x - x\rangle \leqslant f(x) - f(x). \end{cases}$$

Fix $t \in [0,1]$ and $x \in D$. It follows that

$$\begin{aligned}\langle ty^* + (1-t)z^*, x - x\rangle = t\,\langle y^*, x - x\rangle + (1-t)\,\langle z^*, x - x\rangle &\leqslant \\ \leqslant t[f(x) - f(x)] + (1-t)[f(x) - f(x)] &= \\ = f(x) - f(x),&\end{aligned}$$

thus $ty^* + (1-t)z^* \in \partial f(x)$ and hence $\partial f(x)$ is convex.

To prove weak*-closedness, we consider the decomposition

$$\begin{aligned}\partial f(x) &= \{x^* \in E^* \colon \forall x \in D, \langle x^*, x - x\rangle \leqslant f(x) - f(x)\} = \\ &= \bigcap_{x\in D} \{x^* \in E^* \colon \langle x^*, x - x\rangle \leqslant f(x) - f(x)\} = \\ &= \bigcap_{x\in D} L(x)^{-1}(-\infty, f(x) - f(x)],\end{aligned}$$

where

$$L : E \to E^{**} L(x)(x^*) = \langle x^*, x - x\rangle\,.$$

For each $x \in E$, the functionals $L(x)$ are weak*-to-weak continuous because the image $L(E) \subseteq E^{**}$ is isometrically isomorphic to a translation of E. Hence the preimage $L(x)^{-1}(-\infty, f(x) - f(x)]$ is closed and $\partial f(x)$ is weak*-closed as the intersection of weak*-closed sets. □

Lemma 25. *For every point $x \in X$ and every direction $h \in S_X$ the difference quotient is a monotone function of $t > 0$, i.e. for $0 < s < t$*

$$\frac{f(x+sh) - f(x)}{s} \leqslant \frac{f(x+th) - f(x)}{t}$$

Proof.

$$\begin{aligned}\frac{f(x+sh) - f(x)}{s} = \frac{t}{s}\frac{f(x+\frac{s}{t}th) - f(x)}{t} = \frac{t}{s}\frac{f\left(\frac{s}{t}(x+th) + (1-\frac{s}{t})x\right) - f(x)}{t} \leqslant \\ \leqslant \frac{t}{s}\frac{\frac{s}{t}f(x+th) + (1-\frac{s}{t})f(x) - f(x)}{t} = \frac{t}{s}\frac{s}{t}\frac{f(x+th) - f(x)}{t} = \frac{f(x+th) - f(x)}{t}\end{aligned}$$

□

Proposition 26. *For every point $x \in X$ and every direction $h \in S_X$ the one-sided derivative $f'_+(x)(h)$ exists.*

Proof. We use the convexity of f to obtain

$$\begin{aligned}f(x) = f\left(x + \frac{th}{2} - \frac{th}{2}\right) &\leqslant \frac{f(x+th) + f(x-th)}{2}, \\ 0 &\leqslant [f(x-th) - f(x)] + [f(x+th) - f(x)], \\ -[f(x-th) - f(x)] &\leqslant [f(x+th) - f(x)], \\ -\frac{f(x+t(-h)) - f(x)}{t} &\leqslant \frac{f(x+th) - f(x)}{t},\end{aligned}$$

thus the difference quotient in $f'_+(x)(h)$ is bounded below by the difference quotient for $-f'_+(x)(-h)$.

Lemma 25 implies that the right difference quotient is non-increasing, thus both limits exist and

$$-f'_+(x)(-h) \leqslant f'_+(x)(h).$$

□

Proposition 27. *For every point $x \in X$ and every direction $h \in S_X$ the one-sided derivative $f'_+(x)(h)$ is a sublinear functional.*

Proof.

Positive homogeneity For $\lambda > 0$ the equality $f'_+(x)(\lambda h) = \lambda f'_+(x)(h)$ follows from

$$\frac{f(x + t\lambda h) - f(x)}{t} = \lambda \frac{f(x + t\lambda h) - f(x)}{t\lambda}$$

Subadditivity It follows directly from

$$\begin{aligned}\frac{f(x + t(a + b)) - f(x)}{t} &= \frac{f(\frac{1}{2}(x + 2ta) + \frac{1}{2}(x + 2tb)) - f(x)}{t} \leqslant \\ &\leqslant \frac{\frac{1}{2}f(x + 2ta) + \frac{1}{2}f(x + 2tb) - f(x)}{t} = \\ &= \frac{f(x + 2ta) - f(x)}{2t} + \frac{f(x + 2tb) - f(x)}{2t}.\end{aligned}$$

□

Corollary 28.

$$-f'_+(x)(-h) \leqslant f'_+(x)(h)$$

Proof.

$$0 = f'_+(x)(h + (-h)) \leqslant f'_+(x)(h) + f'_+(x)(-h)$$

□

Proposition 29. *The continuous function $f : D \to X$ is convex if and only if its subdifferential $\partial f(x)$[8] [(a)] is nonempty for every x in D.*

Proposition 30. *For every direction $h \in S_X$, we have that*

$$f'_+(x)(h) = \max\{\langle x^*, h\rangle : x^* \in \partial f(x)\}.$$

Theorem 31. *If f is continuous and if the subdifferential $\partial f(x)$ at $x \in X$ is a singleton with element x^*, then f is Gateaux differentiable at x and $f'_G(x) = x^*$.*

Proof. Let $h \in S_X$ be arbitrary. Proposition 26 implies that the one-sided derivatives $f'_+(x)(-h)$ and $f'_+(x)(h)$ exist and

$$-f'_+(x)(-h) \leqslant f'_+(x)(h).$$

Assume$^{\text{LEM}}$ that f is not Gateaux differentiable at x, i.e. for some $h_0 \in X$, we have a strict inequality. Then by proposition 30

$$\min\{\langle x^*, h_0\rangle : x^* \in \partial f(x)\} = -\max\{\langle x^*, -h_0\rangle : x^* \in \partial f(x)\} = -f'_+(x)(-h_0) < \\ < f'_+(x)(h_0) = \max\{\langle x^*, h_0\rangle : x^* \in \partial f(x)\},$$

which implies that there is more that one functional $x^* \in \partial_C f(x)$. This contradicts the assumption of the theorem.

Thus f is Gateaux differentiable at x. □

Theorem 32. *[Phe93, exercise 1.15(b]) In $\mathbb{R}^n$, the existence of the partial derivatives at x for a continuous convex function $f : D \to \mathbb{R}$ at a point $x \in D$ implies Gateaux differentiability.*

Proof. Let $D \subseteq \mathbb{R}^n$ be an open and convex set and let $f : D \to \mathbb{R}$ be continuous and convex. Then $f'_+(x)$ exists everywhere by proposition 26 and is a subdifferential functional by proposition 27.

Let $e_1, \ldots, e_n$ be the canonical basis for $\mathbb{R}^n$.

The partial derivatives

$$\frac{\partial f}{\partial x_i}(x) := \lim_{t \to 0} \frac{f(x + te_i) - f(x)}{t} = f'_+(x)(e_i)$$

exist, hence the projections of $f'_+(x)$ along the coordinate exes are linear.

Define line linear functional

$$l(h) := \sum_{i=1}^{n} h_i \left\langle \frac{\partial f}{\partial x_i}(x), h \right\rangle,$$

where $h_1, \ldots, h_n$ are the coordinates of h along $e_1, \ldots, e_n$.

We will show that $l \equiv f'_+$. Fix $h \in S_X$. We have

$$\begin{aligned} f'_+(x)(h) &= f'_+(x)\left(\sum_{i=1}^{n} h_i e_i\right) \overset{\text{sublinearity}}{\leqslant} \\ &\leqslant \sum_{i=1}^{n} f'_+(x)(h_i e_i) \overset{\text{linearity along } e_i}{=} \\ &= \sum_{i=1}^{n} h_i f'_+(x)(e_i) = \sum_{i=1}^{n} h_i \left\langle \frac{\partial f}{\partial x_i}(x), h \right\rangle. \end{aligned} \tag{1.4}$$

Thus

$$\langle l, h\rangle = -\langle l, -h\rangle \overset{\textit{eq. } (1.4)}{\leqslant} -f'_+(x)(-h) \overset{\text{corollary 28}}{\leqslant} f'_+(x)(h) \overset{\textit{eq. } (1.4)}{\leqslant} \langle l, h\rangle,$$

i.e. $f'_+(x)(h) = \langle l, h\rangle$ for all $h \in S_X$, hence $f'_+(x)$ is a linear functional and f is Gateaux differentiable at x. □

Theorem 33. *[Phe93, exercise 1.15(a]) In $\mathbb{R}^n$, Gateaux differentiability of a continuous convex function $f : D \to \mathbb{R}$ at a point $x \in D$ implies Frechet differentiability.*

Proof. Since f is Gateaux differentiable (definition 7 (c)) at x, the derivative $f'(x) = f'_+(x)$ is linear.

Because f is continuous and convex, it is locally Lipschitz with constant L in some δ-ball with center x.

Suppose$^{\text{LEM}}$ that f is not Frechet differentiable at x. Inverting the condition in definition 7 (d), we obtain that there exist $\varepsilon > 0$ and a sequence $\{h_n\}_n \subseteq B(x, \delta)\backslash\{0\}$ such that $\|h_n\| \to 0$ and yet for all $n \in \mathbb{Z}^{>0}$,

$$\left|f(x + h_n) - f(x) - \left\langle f'(x), h_n\right\rangle\right| > \varepsilon \left\|h_n\right\|. \tag{1.5}$$

Define

$$t_n := \|h_n\| \qquad\qquad u_n := \frac{h_n}{\|h_n\|}.$$

Obviously $t_{n_k} \downarrow 0$. The vectors h_n are linearly independent since otherwise f would not be Gateaux differentiable at x, hence u_n are not all equal.

Since $S_{\mathbb{R}^n}$ is compact$^{\text{USC}}$, by the Bolzano-Weierstrass theorem, there exists a convergent subsequence $\{u_{n_k}\}_k \underset{k\to\infty}{\to} u_0$ of $\{u_n\}_n$. We have

$$\begin{aligned}
&\left|\frac{f(x + t_{n_k}u_{n_k}) - f(x)}{t_{n_k}} - \left\langle f'(x), u_{n_k}\right\rangle\right| \leqslant \left|\frac{f(x + t_{n_k}u_{n_k}) - f(x + t_{n_k}u_0)}{t_{n_k}}\right| + \\
&+ \left|\frac{f(x + t_{n_k}u_0) - f(x)}{t_{n_k}} - \left\langle f'(x), u_0\right\rangle\right| + \left|\left\langle f'(x), u_0 - u_{n_k}\right\rangle\right| \leqslant \\
&\leqslant L\left\|u_{n_k} - u_0\right\| + \left|\frac{f(x + t_{n_k}u_0) - f(x)}{t_{n_k}} - \left\langle f'(x), u_0\right\rangle\right| + \left\|f'(x)\right\|\left\|u_0 - u_{n_k}\right\|.
\end{aligned} \tag{1.6}$$

Fix $\delta > 0$. Because of the Gateaux differentiable of f at x, we can pick k_0 such that

$$\left|\frac{f(x + t_{n_{k_0}}u_0) - f(x)}{t_{n_{k_0}}} - \left\langle f'(x), u_0\right\rangle\right| < \delta.$$

Because $\{u_{n_k}\}_k$ converges to u_0, we can choose k_1 such that

$$\left\|u_0 - u_{n_{k_1}}\right\| < \delta.$$

Thus for $k > \max\{k_0, k_1\}$, eq. (1.6) is bounded by

$$\left|\frac{f(x + t_{n_k}u_{n_k}) - f(x)}{t_{n_k}} - \left\langle f'(x), u_{n_k}\right\rangle\right| \leqslant (L + 1 + \left\|f'(x)\right\|)\delta.$$

It suffices to choose $\delta > 0$ so that

$$\delta < \frac{1}{L + 1 + \|f'(x)\|}$$

in order to have, for $k > \max\{k_0, k_1\}$,

$$\left| \frac{f(x + t_{n_k} u_{n_k}) - f(x)}{t_{n_k}} - \langle f'(x), u_{n_k} \rangle \right| < \varepsilon.$$

But this contradicts eq. (1.5), hence f is Frechet differentiable at x. □

Corollary 34. *In $\mathbb{R}^n$, the existence of the partial derivatives at x for a continuous convex function $f : D \to \mathbb{R}$ at a point $x \in D$ is equivalent to Frechet differentiability.*

Proof. A direct consequence of and theorem 32 and theorem 33. □

Theorem 35. *[Phe93, exercise 1.17] In $\mathbb{R}^n$, continuous convex functions $f : D \to \mathbb{R}$ are differentiable almost everywhere.*

Proof. For all $h \in S_X$ and small enough $t > 0$ we define

$$\varphi_h^t : D \to \mathbb{R} \qquad\qquad \varphi_h^t(x) := \frac{f(x + th) - f(x)}{t}$$

and $\varphi_h(x) := f'_+(x)(h) = \lim_{t \downarrow 0} \varphi_h^t(x)$.

Considered as functions of x, φ_h^t are obviously continuous hence Borel measurable and so φ_h is also Borel measurable.

Denote by

$$B_h := \{x \in D \colon -f'_+(x)(-h) < f'_+(x)(h)\} = \{x \in D \colon -\varphi_{-h}(x) - \varphi_h(x) < 0\}$$

the set of points $x \in D$ where the one-sided derivative $f'_+(x)(h)$ is not linear, given a fixed direction $h \in S_X$. If B_h is nonempty, f is not differentiable at x.

The sets B_h are Borel sets since they are the preimages of $(-\infty, 0)$ under a Borel function. We will show that it is a null set for every direction h.

Fix $h \in S_X$. Denote by $\delta_x := \sup\{t > 0 \colon x + th \in D\}$.

The function $t \mapsto f(x + th)$ is a convex function of one variable. By [Phe93, theorem 1.16], it is differentiable μ_1-almost everywhere in $[0, \delta_x)$, where μ_m is the Lebesgue m-measure.

Denote

$$H := \operatorname{span}\{h\} \equiv \mathbb{R}^1,$$
$$H^\perp \equiv \mathbb{R}^{n-1} \text{ - the orthogonal complement of } H \text{ in } \mathbb{R}^n,$$
$$L_x := \{x + th, 0 \leqslant t < \delta_x\} - half - opensegmentsinD.$$

THe whole domain D can be represented as $D = \cup\{L_x \colon x \in H^\perp\}$.

We can now use Fubini's theorem to show that B_h is a null set:

$$\mu_n(B_h) = \int_{B_h} dz = \int_{\mathbb{R}^n = H^\perp \oplus H} \mathrm{I}_{B_h}(z)dz = \int_{H^\perp} \int_{L_x} \mathrm{I}_{B_h}(y)dydx =$$
$$= \int_{H^\perp} \mu_1(B_h \cap L_x)dx = \int_{H^\perp} 0dx = 0.$$

Hence for all $h \in S_X$, $-f'_+(x)(-h) = f'_+(x)(h)$ for almost all $x \in D$.

In particular, if $e_1, \ldots, e_n$ is the canonical basis of $\mathbb{R}^n$, the i-th partial derivative $\frac{\partial f}{\partial x_i}(x)$ exists only in $D\ B_{e_i}$.

The gradient

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)$$

then exists in

$$\hat{D} := (D\ B_{e_1}) \cap \ldots \cap (D\ B_{e_n}) = D \setminus \left(\bigcup_{i=1}^{n} B_{e_i} \right).$$

Corollary 34 then implies that f is Frechet differentiable in $\hat{D}$, i.e. almost everywhere in D. □

1.9. Clarke generalized gradients

Let X be a Banach space and $f : X \to \mathbb{R}$ be locally Lipschitz.

Definition 36. [Cla13, definition 10.3] Let $x \in X$ and $U \subseteq X$ be a neighborhood of x where f is L-Lipschitz, i.e.

$$\forall y, z \in U, |f(y) - f(z)| \leqslant L \|y - z\|.$$

We use the Clarke generalized derivative[6] [(d)],

$$f^\circ(x)(h) := \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + th) - f(y)}{t}$$

We define the generalized gradient of f at x to be the set

$$\partial_C f(x) := \{x^* \in X^* : \forall h \in X, f^\circ(x)(h) \geqslant \langle x^*, h \rangle\}.$$

We say that the vector h is a descent direction of f at x if

$$\limsup_{t \downarrow 0} \frac{f(x + th) - f(x)}{t} < 0.$$

Proposition 37. *The generalized derivative of a locally Lipschitz function $f : X \to \mathbb{R}$ exists for every $x \in X$.*

Proof. Let $x, h \in X$ and let U be a neighborhood of x where the Lipschitz condition holds with the constant L_U. Then there exists $\delta_0 > 0$ such that $B(x, \delta_0) \subseteq U$.

Define $\delta_1 := \frac{1}{2} \min \left\{ \delta_0, \frac{\delta_0}{\|h\|} \right\} < \delta_0$, so that for $y \in B(x, \delta_1)$ and $t \in (0, \delta_1)$ we have

$$\|(y + th) - x\| \leqslant \|y - x\| + t \|h\| \leqslant \delta_1 + \delta_1 \|h\| \leqslant \begin{cases} \frac{\delta_0}{2}(1 + \|h\|), & \|h\| \leqslant 1 \\ \frac{\delta_0}{2\|h\|}(1 + \|h\|), & \|h\| > 1. \end{cases}$$

In both cases we get that $y + th \in B(x, \delta_0)$.

The generalized derivative in x in the direction $h \in X$ is then norm-bounded by

$$\begin{aligned} |f^\circ(x)(h)| &= \left| \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + th) - f(y)}{t} \right| = \left| \lim_{\delta \to 0} \sup_{\substack{y \in B(x,\delta) \\ t \in (0,\delta)}} \frac{f(y + th) - f(y)}{t} \right| \leqslant \\ &\leqslant \left| \sup_{\substack{y \in B(x,\delta_1) \\ t \in (0,\delta_1)}} \frac{f(y + th) - f(y)}{t} \right| \leqslant \sup_{\substack{y \in B(x,\delta_1) \\ t \in (0,\delta_1)}} \frac{|f(y + th) - f(y)|}{t} \leqslant \\ &\leqslant \sup_{\substack{y \in B(x,\delta_1) \\ t \in (0,\delta_1)}} \frac{\|(y + th) - (y)\|}{t} = \|h\|. \end{aligned}$$

The fact that f is locally Lipschitz gave us that the supremum is taken over a bounded set and thus the generalized derivative exists. ☐

2. Algebra

2.1. Groups

Definition 38. (See example 129) A magma is a nonempty set G with a function $\cdot : G \times G \to G$, called the magma operation. Denote $\cdot(a, b)$ by juxtaposition, that is, ab.

Associativity If $(ab)c = a(bc)$ for all $a, b, c \in G$, we say that the operation is associative and we call the pair $(G, \cdot)$ a semigroup.

Identity An identity is an element $e \in G$ such that $ea = ae = a$ for every $a \in G$. If an identity exists, it is unique by definition 41 (a). If $(G, \cdot)$ is a semigroup with an identity, we call it a monoid (see also definition 239 and definition 43). We can also consider the identity in a monoid to be a function with arity 0. If the monoid is not obvious from the context, we write e_G.

Inverse For each $a \in G$, the inverse is an element $a^{-1} \in G$ such that $aa^{-1} = e$. It is unique for each element by definition 41 (b). If $(G, \cdot)$ is a monoid and each element has an inverse, we call it a group. We can also consider the inverse in a group to be a function with arity 1.

Commutativity If $ab = ba$ for all $a, b \in G$, we say that the magma operation is commutative. If $(G, \cdot)$ is a group with a commutative operation, we call it a commutative group or abelian group.

We define a^n for integer n as

$$a^n := \begin{cases} e, & n = 0 \\ a \cdot a^{n-1}, & n > 0 \\ (a^{-n})^{-1}, & n < 0. \end{cases}$$

We say that

a) the group $\{e\}$ is called the trivial group.

b) the subset $H \subseteq G$ is a submagma (resp. subsemigroup, submonoid, subgroup, etc) of G if H is closed under the magma operation, that is, $ab \in H$ whenever $a, b \in H$.

c) the subgroup $\{e_G\}$ of G is called the trivial subgroup

d) all subgroups except for G itself are proper subgroups.

e) a is of finite order $\operatorname{ord}(a) = n$ if n is the smallest positive integer such that $a^n = e$.

f) a is of infinite order $\operatorname{ord}(a) = \infty$ if $a^n \neq e$ for any positive integer n.

g) a is an absorbing element if $ba = ab = a$ for any $b \in G$.

Note 39. Groups are often used to describe sets of invertible functions[169] where the group operation is composition (see definition 43 for a categorical viewpoint). As such, the group operation is usually denoted by juxtaposition as in definition 38.

Since composition of functions is not commutative in general, abelian groups are usually not sets of invertible functions. Since abelian groups are $\mathbb{Z}$-modules [73], we usually denote the group operation in abelian groups by $a + b$ instead of ab, the inverse by $-a$ instead of a^{-1}, and the unit by 0.

To make a further distinction, if the operation is denoted by juxtaposition, we say that the group is a multiplicative group, and if the operation is denoted by $+$, we say that the group is an additive group. This terminology usually, but not necessarily, coincides with the group being abelian.

Definition 40. Let A be an arbitrary set. We associate with A the symmetric group $S(A)$ whose elements are bijections from A to itself and whose group operation is function composition.

Proposition 41. *Any group G has the following basic properties:*

a) The identity e is unique (also holds for monoids).

b) The inverse a^{-1} of every element is unique.

c) The identity e is its own inverse.

d) $(ab)^{-1} = b^{-1}a^{-1}$.

e) For any $a \in G$, $a = (a^{-1})^{-1}$

f) For any $a \in G$ and positive integer n, $a^{-n} = (a^n)^{-1} = (a^{-1})^n$

g) For any $a, b, c \in G$, the cancellation laws $a = b \iff ac = bc \iff ca = cb$ hold.

Proof.

41 (a) If e' is another identity element of G, then $e' = e'e = e$.

41 (b) If b and c are both inverses of a, then $b = eb = cab = ce = c$.

41 (c) $ee = e$.

41 (d)

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = e = b^{-1}(a^{-1}a)b = (b^{-1}a^{-1})(ab).$$

41 (e)

$$(a^{-1})^{-1} = aa^{-1}(a^{-1})^{-1} = a.$$

41 (f) Using 41 (e),

$$a^{-n} = (a^n)^{-1} = a^{-1} \cdots a^{-1} = (a^{-1})^n.$$

41 (g) If $a = b$, obviously $ac = bc$ and $ca = cb$. Now if $ac = bc$, we have

$$a = acc^{-1} = bcc^{-1} = b.$$

The case $ca = bc$ is analogous.

□

Definition 42. A groupoid is a category[175] in which all morphisms are isomorphisms[179].

Definition 43. Let $\mathbf{G}$ be a locally small with a single object g. Then the endomorphisms of g form a monoid[38] under composition and the subcategory of $\mathbf{G}$ in which all morphisms are isomorphisms forms a group[38]. Thus, if $\mathbf{G}$ is a locally small groupoid with a single object g, then the endomorphisms of g are automorphisms and, thus, they form a group under composition.

Definition 44. Let $H \subseteq G$ be a subgroup of G and let $a \in G$. Consider the sets

$$aH := \{ah \colon h \in H\} \qquad Ha := \{ha \colon h \in H\}$$

called the left and right cosets of H with respect to a.

Proposition 45. *The left (resp. right) cosets of a subgroup H of G partition[163] G.*

Proof. To each element $a \in G$ there corresponds a coset $a \in aH$ (since H contains the identity as a subgroup).

Two cosets aH and bH are either disjoint or equal. Indeed, if they are not disjoint, then there exists $g \in aH \cap bH$ and thus $g = ax = by$ for some $x, y \in H$. Thus $a = axx^{-1} = byx^{-1}$ and since $yx^{-1} \in H$, we have that $a \in bH$. Furthermore, for any $z \in H$, we have $az = b(yx^{-1}z) \in bH$, hence $aH \subseteq bH$. After obtaining the converse inclusion, we conclude $aH = bH$. □

Definition 46. Let M and N be magmas. We say that the function $f : M \to N$ is a magma homomorphism (resp. semigroup, monoid or group homomorphism) if it is compatible with the magma structure on M and N, that is,

Compatibility $f(ab) = f(a)f(b)$ for all $a, b \in M$.

If M and N are monoids, the kernel of f is defined as the preimage[169] $f^{-1}(e)$ of the identity e_N.

The terminology from definition 179 applies to group homomorphisms because of the category $\mathbf{Grp}$ of groups.

Proposition 47. *Any group homomorphism $f : G \to K$ has the following basic properties:*

a) f preserves the identity, that is, $f(e_G) = e_K$ (also holds for monoids).

b) f preserves inverses, that is, $f(a^{-1}) = f(a)^{-1}$ for any $a \in G$.

c) *If H is a subgroup of G, its image $f(H)$ is a subgroup of K (holds for any submagma).*

d) *The kernel* $\ker(f)$ *is a subgroup of G.*

Proof.

47 (a) We have

$$e_K f(e_G) = f(e_G e_G) = f(e_G) f(e_G).$$

Cancelling on the right, we obtain

$$e_K = f(e_G).$$

47 (b)

$$f(a^{-1}) = f(a^{-1})e_K = f(a^{-1})f(a)f(a)^{-1} = f(a^{-1}a)f(a)^{-1} = f(a)^{-1}.$$

47 (c) If $a, b \in G$, then

$$f(a)f(b) = f(ab) \in f(G),$$

thus $f(G)$ is closed under the group operation in K.

47 (d) If $a, b \in \ker(f)$, then

$$f(ab) = f(a)f(b) = e_K e_k = e_k.$$

Thus $ab \in \ker(f)$ and $\ker(f)$ is closed under the group operation.

□

Definition 48. Let N be a subgroup of G. We say that N is a normal subgroup and write $N \trianglelefteq G$ if any of the following equivalent conditions hold:

a) For any $a \in G$, we have the set equality $aNa^{-1} = N$.

b) The partitions induced by the left and rights cosets of N coincide ($aN = Na$) and form the quotient group G/N.

c) N is the kernel[46] of some group homomorphism (in particular, kernels are always normal subgroups).

Proof. This is the group-theoretic analog to proposition 164.

(48 (a) $\implies$ 48 (b)) For any $a \in G$

$$Na = (aNa^{-1})a = aN(a^{-1}a) = aN,$$

thus every left coset is a right coset and vice versa.

(48 (b) $\implies$ 48 (c)) Denote by G/N the family of all cosets of N. Then G/N is itself a group with the inherited from G group structure

- $aN \cdot bN := (ab)N$
- $N = eN$ is an identity element of P
- $a^{-1}N$ is the inverse of aN

Since it is possible for two elements $a, x \in G$ to have the same coset $aN = xN$, the group operation in G/N depends on the choice of representatives for each coset. In order for the operation to be well-defined, we need to make sure that the result does not depend on the choice of representatives. This happens to be true if and only if the subgroup N is normal.

Indeed, let $aN = xN$ and $bN = yN$. If N is normal (in the sense of 48 (b)), we have

$$(ab)N = a(yN) = a(Ny) = (aN)y = x(Ny) = (xy)N.$$

Conversely, if the operation is well defined, then for any $g \in N$

$$N = a^{-1}aN = (a^{-1}N)(aN) = (a^{-1}N)N(aN) = (a^{-1}N)(gN)(aN) = (a^{-1}ga)N.$$

Hence $a^{-1}ga \in N$ and $a^{-1}Na \subseteq N$. Thus

$$\begin{aligned} Na &= (aa^{-1}N)a = (aN)(a^{-1}Na) \subseteq (aN)N = aN, \\ aN &= a(Na^{-1}a) = (aNa^{-1})a \subseteq Na. \end{aligned}$$

Now define the homomorphism

$$\begin{aligned} &\varphi : G \to G/N \\ &\varphi(a) = aN \end{aligned}$$

The preimage of the coset N consists of N itself. Since N (as an element of G/N) is the identity of G/N, we conclude that N (as a subset of G) is the kernel of φ.

(48 (c) $\implies$ 48 (a)) Let $f : G \to K$ be a group homomorphism and fix any $a \in G$. Denote $N := \ker(f)$. Then $aN = Na$ since

$$f(aN) = f(a)f(N) = f(a)f(e_G) = f(a) = f(N)f(a) = f(Na).$$

Thus

$$N = aa^{-1}N = aNa^{-1}.$$

□

Proposition 49. *All subgroups of an abelian group are normal.*

Proof. Let G be abelian and H be a subgroup of G. Then $aGa^{-1} = aa^{-1}H = H$ for any $a \in G$ and thus H is normal. □

Definition 50. [Kna16, p. 306] Let A be an arbitrary set. We associate with A its free monoid $FM(A) := (\mathcal{A}^*, \cdot)$[119]. It is a monoid due to proposition 120. In category-theoretic terms, we define a functor[190] $FM : \mathbf{Set} \to \mathbf{Mon}$ that is left-adjoint[211] to the corresponding forgetful functor.

Definition 51. [Kna16, p. 306] Let A be an arbitrary set. We associate with A a free group $F(A)$. As indefinition 50, $F(S)$ is left-adjoint to the corresponding forgetful functor.

First, regard A as an alphabet[119] and define the set A^{-1} of words of the type a^{-1}, where $a \in A$ and $^{-1}$ is a symbol not in A. Consider the language $W := (A \cup A^{-1})^*$. If $w = a_1 \dots a_n$ is a word, its inverse word is

$$w^{-1} := a_n^{-1} \dots a_1^{-1}.$$

By proposition 120, W is a monoid. It is not a group, however, since $ww^{-1} \neq \varepsilon$. This is why we define a different group operation on a subset of W.

We define the reduction function

$$\begin{aligned} &r : W \to W \\ &r(w) := \begin{cases} r(ps), & w = pvs, \text{ where } v = aa^{-1} \text{ or } v = a^{-1}a \text{ for some } a \in A, \\ w, & \text{otherwise.} \end{cases} \end{aligned}$$

The set of reduced words is the image $F(A) := \operatorname{im} r(W)$. It forms a group under the operation

$$u \cdot_{F(A)} w := r(u \cdot_W w).$$

The group $(F(A), \cdot_{F(A)})$ is called the free group generated by A.

Definition 52. Let $\{X_i\}_{i \in I}$ be a nonempty family of groups.

We define their direct product as the group $\prod_{i \in I} X_i$, the group operation defined componentwise as

$$\{x_i\}_{i \in I} \cdot \{y_i\}_{i \in I} := \{x_i \cdot y_i\}_{i \in I}.$$

We define their direct sum as the subgroup of $\prod_{i \in I} X_i$[52] where only finitely many components of any group element are different from zero.

Note 53. [Kna16, p. 126] If we are given a family of groups as in definition 52, their sum $\bigoplus_{i \in I} X_i$ is sometimes called an external direct sum.

If instead we are given a group X and a family of subgroups $\{X_i\}_{i \in I}$, we say that X is their internal direct sum if the homomorphism

$$\begin{aligned} &\varphi : \prod_{i \in I} X_i \to X \\ &\varphi(\{x_i\}_{i \in I}) := \cdot_{i \in I} x_i \end{aligned}$$

is an isomorphisms.

The sum is well-defined since by definition there are only finitely many non-identity summands.

This terminology also applies to finite direct products[52] of groups, as well as similar constructions for other algebraic structures.

Definition 54. [Kna16, p. 323] Let $\{X_i\}_{i \in I}$ be a nonempty family of groups.

Similarly to definition 51, consider the language[119]

$$W := \left(\bigcup_{i \in I} X_i \right)^* .$$

We define the reduction function

$$r : W \to W$$

$$r(w) := \begin{cases} r(ps), & w = pe_i s \text{ for some } i \in I, \\ r(pts), & w = puvs, \text{ where } u \neq e_i \text{ and } v \neq e_i \text{ and } t = u \cdot_i v \text{ for some } i \in I, \\ w, & \text{otherwise.} \end{cases}$$

The set of reduced words is the image $\ast_{i \in I} X_i := \operatorname{im} r(W)$. It forms a group under the operation

$$u \cdot_\ast w := r(u \cdot_W w).$$

The group $(\ast_{i \in I} X_i, \cdot_\ast)$ is called the free product of the family $\{X_i\}_{i \in I}$.

Definition 55. The class[139] of all groups forms the category[175] **Grp**, where for every two groups $X, Y \in \mathbf{Grp}$, the morphisms $\mathbf{Grp}(X, Y)$ are the homomorphisms[46] from X to Y and composition is the usual function composition[168].

The category **Ab** of abelian groups is a full subcategory of **Grp**.

Both categories are concrete[213], while **Ab** is abelian[250].

Proposition 56. *We are interested in categorical limits[223] and colimits[231] in* **Grp**. *If* $\{X_i\}_{i \in I}$ *is an indexed family of groups, then*

a) their categorical product[224] is their direct product[52] $\prod_{i \in I} X_i$*, the projection morphisms being inherited from definition 159 (a).*

b) their categorical coproduct[233] is their free product[54] $\ast_{i \in I} X_i$*, the injection morphisms being*

$$\iota_j : X_j \to \ast_{i \in I} X_i$$
$$\iota_j(x_j) := x_j .$$

Proposition 57. *We are interested in categorical limits[223] and colimits[231] in* **Ab**. *If* $\{X_i\}_{i \in I}$ *is an indexed family of abelian groups, then*

a) *their categorical product*[224] *is the direct product as inherited from proposition 56.*

b) *their categorical coproduct*[233] *is the direct sum*[52] $\oplus_{i \in I} X_i$*, the injection morphisms being*

$$\iota_j : X_j \to \oplus_{i \in I} X_i$$
$$\iota_j(x_j) := \left\{ \begin{matrix} x_j, & i = j \\ e_i, & i \neq j \end{matrix} \right\}_{i \in \mathbf{I}}.$$

Since **Ab** *is a subcategory of* **Grp***, by proposition 56 we have that for abelian groups the notions of free product*[54] *and direct sum coincide.*

Note 58. By theorem 248, finite direct products and finite direct sums of abelian groups coincide as biproducts. This is also obvious by definition, even for nonabelian groups. What is not obvious, however, is that finite free products and finite direct products coincide for abelian groups.

Definition 59. As a special case of definition 80, since abelian groups are $\mathbb{Z}$-modules by definition 72, we define free abelian groups to be free $\mathbb{Z}$-modules.

This definition is different from free groups[51].

2.2. Rings

Definition 60. A ring is an (additive[39]) abelian group[38] $(R, +)$ with an additional multiplication operation $\cdot : R \times R \to R$ (denoted by juxtaposition), such that for all $a, b, c \in R$ the following axioms hold

Associativity $(ab)c = a(bc)$

Left distributivity $(a + b)c = ab + bc$

Right distributivity $a(b + c) = ab + ac$

We say that

- the ring $\{0\}$ is the zero ring or trivial ring.
- the subset $S \subseteq R$ is a subring of R if S is closed under the ring operations.
- the ring $\{0_R\}$ is the trivial subring of R.
- all subrings except for R itself are proper subrings.
- $a \neq 0$ is a (left) zero divisor (resp. right zero divisor) if there exists $b \neq 0$ such that $ab = 0$ (resp. $ba = 0$).
- a is a (left) unit (resp. right unit) if there exists a^{-1} such that $a \cdot a^{-1} = 1$ (resp $a^{-1} \cdot a = 1$).
- a is nilpotent if a^n for some nonnegative integer n.
- a is idempotent if $aa = a$.

Additionally, the following axioms define different types of rings

Identity If $(R, \cdot)$ is a monoid, that is if there exists a multiplicative identity 1_R such that $1_R a = a 1_R = a$ for all $a \in R$, we say that $(R, \cdot)$ is a ring with identity or unital ring. It is unique by definition 41 (a). This is sometimes taken to be part of the definition of a ring.

Commutativity If $(R, \cdot)$ is a commutative semigroup, i.e. $ab = ba$ for all $a, b \in R$, we say that $(R, \cdot)$ is a commutative ring.

No zero divisors If the ring is a commutative ring and there are no zero divisors in an, we say that $(R, \cdot)$ is an integral domain.

Divisibility If all nonzero elements are units, we say that $(R, \cdot)$ is a division ring.

If we only require the ring to be a monoid under addition (i.e. no inverse elements), we say that $(R, +, \cdot)$ is a semiring.

Proposition 61. *Any ring R has the following basic properties:*

a) *Multiplication by* 0 *is* *absorbing*[38] [(g)], *that is,* $a0 = 0a = 0$ *for any* $a \in R$.

Proof.

61 (a) We have that $0a = (0 + 0)a = 0a + 0a$, thus $0a$ is an additive identity and $0a = 0$. We obtain $a0 = 0$ analogously.

□

Definition 62. Let R and T be rings. We say that the function $f : R \to T$ is a ring homomorphism if it is compatible with the ring structures on S and T, that is, it is a group homomorphism on $(R, +)$ and a semigroup homomorphism on $(R, \cdot)$.

If R is a ring with identity, we additionally require $f(1_R) = 1_T$.

The kernel of f is defined as the preimage[169] $f^{-1}(0)$.

The terminology from definition 179 applies to ring homomorphisms because of the category **Ring** of rings.

Definition 63. Let R be a ring and I be a subset of R (not necessarily a subring). We say that I is a (two-sided) ideal of R and write $I \trianglelefteq R$ if any of the following equivalent conditions hold:

a) $(I, +)$ is a subgroup of $(R, +)$ and the inclusions $RI \subseteq I$ and $IR \subseteq I$ hold.

b) I is the kernel[62] of some ring homomorphism.

We can weaken the condition in 63 (a) to define left ideals (resp right ideals) if only $RI \subseteq I$ (resp. $IR \subseteq I$) holds. If I is a subring, being either a left or right ideal is equivalent to I being a two-sided ideal. If R is a commutative ring, left and right ideals coincide with two-sided ideals.

If R is a ring without identity, all two-sided ideals are subrings. If R has an identity element, however, ideals are not necessarily subrings (see proposition 64).

As with subrings, the trivial ideal of R is the trivial subring and all ideals except for R itself are called proper ideals.

Proof. (63 (a) $\iff$ 63 (b)) Definition 48 implies that $(I, +)$ is a normal subgroup if and only if there exists a group homomorphism $f : (R, +) \to (T, +)$, where $(T, +, \cdot)$ is some ring, such that $I = \ker(f)$. Note that since the additive group is abelian, by proposition 49, all subgroups are normal.

Additionally, we have $f(RI) = f(R)f(I)$, thus $RI \subseteq I$ if and only if I is the kernel of f. □

Proposition 64. *If* R *is a ring with identity, the ideal* I *is proper if and only if* $1 \notin I$.

Proof. We will prove that $1 \in I \iff I = R$.

($\Longrightarrow$) Let $1 \in I$. Then $r1 = r$ for any $r \in R$, thus $RI = R$. Since I is an ideal, we have that $I = R$. ($\Longleftarrow$) If $I = R$, then obviously $1 \in I = R$. □

Definition 65. Let $\{X_i\}_{i \in I}$ be a nonempty family of rings.

Analogously to definition 52, we define their direct product as the ring $\prod_{i \in I} X_i$, the operations defined componentwise as

$$\{x_i\}_{i \in I} + \{y_i\}_{i \in I} := \{x_i + y_i\}_{i \in I},$$
$$\{x_i\}_{i \in I} \cdot \{y_i\}_{i \in I} := \{x_i \cdot y_i\}_{i \in I}.$$

We define their direct sum as the subring of $\prod_{i \in I} X_i$[65] where only finitely many components of any ring element are different from zero.

Definition 66. The class[139] of all rings forms the category[175] **Ring**, where for every two rings $X, Y \in \mathbf{Ring}$, the morphisms $\mathbf{Ring}(X, Y)$ are the homomorphisms[62] from X to Y and composition is the usual function composition[168].

Furthermore, **Ring** is locally small[177] and concrete[213].

Proposition 67. *We are interested in categorical limits[223] and colimits[231] in* **Ring**. *If* $\{X_i\}_{i \in I}$ *is an indexed family of rings, then*

a) their categorical product[224] is their direct product[65] $\prod_{i \in I} X_i$*, the projection morphisms being inherited from definition 159 (a).*

2.3. Fields

Definition 68. A field $(F, +, \cdot)$ is a nontrivial commutative division ring with identity[60].

2.4. Modules

Definition 69. [Kna16][374] Let R be a commutative ring (if R is not commutative, we can separately define left R-modules and right R-modules in the obvious way). An R-module M is an (additive) abelian group $(M,+)$ along with an operation $\circ : R \times M \to M$, called scalar multiplication and denoted by juxtaposition, such that for all $u, v \in M$ and all $a, b \in R$,

Associativity $a(bu) = (ab)u$.

Scalar distributivity $(a + b)u = au + bu$.

Vector distributivity $a(u + v) = au + av$.

Identity If the ring R has an identity, then we require the additional axiom $1_R u = u$.

In analogy with linear algebra, we call elements of R scalars and elements of M vectors. We say that

- the subset $N \subseteq M$ is an R-submodule of M if N is closed under the module operations.
- the module $\{0_R\}$ is the zero R-module or trivial R-module or the trivial submodule since it is a submodule of every R-module.
- all submodules except for M itself are proper submodules.

Example 70. Every commutative ring R is a module over itself. Every ideal $I \trianglelefteq R$ is an R-module since it is closed under multiplication with "scalars" from R.

Definition 71. [Kna16][375] Let M and N be two R-modules. We say that the function $f : M \to N$ is a module homomorphism if it is linear, that is,

Homogeneity $af(u) = f(au)$ for any $a \in R$ and $u \in M$.

Additivity $f(a + b) = f(a) + f(b)$ for any (that is, f is a group homomorphism between $(M,+)$ and $(N,+)$).

The kernel of f is defined as the preimage[169] $f^{-1}(0_M)$.
The terminology from definition 179 applies to module homomorphisms because of the category $\mathbf{Mod}_R$ of R-modules.

Definition 72. [Kna16][375] Let G be an abelian group. Define the $\mathbb{Z}$-module M, associated with G, with scalar multiplication

$$nu := \begin{cases} 0, & n = 0 \\ u + \ldots + u, & n > 0 \\ -((-n)u), & n < 0. \end{cases}$$

Thus, abelian groups can be regarded as modules.

Theorem 73. *[Kna16][375] Every abelian group is isomorphic to exactly one $\mathbb{Z}$-module.*

Proof. We already saw in definition 72 how every abelian group can be regarded as a $\mathbb{Z}$-module. Every $\mathbb{Z}$-module can then be identified with its additive group.

Scalar multiplication ensures that there is exactly one way to define a $\mathbb{Z}$-module structure on an abelian group since $na = (n-1)a + a$ and $0a = 0$. □

Definition 74. Let $\{X_i\}_{i \in I}$ be a nonempty family of R-modules.

Analogously to definition 52, we define their direct product as the module $\prod_{i \in I} X_i$, the operations defined componentwise as

$$\begin{aligned} \{x_i\}_{i \in I} + \{y_i\}_{i \in I} &\coloneqq \{x_i + y_i\}_{i \in I}, \\ \alpha\{x_i\}_{i \in I} &\coloneqq \alpha\{\alpha x_i\}_{i \in I}. \end{aligned}$$

We define their direct sum as the submodule of $\prod_{i \in I} X_i$[74] where only finitely many components of any module element are different from zero.

Definition 75. Fix a ring R. The class[139] of all R-module forms the category[175] $\mathbf{Mod}_R$, where for every two modules $X, Y \in \mathbf{Mod}_R$, the morphisms $\mathbf{Mod}_R(X, Y)$ are the homomorphisms[71] from X to Y and composition is the usual function composition[168].

Furthermore, $\mathbf{Mod}_R$ is concrete[213] and abelian[250].

Proposition 76. *We are interested in categorical limits[223] and colimits[231] in $\mathbf{Mod}_R$. If $\{X_i\}_{i \in I}$ is an indexed family of R-modules, then*

a) their categorical product[224] is their direct product[74] $\prod_{i \in I} X_i$, the projection morphisms being inherited from definition 159 (a).

b) their categorical coproduct[233] is the direct sum[52] $\bigoplus_{i \in I} X_i$, the injection morphisms being inherited from definition 57 (b).

Definition 77. Let M be an R-module and let $\alpha_1, \ldots, \alpha_n \in R$ and $x_1, \ldots, x_n \in M$. We call

$$x \coloneqq \sum_{k=1}^{n} \alpha_k x_k$$

their linear combination with coefficients or scalars $\alpha_1, \ldots, \alpha_n$ and vectors $x_1, \ldots, x_n$.

For a subset $A \subseteq M$, the set of all linear combinations of finite subsets of A is called its span and denoted by $\operatorname{span} A$.

Definition 78. Let M be an R-module and let $A \subseteq M$. We say that the set A is linearly dependent if there exists $x \in A$ such that

$$x \in \operatorname{span} A \backslash \{x\}.$$

If A is not linearly dependent, we say that it is linearly independent.

Definition 79. The subset B of the R-module M is called a basis of M if B is linearly independent and

$$M = \operatorname{span} B.$$

Definition 80. [32][KC16] Let R be a commutative unital ring. We say that the R-module M is a free module if it has a basis[79].

If S is any set, then the direct sum[76 (b)]

$$M := \oplus_{s \in S} R$$

with injections $\{\iota_s\}_{s \in S}$ is called the free module generated by S. Define the function

$$\varphi : S \to M$$
$$\varphi(s) := \iota_s(1_R).$$

The image $\varphi(S)$ is then a basis of M.

If the free module M has a finite basis, we say that it is finitely generated.

2.5. Vector spaces

Definition 81. A vector space $(V, +, \cdot)$ is a module[69] over a field F.

Note 82. Outside of algebra, we are usually only interested in vector spaces over the fields $\mathbb{R}$ or $\mathbb{C}$. We call them real vector spaces and complex vector spaces, respectively.

3. General topology

3.1. Topological spaces

Definition 83. [Eng89, p. 11] Let X be any set and $\mathcal{T} \subseteq \mathcal{P}(X)$ be a family of subsets of X. $\mathcal{T}$ is called a topology on X and the tuple $(X, \mathcal{T})$ is said to be a topological space if the following axioms are satisfied:

O1 $\varnothing, X \in \mathcal{T}$

O2 $U, V \in \mathcal{T} \implies U \cap V \in \mathcal{T}$

O3 $\mathcal{T}' \subseteq \mathcal{T} \implies \bigcap \mathcal{T}' \in \mathcal{T}$

If the topology is obvious from the context, we say that X is a topological space.

Elements of the set X are called points of the topological space, elements of $\mathcal{T}$ are called open sets and set-theoretic complements of open sets are called closed sets.

If $x \in U \in \mathcal{T}$, we say that U is a neighborhood of x. Note that some authors (e.g. [Kel55, p. 38]) alternatively define neighborhoods as arbitrary sets that contain an open set that contains x.

Dually, we can define the family $\mathcal{F}$ of closed sets, where

F1 $\varnothing, X \in \mathcal{F}$

F2 $U, V \in \mathcal{F} \implies U \cup V \in \mathcal{F}$

F3 $\mathcal{F}' \subseteq \mathcal{F} \implies \bigcup \mathcal{F}' \in \mathcal{F}$

If $(X, \mathcal{T})$ is a topological space, we denote the corresponding family of closed sets by

$$\mathcal{F}_{\mathcal{T}} := \{X \backslash U : U \in \mathcal{T}\}.$$

It is sometimes easier to define a topology $\mathcal{T}$ via a subset of $\mathcal{T}$. We will gradually construct a topology from a bare family of sets in X. First, we will give two definitions for a base, one on which does not require an existing topology.

Definition 84. [Eng89, p. 12] Fix a topological space $(X, \mathcal{T})$. We say that the family $\mathcal{B} \subseteq \mathcal{T}$ is a base for the topology $\mathcal{T}$ if $\mathcal{B}$ satisfies any of the equivalent conditions

a) Every open set $U \in \mathcal{T}$ is the union $U = \bigcup \mathcal{B}'$ of some subset $\mathcal{B}' = \mathcal{B}$

b) For any point $x \in X$ and for any neighborhood U of x there exists a set $V \in \mathcal{B}$ in the base such that $x \in V \subseteq U$

Proof. (84 (a) $\implies$ 84 (b)) Fix a point $x \in X$ and a neighborhood $U \in \mathcal{T}$ of x. Let $\mathcal{B}'$ be a subfamily of $\mathcal{B}$ such that

$$U = \bigcup \mathcal{B}'.$$

Then $x \in V$ for at least one $V \in \mathcal{B}'$.

(84 (b) $\implies$ 84 (a)) Fix an open set $U \in \mathcal{T}$. Then for every $x \in U$, there exists a set $V_x \in \mathcal{B}$ such that $x \in V_x \subseteq U$. We have

$$\bigcup_{x \in U} V_x \subseteq U \subseteq \bigcup_{x \in U} V_x,$$

thus

$$U = \bigcup_{x \in U} V_x.$$

□

Proposition 85. *[Eng89, p. 12] Let X be an arbitrary set and let $\mathcal{B}$ be a family of subset that satisfies*

B1 $\bigcup \mathcal{B} = X$

B2 $\forall U, V \in \mathcal{B}, \forall x \in U \cap V, \exists W \in \mathcal{B} : x \in W \subseteq U \cap V$

Then the family

$$\mathcal{T} := \left\{ \bigcup \mathcal{B}' : \mathcal{B}' \subseteq \mathcal{B} \right\} \tag{3.1}$$

is a topology on X. Furthermore, $\mathcal{B}$ is a base[84] *of $\mathcal{T}$.*

Proof. We will first prove that $\mathcal{T}$ is indeed a topology.

O1 $\varnothing = \bigcup \varnothing \in \tau$ and $X = \bigcup \mathcal{B} \in \mathcal{T}$ (by B1)

O3 Fix $\mathcal{T}' = \{U_\alpha : \alpha \in A\} \subseteq \mathcal{T}$. By definition 84 (a), every set U_α has a corresponding subfamily $\mathcal{B}_\alpha$ of $\mathcal{B}$ such that $U_\alpha = \bigcup \mathcal{B}_\alpha$.

Define $\mathcal{B}' := \bigcup_{\alpha \in A} \mathcal{B}_\alpha$. Obviously $\mathcal{B}' \subseteq \mathcal{B}$ and thus, by B1, $\bigcup \mathcal{B} \in \mathcal{T}$.

O2 Fix $U, V \in \mathcal{T}$ and families $\mathcal{B}_U, \mathcal{B}_V \subseteq \mathcal{B}$ such that $U = \bigcup \mathcal{B}_U$ and $V = \bigcup \mathcal{B}_V$.

Fix arbitrary $U' \in \mathcal{B}_U$ and $V' \in \mathcal{B}_V$. We will show that $U' \cap V' \in \tau$.

By B2, for every $x \in U' \cap V'$ there exists a neighborhood W_x of x such that $W \subseteq U' \cap V'$.

The family $\mathcal{B}_{U',V'} := \{W_x : x \in U' \cap V'\}^{\text{AOC}}$ is a subfamily of $\mathcal{B}$ and thus $U' \cap V' = \bigcup \mathcal{B}_{U',V'} \in \mathcal{T}$.

Hence, by O3, $U \cap V \in \tau$.

Now, for any $U \in \mathcal{T}$, by eq. (3.1), there exists a subfamily $\mathcal{B}' \subseteq \mathcal{B}$ such that

$$U = \bigcup \mathcal{B}'.$$

Hence $\mathcal{B}$ is a base for $\mathcal{T}$. □

Definition 86. [Eng89, p. 12] Fix a topological space $(X, \mathcal{T})$. We say that the family $\mathcal{P} \subseteq \mathcal{T}$ is a subbase for the topology $\mathcal{T}$ if $\mathrm{FI}(\mathcal{P})^{154}$) is a base[84] of $\mathcal{T}$.

Proposition 87. *Fix a set X and a family of subsets $\mathcal{P} \subseteq \mathcal{P}(X)$. The family $\mathcal{P}' := \mathcal{P} \cup X$ is then a subbase[86] of some topology on X.*

Definition 88. [Eng89, p. 12] Fix a topological space $(X, \mathcal{T})$ and a point $x \in X$. We say that the family $\mathcal{B}(x)$ is a local base for $\mathcal{T}$ at x if for every neighborhood U of x there exists a set $V \in \mathcal{B}(x)$ such that $x \in V \subseteq U$.

The indexed family of local bases $\{\mathcal{B}(x) \colon x \in X\}$ is called a neighborhood system of $\mathcal{T}$.

Proposition 89. *[Eng89, p. 13] Let X be an arbitrary set and let $\{\mathcal{B}(x) \subseteq \mathcal{P}(X) \colon x \in X\}$ be an indexed family of families of subsets of X that satisfies*

BP1 *For every $x \in X$, $\mathcal{B}(x) \neq \varnothing$ and $x \in U$ for every $U \in \mathcal{B}(x)$.*

BP2 *For every $x \in X$ and for all $U, V \in \mathcal{B}(x)$, $\exists W \in \mathcal{B} : W \subseteq U \cap V$.*

BP3 *For all $x, y \in X$, $x \in U \in \mathcal{B}(y)$ implies that there exists $V \in \mathcal{B}(x)$ such that $U \subseteq V$.*

Then the family

$$\mathcal{B} := \bigcup_{x \in X} \mathcal{B}(x)$$

is the base[85] of some topology $\mathcal{T}$ on X. Furthermore, $\{\mathcal{B}(x) \subseteq \mathcal{P}(X) \colon x \in X\}$ is a neighborhood system[88] for $(X, \mathcal{T})$.

Definition 90. [Eng89, p. 33] Let $(X, \mathcal{T})$ be a topological space. Define the closure operator

$$\begin{aligned} \mathrm{cl} &: \mathcal{P}(X) \to \mathcal{P}(X) \\ \mathrm{cl}(A) &:= \bigcap \{F : F \in \mathcal{F}_{\mathcal{T}}, A \subseteq F\}. \end{aligned}$$

Proposition 91. *[Eng89, p. 14] Let X be an arbitrary set and let $\mathrm{cl} : \mathcal{P}(X) \to \mathcal{P}(X)$ be a function that satisfies*

CO1 $\mathrm{cl}(\varnothing) = \varnothing$

CO2 $\forall A \in \mathcal{P}(X), A \subseteq \mathrm{cl}(A)$

CO3 $\forall A, B \in \mathcal{P}(X), \mathrm{cl}(A \cup B) = \mathrm{cl}(A) \cup \mathrm{cl}(B)$

CO4 $\forall A \in \mathcal{P}(X), \mathrm{cl}(\mathrm{cl}(A)) = \mathrm{cl}(A)$

Then the family

$$\mathcal{T} := \{X \backslash F \colon F = \mathrm{cl}(F)\}$$

is a topology on X. Furthermore, $\mathrm{cl} = \mathrm{cl}_{\mathcal{T}}$, where $\mathrm{cl}_{\mathcal{T}}$ is the closure operator[90] on $(X, \mathcal{T})$.

Definition 92. [Eng89, p. 15] Let $(X, \mathcal{T})$ be a topological space. Define the interior operator

$$\begin{aligned} \mathrm{int} &: \mathcal{P}(X) \to \mathcal{P}(X) \\ \mathrm{int}(A) &:= \bigcup \{U : U \in \mathcal{T}, U \subseteq A\}. \end{aligned}$$

Proposition 93. *Let X be an arbitrary set and let $\mathrm{int} : \mathcal{P}(X) \to \mathcal{P}(X)$ be a function that satisfies*

IO1 $\mathrm{int}(X) = X$

IO2 $\forall A \in \mathcal{P}(X), \mathrm{int}(A) \subseteq A$

IO3 $\forall A, B \in \mathcal{P}(X), \mathrm{int}(A \cap B) = \mathrm{int}(A) \cap \mathrm{int}(B)$

IO4 $\forall A \in \mathcal{P}(X), \mathrm{int}(\mathrm{int}(A)) = \mathrm{int}(A)$

Then the family

$$\mathcal{T} := \{U : U = \mathrm{int}(U)\}$$

is a topology on X. Furthermore, $\mathrm{int} = \mathrm{int}_{\mathcal{T}}$, where $\mathrm{int}_{\mathcal{T}}$ is the interior operator[92] on $(X, \mathcal{T})$.

3.2. Continuous functions

Definition 94. Let $(X, \mathcal{T})$ and $(Y, \mathcal{O})$ be topological spaces[83]. We say that the function[167] $f : X \to Y$ is continuous if any of the equivalent conditions hold:

a) For every open set $V \in \mathcal{O}$, the preimage[169] $f^{-1}(V)$ is open.

b) For every closed set $V \in \mathcal{F}_{\mathcal{O}}$, the preimage $f^{-1}(V)$ is closed.

c) There exists a base[84] $\mathcal{B}_{\mathcal{O}} \subseteq \mathcal{O}$, such that for every $V \in \mathcal{B}_{\mathcal{O}}$, the preimage $f^{-1}(V)$ is open.

d) There exists a subbase[86] $\mathcal{P}_{\mathcal{O}} \subseteq \mathcal{O}$, such that for every $V \in \mathcal{P}_{\mathcal{O}}$, the preimage $f^{-1}(V)$ is open.

e) There exist neighborhood systems[88] $\{\mathcal{B}_{\mathcal{T}}(x) \colon x \in X\}$ and $\{\mathcal{B}_{\mathcal{O}}(y) \colon y \in Y\}$, such that for every point $x \in X$ and for any $V \in \mathcal{B}_{\mathcal{O}}(f(x))$, there exists a set $U \in \mathcal{B}_{\mathcal{O}}(x)$ such that $f(U) \subseteq V$.

f) For every set $A \subseteq X$, $f(\text{cl}(A)) \subseteq \text{cl}(f(A))$.

Definition 95. We say that the continuous function $f : (X, \mathcal{T}) \to (Y, \mathcal{O})$ is a open (resp. closed), if the image $f(U)$ of an open set (resp. closed set) in $\mathcal{T}$ is open (resp. closed) in $\mathcal{O}$.

If f is an open bijection, we say that f is a homeomorphism.

3.3. Initial and final topologies

Definition 96. The class[139] of all topological spaces forms the category[175] **Top**, where for every two topological spaces $X, Y \in \mathbf{Top}$, the set of morphism $\mathbf{Top}(X, Y)$ are the homeomorphisms[95] from X to Y and composition is the usual function composition[168].

Furthermore, **Top** is locally small[177] and concrete[213].

Theorem 97. *The category* **Top** *of is both complete* [223] *and cocomplete* [231].

Definition 98. [nLa20e] Let $\{(X_i, \mathcal{T}_i)\}_{i\in I}$ be a family[171] of topological spaces. Let X be a bare set and let

$$\{f_i : X \to X_i\}_{i\in I}$$

be a family of functions.

The topology on X generated by the subbase

$$\mathcal{P} := \{f_i^{-1}(U) \colon i \in I, U \in \mathcal{T}_i\}$$

is called the initial (or weak) topology on X generated by the family $\{f_i\}_{i\in I}$.

It is the weakest topology that makes all functions in the family $\{f_i\}_{i\in I}$ continuous.

Definition 99. [nLa20e] Dually, if the family of functions is of the type

$$\{f_i : X_i \to X\}_{i\in I},$$

then we define the final (or strong) topology on X generated by the family $\{f_i\}_{i\in I}$ as the topology

$$\mathcal{T} := \{U \subseteq X \colon \forall i \in I, f_i^{-1}(U) \in \mathcal{T}_i\}.$$

It is the strongest topology that makes all functions in the family $\{f_i\}_{i\in I}$ continuous.

Proposition 100. *[nLa20e] Let $D : \mathbf{I} \to$* **Top** *be a small diagram*[192]*. For each space in the image $D(\mathbf{I})$, denote the set corresponding by X_i and the corresponding topology by $\mathcal{T}_i$.*

The limit (resp. colimit) $(X, \mathcal{T})$ of D can then be described as

a) $(X, \{f_i\}_{i\in\mathbf{I}}) = \varprojlim UD$ (resp. $\varinjlim UD$) is the limit (resp. colimit) in **Set** *of $U \circ D$, where U :* **Top** *→* **Set** *is the forgetful functor.*

b) $\mathcal{T}$ is the initial[98] *(resp. final*[99]*) topology on X generated by the family of functions $\{f_i\}_{i\in\mathbf{I}}$.*

In particular, the functor U lifts limits and colimits[237 (d)].

Definition 101. Let $(X, \mathcal{T})$ be a topological space and let $M \subseteq X$ be a subset of X. The topological subspace $(M, \mathcal{T}_M)$ is obtained by endowing M with the topology

$$\mathcal{T}_M := \{U \cap M \colon U \in \mathcal{T}\}.$$

It is the initial topology generated by the canonical injection map $\iota : M \to X$.

Definition 102. The topological product or Tychonoff product $(\prod_{i\in I} X_i, \prod_{i\in I} \mathcal{T}_i)$ of the family $(X_i, \mathcal{T}_i)_{i\in I}$ is simply the categorical product in the category **Top**[224]. The underlying set $\prod_{i\in I} X_i$ is the Cartesian product[149159 (a)] and the topology $\prod_{i\in I} \mathcal{T}_i$ is called the product topology.

Let $(X_i, \mathcal{T}_i)_{i\in I}$ and $(Y_i, \mathcal{O}_i)_{i\in I}$ be two families of topological spaces and let $\{f_i : X_i \to Y_i\}_{i\in I}$ be a family of arbitrary functions between them.

We define the product $\prod_{i\in I} f_i$ of $\{f_i\}_{i\in I}$ as the function

$$\left(\prod_{i\in I} f_i\right) : \prod_{i\in I} X_i \to \prod_{i\in I} Y_i$$
$$\left(\prod_{i\in I} f_i\right)(\{x_i\}_{i\in I}) := \{f_i(x_i)\}_{i\in I}.$$

If all of the spaces $(X_i, \mathcal{T}_i)$ are equal to some space $(X, \mathcal{T})$, we call the product of $\{f_i\}_{i\in I}$ the diagonal product and denote it by

$$\Delta_{i\in I} f_i : X \to \prod_{i\in I} Y_i.$$

Definition 103. [Eng89, p. 90] Let $(X, \mathcal{T})$ be a topological space and let $\cong$ be an equivalence relation[161 (c)] on X. The quotient space $(X, \mathcal{T})/\sim$ is obtained by endowing the quotient set $X/\cong$ with the final topology given by the canonical projection map $x \mapsto [x]$.

Definition 104. [Eng89, p. 74] The topological direct sum $(\oplus_{i\in I} X_i, \oplus_{i\in I} \mathcal{T}_i)$ of the family $(X_i, \mathcal{T}_i)_{i\in I}$ is simply the categorical coproduct in the category **Top**[233]. The underlying set $\oplus_{i\in I} X_i$ is the disjoint union[150159 (b)] and the topology $\oplus_{i\in I} \mathcal{T}_i$ is called the direct sum topology.

Let $(X_i, \mathcal{T}_i)_{i\in I}$ and $(Y_i, \mathcal{O}_i)_{i\in I}$ be two families of topological spaces and let $\{f_i : X_i \to Y_i\}_{i\in I}$ be a family of arbitrary functions between them. Let $\iota_{X_i} : X_i \to \oplus_{i\in I} X_i$ and $\iota_{Y_i} : Y_i \to \oplus_{i\in I} Y_i$ be the corresponding canonical injections.

We define the direct sum $\oplus_{i\in I} f_i$ of $\{f_i\}_{i\in I}$ as the function

$$(\oplus_{i\in I} f_i) : \oplus_{i\in I} X_i \to \oplus_{i\in I} Y_i$$
$$(\oplus_{i\in I} f_i)\restriction_{X_i} := \iota_{Y_i} \circ f_i.$$

Obviously $\oplus_{i\in I} f_i$ is continuous whenever all f_i are continuous.

If all of the spaces $(Y_i, \mathcal{O}_i)$ are equal to some space $(Y, \mathcal{O})$, we call the direct sum of $\{f_i\}_{i\in I}$ simply a sum and denote it by

$$\sum_{i\in I} f_i : \oplus_{i\in I} X_i \to Y.$$

3.4. Separation axioms

Definition 105. We can classify topological spaces using the following separation axioms. We say that $(X, \mathcal{T})$ is

Regular every point $x \in X$ and every closed set $F \in \mathcal{F}_{\mathcal{T}}$ can be separated using neighborhoods, i.e. there exist disjoint open sets $U \ni x$ and $V \supseteq F$.

Completely regular (Tychonoff) every point $x \in X$ and every closed set $F \in \mathcal{F}_{\mathcal{T}}$ can be functionally separated, i.e. there exists a continuous function $f : X \to [0, 1]$ such that $f(x) = 0$ and $f(F) = 1$.

Normal (Urysohn) every two closed sets $F, G \in \mathcal{F}_{\mathcal{T}}$ can be separated using neighborhoods, i.e. there exist disjoint open sets $U \supseteq F$ and $V \supseteq G$.

T0 (Kolmogorov) for every two different points $x, y \in X$, there exists an open set $U \in \mathcal{T}$ such that either $x \in U$ or $y \in U$.

T0.5 every singleton set $\{x\}$ is either open or closed.

T1 (Frechet) every singleton set $\{x\}$ is closed.

T2 (Hausdorff) every two different points $x, y \in X$ can be separated using neighborhoods, i.e. there exist disjoint open sets $U \ni x$ and $V \ni y$.

T3 the space is T0 and Regular

T3.5 the space is T0 and Completely regular

T4 the space is T1 and Normal

3.5. Compact sets

Let $(X, \mathcal{T})$ be a topological space.

Definition 106. [Dei85, p. 40] The set $A \subseteq X$ is called compact if any of the following equivalent conditions hold:

a) (Finite union property) Every open cover of X has a finite subcover.

b) (Finite intersection property) The intersection is nonempty for every family of sets such that the intersection of any finite subfamily is nonempty.

If the closure if A is compact, we call A relatively compact or precompact (although the term "precompact" is also used for totally bounded sets, see 111).

3.6. Baire spaces

Definition 107. A topological space is called a Baire space if every nonempty set is not meager.

Proposition 108. *Every open subspace of a Baire space is a Baire space.*

Proof. Let $(X, \mathcal{T})$ be a Baire space.

Let $A \subseteq X$ be a nonempty open set and $\mathcal{T}_A$ be the induced topology.

Let $B \in \mathcal{T}_A$ be nonempty. Then there exists a set $B' \in \mathcal{T}$ such that $B = A \cap B'$. The sets A and B' are open and so $B = B' \cap A$ is open.

Then A, B and B' are nonempty open sets, hence they are not meager in $\mathcal{T}$.

Suppose that B in meager in $\mathcal{T}_A$. Then there exists a sequence $B_1, B_2, \ldots$ of nowhere dense in $\mathcal{T}_A$ sets such that

$$B = \cup_{k=1}^{\infty} B_k.$$

We will now show that all B_k are nowhere dense in $\mathcal{T}$.

Fix any B_k and $U \in \mathcal{T}$. Since B_k is nowhere dense and $A \cap U$ is open, there exists a $V \in \mathcal{T}_A$ such that $V \subseteq A \cap U$ and $V \cap B_k = \varnothing$.

Let $V' \in \mathcal{T}$ be such that $V = V' \cap A$. Since V' and A are open in $\mathcal{T}$ and $V' \cap B_k = \varnothing$, we have that B_k is nowhere dense in $\mathcal{T}$.

This implies that B is meager as the union of countable nowhere dense subsets. But we have already proved that B is nonmeager in $\mathcal{T}$.

The obtained contradiction proves that B is nonmeager in $\mathcal{T}_A$ also. Hence $(A, \mathcal{T}_A)$ is a Baire space. $\square$

4. Metric spaces

4.1. Hausdorff distance

Let (X, ρ) be a metric space.

Definition 109. [DR14, p. 144] Fix two sets $E \subseteq X$ and $F \subseteq X$.
The ordinary distance from a point $x \in X$ to E is defined as

$$d : X \times \mathcal{P}X \to \mathbb{R} \cup \{\infty\}$$
$$d(x, E) := \begin{cases} +\infty, & E = \varnothing, \\ \inf_{y \in F} \rho(x, y), & E \neq \varnothing. \end{cases}$$

The excess of E beyond F is defined as

$$e : \mathcal{P}X \times \mathcal{P}X \to \mathbb{R} \cup \{\infty\} \tag{4.1}$$
$$e(E, F) := \begin{cases} +\infty, & E = \varnothing, D = \varnothing \\ 0, & E = \varnothing, D \neq \varnothing \\ \sup_{x \in E} d(x, F) \overset{(4.1)}{=} \inf\{\delta > 0 \colon E \subseteq F_\delta\}, & E \neq \varnothing \end{cases}$$

where $F_\delta := \{y \in X \colon d(y, F) \leqslant \delta\}$.

The Pompeiu-Hausdorff distance or simply Hausdorff distance between them is then defined as

$$h(E, F) := \max\{e(E, F), e(F, E)\} = \inf\{\delta > 0 \colon E \subseteq F_\delta, F \subseteq E_\delta\}.$$

Proof. (of 4.1) Note that the set

$$F_{e(E,F)} = \{x \in X \colon d(x, F) \leqslant \sup_{x \in E} d(x, F)\}$$

obviously includes E.

Now let $\delta > 0$ be any real number that satisfies $E \subseteq F_\delta$, i.e.

$$E \subseteq F_\delta = \{x \in X \colon d(x, F) \leqslant \delta\},$$

which implies that

$$e(E, F) = \sup_{x \in E} d(x, F) \leqslant \delta.$$

□

Proposition 110. *The Hausdorff distance is a metric on the nonempty compact subsets of X.*

Proof. Let E, F and G be nonempty compact subsets of X.

The function h is nonnegative. Since we exclude empty and unbounded sets, We do not care about infinite values.

Identity Obviously $h(E, E) = 0$. If $h(E, F) = 0$, then there exists no point of E outside of F and vice versa, hence $E = F$.

Symmetry This follows from the symmetry of the max function.

Subadditivity For any point $y \in X$, we have

$$d(x, G) = \inf_{z \in G} \rho(x, z) \leqslant \rho(x, y) + \inf_{y \in G} \rho(y, z) = \rho(x, y) + d(y, G).$$

Select $y \in F$ that minimizes the distance $\rho(x, y)$ over F (compactness allows us), so that

$$d(x, G) \leqslant \rho(x, y) + d(y, G) = d(x, F) + d(y, G) \leqslant d(x, F) + e(F, G).$$

It now follows that

$$\begin{aligned} e(E, G) &= \inf\{\delta > 0 \colon E \subseteq G_\delta\} = \\ &= \inf\{\delta > 0 \colon E \subseteq \{x \in X \colon d(x, G) \leqslant \delta\} \leqslant \\ &\leqslant \inf\{\delta > 0 \colon E \subseteq \{x \in X \colon d(x, F) + e(F, G) \leqslant \delta, y \in X\} = \\ &= e(F, G) + \inf\{\delta > 0 \colon E \subseteq F_\delta\} = \\ &= e(F, G) + e(E, F). \end{aligned}$$

$\square$

4.2. Totally bounded sets

Let (X, ρ) be a metric space. Let $\mathcal{B}$ be the family of bounded sets in X.

Definition 111. The space $A \subseteq X$ is called totally bounded if any of the following equivalent conditions hold:

a) For every $\varepsilon > 0$ there exists a finite cover of A with sets with diameter at most ε.

b) For every $\varepsilon > 0$ there exists a finite cover of A with balls of radius ε.

c) Kuratowski's noncompactness measure[116 (a)] $\alpha(A)$ is zero.

d) The ball noncompactness measure[116 (b)] $\beta(A)$ is zero.

e) Every sequence in A admits a fundamental subsequence.

Totally bounded sets are sometimes called precompact[106] because of theorem 115. This equivalence requires the metric space to be complete, however.

Proof. The equivalences 111 (a) $\iff$ 111 (c) and 111 (b) $\iff$ 111 (d) are straightforward.

(111 (b) $\implies$ 111 (a)) Given $\varepsilon > 0$, any cover of A with balls of radius $\frac{\varepsilon}{2}$ is a cover with sets of diameter ε.

(111 (a) $\implies$ 111 (b)) Fix $\varepsilon > 0$ and $\mu \in (0, \varepsilon)$ and let $A_1, \ldots, A_n \subseteq \mathcal{P}\, X$ be a finite cover of A with sets of diameter at most μ.

Choose[AOC] a point x_k from every A_k, $k = 1, \ldots, n$. We then have that for every $k = 1, \ldots, n$,

$$\begin{aligned} A_k &\subseteq \operatorname{cl} B(x_k, \mu) \subsetneq B(x_k, \varepsilon) \\ \implies A \subseteq \bigcup_{k=1}^{n} A_k &\subseteq \bigcup_{k=1}^{n} B(x_k, \mu) \subsetneq \bigcup_{k=1}^{n} B(x_k, \varepsilon), \end{aligned}$$

hence $x_1, \ldots, x_n$ are centers of ε-balls that cover A.

(111 (b) $\implies$ 111 (e)) Let $\{x_n\} \subseteq A$ be any sequence.

If we assume[LEM] that $\{x_n\}$ has no fundamental subsequence, then there exists $\varepsilon_0 > 0$ such that $\rho(x_k, x_m) > \varepsilon_0$ for any $n, m \in \mathbb{Z}^{>0}$.

Consider a finite cover of A with ε_0-balls. By the pigeonhole principle, at least one of the balls contains more than one element of the sequence, which contradicts the assumption that all elements of the sequence have a distance of at least ε_0.

Hence an arbitrary sequence in A has a fundamental subsequence.

(111 (e) $\implies$ 111 (b)) Assume[LEM] that there exists $\varepsilon_0 > 0$, such that A admits no finite cover by ε_0-balls.

Define $x_1 \in X, x_2 \in X \backslash B(x_1, \varepsilon_0), \ldots$, so that every two elements of the sequence $\{x_n\}$ have a distance of at least ε_0. But then the sequence is does not admit a fundamental subsequence, which contradicts our assumption.

This contradiction proves that A admits a finite cover by ε-balls for every $\varepsilon > 0$. ☐

Corollary 112. *Assume that X is complete. The set $A \subseteq X$ is sequentially compact if and only if it is closed and totally bounded.*

Proof. The property that every sequence has a fundamental subsequence is equivalent to sequential compactness for a closed set in a complete metric space. □

Proposition 113. *If a set $A \subseteq X$ is totally bounded, then so is its closure* $\operatorname{cl} A$.

Proof. Let $\varepsilon > 0$ and $\mu \in (0, \varepsilon)$ and let $x_1, \ldots, x_n \in X$ be the centers of a cover of A with μ-balls.

If y is a point in $\operatorname{cl} A \backslash A$, there exists a point $z \in A$ with $\rho(y, z) < \varepsilon - \mu$. Let $x_k \in A$ be one of the centers whose μ-balls contain z. We then have that $y \in B(x_k, \varepsilon)$ since

$$\rho(x_k, z) \leqslant \rho(x_k, y) + \rho(y, z) < \mu + \varepsilon - \mu = \varepsilon.$$

Hence the balls $\operatorname{cl} B(x_k, \varepsilon)$ cover $\operatorname{cl} A$, i.e.

$$\operatorname{cl} A \subseteq \bigcup_{k=1}^{n} B(x_k, \varepsilon).$$

□

Lemma 114 (Lebesgue's covering lemma)**.** *Assume that X is complete. Let $A \subseteq X$ be sequentially compact. Given an open cover $\mathcal{F} \subseteq \mathcal{P} A$, there exists a number $\delta > 0$ such that every δ-ball with a center in A is contained in some set of the cover $\mathcal{F}$.*

Proof. Assume$^{\text{LEM}}$ that no such number $\delta > 0$ exists. Then for any natural number $n \in \mathbb{Z}^{>0}$, there exists an element $x_n \in A$ such that the ball $B(x_n, \frac{1}{n})$ is not contained in any set of the cover $\mathcal{F}$. Since A is sequentially compact, the sequence $\{x_n\}_n$ contains a convergent subsequence $\{x_{n_k}\}_k$.

Define

$$x := \lim_{k \to \infty} x_{n_k}.$$

Let$^{\text{AOC}}$ E be a set in $\mathcal{F}$ that contains x. Since E is open, there exists some radius $r > 0$ such that $B(x, r) \subseteq E$.

Choose any $k_0 > \frac{2}{r}$ such that $\rho(x_{n_{k_0}}, x) < \frac{r}{2}$. By the triangle inequality,

$$B\left(x_{n_k}, \frac{1}{k}\right) \subsetneq B\left(x_k, \frac{r}{2}\right) \subseteq B(x, r) \subseteq E,$$

which contradicts the choice of the sequence $\{x_n\}_n$.

Hence there exists a $\delta > 0$ such that for every $x \in A$, the ball $B(x, \delta)$ is contained in some element E of the cover $\mathcal{F}$. □

Theorem 115. *Assume that X is complete. The set $A \subseteq X$ is compact if and only if it is sequentially compact*

Proof. ($\Longrightarrow$) Let $\mathcal{F} \subseteq \mathcal{P}X$ be an open cover of A.

By the Lebesgue covering lemma (lemma 114), there exists $\delta > 0$ such that for every $x \in A$, the ball $B(x, \delta)$ is contained in some set of the cover $\mathcal{F}$. Let $x_1, \ldots, x_n$ be a cover of A with δ-balls.

For each $k = 1, \ldots, n$ we have that the ball $B(x_k, \delta)$ is contained in some set $E_k \in \mathcal{F}$. Hence $E_1, \ldots, E_n$ is a finite subcover of A, because

$$A \subseteq \bigcup_{k=1}^{\infty} B(x_k, \delta) \subseteq \bigcup_{k=1}^{\infty} E_k.$$

Thus A is compact.

($\Longleftarrow$) Let A be compact. Fix $\varepsilon > 0$ and take the cover

$$\mathcal{F} := \{B(a, \varepsilon) \colon a \in A\}.$$

By compactness of A, there exists a finite subcover. Thus a finite cover of A with ε-balls exists for every $\varepsilon > 0$. Definition 111 then implies that total boundedness is equivalent to sequential compactness because X is complete and A is closed. $\square$

4.3. Noncompactness measures

Let (X, ρ) be a metric space. Let $\mathcal{B}$ be the family of bounded sets in X.

Definition 116. [Dei85, definition 7.1] We define the following functions

a) The Kuratowski measure of noncompactness,

$$\begin{aligned}&\alpha : \mathcal{B} \to \mathbb{R}^{>0}\\&\alpha(A) := \inf\{d > 0 \colon \exists U_1, \dots, U_n \subseteq X : \operatorname{diam} U_k < d \wedge A \subseteq \cup_{k=1}^{n} U_k\}\end{aligned}$$

b) The ball measure of noncompactness,

$$\begin{aligned}&\beta : \mathcal{B} \to \mathbb{R}^{>0}\\&\beta(A) := \inf\{r > 0 \colon \exists x_1, \dots, x_2 \in X : A \subseteq \cup_{k=1}^{n} B(x_k, r)\}\end{aligned}$$

Example 117. [Dei85, exercise 7.3] Consider the subsets $A_2 \subseteq A_3 \subseteq A_1 \subseteq C([0,1])$, defined by

$$A_1 := \left\{ x \in C([0,1]) \colon \begin{matrix} 0 \leqslant t \leqslant 1 \implies 0 \leqslant x(t) \leqslant 1 \\ x(0) = 0, x(1) = 1 \end{matrix} \right\}$$

$$A_2 := \left\{ x \in A_1 \colon \begin{matrix} 0 \leqslant t \leqslant \frac{1}{2} \implies 0 \leqslant x(t) \leqslant \frac{1}{2} \\ \frac{1}{2} \leqslant t \leqslant 1 \implies \frac{1}{2} \leqslant x(t) \leqslant 1 \end{matrix} \right\}$$

$$A_3 := \left\{ x \in A_1 \colon \begin{matrix} 0 \leqslant t \leqslant \frac{1}{2} \implies 0 \leqslant x(t) \leqslant \frac{2}{3} \\ \frac{1}{2} \leqslant t \leqslant 1 \implies \frac{1}{3} \leqslant x(t) \leqslant 1 \end{matrix} \right\}$$

Then $\alpha(A_1) = 1, \alpha(A_2) = \frac{1}{2}, \alpha(A_3) = \frac{1}{3}$ and $\beta(A_1) = \beta(A_2) = \beta(A_3) = \frac{1}{2}$.

Proof. Since the distance between any two functions from B_1 is at most 1, we have that $\operatorname{diam} B_1 = 1$ and $\alpha(B_1) \leqslant 1$.

Fix $\varepsilon > 0$. For any function $f \in B_1$, continuity of f gives us a radius $\delta_f > 0$ such that

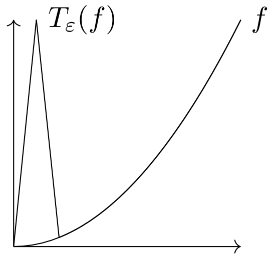
$$x < 2\delta_f \implies f(x) < \varepsilon.$$

Define

$$T_\varepsilon(f)(x) := \begin{cases} \frac{x}{\delta_f}, & 0 \leqslant x < \delta_f \\ f(\delta_f) + [1 - f(\delta_f)](2 - \frac{x}{\delta_f}), & \delta_f \leqslant x < 2\delta_f \\ f(x), & x \geqslant 2\delta_f, \end{cases}$$

so that

$$\|T_\varepsilon(f) - f\| \geqslant T_\varepsilon(f)(\delta_f) - f(\delta_f) = 1 - f(\delta_f) > 1 - \varepsilon.$$



Additionally, because $\delta_{T_\varepsilon(f)} < \delta_f$, we have that $f(\delta_{T_\varepsilon(f)}) < \varepsilon$ and

$$\|T_\varepsilon(T_\varepsilon(f)) - f\| \geqslant T_\varepsilon(T_\varepsilon(f))(\delta_{T_\varepsilon(f)}) - f(\delta_{T_\varepsilon(f)}) = 1 - f(\delta_{T_\varepsilon(f)}) > 1 - \varepsilon.$$

Thus, proceeding by induction, we see that for any $m = 1, 2, \dots$

$$\|T_\varepsilon^m(f) - f\| > 1 - \varepsilon,$$

where T_ε^m denotes repeated application of T_ε.

Consider the sequence

$$\{T_\varepsilon^k(f)\}_{k=0}^\infty = \{f, T_\varepsilon(f), T_\varepsilon(T_\varepsilon(f)), \dots\}.$$

We can easily see that the distance between any two elements of the sequence, say $T_\varepsilon^k(f)$ and $T_\varepsilon^{k+m}(f)$, is strictly greater that $1 - \varepsilon$, i.e.

$$\left\|T_\varepsilon^k(f) - T_\varepsilon^{k+m}(f)\right\| = \left\|T_\varepsilon^k(f) - T_\varepsilon^m(T_\varepsilon^k(f))\right\| > 1 - \varepsilon.$$

Hence B_1 cannot be covered by a finite $(1-\varepsilon)$-net and $\alpha(B_1) \geqslant 1 - \varepsilon$. Since $\varepsilon > 0$ can be made arbitrarily small, this implies that $\alpha(B_1) \geqslant 1$ and, because we already have the reverse inequality, $\alpha(B_1) = 1$.

In the set B_2, the maximum distance between two functions is $\frac{1}{2}$, thus $\operatorname{diam}(B_2) = \frac{1}{2}$ and $\alpha(B_2) \leqslant \frac{1}{2}$. We can then define an operator similar to T_ε that creates "spikes" of height $\frac{1}{2}$ to prove the reverse inequality, obtaining

$$\alpha(B_2) = \frac{1}{2}.$$

Finally, the set B_3 has diameter $\frac{2}{3}$ and hence $\alpha(B_3) = \frac{2}{3}$.

The ball measure for B_1 satisfies the inequalities

$$\frac{1}{2} \leqslant \beta(B_1) \leqslant 1.$$

Additionally, B_1 is strictly contained in the ball centered in the constant function $\frac{1}{2}$ with radius $\frac{1}{2}$, which implies that $\beta(B_1) \leqslant \frac{1}{2}$, hence $\beta(B_1) = \frac{1}{2}$.

For B_2 we have

$$\frac{1}{4} \leqslant \beta(B_2) \leqslant \frac{1}{2}.$$

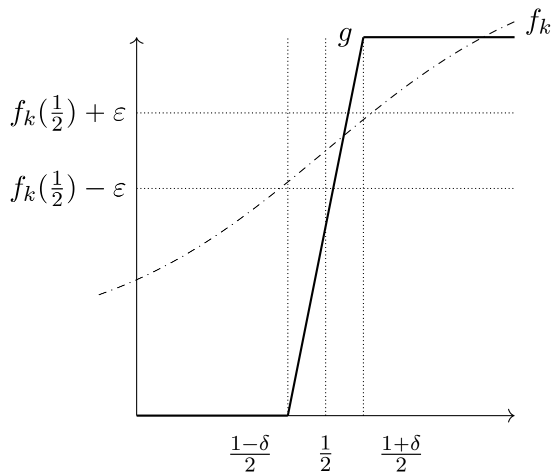
Assume$^{\mathrm{LEM}}$ that for some $\varepsilon > 0$ the set B_2 can be covered by a finite set of balls with centers $\{f_1, \ldots, f_n\} \subsetneq C([0,1])$ and radius $\frac{1}{2} - \varepsilon$.

Because of continuity, we can find a radius $\delta > 0$ such that for all $f_k, k = 1, \ldots, n$ we have

$$x \in \left[\tfrac{1-\delta}{2}, \tfrac{1+\delta}{2}\right] \implies \left|f_k(x) - f_k(\tfrac{1}{2})\right| < \varepsilon.$$

Consider the function

$$g(x) := \begin{cases} 0, & 0 \leqslant x < \frac{1-\delta}{2}, \\ \frac{2x+\delta-1}{2\delta}, & \frac{1-\delta}{2} \leqslant x \leqslant \frac{1+\delta}{2}, \\ 1, & \frac{1+\delta}{2} < x \leqslant 1. \end{cases}$$



If $f_k(\frac{1}{2}) \geqslant \frac{1}{2}$, then $f_k(\frac{1-\delta}{2}) > \frac{1}{2} - \varepsilon$ and

$$\|f_k - g\| \geqslant f_k(\tfrac{1-\delta}{2}) - g(\tfrac{1-\delta}{2}) = f_k(\tfrac{1-\delta}{2}) > \tfrac{1}{2} - \varepsilon.$$

Analogously, if $f_k(\frac{1}{2}) < \frac{1}{2}$, then $f_k(\frac{1+\delta}{2}) < \frac{1}{2} + \varepsilon$ and

$$\|g - f_k\| \geqslant g(\tfrac{1+\delta}{2}) - f_k(\tfrac{1+\delta}{2}) = 1 - f_k(\tfrac{1+\delta}{2}) > \tfrac{1}{2} - \varepsilon.$$

Thus, for every $k = 1, \ldots, n$ we have

$$\|g - f_k\| > \frac{1}{2} - \varepsilon,$$

i.e. g in not contained in a ball of radius $\frac{1}{2} - \varepsilon$ around any of the centers $f_1, \dots, f_n$.

Hence $\beta(B_2) \geqslant \frac{1}{2}$, which implies $\beta(B_2) = \frac{1}{2}$. Because of the inclusion $B_2 \subsetneq B_3 \subsetneq B_1$, we have

$$\frac{1}{2} = \beta(B_2) \leqslant \beta(B_3) \leqslant \beta(B_1) = \frac{1}{2},$$

hence $\beta(B_3) = \frac{1}{2}$. □

Theorem 118. *[Dei85, exercise 7.4], Kuratowski lemma Let X be a Banach space and $\{A_n\}_n$ be a decreasing sequence of nonempty closed subsets such that $\alpha(A_n) \to 0$. Then $A := \bigcap_n A_n$ is nonempty and compact.*

Proof. The set A is compact because it is closed as the intersection of closed sets and $\alpha(A) \leqslant \alpha(A_n) \to 0$, hence $\alpha(A) = 0$.

It remains to show that A is nonempty. Choose$^{\mathrm{AOC}}$ any sequence $\{x_n\}_n$ where $x_n \in A_n$. Since any finite set is compact, we have that for any $k \geqslant 1$

$$\alpha(\{x_n\}_{n\geqslant 1}) = \max\{\alpha(\{x_n\}_{n<k}), \alpha(\{x_n\}_{n\geqslant k})\} = \alpha(\{x_n\}_{n\geqslant k}) \leqslant \alpha(A_k) \to 0,$$

hence the set $\{x_n\colon n \geqslant 1\}$ is compact and thus sequentially compact. We can choose a convergent subsequence $\{x_{n_k}\}_k$ of $\{x_n\}_n$ whose limit lies in every A_n (since they are closed) and, consequently, in their intersection A. So A is nonempty. □

5. Logic

5.1. Languages

Languages are used to define formulas for expressing the axioms of set theory[139]. Here, sets are used to formally define languages. This vicious cycle is left to logicians.

Definition 119. Given a set $\mathcal{A}$, called an alphabet, whose elements are called symbols, we define a word or string over $\mathcal{A}$ to be any tuple[149] of symbols. Words are written simply as strings of symbols, that is, abc instead of (a, b, c). The empty word with no symbols is usually denoted by ε.

The set of all (finite) words over $\mathcal{A}$ is denoted by $\mathcal{A}^*$. A language $\mathcal{L}$ is any subset of $\mathcal{A}^*$.

We define two functions:

$$\begin{aligned} &\operatorname{len} : \mathcal{A}^* \to \mathbb{Z}^{\geqslant 0} && \cdot : \mathcal{A}^* \to \mathbb{Z}^{\geqslant 0} \\ &\operatorname{len}(w) := \text{length of the tuple } w && v \cdot w := (v_1, \dots, v_{\operatorname{len}(v)}, w_1, \dots, w_{\operatorname{len}(w)}). \end{aligned}$$

The function $v \cdot w$ is called concatenation and is usually denoted by juxtaposition. It is obviously associative.

We say that p is a prefix of w if the first $\operatorname{len}(p)$ symbols of w are identical to those of p, that is,

$$w = (p_1, \dots, p_{\operatorname{len}(p)}, w_{\operatorname{len}(p)+1}, \dots, w_{\operatorname{len}(w)}).$$

Suffixes are defined analogously. We say that v is a subword of w if there exists a prefix p and a suffix s such that $w = pvs$. We define the partial order[161 (d)] $v \leqslant w \iff v$ is a subword of w.

Evidently both prefixes and suffixes are subwords and $v \leqslant w \iff \operatorname{len}(v) \leqslant \operatorname{len}(w)$.

For convenience, we denote runs of length n of some letter a as a^n, that is,

$$a^n := \begin{cases} \varepsilon, & n = 0, \\ aa^{n-1}, & n > 1. \end{cases}$$

Thus we do not distinguish between the words $aaabbaa$ and $a^3b^2a^2$.

Proposition 120. *For any alphabet $\mathcal{A}$, the language $(\mathcal{A}^*, \cdot)$ is a monoid.*

5.2. Propositional logic

Propositional logic is a simple framework for describing relationships between statements. It is sometimes called boolean logic because of theorem 127.

Definition 121. [Ner12, p. 12] The language[119] of propositional logic consists of propositional formulas, which are certain well-formed words over the alphabet consisting of

a) A nonempty set **Prop** of propositional variables.

b) The negation symbol $\neg$.

c) The propositional connectives conjunction $\wedge$ and disjunction $\vee$ (we may additionally define other connectives like $\implies$ and $\iff$, however this will only clutter our formal language and make proofs more difficult).

d) Parentheses (and) for defining the order of operations unambiguously.

The propositional formulas $\mathcal{F}_B$ are defined inductively as

- the variables in **Prop** are formulas.
- if φ is a formula, then $\neg\varphi$ is a formula.
- if φ and ψ are formulas, so are $(\varphi \wedge \psi)$ and $(\varphi \vee \psi)$.

Furthermore, we are able to determine every formula's constituent parts uniquely.
If φ and ψ are formulas and ψ is a subword of φ, we say that ψ is a subformula of φ.

Definition 122. We define literals to either be propositional variables $L = P$ or negations $L = \neg P$ of propositional variables.
We define disjuncts (resp. conjuncts) to be finite disjunctions (resp. conjunctions) of literals, i.e.

$$(L_1 \vee (L_2 \vee (\ldots \vee L_n)\ldots)).$$

We say that a propositional formula φ is in conjunctive (resp. disjunctive) normal form (CNF) if φ is a finite conjunction of disjunctions (resp. finite disjunction of conjunctions).

Proposition 123. *Every propositional formula φ is Boolean equivalent[125] to a formula in conjunctive normal form[122].*

Proof. We define the negation function

$$n(\varphi) := \begin{cases} \neg P, & \varphi = P \in \mathbf{Prop} \\ \psi, & \varphi = \neg\psi \\ n(\psi) \wedge n(\theta), & \varphi = \psi \vee \theta \\ n(\psi) \vee n(\theta), & \varphi = \psi \wedge \theta \end{cases}$$

and the reduction function

$$r(\varphi) := \begin{cases} \varphi, & \varphi \in \mathbf{Prop} \\ n(r(\psi)), & \varphi = \neg\psi \\ r(\psi) \vee r(\theta), & \varphi = \psi \vee \theta \\ r(\psi) \wedge r(\theta), & \varphi = \psi \wedge \theta \\ (r(\psi) \vee r(\theta)) \wedge (r(\psi) \vee r(\kappa)), & \varphi = \psi \vee (\theta \wedge \kappa) \end{cases}$$

Given a formula φ, the function $r(\varphi)$ gives a formula in CNF. □

Definition 124. We define the following auxiliary functions using truth tables

x	$H_\neg$	x	y	$H_\vee$	$H_\wedge$
$\top$	$\bot$	$\top$	$\top$	$\top$	$\top$
$\bot$	$\top$	$\top$	$\bot$	$\bot$	$\top$
		$\bot$	$\top$	$\bot$	$\top$
		$\bot$	$\bot$	$\bot$	$\bot$

Note that, as operations over the set $\{\top, \bot\}$, $H_\vee$ and $H_\wedge$ are both associative and commutative.

Definition 125. A propositional interpretation is a function $I : \mathbf{Prop} \to \{\top, \bot\}$.
We define interpretation for formulas inductively as

$$\varphi[I] := \begin{cases} I(P), & \varphi = P \in \mathbf{Prop} \\ H_\neg(\psi[I]), & \varphi = \neg\psi \\ H_\wedge(\psi[I], \theta[I]), & \varphi = \psi \wedge \theta \\ H_\vee(\psi[I], \theta[I]), & \varphi = \psi \vee \theta. \end{cases}$$

We say that

a) I is a Boolean model of φ and write $I \models_B \varphi$ if $\varphi[I] = \top$.

b) If all interpretations are models for φ, we say that φ is a tautology.

c) If no interpretation is a model for φ, we say that φ is a contradiction.

d) φ and ψ are Boolean equivalent and write $\varphi \equiv_B \psi$ if $\varphi[I] = \psi[I]$ for any interpretation I.

Proposition 126. *The Boolean equivalence[125] $\equiv_B$ is an equivalence relation on the set $\mathcal{F}_B$ of propositional formulas.*

Theorem 127. *The quotient set[161 (c)] of propositional formulas $\mathcal{F}_B/\cong$ forms a boolean algebra[274] with*

- *top being the equivalence class of tautologies*

- *bottom being the equivalence class of contradictions*
- *joins given by disjunctions $\vee$ of any representatives of the equivalence classes*
- *meets given by conjunctions $\wedge$*
- *complements given by negation $\neg$*

Proof. See note 266 about handling infinitary joins and meets. Once we prove Associativity, Commutativity and Absorption, we can define a partial order on $\mathcal{F}_B/\cong$ that allows us to extend $\vee$ and $\wedge$ to handle infinite arguments.

Associativity The functions[124] $H_\vee$ and $H_\wedge$ are associative, hence the lattice operations are associative.

Commutativity The proof is analogous to Associativity.

Absorption Let φ and ψ be propositional formulas and I be a propositional interpretation. Then

$\varphi[I]$	$\psi[I]$	$H_\wedge(\psi[I],\varphi[I])$	$H_\vee(\varphi[I],H_\wedge(\psi[I],\varphi[I]))$
$\top$	$\top$	$\top$	$\top$
$\top$	$\bot$	$\bot$	$\top$
$\bot$	$\top$	$\bot$	$\bot$
$\bot$	$\bot$	$\bot$	$\bot$

hence $\varphi[I] = H_\vee(\varphi[I],H_\wedge(\psi[I],\varphi[I]))$.

The proof of the dual law is analogous.

Distributivity Let φ, ψ and θ be propositional formulas and I be a propositional interpretation. Then

$\varphi[I]$	$\psi[I]$	$\theta[I]$	$H_\wedge(\varphi[I],H_\vee(\psi[I],\theta[I]))$	$H_\vee(H_\wedge(\varphi[I],\psi[I]),H_\wedge(\varphi[I],\theta[I]))$
$\top$	$\top$	$\top$	$\top$	$\top$
$\top$	$\top$	$\bot$	$\top$	$\top$
$\top$	$\bot$	$\top$	$\top$	$\top$
$\top$	$\bot$	$\bot$	$\bot$	$\bot$
$\bot$	$\top$	$\top$	$\bot$	$\bot$
$\bot$	$\top$	$\bot$	$\bot$	$\bot$
$\bot$	$\bot$	$\top$	$\bot$	$\bot$
$\bot$	$\bot$	$\bot$	$\bot$	$\bot$

The join and meet induce the partial order $\varphi \leqslant \psi \iff \varphi \vee \psi \equiv \psi$.

Bottom Fix an interpretation I. A formula ω should belong to the supremum $\sup \mathcal{F}_B$ if and only if $\varphi \vee \omega \equiv \omega$ for any formula $\varphi \in \mathcal{F}_B$. If φ is a tautology, $\varphi[I] = \top$ and thus

$$(\varphi \vee \omega)[I] := H_\vee(\varphi[I],\omega[I]) = \top.$$

It follows that $\omega[I] = \top$. Since the interpretation I was arbitrary, ω is also a tautology. Hence the top element is the equivalence class of all tautologies.

Top The proof is analogous to Bottom.

□

5.3. First order logic

The idea of first-order logic (FOL) is to create a formal language whose semantics (given by structures) support boolean operations and can quantify over all elements of an ambient universe. Unlike in propositional logic[5.2], there are many FOL languages.

Definition 128. [Ner12, definition 2.1] The alphabet for a first-order predicate language[119] L consists of:

Logical symbols

1. A countable alphabet of variables $\mathbf{Var}_L$, usually denoted by $x_0, x_1, \dots$ or x, y, z
2. Boolean operations[121 (c)]
   - $\neg$ (negation)
   - $\wedge$ (conjunction)
   - $\vee$ (disjunction)
3. Quantifiers
   - $\forall$ (universal quantifier)
   - $\exists$ (existential quantifier)
4. Parentheses for highlighting the order of operations
5. Optionally, an equality symbol $=$

Non-logical symbols

1. A set of functional symbols, $\mathbf{Func}_L$, whose elements are usually denoted by $f_0, f_1, \dots$ or f, g, h. Each functional symbol has an associated natural number called its arity, denoted by $\#_L f$. Functional symbols with a zero arity are called constants.
2. A set of predicate symbols, $\mathbf{Pred}_L$, whose elements are usually denoted by $p_0, p_1, \dots$ or by symbols like $\oplus$ or $\circ$. Predicate symbols also have an associated arity. Predicate symbols with zero arity are called propositional variables.

Example 129. [Lei14, remark 2.1.4] Most algebraic structures (with the notable exception of fields) can be defined as first-order languages with equality, no predicates and a set of functional symbols called algebraic operations.

- Group theory[2.1] has
  - one zero-arity operation called its identity element e
  - one unitary operation $(\cdot)^{-1}$ called the inverse element
  - one binary operation $\oplus$ called the group operation
- Linear algebra[2.5] has
  - one zero-arity operation called its zero element 0

  - one binary operation + called vector sum
  - for every scalar λ in the underlying field, a unitary operation $\lambda\cdot$ called scalar multiplication by λ

Definition 130. [Ner12, definition 2.2] Given a FOL language L, the set $\mathcal{T}_L$ of terms is defined by structural induction as

- Each variable is a term
- If $\tau_1, \ldots, \tau_n$ are terms and f is a functional symbol with arity n, then the following word is also a term:

$$f(\tau_1, \ldots, \tau_n)$$

In particular, constants are also terms.
Furthermore, we are able to determine every term's constituent parts uniquely.
For each term τ, we define its variables as

$$\mathbf{Free}(\tau) := \begin{cases} x, & \tau = x \in \mathbf{Var}_L, \\ \mathbf{Free}(\tau_1) \cup \ldots \cup \mathbf{Free}(\tau_n), & \tau = f(\tau_1, \ldots, \tau_n). \end{cases}$$

Definition 131. [Ner12, definition 2.5] Given a FOL language L, we define the set of atomic formulas as

- If p is an n-ary predicate symbol and if $\tau_1, \ldots, \tau_n$ are terms, then $p(\tau_1, \ldots, \tau_n)$ is an atomic formula.
- If L has an equality symbol and if τ_1, τ_2 are terms, then $(\tau_1 = \tau_2)$ is an atomic formula.

The set $\mathcal{F}_L$ of predicate formulas is then defined as

- All atomic formulas are formulas
- If φ is a formula, its negation $\neg\varphi$ is also a formula
- If φ and ψ are formulas, then the following are also formulas:
  - $(\varphi \wedge \psi)$
  - $(\varphi \vee \psi)$
- If φ is a formula and x is a variable, then the following are also formulas:
  - $\forall x\varphi$
  - $\exists x\varphi$

Furthermore, we are able to determine every formula's constituent parts uniquely.
Collections of formulas are often called first-order theories.
For each formula φ, we define its free and bound variables as

$$\mathbf{Free}(\varphi) := \begin{cases} \mathbf{Free}(\tau_1) \cup \ldots \cup \mathbf{Free}(\tau_n), & \varphi = p(\tau_1, \ldots, \tau_n) \\ \mathbf{Free}(\tau_1) \cup \mathbf{Free}(\tau_2), & \varphi = (\tau_1 = \tau_2), \\ \mathbf{Free}(\psi), & \varphi = \neg\psi, \\ \mathbf{Free}(\psi_1) \cup \mathbf{Free}(\psi_2), & \varphi = \psi_1 \circ \psi_2, \circ \in \{\wedge, \vee\}, \\ \mathbf{Free}(\psi)\backslash\{x\}, & \varphi = Qx\psi, Q \in \{\forall, \exists\} \end{cases}$$

and

$$\mathbf{Bound}(\varphi) := \begin{cases} \mathbf{Free}(\tau_1) \cup \ldots \cup \mathbf{Free}(\tau_n), & \varphi = p(\tau_1, \ldots, \tau_n) \\ \mathbf{Free}(\tau_1) \cup \mathbf{Free}(\tau_2), & \varphi = (\tau_1 = \tau_2), \\ \mathbf{Bound}(\psi), & \varphi = \neg\psi, \\ \mathbf{Bound}(\psi_1) \cup \mathbf{Bound}(\psi_2), & \varphi = \psi_1 \circ \psi_2, \circ \in \{\wedge, \vee\}, \\ \mathbf{Bound}(\psi) \cup \{x\}, & \varphi = Qx\psi, Q \in \{\forall, \exists\}. \end{cases}$$

A formula is called closed if it has no bound variables.
If a formula φ has free variables $\mathbf{Free} = \{x_1, \ldots, x_n\}$, a common convention is to write it as

$$\varphi(x_1, \ldots, x_n).$$

This highlights that formulas with free variables can act as predicates, however their semantics are completely determined, unlike the semantics of predicates.

Definition 132. Let φ be a first-order formula with a free variable y and ρ be a term. We define the substitions

$$\tau[y \to \rho] := \begin{cases} \rho, & \tau = y, \\ x, & \tau = x \in \mathbf{Var}_L \backslash\{y\}, \\ f(\tau_1[y \to \rho], \ldots, \tau_n[y \to \rho]), & \tau = f(\tau_1, \ldots, \tau_n). \end{cases}$$

$$\varphi[y \to \rho] := \begin{cases} p(\tau_1[y \to \rho], \ldots, \tau_n[y \to \rho]), & \varphi = p(\tau_1, \ldots, \tau_n) \\ (\tau_1[y \to \rho] = \tau_2[y \to \rho]), & \varphi = (\tau_1 = \tau_2), \\ \neg\psi[y \to \rho], & \varphi = \neg\psi, \\ \psi_1[y \to \rho] \circ \psi_2[x \to \rho], & \varphi = \psi_1 \circ \psi_2, \circ \in \{\wedge, \vee\}, \\ Qx\psi[y \to \rho] = T, & \varphi = Qx\psi, Q \in \{\forall, \exists\}, x \notin \mathbf{Free}(\rho), \\ Qx\psi[y \to \rho[x \to z]] = T, & \varphi = Qx\psi, Q \in \{\forall, \exists\}, x \in \mathbf{Free}(\rho) \end{cases}$$

where in the last step $z \in \mathbf{Var} \setminus \mathbf{Free}(\rho)$.
In order to be able to substitute more that one variable simultaneously, we preliminarily rename colliding free variables and then rename them back after substitution.

Definition 133. [Ner12, definition 4.1] Fix a FOL language L. A structure for L consists of:

1. A nonempty set A.
2. For every n-ary function f, a function[167] $I(f) : A^n \to A$ called the interpretation or evaluation of f.
3. For every n-ary predicate p, a family of tuples[149] $I(p) \subseteq A^n$ called the interpretation of p, i.e. all tuples of values that satisfy the predicate within the structure.

Definition 134. Fix a structure $\mathcal{A} = (A, I)$ for a FOL language L. An evaluation for the variables of L is any function $v : \mathbf{Var}_L \to A$.

For every variable x and every universe element $a \in A$ we also define the modified at x with a evaluation

$$v_a^x(y) := \begin{cases} a, & y = x, \\ v(y), & y \neq x. \end{cases}$$

This allows us to define semantics for all terms:

$$\tau[v] := \begin{cases} v(x), & \tau = x \in \mathbf{Var}_L, \\ I(f)(\tau_1[v], \dots, \tau_n[v]), & \tau = f(\tau_1, \dots, \tau_n). \end{cases}$$

and all formulas:

$$\varphi[v] := \begin{cases} (\tau_1[v], \dots, \tau_n[v]) \in I(p), & \varphi = p(\tau_1, \dots, \tau_n) \\ \tau_1[v] = \tau_2[v], & \varphi = (\tau_1 = \tau_2), \\ \neg\psi[v], & \varphi = \neg\psi, \\ \psi_1[v] \circ \psi_2[v], & \varphi = \psi_1 \circ \psi_2, \circ \in \{\wedge, \vee\}, \\ Qa \in A, \psi[v_a^x] = T, & \varphi = Qx\psi, Q \in \{\forall, \exists\}. \end{cases}$$

If $\varphi[v] = T$, we say that φ is true in $\mathcal{A}$ under the evaluation v and we write $\mathcal{A} \models_v \varphi$. If φ is true in $\mathcal{A}$ under every evaluation, we say that φ is true or valid in $\mathcal{A}$ and we write $\mathcal{A} \models \varphi$.

Given a formula $\varphi(x_1, \dots, x_n)$, a common convention is to write

$$\varphi(a_1, \dots, a_n) := \varphi(x_1, \dots, x_n)[v_{a_1, \dots, a_n}^{x_1, \dots, x_n}].$$

Definition 135. [Ner12, definition 4.4] A model for a first-order theory Γ in the FOL language L is a structure $\mathcal{A}$ such that there exists a single evaluation v so that for every formula $\gamma \in \Gamma$, we have $\mathcal{A} \models_v \gamma$. We write $\mathcal{A} \models_v \Gamma$ or simply $\mathcal{A} \models \Gamma$.

Definition 136. A first-order theory is consistent if, under any evaluation in any structure, every formula is either true or false.

6. Set theory

6.1. Sets

There are certain technical subtleties when defining sets and many mathematicians use a mostly informal approach towards set theory that is based on the axioms of Zermelo and Fraenkel, including (or sometimes excluding) the axiom of choice (see 139 and chapter 10).

Definition 137. [**Enderton1997**] Naive set theory is not based on a strict axiom set but rather on the intuitive notion of a set as an unordered collection without repetition. Set equality $A = B$, set membership $x \in A$ and set inclusion $A \subseteq B$ are assumed to be understood. Sets can be explicitly constructed by specifying their elements, e.g.

$$\{3, 7, 31, 127, 8191\}$$

or by specifying a logical formula $\varphi(x)$ in an implicitly assumed logical language:

$$\{x\colon \varphi(x)\}$$

If $\varphi(x) = x \in A \wedge \psi(x)$, we often write

$$\{x \in A\colon \psi(x)\}.$$

In a suitable context, the definitions can be made precise. For example, in the ring of integers $\mathbb{Z}$ with equality, addition, multiplication and predicates partial ordering $\leqslant$ and divisibility $|$, each set can be thought of a formula in the corresponding first-order language[128]. Given formulas φ_A and φ_B with a free variable x and sets

$$A := \{x\colon \varphi_A(x)\} \qquad B := \{x\colon \varphi_B(x)\}$$

- the membership relation $x \in A$ holds precisely when $\mathbb{Z} \models \varphi_A(x)$.
- the inclusion relation $A \subseteq B$ holds when for any evaluation[134] v in $\mathbb{Z}$ and any integer x, we have $\varphi_A(x) \implies \varphi_B(x)$.
- set equality $A = B$ holds precisely when $A \subseteq B$ and $B \subseteq A$

Naive set theory easily leads to paradoxes[138]) and so some axiomatization (e.g. definition 139) is required.

Example 138. Define

$$R := \{x\colon x \neq x\}.$$

We have both $R \in R$ and $R \notin R$.

Definition 139. [**Enderton1997**] In contrast to naive set theory[137]), **Z**ermelo – **F**raenkel set theory with the axiom of choice (ZF**C**) can be made precise. Consider the first-order language[128] with equality $=$, no functional symbols and a single predicate $\in$.

Given a formula $\varphi(x_1, \ldots, x_n)$, we can construct a (syntactic) object

$$A = \{a_1, \ldots, a_n \colon \varphi(a_1, \ldots, a_n)\}$$

that we call a class. Not all classes can be defined to have meaningful semantics (e.g. the class of all classes easily leads to paradoxes like example 138). We define sets in ZFC as classes with semantics given by a model for the following axioms (exclude A8 to obtain ZF). Classes that do not satisfy these axioms are called proper classes and are often said to be too big to be sets, e.g. the class of all sets or the class of all vector spaces).

A1 (extensionality) Two sets are equal if they have the same elements (given by set membership)

A2 (empty set) The following class is a set

$$\varnothing := \{x \colon x \neq x\}.$$

A3 (pairing) If A and B are sets, then

$$\{A, B\}$$

is also a set. In particular, $\{A\} = \{A, A\}$ is a set.

A4 (union) If A is a set, then $\bigcup A$[142] is also a set.

A5 (power set) If A is a set, $\mathcal{P}(A)$[144] is also a set.

A6 (specification) If A is a set and φ is a formula, then

$$\{x \in A \colon \varphi(x)\}$$

is a set.

A7 (infinity) There exists an inductive set[152].

A8 (choice; see chapter 10) Let $J \neq 0$ and for all $j \in J$, X_j is a nonempty set and $X_i \cap X_j = \varnothing$ when $i \neq j$. Then there exists a set M such that for every $j \in J$, the intersection $M \cap X_j$[141] has exactly one member.

A9 (replacement) Given a set X and a formula $\varphi(x, y)$, if for every set $x \in X$ there exists a unique set y such that $\varphi(x, y)$ holds, then

$$Y := \{y \colon \exists x \in X, \varphi(x, y)\}$$

is a set.

A10 (regularity) For every nonempty set A, there exists a member $a \in A$ such that

$$a \cap A \neq \varnothing.$$

Note 140. In ZFC definition 139, everything is a set. However, it is often the case that we are not interested in how a set's elements are represented and only in how they behave, e.g. when working with natural numbers[153] we are interested in the elements of $\mathbb{N}$ and not in the way every element of $\mathbb{N}$ is encoded as a set.

In order to reduce repetitiveness, sets whose elements we consider to be other sets, are often called families of sets. In particular if all (different) sets are disjoint[141], we say that the family is a disjoint family. We usually assume that the sets are nonempty.

Definition 141. [**Enderton1997**] If A is a set, define their intersection as

$$\bigcap A := \{x \colon \forall a \in A, x \in a\}.$$

We leave $\bigcap \varnothing$ undefined.
By A6, $\bigcap A$ is a set.
For two sets A and B, we define the binary intersection as

$$A \cup B := \bigcap \{A, B\} = \{x \colon x \in A \wedge x \in B\}.$$

The class $\{A, B\}$ is a set by A3 and $A \cup B$ is a set by A6.
If $A \cap B = \varnothing$, we say that A and B are disjoint.

Definition 142. [**Enderton1997**] If A is a set, define its union as

$$\bigcup A := \{x \colon \exists a \in A, x \in A\}.$$

In particular, $\bigcup \varnothing = \varnothing$.
By A4, $\bigcup A$ is a set.
For two sets A and B, we define the binary union as

$$A \cup B := \bigcup \{A, B\} = \{x \colon x \in A \vee x \in B\}.$$

The class $\{A, B\}$ is a set by A3 and $A \cup B$ is a set by A4.

Definition 143. [**Enderton1997**] If A and B are sets, define their difference as

$$A \backslash B := \{a \in A \colon a \notin B\}.$$

By A6, $A \backslash B$ is a set.

Definition 144. [**Enderton1997**] If A is a set, define its power set as

$$\mathcal{P}(A) := \{B \colon B \subseteq A\}.$$

By A5, $\mathcal{P}(A)$ is a set.

Note 145. We give two pairs of definitions for tuples and Cartesian products. The first pair, definitions 146 and 147, is quite restricted and is mostly necessary for defining functions[167] and ensuring that everything along the way is indeed a set. The second pair of definitions, given in definition 149, can then be used freely.

Definition 146. [**Enderton1997**] If A and B are sets, define the (binary) tuple or Kuratowski pair as

$$(A, B) := \{\{A\}, \{A, B\}\}.$$

By A3, (A, B) is a set.

Definition 147. [**Enderton1997**] If A and B are sets, define their binary Cartesian product as

$$A \times B := \{(a, b) \colon a \in A \wedge b \in B\}.$$

Proposition 148. *If A and B are sets, their product $A \times B$ is also a set.*

Proof. Fix $a \in A$ and $b \in B$.

- $\{a\}$ is a set by A6 since $\{a\} \subseteq A$
- $A \cup B$ is a set by definition 142
- $\{a, b\}$ is a set by A6 since $\{a\} \subseteq A \cup B$
- $(a, b) = \{\{a\}, \{a, b\}\}$ is a set by A6 since $(a, b) \subseteq \mathcal{P}(A \cup B)$.

Thus $A \times B$ is a set since $A \times B \subseteq \mathcal{P}(\mathcal{P}(A \cup B))$. □

Definition 149. [**Enderton1997**] Let $\{X_i\}_{i \in I}$ be a nonempty family of nonempty sets[171].

We define their Cartesian product as

$$\prod_{i \in I} X_i := \left\{ f : I \to \bigcup_{j \in I} X_j \colon \forall j \in I, f(j) \in X_j \right\}.$$

Any element of the Cartesian product is called a tuple.

Definition 150. Let $\{X_i\}_{i \in I}$ be a nonempty family of nonempty sets[171].

We define their disjoint union as

$$\coprod_{i \in I} X_i := \{(i, x) \colon i \in I, x \in X_i\}.$$

Definition 151. [**Enderton1997**] For any set X, we define the successor operation

$$S(X) := X \cup \{X\}.$$

Definition 152. [**Enderton1997**] A set A is called inductive if

a) $\varnothing \in A$

b) $a \in A \implies S(a) \in A$

Example 153. The smallest inductive set[152] is

$$\omega := \bigcap\{A\colon A \text{ is an inductive set}\}.$$

Since its elements are $\varnothing, S(\varnothing), S(S(\varnothing)), \ldots$, we can identify ω with the set $\mathbb{N}$ of natural numbers.

Whether 0 is a natural number or not, i.e. whether $\varnothing$ encodes 0 or 1, is a matter of convention and notation like $\mathcal{Z}^{\geqslant 0}$ and $\mathcal{Z}^{>0}$ or $\mathcal{N}^0$ and $\mathcal{N}^+$ is sometimes used.

Definition 154. Let X be a nonempty set. We define the finite intersection operator

$$\begin{aligned} \mathrm{FI} &: \mathcal{P}(X) \mapsto \mathcal{P}(X) \\ \mathrm{FI}(P) &:= \left\{\bigcap P'\colon P' \text{ is a nonempty finite subset of } P\right\} \end{aligned}$$

Proposition 155. *The finite intersection operator* FI[154] *satisfies the following*

a) $P \subseteq \mathrm{FI}(P)$

b) $\mathrm{FI}(\mathrm{FI}(P)) = \mathrm{FI}(P)$

c) $\bigcap \mathrm{FI}(P) = \bigcap P$

Definition 156. [**Enderton1997**] A set A is called transitive if $a \in A$ implies $a \subseteq A$.

Definition 157. [**Enderton1997**] A set A is called an ordinal if it is well-ordered[260 (1)] under set membership.

Definition 158. The class[139] of all sets forms the category[175] **Set**, where for every two sets $X, Y \in \mathbf{Set}$, the morphisms $\mathbf{Set}(X, Y)$ are the functions[167] from X to Y and composition is the usual function composition[168].

Furthermore, **Set** is locally small[177] and concrete[213].

Proof. **Identity** For any set $X \in \mathbf{Set}$, we have the identity function

$$\begin{aligned} \mathrm{id}_X &: X \to X \\ \mathrm{id}_X(x) &:= x. \end{aligned}$$

If $f : X \to Y$ is any function, for all $x \in X$ we have

$$[\mathrm{id}_Y \circ f](x) = \mathrm{id}_Y(f(x)) = f(x), \qquad [f \circ \mathrm{id}_X](x) = f(\mathrm{id}_X(x)) = f(x),$$

thus id_x and id_Y are indeed identity morphisms.

Identity Let $f : A \to B$, $g : B \to C$ and $h : C \to D$ be arbitrary functions. For any $x \in A$, we have

$$[[h \circ g] \circ f](x) = [h \circ g](f(x)) = h(g(f(x))) = h([g \circ f](x)) = [h \circ [g \circ f]](x).$$

Since a function $f : X \to Y$ is formally[167] a subset of the product, $X \times X$[149], which is a set, it is itself a set.

The class of all functions $\mathbf{Set}(X, Y)$ from X to Y is then a subset of $\mathcal{P}(X \times Y)$, which is also a set. Thus **Set** is a locally small category.

It is concrete since it is equipped with the identity functor $\mathrm{id}_{\mathbf{Set}}$. □

Theorem 159. *We are interested in categorical limits*[223] *and colimits*[231] *in* **Set**. *If* $\{X_i\}_{i\in I}$ *is an indexed family*[171] *of sets, then*

a) *their categorical product*[224] *is their Cartesian product*[149] $\prod_{i\in I} X_i$*, the projection morphisms being*

$$\begin{aligned}\pi_j : \prod_{i\in I} X_i &\to X_j \\ \pi_j(\{x_i\}_{i\in I}) &:= x_j.\end{aligned}$$

b) *their categorical coproduct*[233] *is their disjoint union*[150] $\coprod_{i\in I} X_i$*, the injection morphisms being*

$$\begin{aligned}\iota_j : X_j &\to \coprod_{i\in I} X_i \\ \iota_j(x_j) &:= (j, x_j).\end{aligned}$$

c) *An equalizer of two functions* $f, g : X \to Y$ *in* **Set** *is the set*

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

Compare this with pullbacks in **Set**[159 (d)].

d) *The pullback of two functions* $f : X \to Z$ *and* $g : Y \to Z$ *in* **Set** *is the set*

$$\{(x, y) \in X \times Y : f(x) = g(y)\}.$$

Compare this with equalizers in **Set**[159 (c)].

e) *A coequalizer of two functions* $f, g : X \to Y$ *in* **Set** *is the quotient space formed by the reflexive, symmetric and transitive closure of the relation* $x \sim y \iff f(x) = g(x)$.

In particular, if $\sim\subseteq X^2$ *is an equivalence relation*[161 (c)]*) on* X*, then the coequalizer of the two projection maps of the product* X^2 *is the pair* $(X/\ ,\pi)$*, where* π *is the quotient map*

$$\begin{aligned}\pi : X &\to X/ \\ \pi(x) &= [x].\end{aligned}$$

f) *Let* X *and* Y *be two sets, let* Z *be a subset of* X *and let* $i : Z \to X$ *be the inclusion map. For any function* $f : Z \to Y$*, we define a pushout of* i *and* f *in* **Set** *to be the set obtained as the quotient of the coproduct* $X \coprod Y$ *and the relation* $x \sim y \iff i^{-1}(x) = f^{-1}(y)$.

6.2. Relations

Definition 160. Let $\{X_i\}_{i\in I}$ be a family of nonempty sets. Subsets of the form

$$\sim\, \subseteq \times_{i\in I} X_i$$

are called relations. If I is a finite set of cardinality n, the relation is called n-ary (binary for $n = 2$, ternary for $n = 3$). If all X_i are the same set, we say that $\sim$ is relation on X.

It is customary to write binary relations as $a \sim b$ instead of $(a, b) \in \sim$.

Definition 161. We will consider binary relations $\sim\, \subseteq X \times X$ on a nonempty set X.

a) The relation $<$ is called a strict partial order if

   Antireflexivity $\neg(a < a)$

   Transitivity $a < b \wedge b < c \implies a < c$

   The binary relation $>$ is defined as $a > b \iff b < a$. Strict partial orders are rarely used compared to nonstrict partial orders (definition 161 (d)).

   If any two elements are in a relation, we call the strict partial order a strict total order or a strict linear order.

b) The relation $\sim$ is called a preorder if:

   Reflexivity $a \sim a$

   Transitivity $a \sim b \wedge b \sim c \implies a \sim c$

   The pair $(X, \sim)$ is said to be a preordered set.

c) The preorder $\cong$ is called an equivalence relation if is is symmetric, i.e.

   Reflexivity $a \cong a$

   Symmetry $a \cong b \implies b \cong a$

   Transitivity $a \cong b \wedge b \cong c \implies a \cong c$

   We define equivalence classes to be sets of the form

$$[a] := \{b \in X : a \cong b\}.$$

   and the quotient set of X by $\cong$ to be the family

$$X/\cong := \{[a] : a \in X\}.$$

   We call the function

$$\pi : X \to X/\cong$$
$$\pi(a) := [a].$$

   the canonical projection. See proposition 164.

d) Assuming we have defined a notion of equality in our formal language[128], the preorder $\leqslant$ is called a (nonstrict) partial order if it is antisymmetric, i.e.

Reflexivity $a \leqslant a$

Antisymmetry $a \leqslant b \wedge b \leqslant a \implies a = b$

Transitivity $a \leqslant b \wedge b \leqslant c \implies a \leqslant c$

A set with a partial order is called a partially ordered set or poset. See definition 260.

Note 162. Equality is a concept that implies that two objects are completely indistinguishable. An example of a formal definition of equality is definition 139. Sometimes equality is meaningless in a certain context, in which case we speak of a language without an equality symbol[128]. When restricted to a set X, it is an equivalence relation[161 (c)]. Furthermore, it is the intersection of all equivalence relations on X.

Definition 163. Let X be a set. A partition of X is a disjoint family[140] $P \subseteq \mathcal{P}(X)$ of nonempty sets such that $X = \bigcup P$. In other words, each element of X belong to exactly one set in P.

Proposition 164. *Fix a set X. The following three constructions are equivalent:*

a) A partition[163] P of X

b) An equivalence relation[161 (c)] $\cong$ on X

c) A function $f : X \to Y$ (where Y is arbitrary)

Proof. (164 (b) $\implies$ 164 (a)) Let $\cong$ be an equivalence relation on X. The quotient set $X/\cong$ is a partition since

- Every element $a \in X$ belongs to the equivalence class $[a]$.
- Let $[a] \cap [b] \neq \varnothing$ and $c \in [a] \cap [b]$. Assume[LEM] that $a \not\cong b$. Then $c \cong a$ and $c \cong b$, thus $a \cong c \cong b$ and $a \cong b$, which is a contradiction. Thus either $[a] = [b]$ or $[a] \cap [b] = \varnothing$.

(164 (a) $\implies$ 164 (c)) Let P be a partition of X. Denote by P_x the set in P which contains x and define the function

$$f : X \to P$$
$$f(x) = P_x.$$

This function is well defined since since all sets in P are disjoint and thus x belongs to exactly one set in P.

(164 (c) $\implies$ 164 (b)) Let $f : X \to Y$ be a function. Define the relation

$$a \cong b \iff f(a) = f(b).$$

It is an equivalence relation since it is induced by the equivalence relation $=$[162]. □

Definition 165. Let $\sim$ be a binary relation on X. Define

a) the <u>converse of $\sim$</u> as

$$(\sim)^{-1} :=\sim \cup \{(y,x)\colon (x,y) \in \sim\}.$$

b) the <u>reflexive closure of $\sim$</u> as

$$(\sim)^{R} :=\sim \cup \{(x,x)\colon x \in X\}.$$

c) the <u>symmetric closure of $\sim$</u> as

$$(\sim)^{S} :=\sim \cup \sim^{-1}.$$

d) the <u>transitive closure of $\sim$</u> as

$$(\sim)^{T} := \bigcap \{\approx\colon \approx \text{ is a binary transitive relation on } X \text{ such that } \sim \subseteq \approx\}.$$

Proof. Let $(\sim)$ be an arbitrary binary relation on X.

The reflexive closure of $\sim$ is reflexive since it contains the diagonal relation $\{(x,x)\colon x \in X\}$.

165 1(b) (c) The symmetric closure of $\sim$ is symmetric since $x\ (\sim)^S y \iff y\ (\sim)^S x$.

165 (d) The transitive closure of $\sim$ is transitive as an intersection of transitive relations.

□

Proposition 166. *Let $(X, \sim)$ be a preordered set. Use the symmetric closure to define the equivalence relation*

$$\cong := (\sim)^S.$$

The quotient set $X/\sim$ along with the induced relation

$$[x] \leqslant [y] \iff x \sim y$$

is then a partially ordered set[161 (d)].

6.3. Functions

It is not straightforward to formalize the notion of correspondence between two values. We will reserve the term mapping for this informal notion and use function in the sense of definition 167. There are several drawbacks of using set theory for defining functions:

1. Mappings are often more general than what can be formalized, i.e. there exist correspondences between logical formulas[131] and between proper classes[139] that cannot be defined in set theory without reaching contradictions.

2. The ambient space often has an additional structure, e.g. algebraic or topological, that is not carried by functions. This leads to definitions such as homomorphism[46] and isometry. This is a motivating example for the benefits category theory[7], where the notion of morphism[175] is able to capture this additional structure (see definition 158).

3. Several generalizations of set-theoretic functions are often used, e.g. multi-valued[167] or partial functions[167], however most formalisms of set theory often only concern functions.

4. Set-theoretic functions are often used in contexts where they do not refer to the intuitive notion of a mapping, e.g. for Cartesian products or for indexing.

Although definitions in terms of set-valued mappings appear simpler and more general, they are also more cumbersome to work with, so we will start with the standard notion of a function. Assume that we have defined a notion of equality in our formal language.

Definition 167. Let X and Y be (potentially empty) sets.

Let $f \subseteq X \times Y$ be a relation on the nonempty sets X and Y. We call the triple (f, X, Y) a function from X to Y if for every $x \in X$, there exists exactly one $y \in Y$ such that $(x, y) \in f$.

We call X the domain of f and denote it by $\operatorname{dom} f = X$ and we call Y the range or codomain of f and denote it by $\operatorname{range} f = Y$. It is customary to define a function solely in terms of the relation f, however in practice the domain and range are important and the range is impossible to recover given only the relation f.

We usually write $f(x) = y$ instead of $(x, y) \in f$ and

$$f : X \to Y,$$
$$f(x) = \dots,$$

where the ellipsis is called the definition of f and is part of a formula that is true whenever $(x, f(x)) \in f$. $f(x)$ is called the value of f at x or the action of f on x or the image of x under f. For a set $A \subseteq X$, we define

$$f(A) := \cup_{a \in A} \{f(a)\}$$

and call $f(A)$ the image of A under f or the action of f on A. We call $f(X) \subset Y$ simply the image of f and denote it by $\operatorname{im} f$. (even if Y is a proper class, $f(X)$ is a set by definition 139)

In analogy to programming languages, we can call $f : X \to Y$ the type or type signature of f (although formally we use sets rather than types) or the declaration of f (see, for example, [Bri88, section 2.4]). It is often enough to only declare f (i.e. specify its type) without defining it in order to use it in practice.

Functions are often called maps, mapping or operators, however:

- Mapping is often used as a more general informal notion of a correspondence between values
- Operator is usually used to refer to functions, especially linear functions, that act on sets (rather than points) of a certain ambient space

The following deviations from the classical notion of a function are commonly used:

- The function $f : X \to Y$ is called apartial function if there may exist points $x \in X$ without a value $f(x)$. In this context,
  - the domain of f refers only to the points of X that have values.
  - standard functions are called total functions.

  These notions are rarely used outside of logic.
- Functions of the type $f : X \to \mathcal{P}Y$, usually denoted as $f : X \rightrightarrows Y$, are called multi-valued or set-valued mappings between X and Y (these are usually referred to as mappings rather than functions). In this context,
  - the domain of f is defined as

    $$\operatorname{dom} f := \{x \in X : f(x) \neq \varnothing\}$$

    and if $\operatorname{dom} f = X$, we call f a total multi-valued function.
  - the image of a set A under f is defined as

    $$f(A) = \cup_{a \in A} f(a).$$

  - partial functions are called single-valued functions and if a partial function $g : X \to Y$ agrees with a multi-valued function $f : X \rightrightarrows Y$, i.e.

    $$\forall x \in \operatorname{dom} f, g(x) \in f(x),$$

    then g is called a selection of f.
  - if f is total then the selections of f are total functions
  - if the value $f(x)$ of each point $x \in X$ is a singleton set, we call f single-valued and generally do not distinguish between f or its selection.

Definition 168. If $f : A \to B$ and $g : B \to C$ are functions, we define their composition $gf : A \to C$ by

$$(gf)(x) := g(f(x))$$

for all $x \in A$.

Definition 169. (Compare with definition 179) Let $f : X \to Y$ be a function. We define the inverse multi-valued function as

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

The set $f^{-1}(y)$ is called the preimage of y.

We list equivalent conditions for three types of invertibility:

- f is called injective, left-invertible or one-to-one if either
  a) different points in X have different images under f
  b) the preimage of any point in Y is either empty or a singleton
  c) there exists a function $g : Y \to X$ such that $g \circ f = \mathrm{id}_Y$
  d) the inverse is a single-valued partial function

  Special notations include $f : X \rightarrowtail Y$ and $f : X \hookrightarrow Y$.

- f is called surjective, right-invertible or onto if either
  a) each point in Y is the image of at least one point in X
  b) the image of f equals the range of f
  c) there exists a function $g : Y \to X$ such that $f \circ g = \mathrm{id}_X$
  d) the inverse is total

  Special notations include $f : X \twoheadrightarrow Y$.

- f is called bijective or simply invertible if either
  a) it is both injective and surjective
  b) each point in Y is the image of exactly one point in X
  c) the preimage of any point in Y is a singleton
  d) there exists a function $g : Y \to X$ such that both $f \circ g = \mathrm{id}_X$ and $g \circ f = \mathrm{id}_Y$
  e) the inverse is a single-valued total function

  Special notations include $f : X \longleftrightarrow Y$.

If $f : X \rightrightarrows Y$ is a multi-valued mapping, we can define two types of preimages for a set $B \subseteq Y$:

a) The small preimage,

$$f_{-1}(B) := \{x \in X : f(x) \subseteq B\}.$$

b) The large preimage,

$$f^{-1}(B) := \{x \in X : f(x) \cap B \neq \varnothing\}.$$

Obviously $f_{-1}(B) \subseteq f^{-1}(B)$. The two types of preimages coincide for single-valued mappings.

Definition 170. Let $f : X \to Y$ be a function. We define the graph of f to be the set

$$\operatorname{gr} f := \{(x, y) \in X \times Y : f(x) = y\},$$

i.e. the underlying relation itself.

Function graphs allows to study functions geometrically, i.e. as subsets of a space with a certain geometric structure. In low-dimensional spaces, function graphs can be plotted graphically to ease this study.

In the case that Y is an ordered set, usually $\mathbb{R}$, we also define the epigraph of f

$$\operatorname{epi} f := \{(x, y) \in X \times Y : y \geqslant f(x)\},$$

and the hypograph of f

$$\operatorname{hyp} f := \{(x, y) \in X \times Y : y \leqslant f(x)\},$$

Definition 171. When considering finite families of sets, it is enough to consider n-tuples. For example, given sets $X_1, \ldots, X_n$, we can think of the family $\{X_k\}_k$ as the ordered tuple

$$(X_1, \ldots, X_n)$$

where the i-th coordinate of the tuple gives us the i-th set of the family.

This approach has two flaws:

- The family **must** be ordered since the natural numbers are ordered. Families of sets often have no obvious ordering.
- The family **must** be at most countable.

A more natural approach to indexed families is given by functions. We choose an arbitrary set I, called the index set. Every function $f : I \to \mathcal{C}$ from I into the class $\mathcal{C}$ of all sets is then called an indexed family. The function f maps every element i of I into a set $X_i := f(i)$. For convenience, this family is denoted as

$$\{X_i\}_{i \in I}.$$

A more general framework than indexed families that also considers relations between the family's elements is given by diagrams in category theory[192].

Definition 172. A sequence $\{x_i\}_{i=1}^{\infty}$ is an indexed family with domain $I = \mathbb{N}$. Sometimes finite n-tuples are referred to as finite sequences, in which case the usual sequences are referred to as infinite sequences.

Example 173.

a) Every n-tuple $(x_1, \ldots, x_n)$ is an indexed family with domain $I = \{1, \ldots, n\}$.

b) An important corner case is when I is the empty set. Since the only possible indexing function is then the empty function, we simply say that the resulting family is empty.

c) In continuous stochastic processes, it is convenient to consider families of random variables $\{X_t\}_{t \geqslant 0}$ indexed by $I = \mathbb{R}^+$. The indexing parameter is often denoted by $t \geqslant 0$ is often interpreted as time.

d) An $n \times m$ matrix $A = \{a_{i,j}\}$ is a family of scalars indexed by the unordered set $I = \{1, \ldots, n\} \times \{1, \ldots, m\}$.

7. Category theory

7.1. Categories

Note 174. The definitions here are somewhat informal because of set-theoretic difficulties (see chapter 6).

Definition 175. [Lei14, definition 1.1.1] A category $\mathbf{C}$ consists of

- a set-theoretic class[139] of objects, where "A is an object in $\mathbf{C}$" is denoted as $A \in \mathbf{C}$
- for each pair of objects $A, B \in \mathbf{C}$, a class $\mathbf{C}(A, B)$ of morphisms (also called arrows)
- for each triple of objects $A, B, C \in \mathbf{C}$, a function

  $$\circ : \mathbf{C}(B, C) \times \mathbf{C}(A, B) \to \mathbf{C}(A, C)$$

  called the composition $g \circ f$ of $f \in \mathbf{C}(A, B)$ and $g \in \mathbf{C}(B, C)$ (the order reversal notation comes from composition of functions)

such that

Identity for each object $A \in \mathbf{C}$, there exists an identity morphism $\mathrm{id}_A \in \mathbf{C}(A, A)$, such that whenever $B \in \mathbf{C}$ and $f : A \to B$, we have

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

Identity composition is associative, i.e. for each $f \in \mathbf{C}(A, B)$, $g \in \mathbf{C}(B, C)$ and $h \in \mathbf{C}(C, D)$, we have

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

If there are no morphisms in $\mathbf{C}$ besides identity morphisms, we say that $\mathbf{C}$ is a discrete category [Lei14, example 1.1.18(b]).

Given a morphism $f : A \to B$, we say that A is the domain of f and that B is the codomain of f.

Example 176. Examples of categories include

a) The category **Set** of sets with functions (see definition 158).

b) The category **Top** of topological spaces with homomorphisms (see definition 96).

c) The category **Grp** of groups with homeomorphisms (see definition 158).

d) Thin categories (see definition 261).

Definition 177. Let $\mathbf{C}$ be a category. If for each pair $A, B \in \mathbf{C}$ the class $\mathbf{C}(A, B)$ is a set, we say that $\mathbf{C}$ is locally small. If, in addition to this, the class of objects is a set, we say that $\mathbf{C}$ is small.

Definition 178. [Lei14, definition 4.1.25] Let $\mathbf{C}$ be a category and $A, B \in \mathbf{C}$. We say that the morphism $f : A \to B$ is a generalized element of B of shape A. In the category **Set**, the morphism $\in: 1 \to B$ is the standard element of the set B since there is a bijection between maps $1 \to B$ and elements of B.

Definition 179. We introduce invertibility for morphisms in some category $\mathbf{C}$ (compare with function invertibility, definition 169).

- $f : A \to B$ is called left-invertible if there exists a morphism $g : B \to A$ such that $g \circ f = \mathrm{id}_A$. In this case we call g a left inverse of f.
- $f : A \to B$ is called right-invertible if there exists a morphism $g : B \to A$ such that $f \circ g = \mathrm{id}_B$. In this case we call g a right inverse of f.
- $f : A \to B$ is called invertible or an isomorphism if there exists a morphism $g : B \to A$ that is both a left and a right inverse of f. In this case we call g a (two-sided) inverse of f and we say that the objects A and B are isomorphic. We denote isomorphisms by $A \cong B$ or $A \overset{f}{\cong} B$. We denote isomorphisms by $f : A \cong B$. A morphism $f : A \to A$ from an object to itself is called an endomorphism and if an endomorphism is an isomorphism, we call it an automorphism.

Closely related notions are

- $f : C \to B$ is called a monomorphism or monic morphism or left-cancellative morphism if for any $g, h : B \to A$ the equality $f \circ g = f \circ h$ implies $g = h$. We denote monomorphisms by $f : C \hookrightarrow B$.
- $f : A \to B$ is called an epimorphism or epic morphism or right-cancellative morphism if for any $g, h : B \to C$ the equality $g \circ f = h \circ f$ implies $g = h$. We denote epimorphisms by $f : C \twoheadrightarrow B$.

Proposition 180. *[Lei14, exercise 1.1.13] A morphism $f : A \to B$ in any category* $\mathbf{C}$ *can have at most one inverse.*

Proof. If f has no inverse, it has at most one inverse and the theorem follows.

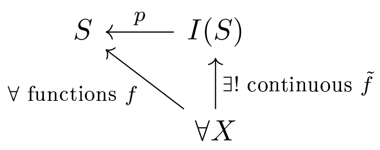
Now assume that f has two inverses g and h, i.e.

$$g \circ f = \mathrm{id}_A \qquad f \circ g = \mathrm{id}_B,$$
$$h \circ f = \mathrm{id}_A \qquad f \circ h = \mathrm{id}_B\,.$$

It follows that $g = h$ since

$$g = g \circ \mathrm{id}_B = g \circ (f \circ h) = (g \circ f) \circ h = \mathrm{id}_A \circ h = h.$$

□



Example 181. [Lei14, exercise 0.10] Let S be a set. The indiscrete topological space $I(S)$ and the canonical projection $p : I(S) \to S$ are characterized by the universal property "for any topological space X and any function $f : X \to S$, there exists a unique continuous function $\tilde{f}$ such that $p \circ \tilde{f} = f$" AOC

Proof. Obviously $I(S)$ and p exist. Assume they are not unique. Let the topological space Y and the function $r : Y \to S$ satisfy the same universal property.

Then by the universal property, there exist unique continuous functions $\tilde{p} : I(S) \to Y$ and $\tilde{r} : Y \to I(S)$ such that

$$r \circ \tilde{p} = p \qquad\qquad p \circ \tilde{r} = r.$$

Hence $p = r \circ \tilde{p} = p \circ \tilde{r} \circ \tilde{p}$ and $\tilde{r} \circ \tilde{p} = \mathrm{id}_{I(S)}$.
Analogously, $r = p \circ \tilde{r} = r \circ \tilde{p} \circ \tilde{r}$, so $\tilde{p} \circ \tilde{r} = \mathrm{id}_Y$.
Thus $\tilde{r}$ and $\tilde{p}$ are mutually inverse and $I(S)$ is isomorphic to Y. □

Definition 182. [Lei14, construction 1.1.9] The opposite or dual category of $\mathbf{C}$ is the category $\mathbf{C}^{\mathrm{op}}$ such that

- The objects in $\mathbf{C}^{\mathrm{op}}$ are the objects in $\mathbf{C}$.
- $f^{\mathrm{op}} \in \mathbf{C}^{\mathrm{op}}(A, B) \iff f \in \mathbf{C}(B, A)$, i.e. the morphisms are reversed.

Example 183. The category $\mathbf{Set}^{\mathrm{op}}$ has a morphism $f : A \to B$ precisely when there exists a function f from the set B to the set A. If $f : A \to B$ is not invertible, then f is not a function.

Definition 184. [Lei14, definition 1.2.18] We call the category $\mathbf{B}$ a subcategory of $\mathbf{A}$ if

- All objects in $\mathbf{B}$ are objects in $\mathbf{A}$.
- All morphisms $f \in \mathbf{B}(A, B)$ are morphisms in $\mathbf{A}(A, B)$.

In case $\mathbf{B}(A, B) = \mathbf{A}(A, B)$ for all objects $A, B \in \mathbf{B}$, we say that $\mathbf{B}$ is a full subcategory.

Definition 185. [Lan94, p. 91] A subcategory $\mathbf{S}$ of $\mathbf{A}$ is called skeletal or a skeleton of $\mathbf{A}$ if it is full and if each object in $\mathbf{A}$ is isomorphic to exactly one object in $\mathbf{B}$.

A category $\mathbf{A}$ is called skeletal if it is its own skeleton, i.e. the only isomorphisms in $\mathbf{A}$ are equalities.

Note 186. A skeletal subcategory $\mathbf{S}$ of $\mathbf{A}$ can be constructed using the axiom of choice by only selecting one object from each isomorphism class within $\mathbf{A}$.

Definition 187. [Lei14, exercise 1.1.14] Let $\mathbf{A}$ and $\mathbf{B}$ be categories. We define their product category $\mathbf{A} \times \mathbf{B}$ component-wise as

- The objects in $\mathbf{A} \times \mathbf{B}$ are pairs (A, B) where $A \in \mathbf{A}$ and $B \in \mathbf{B}$.
- The morphisms in $(\mathbf{A} \times \mathbf{B})[(A, B), (A', B')]$ are pairs (f, g) where $f \in \mathbf{A}(A, A')$ and $g \in \mathbf{B}(B, B')$.

with identities $\mathrm{id}_{(A,B)} = (\mathrm{id}_A, \mathrm{id}_B)$ and composition also defined component-wise.

The definition naturally extends to any finite number of categories.

For a special case, see the notes in definition 206.

Definition 188. [Lei14, definitions 2.1.7] Let $\mathbf{C}$ be a category. The (unique up to an isomorphism, if it exists) object $X \in \mathbf{C}$ is called initial (resp. final or terminal) if for any other object $Y \in \mathbf{C}$ there exists exactly one morphism $f : X \to Y$ (resp. $f : Y \to X$).

If an object is both initial and final, it is called a zero object. A category with a zero object is called a pointed category.

Definition 189. [Lan94][122] Let $\mathbf{C}$ be a category and $X \in \mathbf{C}$ be any object.

Let $u : Y \to X$ and $v : Z \to X$ be monomorphisms[179]. If $u = v \circ u'$ for some monomorphism $u' : Y \to Z$, we say that u factors through v and write $u \leqslant v$. If both $u \leqslant v$ and $v \leqslant u$, we say that u and v are equivalent and write $u \equiv v$.

The equivalence classes among the monomorphisms with a common codomain X are called subobjects of X.

7.2. Functors

Definition 190. [Lei14, definitions 1.2.1, 1.2.10] Let $\mathbf{A}$ and $\mathbf{B}$ be categories. A (covariant) functor $F : \mathbf{A} \to \mathbf{B}$ consists of:

- a function $\mathbf{A} \to \mathbf{B}$, written as $A \mapsto F(A)$.
- for each $A, A' \in \mathbf{A}$, a function
$$\mathbf{A}(A, A') \to \mathbf{B}(F(A), F(A')),$$
written as $f \mapsto F(f)$.

such that

a) $A \overset{f}{\mapsto} B \overset{g}{\mapsto} C$ implies $F(g \circ f) = F(g) \circ F(f)$.

b) $A \in \mathbf{A}$ implies $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$.

If we replace the axiom definition 190 (a) with

b') $A \overset{f}{\mapsto} B \overset{g}{\mapsto} C$ implies $F(g \circ f) = F(f) \circ F(g)$,

we call F a contravariant functor. Equivalently, $F : \mathbf{A} \to \mathbf{B}$ is contravariant if and only if $F : \mathbf{A}^{\mathrm{op}} \to \mathbf{B}$ is covariant.

The identity functor $\mathrm{id}_A : \mathbf{A} \to \mathbf{A}$ simply maps a category to itself.

Note 191. The image $F(\mathbf{A})$ of a category $\mathbf{A}$ under a functor $F : \mathbf{A} \to \mathbf{B}$ may not be a subcategory of $\mathbf{B}$. A simple example is can be constructed as follows:

Let $\mathbf{A}$ be a category with four objects A, B, C, D and two morphisms $f : A \to B$ and $g : C \to D$. If $F(B) = F(C)$, then $F(f) : F(A) \to F(B)$ and $F(g) : F(B) \to F(D)$, however there is no morphism from $F(A)$ to $F(D)$. Thus the image $F(\mathbf{A})$ is not itself category.

Definition 192. A generalization of set-indexed families[171] is given by diagrams. We fix a category $\mathbf{I}$, called an index category, which is often assumed to be small. A diagram of shape $\mathbf{I}$ is then any functor $D : \mathbf{I} \to \mathbf{A}$, where $\mathbf{A}$ is any other category.

It is often convenient to think of diagrams in terms of their images $D(\mathbf{I})$, which are selections of objects and morphisms in $\mathbf{A}$. Note the image $D(\mathbf{I})$ may not be a subcategory of $\mathbf{A}$[191].

If the category $\mathbf{I}$ is small, we say that the diagram is a small diagram.

Example 193.

a) In the case when $\mathbf{I}$ is a small discrete category, a diagram $D : \mathbf{I} \to \mathbf{A}$ is simply a mapping of each element i of $\mathbf{I}$ into an element of $\mathbf{A}$, i.e. we can interpret any diagram of shape $\mathbf{I}$ as a set-indexed family $\{A_i\}_{i \in I}$, where all A_i are objects in $\mathbf{A}$.

b) If $\mathbf{I}$ is not discrete, a diagram $D : \mathbf{I} \to \mathbf{A}$ also involves morphisms. For example, if $\mathbf{I}$ is a three-object category with two morphisms as in the following picture

$$\bullet \longrightarrow \bullet \longrightarrow \bullet,$$

we can interpret a diagram D of shape $\mathbf{I}$ as a selection of objects and morphisms in $\mathbf{A}$ that satisfy the same relations as in $\mathbf{I}$:

$$A \xrightarrow{f} B \xrightarrow{g} C.$$

Definition 194. Let N be a subset of $\mathbb{Z}$ and let $\mathbf{C}$ be any category. A tower diagram in $\mathbf{C}$ is an injective on objects (as a function) diagram $D : N \to \mathbf{C}$ over the poset category[262], i.e.

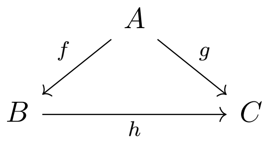
$$\ldots \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \ldots \tag{7.1}$$

Definition 195. A diagram D is said to be commutative if, whenever we have two chains of morphisms $X \overset{f_1}{\to} A_1 \overset{f_2}{\to} \ldots \overset{f_{n-1}}{\to} A_{n-1} \overset{f_n}{\to} Y$ and $X \overset{g_1}{\to} B_1 \overset{g_2}{\to} \ldots \overset{g_{m-1}}{\to} B_{m-1} \overset{f_m}{\to} Y$ in the diagram, where $n > 0$ and $m > 0$, then necessarily

$$f_n \circ \ldots \circ f_1 = g_m \circ \ldots \circ g_1.$$

We also say that the diagram D commutes.

Example 196. Consider the diagram



It is commutative if and only if $h \circ f = g$.

For a more convoluted example, see definition 229.

Definition 197. [Lei14, definition 5.2.1] Given a functor $F : \mathbf{A} \to \mathbf{B}$, we define opposite or dual functor $F^{\mathrm{op}} : \mathbf{A}^{\mathrm{op}} \to \mathbf{B}^{\mathrm{op}}$ as

- $F^{\mathrm{op}}(A) = F(A)$
- $F^{\mathrm{op}}(f^{\mathrm{op}} : A' \to A) = F(f : A \to A')$

Proposition 198. *[Lei14, exercise 1.2.21] Functors preserve isomorphisms, i.e. if $F : \mathbf{A} \to \mathbf{B}$ is a (covariant) functor and $A \cong A'$, then $F(A) \cong F(A')$.*

Proof. Let $f : A \to A'$ be an isomorphism with inverse f^{-1}. From definition 190, we have

$$F(f^{-1}) \circ F(f) \overset{190\_(a)}{=} F(f^{-1} \circ f) = F(\mathrm{id}_A) \overset{190\_(b)}{=} \mathrm{id}_{F(A)}\,.$$

Analogously, $F(f) \circ F(f^{-1}) = \mathrm{id}_{F(A')}$. Thus $F(f) : F(A) \to F(A')$ is an isomorphism with inverse $F(f^{-1})$. □

Note 199. [Lei14, examples 1.2.3, 1.2.4] An informal notion is that of the forgetful functor. A functor $F : \mathbf{A} \to \mathbf{B}$ is called forgetful if the images $F(A)$ of objects $A \in \mathbf{A}$ have "less structure" than A. For example, a functor which maps topological spaces to their underlying sets is forgetful since it "forgets" about the topological structure.

A dual informal notion is that of a free functor. In contrast to forgetful functors which "remove structure", free functors "add structure". For example, a functor which maps a set to its corresponding discrete topological space is a free functor.

Definition 200. [Lei14, definition 1.2.15] A presheaf on the category $\mathbf{A}$ is a contravariant functor

$$F : \mathbf{A} \to \mathbf{Set}\,.$$

Example 201. [Lei14, p. 24] Let (X, τ) be a topological space. Form the category $\mathbf{C}$ from the poset $(\tau, \subseteq)$ as in proposition 263. Presheaves on $\mathbf{C}$ are also called presheaves on the topological space (X, τ).

Let (Y, ρ) be another topological space. Then the map

$$\begin{aligned} &F : \tau \rightrightarrows C(\tau, Y) \\ &F(U) = C(U, Y) = \{f : U \mapsto Y, f \text{ is continuous}\} \end{aligned}$$

is a presheaf.

Definition 202. [Lei14, definition 1.2.16] A functor $F : \mathbf{A} \to \mathbf{B}$ is called faithful (resp. full) if the map

$$\begin{aligned} \mathbf{A}(A, A) &\to \mathbf{B}(F(A), F(A')) \\ f &\mapsto F(f) \end{aligned}$$

is injective (resp. surjective)[169].

Example 203. [Lei14, p. 25] Let $\mathbf{B}$ be a subcategory of $\mathbf{A}$. We define the inclusion functor $I : \mathbf{B} \to \mathbf{A}$ by sending each object and each morphism of $\mathbf{B}$ to itself within $\mathbf{A}$.

Then I is faithful and, if the subcategory $\mathbf{B}$ is full[184], then I is also full.

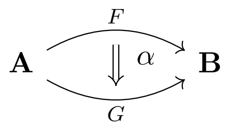
Definition 204. [Lei14, definition 1.3.1] Let $\mathbf{A}$ and $\mathbf{B}$ be categories and let F and G be functors from $\mathbf{A}$ to $\mathbf{B}$.

A natural transformation $\alpha : F \to G$ is a family $\{\alpha_A : F(A) \to G(A)\}_{A \in \mathbf{A}}$ of morphisms in $\mathbf{B}$ such that for every morphism $f : A \to A'$ in $\mathbf{A}$, the diagram

$$\begin{array}{ccc} F(A) & \xrightarrow{F(f)} & F(A') \\ \downarrow{\scriptstyle \alpha_A} & & \downarrow{\scriptstyle \alpha_{A'}} \\ G(A) & \xrightarrow{G(f)} & G(A') \end{array}$$

commutes.

The morphisms α_A are called the components of α. We denote natural transformations using



The natural transformation from F to F composed of identity morphisms is called the identity natural transformation.

Definition 205. Let $F : \mathbf{A} \to \mathbf{B}$, $G : \mathbf{A} \to \mathbf{B}$ and $H : \mathbf{A} \to \mathbf{B}$ be functors and let $\alpha : F \to G$ and $\beta : G \to H$ be natural transformations.

We define the composition (sometimes called vertical composition) of the natural transformations β and α component-wise for $A \in \mathbf{A}$ as

$$(\beta \circ \alpha)_A \coloneqq \beta_A \circ \alpha_A.$$

Definition 206. Given categories $\mathbf{A}$ and $\mathbf{B}$, we define their functor category $[\mathbf{A}, \mathbf{B}]$ by

- the objects in $[\mathbf{A}, \mathbf{B}]$ are functors $F : \mathbf{A} \to \mathbf{B}$.
- the morphisms in $[\mathbf{A}, \mathbf{B}](F, G)$ are the natural transformations from F to G.

The functor category $[\mathbf{A}, \mathbf{B}]$ is often denoted by $\mathbf{B}^{\mathbf{A}}$ since, if $\mathbf{A}$ is a finite discrete category of cardinality n, it is equivalent[207]) to the product category[187] $\mathbf{B} \times \mathbf{B}$

$$\mathbf{B}^{\mathbf{A}} = \mathbf{B}^n = \mathbf{B} \times \ldots \times \mathbf{B}\,.$$

If the natural transformation α is an isomorphism in $[\mathbf{A}, \mathbf{B}]$, we say that the categories $\mathbf{A}$ and $\mathbf{B}$ are naturally isomorphic and write $\mathbf{A} \cong \mathbf{B}$.

Definition 207. [Lei14, definition 1.3.15] An equivalence between the categories $\mathbf{A}$ and $\mathbf{B}$ consists of a pair of functors $F, G : \mathbf{A} \to \mathbf{B}$ and a pair of natural isomorphisms

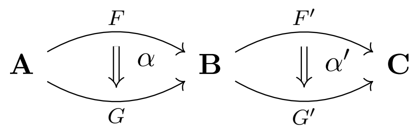
$$\xi : \mathrm{id}_{\mathbf{A}} \to G \circ F, \qquad \eta : F \circ G \to \mathrm{id}_{\mathbf{B}}\,.$$

If an equivalence between $\mathbf{A}$ and $\mathbf{B}$ exists, we say that the categories $\mathbf{A}$ and $\mathbf{B}$ are equivalent and write $\mathbf{A} \simeq \mathbf{B}$.

An equivalence of the form $\mathbf{A}^{\mathrm{op}} \simeq \mathbf{B}$ is called a duality between $\mathbf{A}$ and $\mathbf{B}$ and we say that $\mathbf{A}$ is dual to $\mathbf{B}$ [Lei14, example 1.3.22].

Proposition 208. *[Lan94, p. 91] Every category $\mathbf{A}$ is equivalent to a skeletal subcategory (if one exists; see note 186).*

Definition 209. [Lei14, remarks 1.3.24] Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be categories, $F, G : \mathbf{A} \to \mathbf{B}$ and $F', G' : \mathbf{B} \to \mathbf{C}$ be functors and $\alpha : F \to G$ and $\alpha' : F' \to G'$ be natural transformations.



We define the natural transformation

$$\alpha' * \alpha : F' \circ F \to G' \circ G,$$

called horizontal composition of α and α', defined by

$$(\alpha' * \alpha)_A := \alpha'_{G(A)} \circ F'(\alpha_A) = G'(\alpha_A) \circ \alpha'_{F(A)}.$$

Note 210. We restrict our attention to locally small categories because we need to define an isomorphism of morphism sets.

Definition 211. [Lei14, definition 2.1.1] Let $\mathbf{A}$ and $\mathbf{B}$ be locally small categories and $F : \mathbf{A} \to \mathbf{B}$ and $G : \mathbf{B} \to \mathbf{A}$ be functors. Further assume that for every $A \in \mathbf{A}$ and $B \in \mathbf{B}$ we have an isomorphism

$$\mathbf{A}(A, G(B)) \overset{\varphi_{A,B}}{\cong} \mathbf{B}(F(A), B),$$

where $\mathbf{A}(A, G(B))$ and $\mathbf{B}(F(A), B)$ are regarded as objects in $\mathbf{Set}$.

Given a morphism $f : A \to G(B)$, we define the transpose $\overline{f}$ of f as

$$\begin{aligned} &\overline{f} : F(A) \to B \\ &\overline{f} := \varphi_{A,B}(f). \end{aligned}$$

Dually, given a morphism $g : F(A) \to B$, we define

$$\begin{aligned} &\overline{g} : A \to G(B) \\ &\overline{g} := \varphi^{-1}_{A,B}(g). \end{aligned}$$

We say that the isomorphism $\varphi_{A,B}$ is natural if, given $A' \in \mathbf{A}$ and morphisms $f : A \to G(B)$ and $p : A' \to A$, we have

$$\overline{f \circ p} = \overline{f} \circ F(p),$$

and, given $B' \in \mathbf{B}$ and morphisms $g : F(A) \to B$ and $q : B \to B'$, we have

$$\overline{q \circ g} = G(q) \circ \overline{g}.$$

In this case, we say that F is left-adjoint to G and G is right-adjoint to F, and write $F \dashv G$.

Example 212. [Lei14, example 2.1.5] Consider the functors

- $U : \mathbf{Top} \to \mathbf{Set}$, which maps topological spaces to their underlying sets.
- $D : \mathbf{Set} \to \mathbf{Top}$, which maps sets to topological spaces equipped with the discrete topology.
- $I : \mathbf{Set} \to \mathbf{Top}$, which maps sets to topological spaces equipped with the indiscrete topology.

Let $T \in \mathbf{Top}$ and $S \in \mathbf{Set}$.

Let $f : T \to I(S)$ be any continuous function and $g : U(T) \to S$ be any function.

Denote by $\overline{f} : U(T) \to S$ the function between sets, corresponding to f and by $\overline{g} : T \to I(S)$ the corresponding function between the topological spaces T and $I(S)$. Since any function into an indiscrete topological space is T is continuous, we have that $\overline{g}$ is a morphism $T \to I(S)$.

Thus $\overline{\overline{f}} = f$ and $\overline{\overline{g}} = g$ and we have a natural isomorphism between $\mathbf{Set}(U(T), S)$ and $\mathbf{Top}(T, I(S))$. This proves that $U \dashv I$.

Similarly, since any function from a discrete space is continuous, we have that $D \dashv U$.

Hence $D \dashv U \dashv I$.

Definition 213. A category $\mathbf{C}$ is called concrete is it is equipped with a forgetful[199] faithful[202] functor $F : \mathbf{C} \to \mathbf{Set}$.

A concrete category is said to consist of "sets with extra structure".

Definition 214. [Lei14, definition 2.3.1] Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be categories and $\mathbf{A} \xrightarrow{F} \mathbf{C} \xleftarrow{G} \mathbf{B}$. We define the comma category $(F \downarrow G)$ by

- The objects in $(F \downarrow G)$ are triples (A, h, B) where $A \in \mathbf{A}$, $B \in \mathbf{B}$ and $F(A) \xrightarrow{h} G(B)$.
- The morphisms from (A, h, B) to (A', h', B') are pairs $(f, g) \in \mathbf{A}(A, A') \times \mathbf{B}(B, B')$ such that the following diagram commutes:

$$\begin{array}{ccc} F(A) & \xrightarrow{F(f)} & F(A') \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle h'} \\ G(B) & \xrightarrow[G(g)]{} & G(B') \end{array} \tag{7.2}$$

As a special case, if $\mathbf{A}$ is the one-object category, then F necessarily "selects" an object $C \in \mathbf{C}$. Thus, we can define the comma category $(C \downarrow G)$, in which objects may be regarded as pairs (h, B) rather than triples and the diagram for morphisms looks like

$$\begin{array}{ccccc} & & C & & \\ & {\scriptstyle h}\swarrow & & \searrow{\scriptstyle h'} & \\ G(B) & & \xrightarrow[G(g)]{} & & G(B') \end{array}$$

Analogously, we can also define the category $(F \downarrow C)$ by regarding G and not F as a functor from the one-object category.

Definition 215. [Lei14, definitions 4.1.3, 4.1.16] Let $\mathbf{A}$ be a locally small category and $A \in \mathbf{A}$. Define

$$H^A : \mathbf{A} \to \mathbf{Set},$$
$$H^A(B) := \mathbf{A}(A, B),$$
$$H^A(f : B \to C) := (p : A \to B) \mapsto (f \circ p : A \to C).$$

We say that the functor $F : \mathbf{A} \to \mathbf{Set}$ is <u>representable</u> with <u>representation</u> H^A if $F \cong H^A$.

Analogously, the presheaf $G : \mathbf{A}^{\mathrm{op}} \to \mathbf{Set}$ is representable if for some $A \in \mathbf{A}^{\mathrm{op}}$ we have $G \cong H_A$, where

$$\begin{aligned} &H_A : \mathbf{A}^{\mathrm{op}} \to \mathbf{Set}, \\ &H_A(B) := \mathbf{A}(B, A), \\ &H_A(f : C \to B) := (p : B \to A) \mapsto (p \circ f : C \to A). \end{aligned}$$

Example 216. [Lei14, example 4.1.4] Let $U : \mathbf{Top} \to \mathbf{Set}$ be the forgetful functor which maps a topological space to its underlying set.

Let 1 be the one-element topological space. There is a correspondence between points x in T and continuous functions $p_x : 1 \to T$. Thus the functor H^1 maps

- any topological space T into the set of morphisms

$$H^1(T) = \mathbf{Top}(1, T) = \{p_x : 1 \to T\} \cong U(T).$$

- any continuous function $f : T \to S$ to

$$H^1(f) = p_x \mapsto f \circ p_x \cong x \mapsto f(x) = f.$$

Thus U is representable with representation H^1.

Definition 217. [Lei14, definitions 4.1.15, 4.1.21] Let $\mathbf{A}$ be a locally small category. For each pair $A, B \in \mathbf{A}$ and morphism $f : A \to B$ we define the natural transformation $H^f : H^A \to H^B$ with C-components (note the reversal)

$$\begin{aligned} &H^f(C) : H^A(C) \to H^B(C), \\ &H^f(C) := H_C(f) = p \mapsto p \circ f. \end{aligned}$$

Thus allows us to define the functor $H^\bullet : \mathbf{A}^{\mathrm{op}} \to [\mathbf{A}, \mathbf{Set}]$ by

$$H^\bullet(A) := H^A \qquad\qquad H^\bullet(f) := H^f.$$

Analogously, we define H_f by $H_f(C) = H^C(f), C \in \mathbf{A}$ and the <u>Yoneda embedding</u> $H_\bullet : \mathbf{A} \to [\mathbf{A}^{\mathrm{op}}, \mathbf{Set}]$ by

$$H_\bullet(A) := H_A \qquad\qquad H_\bullet(f) := H_f.$$

Proposition 218. *[Lei14, exercise 4.1.27] Let $\mathbf{A}$ be a locally small category and let $A, A' \in \mathbf{A}$ be such that $H_A \cong H_{A'}$. Then $A \cong A'$.*

Proof. First, let A and A' be arbitrary. Given a natural isomorphism $\eta : H_A \to H_{A'}$, its components are $\alpha_B : H_A(B) \to H_{A'}(B)$.

We are interested in the morphisms

$$\begin{aligned} f &:= \alpha_A(\mathrm{id}_A) : A \to A', \\ g &:= \alpha_{A'}^{-1}(\mathrm{id}'_A) : A' \to A. \end{aligned}$$

We need to show that g is inverse to f. We will use the commutativity of the following diagram:

$$\begin{array}{ccc} H_A(A) & \xleftarrow{H_A(f)} & H_A(A') \\ {\scriptstyle \alpha_A}\downarrow & & \downarrow{\scriptstyle \alpha_{A'}} \\ H_{A'}(A) & \xleftarrow[H_{A'}(f)]{} & H_{A'}(A'), \end{array}$$

where

$$\begin{aligned} H_A(f : A \to A') &= (p : A' \to A) \mapsto (p \circ f : A \to A), \\ H_{A'}(f : A \to A') &= (p : A' \to A') \mapsto (p \circ f : A \to A'). \end{aligned}$$

In particular,

$$\begin{array}{ccc} g \circ f & \xleftarrow{\qquad\qquad H_A(f) \qquad\qquad} & g \\ {\scriptstyle \alpha_A}\downarrow & & \downarrow{\scriptstyle \alpha_{A'}} \\ \alpha_A(g \circ f) = H_{A'}(f)(\mathrm{id}_{A'}) = f & \xleftarrow[H_{A'}(f)]{} & \alpha_{A'}(\alpha_{A'}^{-1}(\mathrm{id}_{A'})) = \mathrm{id}_{A'}, \end{array}$$

i.e.

$$\alpha_A(g \circ f) = f = \alpha_A(\mathrm{id}_A).$$

Since α_A is a bijection, we conclude that $g \circ f = \mathrm{id}_A$.

Analogously, we obtain that $f \circ g = \mathrm{id}_{A'}$. Thus $f : A \to A'$ is an isomorphism, the inverse being $g : A' \to A$. $\square$

Theorem 219. *(Yoneda's lemma)[Lei14, theorem 4.2.1] Let* $\mathbf{A}$ *be a locally small category. Then there is a natural isomorphism between the functors*

$$\begin{aligned} &\mathbf{A}^{\mathrm{op}} \times [\mathbf{A}^{\mathrm{op}}, \mathbf{Set}] \to \mathbf{Set} \\ &(A, X) \mapsto X(A) \end{aligned}$$

and

$$\begin{aligned} &\mathbf{A}^{\mathrm{op}} \times [\mathbf{A}^{\mathrm{op}}, \mathbf{Set}] \to \mathbf{Set} \\ &(A, X) \mapsto [\mathbf{A}^{\mathrm{op}}, \mathbf{Set}](H_A, X). \end{aligned}$$

7.3. Limits

Note 220. Examples of limits and colimits can be found in theorem 159, proposition 56 and section 3.3.

Definition 221. [Lei14, p. 143] Let $\mathbf{I}$ be a small index category and let $\mathbf{A}$ be any category. For each object $A \in \mathbf{A}$, we define the functor $\Delta A : \mathbf{I} \to \mathbf{A}$ as

- For every object $i \in \mathbf{I}$, define $\Delta A(i) = A$
- For every morphism $u : i \to j$, define $\Delta A(u) = \mathrm{id}_A$

We combine these functors for every object $A \in \mathbf{A}$ to obtain the functor $\Delta : \mathbf{A} \to [\mathbf{I}, \mathbf{A}]$.

Definition 222. [Lei14, definition 5.1.19(a)] Let $\mathbf{A}$ be a category and $\mathbf{I}$ be an index category[192]. Let $D : \mathbf{I} \to \mathbf{A}$ be a diagram. A cone on D can be defined equivalently as:

a) a family of projection morphisms $\{f_i : A \to D(i)\}_{i \in \mathbf{I}}$ from the vertex A such that for all morphisms $u : i \to j$ in $\mathbf{I}$, the following diagram commutes:

$$\begin{array}{ccc} & A & \\ {}^{f_i}\swarrow & & \searrow^{f_j} \\ D(i) & \xrightarrow[D(u)]{} & D(j) \end{array} \tag{7.3}$$

b) a natural transformation in $[\mathbf{I}, \mathbf{A}](\Delta A, D)$.

c) an object of the comma category $(\Delta \downarrow D)$ (see the equivalence proof for details).

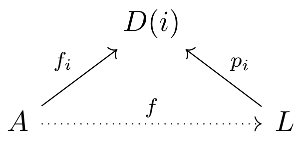
Proof. (222 (a) $\iff$ 222 (b)) Let $i, j \in \mathbf{I}$ and $u : i \to j$. Then a natural transformation f in satisfies the following commutative diagram:

$$\begin{array}{ccc} \Delta A(i) & \xrightarrow{\Delta A(u)} & \Delta A(j) \\ \downarrow{\scriptstyle f_i} & & \downarrow{\scriptstyle f_j} \\ D(i) & \xrightarrow{D(u)} & D(j) \end{array}$$

Since $\Delta A(i) = \Delta A(j) = A$, the above diagram is the same as eq. (7.3).

(222 (b) $\iff$ 222 (c)) We can regard $D : \mathbf{I} \to \mathbf{A}$ as an object in the functor category $[\mathbf{I}, \mathbf{A}]$. Since $\Delta : \mathbf{A} \to [\mathbf{I}, \mathbf{A}]$, an object (A, h) in $(\Delta \downarrow D)$ consists of an object A of $\mathbf{A}$ and a natural transformation from ΔA to D. The converse also applies. ☐

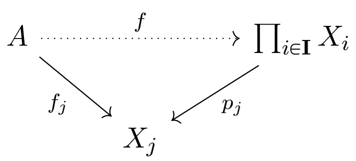
Definition 223. [Lei14, definitions 5.1.19(b), definition 6.3.6] Let $\mathbf{A}$ be a category and $\mathbf{I}$ be an index category. The (unique up to an isomorphism, if it exists) limit or limit cone $\varprojlim D$ of D is a cone $\{L \xrightarrow{p_i} D(i)\}_{i \in \mathbf{I}}$ such that for every cone $\{A \xrightarrow{f_i} D(i)\}_{i \in \mathbf{I}}$ there exists exactly one morphism $f : A \to L$ such that $f \circ p_i = f_i, i \in \mathbf{I}$, i.e. the following diagram commutes:



If the diagram $\mathbf{I}$ is small, its limit is called a small limit. If a category $\mathbf{A}$ has all small limits, it is called complete.

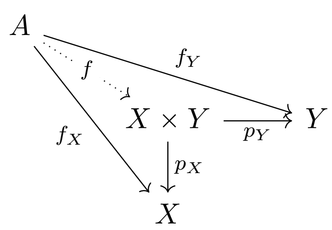
Definition 224. [Lei14, definition 5.1.1, 5.1.7] If the index category $\mathbf{I}$ is discrete, then any diagram $D : \mathbf{I} \to \mathbf{A}$ is simply an indexed family $\{X_i\}_{i \in \mathbf{I}}$ of objects of $\mathbf{A}$. In this case, the limit L does not depend on the functor D. We call it the product in $\mathbf{A}$ indexed by $\mathbf{I}$ and denote it by $\prod_{i \in \mathbf{I}} X_i$.

Explicitly, the product of $\{X\}_{i \in \mathbf{I}}$ is an object $\prod_{i \in \mathbf{I}} X_i$ with associated projection morphisms $\{p_j : \prod_{i \in \mathbf{I}} X_i \to X_j\}_{j \in \mathbf{I}}$, that satisfy the following universal property: for any object A and any family of morphisms $\{f_j : A \to X_j\}_{j \in \mathbf{I}}$ there exists exactly one morphism $f : A \to \prod_{i \in \mathbf{I}} X_i$ such that for every $j \in \mathbf{I}$ we have $p_j \circ f = f_j$, i.e. the following diagram commutes:



The function f is also denoted as $\{f_i\}_{i \in \mathbf{I}}$.

In particular, for two objects $X, Y \in \mathbf{A}$ (i.e. when $\mathbf{I}$ is a two-object discrete category), the product is an object $X \times Y$ with projections $p_X : X \times Y \to X$ and $p_Y : X \times Y \to Y$ such that the following diagram commutes:



Note 225. If the discrete category $\mathbf{I}$ is small, denote the set of its objects by I. This allows us to talk about products of families $\{X_i\}_{i \in I}$ indexed by the set I rather than the category $\mathbf{I}$.

Note 226. The product $\prod_{i \in \varnothing} X_i$ of an empty family of objects is the terminal object of the category.

Definition 227. [Lei14, p. 112] A fork in the category $\mathbf{A}$ is a commutative diagram of the form

$$A \xrightarrow{f} X \overset{s}{\underset{t}{\rightrightarrows}} Y$$

Commutativity simply means that $s \circ f = t \circ f$.

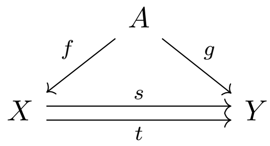
Definition 228. [Lei14, definition 5.1.11] Assume that the index category $\mathbf{I}$ consists of two objects and two unidirectional morphisms:

$$\bullet \rightrightarrows \bullet$$

Diagrams D of shape $\mathbf{I}$ are simply subcategories of $\mathbf{A}$ of the shape

$$X \underset{t}{\overset{s}{\rightrightarrows}} Y$$

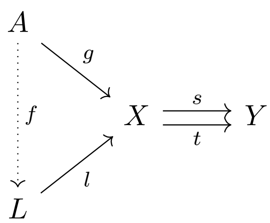
Cones with vertex A are then given by commutative diagrams of shape



Since the morphism $g : A \to Y$ is determined uniquely by f and s, the cones are actually forks:

$$A \xrightarrow{f} X \underset{t}{\overset{s}{\rightrightarrows}} Y$$

The limit (L, l) of D then satisfies the universal property: for any fork (A, g), there exists a unique morphism $f : A \to L$ such that the following diagram commutes:

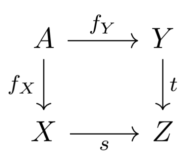


This limit is called the equalizer of s and t.

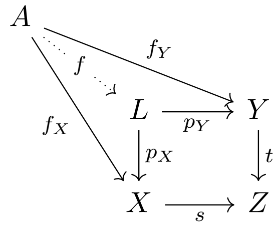
Definition 229. [Lei14, definition 5.1.16] Assume that the index category $\mathbf{I}$ has the shape

$$\bullet \longrightarrow \bullet \longleftarrow \bullet$$

Cones of shape $\mathbf{I}$ with vertex A are then given by commutative diagrams of shape



The limit (L, p_X, p_Y) then satisfies the universal property: for any $\mathbf{I}$-cone with a vertex in $A \in \mathbf{A}$, there exists a unique morphism $f : A \to L$ such that the following diagram commutes:



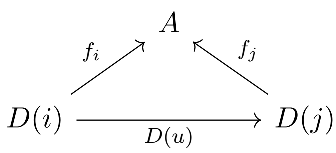
This limit is called the pullback or fibered product of s and t.

Definition 230. [Lei14, definition 5.2.1] The dual notion of a cone[222] is that of a cocone. Given a category $\mathbf{A}$, an index category $\mathbf{I}$ and a diagram $D : \mathbf{I} \to \mathbf{A}$, we say that the family of morphisms $\{D(i) \xrightarrow{f_i} A\}_{i \in \mathbf{I}}$ is a cocone for D if it is a cone for $D^{\mathrm{op}} : \mathbf{I}^{\mathrm{op}} \to \mathbf{A}^{\mathrm{op}}$.

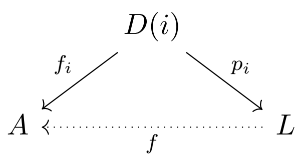
Explicitly, a cocone on D consists of

- an object $A \in \mathbf{A}$, called the vertex of the cocone
- a family of coprojection morphisms $\{f_i : D(i) \to A\}_{i \in \mathbf{I}}$

such that for all morphisms $u : i \to j$ in $\mathbf{I}$, the following diagram commutes:



Definition 231. [Lei14, definition 5.1.19(b)] Analogously to limits[223], we define the colimit $\varinjlim D$ of D to be a cocone $\{D(i) \xrightarrow{p_i} L\}_{i \in \mathbf{I}}$ such that for every cocone $\{D(i) \xrightarrow{f_i} A\}_{i \in \mathbf{I}}$ there exists exactly one morphism $f : L \to A$ such that $f_i = f \circ p_i, i \in \mathbf{I}$, i.e. the following diagram commutes:

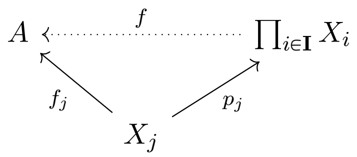


If all small colimits exist, we say that $\mathbf{A}$ is a cocomplete category.

Definition 232. If a category is both complete[223] and cocomplete[231], it is said to be a cocomplete category.

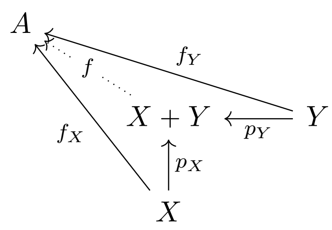
Definition 233. [Lei14, definition 5.2.2] If the index category $\mathbf{I}$ is discrete, specifying a functor $D : \mathbf{I} \to \mathbf{A}$ is analogous to specifying a $\mathbf{I}$-indexed family $\{X\}_{i \in \mathbf{I}}$ of objects in $\mathbf{A}$[224].

The coproduct $\coprod_{i \in \mathbf{I}} X_i$ or categorical sum $\sum_{i \in \mathbf{I}} X_i$ satisfies the following universal property: for any object A and any family of morphisms $\{f_j : X_j \to A\}_{j \in \mathbf{I}}$ there exists exactly one morphism $f : \coprod_{i \in \mathbf{I}} X_i \to A$ such that for every $j \in \mathbf{I}$ we have $f \circ p_j = f_j$, i.e. the following diagram commutes:



The function f is also denoted as $\{f_i\}_{i \in \mathbf{I}}$.

In particular, for two objects $X, Y \in \mathbf{A}$, the product is an object $X + Y$ with co-projections $p_X : X \to X \times Y$ and $p_Y : Y \to X \times Y$ such that the following diagram commutes:



Note 234. The coproduct $\coprod_{i \in \varnothing} X_i$ of an empty family of objects is the initial object of the category.

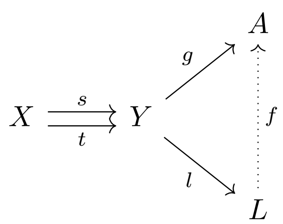
Definition 235. [Lei14, definition 5.2.7] As for equalizers,[235], assume that the index category $\mathbf{I} \cong \mathbf{I}^{\text{op}}$ consists of two objects and two unidirectional morphisms:

$$\bullet \rightrightarrows \bullet$$

Cocones with vertex A are then given by commutative diagrams of shape

$$X \underset{t}{\overset{s}{\rightrightarrows}} Y \xrightarrow{f} A$$

The coequalizer (L, l) then satisfies the universal property: for any $\mathbf{I}$-cocone (A, g), there exists a unique morphism $f : L \to A$ such that the following diagram commutes:

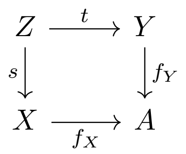


Definition 236. [Lei14, definition 5.2.11] A pushout in $\mathbf{A}$ is a pullback in $\mathbf{A}^{\text{op}}$.

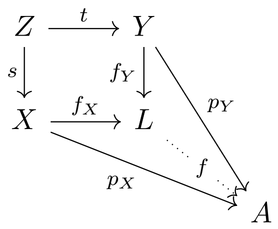
Explicitly, the index category $\mathbf{I}$ has the shape

$$\bullet \longleftarrow \bullet \longrightarrow \bullet$$

Cocones of shape $\mathbf{I}$ with vertex A are then given by commutative diagrams of shape



The pushout (L, p_X, p_Y) of D then satisfies the universal property: for any $\mathbf{I}$-cocone with a vertex in $A \in \mathbf{A}$, there exists a unique morphism $f : L \to A$ such that the following diagram commutes:



Definition 237. [Lei14, definitions 5.3.1, 5.3.5] Let $F : \mathbf{A} \to \mathbf{B}$ be a functor. We say that

a) F preserves limits of shape $\mathbf{I}$ for some index category $\mathbf{I}$ if, given a $\mathbf{I}$-shaped limit cone $\{L \xrightarrow{p_i} D(i)\}_{i \in \mathbf{I}}$, its image $\{F(L) \xrightarrow{F(p_i)} F(D(i))\}_{i \in \mathbf{I}}$ is also a limit cone. We say that F simply preserves limits if it preserves limits for every index category $\mathbf{I}$.

b) F reflects limits of shape $\mathbf{I}$ if, given any $\mathbf{I}$-shaped cone, if its image is a limit cone, then is it itself a limit cone.

c) F creates limits of shape $\mathbf{I}$ if it both preserves and reflects limits.

d) F lifts limits of shape $\mathbf{I}$ if, given a diagram $D : \mathbf{I} \to \mathbf{B}$, any limit cone $\varprojlim D$ is the image of some limit cone in A.

Note 238. Analogous definitions can be given for colimits.

7.4. Abelian categories

Definition 239. [Lan94][158] A monoidal category is a generalization of A monoid[38] from sets to categories. Formally, it is a category $\mathbf{M}$ along with

- a functor $\otimes : C \times C \to C$
- an identity object $1 \in \mathbf{M}$
- natural transformations
  $\alpha : ((-) \otimes (-)) \otimes (-) \cong (-) \otimes ((-) \otimes (-))$
  $\lambda : 1 \times (-) \cong (-)$
  $\rho : (-) \times 1 \cong (-)$

such that

a) for every object $A \in \mathbf{M}$,

$$1 \otimes A \overset{\lambda_a}{\cong} A$$
$$A \otimes 1 \overset{\rho_a}{\cong} A$$

b) for all objects $A, B, C \in \mathbf{M}$,

$$A \otimes (B \otimes C) \overset{\alpha_{A,B,C}}{\cong} (A \otimes B) \otimes C$$

c) the following diagram commutes for all objects $A, B, C, D \in \mathbf{M}$

$$\begin{array}{ccccc} & & (A \otimes B) \otimes (C \otimes D) & & \\ & \overset{\alpha_{A,B,(C\otimes D)}}{\nearrow} & & \overset{\alpha_{(A\otimes B),C,D}}{\searrow} & \\ A \otimes (B \otimes (C \otimes D)) & & & & ((A \otimes B) \otimes C) \otimes D \\ \downarrow {\scriptstyle \mathrm{id} \otimes \alpha_{B,C,D}} & & & & \uparrow {\scriptstyle \alpha_{A,B,C} \otimes \mathrm{id}} \\ A \otimes ((B \otimes C) \otimes D) & & \xrightarrow{\alpha_{A,(B\otimes C),D}} & & (A \otimes (B \otimes C)) \otimes D \end{array}$$

d) the following diagram commutes for all objects $A, B \in \mathbf{M}$

$$\begin{array}{ccc} A \otimes (1 \otimes B) & \xrightarrow{\alpha_{A,1,B}} & (A \otimes 1) \otimes B \\ & {\scriptstyle \mathrm{id} \otimes \lambda_b} \searrow \quad \swarrow {\scriptstyle \rho_a \otimes \mathrm{id}} & \\ & A \otimes B & \end{array}$$

If the natural isomorphisms α, λ and ρ are identities, we say that $\mathbf{M}$ is a strict monoidal category.

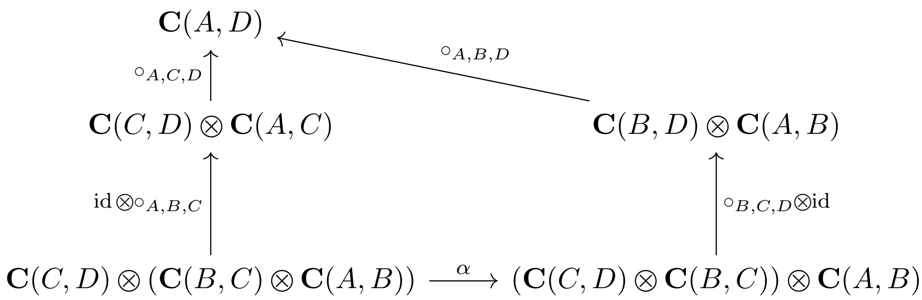
Definition 240. [Lan94][180],[nLa20c] Enriched categories provide additional structure to the morphism sets of locally small categories. The definition can be compared with-definition 175. We say that $\mathbf{C}$ is an enriched category over the small monoidal category $\mathbf{M}$ if

- there exists a class of objects, where the membership is denoted as $A \in \mathbf{C}$
- for each object $A \in \mathbf{C}$, there exists an identity morphism $j_A : 1 \to \mathbf{C}(A, A)$
- for each pair of objects $A, B \in \mathbf{C}$, there exists an object $\mathbf{C}(A, B)$ in $\mathbf{M}$
- for each triple of objects $A, B, C \in \mathbf{C}$, there exists a composition morphism in $\mathbf{M}$:

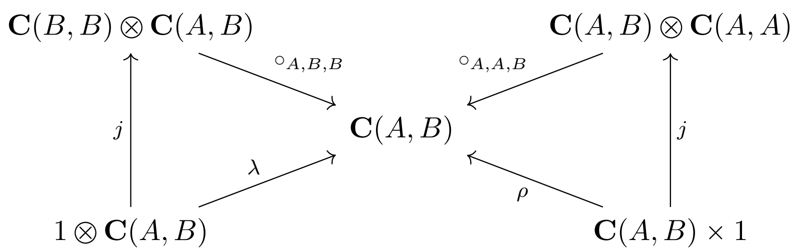
$$\circ_{A,B,C} : \mathbf{C}(B, C) \times \mathbf{C}(A, B) \to \mathbf{C}(A, C).$$

such that

a) the following diagram commutes for all objects $A, B, C, D \in \mathbf{C}$



b) the following diagram commutes for all objects $A, B \in \mathbf{M}$



In order for monoidal categories to actually be categories (more specifically, locally small categories), formally we need a functor $U : \mathbf{M} \to \mathbf{Set}$ so that morphism objects $\mathbf{C}(A, B)$ become sets $U(\mathbf{C}(A, B))$. This is usually defined implicitly, for example $U(\mathbf{C}(A, B)) := \mathbf{M}(1, C(A, B))$.

Definition 241. [Lan94][28] A preadditive category $\mathbf{C}$ is any category enriched over the category $\mathbf{Ab}$ of abelian groups[55], such that composition

$$\circ_{A,B,C} : \mathbf{Ab}(B, C) \times \mathbf{Ab}(A, B) \to \mathbf{Ab}(A, C)$$

is bilinear, e.g. given group homomorphisms $f, f' : A \to B$ and $g, g' : B \to C$, we have

$$(g + g') \circ (f + f') = g \circ f + g \circ f' + g' \circ f + g' \circ f'.$$

Definition 242. Let $\mathbf{C}$ be a category. We say that the morphism $f : A \to B$ is

a) a left-zero morphism or a constant morphism if $f \circ g = f \circ h$ for any two morphisms $g, h : A' \to A$ for any object A'.

b) a right-zero morphism or a coconstant morphism if $g \circ f = h \circ f$ for any two morphisms $g, h : B \to B'$ for any object B'.

c) a zero morphism if it is both a left-zero and a right-zero morphism. We denote it by $0_{A,B}$ if it is unique (for example, in preadditive categories[241]).

Proposition 243. *If* $\mathbf{C}$ *is a preadditive category*[241] *and* $A, B \in \mathbf{C}$*, the identity of* $\mathbf{C}(A, B)$ *is the unique zero morphism*[242] *from* A *to* B*.*

Proof. Denote the identity of $\mathbf{C}(A, B)$ by $0_{A,B}$. We will show that it is a zero morphism in the sense of definition 242.

Let $C \in \mathbf{C}$ and fix a morphism $f : B \to C$. Then, by linearity,

$$f \circ 0_{A,B} + f \circ 0_{A,B} = f \circ (0_{A,B} + 0_{A,B}) = f \circ 0_{A,B}.$$

Thus $f \circ 0_{A,B} = 0_{A,C}$. Since this holds for any function, we conclude that $g \circ 0_{A,B} = h \circ 0_{A,B} = 0_{A,C}$ for any two morphisms in $g, h \in \mathbf{C}(B, C)$ and hence $0_{A,B}$ is a left zero morphism. The proof that $0_{A,B}$ is a right zero morphism is identical. Hence $0_{A,B}$ is a zero morphism.

Now we will show that these are the only zero morphisms in $\mathbf{C}$. Assume that $z : A \to B$ is a zero morphism. Then

$$z = 0_{B,B} \circ z = (0_{B,B} + 0_{B,B}) \circ z = z + z,$$

hence $z = 0_{A,B}$. □

Proposition 244. *If* $\mathbf{C}$ *is a preadditive category, the vertices of nonempty finite products and coproducts coincide.*

Proof. Let $X : \mathbf{I} \to \mathbf{C}$ be a finite discrete diagram. Denote the objects $X(i)$ by X_i and their product by

$$(X, \pi) := \varprojlim D$$

where X is an object in C and

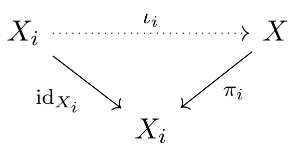
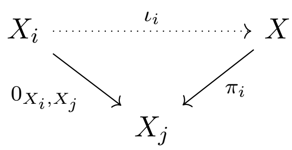
$$\pi = \{\pi_i : X \to X_i\}_{i \in \mathbf{I}}$$

is the family of projections.

Consider the object $X_i \in \mathbf{C}$ with the family of morphisms

$$\begin{Bmatrix} \mathrm{id}_{X_i}, & j = i \\ 0_{X_i, X_j}, & j \neq i \end{Bmatrix}_{j \in \mathbf{I}}$$

By the definition of product[224], there exists a unique map ι_i such that the following diagram commutes

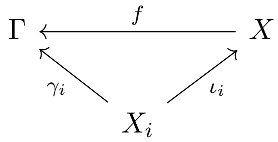
Define $\iota := \{\iota_i\}_{i \in \mathbf{I}}$. We will prove that (X, ι) is a coproduct[233].
Let $\Gamma \in \mathbf{C}$ be an arbitrary object such that there exists a family of morphisms

$$\{\gamma_i : X_i \to \Gamma\}_{i \in \mathbf{I}}.$$

Define

$$f := \sum_{i \in I} (\gamma_i \circ \pi_i) : X \to \Gamma.$$

Fix $i \in \mathbf{i}$. Now we show that the following diagrams commute:



Indeed,

$$f \circ \iota_i = \left(\sum_{j \in \mathbf{I}} \gamma_j \circ \pi_j \right) \circ \iota_i = \sum_{i \in \mathbf{I}} (\gamma_j \circ (\pi_j \circ \iota_i)) = \gamma_i \circ \mathrm{id}_{X_i} + \sum_{\substack{j \in \mathbf{I} \\ j \neq i}} \gamma_j \circ 0_{X_i, X_j} = \gamma_i.$$

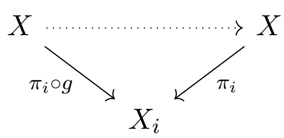
Note that the sum is well-defined since the indexing category $\mathbf{I}$ is finite.
Now we will show that the morphism f is unique.
First define

$$g := \sum_{j \in \mathbf{I}} \iota_j \circ \pi_j : X \to X.$$

Note that for each $i \in \mathbf{I}$,

$$\pi_i \circ g = \pi_i \circ \left(\sum_{j \in \mathbf{I}} \iota_j \circ \pi_j \right) = \sum_{j \in \mathbf{I}} ((\pi_i \circ \iota_j) \circ \pi_j) = \mathrm{id}_i \circ \pi_i + \sum_{\substack{j \in \mathbf{I} \\ j \neq i}} 0_{X, X_j} = \pi_i.$$

We claim that $g = \mathrm{id}_X$. Since X is a product, there exists a unique morphism such that the following diagram commutes for each $i \in \mathbf{I}$:



(7.4)

Both g and id_X satisfy the universal property in eq. (7.4), hence they are equal.

To show that the morphism f is unique, assume that there exists $f' : \Gamma \to X$ such that for each $i \in \mathbf{I}$,

$$f' \circ \iota_i = \gamma_i.$$

But

$$f - f' = (f - f') \circ \mathrm{id}_X = (f - f') \circ \left(\sum_{i \in \mathbf{I}} \iota_i \circ \pi_i \right) = \sum_{i \in \mathbf{I}} ((f \circ \iota_i) \circ \pi_i - (f' \circ \iota_i) \circ \pi_i) = $$
$$= \sum_{i \in \mathbf{I}} (\gamma_i \circ \pi_i - \gamma_i \circ \pi_i) = 0_{\Gamma,X},$$

thus $f = f'$.

Hence the definition of coproduct is satisfied by (X, ι). □

Definition 245. Let $\mathbf{C}$ be a preadditive category. A biproduct of the finite family $\{X_i\}_{i \in I}$ of objects in $\mathbf{C}$ is a triple (X, π, ι), such that (X, π) is a product, (X, ι) is a coproduct.

Note 246. By proposition 244, if a nonempty finite product exists in a preadditive category, so does the corresponding coproduct, hence it is a biproduct. If the empty product exists, however, it may not be a coproduct.

In order to ensure some regularity, additive categories[247] are introduced.

Definition 247. [Lan94][196] A preadditive category[241] is called additive if it has all finite biproducts[245], including empty biproducts (see theorem 248).

Theorem 248. *If* $\mathbf{C}$ *is an additive category, the vertices of finite products and coproducts coincide, that is, they are biproducts.*

Proof. The proof follows from proposition 244 and the fact that the initial[234] and terminal[226] object coincide. □

Definition 249. Let $\mathbf{C}$ be a preadditive category and $f : A \to B$ be a morphism in $\mathbf{C}$. We define the kernel $\ker(f)$ of f as the equalizer[228] of f and $0_{A,B}$. Thus $\ker(f)$ is a morphism from L (the limit vertex) to A.

Analogously, we define the cokernel $\mathrm{coker}(f)$ of f as the coequalizer[235] of f and $0_{A,B}$. Thus $\mathrm{coker}(f) : B \to C$, where C is the colimit vertex.

Definition 250. [Lan94][196] An additive category[247] $\mathbf{C}$ is called an abelian category if:

a) $\mathbf{C}$ has a kernel and a cokernel for every morphism[249]

b) every monomorphism is a kernel and every epimorphism is a cokernel[179]

Proposition 251. *[Lan94][proposition 8.3.1] In an abelian category* $\mathbf{C}$*, every morphism* $f : A \to B$ *has a factorization* $f = \mathrm{im}\, f \circ \mathrm{coim}\, f$*, where*

- $\operatorname{im} f := \ker(\operatorname{coker} f : B \to C_1) : L_1 \to B$ *is a monomorphism*
- $\operatorname{coim} f := \operatorname{coker}(\ker f : L_2 \to A) : A \to C_2$ *is an epimorphism*

Here L_1 and L_2 are the limit vertices and C_1 and C_2 are the colimit vertices as in definition 249. Necessarily $L_1 \cong C_2$.

Definition 252. [Lan94][196] In an abelian category $\mathbf{C}$, a composable pair of morphisms $f : A \to B$ and $g : B \to C$ is said to be exact at B if $\operatorname{im} f \equiv \ker g$ as subobjects of B (or, equivalently, $\operatorname{coker} f \equiv \operatorname{coim} g$; see definition 189).

Definition 253. [Lan94][196] In an abelian category $\mathbf{C}$, the tower diagram[194]

$$0 \xrightarrow{\iota} A \xrightarrow{i} B \xrightarrow{p} C \xrightarrow{\pi} 0 \tag{7.5}$$

is called a short exact sequence (SES) if it is exact at A, B and C (in the sense of definition 252).

Equivalently, eq. (7.5) is short exact if and only if $f \equiv \ker g$ as subobjects of B and $g \equiv \operatorname{coker} f$ as subobjects of C.

Note 254. Since 0 is an initial object, the morphism $\iota : 0 \to A$ exists and is unique. Analogously, $\pi : C \to 0$ exists and is unique. This is why ι and π can be skipped entirely when defining short exact sequences.

The morphism i is necessarily a monomorphism ("i" stands for "injection") since it is equivalent to a kernel and p is necessarily an epimorphism ("p" stands for "projection"). When either i or p is obvious, they may also be skipped.

This makes SES a good framework for describing factorization of algebraic structures, as can be seen in example 257.

Definition 255. [Lan94][198] Consider the two short exact sequences over the same category $\mathbf{C}$:

$$0 \longrightarrow A \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0$$

$$0 \longrightarrow A' \xrightarrow{i'} B' \xrightarrow{p'} C' \longrightarrow 0$$

We say that the triple $f = (f_A : A \to A', f_B : B \to B', f_C : C \to C')$ is a morphism of the short exact sequences if the following diagram commutes:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \xrightarrow{i} & B & \xrightarrow{p} & C & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle f_A} & & \downarrow{\scriptstyle f_B} & & \downarrow{\scriptstyle f_C} & & \\ 0 & \longrightarrow & A' & \xrightarrow{i'} & B' & \xrightarrow{p'} & C' & \longrightarrow & 0 \end{array}$$

If each component of f is an isomorphism, we say that the short exact sequences are isomorphic.

Definition 256. [nLa15] A short exact sequence

$$0 \longrightarrow A \xrightarrow{i} B \xrightarrow{p} C \longrightarrow 0 \tag{7.6}$$

is said to be splitting or split exact if any of the following equivalent conditions hold:

a) i has a left inverse

b) p has a right inverse

c) the sequence eq. (7.6) is isomorphic to the SES

$$0 \longrightarrow A \longrightarrow A \otimes C \longrightarrow C \longrightarrow 0 \tag{7.7}$$

with the canonical injection and projection morphisms

The equivalence of the three conditions is called the splitting lemma.

Example 257.

a) Fix a natural number $n > 0$ and consider the category of **Ab** of abelian groups and the following short exact sequence:

$$0 \longrightarrow \mathbb{Z} \xrightarrow{n\cdot} \mathbb{Z} \xrightarrow{[\cdot]_n} \mathbb{Z}/n\mathbb{Z} \longrightarrow 0$$

where

- $i(x) := nx$ multiplies any integer by n to obtain the subgroup $n\mathbb{Z}$
- $p(x) := [x]_n$ projects any integer into the corresponding remainder when divided by n

The (group-theoretic) image $n\mathbb{Z}$ of i is precisely the (group-theoretic) kernel of $[\cdot]_n$. The sequence does not split since i does not have a left inverse.

b) Consider the additive groups $\mathbb{Z}$, $\mathbb{R}$ and the unit circle group $S_{\mathbb{R}^2}$ with the group operation given by addition of polar angles and with the vector $(1, 0)^T$ as a unit.

$$0 \longrightarrow \mathbb{Z} \xrightarrow{i} \mathbb{R} \xrightarrow{p} S_{\mathbb{R}^2} \longrightarrow 0$$

where

- i is the canonical embedding of $\mathbb{Z}$ is $\mathbb{R}$
- $p := f \circ g$ where $g(x) := \{x\}$ is the fractional part of x (modulo 1) and $f(x) := (\cos(x), \sin(x))^T$ is an embedding of the interval $[0, 1)$ into the unit circle.

Since each integer has fractional part 0 and $p(0) = (1, 0)^T$, the image $\mathbf{Z}$ of $\mathbf{Z}$ under i is the kernel of the group homomorphism p.

The sequence does not split since i is not left-invertible.

c) The following SES of real vector spaces splits

$$0 \longrightarrow \mathbb{R} \xrightarrow{\begin{pmatrix}1\\0\end{pmatrix}} \mathbb{R}^2 \xrightarrow{\begin{pmatrix}0 & 1\end{pmatrix}} \mathbb{R} \longrightarrow 0$$

since all of the following equivalent conditions hold

- $\begin{pmatrix}1 & 0\end{pmatrix}$ is a left inverse to $\begin{pmatrix}1\\0\end{pmatrix}$
- $\begin{pmatrix}0\\1\end{pmatrix}$ is a right inverse to $\begin{pmatrix}0 & 1\end{pmatrix}$
- $\mathbb{R}^2$ is a direct product and a biproduct of two copies of $\mathbb{R}$

d) The fundamental theorem of calculus is a splitting of the SES of vector spaces

$$0 \longrightarrow \mathbb{R} \longrightarrow C^n(\mathbb{R},\mathbb{R}) \xrightarrow{\frac{d}{dx}} C^{n-1}(\mathbb{R},\mathbb{R}) \longrightarrow 0.$$

Definition 258. [nLa20a] In an abelian category $\mathbf{C}$, the tower diagram[194] with objects $\{C_n\}_{n\in\mathbb{Z}}$ and morphisms $\partial_n : C_n \to C_{n-1}$

$$\dots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} C_{-1} \xrightarrow{\partial_{-1}} \dots \tag{7.8}$$

is called a chain complex if for every n,

$$\partial_n \circ \partial_{n+1} = 0_{C_{n+1},C_{n-1}}.$$

Chain complexes may be finite or infinite in one or both directions. The morphisms ∂_n are called boundary maps.

A cochain complex is a chain complex on $\mathbf{C}^{\text{op}}$, i.e.

$$\dots \xleftarrow{\partial_1} C_1 \xleftarrow{\partial_0} C_0 \xleftarrow{\partial_{-1}} C_{-1} \xleftarrow{\partial_{-2}} \dots \tag{7.9}$$

such that for any n,

$$\partial_{n+1} \circ \partial_n = 0_{C_{n-1},C_{n+1}}.$$

8. Order theory

8.1. Posets

Note 259. The definitions in definition 260 mostly make sense for preorders. The study of partial orders[161 (d)], however, appears to be much broader than the study of preorders[161 (b)]. Especially considering proposition 166, which gives us a "normal form" for preorders.

Definition 260. Let $(X, \leqslant)$ be a partially ordered set (poset) as defined in definition 161 (d). We say that

a) the order $<$ defined by $x < y \iff (x \leqslant y \wedge x \neq y)$ is the corresponding strict order[161 (a)].

b) the partially ordered set $(X, \geqslant)$ where $\geqslant$ is the converse relation[165 (a)], is the called the dual poset. See definition 261 for a discussion of the duality.

c) for every subset $A \subseteq X$ the pair $(A, \leqslant_A)$ is a partially ordered set and the restriction

$$\leqslant_A := \{(a, b) \in \leqslant : a \in A \wedge b \in A\}.$$

is called the subset order.

d) $x, y \in X$ are comparable if either $x \leqslant y$ or $y \leqslant x$.

e) $\leqslant$ is a total order on X or a linear order if any two elements are comparable.

f) The subset $A \subseteq X$ is a chain in X if $(A, \leqslant_A)$ is a total order.

g) The subset $A \subseteq X$ is an antichain in X if no two elements of A are compatible.

h) $x \in X$ is an upper bound for $A \subseteq X$ (resp. lower bound for A in the dual poset $(X, \geqslant)$) if $y \leqslant x$ for any $y \in A$.

i) $x \in X$ is a greatest element of X (resp. least element of $(X, \geqslant)$) if $y \leqslant x$ any $y \in X$.

j) $x \in X$ is a maximal element of X (resp. minimal element of $(X, \geqslant)$) if $x \leqslant y$ implies $x = y$ for any $y \in X$.

k) $x \in X$ is a supremum for $A \subseteq X$ (resp. infimum for A in $(X, \geqslant)$) if x is the least upper bound of A, i.e. the least element of the set $U \subset X$ consisting of all upper bounds of X.

l) $\leqslant$ is a well-order on X of $X \neq \varnothing$ and if every nonempty subset has a least element.

Definition 261. [nLa20d] A category[175] $\mathbf{P}$ is called a thin category if, for every two objects $A, B \in \mathbf{P}$, whenever $f, g \in \mathbf{P}(A, B)$, we have $f = g$.

If $\mathbf{P}$ is locally small, this is equivalent to saying that any set of morphisms $\mathbf{P}(A, B)$ is at most a singleton.

Definition 262. We say that $\mathbf{P}$ is a poset category if it is a small[177] thin[261] skeletal[185] category.

See proposition 263.

Proposition 263. *Let $(P, \leqslant)$ be a poset[260]. Then P is also a poset category[262] with objects P and morphisms ordered pairs from $\leqslant$ (viewed as a relation[160]).*

Conversely, if $\mathbf{P}$ is poset category, the pair $\left(\mathbf{P}, \bigcup_{x,y \in \mathbf{P}} \mathbf{P}(x, y)\right)$ is a preordered set.

Furthermore, infima[260 (k)] correspond to categorical products[224], while suprema to coproducts[233].

Proof. We will only prove the equivalence of products and infima since the argument for suprema and coproducts is completely dual.

($\Longrightarrow$) The relation $\leqslant$ satisfies

Transitivity hence composition of morphisms is well defined and is associative and P is a thin category[261].

Reflexivity hence there exist identity morphisms.

Antisymmetry hence the category is skeletal[185].

Let p be the product of the set $A \subseteq X$. Then $p \leqslant x$ for all $x \in A$, hence it is a lower bound. If q is another lower bound, then by definition of product[224], there exists a unique morphism $q \leqslant p$. Hence p is the infimum.

($\Longleftarrow$)

Reflexivity $x \leqslant x$ because of the existence of identity morphisms.

Antisymmetry If $x \leqslant y$ and $y \leqslant x$ implies that there exist morphisms $f : x \to y$ and $g : y \to x$ and, furthermore, these morphisms are unique because $\mathbf{P}$ is thin. So we necessarily have that $f \circ g = \mathrm{id}_x$ and $g \circ f = \mathrm{id}_y$ so x and y are isomorphic and, since $\mathbf{P}$ is skeletal, $x = y$.

Transitivity If $x \leqslant y$ and $y \leqslant z$, then, by composition of morphisms, $x \leqslant z$.

Now since the category is thin, the infimum of A (if it exists) has a unique morphism $\inf A \leqslant x$ for any $x \in A$. If $b \leqslant x$ for all $x \in A$ if another cone[222], then necessarily $b \leqslant \inf A$. Thus the infimum is the categorical product.

□

Note 264. A more general result than proposition 263 states that any small thin category is a preordered set. The proof is the same except we only have isomorphisms in the antisymmetry, not equality.

Proposition 265. *Dual posets correspond to dual poset categories.*

8.2. Lattices

Note 266. Suprema and infima in posets can be used to define operations named joins and meets[270], however there are also axioms for binary joins and meets[267]. If we are interested in infinitary joins and meets, however, we need to use the poset definition. This can be accomplished indirectly by

1. defining joins and meets axiomatically
2. using them to define a partial order
3. using the partial order to define infinitary joins and meets

Definition 267. Fix an arbitrary set X and let $x, y \in X$. Define two binary operations

- the join of x and y, $x \vee y$
- the meet of x and y, $x \wedge y$

such that,

Associativity

$$(x \vee y) \vee z = x \vee (y \vee z) \qquad (x \wedge y) \wedge z = x \wedge (y \wedge z).$$

Commutativity

$$x \vee y = y \vee x \qquad x \wedge y = y \wedge x.$$

Absorption

$$x \vee (y \wedge x) = x \qquad x \wedge (y \vee x) = x.$$

We can use joins to define the partial order relation[161 (d)]

$$x \leqslant y \iff x \vee y = y,$$

thus X is automatically a poset and this structure is compatible with definition 270.

Proof. We will prove that $\leqslant$ is indeed a partial order.

Reflexivity Direct consequence of definition 269 (a).

Antisymmetry Let $x \leqslant y$ and $y \leqslant x$, that is, $x \vee y = y$ and $y \vee x = x$. By Commutativity, $x = y \vee x = x \vee y = y$.

Transitivity Let $x \leqslant y$ and $y \leqslant z$, that is, $x \vee y = y$ and $y \vee z = z$. Then, by Associativity,

$$z = y \vee z = (x \vee y) \vee z = x \vee (y \vee z) = x \vee z.$$

□

Note 268. We can analogously define $x \leqslant y \iff x \wedge y = x$. The resulting partial order would be the same.

Proposition 269. *If* $(X, \vee, \wedge)$ *is a set with a binary join and a meet*[267]*, the following properties hold:*

a) both operations are idempotent, i.e. $x \vee x = x = x \wedge x$

Proof. **269 (a)** Absorption implies that

$$x \vee x = x \vee (x \wedge (x \vee x)) = x$$

and analogously $x \wedge x = x$.

□

Definition 270. Let $(X, \leqslant)$ be a poset[260]. We define joins $\vee$ and meets $\wedge$ as the partial[167] functions

$$\vee : \mathcal{P}(X) \to X \qquad\qquad \wedge : \mathcal{P}(X) \to X$$
$$\vee\,(A) := \sup X \qquad\qquad \wedge\,(A) := \inf X.$$

For finite sets, we usually use the infix notation $x_1 \vee \ldots \vee x_n$ instead of $\vee\{x_1, \ldots, x_n\}$.

Proof. We first show that $\vee$ and $\wedge$ satisfy definition 267. Since suprema and infima are obviously associative and commutative, it remains only to show Absorption, that is, for any comparable $x, y \in X$,

$$x = \sup\{x, \inf\{x, y\}\} \qquad\qquad x = \inf\{x, \sup\{x, y\}\}.$$

If $x \leqslant y$, then

$$\sup\{x, \inf\{x, y\}\} = \sup\{x, x\} = x \qquad\qquad \inf\{x, \sup\{x, y\}\} = \inf\{x, y\} = x.$$

If $y \leqslant x$, then

$$\sup\{x, \inf\{x, y\}\} = \sup\{x, y\} = x \qquad\qquad \inf\{x, \sup\{x, y\}\} = \inf\{x, x\} = x.$$

We now show that if the partial order $\leqslant$ was defined using binary joins and meets as in definition 267, then the original join $\vee$ and meet $\wedge$ are compatible with the binary sup and inf.

Fix $x, y \in X$. Since the functions $\vee$ and $\wedge$ are total, all binary suprema and infima exist. If $\sup\{x, y\} = z$, then z is the least element such that both $x \leqslant z$ and $y \leqslant z$. Thus

$$x \vee z = z = y \vee z.$$

Hence

$$x \vee y = (x \vee (z \wedge x)) \vee y = x \vee z \vee y = x \vee (z \vee z) \vee y = \\ = (x \vee z) \vee (z \vee y) = z \vee z = z.$$

Conversely, if $x \vee y = z$, by Absorption,

$$x \vee z = (x \wedge (y \vee x)) \vee z = (x \wedge z) \vee z = z,$$

thus $x \leqslant z$. Analogously, $y \leqslant z$.

If we assume that there exists $t \in X$ such that $x \leqslant t \leqslant z$ and $y \leqslant t \leqslant z$, then

$$t = t \vee x = t \vee x \vee y = t \vee z = z.$$

Thus $z = \sup\{x, y\}$.

The equivalence between binary inf and $\wedge$ can be obtained analogously. □

Definition 271. A poset $(X, \leqslant)$ is called a lattice if it has

Bottom a top element $\top$, such that $\top = \vee X$ (in particular, $\vee X$ exists).

Top a bottom element $\bot$, such that $\bot = \wedge X$.

Finite joins all finite joins[267] exist.

Finite meets all finite meets[267] exist.

If the join and meet are defined axiomatically, all finite joins and meets necessarily exist.

If the last two properties hold for all joins and meets, not necessarily finite, we say that the lattice is a full lattice.

Note 272. The existence of finite joins and meets is equivalent to the existence of finite products and coproducts in the corresponding category by proposition 263.

Definition 273. [nLa20b] A lattice $(X, \top, \bot, \vee, \wedge)$ is called a distributive lattice if any of the following two equivalent distributive axioms hold for all $x, y, z \in X$:

Distributivity

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z) \qquad x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z).$$

Definition 274. [nLa16] Let X be a distributive lattice. A complement of x is an element y of X such that

$$x \vee y = \top \qquad x \wedge y = \bot.$$

Since in a distributive lattice complements are unique[276 (a)], the complement of x is denoted by $\neg x$. If all elements of X have complements, then $(X, \top, \bot, \vee, \wedge, \neg)$ is called a Boolean algebra.

Example 275. By theorem 127, the equivalence classes of propositional formulas under semantic equivalence form a Boolean algebra.

Proposition 276. *A Boolean algebra X has the following basic properties:*

a) For each $x \in X$, there exists a unique complement $\neg x$.

b) For each $x \in X$, we have $x = \neg\neg x$.

Proof.

276 (a) If y and z are both complements of x, then

$$\begin{aligned} y &= y \wedge \top = y \wedge (z \vee x) = (y \wedge z) \vee (y \wedge x) = \\ &= y \wedge z = (x \wedge z) \vee (y \wedge z) = (x \vee y) \wedge z = z. \end{aligned}$$

276 (b) Fix $x \in X$. We have

$$\begin{aligned} \neg\neg x &= \neg\neg x \wedge \top = \neg\neg x \wedge (\neg x \vee x) = (\neg\neg x \wedge \neg x) \vee (\neg\neg x \wedge x) = \\ &= \neg\neg x \wedge x = (\neg\neg x \wedge x) \vee (\neg x \wedge x) = (\neg\neg x \vee \neg x) \wedge x = x. \end{aligned}$$

□

9. Combinatorics

9.1. Abstract simplicial complexes

Definition 277. [Car09, definition 2.1] An abstract simplicial complex is a pair (V, Σ), where

- V is a finite set
- $\Sigma \subseteq \mathcal{P}(V)$ such that $\sigma \in \Sigma$ and $\tau \subseteq \sigma$ implies $\tau \in \Sigma$.

Due to the equivalence with families of simplices (see proposition 279), elements of V are called vertices and elements Σ are called simplices.

Denote by Σ_k the family of all simplices S with $|S| = k + 1$, that is, all k-simplices.

Definition 278. A simplicial complex in $\mathbb{R}^n$ is a set K of simplices[3], such that

- For any simplex $S \in K$, any face of S is also in K.
- The intersection of any two simplices S_1 and S_2 of K is either empty or is a face of both S_1 and S_2.

Denote by K_k the family of all k-simplices in K.

Proposition 279. *Let (V, Σ) be an abstract simplicial complex[277] and let $v_1 < \ldots < v_n$ be an ordering of elements of V. Define the map $E(v_k) := e_k, k = 1, \ldots, n$, where e_k are the corresponding basis vectors in $\mathbb{R}^n$. Then the set*

$$K := \{\operatorname{conv} E(S) \colon S \in \Sigma\}$$

is a simplicial complex[278].

Conversely, if K is a simplicial complex in $\mathbb{R}^n$, denote by V all 0-simplices (vertices) in K and

$$\Sigma := \{U \subseteq V \colon \operatorname{conv} U \in K\}.$$

Then (V, Σ) is an abstract simplicial complex.

Definition 280. [Car09, p. 262] Let $X = (V, \Sigma)$ be an abstract simplicial complex. For each nonnegative integer k, define the corresponding group of k-chains $C_k(X)$ as the free abelian group[59] generated by the k-simplices Σ_k.

Let $v_1 < \ldots < v_n$ be a total order[260 (e)] on the vertex set V. We define the functions

$$\begin{aligned} &d_i : \Sigma \to \Sigma \\ &d_i(S) := S \backslash \{v_i\}. \end{aligned}$$

and the homomorphisms

$$\begin{aligned} &\partial_k : C_k(X) \to C_{k-1}(X) \\ &\partial_k(S) := \sum_{i=1}^{k} (-1)^i d_i(S) \end{aligned}$$

We can use the induced ordering to represent the operators ∂_k via matrices.

Proposition 281. *In an abstract simplicial complex $X = (V, \Sigma)$, the homomorphisms $\partial_k : C_k(X) \to C_{k-1}(X)$ form a chain complex[258].*

10. Lists

References to the axiom of choice or equivalent statements

References to the law of the excluded middle

References to the compactness of unit spheres

11. Bibliography

[Bri88] Denis Richie Brien Kernighan. *The C programming language.* Englewood Cliffs, N.J: Prentice Hall, 1988. ISBN: 0131103628.

[Car09] Gunnar Carlsson. "Topology and data". In: *Bulletin of the American Mathematical Society* 46.2 (Jan. 29, 2009), pp. 255–308. DOI: 10.1090/s0273-0979-09-01249-x. URL: https://doi.org/10.1090%2Fs0273-0979-09-01249-x.

[Cla13] Francis Clarke. *Functional Analysis, Calculus of Variations and Optimal Control.* Springer London, 2013. DOI: 10.1007/978-1-4471-4820-3. URL: https://doi.org/10.1007/978-1-4471-4820-3.

[Dei85] Klaus Deimling. *Nonlinear functional analysis.* Springer-Verlag, 1985. ISBN: 0387139281.

[DR14] Asen L. Dontchev and R. Tyrrell Rockafellar. *Implicit Functions and Solution Mappings.* Springer New York, 2014. DOI: 10.1007/978-1-4939-1037-3. URL: https://doi.org/10.1007%2F978-1-4939-1037-3.

[Eng89] Ryszard Engelking. *General topology.* Heldermann, 1989. ISBN: 9783885380061.

[Kel55] John L. Kelley. *General Topology.* 1955.

[Kna16] Anthony Knapp. *Basic Algebra (Digital Second Edition).* 2016. URL: http://www.math.stonybrook.edu/~aknapp/.

[Lan94] Saunders Mac Lane. *S. Mac Lane - Categories for the Working Mathematician.* 6th ed. Springer-Verlag, 1994. ISBN: 9780387900353.

[Lei14] Tom Leinster. *Basic Category Theory.* 2014. URL: https://arxiv.org/abs/1612.09375.

[Ner12] Anil Nerode. *Logic for Applications.* Springer, 2012. ISBN: 9781461268550.

[nLa15] nLab authors. *split exact sequence.* Dec. 20, 2015. URL: https://ncatlab.org/nlab/revision/split+exact+sequence/16.

[nLa16] nLab authors. *Boolean algebra.* Aug. 28, 2016. URL: https://ncatlab.org/nlab/revision/Boolean+algebra/21.

[nLa20a] nLab authors. *chain complex.* Aug. 25, 2020. URL: https://ncatlab.org/nlab/show/chain+complex/33.

[nLa20b] nLab authors. *distributive lattice.* Mar. 20, 2020. URL: https://ncatlab.org/nlab/revision/distributive+lattice/22.

[nLa20c] nLab authors. *enriched category*. Apr. 16, 2020. URL: http://ncatlab.org/nlab/revision/enriched+category/93.

[nLa20d] nLab authors. *thin category*. July 1, 2020. URL: https://ncatlab.org/nlab/revision/thin+category/8.

[nLa20e] nLab authors. *Top*. June 6, 2020. URL: https://ncatlab.org/nlab/revision/Top/42.

[Phe93] Robert Phelps. *Convex functions, monotone operators, and differentiability*. Springer-Verlag, 1993. ISBN: 0387567151.

[ЙТ74] Александр Давидович Йоффе and Владимир Михайлович Тихомиров. *Теория экстремальных задач*. Нелинейный анализ и его приложения. Наука, 1974.

[КС16] Борис Коцев and Пламен Сидеров. *Комутативна алгебра (второ издание)*. 2016. ISBN: 9789548857383.