Prerequisites to the Langlands program

A mammoth project

Tien Nguyen Thanh

25 July 2026

©Tien Nguyen Thanh

PARTS SUMMARY

INTRODUCTION

The goal of this ambitious mammoth project is to absorb or reabsorb as much prerequisites as possible in order to even start to understand what the *Langlands Program* is about. This idea was mainly born during my Bachelor's thesis, which was about the fundamental lemma for quadratic base change in GL_2 (thanks to my advisor for giving me this opportunity). Personally, I think that there is much to do in this field, albeit quite difficult.

Moreover, I felt the need to relearn and to reorder much of the knowledge I have acquired through lecture courses at the University of Bonn in one central place. Most lecturers did a very good job at teaching one semester's worth of coherent content, yet at some point, all of these need to be connected. Furthermore, a semester is just not enough time to cover the standard canon on any specific topic, so this project also wants to remedy that. Lastly, I forget a lot of stuff over time, so this project gives me a chance to recall a lot of things.

The *modus operandi* will be me reading a few books at a time and trying to integrate the content bit-by-bit into this document. I will jump between different parts, however, depending on whatever motivates me at the moment. Topics are inspired by the article [Kna10] and include the following:

(i) Commutative algebra. Maybe non-commutative algebra, but I am not sure.
(ii) Algebraic number theory and class field theory.
(iii) Algebraic geometry, especially scheme theory.
(iv) Infinite-dimensional representation theory of Lie groups and Lie algebras.
(v) Modular forms, maybe some complex analysis.

I have also added a chapter on category theory.

The text layout is mostly inspired by my `grothendieck` package with a few tweaks.

PART I

COMMUTATIVE ALGEBRA

CHAPTER CONTENTS

INTRODUCTION

Reference **.0.0.1.** — Mainly [Mat86]; [Mat80]. Supplemented by [AM16].

Notation **.0.0.2.** — – Despite how [Mat86]; [Mat80] denote extensions and retractions, we do not follow this convention. Namely, if $f\colon A \to B$ is a ring homomorphism, then $f^{-1}(\mathfrak{b})$ is denoted as is for an ideal $\mathfrak{b}$ of B, and $f(\mathfrak{a})B$ is denoted as is for an ideal $\mathfrak{a}$ of A (instead of $A \cap \mathfrak{b}$ and $\mathfrak{a}B$).

– Ideals are written with Fraktur letters, not capital ones, following [AM16]. If $\mathfrak{a}$ is an ideal of a ring A and if $x \in A$, then $x\mathfrak{a} = \{xa \mid a \in \mathfrak{a}\}$.

– We denote the evaluation map by $\mathrm{ev}_a\colon A[x_i \mid i \in I] \to A$, $x_i \mapsto a_i$ for $a \coloneqq (a_i)_{i \in I}$ a family of elements $a_i \in A$.

CHAPTER I

RINGS AND MODULES

§ I.1. RINGS

I.1.1. Commutative rings and ring homomorphisms.

***Definition* I.1.1.1.** — *A **ring** is a set A with two binary operations addition $+\colon A \times A \to A$ and multiplication $\cdot\colon A \times A \to A$ satisfying the following conditions.*

(i) *The set A forms an abelian group w. r. t. addition. In particular, there is a zero element 0 in A, and each $x \in A$ has an additive inverse $-x \in A$.*

(ii) *Multiplication is associative and distributive over addition, that is for all $x, y, z \in A$, we have the following properties:*

$$(xy)z = x(yz), \qquad x(y+z) = xy + xz, \qquad (x+y)z = xz + yz.$$

We will always require the following additional properties.

(iii) *The ring A is **commutative** if $xy = yx$ for all $x, y \in A$.*

(iv) *The ring A has an **identity element** 1 such that $x1 = 1x = x$ for all $x \in A$. Note that this element is unique.*

***Remark* I.1.1.2.** — It is still possible that $0 = 1$ in a ring A. In this case, $x = x1 = x0 = 0$ for all $x \in A$, so A has precisely one element, namely 0. We call A the *zero ring* and write $A = 0$. Sometimes, we will forget to exclude the zero ring from theorem statements, just to be aware.

***Definition* I.1.1.3.** — *A **ring homomorphism** is a map $f\colon A \to B$ between rings satisfying the following conditions.*

(i) *It is additive, i. e., $f(x+y) = f(x) + f(y)$ for all $x, y \in A$. In other words, f is a group homomorphism of abelian groups. In particular, $f(-x) = -f(x)$ for all $x \in A$ and $f(0) = 0$.*

(ii) *It is multiplicative, i. e., $f(xy) = f(x)f(y)$ for all $x, y \in A$.*

(iii) *It respects the identity element, i. e., $f(1) = 1$.*

*If f is bijective, then f is called a **ring isomorphism**.*

***Convention* I.1.1.4.** — *From now on, the word* ring *should always mean a commutative ring with an identity element. Similarly, the word* ring map *should always mean a ring homomorphism.*

***Fact* I.1.1.5.** — *The composition of two ring maps is again a ring map.*

***Definition* I.1.1.6.** — *A subset S of a ring A is a **subring** if it is closed under addition and multiplication and if it contains the identity element $1 \in A$. In this case, S is again a ring, and the (injective) canonical inclusion $\iota\colon S \hookrightarrow A$ is a ring map. We also say that A is an **extension ring** of S.*

***Definition* I.1.1.7.** — *Let A be a ring.*

(i) *An element $x \in A$ is a **zero divisor** if it divides zero, i. e., there exists some $y \in A \setminus \{0\}$ such that $xy = 0$. If $A \neq 0$ and if A does not contain any zero divisors except zero, then we call A an **integral domain** or just* domain.

(ii) *An element $x \in A$ is **nilpotent** if $x^n = 0$ for some $n > 0$. If A does not contain any nilpotent elements except zero, then we call A **reduced**.*

(iii) *An element $x \in A$ is a **unit** or **invertible** if it divides one, i. e., there exists some $y =: x^{-1} \in A$ such that $xy = 1$. The inverse of x is unique, and the units of A form a multiplicative abelian group $A^\times$. If $A \neq 0$ and if every element in A except zero is invertible, then we call A a **field**.*

***Remark* I.1.1.8.** — Every field is an integral domain: if $xy = 0$ with $x \neq 0$, then $0 = x^{-1}(xy) = y$. The converse is not true: $\mathbb{Z}$ is an integral domain, but not a field.

***Fact* I.1.1.9.** — *Let $u \in A^\times$ be a unit and let $x \in A$ be nilpotent. Then $u + x$ is a unit.*

Proof. — Say $x^n = 0$, and set $y := -u^{-1}x$. Then $y^n = 0$. Recall the famous factorisation $(1 - y)(1 + y + \dots + y^{n-1}) = 1 - y^n = 1$, which says that $1 - y$ is a unit. Thus $u + x = u(1 - y)$ is a unit as well.

I.1.2. Ideals.

Many properties of rings can be conveniently expressed through the concept of *ideals*.

***Definition* I.1.2.1.** — *An **ideal** $\mathfrak{a}$ of a ring A is a subset satisfying the following properties.*

(i) *The set $\mathfrak{a}$ is an additive subgroup of A.*

(ii) *It is closed under multiplication with A, i. e., $A\mathfrak{a} \subseteq \mathfrak{a}$ (i. e., $xy \in \mathfrak{a}$ for all $x \in A$ and $y \in \mathfrak{a}$).*

If S is a subset of A, then

$$(S) := \bigcap_{\textit{ideals } \mathfrak{a} \supseteq S} \mathfrak{a} = \left\{ \sum_{i=1}^n x_i s_i \;\middle|\; x_i \in A, s_i \in S \right\}$$

denotes the smallest ideal in A containing S. If $S = \{s_1, \dots, s_n\}$ is finite, we also write $(S) = (s_1, \dots, s_n) = \sum_{i=1}^n As_i$. We also say that (S) is the ideal generated *by S.*

***Example* I.1.2.2.** — There are always two trivial ideals: the *zero ideal* (0) and the *unit ideal* $A = (1)$. An element $x \in A$ is a unit iff $(u) = (1)$.

***Construction* I.1.2.3.** — Let A be a ring and let $\mathfrak{a}$ be an ideal of A. Then the abelian quotient group $A/\mathfrak{a}$ inherits a multiplication from A, defined by $(x + \mathfrak{a})(y + \mathfrak{a}) = xy + \mathfrak{a}$ for all $x, y \in A$. This makes $A/\mathfrak{a}$ a ring, called the **quotient ring** or just *quotient* of A by $\mathfrak{a}$. It comes with a (surjective) canonical projection $\pi\colon A \twoheadrightarrow A/\mathfrak{a}$, $x \mapsto x + \mathfrak{a}$.

It is common to denote the image of any element $x \in A$ in $A/\mathfrak{a}$ by $\bar{x} \in \bar{A} := A/\mathfrak{a}$. For two elements $x, y \in A$, we also write $x \equiv y \mod \mathfrak{a}$ if $\bar{x} = \bar{y}$ or equivalently if $x - y \in \mathfrak{a}$.

***Proposition* I.1.2.4** (ideal correspondence)**.** — *Let A be a ring and let $\mathfrak{a}$ be an ideal in A. Let $\pi\colon A \twoheadrightarrow A/\mathfrak{a}$ be the canonical projection. Then there is a one-to-one, inclusion-preserving correspondence*

$$\begin{aligned} \{\textit{ideals } \mathfrak{b} \textit{ of } A \textit{ containing } \mathfrak{a}\} &\longleftrightarrow \{\textit{ideals } \bar{\mathfrak{b}} \textit{ of } A/\mathfrak{a}\}, \\ \mathfrak{b} &\longmapsto \pi(\mathfrak{b}) = \mathfrak{b}/\mathfrak{a}, \\ \pi^{-1}(\bar{\mathfrak{b}}) &\longleftarrow\!\shortmid \bar{\mathfrak{b}}. \end{aligned}$$

Moreover, if $\mathfrak{b}$ is any ideal in A, then $\pi(\mathfrak{b}) = (\mathfrak{b} + \mathfrak{a})/\mathfrak{a}$ is an ideal of $A/\mathfrak{a}$.

This is powerful because if we want to consider ideals of a ring which contain another fixed ideal, we can simply pass to quotients in order to remove this condition.

***Homomorphism theorem* I.1.2.5.** — *Let $f\colon A \to B$ be a ring map. Then the so-called **kernel** $\ker f := f^{-1}(0)$ of f is an ideal of A, and the so-called **image** $\operatorname{im} f := f(A)$ of f is a subring of B. Moreover, there exists a canonical isomorphism*

$$A/\ker f \xrightarrow{\sim} \operatorname{im} f$$

induced by f.

***Remark* I.1.2.6.** — In particular, as with groups, a ring map $f\colon A \to B$ is injective if $\ker f = (0)$, and it is surjective if $\operatorname{im} f = B$. Also, if $\mathfrak{b}$ is an ideal in B, then $f^{-1}(\mathfrak{b})$ is an ideal in A since it is the kernel of $A \to B \twoheadrightarrow B/\mathfrak{b}$.

***First isomorphism theorem* I.1.2.7.** — *Let A be a ring, let S be a subring of A, and let $\mathfrak{a}$ be an ideal of A. Then sum $S + \mathfrak{a}$ is a subring of A, the intersection $S \cap \mathfrak{a}$ is an ideal of S, and there exists a canonical isomorphism*

$$S/(S \cap \mathfrak{a}) \xrightarrow{\sim} (S + \mathfrak{a})/\mathfrak{a}.$$

***Second isomorphism theorem* I.1.2.8.** — *Let A be a ring with two ideals $\mathfrak{a}$ and $\mathfrak{b}$ such that $\mathfrak{a} \subseteq \mathfrak{b}$. Then $\mathfrak{b}/\mathfrak{a}$ is an ideal of $A/\mathfrak{a}$ (cf. proposition I.1.2.4), and there is a canonical isomorphism*

$$(A/\mathfrak{a})/(\mathfrak{b}/\mathfrak{a}) \xrightarrow{\sim} A/\mathfrak{b}.$$

Proof of the above statements. — The above statements hold on the level of abelian groups and can be easily extended to the case of rings by checking multiplicativity of the ring maps. For more details, consider the proofs of the homomorphism theorem I.2.2.8, first isomorphism theorem I.2.2.9, and second isomorphism theorem I.2.2.10 in the case of modules. ◻

We will use these basic statements repeatedly and will not reference them.

***Proposition* I.1.2.9** (characterisation of fields)**.** — *Let $A \neq 0$ be a ring. Then the following are equivalent:*

(i) *A is a field.*
(ii) *The only ideals of A are the trivial ones (0) and (1).*
(iii) *Any ring map $A \to B \neq 0$ is injective.*

Proof. — Assume (i). Let $\mathfrak{a} \neq (0)$ be an ideal in A. Then there exists $x \in \mathfrak{a} \setminus (0)$, which must be a unit. Hence $(1) = (x) \subseteq \mathfrak{a}$, and thus $\mathfrak{a} = (1)$. This shows (ii).

Now assume (ii), and let $f\colon A \to B \neq 0$ be any ring map. Since f is not the zero map, $\ker f$ is a proper ideal of A. Hence $\ker f = (0)$, and thus (iii) holds.

Lastly assume (iii). Suppose that there exists an element $x \in A$ that is not a unit. Then $(x) \neq (1)$ and $A/(x) \neq 0$. By hypothesis, the projection $A \twoheadrightarrow A/(x)$ is injective with kernel $(x) = (0)$; thus $x = 0$, which shows (i). ◻

***Fact* I.1.2.10.** — *Let $(\mathfrak{a}_i)_{i\in I}$ be a family of ideals in A. Then the following sets are again ideals in A.*

(i) *The **sum***

$$\sum_{i\in I} \mathfrak{a}_i := \left\{ \sum_{i\in I} x_i \;\middle|\; x_i \in \mathfrak{a}_i, x_i = 0 \text{ for almost all } i \in I \right\},$$

which is the smallest ideal containing all $\mathfrak{a}_i$.

(ii) *The intersection* $\bigcap_{i\in I}\mathfrak{a}_i$.
(iii) *If I is finite, the* ***product***

$$\prod_{i\in I}\mathfrak{a}_i := \left\{\textit{finite sums of elements } \prod_{i\in I} x_i \;\middle|\; x_i\in\mathfrak{a}_i\right\}.$$

In particular, if $\mathfrak{a}$ is an ideal of A, then $\mathfrak{a}^n$ for $n>0$ is the ideal generated by all products of the form $x_1\cdots x_n$ with $x_i\in\mathfrak{a}$.

In particular, the ideals of A form a complete lattice w. r. t. inclusion, i. e., infima and suprema always exist.

Proof. — Trivial calculation.

***Example* I.1.2.11.** — (i) Let $A=\mathbb{Z}$, $\mathfrak{a}=(m)$ and $\mathfrak{b}=(n)$ for $m,n\in\mathbb{Z}$. Then $\mathfrak{a}+\mathfrak{b}=(\gcd(m,n))$, $\mathfrak{a}\cap\mathfrak{b}=(\operatorname{lcm}(m,n))$ and $\mathfrak{a}\mathfrak{b}=(mn)$. In particular, $\mathfrak{a}\mathfrak{b}=\mathfrak{a}\cap\mathfrak{b}$ iff $\operatorname{lcm}(m,n)\mid\gcd(m,n)$ iff m and n are coprime.

(ii) Let $A=k[x_1,\dots,x_n]$ be a polynomial ring over a field k, and let $\mathfrak{a}=(x_1,\dots,x_n)$. Then $\mathfrak{a}^m$ is the set of all polynomials whose monomial terms have degree at least m.

***Warning* I.1.2.12.** — In general, the union $\mathfrak{a}\cup\mathfrak{b}$ of two ideals is not an ideal. In fact, $(\mathfrak{a}\cup\mathfrak{b})=\mathfrak{a}+\mathfrak{b}$.

Definition* I.1.2.13.** — *Two ideals $\mathfrak{a}$ and $\mathfrak{b}$ of a ring are said to be* ***coprime *if $\mathfrak{a}+\mathfrak{b}=(1)$, or equivalently, if there are $x\in\mathfrak{a}$ and $y\in\mathfrak{b}$ such that $x+y=1$.*

***Proposition* I.1.2.14** (laws for sums, intersections and products of ideals)**.** — *Let $\mathfrak{a}$, $\mathfrak{b}$ and $\mathfrak{c}$ be ideals of a ring.*

(i) Distributive law: $\mathfrak{a}(\mathfrak{b}+\mathfrak{c})=\mathfrak{a}\mathfrak{b}+\mathfrak{a}\mathfrak{c}$.
(ii) Modular law: $\mathfrak{a}\cap(\mathfrak{b}+\mathfrak{c})=(\mathfrak{a}\cap\mathfrak{b})+\mathfrak{c}$ *if* $\mathfrak{a}\supseteq\mathfrak{c}$.
(iii) $\mathfrak{a}\cap\mathfrak{b}=\mathfrak{a}\mathfrak{b}$ *if $\mathfrak{a}$ and $\mathfrak{b}$ are coprime. In general, $\bigcap_{i=1}^n\mathfrak{a}_i=\prod_{i=1}^n\mathfrak{a}_i$ if the ideals $\mathfrak{a}_i$ are pairwise coprime.*

Proof. — (i) and (ii) follow by checking the equalities elementwise. We only show (iii). Suppose that $\mathfrak{a}+\mathfrak{b}=(1)$. Then

$$\mathfrak{a}\mathfrak{b}\subseteq\mathfrak{a}\cap\mathfrak{b}=(\mathfrak{a}+\mathfrak{b})(\mathfrak{a}\cap\mathfrak{b})=\mathfrak{a}(\mathfrak{a}\cap\mathfrak{b})+\mathfrak{b}(\mathfrak{a}\cap\mathfrak{b})\subseteq\mathfrak{a}\mathfrak{b}.$$

***Construction* I.1.2.15.** — Let $A_1,\dots,A_n$ be rings. Then the direct product of abelian groups $A=\prod_{i=1}^n A_i$ becomes a ring via not only componentwise addition, but also componentwise multiplication, called the **direct product** of the $A_1,\dots,A_n$. The identity element is $(1,\dots,1)\in A$. It comes with (surjective) canonical projections $\pi_i\colon A\twoheadrightarrow A_i$, $(x_1,\dots,x_n)\mapsto x_i$.

***Chinese remainder theorem* I.1.2.16.** — *Let A be a ring and let $\mathfrak{a}_1,\dots,\mathfrak{a}_n$ be ideals of A. Define the ring map*

$$\phi\colon A\longrightarrow\prod_{i=1}^n(A/\mathfrak{a}_i),\qquad x\longmapsto(x+\mathfrak{a}_1,\dots,x+\mathfrak{a}_n).$$

Then the following hold:

(i) *ϕ is surjective iff $\mathfrak{a}_1,\dots,\mathfrak{a}_n$ are pairwise coprime.*
(ii) *ϕ is injective iff $\bigcap_{i=1}^n\mathfrak{a}_i=(0)$.*
(iii) *If $\mathfrak{a}_1,\dots,\mathfrak{a}_n$ are pairwise cooprime, then $\prod_{i=1}^n\mathfrak{a}_i=\bigcap_{i=1}^n\mathfrak{a}_i$.*

Under the hypothesis that $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ are pairwise coprime, we have the isomorphism

$$A/\prod_{i=1}^{n} \mathfrak{a}_i \xrightarrow{\sim} \prod_{i=1}^{n} (A/\mathfrak{a}_i).$$

Proof. — (ii) is clear since $\ker \phi = \bigcap_{i=1}^{n} \mathfrak{a}_i$. We now show (i). Suppose that ϕ is surjective. W. l. o. g., we will only show that $\mathfrak{a}_1$ and $\mathfrak{a}_2$ are coprime. There is some $x \in A$ with $\phi(x) = (1, 0, \dots, 0)$, so that $x \equiv 1 \mod \mathfrak{a}_1$ and $x \equiv 0 \mod \mathfrak{a}_2$. Thus $1 = (1 - x) + x \in \mathfrak{a}_1 + \mathfrak{a}_2$ as desired.

Conversely, suppose that the $\mathfrak{a}_i$ are pairwise coprime. Observe that since the product ring is additive, it suffices to find preimages for elements with only one non-zero component. Let us do this for $(\bar{y}, 0, \dots, 0)$ where $y \in A$. By hypothesis, we have relations $u_i + v_i = 1$ for all $2 \leq i \leq n$ with $u_i \in \mathfrak{a}_1$ and $v_i \in \mathfrak{a}_i$. Take $x := y \prod_{i=2}^{n} v_i$; then $x = y \prod_{i=2}^{n} (1 - u_i) \equiv y \mod \mathfrak{a}_1$ and $x \equiv 0 \mod \mathfrak{a}_i$ for all $i \geq 2$. Thus $\phi(x) = (\bar{y}, 0, \dots, 0)$.

Finally to (iii). The induction start is proposition I.1.2.14. For the induction step, suppose that $\prod_{i=1}^{n-1} \mathfrak{a}_i = \bigcap_{i=1}^{n-1} \mathfrak{a}_i =: \mathfrak{b}$. Since $\mathfrak{a}_i + \mathfrak{a}_n = (1)$ for all $1 \leq i < n$ by hypothesis, there are relations $x_i + y_i = 1$ with $x_i \in \mathfrak{a}_i$ and $y_i \in \mathfrak{a}_n$. For $\prod_{i=1}^{n} x_i \in \mathfrak{b}$, it follows that $\prod_{i=1}^{n-1} x_i = \prod_{i=1}^{n-1} (1 - y_i) \equiv 1 \mod \mathfrak{a}_n$, and hence $\mathfrak{a}_n + \mathfrak{b} = (1)$. Thus $\mathfrak{b}\mathfrak{a}_n = \mathfrak{b} \cap \mathfrak{a}_n$. ∎

***Definition* I.1.2.17.** — *Let $f\colon A \to B$ be a ring map.*

(i) *If $\mathfrak{a}$ is an ideal of A, then $Bf(\mathfrak{a})$ is an ideal of B which consists of finite sums $\sum y_i f(x_i)$ where $x_i \in \mathfrak{a}$ and $y_i \in B$. We call $Bf(\mathfrak{a})$ the **extended ideal** of $\mathfrak{a}$ along f.*

(ii) *If $\mathfrak{b}$ is an ideal of B, then the preimage $f^{-1}(\mathfrak{b})$ is an ideal of A, called the **contracted ideal** of $\mathfrak{b}$ along f.*

***Warning* I.1.2.18.** — If $f\colon A \to B$ is a ring map and if $\mathfrak{a}$ is an ideal of A, then $f(\mathfrak{a})$ is generally not an ideal of B. For example, take the embedding $f\colon \mathbb{Z} \hookrightarrow \mathbb{Q}$. If $\mathfrak{a} = (n)$ is a non-zero ideal of $\mathbb{Z}$ with $n \in \mathbb{Z} \setminus \{0\}$, then $f(\mathfrak{a}) = \{mn \mid m \in \mathbb{Z}\}$ is not an ideal of $\mathbb{Q}$: the only ideals of the field $\mathbb{Q}$ are (0) and $(1) = \mathbb{Q}$.

***Fact* I.1.2.19** (laws for extensions and contractions)**.** — *Let $f\colon A \to B$ be a ring map, and let $(-)^e$ and $(-)^c$ denote the extension and contraction along f resp. Let $\mathfrak{a}$ be an ideal of A, and let $\mathfrak{b}$ be an ideal of B. Then the following hold:*

(i) $\mathfrak{a} \subseteq \mathfrak{a}^{ec}$ *and* $\mathfrak{b} \supseteq \mathfrak{b}^{ce}$.
(ii) $\mathfrak{a}^e = \mathfrak{a}^{ece}$ *and* $\mathfrak{b}^c = \mathfrak{b}^{cec}$.
(iii) *There is an inclusion preserving bijection*

$$\{\textit{contracted ideals in } A\} = \{\mathfrak{a} \mid \mathfrak{a}^{ec} = \mathfrak{a}\} \longleftrightarrow \{\mathfrak{b} \mid \mathfrak{b}^{ce} = \mathfrak{b}\} = \{\textit{extended ideals in } B\},$$
$$\mathfrak{a} \longmapsto \mathfrak{a}^e,$$
$$\mathfrak{b}^c \longleftarrow\!\shortmid \mathfrak{b}.$$

For ideals $\mathfrak{a}_1, \mathfrak{a}_2$ of A and $\mathfrak{b}_1, \mathfrak{b}_2$ of B, we have the following rules:[I.1]

(iv) $(\mathfrak{a}_1 + \mathfrak{a}_2)^e = \mathfrak{a}_1^e + \mathfrak{a}_2^e$ *and* $(\mathfrak{b}_1 + \mathfrak{b}_2)^c \supseteq \mathfrak{b}_1^c + \mathfrak{b}_2^c$.
(v) $(\mathfrak{a}_1 \cap \mathfrak{a}_2)^e \subseteq \mathfrak{a}_1^e \cap \mathfrak{a}_2^e$ *and* $(\mathfrak{b}_1 \cap \mathfrak{b}_2)^c = \mathfrak{b}_1^c \cap \mathfrak{b}_2^c$.
(vi) $(\mathfrak{a}_1 \mathfrak{a}_2)^e = \mathfrak{a}_1^e \mathfrak{a}_2^e$ *and* $(\mathfrak{b}_1 \mathfrak{b}_2)^c \supseteq \mathfrak{b}_1^c \mathfrak{b}_2^c$.

Proof. — (i) is easy set theory. (ii) follows from (i): we have $\mathfrak{a}^e \subseteq (\mathfrak{a}^{ec})^e = (\mathfrak{a}^e)^{ce} \subseteq \mathfrak{a}^e$; similarly for $\mathfrak{b}^c$. (iii) follows immediately from (ii). The other statements follow by direct computation. ∎

[I.1] After introducing radicals and ideal quotients, there are also the following laws: (i) $(\mathfrak{a}_1 : \mathfrak{a}_2)^e \subseteq (\mathfrak{a}_1^e : \mathfrak{a}_2^e)$ and $(\mathfrak{b}_1 : \mathfrak{b}_2)^c \subseteq (\mathfrak{b}_1^c : \mathfrak{b}_2^c) = (\mathfrak{b}_1 : \mathfrak{b}_2^{ce})^c$. (ii) $\sqrt{\mathfrak{a}} \subseteq \sqrt{\mathfrak{a}^e}$ and $\sqrt{\mathfrak{b}}^c = \sqrt{\mathfrak{b}^c}$.

I.1.3. Prime and maximal ideals.

Prime ideals are fundamental in commutative algebra, as they have a striking correspondence to the geometric side in algebraic geometry. They also behave quite well.

***Definition* I.1.3.1.** — *Let A be a ring.*

(i) *An ideal $\mathfrak{p}$ of A is **prime** if $\mathfrak{p} \neq (1)$ and if $xy \in \mathfrak{p}$ implies $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$. The set of all prime ideals is denoted by* $\operatorname{Spec} A$ *and is called the **prime spectrum** or simply **spectrum** of A.*

(ii) *An ideal $\mathfrak{m}$ of A is **maximal** if $\mathfrak{m} \neq (1)$ and there is no ideal $\mathfrak{a}$ such that $\mathfrak{m} \subsetneq \mathfrak{a} \subsetneq (1)$. The set of all maximal ideals is denoted by* $\operatorname{MaxSpec} A$ *and is called the **maximal spectrum** of A.*

***Proposition* I.1.3.2** (characterisation of prime and maximal ideals)**.** — *Let A be a ring.*

(i) *An ideal $\mathfrak{p}$ of A is prime iff $A/\mathfrak{p}$ is an integral domain. In particular, the zero ideal is prime iff A is an integral domain.*

(ii) *An ideal $\mathfrak{m}$ of A is maximal iff $A/\mathfrak{m}$ is a field. In particular, maximal ideals are prime.*

Proof. — For (i), let $xy \in \mathfrak{p}$, i. e., $\bar{x}\bar{y} = 0$. Then $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$ iff $\bar{x} = 0$ or $\bar{y} = 0$ iff $A/\mathfrak{p}$ is an integral domain. On the other hand, (ii) follows directly from the ideal correspondence and proposition I.1.2.9. Since fields are integral domains, maximal ideals are prime. ∎

***Warning* I.1.3.3.** — The converse that prime ideals are maximal is not always true. Take for a counterexample any integral domain that is not a field, e. g., $A = \mathbb{Z}$.

***Fact* I.1.3.4** (Spec is functorial)**.** — *Let $f\colon A \to B$ be a ring map. If $\mathfrak{q}$ is a prime ideal of B, then $\mathfrak{p} = f^{-1}(\mathfrak{q})$ is a prime ideal of A. In this case, we say that $\mathfrak{q}$ **lies over** $\mathfrak{p}$ or that $\mathfrak{p}$ **lies under** $\mathfrak{q}$. Hence, we obtain a map* $\operatorname{Spec} f\colon \operatorname{Spec} B \to \operatorname{Spec} A$.

Proof. — Consider the composition $A \to B \twoheadrightarrow B/\mathfrak{q}$ with kernel $f^{-1}(\mathfrak{q})$. Then $A/f^{-1}(\mathfrak{q})$ is isomorphic to a subring of $B/\mathfrak{q}$, which is an integral domain by proposition I.1.3.2. So $A/f^{-1}(\mathfrak{q})$ is an integral domain as well, i. e., $f^{-1}(\mathfrak{q})$ is prime. ∎

***Warning* I.1.3.5.** — Contracted maximal ideals are generally not maximal: consider the ring map $\mathbb{Z} \hookrightarrow \mathbb{Q}$ with maximal ideal $\mathfrak{n} = (0)$ in $\mathbb{Q}$. The pullback is (0) in $\mathbb{Z}$, which is not maximal, but prime.

***Warning* I.1.3.6.** — Extended prime ideals are generally not prime. For example, consider the ring map $\mathbb{Z} \hookrightarrow \mathbb{Q}$, and let $\mathfrak{p}$ be a non-zero prime ideal of $\mathbb{Z}$. Then $\mathbb{Q}\mathfrak{p} = \mathbb{Q}$ is not a prime ideal of $\mathbb{Q}$.

***Example* I.1.3.7.** — A general problem in algebraic number theory is how prime ideals behave under extension. For example, consider the inclusion $\mathbb{Z} \hookrightarrow \mathbb{Z}[i]$ with $\mathrm{i} = \sqrt{-1}$. One can show that $\mathbb{Z}[\mathrm{i}]$ is a euclidean domain, whence it is a principal ideal domain. Moreover, the extension of primes is as follows:

(i) If $p = 2$, then $p\mathbb{Z}[\mathrm{i}] = (1+\mathrm{i})^2$ is the square of the prime ideal $(1+\mathrm{i})$ in $\mathbb{Z}[\mathrm{i}]$.

(ii) If $p \equiv 1 \mod 4$ is prime, then $p\mathbb{Z}[\mathrm{i}]$ is the product of two distinct prime ideals in $\mathbb{Z}[\mathrm{i}]$, e. g., $5\mathbb{Z}[\mathrm{i}] = (2+\mathrm{i})(2-\mathrm{i})$.

(iii) If $p \equiv 3 \mod 4$ is prime, then $p\mathbb{Z}[\mathrm{i}]$ is prime in $\mathbb{Z}[\mathrm{i}]$.

This result is non-trivial. In particular, (ii) is essentially the reason why each prime $p \equiv 1 \mod 4$ can be uniquely written as a sum of two squares, e. g., $5 = 2^2 + 1^2$, $97 = 9^2 + 4^2$, etc.; this is a classical theorem due to Fermat.

***Proposition* I.1.3.8.** — *Let $A_1, \dots, A_n$ be rings. The prime ideals in $\prod_{i=1}^n A_i$ are of the form $A_1 \times \dots \times A_{i-1} \times \mathfrak{p}_i \times A_{i+1} \times \dots \times A_n$ where $\mathfrak{p}_i$ is a prime ideal in A_i. More succinctly,* $\operatorname{Spec} \prod_{i=1}^n A_i = \coprod_{i=1}^n \operatorname{Spec} A_i$.

Proof. — Let $\mathfrak{p}$ be a prime ideal of $A := \prod_{i=1}^n A_i$. Write $e_i := (0, \ldots, 0, 1, 0, \ldots, 0)$ for the standard vectors in A. Since $\mathfrak{p} \neq (1)$, w. l. o. g. we have $e_1 \notin \mathfrak{p}$. Since $\mathfrak{p}$ is prime, $e_1 e_i = 0 \in \mathfrak{p}$ for all $i \geq 1$ implies that $e_i \in \mathfrak{p}$ for all $i \geq 2$. Hence $\mathfrak{p} = \mathfrak{p}_1 \times \prod_{i=2}^n A_i$ for some set $\mathfrak{p}_1 \subseteq A_1$. Now we pass to the quotient $A/0 \times \prod_{i=2}^n A_i \cong A_1$. Then $\mathfrak{p} \equiv \mathfrak{p}_1 \mod \prod_{i=2}^n A_i$, so $\mathfrak{p}_1$ is an ideal of A_1 by ideal correspondence. We finally see that $\mathfrak{p}_1$ must be prime in order for $\mathfrak{p}$ to be prime. ◻

***Prime avoidance lemma* I.1.3.9.** — *Let A be a ring.*

(i) *Let $\mathfrak{a}, \mathfrak{p}_1, \ldots, \mathfrak{p}_n$ be ideals of A such that $\mathfrak{p}_3, \ldots, \mathfrak{p}_n$ are prime. If $\mathfrak{a} \nsubseteq \mathfrak{p}_i$ for all i, then $\mathfrak{a} \nsubseteq \bigcup_{i=1}^n \mathfrak{p}_i$.*

(ii) *Let $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ be ideals of A, and let $\mathfrak{p}$ be a prime ideal of A. If $\mathfrak{p} \nsupseteq \mathfrak{a}_i$ for all i, then $\mathfrak{p} \nsupseteq \bigcap_{i=1}^n \mathfrak{a}_i$. In fact, $\mathfrak{p} \nsupseteq \prod_{i=1}^n \mathfrak{a}_i$.*

Proof. — We first show (i) by induction on n. The case $n = 1$ is trivial, so let $n \geq 2$. We may assume that $\mathfrak{p}_i \nsubseteq \mathfrak{p}_j$ for all $i \neq j$ by omitting those $\mathfrak{p}_i$ which do not satisfy this. If $n = 2$, take $x \in \mathfrak{a} \setminus \mathfrak{p}_2$ and $y \in \mathfrak{a} \setminus \mathfrak{p}_1$, and suppose that $\mathfrak{a} \subseteq \mathfrak{p}_1 \cup \mathfrak{p}_2$. Then $x \in \mathfrak{p}_1 \setminus \mathfrak{p}_2$ and $y \in \mathfrak{p}_2 \setminus \mathfrak{p}_1$, so $x + y \notin \mathfrak{p}_1 \cup \mathfrak{p}_2$, a contradiction.

If $n > 2$, take $x \in S := \mathfrak{a} \setminus \bigcup_{i=1}^{n-1} \mathfrak{p}_i$ by induction hypothesis. Since $\mathfrak{p}_n$ is prime and since $\mathfrak{a}, \mathfrak{p}_i \nsubseteq \mathfrak{p}_n$ for all $i < n$, we have $\mathfrak{a}\mathfrak{p}_1 \cdots \mathfrak{p}_{n-1} \nsubseteq \mathfrak{p}_n$, so take $y \in \mathfrak{a}\mathfrak{p}_1 \cdots \mathfrak{p}_{n-1} \setminus \mathfrak{p}_n$. Suppose that $\mathfrak{a} \subseteq \bigcup_{i=1}^n \mathfrak{p}_i$. Then $x \in \mathfrak{p}_n \setminus \bigcup_{i=1}^{n-1} \mathfrak{p}_i$ and $y \in \bigcap_{i=1}^{n-1} \mathfrak{p}_i \setminus \mathfrak{p}_n$. Hence, $x + y \notin \bigcup_{i=1}^n \mathfrak{p}_i$, a contradiction.[I.2]

Now (ii). By hypothesis, there are $x_i \in \mathfrak{a}_i \setminus \mathfrak{p}_i$ for all i. Hence $\prod_{i=1}^n x_i \in \prod_{i=1}^n \mathfrak{a}_i \subseteq \bigcap_{i=1}^n \mathfrak{a}_i$, yet $\prod_{i=1}^n x_i \notin \mathfrak{p}$ since $\mathfrak{p}$ is prime. Thus $\mathfrak{p} \nsupseteq \bigcap_{i=1}^n \mathfrak{a}_i$. ◻

***Remark* I.1.3.10.** — If A contains an infinite field k, then the condition that $\mathfrak{p}_3, \ldots, \mathfrak{p}_n$ are prime in (i) is superfluous. Namely, all ideals are k-vector spaces. So $\mathfrak{a} \nsubseteq \mathfrak{p}_i$ means that $\mathfrak{a} \cap \mathfrak{p}_i$ is a proper subspace of $\mathfrak{a}$. Since k is infinite, we have $\bigcup_{i=1}^n (\mathfrak{a} \cap \mathfrak{p}_i) \subsetneq \mathfrak{a}$, and so $\mathfrak{a} \nsubseteq \bigcup_{i=1}^n \mathfrak{p}_i$.

Definition* I.1.3.11.** — *A subset S of a ring is* ***multiplicatively closed *or just* ***multiplicative*** *if $x, y \in S$ implies $xy \in S$ and if $1 \in S$. Equivalently, a multiplicatively closed subset of a ring is a submonoid.*

***Krull's theorem* I.1.3.12.** — *Let $A \neq 0$ be a ring. Let S be a multiplicative subset of A, and let $\mathfrak{a}$ be an ideal disjoint from S. Then there exists a prime ideal $\mathfrak{p}$ in A that contains $\mathfrak{a}$ and that is disjoint from S, and we can choose $\mathfrak{p}$ maximally.*

Proof. — This is a standard application of *Zorn's lemma*.[I.3] Let Σ denote the set of all ideals in A that contain $\mathfrak{a}$ but are disjoint from S. Then Σ naturally admits a partial order by inclusion. Of course, Σ is non-empty since $\mathfrak{a} \in S$. Let $(\mathfrak{a}_\alpha)_\alpha$ be any chain in Σ, i. e., $\mathfrak{a}_\alpha \subseteq \mathfrak{a}_\beta$ or $\mathfrak{a}_\beta \subseteq \mathfrak{a}_\alpha$ for any indices α and β. Then this chain admits an upper bound $\mathfrak{a} = \bigcup_\alpha \mathfrak{a}_\alpha$. One checks that this is an ideal, and it indeed belongs to Σ. So by Zorn's lemma, there is a maximal element $\mathfrak{p} \in \Sigma$.

[I.2] The distinction between $n = 2$ and $n > 2$ can be skipped if $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are prime as well. This weaker form of the prime avoidance lemma is certainly easier to remember.

[I.3] A *partial order* $\leq$ on a set S is a relation that is reflexive ($x \leq x$), transitive ($x \leq y$ and $y \leq z$ implies $x \leq z$) and antisymmetric ($x \leq y$ and $y \leq x$ implies $x = y$). A subset T of S is a *chain* if $x \leq y$ or $y \leq x$ for any $x, y \in T$. An *upper bound* of a subset T of S is some element $x \in S$ such that $t \leq x$ for all $t \in T$.
Zorn's lemma states the following: Let S be a non-empty partially ordered set. If every chain of S has an upper bound in S, then S has (at least) one maximal element. It is well-known that Zorn's lemma is equivalent to the axiom of choice.

It remains to see that $\mathfrak{p}$ is prime. Let $x, y \notin \mathfrak{p}$. By maximality of $\mathfrak{p}$ in Σ, both $\mathfrak{p} + (x)$ and $\mathfrak{p} + (y)$ meet S, i. e., intersect S non-trivially. Then the product $(\mathfrak{p} + (x))(\mathfrak{p} + (y))$ meets S too. Since this product lies in $\mathfrak{p} + (xy)$, we have $\mathfrak{p} + (xy) \notin \Sigma$, and thus $xy \notin \mathfrak{p}$ as desired.(I.4)

***Corollary* I.1.3.13.** — *Let $A \neq 0$ be a ring. Then each ideal not being the unit ideal is contained in a maximal ideal. Or more succinctly, A admits a maximal ideal.*

Proof. — Choose $S = \{1\}$ or in the second case, $S = \{1\}$ and $\mathfrak{a} = (0)$.

***Remark* I.1.3.14.** — If A is *noetherian* (see definition I.5.1.3), then the set of all prime ideals of A disjoint from S has a maximal element by fact I.5.1.2. In this case, we do not need to appeal to Zorn's lemma, i. e., the axiom of choice.

***Proposition* I.1.3.15.** — *Let $A \neq 0$ be a ring, and let $\mathfrak{a}$ be a proper ideal of A. Then there is a minimal prime ideal above $\mathfrak{a}$.*

Proof. — This is again a standard application of Zorn's lemma to the set of all prime ideals containing $\mathfrak{a}$. One needs to check that for a chain of prime ideals $(\mathfrak{p}_\alpha)_\alpha$, the intersection $\mathfrak{p} = \bigcap_\alpha \mathfrak{p}_\alpha$ is again a prime ideal: for any $x, y \notin \mathfrak{p}$, there are β and γ such that $x \notin \mathfrak{p}_\beta$ and $y \notin \mathfrak{p}_\gamma$. Say $\mathfrak{p}_\beta \subseteq \mathfrak{p}_\gamma$; then $x, y \notin \mathfrak{p}_\gamma$, and hence $xy \notin \mathfrak{p}_\gamma \supseteq \bigcap_\alpha \mathfrak{p}_\alpha$.

Definition* I.1.3.16.** — *Given an ideal $\mathfrak{a}$ in a ring A, the* ***radical *of $\mathfrak{a}$ is the set*

$$\sqrt{\mathfrak{a}} := \{x \in A \mid x^n \in \mathfrak{a} \text{ for some } n > 0\}.$$

In particular, $\operatorname{nil} A := \sqrt{(0)}$ *is the* ***nilradical*** *of A, the set of all nilpotent elements.*

***Proposition* I.1.3.17** (characterisation of radical ideals)**.** — *Then the radical $\sqrt{\mathfrak{a}}$ of an ideal $\mathfrak{a}$ is always an ideal. In fact, we have $\sqrt{\mathfrak{a}} = \bigcap_{\mathfrak{p} \supseteq \mathfrak{a}} \mathfrak{p}$ where $\mathfrak{p}$ runs through all prime ideals.*(I.5)

Proof. — It is instructive to show that $\sqrt{\mathfrak{a}}$ is an ideal first. Let $x, y \in \sqrt{\mathfrak{a}}$ and $a \in A$, say $x^n, y^m \in \mathfrak{a}$ for $m, n > 0$. Obviously, $(ax)^n \in \mathfrak{a}$, and so $ax \in \sqrt{\mathfrak{a}}$. To show $x + y \in \sqrt{\mathfrak{a}}$, expand the term $(x+y)^{n+m-1}$. This is a finite sum of multiples of products $x^r y^s$ with $r + s = m + n - 1$; in particular, we always have either $r \geq n$ or $s \geq m$. So this term is in $\mathfrak{a}$, whence $x + y \in \sqrt{\mathfrak{a}}$.

Now the stronger statement. Denote by $\mathfrak{p}'$ the intersection of all prime ideals containing $\mathfrak{a}$. First, let $\mathfrak{p}$ be a prime ideal containing $\mathfrak{a}$, and let $x \in \sqrt{\mathfrak{a}}$, say $x^n \in \mathfrak{a} \subseteq \mathfrak{p}$. Then $x \in \mathfrak{p}$, i. e., $\sqrt{\mathfrak{a}} \subseteq \mathfrak{p}'$. Conversely, for any $x \notin \sqrt{\mathfrak{a}}$, consider the multiplicative subset $S = \{1, x, x^2, \dots\}$, which is disjoint from $\mathfrak{a}$. By Krull's theorem I.1.3.12, there is a prime ideal $\mathfrak{p}$ that contains $\mathfrak{a}$ but not x, i. e., $x \notin \mathfrak{p}'$. This proves $\sqrt{\mathfrak{a}} = \mathfrak{p}'$.

***Example* I.1.3.18.** — Let $A = \mathbb{Z}$ and let $\mathfrak{a} = (m)$ be an ideal. If $p_1, \dots, p_r$ are the distinct prime divisors of m, then $\sqrt{\mathfrak{a}} = \bigcap_{i=1}^r (p_i) = (p_1 \cdots p_r)$.

***Corollary* I.1.3.19** (characterisation of reducedness)**.** — *Let A be ring. Then $A_{\text{red}} := A / \operatorname{nil} A$ is reduced. In particular, A is reduced if and only if* $\operatorname{nil} A = (0)$.

(I.4) There is an alternative proof with more advanced methods. One shows via Zorn's lemma that non-zero rings have maximal ideals, which are prime. Now given $\mathfrak{a}$ and S, consider the *localisation* of $A/\mathfrak{a}$ at the image of S. This is again a non-zero ring (since $\mathfrak{a} \cap S = \varnothing$, the image $\bar{S}$ of S cannot contain any zero divisors, so the localisation is non-zero). By the original Krull's theorem, this contains a maximal ideal, from which we can reconstruct a prime ideal in A.

(I.5) Alternatively, one can first show that $\operatorname{nil} A$ is an ideal and that it is the intersection of all prime ideals in A. Then $\sqrt{\mathfrak{a}} = \operatorname{nil}(A/\mathfrak{a})$ and the statement follows.

Proof. — Let $\bar{x} \in A/\operatorname{nil} A$ with lift $x \in A$. Suppose that $\bar{x}$ is nilpotent, say $\bar{x}^n = 0$. It follows that $x^n \in \operatorname{nil} A$, say $x^{nm} = 0$ for some $m > 0$. Hence $x \in \operatorname{nil} A$, i. e., $\bar{x} = 0$.

***Proposition* I.1.3.20** (laws for radicals). — *Let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals. Then the following hold:*

(i) $\mathfrak{a} \subseteq \sqrt{\mathfrak{a}}$.
(ii) $\sqrt{\mathfrak{a}} = \sqrt{\sqrt{\mathfrak{a}}}$.
(iii) $\sqrt{\mathfrak{a}\mathfrak{b}} = \sqrt{\mathfrak{a} \cap \mathfrak{b}} = \sqrt{\mathfrak{a}} \cap \sqrt{\mathfrak{b}}$.
(iv) $\sqrt{\mathfrak{a}} = 1$ *iff* $\mathfrak{a} = (1)$.
(v) $\sqrt{\mathfrak{a} + \mathfrak{b}} = \sqrt{\sqrt{\mathfrak{a}} + \sqrt{\mathfrak{b}}}$.
(vi) *If $\mathfrak{p}$ is a prime ideal, then $\sqrt{\mathfrak{p}^n} = \mathfrak{p}$ for all $n > 0$.*

Proof. — These are shown by explicit calculation.

***Corollary* I.1.3.21.** — *Let $\mathfrak{a}$ and $\mathfrak{b}$ be two ideals. Then if $\sqrt{\mathfrak{a}}$ and $\sqrt{\mathfrak{b}}$ are coprime, then $\mathfrak{a}$ and $\mathfrak{b}$ are too.*

Proof. — $\sqrt{\mathfrak{a} + \mathfrak{b}} = \sqrt{\sqrt{\mathfrak{a}} + \sqrt{\mathfrak{b}}} = \sqrt{(1)} = (1)$ by proposition I.1.3.20, and so $\mathfrak{a} + \mathfrak{b} = (1)$.

Definition* I.1.3.22.** — *Let $A \neq 0$ be a ring. The intersection of all maximal ideals in A is called the* ***Jacobson radical $\operatorname{rad} A$ *of A.*[I.6]

***Proposition* I.1.3.23** (characterisation of $\operatorname{rad} A$). — *Let $A \neq 0$ be a ring. Then $x \in \operatorname{rad} A$ iff $1 - xy \in A^\times$ for all $y \in A$.*

Proof. — Suppose that $x \in \operatorname{rad} A$, i. e., x is contained in every maximal ideal. Suppose that $1 + xy$ is not a unit. Then by corollary I.1.3.13, there is some maximal ideal $\mathfrak{m}$ with $1 - xy \in \mathfrak{m}$. Hence $x \notin \mathfrak{m}$, for otherwise, $1 \in \mathfrak{m}$. In other words, $x \notin \operatorname{rad} A$.

For the converse, suppose that $x \notin \operatorname{rad} A$, i. e., $x \notin \mathfrak{m}$ for some maximal ideal $\mathfrak{m}$. So $\mathfrak{m} + (x) = (1)$ by maximality, and $u + xy = 1$ for some $u \in \mathfrak{m}$ and some $y \in A$. It follows that $1 - xy \in \mathfrak{m}$, i. e., $1 - xy$ is not a unit.

Definition* I.1.3.24.** — *A ring A is* ***local *if it has a unique maximal ideal $\mathfrak{m}$. The field $k = A/\mathfrak{m}$ is called the* ***residue field*** *of A. We often write $(A, \mathfrak{m}, k)$ or just $(A, \mathfrak{m})$ for a local ring. Likewise, a ring is* ***semilocal*** *if it has finitely many maximal ideals.*

***Proposition* I.1.3.25** (characterisation of local rings). — *Let $A \neq 0$ be a ring.*

(i) *Let $\mathfrak{m}$ be a proper ideal of A. Then $(A, \mathfrak{m})$ is local iff $A \setminus \mathfrak{m}$ only consists of units.*
(ii) *Let $\mathfrak{m}$ be a maximal ideal of A. Then $(A, \mathfrak{m})$ is local iff $1 + x \in A^\times$ for all $x \in \mathfrak{m}$.*

Proof. — Let $\mathfrak{m}$ be a proper ideal of A. Then $(A, \mathfrak{m})$ is local iff every proper ideal of A is contained in $\mathfrak{m}$ iff $A \setminus \mathfrak{m}$ consists of units (since proper ideals do not contain units). This shows (i).

For (ii), let $\mathfrak{m}$ be a maximal ideal of A. Then by proposition I.1.3.23, $(A, \mathfrak{m})$ is local iff $\operatorname{rad} A = \mathfrak{m}$ iff $1 + (x) \subseteq A^\times$ only for all $x \in \mathfrak{m}$. We can drop 'only' since we always have $\operatorname{rad} A \subseteq \mathfrak{m}$. Now since $A\mathfrak{m} = \mathfrak{m}$, we have $1 + (x) \subseteq A^\times$ for all $x \in \mathfrak{m}$ iff $1 + \mathfrak{m} \subseteq A^\times$.

Definition* I.1.3.26.** — *In a ring A, an element e is an* ***idempotent *if $e^2 = e$.*

***Example* I.1.3.27.** — Any ring has always the trivial idempotents 0 and 1.

***Fact* I.1.3.28.** — *In a local ring $(A, \mathfrak{m})$, all idempotents are trivial.*

Proof. — Let $e \in A$ be an idempotent. Then $e(1 - e) = 0 \in \mathfrak{m}$, so $e \in \mathfrak{m}$ or $1 - e \in \mathfrak{m}$. By proposition I.1.3.25, we have $1 - e \in A^\times$ or $e \in A^\times$ resp. Hence $e(1 - e) = 0$ implies $e = 0$ or $1 - e = 0$ resp.

[I.6] The term 'radical' in 'Jacobson radical' does not imply that $\operatorname{rad} A$ can be written as a radical $\sqrt{\mathfrak{a}}$. The term just arose historically.

I.1.4. Examples of rings.

Definition* I.1.4.1.** — *Let A be a ring. The* ***ring of formal power series *in one variable x with coefficients in A is*

$$A[\![x]\!] := \left\{ \sum_{n=0}^{\infty} a_n x^n \;\middle|\; a_n \in A \right\}.$$

Addition and multiplication is defined as usual. Note that since multiplication is defined distributively, the coefficients of a product of two power series involves only finite sums:

$$\left(\sum_{n=0}^{\infty} a_n x^n\right)\left(\sum_{n=0}^{\infty} b_n x^n\right) = \sum_{n=0}^{\infty}\left(\sum_{k=0}^{n} a_k b_{n-k}\right) x^n.$$

Similarly, the ring of formal power series in multiple variables $x_1, \dots, x_n$ can be defined inductively by

$$A[\![x_1, \dots, x_n]\!] = \left\{ a + \sum_{i=1}^{n} a_i x_i + \sum_{i,j=1}^{n} a_{ij} x_i x_j + \cdots \;\middle|\; a, a_i, a_{ij}, \dots \in A \right\} = A[\![x_1, \dots, x_{n-1}]\!][\![x_n]\!].$$

***Lemma* I.1.4.2.** — *A power series $f = \sum_{n=0}^{\infty} a_n x^n$ is a unit in $A[\![x]\!]$ iff $a_0 \in A^\times$. More generally, a power series $f = a + \sum_{i=0}^{n} a_i x_i + \cdots$ is a unit in $A[\![x_1, \dots, x_n]\!]$ iff $a \in A^\times$.*

Proof. — Suppose that $f \in A[\![x]\!]^\times$. If $f^{-1} = \sum_{n=0}^{\infty} b_n x^n$, then $ff^{-1} = 1$ implies that $a_0 b_0 = 1$, i. e., $a_0 \in A^\times$.

Conversely, suppose that $a_0 \in A^\times$. We want to construct the inverse $g = \sum_{n=0}^{\infty} b_n x^n \in A[\![x]\!]$. Explicitly expanding the product $1 = fg$ and comparing coefficients, we would need to successively solve equations $a_0 b_0 = 1$, $a_0 b_1 + a_1 b_0 = 0$, etc. This is possible, so $g = f^{-1}$ exists. The general case is proven identically.

An alternative argument for $f \in A[\![x]\!]$ follows. Since $a_0 \in A^\times$, we may consider $a_0^{-1} f$. So, w. l. o. g., write $f = 1 - gx$ for some $g \in A[\![x]\!]$. Then by a geometric series, $f^{-1} = (1 - gx)^{-1} = \sum_{n=0}^{\infty} (gx)^n$. Expanding this product yields the desired inverse. □

***Corollary* I.1.4.3.** — *A power series $f = a + \cdots \in A[\![x_1, \dots, x_n]\!]$ is in $\operatorname{rad} A[\![x_1, \dots, x_n]\!]$ iff $a \in \operatorname{rad} A$.*

Proof. — By proposition I.1.3.23 and lemma I.1.4.2, we have $f \in \operatorname{rad} B := A[\![x_1, \dots, x_n]\!]$ iff $1 + (f) \subseteq B^\times$ iff $1 + (a) \subseteq A^\times$ iff $a \in \operatorname{rad} A$. □

***Example* I.1.4.4.** — What are the maximal ideals in $B := A[\![x_1, \dots, x_n]\!]$? By corollary I.1.4.3, we have $\mathfrak{a} := (x_1, \dots, x_n) \subseteq \operatorname{rad} B$. So the maximal ideals of B correspond to those in $B/\mathfrak{a} \cong A$. Therefore, the maximal ideals in B are precisely all ideals $\mathfrak{m}B + \mathfrak{a}$ for some maximal ideal $\mathfrak{m}$ of A (we identified A with the constant power series in B here). Conversely, $(\mathfrak{m}B + \mathfrak{a}) \cap A = \mathfrak{m}$.

However, the case for $A[x]$ is much more complicated. It is not true that each maximal ideal of $A[x]$ must contain (x). Namely, $x - 1 \in A[x]$ is not a unit, so by Krull's theorem I.1.3.12, there is a maximal ideal $\mathfrak{m}$ containing $x - 1$; thus $x \notin \mathfrak{m}$. Moreover, if $\mathfrak{m}$ is a maximal ideal in $A[x]$, then $A \cap \mathfrak{m}$ might not be maximal in A (cf. warning I.1.3.5).

***Proposition* I.1.4.5.** — *If A is a local ring, then $A[\![x_1, \dots, x_n]\!]$ is local too. In particular, if k is a field, then the ideals of $k[\![x]\!]$ are (x^n) for $n \geq 0$ and (0).*

Proof. — Let $(A, \mathfrak{m})$ be a local ring. By induction, it suffices to show that $A[\![x]\!]$ is local. Note that $A[\![x]\!]/(\mathfrak{m} + (x)) \cong A/\mathfrak{m}$ is a field, so $\mathfrak{m} + (x)$ is a maximal ideal in $A[\![x]\!]$ by proposition I.1.3.2. Hence with lemma I.1.4.2 and proposition I.1.3.25, we see that $A[\![x]\!] \setminus (\mathfrak{m} + (x))$ are the units of $A[\![x]\!]$, and so $A[\![x]\!]$ is local.

In the case that $(A, \mathfrak{m}) = (k, (0))$, the maximal ideal of $k[\![x]\!]$ is (x), and all other proper ideals lie below (x). Let $\mathfrak{a} \subseteq (x)$ be a proper non-zero ideal of $k[\![x]\!]$. Then each $f \in \mathfrak{a}$ can be uniquely written as $f = x^n u$ where $u \in k[\![x]\!]^\times$; define $\deg f := n$. Set $m := \min\{\deg f \mid f \in \mathfrak{a}\} \geq 1$; there is some $f_0 \in \mathfrak{a}$ such that $\deg f_0 = m$. Then $(x^m) = (f_0) \subseteq \mathfrak{a} \subseteq (x^m)$ by minimality of m.

***Fact* I.1.4.6.** — *If A is an integral domain, then $A[x]$ and $A[\![x]\!]$ are integral domains too.*

Proof. — If $f, g \neq 0$ are polynomials or power series, consider the non-zero coefficients of lowest degree. Then fg has a non-zero coefficient of lowest degree, i. e., $fg \neq 0$, since A is an integral domain.

***Example* I.1.4.7.** — Let A be a ring and let $\mathfrak{a}$ be an ideal of A. Denote by $\mathfrak{a}[x]$ the polynomials in $A[x]$ with coefficients in $\mathfrak{a}$. Then $\mathfrak{a}[x]$ is an ideal since it is the kernel of the reduction of coefficients modulo $\mathfrak{a}$, i. e., ring maps $A[x] \twoheadrightarrow (A/\mathfrak{a})[x]$. In particular, there is the isomorphism $A[x]/\mathfrak{a}[x] \xrightarrow{\sim} (A/\mathfrak{a})[x]$. With proposition I.1.3.2, we see that $\mathfrak{p}[x]$ is a prime ideal of $A[x]$ if $\mathfrak{p}$ is a prime ideal of A. A similar statement exists in the case of $A[\![x]\!]$ and $\mathfrak{p}[\![x]\!]$.

***Example* I.1.4.8.** — Let A be a ring and let $\mathfrak{a}$ be an ideal of A. If $\mathfrak{a} = (a_1, \ldots, a_n)$ is finitely generated, then $\mathfrak{a}[\![x]\!] = \sum_{i=1}^n a_i A[\![x]\!] = \mathfrak{a} \cdot A[\![x]\!]$. If $\mathfrak{a} = (a_\alpha)_\alpha$ is infinitely generated, then $\mathfrak{a}[\![x]\!]$ is bigger than $\mathfrak{a} \cdot A[\![x]\!]$. For example, take $f = \sum_{i=0}^\infty a_{\alpha_i} x^i \in \mathfrak{a}[\![x]\!]$ for pairwise distinct a_{α_i}, yet $\mathfrak{a} \cdot A[\![x]\!]$ has only finitely many elements from $\mathfrak{a}$ as coefficients. This problem does not appear with polynomials: we always have $\mathfrak{a}[x] = \mathfrak{a} \cdot A[x]$.

***Definition* I.1.4.9.** — *Let A be an integral domain.*

(i) *An element $a \in A \setminus \{0\}$ is **irreducible** if $a \notin A^\times$ and if $a = bc$ implies $b \in A^\times$ or $c \in A^\times$.*

(ii) *For two elements $a, b \in A$, we write $a \mid b$ if $b = ac$ for some $c \in A$. An element $a \in A \setminus \{0\}$ is **prime** if $a \notin A^\times$ and if $a \mid bc$ implies $a \mid b$ or $a \mid c$.*

***Lemma* I.1.4.10.** — *Let A be an integral domain and let $a \in A \setminus \{0\}$.*

(i) *(a) is prime iff a is prime.*

(ii) *(a) is maximal among all principal ideals iff a is irreducible.*

Proof. — Firstly, note that $(a) \neq (1)$ iff $a \notin A^\times$. We will use the following fact: $a \mid b$ iff $(b) \subseteq (a)$.

For (i), $a \mid bc$ means $bc \in (a)$, and $a \mid b$ or $a \mid c$ means $b \in (a)$ or $c \in (a)$. This shows (i).

For (ii), a principal (b) contains (a) means $a = bc$ for some $c \in A$. Then $(b) = (1)$ means $b \in A^\times$, and $(b) = (a)$ means $c \in A^\times$. This shows (ii).

***Warning* I.1.4.11.** — It is true that prime elements are irreducible: this follows by definition. The converse, however, is generally not true; see example I.1.4.19.

***Definition* I.1.4.12.** — *An ideal $\mathfrak{a}$ of a ring A is **principal** if it can be written as $\mathfrak{a} = (x)$ for some $x \in A$. If A is an integral domain and if all ideals of A are principal, then A is a **principal ideal domain**.*

***Corollary* I.1.4.13.** — *In a principal ideal domain, every non-zero prime ideal is maximal.*

Proof. — This follows immediately from lemma I.1.4.10.

***Example* I.1.4.14.** — (i) The ring $A = k[x]$ for a field k is a principal ideal domain. The prime ideals are precisely given by (f) for irreducible polynomials f or $f = 0$ because of the unique prime factorisation.

(ii) $A = \mathbb{Z}$ is a principal ideal domain, where each ideal is given by (n) for some $n \geq 0$. The prime ideals are the ones for which n is a prime number of $n = 0$. The non-zero prime ideals are maximal: $\mathbb{Z}/(p) \cong \mathbb{F}_p$ for primes p are fields.

(iii) Of course, maximal ideals need not be principal. Consider $A = k[x_1, \dots, x_n]$ for a field k and $n \geq 2$. Then the set $\mathfrak{m}$ of all polynomials with no constant term is a maximal ideal since $\mathfrak{m}$ is the kernel of the evaluation $\mathrm{ev}_0 \colon A \twoheadrightarrow k$, $f \mapsto f(0)$. However, as it turns out, $\mathfrak{m}$ needs at least n generators. On the other hand, principal ideals (f) for irreducible irreducible $f \in A$ are prime, yet not maximal.

(iv) By proposition I.1.4.5, $k[\![x]\!]$ for a field k is a principal ideal domain. The only prime ideals are (x) and (0).

Definition* I.1.4.15.** — *An integral domain A is a* ***unique factorisation domain *if each element $a \in A \setminus (A^\times \cup \{0\})$ admits a prime factorisation $a = \prod_{i=1}^n p_i$ with p_i prime.*

***Observation* I.1.4.16.** — A prime factorisation in an integral domain, if it exists, is always unique in the following sense. Let $a = \prod_{i=1}^n p_i = \prod_{j=1}^m q_j$ with primes p_i and q_j. Then $n = m$ and $(p_i) = (q_{\sigma(i)})$ for some permutation σ.

This is seen as follows. Since p_1 is prime, w. l. o. g. we have $p_1 \mid q_1$. Since q_1 is irreducible, we must have $(p_1) = (q_1)$, i. e., we have $q_1 = up_1$ for a unit u. This yields $\prod_{i=2}^n p_i = u\prod_{j=2}^m q_j$. Note that uq_2 is again prime, so by the same above argument, we have $(p_2) = (uq_2) = (q_2)$. Induction shows the statement.

***Lemma* I.1.4.17.** — *In a unique factorisation domain, irreducible elements are prime.*

Proof. — Let A be a unique factorisation domain, and let $a \in A \setminus (A^\times \cup \{0\})$. Then a has a unique prime factorisation $a = \prod_{i=1}^n p_i$. Since p_1 is prime and since a is irreducible, we have $\prod_{i=2}^n p_i$ a unit. If $n \geq 2$, then $(1) = (\prod_{i=2}^n p_i) \subseteq (p_2) \subsetneq (1)$, a contradiction. Hence $n = 1$ and $a = p_1$ is prime. ∎

***Example* I.1.4.18.** — (i) Principal ideal domains are unique factorisation domains.

(ii) $k[x_1, \dots, x_n]$ for $n \geq 1$ are unique factorisation domains.

***Example* I.1.4.19.** — $A = \mathbb{Z}[\sqrt{-5}]$ is not a unique factorisation domain; we show this by examining $2 \in \mathbb{Z}[\sqrt{-5}]$. First note that 2 is not prime. Indeed, we have $2 \cdot 3 = (1 + \sqrt{-5})(1 + \sqrt{-5})$, but $\frac{1}{2}(1 \pm \sqrt{-5}) \notin A$. Further note that 2 is irreducible. Indeed, suppose that $2 = \alpha\beta$ with $\alpha = a + b\sqrt{-5}$ and $\beta = c + d\sqrt{-5}$. Then $4 = \alpha\bar{\alpha}\beta\bar{\beta} = (a^2 + 5b^2)(c^2 + 5d^2)$ in $\mathbb{Z}$. Since $a^2 + 5b^2 = 2$ has no solutions in $\mathbb{Z}$, w. l. o. g. $a^2 + 5b^2 = 1$. In this case, $\alpha \in A^\times$. Thus A is not a unique factorisation domain by lemma I.1.4.17.

So (2) cannot be maximal in A. Since $A \cong \mathbb{Z}[x]/(x^2 + 5)$, we have

$$A/(2) \cong \mathbb{Z}[x]/(2, x^2 + 5) \cong \mathbb{F}_2[x]/(x^2 - 1) \cong \mathbb{F}_2[x]/(x - 1)^2.$$

Observe that $\mathbb{F}_2[x]/(x - 1)^2$ has a unique maximal ideal $(x - 1)$; it corresponds to the maximal ideal $(\sqrt{-5} - 1)$ in $A/(2)$. Thus the unique maximal ideal in A containing (2) is $(2, \sqrt{-5} - 1)$.

As an application, we prove a very important theorem from field theory.

***Theorem* I.1.4.20** (Artin)**.** — *Every field K admits an algebraic closure $\bar{K}$.*

Proof. — Let Σ be the set of all irreducible monic polynomials $f \in K[x]$. Let $A = K[x_f : f \in \Sigma]$ be the polynomial ring where the x_f are indeterminates. Let $\mathfrak{a} = (f(x_f) : f \in \Sigma)$, which is an ideal of A. Suppose that $\mathfrak{a} = (1)$, say there are $f_1, \dots, f_n \in \Sigma$ and $g_1, \dots, g_n \in A$ such that $\sum_{i=1}^n g_i f_i(x_{f_i}) = 1$.

Let K'/K be a finite field extension such that each f_i splits. Pick for $i = 1, \dots, n$ a root $\alpha_i \in K'$ of f_i. Then we obtain $\sum_{i=1}^{n} g_i f_i(\alpha_i) = 0$ (plugging in arbitrary values for g_i), which is absurd.

Hence $\mathfrak{a} \neq (1)$, and there is a maximal ideal $\mathfrak{m}$ of A containing $\mathfrak{a}$ by corollary I.1.3.13. Then $K_1 = A/\mathfrak{m}$ is a field extension over K such that for any $f \in \Sigma$, there is a root $\bar{x}_f \in K_1$. Now repeat the construction with K_1 instead of K to obtain K_2, etc. By induction, we obtain the field tower $K \subseteq K_1 \subseteq K_2 \subseteq \cdots$, so that $L := \bigcup_{n=1}^{\infty} K_n$ is a field. It has the property that any $f \in \Sigma$ completely splits in linear factors over L. If $\bar{K}$ is the subset of all elements of L which are algebraic over K, then $\bar{K}$ is an algebraic closure of K. ◻

§ I.2. MODULES

Modern commutative algebra has a greater emphasise on modules, which generalise rings and ideals alike. The categories of modules is a so-called *abelian category*, which has very nice properties.

I.2.1. Modules and module homomorphisms.

***Definition* I.2.1.1.** — *Let A be a (commutative) ring. An A-**module** M is a set with a binary operation called* addition $+\colon M \times M \to M$ *and an A-action called* scalar multiplication $\cdot\colon A \times M \to M$ *satisfying the following conditions.*

(i) *The set M forms an abelian group w. r. t. addition. In particular, there is a zero element 0 in M, and each $x \in M$ has an additive inverse $-x \in M$.*

(ii) *Scalar multiplication is an A-linear action. That is for all $a, b \in A$ and $x \in M$, we have the following properties:*

$$a(x + y) = ax + ay, \qquad (a + b)x = ax + bx, \qquad (ab)x = a(bx), \qquad 1x = x.$$

Equivalently, an A-module M is an abelian group together with a ring map $A \to \mathrm{End}_{\mathsf{AbGrp}}(M)$.

***Example* I.2.1.2.** — (i) An ideal $\mathfrak{a}$ of a ring A is an A-module. In particular, $A = (1)$ is an A-module itself.

(ii) If $A = k$ is a field, then the A-modules are precisely the k-vector spaces.

(iii) If $A = k[x]$ with k a field, then the A-modules are the k-vector spaces V together with a linear transformation $x\colon V \to V$.

(iv) Let G be a finite group, and let $A = k[G]$ be the group algebra of G over the field k, i. e., A is a polynomial ring over k with the elements of G as generators where the product of generators is defined as in G. Note that A is not commutative, unless G is abelian. Then the A-modules are precisely the k-representations of G, i. e., the group homomorphisms $G \to \mathrm{GL}(V)$ for k-vector spaces V.

***Definition* I.2.1.3.** — *A map $f\colon M \to N$ between A-modules is an A-**module homomorphism** or is A-**linear** if it satisfies the following conditions.*

(i) *It is additive, i. e., $f(x + y) = f(x) + f(y)$ for all $x, y \in M$. In particular, f is a group homomorphism of abelian groups with $f(-x) = -f(x)$ for all $x \in M$ and $f(0) = 0$.*

(ii) *It commutes with the A-action on M, i. e., $f(ax) = af(x)$ for all $a \in A$ and $x \in M$.*

*If f is bijective, then f is called an A-**module isomorphism**.*

***Convention* I.2.1.4.** — *From now on, we always mean* A-module homomorphisms *when we say* A-module map*.*

***Example* I.2.1.5.** — For $A = k$ a field, then the A-module maps are precisely the k-linear transformations.

***Remark* I.2.1.6.** — Let M be an A-module. If we consider M as an additive abelian group, then the set $\mathrm{End}_{\mathsf{AbGrp}}(M)$ of group endomorphisms on M forms an (in general non-commutative) ring by pointwise addition and composition. In other words, for any $f, g \in \mathrm{End}_{\mathsf{AbGrp}}(M)$, we have $(f+g)(x) := f(x) + g(x)$ and $(gf)(x) := (g \circ f)(x)$ for all $x \in M$.

Let $A \to \mathrm{End}_{\mathsf{AbGrp}}(M)$ be a ring map. If we denote by $a_L : M \to M$ the image of $a \in A$, then for all $a, b \in A$, we obtain $(ab)_L = a_L b_L$, $(a+b)_L = a_L + b_L$ and $1_L = \mathrm{id}_M$. Thus defining an A-module structure on M is the same as giving a ring map $A \to \mathrm{End}_{\mathsf{AbGrp}}(M)$. The action of a on M is the same as the map a_L on M, i. e., $a_L : M \to M$ is given by $m \mapsto am$.

Let $f : M \to M$ be a map. Then f is A-linear iff $f \in \mathrm{End}_{\mathsf{AbGrp}}(M)$ and if f commutes with a_L for all $a \in A$. Since A is commutative, we have $a_L b_L = b_L a_L$ for all $a, b \in A$; thus a_L is itself A-linear.

***Notation* I.2.1.7.** — Let M be an A-module. We denote $a_L : M \to M$, $m \mapsto am$ simply as $a = a_L$ in the future.

***Fact* I.2.1.8.** — *The composition of A-module maps is again again A-linear.*

***Construction* I.2.1.9.** — Let $\mathrm{Hom}_A(M, N)$ denote the set of all A-module homomorphisms $M \to N$. This can be naturally turned into a A-module by defining addition and scalar multiplication pointwise: for all $f, g \in \mathrm{Hom}_A(M, N)$ and all $a \in A$, we have

$$(f+g)(x) := f(x) + g(x), \qquad (af)(x) := af(x)$$

for all $x \in M$. We will also write $\mathrm{Hom}(M, N) := \mathrm{Hom}_{A\text{-}\mathsf{Mod}}(M, N) := \mathrm{Hom}_A(M, N)$.

Let $u : M' \to M$ and $v : N \to N'$ be A-module maps. They induce A-module maps

$$\begin{aligned} u^* &: \mathrm{Hom}(M, N) \longrightarrow \mathrm{Hom}(M', N), \qquad f \longmapsto f \circ u; \\ v_* &: \mathrm{Hom}(M, N) \longrightarrow \mathrm{Hom}(M, N'), \qquad f \longmapsto v \circ f \end{aligned}$$

by pre- and post-composition resp.

***Fact* I.2.1.10.** — *For any A-module M, there is a natural isomorphism*

$$\mathrm{Hom}(A, M) \xrightarrow{\sim} M, \qquad f \longmapsto f(1).$$

***Remark* I.2.1.11.** — Let M be a B-module, and let $f : A \to B$ be a ring map. Then M is naturally an A-module with $am := f(a)m$ for all $a \in A$ and $m \in M$. Equivalently, if the B-module structure on M is given by the ring map $B \to \mathrm{End}_{\mathsf{AbGrp}}(M)$, then we obtain an A-module structure through $A \to B \to \mathrm{End}_{\mathsf{AbGrp}}(M)$.

I.2.2. Operations on modules.

***Definition* I.2.2.1.** — *A **submodule** M' of an A-module M is a subset satisfying the following properties.*

(i) *The set M' is an additive subgroup of M.*
(ii) *It is closed under scalar multiplication, i. e., $AM' \subseteq M'$.*

Then N is again an A-module. It comes with an (injective) canonical inclusion $\iota : M' \hookrightarrow M$.

If S is a subset of M, then

$$(S) := \bigcap_{\text{submodules } N \supseteq S} N = \left\{ \sum_{i=1}^{n} x_i s_i \;\middle|\; x_i \in A, s_i \in S \right\}$$

denotes the smallest submodule of M containing S. If $S = \{s_1, \dots, s_n\}$ is finite, then we also write $(S) = (s_1, \dots, s_n) = \sum_{i=1}^{n} A s_i$. We say that S is the submodule generated *by S.*

***Example* I.2.2.2.** — Ideals of a ring A are A-submodules of A.

***Fact* I.2.2.3.** — *Let $(M_i)_{i\in I}$ be a family of submodules of an A-module M. Then the following sets are again submodules of M.*

(i) *The* ***sum***

$$\sum_{i\in I} M_i := \left\{ \sum_{i\in I} x_i \;\middle|\; x_i \in M_i, x_i = 0 \textit{ for almost all } i \in I \right\},$$

which is the smallest ideal containing all M_i.

(ii) *The intersection $\bigcap_{i\in I} M_i$.*

(iii) *If $\mathfrak{a}$ is an ideal of A, then the product*

$$\mathfrak{a}M := \left\{ \textit{finite sums } \sum a_i x_i \;\middle|\; a_i \in \mathfrak{a}, x_i \in M \right\}.$$

In particular, the submodules of M form a complete lattice w. r. t. inclusion, i. e., infima and suprema always exist.

Proof. — Trivial calculation. □

Definition* I.2.2.4.** — *Let N' be a submodule of an A-module M. Then the abelian quotient group M/M' inherits an A-module structure from M, namely $a(x + M') = ax + M'$ for all $a \in A$ and $x \in M$. We call M/M' the* ***quotient module *or just* ***quotient*** *of M by M'. It comes with a (surjective) canonical projection $\pi\colon M \twoheadrightarrow M/M'$, $x \mapsto x + M'$.*

As in the case of ideals, we usually denote the image of an element $x \in M$ in M/M' by $\bar{x}$. Moreover, we write $x \equiv y \mod M'$ if $\bar{x} = \bar{y}$ or equivalently if $x - y \in M'$.

***Proposition* I.2.2.5** (submodule correspondence, cf. proposition I.1.2.4)**.** — *Let M be an A-module and let M' be a submodule. Let $\pi\colon M \twoheadrightarrow M/M'$ be the canonical projection. Then there is a one-to-one, inclusion preserving correspondence*

$$\begin{aligned} \{\textit{submodules } N \textit{ of } M \textit{ containing } M'\} &\longleftrightarrow \{\textit{submodules } \bar{N} \textit{ of } M/M'\}, \\ N &\longmapsto \pi(N) = N/M', \\ \pi^{-1}(\bar{N}) &\longleftarrow\!\shortmid \bar{N}. \end{aligned}$$

Moreover, if N is any submodule of M, then $\pi(N) = (N + M')/M'$ is a submodule of M/M'.

Definition* I.2.2.6.** — *Let $f\colon M \to N$ be an A-module map. The* ***kernel *of f is the submodule $\ker f := f^{-1}(0)$ of M. The* ***image*** *of f is the submodule $\operatorname{im} f := f(M)$ of N. The* ***cokernel*** *of f is the quotient module $\operatorname{coker} f := N/\operatorname{im} f$ of N.*

***Remark* I.2.2.7.** — (i) As with groups, an A-module map $f\colon M \to N$ is injective iff $\ker f = 0$ and surjective iff $\operatorname{im} f = N$.

(ii) In general, preimages of submodules are again submodules. Let $f\colon M \to N$ be an A-module map, and let N' be a submodule of N. Then the kernel of the composition $M \to N \twoheadrightarrow N/N'$ is $f^{-1}(N')$ and hence a submodule of M.

***Homomorphism theorem* I.2.2.8.** — *Let $f\colon M \to N$ be an A-module map, and let $M' \subseteq \ker f$ be a submodule of M. Then f induces a unique A-linear map*

$$\bar{f}\colon M/M' \longrightarrow N, \qquad \bar{x} \longmapsto f(x)$$

with kernel $\ker f/M'$ *such that the following diagram commutes:*

$$\begin{tikzcd} M \arrow[r, two heads] \arrow[d, "f"'] & M/M' \arrow[dl, dashed, "\bar{f}"] \\ N. & \end{tikzcd}$$

(This property is also called the ***universal property of quotient modules****.) Moreover, there exists an induced canonical isomorphism*

$$M/\ker f \xrightarrow{\sim} \operatorname{im} f.$$

Proof. — One easily checks that $\bar{f}$ is A-linear, that it is unique such that the diagram commutes and that $\ker \bar{f} = \ker f/M'$. It is also well-defined: for $x, y \in M$ with $\bar{x} = \bar{y}$, we have $f(\bar{x}) = f(\bar{y})$ since $M' \subseteq \ker f$. ☐

***First isomorphism theorem* I.2.2.9.** — *Let M be a module with submodules M_1 and M_2. Then there is a canonical isomorphism*

$$M_2/(M_1 \cap M_2) \xrightarrow{\sim} (M_1 + M_2)/M_1.$$

Proof. — Consider the composite $M_2 \hookrightarrow M_1 + M_2 \twoheadrightarrow (M_1 + M_2)/M_1$, which is surjective with kernel $M_1 \cap M_2$. The statement follows from the homomorphism theorem I.2.2.8. ☐

***Second isomorphism theorem* I.2.2.10.** — *Let $L \supseteq M \supseteq N$ be three A-modules. Then there is a canonical isomorphism*

$$(L/N)/(M/N) \xrightarrow{\sim} L/M.$$

Proof. — Consider the map $\bmod M \colon L/N \to L/M$, $\bar{x} \mapsto \bar{x} + M$ obtained by reduction modulo M. One easily checks that $\bmod M$ is a well-defined, surjective A-module map whose kernel is M/N. The statement follows from the homomorphism theorem I.2.2.8. ☐

***Definition* I.2.2.11.** — *Let N and P be submodules of an A-module M. Define*

$$(N : P) := \{a \in A \mid aP \subseteq N\},$$

which is an ideal of A. If $P = (x)$, then we abbreviate $(N : x) := (N : (x))$. Similarly for an ideal $\mathfrak{a}$ of A, define

$$(N : \mathfrak{a}) := \{x \in M \mid \mathfrak{a}x \subseteq N\},$$

which is a submodule of M. Note that in the case where $N = \mathfrak{a}$ and $P = \mathfrak{b}$ are ideals of $M = A$, both notations coincide. We call $(\mathfrak{a} : \mathfrak{b}) = \{x \in A \mid x\mathfrak{b} \subseteq \mathfrak{a}\}$ also the ***ideal quotient*** *of $\mathfrak{a}$ by $\mathfrak{b}$.*

In particular,

$$\operatorname{ann} M := (0 : M) = \{a \in A \mid aM = 0\}$$

is the ***annihilator*** *of M. We say that M is* ***faithful*** *if $\operatorname{ann} M = 0$. For $x \in M$, we write $\operatorname{ann} x := (0 : x)$.*

***Example* I.2.2.12.** — (i) Let A be a ring. Then the set of all zero divisors is $\bigcup_{x \in A \setminus \{0\}} \operatorname{ann} x$.

(ii) Consider $A = \mathbb{Z}$ together with ideals $\mathfrak{a} = (m)$ and $\mathfrak{b} = (n)$. Say $m = \prod_p p^{m_p}$ and $n = \prod_p p^{n_p}$. Then $(\mathfrak{a} : \mathfrak{b}) = (q)$ with $q = \prod_p p^{q_p}$ with $q_p = \max(m_p - n_p, 0) = m_p - \min(m_p, n_p)$.

***Remark* I.2.2.13.** — Let M be an A-module. For any ideal $\mathfrak{a} \subseteq \operatorname{ann} M$, we can consider M as an $A/\mathfrak{a}$-module via $\bar{x}m := xm$ for all $\bar{x} \in A/\mathfrak{a}$ and $m \in M$. This is well-defined since for any $x \in \mathfrak{a}$, we have $xM \subseteq \mathfrak{a}M = 0$. In particular, M is faithful over $A/\operatorname{ann} M$.

***Fact* I.2.2.14** (laws for ideal quotients)**.** — *Let $\mathfrak{a}, \mathfrak{a}_i, \mathfrak{b}, \mathfrak{b}_i, \mathfrak{c}$ be ideals of a ring A for all $i \in I$. Then the following hold:*

(i) $\mathfrak{a} \subseteq (\mathfrak{a} : \mathfrak{b})$.
(ii) $(\mathfrak{a} : \mathfrak{b})\mathfrak{b} \subseteq \mathfrak{a}$.
(iii) $((\mathfrak{a} : \mathfrak{b}) : \mathfrak{c}) = (\mathfrak{a} : \mathfrak{b}\mathfrak{c}) = ((\mathfrak{a} : \mathfrak{c}) : \mathfrak{b})$.
(iv) $(\bigcap_{i\in I} \mathfrak{a}_i : \mathfrak{b}) = \bigcap_{i\in I}(\mathfrak{a}_i : \mathfrak{b})$.
(v) $(\mathfrak{a} : \sum_{i\in I} \mathfrak{b}_i) = \bigcap_{i\in I}(\mathfrak{a} : \mathfrak{b}_i)$.

Proof. — This follows after explicit calculation.

***Fact* I.2.2.15** (laws for annihilators)**.** — *Let N, P be submodules of some module. Then the following hold:*

(i) $\operatorname{ann}(N + P) = \operatorname{ann} N \cap \operatorname{ann} P$.
(ii) $(N : P) = \operatorname{ann}((N + P)/N)$.

Proof. — This follows directly from the definitions.

***Construction* I.2.2.16.** — Let $(M_i)_{i\in I}$ be a family of A-modules. The **direct sum** of the M_i is

$$\bigoplus_{i\in I} M_i := \{(x_i)_{i\in I} \mid x_i = 0 \text{ for almost all } i \in I\}.$$

It carries an A-module structure via componentwise addition and scalar multiplication. It comes with (injective) canonical inclusions $\iota_i \colon M_i \hookrightarrow \bigoplus_{i\in I} M_i$ for all $i \in I$.

***Construction* I.2.2.17.** — Let $(M_i)_{i\in I}$ be a family of A-modules. The **direct sum** of the M_i is the Cartesian product $\prod_{i\in I} M_i$. It carries an A-module structure via componentwise addition and scalar multiplication. It comes with (surjective) canonical projections $\pi_i \colon \prod_{i\in I} M_i \twoheadrightarrow M_i$ for all $i \in I$.

***Warning* I.2.2.18.** — If I is finite, then $\bigoplus_{i\in I} = \prod_{i\in I}$. However, we only have $\bigoplus_{i\in I} \subseteq \prod_{i\in I}$ in general.

***Observation* I.2.2.19.** — Let $A = \prod_{i=1}^n A_i$ be a direct product of rings. Then

$$\mathfrak{a}_i := 0 \times \cdots \times 0 \times A_i \times 0 \times \cdots \times 0$$

is an ideal of A. Notice that $\mathfrak{a}_i$ is not a subring as it does not contain the identity element of A (except when $n = 1$). Then as A-modules, we have the decomposition $A = \bigoplus_{i=1}^n \mathfrak{a}_i$.

Conversely, let A be a ring. Suppose that as A-modules, we have a decomposition $A = \bigoplus_{i=1}^n \mathfrak{a}_i$ of A into ideals. Note that $\mathfrak{b}_i := \bigoplus_{j\neq i} \mathfrak{a}_j$ for all i are again an ideal of A and that they are pairwise coprime. Thus by the Chinese remainder theorem I.1.2.16, we obtain the isomorphism $A \cong \prod_{i=1}^n (A/\mathfrak{b}_i)$, i. e., A is a direct product of rings. In fact, $\mathfrak{a}_i \cong A/\mathfrak{b}_i$ as A-modules, and if e_i is the image in $\mathfrak{a}_i$ of $1 \in A/\mathfrak{b}_i$, then e_i is an idempotent in A and $\mathfrak{a}_i = Ae_i$.

I.2.3. Nakayama's lemma.

***Definition* I.2.3.1.** — (i) *An A-module M is **free** if there exists an isomorphism $M \cong \bigoplus_{i\in I} A =: A^{(I)}$. In particular, M admits an A-linearly independent **basis** of cardinality $\#I$, and we say that M has **rank** $\operatorname{rk} M := \#I$. If $\#I = n$ is finite, then we also write $A^n := A^{(I)}$. By convention, $A^0 := 0$.*

(ii) *An A-module M is **finitely generated**, **finite**, or of **finite type**, if M is generated by a finite set, i. e., $M = (x_1, \dots, x_n)$.*

***Fact* I.2.3.2.** — *If $A^n \cong A^m$, then $m = n$.*

Proof. — For any maximal ideal $\mathfrak{m}$ of A, there is a surjective map $A^n \twoheadrightarrow (A/\mathfrak{m})^n$ with kernel $\mathfrak{m}A^n$, whence $A^n/\mathfrak{m}A^n \cong (A/\mathfrak{m})^n$. Passing to quotients, the hypothesised isomorphism becomes $(A/\mathfrak{m})^n \cong (A/\mathfrak{m})^m$ of $A/\mathfrak{m}$-vector spaces. We obtain $n = m$.

***Definition* I.2.3.3.** — *Let M be an A-module. A generating set S of M is* minimal *if removing any element from S does not generate M anymore.*

***Fact* I.2.3.4** (presentation of modules). — *Any A-module M is isomorphic to a quotient of a free module. In particular, M is finitely generated iff M is isomorphic to a quotient of A^n for some $n > 0$.*

Proof. — Since we always have $M = (M)$, there is a surjective A-linear map

$$\pi\colon A^{(M)} \longrightarrow\!\!\!\!\!\rightarrow M, \qquad e_m \longmapsto m$$

where $e_m \in \bigoplus_{m\in M} A$ is the tuple with entry 1 at index m and zero elsewhere. Then $M \cong A^{(M)}/\ker\pi$.

If $M = (x_1, \dots, x_n)$ is finitely generated, then we can replace $A^{(M)}$ with A^n and map $e_{x_i} = e_i$ to x_i. Conversely, if $M \cong A^n/N$ for some submodule N of A^n, then we obtain a surjective A-linear map $A^n \twoheadrightarrow A^n/N \cong M$.

***Lemma* I.2.3.5** (determinant trick). — *Let M be a finitely generated A-module, let $\mathfrak{a}$ be an ideal of A, and let $\phi\colon M \to M$ be A-linear such that $\operatorname{im}\phi \subseteq \mathfrak{a}M$. Then ϕ satisfies an equation $\phi^n + a_1\phi^{n-1} + \dots + a_n = 0$ with $a_i \in \mathfrak{a}$.*

Proof. — Let $M = (x_1, \dots, x_n)$. By hypothesis, we have $\phi(x_i) \in \mathfrak{a}M$ for all i, say $\phi(x_i) = \sum_{j=1}^n a_{ij}x_j$ with $a_{ij} \in \mathfrak{a}$. If δ_{ij} denotes the Kronecker delta, we can reformulate the equation as $\sum_{j=1}^n (\delta_{ij}\phi - a_{ij})x_j = 0$ for all i. Now consider the matrix $X := (\delta_{ij}\phi - a_{ij})_{ij} \in \operatorname{Mat}_n(\operatorname{End}(M))$. Recall that the *adjugate matrix* $\operatorname{adj} X$ of X satisfies $(\operatorname{adj} X)\cdot X = \det X \cdot \mathrm{id}$. Since $X(x_1, \dots, x_n)^\top = 0$, we obtain $(\det X)x_i = 0$ for all i and hence $(\det X)(M) = 0$. Expanding $\det X$ gives the desired equation.

***Corollary* I.2.3.6.** — *Let M be a finitely generated A-module, and let $\mathfrak{a}$ be an ideal of A such that $\mathfrak{a}M = M$. Then there exists some $x \in A$ with $x \equiv 1 \mod \mathfrak{a}$ such that $xM = 0$.*

Proof. — Take $\phi = \mathrm{id}_M$, so that $\operatorname{im}\phi = M = \mathfrak{a}M$. Thus, $x\mathrm{id}_M = 0$ where $x := 1 + a_1 + \dots + a_n \equiv 1 \mod \mathfrak{a}$.

As a nice by-product, we obtain the Cayley-Hamilton theorem for rings.

***Cayley-Hamilton theorem* I.2.3.7.** — *Let M be a free A-module of finite rank, and let $\phi\colon M \to M$ be an A-linear endomorphism. After a choice of basis for M, the map ϕ satisfies its own characteristic polynomial* $\det(\mathrm{id}_M T - \phi) \in A[T]$.

Proof. — In the proof of lemma I.2.3.5, the constructed polynomial $\det(\delta_{ij}T - a_{ij})_{ij}$ which is satisfied by ϕ, is the characteristic polynomial of the matrix $(a_{ij})_{ij}$. After a choice of basis, we have $\phi = (a_{ij})_{ij} \in \operatorname{Mat}_n(A)$. The result follows.

***Nakayama's lemma* I.2.3.8** (Nakayama, Krull, Azumaya). — *Let M be a finitely generated A-module, and let $\mathfrak{a} \subseteq \operatorname{rad} A$ be an ideal of A. If $\mathfrak{a}M = M$, then $M = 0$.*

First proof. — By corollary I.2.3.6, there exists some $x \in A$ with $x \equiv 1 \mod \operatorname{rad} A$ such that $xM = 0$. Then $x \in A^\times$ by proposition I.1.3.23. Thus $M = x^{-1}xM = 0$.

Second proof. — This proof does not depend on lemma I.2.3.5. By way of contradiction, assume that $M \neq 0$. Let $x_1, \dots, x_n$ be a minimal generating set for M. Since $x_n \in M = \mathfrak{a}M$, write $x_n = \sum_{i=1}^{n} a_i x_i$ with $a_i \in \mathfrak{a}$. Hence $(1 - a_n)x_n = \sum_{i=1}^{n-1} a_i x_i$. Since $a_n \in \operatorname{rad} A$, proposition I.1.3.23 implies that $1 - a_n \in A^\times$. In other words, $x_n \in (x_1, \dots, x_{n-1}) \subsetneq M$, a contradiction.

***Corollary* I.2.3.9.** — *Let M be an A-module, let N be a submodule of M such that M/N is finitely generated, and let $\mathfrak{a} \subseteq \operatorname{rad} A$ be an ideal of A. If $M = \mathfrak{a}M + N$, then $M = N$.*

Proof. — Observe that $\mathfrak{a}(M/N) = (\mathfrak{a}M + N)/N = M/N$. Now apply Nakayama's lemma I.2.3.8 to M/N so that $M/N = 0$, i. e., $M = N$.

***Corollary* I.2.3.10.** — *Let $\phi\colon M \to N$ be an A-module map such that N is finite, and let $\mathfrak{a} \subseteq \operatorname{rad} A$ be an ideal of A. If the induced map $\bar{\phi}\colon M/\mathfrak{a}M \to N/\mathfrak{a}N$ is surjective, then so is ϕ.*

Proof. — For any $x \in N$, there is some $y \in M$ such that $\bar{\phi}(\bar{y}) = \bar{x}$ by hypothesis, i. e., $\phi(y) \equiv x \mod \mathfrak{a}N$. It follows that $N = \mathfrak{a}N + \operatorname{im} \phi$. Since $N/\operatorname{im} \phi$ is finite, corollary I.2.3.9 implies that $N = \operatorname{im} \phi$.

***Proposition* I.2.3.11** (Vasconcelos 1969). — *Let M be a finitely generated A-module, and let $\phi\colon M \to M$ be A-linear. If ϕ is surjective, then ϕ is already an automorphism.*

Proof. — Since ϕ is A-linear, we can view M as an $A[\phi]$-module, given by $\phi x := \phi(x)$ for all $x \in M$. Since $\phi M = M$, by corollary I.2.3.6, there exists some $\psi \in A[\phi]$ such that $(1 + \phi\psi)M = 0$. Hence for all $x \in \ker \phi$, we have $0 = (1 + \phi\psi)(x) = x + \psi(\phi(x)) = x$. Thus $\ker \phi = 0$ and ϕ is injective.

***Warning* I.2.3.12.** — Two minimal generating sets of an A-module might not have the same number of elements. In this regard, modules work very differently from vector spaces. For example, take $M = A$ a ring such that there are $x, y \notin A^\times$ satisfying $x + y = 1$. Then $M = (1) = (x, y)$, but $(x) \neq M \neq (y)$. However, over local rings A, we have similar statements as with vector spaces.

***Proposition* I.2.3.13.** — *Let M be a finite module over a local ring $(A, \mathfrak{m}, k)$. Then $\bar{M} := M/\mathfrak{m}M$ is a finite-dimensional k-vector space.*

(i) *Any lift $\{x_1, \dots, x_n\}$ of a k-basis $\{\bar{x}_1, \dots, \bar{x}_n\}$ of $\bar{M}$ is a minimal generating set of M.*

(ii) *Any minimal generating set of M arises in such a way. In particular, they all have the same cardinality n.*

(iii) *Any two minimal generating sets of M can be transformed into another by an A-linear base change. That is, if $M = (x_1, \dots, x_n) = (y_1, \dots, y_n)$, write $y_i = \sum_{j=1}^{n} a_{ij} x_j$ for all i with $a_{ij} \in A$. Then $(a_{ij})_{ij}$ is in invertible matrix with entries in A.*

Proof. — The idea is to reduce everything to the k-vector space $\bar{M}$.

For (i), let $N = (x_1, \dots, x_n)$. Since $N \hookrightarrow M \twoheadrightarrow \bar{M}$ is surjective, we have $N + \mathfrak{m}M = M$. As M is finitely generated, so is M/N. Therefore, corollary I.2.3.9 implies that $M = N$. If $\{x_1, \dots, x_n\}$ would not be minimal, say $M = (x_2, \dots, x_n)$, then $\{\bar{x}_2, \dots, \bar{x}_n\}$ would generate the vector space $\bar{M}$, a contradiction.

For (ii), suppose that $S \coloneqq \{x_1, \dots, x_m\}$ would be a minimal generating set of M; then $\bar{S} \coloneqq \{\bar{x}_1, \dots, \bar{x}_m\}$ generates $\bar{M}$. Moreover, this generating set of $\bar{M}$ is a k-basis, for otherwise, a proper subset of $\bar{S}$ lifts to a generating set of M which is a proper subset of S, contradicting minimality. Hence $m = n$.

For (iii), we have $\bar{y}_i = \sum_{j=1}^{n} \bar{a}_{ij}\bar{x}_j$ in $\bar{M}$ with $\bar{a}_{ij} \in k$. Then $\det(\bar{a}_{ij})_{ij} = \det(a_{ij})_{ij} \bmod \mathfrak{m} \in k^\times$. By proposition I.1.3.25, we have $\det(a_{ij})_{ij} \in A^\times$. Now Cramer's rule gives an explicit formula for the inverse of $(a_{ij})_{ij}$ in $\mathrm{Mat}_n(A)$.

I.2.4. Kaplansky's theorem on projective modules.

***Definition* I.2.4.1.** — *An A-module P is **projective** if it satisfies the following lifting property. For any A-linear surjection $\pi\colon M \twoheadrightarrow N$ and any A-linear map $f\colon P \to N$, there exists a lifting $\tilde{f}\colon P \to M$ such that the following diagram commutes:*

$$\begin{tikzcd} & M \arrow[d, "\pi"] \\ P \arrow[ur, dashed, "\tilde{f}"] \arrow[r, "f"] & N. \end{tikzcd}$$

***Warning* I.2.4.2.** — This is *not* a universal property since $\bar{f}$ might not be unique.

***Proposition* I.2.4.3** (characterisation of projective modules)**.** — *An A-module P is projective iff it is a direct summand of a free A-module.*

Proof. — Let P be projective, and let $\pi\colon F = A^{(P)} \twoheadrightarrow P$ be the obvious A-module map. Then the identity map $f = \mathrm{id}_P\colon P \to P$ admits a lift $\tilde{f}\colon P \to F$ such that $\pi \circ \tilde{f} = \mathrm{id}_P$. Hence $\tilde{f}$ is injective, and we can write $F = \operatorname{im} \tilde{f} \oplus \ker \pi \cong P \oplus \ker \pi$.

Conversely, suppose that $F = P \oplus Q$ is a free A-module. Let $\pi\colon M \twoheadrightarrow N$ be an A-linear surjection, and let $f\colon P \to N$ be any A-linear map. Consider the A-module maps $\pi' = \pi \oplus \mathrm{id}_Q$ and $f' = f \oplus \mathrm{id}_Q$. Then there exists a lift $\tilde{f}'$ of f' such that the following diagram commutes:

$$\begin{tikzcd} & M \oplus Q \arrow[d, two heads, "\pi'"] \\ F \arrow[ur, dashed, "\tilde{f}'"] \arrow[r, "f'"] & N \oplus Q. \end{tikzcd}$$

We can do this as follows: let $(e_i)_{i \in I}$ be a free basis of F, i. e., $F = \bigoplus_{i \in I} Ae_i$. For each $i \in I$, pick any $x_i \in (\pi')^{-1}(f'(e_i))$ and define $\tilde{f}'(e_i) \coloneqq x_i$. Finally, taking the whole diagram modulo Q gives gives the lift $\tilde{f}$ of f.

***Example* I.2.4.4.** — Free modules are always projective.

***Kaplansky's theorem* I.2.4.5** (1958)**.** — *Projective modules over a local ring $(A, \mathfrak{m})$ are free.*

Proof for finite modules. — Let M be a finitely generated, projective A-module. By proposition I.2.3.13, there exists a minimal generating set $\{x_1, \dots, x_n\}$ of M. Let $\phi\colon F = A^n \twoheadrightarrow M$ be the natural map which sends the standard basis elements of F to the x_i. Observe that if $\sum_{i=1}^{n} a_i x_i = 0$ for $a_i \in A$, then after passing to the $A/\mathfrak{m}$-vector space $M/\mathfrak{m}M$, we have $a_i \in \mathfrak{m}$ for all i. Hence $K \coloneqq \ker \phi \subseteq \mathfrak{m}F$.

Since M is projective, we can lift id_M along ϕ to obtain an A-linear map $\psi\colon M \to F$ such that $F = \operatorname{im} \psi \oplus K$ with $\operatorname{im} \psi \cong M$. Then $K \subseteq \mathfrak{m}F = \mathfrak{m} \operatorname{im} \psi \oplus \mathfrak{m}K$ and hence $K \subseteq \mathfrak{m}K$ since

$\mathfrak{m} \operatorname{im} \psi \cap K \subseteq \operatorname{im} \psi \cap K = 0$. It follows that $K = \mathfrak{m} K$. On the other hand, $K \cong F / \operatorname{im} \psi$ is finitely generated since F is, and by Nakayama's lemma I.2.3.8, $K = 0$. Thus $F = \operatorname{im} \psi \cong M$ is free. ∎

To obtain a more general proof, we need to do considerably more.

***Lemma* I.2.4.6.** — *Let R be any (non-commutative) ring, and let F be a direct sum of countably generated R-modules. If M is an arbitrary direct summand of F, then M is also a direct sum of countably generated R-modules.*

Proof. — Write $F = \bigoplus_{\omega \in \Omega} E_\omega$ for countably generated modules E_α, and suppose that $F = M \oplus N$. By transfinite induction, we construct a well-ordered family $(F_\alpha)_\alpha$ of submodules of F satisfying the following properties:

(i) $F_\alpha \subsetneq F_\beta$ whenever $\alpha < \beta$.
(ii) $F = \bigcup_\alpha F_\alpha$.
(iii) If α is a limiting ordinal, then $F_\alpha = \bigcup_{\beta < \alpha} F_\beta$.
(iv) $F_{\alpha+1}/F_\alpha$ is countably generated.
(v) $F_\alpha = M_\alpha \oplus N_\alpha$ where $M_\alpha = M \cap F_\alpha$ and $N_\alpha = N \cap F_\alpha$.
(vi) $F_\alpha = \bigoplus_{\omega \in \Omega'} E_\omega$ for a suitable subset $\Omega' \subseteq \Omega$.

Let us construct such $\{F_\alpha\}_\alpha$. Set $F_0 = 0$. By induction hypothesis, suppose that for an ordinal α, all F_β with ordinals $\beta < \alpha$ have been constructed. If α is a limiting ordinal, take $F_\alpha := \bigcup_{\beta < \alpha} F_\beta$.

If $\alpha = \beta + 1$ is a successor ordinal, let Q_1 be any of the E_ω not contained in F_β (if $F = F_\beta$, then the construction stops as Ω was finite). Suppose that $Q_1 = (x_{11}, x_{12}, \dots)$, and decompose $x_{11} \in F = M \oplus N$ into its M- and N-components. Each component can be written as a finite sum in $F = \bigoplus_\omega E_\omega$; let Q_2 be the direct sum of the finitely many E_ω necessary to write these two sums. Suppose that $Q_2 = (x_{21}, x_{22}, \dots)$.

Now repeat the construction with $x_{12} \in Q_1$ in order to obtain $Q_3 = (x_{31}, x_{32}, \dots)$. Then repeat the construction with $x_{21} \in Q_2$ in order to obtain Q_4. We do this construction with all x_{ij} in the order $x_{11}, x_{12}, x_{21}, x_{31}, x_{22}, x_{13}, \dots$, similar to Cantor's first diagonal argument. Define $F_\alpha = F_\beta \oplus (x_{11}, x_{12}, x_{21}, \dots)$, which is a submodule of F. On checks that this construction satisfies our desired properties.

Now observe that $M = \bigcup_\alpha M_\alpha$ (property (ii)) where each M_α is a direct summand of F (properties (v) and (vi)).

Moreover for successor ordinals, $M_\alpha \subsetneq M_{\alpha+1}$ (property (i)) is a direct summand of $M_{\alpha+1}$ (properties (v) and (vi)). We have the same thing for N and N_α, and so $F_{\alpha+1}/F_\alpha = (M_{\alpha+1}/M_\alpha) \oplus (N_{\alpha+1}/N_\alpha)$, which is countably generated (property (iv)). Hence $M_{\alpha+1}/M_\alpha$ is also countably generated, and thus $M_{\alpha+1} = M_\alpha \oplus M'_{\alpha+1}$ with $M'_{\alpha+1}$ countably generated.

On the other hand, for limit ordinals α, we have $M_\alpha = \bigcup_{\beta < \alpha} M_\beta$ (property (iii)). In this case, set $M'_\alpha = 0$. Thus we obtain $M = \bigcup_\alpha M_\alpha = \bigoplus_\alpha M'_\alpha$ with M'_α countably generated. ∎

***Lemma* I.2.4.7.** — *Let M be a projective module over a local ring $(A, \mathfrak{m})$, and let $x \in M$. Then there exists a free direct summand of M containing x.*

Proof. — By proposition I.2.4.3, write $F = M \oplus N$ for a free A-module F. For any basis of F, we can express x as a sum with finitely many non-zero coefficients. Choose any basis $B = \{x_i\}_{i \in I}$ of F such that the number of non-zero coefficients to express x over B is minimal. Say $x = \sum_{i=1}^n a_i x_i$ with $a_i \in A \setminus \{0\}$. Then

$$a_i \notin \sum_{j \neq i} (a_j) \qquad \text{for all } i. \tag{I.2.4.7.1}$$

For example, if we would have $a_n = \sum_{i=1}^{n-1} b_i a_i$ with $b_i \in A$, then $x = \sum_{i=1}^{n-1} a_i (x_i + b x_n)$, contradicting the choice of B.

Now decompose $x_i = y_i + z_i$ with $y_i \in M$ and $z_i \in N$. Then $x = \sum_{i=1}^n a_i x_i = \sum_{i=1}^n a_i y_i \in M$ Further decompose $y_i = \sum_{j=1}^n c_{ij} x_j + t_i$ with $t_i \in (B \setminus \{x_1, \dots, x_n\})$ and $c_{ij} \in A$. Plugging this into the decomposition of x and comparing coefficients of the x_i yields $a_i = \sum_{j=1}^n a_j c_{ji}$ (and $t_i = 0$). If for any i, we would have $c_{ji} \in A^\times$ for $j \neq i$ or $1 - c_{ii} \in A^\times$, then rearranging would contradict equation (I.2.4.7.1). Hence $c_{ij}, 1 - c_{ii} \in \mathfrak{m} = A \setminus A^\times$ for all $i \neq j$.

By the Leibniz formula, we have $\det(c_{ij})_{ij} \equiv \prod_{i=1}^n c_{ii} \equiv 1 \mod \mathfrak{m}$, so $(c_{ij})_{ij}$ is a base change. In other words, we may replace $\{x_1, \dots, x_n\} \subseteq B$ by $\{y_1, \dots, y_n\} \subseteq M$ and still have a basis of F. Therefore, $M' \coloneqq (y_1, \dots, y_n) \subseteq M$ is a free direct summand of F and thus also of M, and M' contains x. ∎

Proof of Kaplansky's theorem I.2.4.5. — Let M be a projective module over a local ring $(A, \mathfrak{m})$. By proposition I.2.4.3, it is a direct summand of a free module. By lemma I.2.4.6, it suffices to show that each countably generated (projective) direct summands of M is free. So w. l. o. g., let $M = (x_1, x_2, \dots)$ be countably generated.

By lemma I.2.4.7, we can write $M = F_1 \oplus M_1$ with F_1 free and $x_1 \in F_1$. Then M_1 is again projective. Let x_2' be the M_1-component of $x_2 \in F_1 \oplus M_1$. Again by lemma I.2.4.7, we can write $M_1 = F_2 \oplus M_2$ with F_2 free and $x_2' \in F_2$. Then M_2 is again projective. Let x_3' be the M_2-component of $x_3 \in F_1 \oplus F_2 \oplus M_2$ etc. By induction, we obtain $M = F_1 \oplus F_2 \oplus \cdots$, which is free. ∎

I.2.5. Exact sequences.

***Definition* I.2.5.1.** — *A (countable) sequence A-modules and A-module maps*

$$\cdots \longrightarrow M_{i-1} \xrightarrow{f_i} M_i \xrightarrow{f_{i+1}} M_{i+1} \longrightarrow \cdots \tag{I.2.5.1.1}$$

*is **exact** at M_i if $\operatorname{im} f_i = \ker f_{i+1}$. The whole sequence is **exact** if it is exact at each M_i.*

***Fact* I.2.5.2.** — (i) *The sequence $0 \to M' \xrightarrow{f} M$ is exact iff f is injective.*

(ii) *The sequence $M \xrightarrow{f} M'' \to 0$ is exact iff f is surjective.*

(iii) *The sequence $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ is exact iff f is injective, g is surjective, and g induces an isomorphism $\operatorname{coker} f \xrightarrow{\sim} M''$. Such a sequence is called a **short exact sequence**.*

Proof. — Directly by definition. ∎

***Remark* I.2.5.3.** — Any long exact sequence (I.2.5.1.1) can be split up into a bunch of short exact sequences:

$$0 \longrightarrow \ker f_{i+1} \hookrightarrow M_i \xrightarrow{f_{i+1}} \operatorname{im} f_{i+1} \longrightarrow 0$$

is short exact for each i.

***Proposition* I.2.5.4** (left exactness of Hom)**.** — (i) *A sequence*

$$M' \xrightarrow{f} M \xrightarrow{g} M'' \longrightarrow 0 \tag{I.2.5.4.1}$$

of A-modules is exact iff the sequence

$$0 \longrightarrow \operatorname{Hom}(M'', N) \xrightarrow{g^*} \operatorname{Hom}(M, N) \xrightarrow{f^*} \operatorname{Hom}(M', N) \tag{I.2.5.4.2}$$

is exact for all A-modules N.

(ii) *A sequence*

$$0 \longrightarrow N' \xrightarrow{f} N \xrightarrow{g} N''$$

of A-modules is exact iff the sequence

$$0 \longrightarrow \operatorname{Hom}(M, N') \xrightarrow{f_*} \operatorname{Hom}(M, N) \xrightarrow{g_*} \operatorname{Hom}(M, N'')$$

is exact for all A-modules M.

Proof. — We will only show (i). The other proof is an easy exercise.

Suppose that sequence (I.2.5.4.1) is exact, and let N be any A-module. For any $\phi\colon M'' \to N$ with $g^*(\phi) = \phi \circ g = 0$, we have $\phi = 0$ since g is surjective. This shows exactness at $\operatorname{Hom}(M'', N)$. For exactness at $\operatorname{Hom}(M, N)$, note that $f^*(g^*(\phi)) = \phi \circ g \circ f = 0$ since $g \circ f = 0$, so $\operatorname{im} g^* \subseteq \ker f^*$. Now let $\phi\colon M \to N$ with $f^*(\phi) = \phi \circ f = 0$. Then $\ker g = \operatorname{im} f \subseteq \ker \phi$, and hence ϕ induces $\bar{\phi}\colon M'' \xrightarrow{\sim} M/\ker g \to N$ where the isomorphism is induced by g. One now checks that $\bar{\phi} \circ g = \phi$. This shows $\ker f^* \subseteq \operatorname{im} g^*$.

Conversely, suppose that sequence (I.2.5.4.2) is exact for all N. To show exactness at M'', take $N = \operatorname{coker} g$ and consider the canonical projection $\pi\colon M \twoheadrightarrow \operatorname{coker} g$. Then $g^*(\pi) = \pi \circ g = 0$, and hence $\pi = 0$ since g^* is injective. This shows that $\operatorname{coker} g = 0$, i. e., that g is surjective. For exactness at M, note that $f^*(g^*(\phi)) = \phi \circ g \circ f = 0$ for all maps $\phi\colon M'' \to N$. Taking $N = M''$ and $\phi = \mathrm{id}$ in particular, we see that $g \circ f = 0$, i. e., $\operatorname{im} f \subseteq \ker g$. Now take $N = \operatorname{coker} f$ and consider the canonical projection $\pi\colon M \twoheadrightarrow \operatorname{coker} f$. Then $f^*(\pi) = \pi \circ f = 0$, i. e., $\pi \in \ker f^* = \operatorname{im} g^*$. In other words, we have $\pi = g^*(\psi) = \psi \circ g$ for some $\psi\colon M'' \to \operatorname{coker} f$. Thus $\ker g \subseteq \ker \pi = \operatorname{im} f$.[I.7] ◻

***Snake lemma* I.2.5.5.** — *Let*

$$\begin{array}{ccccccc}
 & & M' & \xrightarrow{u} & M & \xrightarrow{v} & M'' & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle f'} & & \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle f''} & & \\
0 & \longrightarrow & N' & \xrightarrow{u'} & N & \xrightarrow{v'} & N'' & &
\end{array}$$

be a commutative diagram of A-modules with exact rows. Then there exists a long exact sequence

$$\ker f' \xrightarrow{\bar{u}} \ker f \xrightarrow{\bar{v}} \ker f'' \xrightarrow{\delta} \operatorname{coker} f' \xrightarrow{\bar{u}'} \operatorname{coker} f \xrightarrow{\bar{v}'} \operatorname{coker} f''$$

where $\bar{u}, \bar{v}, \bar{u}', \bar{v}'$ are induced by u, v, u', v'. Moreover,

(i) *$\bar{u}$ is injective if u is;*
(ii) *$\bar{v}'$ is surjective if v' is.*

Proof. — We will only construct the *boundary homomorphism* δ. Let $x'' \in \ker f''$; then $x'' = v(x)$ for some $x \in M$ by exactness at M'', and $v'(f(x)) = f''(v(x)) = 0$. Hence $f(x) \in \ker v' = \operatorname{im} u'$ by exactness in N, i. e., $f(x) = u'(y')$ for some $y' \in N'$. Define $\delta(x'')$ to be the image of y' in $\operatorname{coker} f'$.

This is well-defined: there are two choices, namely for x and for y'. For any other choice $\tilde{y}'$ instead of y', we have $u'(y' - \tilde{y}') = 0$, i. e., $y' - \tilde{y}' \in \ker u' = 0$, and thus $y' = \tilde{y}'$. On the other hand, for any other choice $\tilde{x}$ instead of x, we have $v(x - \tilde{x}) = 0$, whence $x - \tilde{x} \in \ker v = \operatorname{im} u$ by exactness in M. So $u(x') = x - \tilde{x}$ for some $x' \in M'$, and $u'(f'(x')) = f(x) - f(\tilde{x})$. As in the construction of δ, we have $f(x) = u'(y')$ and $f(\tilde{x}) = u'(\tilde{y}')$ for some $y', \tilde{y}' \in N'$. Hence $f'(x') = y' - \tilde{y}'$ since u' is injective, and thus $y' \equiv \tilde{y}' \mod \operatorname{im} f'$, as desired.

The technique above is called a *diagram chase*, i. e., picking an element of interest and repeatedly constructing preimages and images by virtue of exactness in the given diagram. The claimed exactness of long exact sequence is proved by a diagram chase and is left as an exercise. ◻

[I.7] This can be also shown by universal properties of kernels and cokernels.

***Five lemma* I.2.5.6.** — *Let*

$$\begin{array}{ccccccccc} M_1 & \longrightarrow & M_2 & \longrightarrow & M_3 & \longrightarrow & M_4 & \longrightarrow & M_5 \\ \downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_3} & & \downarrow{\scriptstyle f_4} & & \downarrow{\scriptstyle f_5} \\ N_1 & \longrightarrow & N_2 & \longrightarrow & N_3 & \longrightarrow & N_4 & \longrightarrow & N_5 \end{array}$$

be a commutative diagram of A-modules with exact rows. Then the following hold:

(i) *If f_1 is surjective and if f_2 and f_4 are injective, then f_3 is injective.*
(ii) *If f_5 is injective and if f_2 and f_4 are surjective, then f_3 is surjective.*

Proof. — A nasty exercise in diagram chasing. ◻

***Remark* I.2.5.7.** — Both the snake lemma I.2.5.5 and five lemma I.2.5.6 are famous statements from *homological algebra*. In particular, the snake lemma I.2.5.5 is a special case of the long exact homology sequence induced by a short exact sequence of A-chain complexes.

Definition* I.2.5.8.** — *An A-module M is of* ***finite presentation *if there exists an exact sequence of the form $A^m \to A^n \to M \to 0$. In other words, M is finitely generated and the kernel of the obvious projection $A^n \twoheadrightarrow M$ is finitely generated as well.*

***Proposition* I.2.5.9.** — *Let*

$$0 \longrightarrow K \xrightarrow{f} M \xrightarrow{g} Q \longrightarrow 0$$

be a short exact sequence of A-modules

(i) *If K and Q are finitely generated, then so is M.*
(ii) *If M is finitely generated and if Q is finitely presented, then K is finitely generated.*
(iii) *If M is finitely presented and if K is finitely generated, then Q is finitely presented.*
(iv) *If K and Q are finitely presented, then so is M.*

Proof. — For (i), assume w. l. o. g. that $K = \operatorname{im} f$ is a submodule of M and that $Q = \operatorname{coker} f = M/K$. Say $K = (x_1, \dots, x_n)$ and $Q = (\bar{y}_1, \dots, \bar{y}_m)$. Let $z \in M$. Then $z \equiv \sum_{i=1}^m a_i y_i \mod K$ for some $a_i \in A$, and so $\tilde{z} := z - \sum_{i=1}^m a_i y_i \in K$. It follows that $\tilde{z} = \sum_{i=1}^n \tilde{a}_i x_i$ for some $\tilde{a}_i \in A$. Therefore $M = (x_1, \dots, x_n, y_1, \dots, y_m)$.

We now show (ii). By hypotheses, let

$$A^m \xrightarrow{f'} A^n \xrightarrow{g'} Q \longrightarrow 0$$

be a finite presentation of Q. Since free modules are projective, we may lift g' along g, which is surjective, to the map $h_M \colon A^n \to M$. Similarly, we may lift $h_M \circ f'$ along $\bar{f} \colon K \twoheadrightarrow \operatorname{im} f$ to $h_K \colon A^m \to K$. In other words, we obtain the commutative diagram

$$\begin{array}{ccccccccc} & & A^m & \xrightarrow{f'} & A^n & \xrightarrow{g'} & Q & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle h_K} & & \downarrow{\scriptstyle h_M} & & \Vert{\scriptstyle \mathrm{id}} & & \\ 0 & \longrightarrow & K & \xrightarrow{f} & M & \xrightarrow{g} & Q & \longrightarrow & 0 \end{array}$$

with exact rows. By the snake lemma I.2.5.5, we obtain the isomorphism $\operatorname{coker} h_K \xrightarrow{\sim} \operatorname{coker} h_M$. Since $\operatorname{coker} h_M$ is a quotient of the finitely generated M, it is finitely generated as well. Since $\operatorname{im} h_K$ is finitely generated, (i) implies that K is finitely generated.

We proceed with (iii). By hypotheses, let $h_M\colon A^n \twoheadrightarrow M$ be a finite presentation of M, i. e., $\ker h_M$ is finitely generated. Consider the commutative diagram

$$\begin{tikzcd} & 0 \arrow[r] \arrow[d] & A^n \arrow[r, equal, "\mathrm{id}"] \arrow[d, two heads, "h_M"] & A^n \arrow[r] \arrow[d, "h_Q"] & 0 \\ 0 \arrow[r] & K \arrow[r, "f"] & M \arrow[r, "g"] & Q \arrow[r] & 0 \end{tikzcd}$$

with exact rows where $h_Q = g \circ h_M$ is surjective. By the snake lemma I.2.5.5, we obtain an exact sequence

$$0 \longrightarrow \ker h_M \longrightarrow \ker h_Q \longrightarrow K \longrightarrow 0.$$

As $\ker h_M$ and K are finitely generated by hypothesis, (i) implies that $\ker h_Q$ is finitely generated as well. Thus h_Q is a finite presentation of Q.

Lastly (iv). Let $h_K\colon A^n \twoheadrightarrow K$ and $h_Q\colon A^m \twoheadrightarrow Q$ be finite presentations of K and Q, i. e., the kernels of both maps are finitely generated. Since K and Q are finitely generated in particular, (i) gives the commutative diagram

$$\begin{tikzcd} 0 \arrow[r] & A^n \arrow[r, "\iota"] \arrow[d, two heads, "h_K"] & A^{n+m} \arrow[r, "\pi"] \arrow[d, "h_M"] & A^m \arrow[r] \arrow[d, two heads, "h_Q"] & 0 \\ 0 \arrow[r] & K \arrow[r, "f"] & M \arrow[r, "g"] & Q \arrow[r] & 0 \end{tikzcd}$$

where ι and π are the obvious inclusions and projections into or from $A^{n+m} = A^n \oplus A^m$. The snake lemma I.2.5.5 gives the long exact sequence

$$0 \longrightarrow \ker h_K \longrightarrow \ker h_M \longrightarrow \ker h_Q \longrightarrow 0 \longrightarrow \operatorname{coker} h_M \longrightarrow 0.$$

Then immediately $\operatorname{coker} h_M = 0$, i. e., h_M is surjective. Since $\ker h_K$ and $\ker h_Q$ are finitely generated, so is $\ker h_M$ by (i). Therefore h_M is a finite presentation of M. ∎

Definition* I.2.5.10.** — *Let C be a class of A-modules, and let $\lambda\colon C \to \mathbb{Z}$ (or more generally $\lambda\colon C \to G$ for an additive abelian group G) be a function. Then λ is* ***additive *if for any short exact sequence $0 \to M' \to M \to M'' \to 0$ in C, we have $\lambda(M') - \lambda(M) + \lambda(M'') = 0$.*

***Example* I.2.5.11.** — (i) The function $V \mapsto \dim V$ is an additive function on the class of finite-dimensional k-vector spaces by the rank-nullity theorem.

(ii) The function $M \mapsto \operatorname{len} M$ is an additive function on the class of A-modules of finite length (see fact I.5.2.8).

***Proposition* I.2.5.12.** — *Let C be a collection of A-modules. Let $0 \to M_0 \to \cdots \to M_n \to 0$ be an exact sequences in C such that all kernels are in C as well. Then for any additive function λ on C, we have $\sum_{i=0}^{n} (-1)^i \lambda(M_i) = 0$.*

Proof. — Set $M_{i-1} = M_{n+1} = 0$. If $f_i\colon M_i \to M_{i+1}$ for $-1 \le i \le n+1$ is the map in the given exact sequence, then we can split this long exact sequence as many short exact sequences $0 \to \ker f_i \to M_i \to \ker f_{i+1} \to 0$. By additivity, $\lambda(M_i) = \lambda(\ker f_i) + \lambda(\ker f_{i+1})$. Plugging in the telescoping sum yields the statement. ∎

I.2.6. Tensor product.

Definition* I.2.6.1.** — *Let M, N and P be three A-modules. A map $M \times N \to P$ is A-**bilinear

(i) *if $f(x, -)\colon N \to P$ is A-linear for all $x \in M$, and*
(ii) *if $f(-, y)\colon M \to P$ is A-linear for all $y \in N$.*

Here $M \times N$ should be understood as a Cartesian product of sets.

***Construction* I.2.6.2** (universal property of the tensor product)**.** — Let M and N be two A-modules. We will construct an A-module T together with an A-bilinear map $g\colon M \times N \to T$ such that the following universal property is satisfied. For any A-bilinear map $f\colon M \times N \to P$, there exists a unique A-linear map $\tilde{f}\colon T \to P$ such that the following diagram commutes:

$$\begin{tikzcd} M \times N \arrow[r, "g"] \arrow[dr, "f"'] & T \arrow[d, dashed, "\tilde{f}"] \\ & P. \end{tikzcd} \tag{I.2.6.2.1}$$

Let

$$C := A^{(M \times N)} = \left\{ \sum_{i=1}^{n} a_i(x_i, y_i) \;\middle|\; a_i \in A, x_i \in M, y_i \in N \right\}$$

be the free A-module consisting of all formal A-linear combinations of elements of $M \times N$. Let D be the submodule of C generated by the elements

$$(x + x', y) - (x, y) - (x', y), \quad (x, y + y') - (x, y) - (x, y'), \quad (ax, y) - a(x, y), \quad (x, ay) - a(x, y)$$

for all $x, x' \in M$; $y, y' \in N$ and $a \in A$.

Set $T := C/D$; let $x \otimes y$ be the image of $(x, y) \in M \times N$ in T. Observe that D is defined in such a way that

$$g\colon M \times N \longrightarrow T, \qquad (x, y) \longmapsto x \otimes y$$

is A-bilinear.

Note that any map $f\colon M \times N \to P$ of sets extends A-linearly to an A-module map $\bar{f}\colon C \to P$. If f is additionally A-bilinear, then $D \subseteq \ker \bar{f}$. By the homomorphism theorem I.2.2.8, we see that $\bar{f}$ factors through the map $\tilde{f}\colon T = C/D \to P$, $x \otimes y \mapsto f(x, y)$. It is obviously the unique map such that diagram (I.2.6.2.1) commutes. Thus we have found T and g.

We call T the **tensor product** of M and N and write $M \otimes_A N := T$. If it is clear that these are A-modules, we will also just write $M \otimes N$. To summarise, we have the following rules for all $x, x' \in M$; $y, y' \in N$ and $a \in A$:

(i) $(x + x') \otimes y = x \otimes y + x' \otimes y$,
(ii) $x \otimes (y + y') = x \otimes y + x \otimes y'$,
(iii) $(ax) \otimes y = a(x \otimes y) = x \otimes (ay)$.

***Remark* I.2.6.3.** — As always with universal properties, it implies that (T, g) is unique up to unique isomorphism, that is, for another pair (T', g') satisfying the universal property, there exists a unique A-module isomorphisms $\phi\colon T \xrightarrow{\sim} T'$ such that $\phi \circ g = g'$.

To see this, applying the universal property of (T, g) to (T', g'), we obtain a unique map $\phi\colon T \to T'$ such that $g' = \phi \circ g$. Swapping the roles of (T, g) and (T', g'), we obtain a unique map $\phi'\colon T' \to T$ such that $g = \phi' \circ g'$. It follows that $g = \phi' \circ \phi \circ g$ and $g' = \phi \circ \phi' \circ g'$. However, applying the universal property of (T, g) to itself, we see that $\phi' \circ \phi = \mathrm{id}$; similarly $\phi \circ \phi' = \mathrm{id}$, i. e., ϕ is an isomorphism.

***Principle* I.2.6.4.** — As it turns out, the universal property fully characterises the tensor product, so we may work with this property and forget about the explicit construction of the tensor product. The reason is simply that the tensor product is defined precisely such that it satisfies the universal property and nothing more.

***Warning* I.2.6.5.** — The tensor product $M \otimes N$ is generated by the *elementary tensors* $x \otimes y$ with $x \in M$ and $y \in N$. That is, if $(x_i)_i$ and $(y_j)_j$ are generators of M and N resp., then any element in $M \otimes N$ is a finite sum $\sum_{(i,j)} x_i \otimes y_j$. In particular, general elements in $M \otimes N$ are *not* only of the form $x \otimes y$.

***Fact* I.2.6.6.** — *If M and N are finite A-modules, then $M \otimes N$ is finite as well.*

Proof. — Trivial. ∎

***Warning* I.2.6.7.** — The notation $x \otimes y$ is inherently ambiguous unless we specify the modules M and N for which $x \in M$ and $y \in N$. Consider the $\mathbb{Z}$-modules $M = \mathbb{Z}$ and $N = \mathbb{Z}/2\mathbb{Z}$, and consider the submodule $M' = 2\mathbb{Z}$ of M; let $\bar{1} \in N$ be the non-trivial element. Consider the element $2 \otimes \bar{1}$. As an element of $M \otimes N$, we have $2 \otimes \bar{1} = 1 \otimes (2 \cdot \bar{1}) = 1 \otimes 0 = 0.$(I.8) However, as an element of $M' \otimes N$, we have $2 \otimes \bar{1} \neq 0$.

Add example after [AM16, p. 2.18].

***Fact* I.2.6.8.** — *Let M and N be A-modules. Suppose that $\sum_{i=1}^n x_i \otimes y_i = 0$ in $M \otimes N$ with $x_i \in M$ and $y_i \in N$. Then there are finite submodules M_0 and N_0 of M and N resp. such that $\sum_{i=1}^n x_i \otimes y_i = 0$ in $M_0 \otimes N_0$.*

Proof. — Consider construction I.2.6.2 with $M \otimes N = C/D$. Then $\sum_{i=1}^n x_i \otimes y_i = 0$ implies that $z := \sum_{i=1}^n (x_i, y_i) \in D$. Write z as a finite sum of generators of D, and let $\tilde{M}_0$ and $\tilde{N}_0$ be the submodules generated by the finitely many necessary elements to write these generators. Finally, set $M_0 := \tilde{M}_0 + (x_1, \dots, x_n)$ and $N_0 := \tilde{N}_0 + (y_1, \dots, y_n)$. By construction I.2.6.2, we have $\sum_{i=1}^n x_i \otimes y_i = 0$ in $M_0 \otimes N_0$. ∎

***Observation* I.2.6.9.** — We can easily extend this to finite tensor products. Let $M_1, \dots, M_r$ be A-modules. Analogously, we define A-*multilinear* maps $f\colon M_1 \times \dots \times M_r \to P$ which are linear in each variable.

Then we can construct a multitensor product $T = M_1 \otimes \dots \otimes M_r$ with a canonical map $g\colon M_1 \times \dots \times M_r \to T$ in the same way. In particular, it satisfies the following universal property. For any A-multilinear map $f\colon M_1 \times \dots \times M_r \to P$, there exists a unique map $\tilde{f}\colon T \to P$ such that the following diagram commutes:

$$\begin{tikzcd} M_1 \times \dots \times M_r \arrow[r, "g"] \arrow[dr, "f"'] & T \arrow[d, dashed, "\tilde{f}"] \\ & P. \end{tikzcd}$$

I.2.7. Properties of the tensor product.

***Notation* I.2.7.1.** — Let $\mathrm{Bihom}_A(M, N; P)$ denote the set of all A-bilinear maps $M \times N \to P$. This carries an A-module structure by pointwise addition and scalar multiplication (note that A is commutative).

***Tensor-hom adjunction* I.2.7.2.** — *Let M, N, P be three A-modules. Then we obtain a canonical isomorphism of A-modules*

$$\mathrm{Hom}(M \otimes N, P) \cong \mathrm{Bihom}(M, N; P) \cong \mathrm{Hom}(M, \mathrm{Hom}(N, P)).$$

(I.8) The last equality follows since $1 \otimes 0 + 1 \otimes 0 = 2 \otimes 0 = 2(1 \otimes 0) = 1 \otimes 0$.

In fact, this is the incarnation of the universal property of the tensor product $M \otimes N$: if

$$\operatorname{Hom}(M, \operatorname{Hom}(N, P)) \xrightarrow{\sim} \operatorname{Hom}(T, P)$$

is a canonical isomorphism for all A-modules P, then T is the tensor product $M \otimes N$.

Proof. — The first isomorphism is given by the universal property of tensor products. The second isomorphism is given by

$$\begin{aligned}\operatorname{Bihom}(M, N; P) &\xrightarrow{\sim} \operatorname{Hom}(M, \operatorname{Hom}(N, P)),\\ f &\longmapsto (x \longmapsto f(x, -)),\\ ((x, y) \longmapsto f(x)(y)) &\longleftarrow f.\end{aligned}$$

***Fact* I.2.7.3** (laws for tensor products). — *Let M, N, P and $(M_i)_{i \in I}$ be A-modules. Then there are the following unique isomorphisms:*

(i) $M \otimes N \xrightarrow{\sim} N \otimes M$, $x \otimes y \mapsto y \otimes x$.
(ii) $(M \otimes N) \otimes P \xrightarrow{\sim} M \otimes (N \otimes P) \xrightarrow{\sim} M \otimes N \otimes P$, $(x \otimes y) \otimes z \mapsto x \otimes (y \otimes z) \mapsto x \otimes y \otimes z$.
(iii) $(\bigoplus_{i \in I} M_i) \otimes N \xrightarrow{\sim} \bigoplus_{i \in I} M_i \otimes N$, $(x_i)_{i \in I} \otimes y \mapsto (x_i \otimes y)_{i \in I}$.
(iv) $A \otimes M \xrightarrow{\sim} A$, $a \otimes x \mapsto ax$.

Proof. — All statements are proven with universal properties of tensor products, multitensor products and direct sums. For instance the first isomorphism of (ii). For any $z \in P$, the universal property of the tensor product $M \otimes N$ gives

$$M \otimes N \longrightarrow M \otimes N \otimes P, \qquad x \otimes y \longmapsto x \otimes y \otimes z.$$

Next the universal property of $(M \otimes N) \otimes P$ gives

$$(M \otimes N) \otimes P \longrightarrow M \otimes N \otimes P, \qquad (x \otimes y) \otimes z \longmapsto x \otimes y \otimes z. \tag{I.2.7.3.1}$$

Finally for the converse map, note that $(x, y, z) \mapsto (x \otimes y) \otimes z$ is A-trilinear. Hence the universal property of the multitensor product $M \otimes N \otimes P$ gives

$$M \otimes N \otimes P \longrightarrow (M \otimes N) \otimes P, \qquad x \otimes y \otimes z \longmapsto (x \otimes y) \otimes z. \tag{I.2.7.3.2}$$

It is immediate that equations (I.2.7.3.1) and (I.2.7.3.2) are inverses of each other.

Alternatively, we can also use tensor-hom adjunction I.2.7.2; let us prove (iv). For any A-module P, we have the canonical, i. e., unique, isomorphism

$$\begin{aligned}\operatorname{Hom}(M, P) \xrightarrow{\sim} \operatorname{Hom}(A, \operatorname{Hom}(M, P)) &\xrightarrow{\sim} \operatorname{Hom}(A \otimes M, P),\\ f \longmapsto (a \longmapsto af) &\longmapsto (a \otimes m \longmapsto af(m) = f(am)).\end{aligned}$$

This shows that $A \otimes M \xrightarrow{\sim} A$, $a \otimes m \mapsto am$.

***Fact* I.2.7.4.** — *Let M be an A-module, and let P be a B-module. Let N be an (A, B)-**bimodule**, i. e., N is an A-module as well as a B-module such that $a(bx) = b(ax)$ for all $a \in A$, $b \in B$ and $x \in N$. Then the following hold:*

(i) *$M \otimes_A N$ is a B-module, $N \otimes_B P$ is an A-module, and $\operatorname{Hom}_B(N, P)$ is an A-module.*
(ii) *$(M \otimes_A N) \otimes_B P \cong M \otimes_A (N \otimes_B P)$.*
(iii) *$\operatorname{Hom}_A(M, \operatorname{Hom}_B(N, P)) \cong \operatorname{Hom}_B(M \otimes_A N, P)$.*

Proof. — First (i). $M \otimes_A N$ is naturally a B-module via $b(x \otimes y) = x \otimes (by)$ for $b \in B$, $x \in M$ and $y \in N$; similarly for $N \otimes_B P$. For $\operatorname{Hom}_B(N, P)$, the product af for $a \in A$ and $f \in \operatorname{Hom}_B(N, P)$ is given by $(af)(x) = f(ax)$ for all $x \in N$. Finally, the proofs of (ii) and (iii) are analogous to fact I.2.7.3 and the tensor-hom adjunction I.2.7.2.

***Observation* I.2.7.5** (functoriality of tensor products). — Let $f\colon M \to M'$ and $g\colon N \to N'$ be A-module maps. Note that $M \times N \to M' \otimes N'$, $(x, y) \mapsto f(x) \otimes g(y)$ is A-bilinear, whence we obtain an A-linear map

$$f \otimes g\colon M \otimes N \longrightarrow M' \otimes N', \qquad x \otimes y \longmapsto f(x) \otimes g(y)$$

by the universal property of tensor products.

Let $f'\colon M' \to M''$ and $g'\colon N' \to N''$ be further A-module maps. Since $(f' \circ f) \otimes (g' \circ g)$ and $(f' \otimes g') \circ (f \otimes g)$ agrees on all elementary tensors $x \otimes y$ of $M \otimes N$, they are indeed equal on $M \otimes N$, i. e.,

$$(f' \circ f) \otimes (g' \circ g) = (f' \otimes g') \circ (f \circ g).$$

***Proposition* I.2.7.6.** — *Let $f\colon M \twoheadrightarrow M'$ and $g\colon N \twoheadrightarrow N'$ be surjective A-module maps. Then we obtain an isomorphism*

$$M \otimes N/(\ker f \otimes N + M \otimes \ker g) \xrightarrow{\sim} M' \otimes N', \qquad \overline{x \otimes y} \longmapsto f(x) \otimes g(y).$$

In particular, this is applicable to presentations $f\colon A^{(I)} \twoheadrightarrow M'$ and $g\colon A^{(J)} \twoheadrightarrow N'$.

Proof. — Since f and g are surjective, $f \otimes g$ is surjective as well on elementary tensors of $M' \otimes N'$ and hence on all $M' \otimes N'$. For the kernel, consider the set

$$K := \{x \otimes y \in M \otimes N \mid f(x) = 0 \text{ or } g(y) = 0\}.$$

Then $(K) = \ker f \otimes N + M \otimes \ker g$.

Note that $(K) \subseteq \ker(f \otimes g)$, so $f \otimes g$ induces $\phi\colon (M \otimes N)/(K) \to M' \otimes N'$. On the other hand, the map

$$M' \times N' \longrightarrow (M \otimes N)/(K), \qquad (x', y') \longmapsto x \otimes y \bmod (K)$$

is A-bilinear for $f(x) = x'$ and $g(y) = y'$ with $x \in M$ and $y \in N$. Note that this map is well-defined: for another choice $f(\tilde{x}) = x'$ and $g(\tilde{y}) = y'$, we have

$$x \otimes y \equiv x \otimes y + (x' - x) \otimes y + x' \otimes (y' - y) \equiv x' \otimes y' \mod (K).$$

The universal property of tensor products yields a map $\psi\colon M' \otimes N' \to (M \otimes N)/(K)$. By definition, ϕ and ψ are mutual inverses, which proves the statement. ∎

***Proposition* I.2.7.7** (right exactness of tensor products). — *Let*

$$M' \xrightarrow{f} M \xrightarrow{g} M'' \longrightarrow 0 \tag{I.2.7.7.1}$$

be an exact sequence of A-modules. Then for all A-modules N, the sequence

$$M' \otimes N \xrightarrow{f \otimes \mathrm{id}} M \otimes N \xrightarrow{g \otimes \mathrm{id}} M'' \otimes N \longrightarrow 0 \tag{I.2.7.7.2}$$

is exact.

Proof. — Let $M^\bullet$ denote sequence (I.2.7.7.1) so that $M^\bullet \otimes N$ denotes sequence (I.2.7.7.2). By left exactness of Hom, we know that $\mathrm{Hom}(M^\bullet, \mathrm{Hom}(N, P))$ is exact for all A-modules P. Tensor-hom adjunction I.2.7.2 implies that $\mathrm{Hom}(M^\bullet \otimes N, P)$ is exact, whence $M^\bullet \otimes N$ is exact, again by left exactness of Hom. ∎

***Remark* I.2.7.8.** — In the language of category theory, we can generalise this. Let $F = - \otimes N$ and $G = \mathrm{Hom}(N, -)$ be functors on A-Mod, the category A-modules. Then tensor-hom adjunction I.2.7.2 takes the form $\mathrm{Hom}(F(M), P) \cong \mathrm{Hom}(M, G(P))$ for all A-modules M and P. One says that F is a *left adjoint* of U and that G is a *right adjoint* of F, written $F \dashv G$. The proof above shows that left adjoint functors are right exact, whereas right adjoint functors are left exact.

***Warning* I.2.7.9.** — In general, tensoring is *not* exact, that is if $M' \to M \to M''$ is exact, then $M' \otimes N \to M \otimes N \to M'' \otimes N$ is not exact for general A-modules N. For example, consider the exact sequence of $\mathbb{Z}$-modules $0 \to \mathbb{Z} \xrightarrow{n} \mathbb{Z}$, which is multiplication by n. Tensoring with $N = \mathbb{Z}/n\mathbb{Z}$ yields $0 \to N \xrightarrow{n} N$, which is not exact since the kernel is $N \neq 0$.

***Fact* I.2.7.10.** — *Let M be an A-module and let $\mathfrak{a}$ be an ideal of A. Then $(A/\mathfrak{a}) \otimes_A M \xrightarrow{\sim} M/\mathfrak{a}M$, $\bar{a} \otimes m \mapsto a\bar{m}$ is an isomorphism.*

Proof. — Tensoring the short exact sequence $0 \to \mathfrak{a} \to A \to A/\mathfrak{a} \to 0$ with M, we obtain the exact sequence $\mathfrak{a} \otimes M \to A \otimes M \to (A/\mathfrak{a}) \otimes M \to 0$, whence $(A/\mathfrak{a}) \otimes M \cong \operatorname{coker} f$ where $f\colon \mathfrak{a} \otimes M \to A \otimes M$. Since the image of the composition of f with $A \otimes M \xrightarrow{\sim} M$ is $\mathfrak{a}M$, the claim follows.

***Proposition* I.2.7.11.** — *Let $A \neq 0$ be a ring, and let $\phi\colon A^m \to A^n$ be an A-linear map.*

(i) *If ϕ is surjective, then $m \geq n$.*
(ii) *If ϕ is injective, then $m \leq n$.*
(iii) *If ϕ is bijective, then $m = n$ (cf. fact I.2.3.2).*

Proof. — For (i), pick any maximal ideal $\mathfrak{m}$ of A so that $k = A/\mathfrak{m}$ is a field. Tensoring with k yields a surjective A-linear map $\bar{\phi} = \mathrm{id}_k \otimes \phi\colon k \otimes A^m \twoheadrightarrow k \otimes A^n$. By extension, $\mathrm{id}_k \otimes \phi$ is also k-linear, i. e., a surjective map of k-vector spaces $\bar{\phi}\colon k^m \twoheadrightarrow k^n$. Thus $m \geq n$.

For (ii), assume by way of contradiction that $m > n$, and consider A^n as a submodule of A^m generated by the first n coordinates of A^m. Then $\phi\colon A^m \hookrightarrow A^n \subseteq A^m$ is an endomorphism on A^m, and lemma I.2.3.5 gives an equation $\Phi = \phi^l + a_1\phi^{l-1} + \dots + a_l = 0$ over A. Choose l to be minimal among all such equations. Then $a_l \neq 0$, for otherwise, $\phi(\phi^{l-1} + a_1\phi^{l-2} + \dots + a_{l-1}) = 0$ and injectivity of ϕ gives $\phi^{l-1} + a_1\phi^{l-2} + \dots + a_{l-1} = 0$, contradicting the minimality of l. Now for $e_m = (0, \dots, 0, 1) \in A^m \setminus A^n$, we have $\phi(e_m) = 0$, whence $\Phi(e_m) = a_l \neq 0$, which is absurd.

The last statement is obvious. Alternatively, perform the proof of (i) on the isomorphism ϕ.

I.2.8. Algebras.

***Definition* I.2.8.1.** — *Let $f\colon A \to B$ be a ring map. While B is a ring, it also obtains an A-module structure by virtue of f: for $a \in A$ and $b \in B$, define $ab := f(a)b$. Then B or, more precisely, f is called an A-**algebra**.*

***Example* I.2.8.2.** — (i) If $A = k$ is a field and if $B \neq 0$, then any ring map $f\colon A \to B$ is injective according to proposition I.1.2.9. In other words, any k-algebra B contains k as a subring.

(ii) For any ring A, there is a unique ring map $\mathbb{Z} \to A$, $n \mapsto n1$. Hence any ring A is already a $\mathbb{Z}$-algebra.

***Definition* I.2.8.3.** — *Let $f\colon A \to B$ and $g\colon A \to C$ be two A-algebras. Then a ring map $h\colon B \to C$ is an A-**algebra homomorphism** if it is an A-module map, or equivalently, if the following diagram commutes:*

$$\begin{tikzcd} B \arrow[rr, "h"] \arrow[dr, "f"'] & & C \arrow[dl, "g"] \\ & A. & \end{tikzcd}$$

***Convention* I.2.8.4.** — *From now on, we will only say* A-algebra map *instead of A-algebra homomorphism.*

***Definition* I.2.8.5.** — *Let $f\colon A \to B$ be a ring map so that B is an A-algebra.*

(i) *We say that f is **finite** or that B is a **finite** A-algebra if B is finite as an A-module.*

(ii) *We say that f is of **finite type** or that B is a **finitely generated** A-algebra if there are finitely many $x_1, \dots, x_n \in B$ such that*

$$B = \Big\{ \textit{finite sums } \sum b_i x_1^{i_1} \cdots x_n^{i_n} \;\Big|\; b_i \in \operatorname{im} f \Big\},$$

*or equivalently, if there exists a surjective A-algebra map $A[x_1, \dots, x_n] \twoheadrightarrow B$ (cf. fact I.2.3.4). In this respect, a ring A is **finitely generated** if it is finitely generated as a $\mathbb{Z}$-algebra.*

***Warning* I.2.8.6.** — Finite A-algebras and finitely generated A-algebras are different notions! In a finite A-algebra generated by $x_1, \dots, x_n$, each element is a *linear combination* of the x_i. In an A-algebra finitely generated by $x_1, \dots, x_n$, each element is a *polynomial* in the x_i. Finite implies finitely generated, but the converse does not hold, e. g., for $k[x]$.

***Construction* I.2.8.7** (extension and restriction of scalars)**.** — (i) Let $f\colon A \to B$ be an A-algebra, and let M be a B-module. Then M becomes an A-module by virtue of f: define $ax = f(a)x$ for all $a \in A$ and $x \in M$. This operation is called **restriction of scalars**.

(ii) Conversely, let B be an A-algebra, and let M be an A-module. Then the A-module $B \otimes_A M$ becomes a B-module by $b(b' \otimes x) = bb' \otimes x$ for all $b, b' \in B$ and $x \in M$. This operation is called **extension of scalars**. More concisely, we can consider B as an (A, B)-bimodule. Then $B \otimes_A M$ is a B-module by fact I.2.7.4.

***Fact* I.2.8.8.** — (i) *Let B be a finite A-algebra, and let M be a finite B-module. Then M is finite over A.*

(ii) *Let B be any A-algebra, and let M be a finite A-module. Then $B \otimes_A M$ is finite over B.*

Proof. — Let $M = \sum_{j=1}^{m} B y_n$ and $B = \sum_{i=1}^{n} A x_j$. Then $M = \sum_{i=1}^{n} \sum_{j=1}^{m} A x_i y_j$; this proves (i). Let $M = \sum_{i=1}^{n} A x_i$. Then $B \otimes_A M = \sum_{i=1}^{n} B(1 \otimes x_i)$; this proves (ii). ∎

***Construction* I.2.8.9** (tensor product of algebras)**.** — Let $f\colon A \to B$ and $g\colon A \to C$ be two A-algebras. Then the tensor product $D = B \otimes_A C$ as A-modules is well-defined.

In order to turn D into an A-algebra, we need to define a multiplication on D. By the universal property of the multitensor product, we obtain an A-linear map

$$D \otimes D \xrightarrow{\sim} B \otimes C \otimes B \otimes C \longrightarrow D, \qquad (b \otimes c) \otimes (b' \otimes c') \longmapsto bb' \otimes cc'.$$

This gives rise to an A-bilinear multiplication

$$D \times D \longrightarrow D, \qquad (b \otimes c, b' \otimes c') \longmapsto bb' \otimes cc',$$

or more generally,

$$\left(\sum_{i=1}^{n} b_i \otimes c_i \right) \left(\sum_{j=1}^{m} b'_j \otimes c'_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i b'_j \otimes c_i c'_j.$$

With this multiplication and the identity element $1 \otimes 1$, we see that D truly becomes and A-algebra. The corresponding map $A \to D$ of an A-algebra D is given by $a \mapsto f(a) \otimes 1 = 1 \otimes g(a)$.

In fact, the tensor product of A-algebras satisfies the universal property of coproducts of A-algebras. It comes with canonical inclusions $\iota_B\colon B \to B \otimes C$, $b \mapsto b \otimes 1$ and $\iota_C\colon C \to B \otimes C$, $c \mapsto 1 \otimes c$. Note that they might not be injective, i. e., $B \ncong B \otimes 1$ and $C \ncong 1 \otimes C$.

***Example* I.2.8.10.** — Let M be an A-module, and let B be an A-algebra.

(i) Let $\mathfrak{a}$ be an ideal of A, so that we have an exact sequence $\mathfrak{a} \to A \to A/\mathfrak{a} \to 0$. Tensoring with M yields the exact sequence $\mathfrak{a}M \to M \to M \otimes (A/\mathfrak{a}) \to 0$, and we obtain the isomorphism $M \otimes (A/\mathfrak{a}) \cong M/\mathfrak{a}M$. Setting $M = B$ yields $B \otimes_A (A/\mathfrak{a}) \cong B/\mathfrak{a}B$ as A-modules, but one can check that this canonical isomorphism also holds between A-algebras.

(ii) Consider the polynomial ring $A[x]$ for an indeterminate x. Then $A[x] \cong \bigoplus_{n=0}^{\infty} A$ as A-modules, and so $M \otimes A[x] \cong \bigoplus_{n=0}^{\infty} M$. We can write $\bigoplus_{n=0}^{\infty} M = M[x]$ (multiplication on $M[x]$ is not necessary). Setting $M = B$ yields $B \otimes_A A[x] \cong B[x]$ as A-modules. Again, this isomorphism is also an A-algebra isomorphism.

I.2.9. Direct limits.

***Definition* I.2.9.1.** — *A partially ordered set I is a **directed set** if for any two elements $i, j \in I$, there exists some $k \in I$ such that $i, j \leq k$.*

Let $(M_i)_{i \in I}$ be a family of A-modules that is indexed by a directed set I. Moreover, for any two elements $i \leq j$ in I, let $f_{ij} \colon M_i \to M_j$ be an A-module map satisfying the following conditions:

(i) $f_{ii} = \mathrm{id}_{M_i}$ *for all $i \in I$,*
(ii) *Whenever $i \leq j \leq k$, the following diagram commutes:*

$$\begin{tikzcd} M_i \arrow[r, "f_{ij}"] \arrow[dr, "f_{ik}"'] & M_j \arrow[d, "f_{jk}"] \\ & M_k. \end{tikzcd}$$

*We call the collection $\mathcal{M} := (M_i, f_{ij})$ a **direct system** of A-modules over I or indexed by I.*

***Example* I.2.9.2.** — Some directed sets are:

(i) any totally ordered set,
(ii) $\mathbb{N}$ ordered by divisibility $a \mid b$,
(iii) subsets of a set ordered by inclusion.

***Construction* I.2.9.3** (universal property of direct limits)**.** — Let $\mathcal{M} = (M_i, f_{ij})$ be a direct system over I. We will construct an A-module M together with a family of A-module maps $(\mu_i \colon M_i \to M)_{i \in I}$ such that

$$\begin{tikzcd} M_i \arrow[rr, "f_{ij}"] \arrow[dr, "\mu_i"'] & & M_j \arrow[dl, "\mu_j"] \\ & M & \end{tikzcd}$$

commutes whenever $i \leq j$ and such that the following universal property is satisfied. For any A-module N with a family of A-module maps $(\nu_i \colon M_i \to N)_{i \in I}$ such that $\nu_i = \nu_j \circ f_{ij}$ whenever $i \leq j$, there exists a unique A-module map $\nu \colon M \to N$ such that the following diagram commutes whenever $i \leq j$:

$$\begin{tikzcd} M_i \arrow[rr, "f_{ij}"] \arrow[dr, "\mu_i"'] \arrow[ddr, bend right, "\nu_i"'] & & M_j \arrow[dl, "\mu_j"] \arrow[ddl, bend left, "\nu_j"] \\ & M \arrow[d, dashed, "\nu"] & \\ & N. & \end{tikzcd}$$

Let $C := \coprod_{i \in I} M_i$ be the disjoint union of the underlying sets of the M_i. We define the following relation $\sim$ on C: for any $x_i \in M_i$ and $x_j \in M_j$, we set $x_i \sim x_j$ if there is some $k \in I$ with $i, j \leq k$

such that $f_{ik}(x_i) = f_{jk}(x_j)$ in M_k. By definition of direct systems, one sees that $\sim$ is an equivalence relation. Now define the set of equivalence classes $M := C/\sim$. This becomes an A-module as follows:

(i) For all $a \in A$ and $x_i \in M_i$, define $a\bar{x}_i := \overline{ax_i}$.

(ii) For all $x_i \in M_i$ and $x_j \in M_j$, there is some $k \in I$ with $i, j \le k$. Then define $\bar{x}_i + \bar{x}_j := \overline{f_{ik}(x_i) + f_{jk}(x_j)}$.

Once again, one checks that these operations are well-defined, which is a very long calculation. For the canonical maps, define the map of sets $\mu_i\colon M_i \hookrightarrow C \twoheadrightarrow M$ for each $i \in I$, which is easily seen to be A-linear. By construction, we have $\mu_i = \mu_j \circ f_{ij}$ whenever $i \le j$.(I.9)

To construct the map ν, consider the unique map of sets $\tilde{\nu} := \coprod_{i\in I} \nu_i \colon C \to N$. Note that if $x_i \sim x_j$ for $x_i \in M_i$ and $x_j \in M_j$, then $f_{ik}(x_i) = f_{jk}(x_j)$ for some $k \ge i, j$, and so $\nu_i(x_i) = \nu_k \circ f_{ik}(x_i) = \nu_k \circ f_{jk}(x_j) = \nu_j(x_j)$. Thus $\tilde{\nu}$ factors through M, yielding a unique map $\nu\colon M \to N$. One checks that ν is A-linear; this shows the universal property.

We call M the **direct limit** of $\mathcal{M}$ and write $\varinjlim_{i\in I} M_i := M$. Note that $\varinjlim_{i\in I} M_i$ heavily depends on the given maps f_{ij}; however, we drop them from notation.

***Remark* I.2.9.4.** — More generally, we can consider direct systems $\mathcal{M}$ in any category, e. g., direct systems of sets, rings, A-algebras, etc. Their construction is identical to construction I.2.9.3.

***Fact* I.2.9.5.** — *Let (M_i, f_{ij}) be a direct system such that $M_i \subseteq M_j$ whenever $i \le j$, i. e., all f_{ij} are inclusions. Then $\varinjlim_i M_i \cong \bigcup_i M_i$.*

Proof. — This follows from construction I.2.9.3. ∎

***Example* I.2.9.6.** — Every A-module is a direct limit of its finite submodules, ordered by inclusion.

***Definition* I.2.9.7.** — *Let $\mathcal{M} = (M_i, f_{ij})$ and $\mathcal{N} = (N_i, g_{ij})$ be two direct systems of A-modules over the same directed set I. A **morphism of direct systems** $\phi\colon \mathcal{M} \to \mathcal{N}$ is a family of A-module maps $(\phi_i\colon M_i \to N_i)_{i\in I}$ such that the following diagram commutes whenever $i \le j$:*

$$\begin{array}{ccc} M_i & \xrightarrow{f_{ij}} & M_j \\ {\scriptstyle\phi_i}\downarrow & & \downarrow{\scriptstyle\phi_j} \\ N_i & \xrightarrow{g_{ij}} & N_j. \end{array}$$

*A sequence of direct systems $\mathcal{M} \to \mathcal{N} \to \mathcal{P}$ (over the same directed set) is **exact** if the underlying sequence of A-modules $M_i \to N_i \to P_i$ is exact for all $i \in I$.*

***Observation* I.2.9.8** (functoriality of direct limits)**.** — Fix a directed set I. Let $\mathcal{M} = (M_i, f_{ij})$ and $\mathcal{N} = (N_i, g_{ij})$ be two direct systems, and let $\phi = (\phi_i)_{i\in I}\colon \mathcal{M} \to \mathcal{N}$ be a morphism. Let $\mu = (\mu_i)_{i\in I}\colon \mathcal{M} \to M = \varinjlim_{i\in I} M_i$ and $\nu = (\nu_i)_{i\in I}\colon \mathcal{N} \to N = \varinjlim_{i\in I} N_i$ be the canonical maps. Note that for $\nu \circ \phi = (\nu_i \circ \phi_i)_{i\in I}\colon \mathcal{M} \to N$, we have $\nu_i \circ \phi_i = (\nu_j \circ \phi_j) \circ f_{ij}$ whenever $i \le j$. Hence by the universal property of M, we obtain a unique map

$$\bar{\phi} = \varinjlim_{i\in I} \phi_i \colon M = \varinjlim_{i\in I} M_i \longrightarrow N = \varinjlim_{i\in I} N_i.$$

(I.9) [AM16, §2, Exer. 14] gives another construction. Let $C := \bigoplus_{i\in I} M_i$, and let D be the submodule of D generated by all $x_j - f_{ij}(x_i)$ whenever $i \le j$ and $x_i \in M_i$. Then define $M := C/D$, and let $\mu_i\colon M_i \hookrightarrow C \twoheadrightarrow M$. To construct ν given N and $(\nu_i)_i$, consider $\tilde{\nu} := \bigoplus_{i\in I} \nu_i \colon C \to N$. Since $\nu_i = \nu_j \circ f_{ij}$ whenever $i \le j$, one checks that $D \subseteq \ker \tilde{\nu}$, so $\tilde{\nu}$ factors through $\nu\colon M \to N$.
This construction works generally for arbitrary *colimits*. However, we loose the advantage that this colimit is *filtered*; the original construction is preferred.

such that the following diagram commutes for all $i \in I$:

$$\begin{tikzcd} M_i \arrow[r, "\phi_i"] \arrow[d, "\mu_i"'] & N_i \arrow[d, "\nu_i"] \\ \varinjlim_{j\in I} M_j \arrow[r, dashed, "\bar\phi"] & \varinjlim_{j\in I} N_j. \end{tikzcd}$$

Let $\mathcal{P} = (P_i, h_{ij})$ be another direct system, and let $\psi = (\psi_i)_{i\in I} : \mathcal{N} \to \mathcal{P}$ be another morphism. In the same way, we obtain $\bar\psi = \varinjlim_{i\in I} \psi_i$ and $\varinjlim_{i\in I}(\psi_i \circ \phi_i)$ which fit in the following commutative diagram for any $i \in I$:

$$\begin{tikzcd} M_i \arrow[r, "\phi_i"] \arrow[d, "\mu_i"'] & N_i \arrow[r, "\psi_i"] \arrow[d, "\nu_i"'] & P_i \arrow[d, "\pi_i"'] \\ \varinjlim_{j\in I} M_j \arrow[r, dashed, "\bar\phi"] & \varinjlim_{j\in I} N_j \arrow[r, dashed, "\bar\psi"] & \varinjlim_{j\in I} P_j. \end{tikzcd} \tag{I.2.9.8.1}$$

Due to uniqueness of the induced maps, we have

$$\varinjlim_{i\in I}(\psi_i \circ \phi_i) = \varinjlim_{i\in I} \psi_i \circ \varinjlim_{i\in I} \phi_i.$$

The following two properties are the most important for direct limits.

***Proposition* I.2.9.9** (exactness of direct limits). — *Let*

$$(M_i, f_{ij}) \xrightarrow{\phi} (N_i, g_{ij}) \xrightarrow{\psi} (P_i, h_{ij})$$

be an exact sequence of direct systems of A-modules over the same directed set I. Then the induced sequence

$$\varinjlim_{i\in I} M_i \xrightarrow{\bar\phi} \varinjlim_{i\in I} N_i \xrightarrow{\bar\psi} \varinjlim_{i\in I} P_i$$

is exact.

Proof. — We take the same notation as in observation I.2.9.8. Since diagram (I.2.9.8.1) commutes, we have $\bar\psi \circ \bar\phi \circ \mu_i = \pi_i \circ \psi_i \circ \phi_i = 0$ for all $i \in I$. As $\bar\psi \circ \bar\phi$ is the unique induced map $M \to P$, we must have $\bar\psi \circ \bar\phi = 0$. This shows $\operatorname{im}\bar\phi \subseteq \ker\bar\psi$.

For the converse inclusion, let $\bar y \in \ker\bar\psi$. By construction I.2.9.3, there exists some representative $y \in N_i$ of $\bar y$ for some $i \in I$, so that $0 = \bar\psi(\nu_i(y)) = \pi_i(\psi_i(y)) = 0$ by diagram (I.2.9.8.1). Moreover, all representatives of $0 \in P$ are $\coprod_{i\leq j} \ker h_{ij}$; hence there is some $j \geq i$ such that $0 = \pi_i(\psi_i(y)) = h_{ij}(\psi_i(y))$. Also, by definition, $0 = h_{ij}(\psi_i(y)) = \psi_j(g_{ij}(y))$. Since $M_j \to N_j \to P_j$ is exact by hypothesis, there is a preimage $x \in M_j$ of $g_{ij}(y)$, i. e., $\phi_j(x) = g_{ij}(y)$. For $\bar x = \mu_j(x) \in M$, we have $\bar\phi(\bar x) = \nu_j(\phi_j(x)) = \nu_j(g_{ij}(y)) = \nu_i(y) = \bar y$, i. e., $\bar y \in \operatorname{im}\bar\phi$. □

***Proposition* I.2.9.10** (tensor products commute with direct limits). — *Let $\mathcal{M} = (M_i, f_{ij})$ be a direct system of A-modules, and let N be an A-module. Then there is a canonical isomorphism*

$$\varinjlim_{i\in I}(M_i \otimes N) \cong \left(\varinjlim_{i\in I} M_i\right) \otimes N$$

Proof. — Let $M = \varinjlim_{i\in I} M_i$, and let $P = \varinjlim_{i\in I}(M_i \otimes N)$. Let $(\mu_i)_{i\in I} \colon \mathcal{M} \to M$ be the canonical maps. They induce maps $\mu_i \otimes \mathrm{id}_N \colon M_i \otimes N \to M \otimes N$ for all $i \in I$, which in turn induce a map $\phi\colon P \to M \otimes N$ by the universal property of $\mathcal{M} \otimes N$.

To construct the inverse, define for each $y \in N$ and each $i \in I$ the A-linear map $g_i(-,y)\colon M_i \to P$, $x \mapsto \overline{x \otimes y}$. By the universal property, we obtain a map $g(-,y) = \varinjlim_{i\in I} g_i(-,y)\colon M \to P$. Note that this is also A-linear in y, whence $g\colon M \times N \to P$ is A-bilinear, which in turn induces $\psi\colon M \otimes N \to P$. One can now check that ϕ and ψ are mutual inverses. ∎

§ I.3. LOCALISATION

Localisations are one of the most important operations in commutative algebra. Geometrically, they correspond to restricting to *local* data; hence the name.

I.3.1. Localisations of rings.

***Motivation* I.3.1.1.** — Let $\mathbb{Z}$ be the integers. Then we can obtain the rational field $\mathbb{Q}$ as follows. We consider the set $S = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ and define the relation $\sim$ on this set by

$$(a,s) \sim (b,t) \iff at - bs = 0.$$

One directly checks that this is an equivalence relation (note that for transitivity, we need that $\mathbb{Z}$ is an integral domain in order to cancel elements). Then we can identify $S/{\sim}$ with $\mathbb{Q}$: denote the equivalence class of (a,s) by $a/s \in \mathbb{Q}$. Now we can endow $\mathbb{Q}$ with a ring structure via the usual arithmetic laws:

$$a/s + b/t := (at + bs)/st, \qquad (a/s)(b/t) := ab/st.$$

This can be generalised to arbitrary rings.

I.3.1.2. — Recall the definition I.1.3.11 of a *multiplicative subset* S of a ring A: it shall satisfy $1 \in S$ and $xy \in S$ whenever $x, y \in S$. In other words, S is a multiplicative sub-semigroup of A.

***Construction* I.3.1.3** (universal property of localisations of rings). — Let A be a ring, and let S be a multiplicative subset of A. We will construct a ring $S^{-1}A$ (this is purely notation) together with a ring map $f\colon A \to S^{-1}A$ with $f(S) \subseteq (S^{-1}A)^\times$ such that the following universal property is satisfied. For any ring map $g\colon A \to B$ such that $g(S) \subseteq B^\times$, there exists a unique ring map $\tilde{g}\colon S^{-1}A \to B$ such that the following diagram commutes:

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "g"'] & S^{-1}A \arrow[d, dashed, "\tilde{g}"] \\ & B. \end{tikzcd}$$

Define the relation $\sim$ on the set $A \times S$ by

$$(a,s) \sim (b,t) \iff (at - bs)u = 0 \quad \text{for some } u \in S. \tag{I.3.1.3.1}$$

This is clearly reflexive and symmetric. For transitivity, suppose that $(a,s) \sim (b,t)$ and $(b,t) \sim (c,u)$. Then there are $v, w \in S$ such that $(at - bs)v = 0$ and $(bu - ct)w = 0$. By multiplying the equations with suitable elements, we may eliminate b so that $(au - cs)tvw = 0$. Since $tvw \in S$, we obtain $(a,s) \sim (c,u)$. Hence, $\sim$ is an equivalence relation.

Define $S^{-1}A := (A \times S)/{\sim}$. Let $a/s \in S^{-1}A$ denote the equivalence class of (a,s). Endow $S^{-1}A$ with a ring structure by standard addition and multiplication of *fractions*: for all $a/s, b/t \in S^{-1}A$, define

(i) $a/s + b/t := (at + bs)/st$,
(ii) $(a/s)(b/t) := ab/st$.

The identity element is $1/1 \in S^{-1}A$, and the zero element is $0/1 \in S^{-1}A$. One checks by direct computation that these operations are well-defined. We further obtain a canonical ring map $f\colon A \to S^{-1}A$, $x \mapsto x/1$ with $f(S) \subseteq (S^{-1}A)^\times$ (the inverse of $s/1$ is $1/s \in S^{-1}A$).

To check the universal property, let $g\colon A \to B$ with $g(S) \subseteq B^\times$. We define $\tilde{g}\colon S^{-1}A \to B$ as follows. For all $a \in A$, we need $\tilde{g}(a/1) = \tilde{g}(f(a)) = g(a)$, so for all $s \in S$, we need $\tilde{g}(1/s) = \tilde{g}(s/1)^{-1} = g(s)^{-1}$ (this is possible since $g(S) \subseteq B^\times$). Hence, $\tilde{g}$ must be uniquely defined as $\tilde{g}(a/s) := g(a)g(s)^{-1}$. This is indeed well-defined: say $a/s = a'/s'$, i. e., there is some $t \in S$ such that $(as' - a's)t = 0$. Applying g yields $(g(a)g(s') - g(a')g(s))g(t) = 0$. Since $g(t) \in (S^{-1}A)^\times$, after some rearranging, we obtain $\tilde{g}(a/s) = \tilde{g}(a'/s')$.

We call $S^{-1}A$ the **localisation** or the **ring of fractions** of A w. r. t. S. Alternative notations include A_S and $A[S^{-1}]$.

***Remark* I.3.1.4.** — Note that the $u \in S$ in equation (I.3.1.3.1) is necessary for transitivity. If $(a, s) \sim (b, t)$ would only mean $at = bs$, then $(a, s) \sim (b, t)$ and $(b, t) \sim (c, u)$ mean $at = bs$ and $bu = ct$. It follows that $(au - cs)t = 0$. If $t \in S$ is a zero divisor, then we could not deduce $au = cs$.

***Remark* I.3.1.5.** — As always, the localisation of a ring as described by the universal property fully characterises it. In particular, the localisation is unique up to unique isomorphism (cf. remark I.2.6.3).

***Fact* I.3.1.6.** — *Let A be a ring, and let S be a multiplicative subset of A.*

(i) *The map $f\colon A \to S^{-1}A$ is injective iff S does not contain any zero divisors, e. g., if A is an integral domain.*
(ii) *$S^{-1}A = 0$ iff $0 \in S$.*

Proof. — The kernel of f is

$$\ker f = \bigcup_{s \in S} \operatorname{ann} s = \{a \in A \mid sa = 0 \text{ for some } s \in S\}.$$

Hence $\ker f = 0$ iff $\operatorname{ann} s = 0$ for all $s \in S$, i. e., s is a not a zero divisor; this shows (i). Moreover, we have $S^{-1}A = 0$ iff $1/1 = 0/1$ in $S^{-1}A$ iff $(1 \cdot 1 - 0 \cdot 1)u = u = 0$ for some $u \in S$; this is (ii). ∎

***Warning* I.3.1.7.** — The canonical map $f\colon A \to S^{-1}A$ is generally not injective.

***Notation* I.3.1.8.** — Let A be a ring. For any $f \in A$, the set $S = \{f^n \mid n \geq 0\}$ is multiplicative. We write $A_f := S^{-1}A$. Note that $A_f = 0$ iff f is nilpotent, so we generally exclude this case.

***Definition* I.3.1.9.** — *Let $\mathfrak{p}$ be a prime ideal of a ring A. Then $S = A \setminus \mathfrak{p}$ is multiplicative by definition (in fact, this holds only if $\mathfrak{p}$ is prime). Then we call $S^{-1}A$ the **localisation** of A at $\mathfrak{p}$ and write $A_\mathfrak{p} := S^{-1}A$.*

***Fact* I.3.1.10.** — *For a prime ideal $\mathfrak{p}$ of a ring A, the localisation $(A_\mathfrak{p}, \mathfrak{p}A_\mathfrak{p})$ is a local ring.*

Proof. — Note that since $\mathfrak{p}$ is prime, it will contain all zero divisors of A (and we automatically have $A \neq 0$). Hence, $A \hookrightarrow A_\mathfrak{p}$ is injective and we may consider A as a subring of $A_\mathfrak{p}$. Now, consider the extension $\mathfrak{p}A_\mathfrak{p}$ of $\mathfrak{p}$. For all $a/s \notin \mathfrak{p}A_\mathfrak{p}$ we have $a \in A \setminus \mathfrak{p}$, whence a/s is a unit in $A_\mathfrak{p}$. Thus, $(A_\mathfrak{p}, \mathfrak{p}A_\mathfrak{p})$ is a local ring by proposition I.1.3.25. ∎

***Warning* I.3.1.11.** — Some literature like [Mat86, ch. 2, §4, Ex. 2] also write $A_S = S^{-1}A$. In the case of prime ideals $\mathfrak{p}$, however, this yields to the illogical convention that $A_\mathfrak{p}$ and $A_{A \setminus \mathfrak{p}}$ denote the same thing. In this light, we will prefer $S^{-1}A$ over A_S.

***Definition* I.3.1.12.** — *Let A be a ring, and let S be the set of all non-zero divisors of A, which is multiplicative. Then $S^{-1}A$ is called the* **total ring of fractions** *of A. In the case that A is an integral domain so that $S = A \setminus \{0\}$, we call $S^{-1}A$ the* **field of fractions** *or* **quotient field** *of A and write* $\operatorname{Quot} A$. *Another notation for* $\operatorname{Quot} A$ *is* $\operatorname{Frac} A$.

***Example* I.3.1.13.** — (i) Let $A = \mathbb{Z}$. Consider $\mathfrak{p} = (p)$ for a prime number p; then $A_{\mathfrak{p}} = \{m/n \in \mathbb{Q} \mid p \nmid n\}$. Consider $f \in \mathbb{Z} \setminus \{0\}$; then $A_f = \{m/f^k \in \mathbb{Q} \mid k \geq 0\}$.

(ii) Let $A = k[t_1, \dots, t_n]$ where k is a field, and let $\mathfrak{p}$ be a prime ideal of A. Then $A_{\mathfrak{p}}$ is the set of rational functions $f/g \in \operatorname{Quot} A = k(t_1, \dots, t_n)$ with $g \notin \mathfrak{p}$.

Such local rings $A_{\mathfrak{p}}$ arise naturally in algebraic geometry. Let $\mathrm{V}(\mathfrak{p})$ be the set of all points $x \in k^n$ such that $f(x) = 0$ for all $f \in \mathfrak{p}$ (*affine algebraic set*). If k is infinite, then $A_{\mathfrak{p}}$ is the ring of rational functions on k^n that are defined almost everywhere on $\mathrm{V}(\mathfrak{p})$.

***Lemma* I.3.1.14.** — *Let A be an integral domain, so that we can consider any localisation of A as a subring of* $\operatorname{Quot} A$. *Then $A = \bigcap_{\mathfrak{m}} A_{\mathfrak{m}}$ where $\mathfrak{m}$ runs through all maximal ideals of A.*

Proof. — For $x \in \operatorname{Quot} A$ set $\mathfrak{a} := \{a \in A \mid ax \in A\}$, which is the ideal all possible denominators of x together with 0. By definition, we have $x \in A_{\mathfrak{p}}$ iff $\mathfrak{a} \subsetneq \mathfrak{p}$ for all prime ideals $\mathfrak{p}$ of A. Hence, if $x \in \bigcap_{\mathfrak{m}} A_{\mathfrak{m}}$ then $\mathfrak{a} = (1)$, and thus $x \in A$. The converse inclusion is clear. ∎

***Fact* I.3.1.15.** — *Let $f\colon A \to B$ be a ring map, and let S be a multiplicative subset of A. Then $S^{-1}B \cong f(S)^{-1}B$ as $S^{-1}A$-modules and as rings.*

Proof. — Note that $f(S)^{-1}B$ is an $S^{-1}A$-module via $(a/t)(b/f(s)) := ab/f(st)$. Define $\phi\colon S^{-1}B \to f(S)^{-1}B$, $b/s \mapsto b/f(s)$, which is clearly $S^{-1}A$-linear and a ring map. Let $b/s \in \ker\phi$, that is, $f(t)b = tb = 0$ for some $t \in S$. Hence, $b/s = 0$ and ϕ is injective. Since ϕ is surjective by definition, it is an isomorphism. ∎

***Proposition* I.3.1.16** (localisations commute with quotients)**.** — *Let A be a ring, let S be a multiplicative subset of A, and let $\mathfrak{a}$ be an ideal of A. If $\bar{S}$ is the image of S in $A/\mathfrak{a}$, then there is a canonical isomorphism*

$$S^{-1}A/S^{-1}\mathfrak{a} \xrightarrow{\sim} \bar{S}^{-1}(A/\mathfrak{a}), \qquad a/s \bmod S^{-1}\mathfrak{a} \longmapsto \bar{a}/\bar{s}.$$

Here, $S^{-1}\mathfrak{a}$ is the extended ideal of $\mathfrak{a}$ in $S^{-1}A$.

Proof. — Any element in the extension of $\mathfrak{a}$ in $S^{-1}A$ is of the for $\sum_i a_i/s_i$ with $a_i \in \mathfrak{a}$ and $s_i \in S$. Taking common denominators, we see that these elements precisely constitute $S^{-1}\mathfrak{a}$.

For the isomorphism, note that both rings satisfy the following universal property: every ring map $f\colon A \to C$ such that $g(S) \subseteq C^\times$ and $\mathfrak{a} \subseteq \ker g$ factorises uniquely through the canonical map $A \to S^{-1}A \to S^{-1}A/S^{-1}\mathfrak{a}$ (resp. $A \to A/\mathfrak{a} \to \bar{S}^{-1}(A/\mathfrak{a})$). Thus, both rings are canonically isomorphic. ∎

***Proposition* I.3.1.17** (ideal correspondence)**.** — *Let A be a ring, and let S be a multiplicative subset of A. Let $(-)^e$ and $(-)^c$ denote extension and contraction of ideals along the canonical map $A \to S^{-1}A$.*

(i) *All ideals of $S^{-1}A$ are an extensions of ideals of A. More precisely, if $\mathfrak{b}$ is an ideal of $S^{-1}A$, then $\mathfrak{b} = \mathfrak{b}^{ce}$. Hence, there is an inclusion preserving correspondence*

$$\begin{aligned} \{\textit{contracted ideals of } A\} &\longleftrightarrow \{\textit{ideals of } S^{-1}A\}, \\ \mathfrak{a} &\longmapsto \mathfrak{a}^e = S^{-1}\mathfrak{a}, \\ \mathfrak{b}^c &\longleftarrow\!\shortmid \mathfrak{b}. \end{aligned}$$

[Mat86, Thm. 4.1]: also holds for primary ideals.

(ii) *If $\mathfrak{a}$ is an ideal of A, then $\mathfrak{a}^{ec} = \bigcup_{s\in S}(\mathfrak{a} : s)$. In particular, $\mathfrak{a}^e = S^{-1}A$ iff $\mathfrak{a}$ meets S.*
(iii) *An ideal $\mathfrak{a}$ of A is contracted, i. e., $\mathfrak{a} = \mathfrak{a}^{ec}$, iff no element of S is a zero-divisor in $A/\mathfrak{a}$.*
(iv) *The correspondence in (i) restricts to*

$$\{\textit{prime ideals of } A \textit{ disjoint from } S\} \longleftrightarrow \{\textit{prime ideals of } S^{-1}A\}.$$

Proof. — To show (i), let $\mathfrak{b}$ be an ideal of $S^{-1}A$. By fact I.1.2.19, we have $S^{-1}f^{-1}(\mathfrak{b}) \subseteq \mathfrak{b}$. Conversely, let $x/s \in \mathfrak{b}$. Then $x/1 \in \mathfrak{b}$; hence $x \in f^{-1}(\mathfrak{b})$ and therefore $x/s \in S^{-1}f^{-1}(\mathfrak{b})$. The correspondence is known from fact I.1.2.19.

For (ii), we have $x \in \mathfrak{a}^{ec} = (S^{-1}\mathfrak{a})^c$ iff $x/1 = a/s$ for some $a \in \mathfrak{a}$ and $s \in S$ iff $xst = at \in \mathfrak{a}$ for some $a \in \mathfrak{a}$ and $s, t \in S$ iff $x \in \bigcup_{s\in S}(\mathfrak{a} : s)$. In particular, $\mathfrak{a}^e = S^{-1}A$ iff $\mathfrak{a}^{ec} = \bigcup_{s\in S}(\mathfrak{a} : s) = A$ iff $1 \in (\mathfrak{a} : s)$ for some $s \in S$ iff $\mathfrak{a}$ meets S.

For (iii), we have $\mathfrak{a}^{ec} = \bigcup_{s\in S}(\mathfrak{a} : s) \subseteq \mathfrak{a}$ ($\mathfrak{a} \subseteq \mathfrak{a}^{ec}$ is always true) iff: $sx \in \mathfrak{a}$ where $s \in S$ implies $x \in \mathfrak{a}$. This is equivalent to the statement that no $s \in S$ is a zero divisor in $A/\mathfrak{a}$.

Now to (iv). If $\mathfrak{q}$ is a prime ideal of $S^{-1}A$, then we know that $\mathfrak{q}^c$ is a prime ideal of A. Moreover, $\mathfrak{q}^c$ is disjoint from S by (ii). Conversely, let $\mathfrak{p}$ be a prime ideal of A. Let $\bar{S}$ be the image of S in $A/\mathfrak{p}$. Then we have $S^{-1}A/S^{-1}\mathfrak{p} \cong \bar{S}^{-1}(A/\mathfrak{p})$ by proposition I.3.1.16. Since $A/\mathfrak{p}$ is an integral domain, we know that $\bar{S}^{-1}(A/\mathfrak{p})$ is a subring of $\operatorname{Quot}(A/\mathfrak{p})$ and is hence either the zero ring or an integral domain. By (ii), the latter can only happen iff $\mathfrak{p}$ is disjoint from S. In this case, $\bar{S}$ has no zero divisors in $A/\mathfrak{p}$, so $\mathfrak{p}$ is a contracted ideal in A by (iii).

***Corollary* I.3.1.18.** — *Let $A \to B$ be an A-algebra, and let $\mathfrak{p}$ be a prime ideal of A. Then $\mathfrak{p}$ is the contraction of a prime ideal of B iff $\mathfrak{p} = \mathfrak{p}^{ec}$.*

Proof. — If $\mathfrak{p} = \mathfrak{q}^c$ for a prime ideal $\mathfrak{q}$ of B, then $\mathfrak{p}^{ec} = \mathfrak{q}^{cec} = \mathfrak{q}^c = \mathfrak{p}$. Conversely, suppose that $\mathfrak{p}^{ec} = \mathfrak{p}$. Let S be the image of $A \setminus \mathfrak{p}$ in B. Then $\mathfrak{p}^e$ is disjoint from S, so its extension in $S^{-1}B$ is proper and hence contained in a maximal ideal $\mathfrak{m}$ of $S^{-1}B$. If $\mathfrak{q}$ denotes the contraction of $\mathfrak{m}$ in B, then $\mathfrak{q}$ is prime, $\mathfrak{q} \supseteq \mathfrak{p}^e = \mathfrak{p}B$ and $\mathfrak{q} \setminus \mathfrak{p}^e \subseteq \mathfrak{q} \cap S = \varnothing$. Thus, $\mathfrak{p} = \mathfrak{p}^{ec} = \mathfrak{q}^c$.

***Remark* I.3.1.19.** — Localisations and quotients are two very important tools in algebraic geometry. Let A be a ring, and let $\mathfrak{q} \subseteq \mathfrak{p}$ be prime ideals of A. Then passing A to $A/\mathfrak{q}$ cuts out all prime ideals of A not containing $\mathfrak{q}$, and passing A to $A_\mathfrak{p}$ (which is a local ring with maximal ideal $\mathfrak{p}A_\mathfrak{p}$) cuts out all prime ideals not contained in $\mathfrak{p}$. Therefore, passing to $A_\mathfrak{p}/\mathfrak{q}A_\mathfrak{p} \cong (A/\mathfrak{q})_{\mathfrak{p}/\mathfrak{q}}$ only retains prime ideals between $\mathfrak{q}$ and $\mathfrak{p}$.

Definition* I.3.1.20.** — *For a prime ideal $\mathfrak{p}$ of a ring A, the residue field $\kappa(\mathfrak{p}) := A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p} \cong \operatorname{Quot}(A/\mathfrak{p})$ of the local ring $(A_\mathfrak{p}, \mathfrak{p}A_\mathfrak{p})$ is called the* ***residue field *at $\mathfrak{p}$.*

Alternative proof of proposition I.1.3.17. — We want to show that if $\mathfrak{a}$ is an ideal of A, then $\sqrt{\mathfrak{a}} = \bigcap_{\mathfrak{p}\supseteq\mathfrak{a}} \mathfrak{p}$ where $\mathfrak{p}$ runs through all prime ideals of A. Passing to the quotient $A/\mathfrak{a}$, it suffices to show that $\operatorname{nil} A = \bigcap_\mathfrak{p} \mathfrak{p}$.

The inclusion $\operatorname{nil} A \subseteq \bigcap_\mathfrak{p} \mathfrak{p}$ is clear. Conversely, let $f \notin \operatorname{nil} A$. Then $0 \notin \{f^n \mid n \geq 0\}$, so $A_f \neq 0$ has a maximal ideal by Krull's theorem I.1.3.12, whose contraction in A is a prime ideal $\mathfrak{p}$ disjoint from S by proposition I.3.1.17. Thus, $f \notin \mathfrak{p}$.

I.3.2. Localisations of modules.

***Construction* I.3.2.1.** — We can apply the same construction I.3.1.3 to modules. Let M be an A-module, and let S be a multiplicative subset of A. Define on $M \times S$ the relation $\sim$ by

$$(m, s) \sim (n, t) \iff u(tm - sn) = 0 \quad \text{for some } t \in S.$$

This is an equivalence relation by the same calculation as in construction I.3.1.3. Let m/s denote the equivalence class of (m, s) in $S^{-1}M := (M \times S)/{\sim}$. Endow it with an $S^{-1}A$-module structure by:

(i) $m/s + n/t := (tm + sn)/st$,
(ii) $(a/t)(m/s) := am/st$.

Via restriction of scalars, $S^{-1}M$ is also an A-module. We call $S^{-1}M$ with the $S^{-1}A$-module structure the **localisation** of M w. r. t. S. It comes with a canonical A-module map $f\colon M \to S^{-1}M$.

What about the universal property?

***Notation* I.3.2.2.** — Let M be an A-module. If $S = A \setminus \mathfrak{p}$ for a prime ideal $\mathfrak{p}$ of A, then we write $M_{\mathfrak{p}} := S^{-1}M$. If $S = \{f^n \mid n \geq 0\}$ for some $f \in A$, then we write $M_f := S^{-1}M$.

Here are nice applications of localisation.

Alternative proof of the Cayley-Hamilton theorem I.2.3.7. — After a choice of basis for M, write $\phi = (a_{ij})_{ij} \in \mathrm{Mat}_n(A)$, and our goal is to show that ϕ satisfies its characteristic polynomial $\chi = \det(\mathrm{id}_M T - \phi) \in A[T]$.

Firstly, observe that the theorem holds over the ring $B := \mathbb{Z}[x_{ij} \mid 1 \leq i, j \leq n]$: for any matrix in $\mathrm{Mat}_n(B)$, we may consider it as a matrix over the field $\operatorname{Quot} B$ via the inclusion $B \hookrightarrow \operatorname{Quot} B$, and the matrix certainly satisfies its own characteristic polynomial over $\operatorname{Quot} B$, so also over B by injectivity. This is in particular true for the matrix $\tilde{\phi} = (x_{ij})_{ij} \in \mathrm{Mat}_n(B)$ with its characteristic polynomial $\tilde{\chi} \in B[T]$. Now, consider $\mathrm{ev}_\phi\colon B \to A$, $x_{ij} \mapsto a_{ij}$. Since $\tilde{\chi}(\tilde{\phi}) = 0$ in B, we also have $\mathrm{ev}_\phi(\tilde{\chi}(\tilde{\phi})) = \chi(\phi) = 0$ in A. ∎

Alternative proof of corollary I.2.3.6. — This proof uses Nakayama's lemma I.2.3.8 instead of lemma I.2.3.5 (determinant trick). Recall the statement: let M be a finitely generated A-module, and let $\mathfrak{a}$ be an ideal of A such that $\mathfrak{a}M = M$. We want to find some $x \equiv 1 \mod \mathfrak{a}$ such that $xM = 0$.

Take the multiplicative subset $S = 1 + \mathfrak{a}$ of A. Then we have $S^{-1}\mathfrak{a} \subseteq \operatorname{rad}(S^{-1}A)$ by proposition I.1.3.23: for all $a/s \in S^{-1}\mathfrak{a}$ and all $b/t \in S^{-1}A$, we have $1 - ab/st \in (S^{-1}A)^\times$ since the denominator is $st - ab \in S$. Moreover, we have $S^{-1}M = (S^{-1}\mathfrak{a})(S^{-1}M)$ as $M = \mathfrak{a}M$, so Nakayama's lemma I.2.3.8 says that $S^{-1}M = 0$. If $x_1, \dots, x_n$ generate M, then there are $s_i \in S$ such that $s_i x_i = 0$. Setting $x := \prod_{i=1}^n s_i \in S$, that means, $x \equiv 1 \mod \mathfrak{a}$, it follows that $xM = 0$. ∎

***Observation* I.3.2.3** (functoriality of localisations)**.** — Let $f\colon M \to N$ be an A-module map, and let S be a multiplicative subset of A. Then f induces an A-linear map $S^{-1}f\colon S^{-1}M \to S^{-1}N$, $m/s \mapsto f(m)/s$. This is obviously functorial, i. e., for A-module maps $f\colon M \to N$ and $g\colon N \to P$, we have $S^{-1}(g \circ f) = S^{-1}g \circ S^{-1}f$.

***Proposition* I.3.2.4** (exactness of localisations)**.** — *Let*

$$M' \xrightarrow{f} M \xrightarrow{g} M''$$

be an exact sequence of A-modules. Then

$$S^{-1}M' \xrightarrow{S^{-1}f} S^{-1}M \xrightarrow{S^{-1}g} S^{-1}M''$$

is an exact sequence of $S^{-1}A$-modules.

Proof. — Since $g \circ f = 0$, we have $S^{-1}g \circ S^{-1}f = S^{-1}0 = 0$, that is, $\operatorname{im} S^{-1}f \subseteq \ker S^{-1}g$. Conversely, let $m/s \in \ker S^{-1}g$, meaning $g(m)/s = 0$ in $S^{-1}M''$. Then $g(tm) = tg(m) = 0$ in M'' for some $t \in S$, whence $tm \in \ker g = \operatorname{im} f$. Choose a preimage $m' \in M'$ such that $tm = f(m')$. It follows that $m/s = f(m')/st \in \operatorname{im} S^{-1}f$; thus $\ker S^{-1}g \subseteq \operatorname{im} S^{-1}f$. ∎

***Convention* I.3.2.5.** — *By the exactness of localisation, if N is a submodule of a module M, then we may consider $S^{-1}N$ as a submodule of $S^{-1}M$.*

***Fact* I.3.2.6** (laws for localisations)**.** — *Let N and P be submodules of an A-module M. Then the following hold:*

(i) $S^{-1}(N+P) = S^{-1}N + S^{-1}P$.
(ii) $S^{-1}(N \cap P) = S^{-1}N \cap S^{-1}P$.
(iii) $S^{-1}(M/N) \cong S^{-1}M/S^{-1}N$ *as* $S^{-1}A$*-modules.*

Proof. — For (i), we have $N, P \subseteq N+P$, and hence $S^{-1}N + S^{-1}P \subseteq S^{-1}(N+P)$. The other inclusion is obvious.

For (ii), let $y/s = z/t$ for $y/s \in S^{-1}N$ and $z/t \in S^{-1}P$. Then $u(ty - sz) = 0$ for some $u \in S$, whence $uty = usz \in N \cap P$. Therefore $y/s = uty/uts \in S^{-1}(N \cap P)$; we have shown $S^{-1}N \cap S^{-1}P \subseteq S^{-1}(N \cap P)$. The other inclusion is obvious.

Lastly for (iii), we apply S^{-1} to the short exact sequence $0 \to N \to M \to M/N \to 0$. The result follows from the homomorphism theorem.

***Fact* I.3.2.7** (laws for localisations)**.** — *Let $\mathfrak{a}$ and $\mathfrak{b}$ be ideals of a ring A. Then the following hold:*

(i) $S^{-1}(\mathfrak{a}+\mathfrak{b}) = S^{-1}\mathfrak{a} + S^{-1}\mathfrak{b}$.
(ii) $S^{-1}(\mathfrak{a}\mathfrak{b}) = (S^{-1}\mathfrak{a})(S^{-1}\mathfrak{b})$.
(iii) $S^{-1}(\mathfrak{a} \cap \mathfrak{b}) = S^{-1}\mathfrak{a} \cap S^{-1}\mathfrak{b}$.
(iv) $S^{-1}\sqrt{\mathfrak{a}} = \sqrt{S^{-1}\mathfrak{a}}$.

Proof. — (i) and (ii) are always true for extensions of ideals. (iii) is also true since it is true as A-modules. For (iv), we always have $S^{-1}\sqrt{\mathfrak{a}} \subseteq \sqrt{S^{-1}\mathfrak{a}}$ (laws for extensions). The verification of $\sqrt{\mathfrak{a}^e} \subseteq (\sqrt{\mathfrak{a}})^e$ is easy.

***Corollary* I.3.2.8.** — *For a multiplicative subset S of A, we have* $\operatorname{nil}(S^{-1}A) = S^{-1}\operatorname{nil} A$.

Proof. — Follows from proposition I.1.3.17.

***Proposition* I.3.2.9** (localisations commute with annihilators)**.** — *Let M be a finitely generated A-module. Then $S^{-1}(\operatorname{ann} M) = \operatorname{ann}(S^{-1}M)$.*

Proof. — Suppose that this is true for two A-modules M and N. Then, by properties of ideal quotients and localisations, we have

$$\begin{aligned} S^{-1}(\operatorname{ann}(M+N)) &= S^{-1}(\operatorname{ann} M \cap \operatorname{ann} N) \\ &= S^{-1}(\operatorname{ann} M) \cap S^{-1}(\operatorname{ann} N) \\ &= \operatorname{ann}(S^{-1}M) \cap \operatorname{ann}(S^{-1}N) \\ &= \operatorname{ann}(S^{-1}M + S^{-1}N) \\ &= \operatorname{ann}(S^{-1}(M+N)). \end{aligned}$$

So, the statement is again true for $M+N$, and hence it is enough to show that the statement is true for A-modules $M \cong A/\operatorname{ann} M$ generated by one element. Since $S^{-1}M \cong S^{-1}A/S^{-1}(\operatorname{ann} M)$ as $S^{-1}A$-modules, we have $\operatorname{ann}(S^{-1}M) = S^{-1}(\operatorname{ann} M)$.

***Corollary* I.3.2.10.** — *Let N and P be submodules of an A-module. If P is finitely generated, then $S^{-1}(N : P) = (S^{-1}N : S^{-1}P)$. In particular, $S^{-1}(\mathfrak{a} : \mathfrak{b}) = (S^{-1}\mathfrak{a} : S^{-1}\mathfrak{b})$ for ideals $\mathfrak{a}$ and $\mathfrak{b}$ of A where $\mathfrak{b}$ is finitely generated.*

Proof. — Since $(N : P) = \operatorname{ann}((N+P)/N)$ for annihilators, the statement follows from proposition I.3.2.9.

***Proposition* I.3.2.11** (localisations commute with tensor products)**.** — *Let M and N be A-modules. Then the following hold:*

(i) *There is a canonical isomorphism of $S^{-1}A$-modules*

$$S^{-1}A \otimes_A M \xrightarrow{\sim} S^{-1}M, \qquad (a/s) \otimes m \longmapsto am/s.$$

(ii) *There is a canonical isomorphism of $S^{-1}A$-modules*

$$S^{-1}M \otimes_{S^{-1}A} S^{-1}N \xrightarrow{\sim} S^{-1}(M \otimes_A N), \qquad (m/s) \otimes (n/t) \longmapsto (m \otimes n)/st.$$

Proof. — Consider the A-bilinear map $S^{-1}A \times M \to S^{-1}M$, $(a/s, m) \mapsto am/s$. This induces an A-linear map $f\colon S^{-1}A \otimes_A M \to S^{-1}M$ as claimed. By extension of scalars, this is also $S^{-1}A$-linear. This is obviously surjective. For injectivity, let $\sum_{i=1}^n (a_i/s_i) \otimes m \in \ker f$. Set $s := \prod_{i=1}^n s_i \in S$, $t_i := s/s_i \in S$ and $m := \sum_{i=1}^n a_i t_i m \in M$. Then we have

$$\sum_{i=1}^n (a_i/s_i) \otimes m = \sum_{i=1}^n (a_i t_i/s) \otimes m = \sum_{i=1}^n (1/s) \otimes (a_i t_i m) = (1/s) \otimes m.$$

This is sent to $m/s = 0$ under f, that is, $tm = 0$ for some $t \in S$. Therefore $(1/s) \otimes m = (t/st) \otimes m = (1/st) \otimes tm = 0$, whence f is injective and (i) holds.

Alternatively, we may give an inverse map to f. Define the A-module map $g\colon S^{-1}M \to S^{-1}A \otimes_A M$, $m/s \mapsto (1/s) \otimes m$. This is well-defined: for $m/s = n/t$, we have $utm = usn$ for some $u \in S$. Hence

$$(1/s) \otimes m = (ut/uts) \otimes m = (1/uts) \otimes utm = (1/uts) \otimes usn = (1/s) \otimes n.$$

One now checks that f and g are mutual inverse.

For (ii), we will exploit (i). Tensoring f with $S^{-1}N$ yields

$$S^{-1}N \otimes_A M \cong S^{-1}N \otimes_{S^{-1}A} S^{-1}A \otimes_A M \cong S^{-1}M \otimes_{S^{-1}A} S^{-1}N$$

as $S^{-1}A$-modules. Applying (i) to $M \otimes_A N$ yields

$$S^{-1}N \otimes_A M \cong S^{-1}A \otimes_A M \otimes_A N \cong S^{-1}(M \otimes_A N)$$

as $S^{-1}A$-modules. Note that all isomorphisms are canonical, so composing all isomorphisms yields the desired isomorphism. ∎

Insert [AM16, Cor. 3.6].

***Definition* I.3.2.12.** — *A property P of a A-modules M is said to be a* **local property** *if the following is true: M has P iff $M_\mathfrak{p}$ has P for each $\mathfrak{p} \in \operatorname{Spec} A$. The same notion local properties apply to rings A.*

***Proposition* I.3.2.13.** — *Let M be an A-module. Then the following are equivalent:*

(i) $M = 0$*;*
(ii) $M_\mathfrak{p} = 0$ *for all* $\mathfrak{p} \in \operatorname{Spec} A$*;*
(iii) $M_\mathfrak{m} = 0$ *for all* $\mathfrak{m} \in \operatorname{MaxSpec} A$*.*

Proof. — Obviously, (i) implies (ii), and (ii) implies (iii). Suppose that (iii) holds and that $M \neq 0$. Let $x \in M$ be non-zero. Then $\operatorname{ann} x$ is a proper ideal of A as it does not contain 1, so it is contained in a maximal ideal $\mathfrak{m}$ by corollary I.1.3.13. Consider $x/1 \in M_\mathfrak{m}$. Since $M_\mathfrak{m} = 0$ by hypothesis, x is annihilated by some element in $A \setminus \mathfrak{m}$. However, this contradicts $\operatorname{ann} x \subseteq \mathfrak{m}$. ∎

***Corollary* I.3.2.14.** — *Let M be a finitely generated A-module. Then the following are equivalent:*

(i) $M = 0$*;*

(ii) $M_\mathfrak{p}/\mathfrak{p}M_\mathfrak{p} = 0$ *for all* $\mathfrak{p} \in \operatorname{Spec} A$*;*
(iii) $M_\mathfrak{m}/\mathfrak{m}M_\mathfrak{m} = 0$ *for all* $\mathfrak{m} \in \operatorname{MaxSpec} A$*.*

Proof. — Since $M_\mathfrak{p}/\mathfrak{p}M_\mathfrak{p} = 0$ and $\mathfrak{p}A_\mathfrak{p} = \operatorname{rad} A_\mathfrak{p}$ implies $M_\mathfrak{p} = 0$ by Nakayama's lemma I.2.3.8, the statement follows from proposition I.3.2.13.

We can generalise this.

***Proposition* I.3.2.15.** — *Let $f\colon A \to B$ be an A-algebra, and let M be a finitely generated B-module. Then the following are equivalent:*

(i) $M = 0$*;*
(ii) $M_\mathfrak{p}/\mathfrak{p}M_\mathfrak{p} = 0$ *for all* $\mathfrak{p} \in \operatorname{Spec} A$*.*

Proof. — Suppose that $M \neq 0$. Then $M_\mathfrak{q} \neq 0$ for some prime ideal $\mathfrak{q}$ of B by proposition I.3.2.13. By Nakayama's lemma I.2.3.8, we hence have $M_\mathfrak{q}/\mathfrak{q}M_\mathfrak{q} \neq 0$. If $\mathfrak{p} = f^{-1}(\mathfrak{q})$, then $\mathfrak{p}M_\mathfrak{q} \subseteq \mathfrak{q}M_\mathfrak{q}$ and hence $M_\mathfrak{q}/\mathfrak{p}M_\mathfrak{q} \neq 0$. Let $S = A \setminus \mathfrak{p}$. Note that $M_\mathfrak{p}$ as an A-module and $f(S)^{-1}M$ as a B-module coincide: both are $\{m/s \mid m \in M, s \in S\}$. Since $f(S) \subseteq T := B \setminus \mathfrak{q}$, it follows that $M_\mathfrak{q} = T^{-1}M = T^{-1}(f(S)^{-1}M) = T^{-1}M_\mathfrak{p}$ and therefore $0 \neq M_\mathfrak{q}/\mathfrak{p}M_\mathfrak{q} \cong T^{-1}(M_\mathfrak{p}/\mathfrak{p}M_\mathfrak{p})$. Thus $M_\mathfrak{p}/\mathfrak{p}M_\mathfrak{p} \neq 0$. The converse is clear. Cf. [Mat86, Thm. 4.3].

***Remark* I.3.2.16.** — We cannot sharpen proposition I.3.2.15 with $M_\mathfrak{m}/\mathfrak{m}M_\mathfrak{m} = 0$ for all $\mathfrak{m} \in \operatorname{MaxSpec} A$. What we do have is that $M = 0$ iff $M_\mathfrak{m}/\mathfrak{m}M_\mathfrak{m} = 0$ for each ideal $\mathfrak{m}$ which is the contraction of a maximal ideal of B, as seen in the proof. However, if, e.g., $(A, \mathfrak{m}, B)$ is a local integral domain and $M = B$, then $M_\mathfrak{m}/\mathfrak{m}M_\mathfrak{m} = B/\mathfrak{m}B = 0$, but $M \neq 0$.

***Proposition* I.3.2.17** (injectivity/surjectivity is local)**.** — *Let $\phi\colon M \to N$ be an A-module homomorphism. Then the following are equivalent:*

(i) ϕ *is injective (resp. surjective);*
(ii) $\phi_\mathfrak{p}\colon M_\mathfrak{p} \to N_\mathfrak{p}$ *is injective (resp. surjective) for all* $\mathfrak{p} \in \operatorname{Spec} A$*;*
(iii) $\phi_\mathfrak{m}\colon M_\mathfrak{m} \to N_\mathfrak{m}$ *is injective (resp. surjective) for all* $\mathfrak{m} \in \operatorname{MaxSpec} A$*.*

Proof. — We only prove the statement about injectivity because surjectivity works by reversing all arrows. Since (i) means that $0 \to M \to N$ is exact, $0 \to M_\mathfrak{p} \to N_\mathfrak{p}$ is again exact. This implies (ii). Obviously, (ii) implies (iii).

Now, assume that (iii) holds. Since $0 \to \ker\phi \to M \to N$ is exact, $0 \to (\ker\phi)_\mathfrak{m} \to M_\mathfrak{m} \to N_\mathfrak{m}$ is again exact for all maximal ideals $\mathfrak{m}$. By hypothesis, we have $(\ker\phi)_\mathfrak{m} \cong \ker\phi_\mathfrak{m} = 0$, and thus $\ker\phi = 0$ by proposition I.3.2.13. This shows (i).

§ I.4. ASSOCIATED PRIMES AND PRIMARY DECOMPOSITION

Primary decomposition due to Noether was a corner stone of ideal theory. It provides the algebraic analogue of geometrically decomposing an algebraic variety into irreducible components. In the context of algebraic number theory, it generalises the prime factorisation of integers. Currently, however, commutative algebra emphasises localisation, so primary decomposition lost its importance to the notion of *associated primes* by Bourbaki.

***Motivation* I.4.0.1.** — Recall the prototypes for commutative rings: $\mathbb{Z}$ and polynomial rings $k[x_1, \dots, x_n]$ over a field k. Both are factorisation domains, but general rings do not admit prime factorisations, e.g., in $\mathbb{Z}[\sqrt{-5}]$ we have $6 = 2\cdot 3 = (1+\sqrt{-5})(1-\sqrt{-5})$. We can, however, generalise unique prime factorisations to unique factorisations of *ideals* in a wide class of rings, namely noetherian rings. In this sense, prime ideals take role of prime numbers, and *primary* ideals take the role of prime powers.

Definition* I.4.0.2.** — *An ideal $\mathfrak{q}$ in a ring A is* ***primary *if $\mathfrak{q} \neq (1)$ and if $xy \in \mathfrak{q}$ implies $x \in \mathfrak{q}$ or $y^n \in \mathfrak{q}$ for some $n > 0$. In other words, $\mathfrak{q}$ is primary iff every zero divisor in $A/\mathfrak{q} \neq 0$ is nilpotent.*

***Fact* I.4.0.3.** — *The contraction of a primary ideal is primary.*

Proof. — Let $f\colon A \to B$ be a ring map, and let $\mathfrak{q}$ be a primary ideal of B. Then f induces an inclusion of rings $A/f^{-1}(\mathfrak{q}) \hookrightarrow B/\mathfrak{q}$. One now easily checks that $A/f^{-1}(\mathfrak{q}) \neq 0$ has only nilpotent zero divisors.

***Fact* I.4.0.4.** — *Let $\mathfrak{q}$ be a primary ideal of a ring A. Then $\sqrt{\mathfrak{q}}$ is the smallest prime ideal containing $\mathfrak{q}$.*

Proof. — Because of proposition I.1.3.17, we only need to show that $\sqrt{\mathfrak{q}}$ is prime. Let $xy \in \sqrt{\mathfrak{q}}$, that is, $(xy)^m \in \mathfrak{q}$ for some m, so $x^m \in \mathfrak{q}$ or $y^{mn} \in \mathfrak{q}$ for some n. Hence, $x \in \sqrt{\mathfrak{q}}$ or $y \in \sqrt{\mathfrak{q}}$.

***Definition* I.4.0.5.** — *A primary ideal $\mathfrak{q}$ is said to be $\mathfrak{p}$-**primary** if $\mathfrak{p} = \sqrt{\mathfrak{q}}$. In this case, $xy \in \mathfrak{q}$ implies $x \in \mathfrak{q}$ or $y \in \mathfrak{p}$.*

***Example* I.4.0.6.** — (i) Prime ideals are primary.
(ii) The primary ideals of $\mathbb{Z}$ are (0) and (p^n) for a prime p: these are the only ideals with prime radical and are indeed primary.

***Warning* I.4.0.7.** — Primary ideals are not necessarily prime powers: consider $A = k[x, y]$ and $\mathfrak{q} = (x, y^2)$. Then $A/\mathfrak{q} \cong k[y]/(y^2)$ in which the zero divisors are multiples of y and hence nilpotent. Thus, $\mathfrak{q}$ is $\mathfrak{p}$-primary with radical $\mathfrak{p} = (x, y)$. Since $\mathfrak{p}^2 \subsetneq \mathfrak{q} \subsetneq \mathfrak{p}$, primary ideals are not necessarily prime powers.

Conversely, prime powers are not necessarily primary: consider $A = k[x, y, z]/(xy - z^2)$. Then $\mathfrak{p} = (\bar{x}, \bar{z})$ is prime as $A/\mathfrak{p} \cong k[y]$ is an integral domain. We have $\bar{x}\bar{y} = \bar{z}^2 \in \mathfrak{p}^2$, but $\bar{x} \notin \mathfrak{p}^2$ and $\bar{y} \notin \mathfrak{p} = \sqrt{\mathfrak{p}^2}$, whence $\mathfrak{p}^2$ is not primary.

***Proposition* I.4.0.8.** — *If $\sqrt{\mathfrak{q}}$ is maximal, then $\mathfrak{q}$ is a primary ideal. In particular, if $\mathfrak{m}$ is a maximal ideal, then $\mathfrak{m}^n$ is $\mathfrak{m}$-primary for all $n > 0$.*

Proof. — The image of $\sqrt{\mathfrak{q}}$ in $A/\mathfrak{q}$ is nil $A/\mathfrak{q}$, which is maximal, so $A/\mathfrak{q}$ has only one prime ideal by proposition I.1.3.17. Therefore, $(A/\mathfrak{q}, \sqrt{\mathfrak{q}})$ is a local ring, and hence every element of $A/\mathfrak{q}$ is either a unit or nilpotent. Thus, every zero divisor of $A/\mathfrak{q}$ is nilpotent and $\mathfrak{q}$ is primary.

***Lemma* I.4.0.9.** — *If $\mathfrak{q}_1, \dots, \mathfrak{q}_n$ are $\mathfrak{p}$-primary ideals, then $\bigcap_{i=1}^n \mathfrak{q}_i$ is $\mathfrak{p}$-primary.*

Proof. — Let $xy \in \bigcap_{i=1}^n \mathfrak{q}_i$ and $y \notin \bigcap_{i=1}^n \mathfrak{q}_i$. Then $y \notin \mathfrak{q}_i$ for some i and $xy \in \mathfrak{q}_i$. Since $\mathfrak{q}_i$ is primary, we have $x \in \mathfrak{p}$. Moreover, $\sqrt{\bigcap_{i=1}^n \mathfrak{q}_i} = \bigcap_{i=1}^n \sqrt{\mathfrak{q}_i} = \mathfrak{p}$. Thus, $\bigcap_{i=1}^n \mathfrak{q}_i$ is $\mathfrak{p}$-primary.

***Lemma* I.4.0.10.** — *Let $\mathfrak{q}$ be a $\mathfrak{p}$-primary ideal of A, and let $x \in A$. Then:*

(i) *if $x \in \mathfrak{q}$, then $(\mathfrak{q} : x) = (1)$;*
(ii) *if $x \notin \mathfrak{q}$, then $(\mathfrak{q} : x)$ is $\mathfrak{p}$-primary;*
(iii) *$x \notin \mathfrak{p}$, then $(\mathfrak{q} : x) = \mathfrak{q}$.*

Proof. — (i) and (iii) are clear by definition.
For (ii), let $y \in (\mathfrak{q} : x)$, that is, $xy \in \mathfrak{q}$. Since $x \notin \mathfrak{q}$ and since $\mathfrak{q}$ is $\mathfrak{p}$-primary, we then have $y \in \mathfrak{p}$. Hence, $\mathfrak{q} \subseteq (\mathfrak{q} : x) \subseteq \mathfrak{p}$, and taking radicals yields $\sqrt{(\mathfrak{q} : x)} = \mathfrak{p}$. Moreover, let $yz \in (\mathfrak{q} : x)$ with $y \notin \mathfrak{p}$. Then $xyz \in \mathfrak{q}$, so $xz \in \mathfrak{q}$ as $\mathfrak{q}$ is $\mathfrak{p}$-primary. Thus, $z \in (\mathfrak{q} : x)$, that is $(\mathfrak{q} : x)$ is $\mathfrak{p}$-primary.

Definition* I.4.0.11.** — *Let $\mathfrak{a}$ be an ideal of a ring A. A* ***primary decomposition *of $\mathfrak{a}$ is a representation of $\mathfrak{a}$ as a finite intersection $\mathfrak{a} = \bigcap_{i=1}^n \mathfrak{q}_i$ of primary ideals $\mathfrak{q}_i$ of A. This decomposition is generally neither unique nor existent.*[I.10] *If, moreover,*

[I.10] A ring A in which every ideal has a primary decomposition is called a *Lasker ring*.

(i) *the* $\sqrt{\mathfrak{q}_i}$ *are all distinct;*
(ii) *no* $\mathfrak{q}_i$ *can be omitted, i. e.,* $\mathfrak{a} \subsetneq \bigcap_{j \neq i} \mathfrak{q}_j$ *for all* i*;*

the primary decomposition is ***minimal*** *or* ***irredundant*** *(or* reduced*, or* normal*).*

Remark **I.4.0.12.** — Given a primary decomposition $\mathfrak{a} = \bigcap_{i=1}^n \mathfrak{q}_i$, we can always reduce it to a minimal one: by lemma I.4.0.9, we can group terms in the intersection so that the $\sqrt{\mathfrak{q}_i}$ are distinct, and then we can omit all superfluous terms.

Theorem **I.4.0.13** (first uniqueness theorem)**.** — *Let* $\mathfrak{a}$ *be an ideal of a ring* A*, and let* $\mathfrak{a} = \bigcap_{i=1}^n \mathfrak{q}_i$ *be a minimal primary decomposition of* $\mathfrak{a}$*. Then the set of all prime ideals* $\sqrt{\mathfrak{q}_i}$ *for* $i = 1, \dots, n$ *is the set* $\{\sqrt{(\mathfrak{a} : x)} \mid x \in A\}$ *and hence independent of the particular decomposition.*

I.4.1. Associated primes.

Definition **I.4.1.1.** — *Let* M *be an* A*-module. A prime ideal* $\mathfrak{p}$ *of* A *is an* ***associated prime*** *of* M *if* $\mathfrak{p} = \operatorname{ann} x$ *for some* $x \in M$*. The set of all associated primes of* M *is denoted by* $\operatorname{Ass}_A M = \operatorname{Ass} M$*.*

Proposition **I.4.1.2.** — *Let* A *be a noetherian ring, and let* M *be a non-zero* A*-module.*

(i) *Every maximal element of the set* $\{\operatorname{ann} x \mid x \in M \setminus \{0\}\}$ *is an associated prime of* M*. In particular,* $\operatorname{Ass} M$ *is non-empty.*
(ii) *The set of all* $a \in A$ *such that* $ax = 0$ *for some* $x \in M \setminus \{0\}$ *is the union of all associated primes of* M*.*

Proof. — For (i), we must show that if $\operatorname{ann} x$ is maximal among all proper annihilators then it is prime. Let $ab \in \operatorname{ann} x$ with $b \notin \operatorname{ann} x$, that is, $abx = 0$ but $bx \neq 0$. Since $\operatorname{ann}(bx) = \operatorname{ann} x$ by maximality of $\operatorname{ann} x$, we have $ax = 0$, that is, $a \in \operatorname{ann} x$.

For (ii), if $ax = 0$ for some $x \neq 0$, then $a \in \operatorname{ann} x$. By (i), there is an associated prime of M containing $\operatorname{ann} x$. ◻

Proposition **I.4.1.3.** — *Let* S *be a multiplicative subset of a ring* A*, and identify* $\operatorname{Spec}(S^{-1}A)$ *as a subset of* $\operatorname{Spec} A$*. If* N *is an* $S^{-1}A$*-module, then* $\operatorname{Ass}_A N = \operatorname{Ass}_{S^{-1}A} N$*. If* A *is noetherian and if* M *is an* A*-module, then* $\operatorname{Ass}(S^{-1}M) = \operatorname{Ass} M \cap \operatorname{Spec}(S^{-1}A)$*.*

Proof. — Since $\operatorname{ann}_A x = (\operatorname{ann}_{S^{-1}A} x)^c$ and $\operatorname{ann}_{S^{-1}A} x = S^{-1}(\operatorname{ann}_A x)$ for all $x \in N$, the equality $\operatorname{Ass}_A N = \operatorname{Ass}_{S^{-1}A} N$ follows from the ideal correspondence for localisation.

For the second statement, let $\mathfrak{p} := \operatorname{ann}_A x \in \operatorname{Ass} M \cap \operatorname{Spec}(S^{-1}A)$ where $x \in M$, which means that $\mathfrak{p}$ is disjoint from S. Suppose that $(a/s)x = 0$ in $S^{-1}M$. Then $tax = 0$ for some $t \in S$, so $ta \in \mathfrak{p}$ but $t \notin \mathfrak{p}$. Since $\mathfrak{p}$ is prime, we obtain $a \in \mathfrak{p}$, and thus $\operatorname{ann}_{S^{-1}A} x = S^{-1}\mathfrak{p}$ and $S^{-1}\mathfrak{p} \in \operatorname{Ass}(S^{-1}M)$. Conversely, let $\mathfrak{q} := \operatorname{ann}_{S^{-1}A} x \in \operatorname{Ass}(S^{-1}M)$ where $x \in M$. Then $\mathfrak{q}^c$ is finitely generated as A is noetherian, say, $\mathfrak{q}^c = (a_1, \dots, a_n)$. Since $\mathfrak{q} = S^{-1}\mathfrak{q}^c$, we have $(a_i/s_i)x = 0$ for suitable $s_i \in S$. It follows that there are $t_i \in S$ such that $t_i a_i x = 0$, and hence $\mathfrak{q}^c = \operatorname{ann}_A(tx)$ where $t = \prod_{i=1}^n t_i$. Thus, $\mathfrak{q}^c \in \operatorname{Ass} M$. ◻

§ I.5. CHAIN CONDITIONS

Chain conditions are a property with which we can impose some finiteness conditions on modules and rings. Under these conditions, we can make more profound statements. There is a symmetry between ascending and descending chain conditions as they are dual. However, this symmetry disappears if we consider rings.

I.5.1. Noetherian and artinian modules.

Definition* I.5.1.1.** — *Let Σ be a partially ordered set. If any ascending chain $x_1 \leq x_2 \leq \cdots$ in Σ* stabilises *or* becomes stationary*, i. e., if there is some n such that $x_n = x_{n+1} = \ldots$, then Σ satisfies the* ***ascending chain condition*. Dually, if any descending chain $x_1 \geq x_2 \geq \cdots$ in Σ stabilises, then Σ satisfies the* ***descending chain condition****.*

***Fact* I.5.1.2.** — *Let Σ be a partially ordered set. Then it satisfies the ascending chain condition iff any non-empty subset of Σ has a maximal element.*

Proof. — Suppose that Σ has the ascending chain condition, and let Σ' be any non-empty subset of Σ which does not have a maximal element. Then for each $x \in \Sigma'$, there is some $\tilde{x} \in \Sigma'$ such that $x < \tilde{x}$. Now pick any $x \in \Sigma'$, and consider the ascending chain $x < \tilde{x} < \tilde{\tilde{x}} < \cdots$ which does not stabilise, a contradiction.

Conversely, suppose that any non-empty subset of Σ has a maximal element. Let $x_1 \leq x_2 \leq \cdots$ be an ascending chain in Σ. Then $\{x_1, x_2, \ldots\}$ has a maximal element, say x_n. Thus $x_n \leq x_{n+1} \leq \ldots \leq x_n$, i. e., the chain stabilises at x_n. ∎

Definition* I.5.1.3** (Noether 1921)**.** — *An A-module M is* ***noetherian *(resp.* ***artinian****) if the set of submodules of M ordered by inclusion satisfies the ascending (resp. descending) chain condition. The same notion applies to rings A if considered as an A-module, i. e., if the set of ideals ordered by inclusion satisfies the ascending (resp. descending) chain condition.*

***Proposition* I.5.1.4** (characterisation of noetherian modules)**.** — *An A-module M is noetherian iff every submodule of M is finitely generated.*

Proof. — Assume that M is noetherian. Let N be a submodule of M, and let Σ be the set of all finitely generated submodules of N. Then Σ is non-empty since $0 \in \Sigma$, so it has a maximal element, say N_0 (no need for Zorn's lemma). As $N_0 + Ax \in \Sigma$ for all $x \in N$, we must have $N_0 = N_0 + Ax$ for all $x \in N$. In other words, $N = N_0$ is finitely generated.

Conversely, assume that every submodule of M is finitely generated. Let $M_1 \subseteq M_2 \subseteq \cdots$ be an ascending chain of submodules of M. Then $N = \bigcup_{n=1}^{\infty} M_n$ is a submodule of M which is finitely generated by, say, elements $x_1, \ldots, x_n$. There must exist some M_n containing these elements, whence $N = M_n$ and the chain stabilises at M_n. ∎

Proposition I.5.1.4 is the reason that the noetherian property is, as opposed to the artinian property, the right finiteness condition to make many theorems work.

***Proposition* I.5.1.5.** — *Let $0 \to M' \to M \to M'' \to 0$ be a short exact sequence of A-modules. Then M is noetherian (resp. artinian) iff M' and M'' are noetherian (resp. artinian). In particular, the quotient of a noetherian (resp. artinian) ring is again noetherian (resp. artinian).*

Proof. — We will only show the noetherian case as the artinian case is analogous. Suppose that M is noetherian. Then any ascending chain in M' is also one in M and hence stationary. The same is true for M'' by virtue of proposition I.2.2.5.

For the converse, we make the following observation. Let N be a submodule of M. If two submodules $N_1 \subseteq N_2$ of M satisfy $N_1 \cap N = N_2 \cap N$ and $N_1 \bmod N = N_2 \bmod N$, then $N_1 = N_2$. Indeed, for any $x \in N_2$ we have $x + n \in N_1 \subseteq N_2$ for some $n \in N$, whence $n \in N_2 \cap N \subseteq N_1$ and thus, $x \in N_1$.

So, suppose that M' and M'' are noetherian, and let $f\colon M' \to M$ and $g\colon M \to M''$ be the maps in the given exact sequence. If $(M_n)_{n\geq 1}$ is an ascending chain of submodules of M, then $(f^{-1}(M_n))_n$ is a ascending chain in M' and $(g(M_n))_n$ one in M''. Since both stabilise, the original chain $(M_n)_n$ must stabilise by the above observation, too.

In the case of rings, note that any chain of ideals of $A/\mathfrak{a}$ is also a chain of A-submodules of $A/\mathfrak{a}$. So, A being noetherian or artinian implies that $A/\mathfrak{a}$ has the same property as a ring. ∎

***Corollary* I.5.1.6.** — *If $M_1, \dots, M_n$ are noetherian (resp. artinian) A-modules, then so is $\bigoplus_{i=1}^n M_i$. In particular, if A is a noetherian (resp. artinian) ring, then:*

(i) *any finitely generated A-module is again noetherian (resp. artinian);*
(ii) *any finite A-algebra is again noetherian (resp. artinian) as a ring.*

Proof. — The first statement follows by induction on n and by applying proposition I.5.1.5 to the short exact sequence $0 \to M_n \to \bigoplus_{i=1}^n M_i \to \bigoplus_{i=1}^{n-1} M_i \to 0$.

(i) follows from the statement above and from the fact that M is the quotient of A^n for some n. (ii) follows since a finite A-algebra B is noetherian (resp. artinian) as an A-module by (i), so also as a B-module by restriction of scalars. ∎

***Example* I.5.1.7.** — (i) Any finite abelian group (as a $\mathbb{Z}$-module) is both noetherian and artinian.

(ii) Any field is both noetherian and artinian.

(iii) Any principal ideal domain is noetherian by proposition I.5.1.4.

(iv) In particular, the integers $\mathbb{Z}$ and the polynomial ring $k[x]$ over a field k are noetherian. However, they are both not artinian: for any non-zero element x, we have strict inclusions $(x) \supsetneq (x^2) \supsetneq (x^3) \supsetneq \cdots$.

(v) The quotient ring $\mathbb{Z}/(n)$ for $n > 0$ is both noetherian and artinian.

(vi) The Gaussian integers $\mathbb{Z}[i]$ are noetherian by corollary I.5.1.6. More generally, the ring of integers for any algebraic number field is noetherian.

Insert reference to theorem that $[\mathcal{O}_K : \mathbb{Z}]$ is finite for any $K/\mathbb{Q}$.

(vii) Fix a prime number p. Let $G := \mathbb{Z}[p^{-1}]/\mathbb{Z}$ be the subgroup of $\mathbb{Q}/\mathbb{Z}$ of those elements whose order is a power of p. Then $G_n = \langle p^{-n} \rangle$ is the only subgroup of order p^n for each $n \geq 0$, and these G_n are all proper subgroups of G. Since we have $G_0 \subsetneq G_1 \subsetneq \cdots$, we see that G is artinian, but not noetherian.

(viii) Continuing the notation from (vii), consider the group $H = \{m/p^n \mid m \in \mathbb{Z}, n \geq 0\}$. Then proposition I.5.1.5 shows that H is neither noetherian nor artinian, for there is a short exact sequence $0 \to \mathbb{Z} \hookrightarrow H \twoheadrightarrow G \to 0$.

(ix) The polynomial ring $k[x_1, x_2, \dots]$ in infinitely many indeterminates is neither noetherian nor artinian: we have infinite sequences $(x_1) \subsetneq (x_1, x_2) \subsetneq \cdots$ and $(x_1) \supsetneq (x_1^2) \supsetneq \cdots$.

(x) Let X be an infinite compact Hausdorff space, and let $C(X)$ be the ring of real-valued continuous functions on X. Take a decreasing sequence of closed sets $F_1 \supsetneq F_2 \supsetneq \cdots$ in X, and set $\mathfrak{a}_n := \{f \in C(X) \mid f(F_n) = 0\}$. Then $C(X)$ is not noetherian because $\mathfrak{a}_1 \supsetneq \mathfrak{a}_2 \supsetneq \cdots$ is a strictly decreasing chain of ideals.

***Warning* I.5.1.8.** — If a ring A is noetherian or artinian, then a subring of A does not necessarily have the same property (but quotient rings of A do). For example, the integral domain $k[x_1, x_2, \dots]$ is neither noetherian nor artinian (example I.5.1.7), although the field $\operatorname{Quot} k[x_1, x_2, \dots]$ certainly is.

I.5.2. Length of modules.

Here an example from representation theory.

***Definition* I.5.2.1.** — *An A-module $M \neq 0$ is **simple** if the only submodules of M are the trivial ones: 0 and M itself.*

***Observation* I.5.2.2** (classification of simple modules)**.** — In particular, for any $x \in M \setminus \{0\}$, we have $M = (x)$. Note that there is an isomorphism of A-modules $Ax \cong A/\operatorname{ann} x$, which is simple iff $\operatorname{ann} x$ is a maximal ideal of A by ideal correspondence. To summarise, up to isomorphism, the simple A-modules are precisely $A/\mathfrak{m}$ for $\mathfrak{m}$ a maximal ideal of A.

Definition* I.5.2.3.** — *Let M be an A-module. A strictly descending chain $M = M_0 \supsetneq \cdots \supsetneq M_n = 0$ of submodules of M is a **composition series** of M if every quotient M_i/M_{i+1} is simple,* *i. e., if no other submodules of M fit into this chain. Then n is the* ***length *of this composition series.*

***Proposition* I.5.2.4.** — *Suppose that a module M has a composition series of length n. Then every composition series of M has length n, and any strictly descending chain in M can be extended to a composition series.*

Proof. — Let $l(M)$ be the smallest length of a composition series of a module M. We prove two statements first.

(i) *If $N \subsetneq M$, then $l(N) < l(M)$.* Let (M_i) be a composition series of length $l(M)$, and consider the submodules $N_i := N \cap M_i$ of N. Then N_{i-1}/N_i is contained in the simple quotient M_{i-1}/M_i, so either it is M_{i-1}/M_i or we have $N_{i-1} = N_i$. Removing repeated terms N_i, we obtain a composition series of N, whence $l(N) \leq l(M)$. If $l(N) = l(M) =: n$, then $N_{i-1}/N_i = M_{i-1}/M_i$ for all i. In particular, $M_n = N_n = 0$, then $M_{n-1} = N_{n-1}$, then $M_{n-2} = N_{n-2}$, and so on, until $M = M_0 = N_0 = N$.

(ii) *Any strictly descending chain in M has length at most $l(M)$.* Let $M = M_0 \supsetneq \cdots \supsetneq M_k = 0$ be a chain of length k. By (i), we have $l(M) = l(M_0) > \cdots > l(M_k) = 0$, which are k number below $l(M)$, so $k \leq l(M)$.

By minimality of $l(M)$ and by (ii), any composition series of M must have length $l(M)$. Conversely, any strictly descending chain which is not a composition series can be refined by inductively inserting more terms. This process has to stop by (ii), at which point we obtain a composition series. ∎

Definition* I.5.2.5.** — *The* ***length $\operatorname{len}_A M = \operatorname{len} M$ *of an A-module M is the length of any composition series of M. If M has no composition series, then $\operatorname{len} M := \infty$. If $\operatorname{len} M < \infty$, then we say that M has* ***finite length****, and otherwise* ***infinite length****.*

There is an even stronger statement of proposition I.5.2.4.

***Jordan-Hölder theorem* I.5.2.6.** — *Let M be a module with two composition series $M = M_0 \supsetneq \cdots \supsetneq M_m = 0$ and $M = N_0 \supsetneq \cdots \supsetneq N_n = 0$. Then $m = n$ and there exists a permutation σ of $\{1, \ldots, m\}$ such that $M_{i-1}/M_i \cong N_{\sigma(i)-1}/N_{\sigma(i)}$ for all i. In other words, composition series, if they exist, are unique up to isomorphism and permutation of quotients.*(I.11)

Proof. — Let $1 \leq i \leq n$ be the largest index such that $N_{i-1} \supseteq M_{m-1}$. Then the induced map $M_{m-1}/M_m = M_{m-1} \to N_{i-1}/N_i$ is both injective and surjective since both source and target are simple, and so we set $\sigma(m) := i$. It follows that $N_{i-1} = N_i + M_{m-1} \supsetneq N_i$ and, by modularity, that

$$N_{j-1} \supsetneq M_{m-1} \cap N_{j-1} = (M_{m-1} \cap N_{j-1}) + N_j = N_j$$

for all $j > i$ (otherwise, we would have $M_{m-1} = N_j$ for some $j \geq i$). Passing to quotients, we obtain the two chains

$$\begin{gathered} M_0/M_{m-1} \supsetneq \cdots \supsetneq M_{m-1}/M_{m-1}, \\ N_0/M_{m-1} \supsetneq \cdots \supsetneq N_{i-1}/M_{m-1} = (N_i + M_{m-1})/M_{m-1} \supsetneq \cdots \supsetneq (N_n + M_{m-1})/M_{m-1} \end{gathered}$$

from M/M_{m-1} to 0. The first chain has quotients $(M_{j-1}/M_{m-1})/(M_j/M_{m-1}) \cong M_{j-1}/M_j$ for $j < m$, while the second chain similarly has quotients N_{j-1}/N_j for $j < i$ and

$$(N_{j-1} + M_{m-1})/(N_j + M_{m-1}) \cong N_{j-1}/(N_j + (M_{m-1} \cap N_{j-1})) = N_{j-1}/N_j$$

for $j > i$. By induction, we conclude $m = n$. ∎

(I.11) This was originally proved in more generality for non-abelian groups by what are called the *Schreier refinement theorem* and the *Zassenhaus lemma*. However, the proof becomes more complicated as we have to check normality of subgroups.

***Proposition* I.5.2.7.** — *A module has finite length iff it is both noetherian and artinian.*

Proof. — If a module M has finite length, then every chain of M has length bounded by $\operatorname{len} M$, so M is both noetherian and artinian.

Conversely, if M is noetherian and artinian, then we construct a composition series as follows. Start with $M_0 := M$. For each $i \geq 0$, pick any maximal submodule M_{i+1} of M_i, which is possible since M_i is noetherian. We obtain a strictly descending chain $M_0 \supsetneq M_1 \supsetneq \cdots$ which must stabilise since M is artinian. ◻

***Fact* I.5.2.8.** — *The function $M \mapsto \operatorname{len} M$ is an additive function on the class of all A-modules of finite length.*

Proof. — Consider a short exact sequence

$$0 \longrightarrow M' \xrightarrow{f} M \xrightarrow{g} M'' \longrightarrow 0$$

of A-modules of finite length. The preimage of any composition series of M'' under g followed by the image of any composition series of M' under f gives a composition series of M. This shows $\operatorname{len} M = \operatorname{len} M' + \operatorname{len} M''$. ◻

***Proposition* I.5.2.9.** — *For a vector space, the following are equivalent:*

(i) *it is finite-dimensional;*
(ii) *it is of finite length;*
(iii) *it is noetherian;*
(iv) *it is artinian.*

In this case, length and dimension coincide.

Proof. — If V is a finite-dimensional vector space, then we obtain a composition series by successively deleting basis vectors, so V is of length $\dim V < \infty$. Moreover, that finite length implies the noetherian and artinian condition is proposition I.5.2.7. It remains to see that V being noetherian or artinian implies that it is finite-dimensional. Suppose that V is infinite-dimensional, i. e., there are infinitely many linearly independent elements $x_n \in V$ for $n \geq 1$. Then $(\operatorname{span}(x_1, \dots, x_n))_{n\geq1}$ (resp. $(\operatorname{span}(x_n, x_{n+1}, \dots))_{n\geq1}$) is an infinite strictly ascending (resp. descending) chain of subspaces. Hence, V is neither noetherian nor artinian. ◻

***Example* I.5.2.10.** — Let $\mathfrak{m}$ be a finitely generated maximal ideal of a ring A. Then $V_i := \mathfrak{m}^{i-1}/\mathfrak{m}^i$ for $i \geq 1$ is an $A/\mathfrak{m}$-vector space since it is annihilated by $\mathfrak{m}$, and it is finite-dimensional since the images of the finitely many generators of $\mathfrak{m}$ will span V_i. Hence, $\operatorname{len}_A V_i = \dim_{A/\mathfrak{m}} V_i < \infty$ by proposition I.5.2.9 (the A-submodules and the $A/\mathfrak{m}$-subspaces of V_i coincide). The short exact sequences $0 \to V_i \to A/\mathfrak{m}^i \to A/\mathfrak{m}^{i-1} \to 0$ show that $\operatorname{len}(A/\mathfrak{m}^n) = \sum_{i=1}^n \operatorname{len} V_i < \infty$. The function $n \mapsto \operatorname{len} A/\mathfrak{m}^n$ plays a crucial role in algebraic geometry, more specifically in the resolution of singularities.

***Lemma* I.5.2.11.** — *Let M be a noetherian (resp. artinian) A-module. Then a surjective (resp. injective) A-linear map $\phi\colon M \to M$ is an automorphism.*

Proof. — In the noetherian case, consider the ascending chain $\ker\phi \subseteq \ker\phi^2 \subseteq \cdots$ of submodules of M, which stabilises at, say, $\ker\phi^n$. For $x \in \ker\phi^n$, there exists $y \in M$ with $\phi^n(y) = x$ by hypothesis. It follows that $y \in \ker\phi^{2n} = \ker\phi^n$ and hence, $x = \phi^n(y) = 0$. Thus, $\ker\phi \subseteq \ker\phi^n = 0$.

In the artinian case, consider the descending chain $\operatorname{im}\phi \supseteq \operatorname{im}\phi^2 \supseteq \cdots$ of submodules of M, which stabilises at, say $\operatorname{im}\phi^n$. For $x \in M$, we have $\phi^n(x) \in \operatorname{im}\phi^n = \operatorname{im}\phi^{2n}$, so there exists $y \in M$ with $\phi^{2n}(y) = \phi^n(x)$. By hypothesis, we have $x = \phi^n(y) \in \operatorname{im}\phi^n$, whence $M = \operatorname{im}\phi^n \subseteq \operatorname{im}\phi$. ◻

I.5.3. Noetherian rings.

I.5.3.1. — Recall from proposition I.5.1.4 that a ring is noetherian iff every ideal is finitely generated.

***Lemma* I.5.3.2.** — *Any localisation $S^{-1}A$ of a noetherian ring A is noetherian. In particular, $A_{\mathfrak{p}}$ is noetherian for all prime ideals $\mathfrak{p}$ of a noetherian ring A.*

First proof. — By proposition I.3.1.17, the ideals of $S^{-1}A$ are all extended ideals, which are, quite generally, in inclusion preserving correspondence with contracted ideals of A. The latter satisfy the ascending chain condition by hypothesis, so $S^{-1}A$ is noetherian. ◻

Second proof. — Any ideal $\mathfrak{a}$ of A is finitely generated, say, by $x_1, \dots, x_n$, whence $S^{-1}\mathfrak{a}$ is finitely generated by $x_1/1, \dots, x_n/1$. ◻

***Proposition* I.5.3.3** (I. S. Cohen). — *A ring in which every prime ideal is finitely generated is noetherian.*

Proof. — Suppose that A is a non-noetherian ring, and let Σ be the set of all ideals in A which are not finitely generated, ordered by inclusion. Since Σ is non-empty by assumption, it contains a maximal element $\mathfrak{a}$ by Zorn's lemma.

Suppose that $\mathfrak{a}$ is not prime, i. e., there are $x, y \notin \mathfrak{a}$ with $xy \in \mathfrak{a}$. As $\mathfrak{a} + (x)$ strictly contains $\mathfrak{a}$, there is a finitely generated ideal $\mathfrak{a}_0 \subseteq \mathfrak{a}$ such that $\mathfrak{a} + (x) = \mathfrak{a}_0 + (x)$. Observe that $\mathfrak{a} = \mathfrak{a}_0 + x(\mathfrak{a} : x)$: indeed, for any $z \in \mathfrak{a}$ there are $z_0 \in \mathfrak{a}_0$ and $a \in A$ such that $z = z_0 + ax$, but $z - z_0 \in \mathfrak{a}$ and hence $a \in (\mathfrak{a} : x)$. Since $(\mathfrak{a} : x)$ contains $\mathfrak{a} + (y)$, it is finitely generated. Therefore, $\mathfrak{a}$ is finitely generated, a contradiction. Thus, $\mathfrak{a}$ is an infinitely generated prime ideal in A. ◻

***Hilbert's basis theorem* I.5.3.4** (1890). — *If A is a noetherian ring, then $A[x]$ and $A[\![x]\!]$ are noetherian rings.*[I.12] *Then by induction, $A[x_1, \dots, x_n]$ and $A[\![x_1, \dots, x_n]\!]$ are noetherian, too.*

Proof. — Let $\mathfrak{a}$ be an ideal of $A[x]$; it suffices to show that $\mathfrak{a}$ is finitely generated. Let $\mathfrak{l}$ be the ideal of A generated by all leading coefficients a_r of polynomials $a_r x^r + a_{r-1}x^{r-1} + \dots \in \mathfrak{a}$. Since A is noetherian, write $\mathfrak{l} = (a_1, \dots, a_n)$, say. For each a_i, pick a corresponding polynomial $f_i = a_i x^{r_i} + \dots \in \mathfrak{a}$; set $r := \max\{r_1, \dots, r_n\}$. Define the ideal $\mathfrak{a}' := (f_1, \dots, f_n) \subseteq \mathfrak{a}$.

Let $f = ax^m + \dots \in \mathfrak{a}$ be arbitrary. If $m \geq r$, write $a = \sum_{i=1}^n u_i a_i \in \mathfrak{l}$ where $u_i \in A$. Then $f - \sum_{i=1}^n u_i f_i x^{m-r_i} \in \mathfrak{a}$ has smaller degree than m. Proceeding inductively, we arrive at some polynomial $\tilde{f} \in \mathfrak{a}$ of degree smaller than r such that $f = \tilde{f} + g$ for suitable $g \in \mathfrak{a}'$.

To summarise, we have shown that $\mathfrak{a} = (\mathfrak{a} \cap M) + \mathfrak{a}'$ as A-modules where $M := (1, x, \dots, x^{r-1})$ is an A-module. Since M is noetherian by corollary I.5.1.6, write $\mathfrak{a} \cap M = (g_1, \dots, g_m)$ as an A-submodule, say. Thus, $\mathfrak{a}$ is generated by the f_i and the g_j, as desired.

The proof for $A[\![x]\!]$ is almost identical to the one for $A[x]$, but considering the ideal $\mathfrak{l}$ of A generated by the coefficients a_r of *least degree* of polynomials $a_r x^r + a_{r+1}x^{r+1} + \dots \in \mathfrak{a}$. This time, the aim is to transform any $f \in \mathfrak{a}$ into a polynomial $\tilde{f} \in \mathfrak{a} \cap M$ of degree less than r. This is done inductively by subtracting (maybe infinitely many) elements in $\mathfrak{a}'$ in order to kill the monomials of f of degrees $r, r+1, \dots$ ◻

***Remark* I.5.3.5.** — The converse of Hilbert's basis theorem I.5.3.4 is also true: if $A[x_1, \dots, x_n]$ is noetherian, then $A \cong A[x_1, \dots, x_n]/(x_1, \dots, x_n)$ is noetherian by proposition I.5.1.5; similarly for $A[\![x_1, \dots, x_n]\!]$.

***Corollary* I.5.3.6.** — *If A is a noetherian ring, then any finitely generated A-algebra is noetherian as a ring (cf. corollary I.5.1.6). This applies, in particular, to finitely generated rings (over $\mathbb{Z}$) and finitely generated polynomial rings over a field.*

(I.12) Classically, only the statement for $A[x]$ is Hilbert's basis theorem.

Proof. — A finitely generated A-algebra is a quotient of a polynomial ring $A[x_1, \dots, x_n]$, which is noetherian by proposition I.5.1.5. ◻

***Artin-Tate lemma* I.5.3.7** (1951). — *Let A be a noetherian ring, and let $B \subseteq C$ be two A-algebras. If C is finitely generated as an A-algebra and finite as a B-module, then B is finitely generated as an A-algebra.*

Proof. — Let $x_1, \dots, x_m$ generate C as an A-algebra, and let $C = (y_1, \dots, y_n)$ as a B-module. Then we can express $x_i = \sum_j b_{ij} y_j$ and $y_i y_j = \sum_k b_{ijk} y_k$ with coefficients $b_{ij}, b_{ijk} \in B$. Let B_0 be the A-subalgebra of B generated by these coefficients, which is noetherian as a ring by corollary I.5.3.6.

Any element of C is a polynomial in the x_i with coefficients in A. Making repeated use of the expressions for x_i and $y_i y_j$, we see that C is finite as a B_0-module. Since B_0 is noetherian, $B \subseteq C$ is also finite as a B_0-module. Thus, B is a finitely generated as an A-algebra because B_0 is. ◻

***Zariski's lemma* I.5.3.8** (1947). — *Let k be a field, and let E be a finitely generated k-algebra. If E is a field, then E/k is a finite extension.*

Proof. — Let $x_1, \dots, x_n$ be generators of E, so that $E = k[x_1, \dots, x_n]$. Suppose that E/k is not algebraic; w. l. o. g., let $x_1, \dots, x_r$ with $r \geq 1$ be algebraically independent over k, and let $x_{r+1}, \dots, x_n$ be algebraic over k and hence over the field $F := k(x_1, \dots, x_r)$. Hence, E/F is a finitely generated algebraic and hence finite extension, i. e., E is a finite F-module. By Artin-Tate lemma I.5.3.7, F is a finitely generated k-algebra, say $F = k[y_1, \dots, y_s]$ where $y_i = f_i/g_i$ for $f_i, g_i \in k[x_1, \dots, x_r]$. Now, consider the unique factorisation domain $k[x_1, \dots, x_r]$. The element $h = g_1 \cdots g_s + 1$ in this domain is not divisible by any g_i, whence the element $h^{-1} \in F$ cannot a polynomial in $k[y_1, \dots, y_s]$, a contradiction. Therefore, E/k is finitely generated algebraic and thus finite. ◻

A somewhat similar statement to the Artin-Tate lemma I.5.3.7 is the Eakin-Nagata theorem I.5.3.12, for which there are many proofs. The main bulk of the proof presented here is lemma I.5.3.11 due to Formanek, which is based on the original ideas of Eakin and Nagata.

***Lemma* I.5.3.9.** — *If M is a noetherian A-module, then $A/\operatorname{ann} M$ is a noetherian ring. In particular, if a ring A has a faithful noetherian module, then it is noetherian.*

Proof. — Note that M is also noetherian as a module over $\bar{A} := A/\operatorname{ann} M$ because the A-submodules and $\bar{A}$-submodules of M coincide. Write $M = (m_1, \dots, m_n)$, say. Then the $\bar{A}$-linear map $\bar{A} \hookrightarrow M^n$, $\bar{a} \mapsto (am_i)_i$ is injective, whence $\bar{A}$ is noetherian over itself by proposition I.5.1.5. ◻

***Warning* I.5.3.10.** — The artinian analogue of lemma I.5.3.9 is false. Recall from example I.5.1.7 that the $\mathbb{Z}$-module $G := \mathbb{Z}[p^{-1}]/\mathbb{Z}$ is artinian. However, $\operatorname{ann} G = 0$ and $\mathbb{Z}$ is not artinian.

***Lemma* I.5.3.11** (Formanek 1973). — *Let M be a finitely generated and faithful A-module. Assume that the submodules $\mathfrak{a}M$, where $\mathfrak{a}$ is an ideal of A, satisfy the ascending chain condition. Then A is noetherian.*

Proof. — By lemma I.5.3.9, it suffices to show that M is noetherian. By way of contradiction, suppose that it is not. Consider the set of all submodules $\mathfrak{a}M$ such that $M/\mathfrak{a}M$ is non-noetherian. Since this set contains the zero module, it contains a maximal element, say, $\mathfrak{a}_0 M$ by hypothesis. After replacing M with $M/\mathfrak{a}_0 M$ and A with $A/\operatorname{ann}(M/\mathfrak{a}_0 M)$, we may assume, w. l. o. g., that M is additionally non-noetherian, but $M/\mathfrak{a}M$ is noetherian for any non-zero ideal $\mathfrak{a}$ of A.

Next, consider the set Σ of all submodules N of M such that M/N is faithful. Clearly, $0 \in \Sigma$, so it is non-empty. When writing $M = (m_1, \dots, m_n)$, the set Σ is alternatively the set of all submodules N of M such that: for each $a \in A \setminus \{0\}$ we have $am_i \notin N$ for some i. From this, we see that any chain in Σ has an upper bound, and Zorn's lemma yields a maximal element $N_0 \in \Sigma$. If M/N_0 were noetherian, then A would be noetherian by lemma I.5.3.9, and then M would be noetherian by corollary I.5.1.6, contradicting our assumption. Therefore, after replacing M with M/N_0, w. l. o. g., we may assume the following about M:

(i) M is not noetherian;
(ii) $M/\mathfrak{a}M$ is noetherian for all non-zero ideals $\mathfrak{a}$ of A;
(iii) M/N is not faithful for all non-zero submodules N of M.

To finish, let N be any non-zero submodule of M. Then there exists some non-zero $a \in A$ such that $a(M/N) = 0$, i. e., such that $aM \subseteq N$. Since M/aM is noetherian, the submodule N/aM is finitely generated. Moreover, aM is finitely generated like M. It follows from proposition I.2.5.9 that N is finitely generated. Thus, M is noetherian, a contradiction. ∎

The following is a direct corollary and serves as the converse of corollary I.5.1.6 in the case of algebras.

***Eakin-Nagata theorem* I.5.3.12** (Eakin 1968, Nagata 1968). — *Let B be a finite A-algebra. If B is noetherian as a ring, then A is noetherian.*

Proof. — Since B has a unit, it is faithful as an A-module. The statement now follows from lemma I.5.3.11. ∎

***Remark* I.5.3.13.** — There are also many generalisations to non-commutative rings, e. g.:

(i) Let B be a non-commutative right-noetherian (i. e., the right ideals satisfy the ascending chain condition) ring, and let A be a commutative subring of B. If B is finitely generated as a left A-module, then A is noetherian.

(ii) Let B be a non-commutative ring whose two-sided ideals satisfy the ascending chain condition, and let A be a (commutative) subring contained in the centre of B. If B is finitely generated as a (two-sided) A-module, then A is noetherian.

These generalisations also follow from lemma I.5.3.11 after obvious modifications since B, as a ring, is faithful over A.

I.5.4. Artinian rings.

Artinian rings are the simplest kind of rings after a field. We will study them here not because of their usefulness but because of their simplicity.

***Proposition* I.5.4.1.** — *In an artinian ring, every prime ideal is maximal.*

Proof. — Let A be an artinian ring and let $\mathfrak{p}$ be a prime ideal of A. Then $B = A/\mathfrak{p}$ is an artinian integral domain. For any $x \in B \setminus \{0\}$, we have $(x^n) = (x^{n+1})$ for some n as B is artinian, and hence $x^n = x^{n+1}y$ for some $y \in B$. Since B is a domain, we obtain $xy = 1$. Hence, x has an inverse in B, therefore B is a field, and thus $\mathfrak{p}$ was maximal. ∎

***Corollary* I.5.4.2.** — *For an artinian ring A, we have* $\operatorname{nil} A = \operatorname{rad} A$.

Proof. — This follows from proposition I.1.3.17 and the definitions. ∎

***Proposition* I.5.4.3.** — *An artinian ring has only finitely many maximal ideals.*

Proof. — Consider the set of all finite intersections $\bigcap_{i=1}^{n} \mathfrak{m}_i$ where the $\mathfrak{m}_i$ are maximal ideals. This set has a minimal element, say, $\bigcap_{i=1}^{n} \mathfrak{m}_i$ as the ring is artinian. Hence, for any maximal ideal $\mathfrak{m}$ we have $\mathfrak{m} \cap \bigcap_{i=1}^{n} \mathfrak{m}_i = \bigcap_{i=1}^{n} \mathfrak{m}_i$, that is, $\mathfrak{m} \supseteq \bigcap_{i=1}^{n} \mathfrak{m}_i$. By the prime avoidance lemma I.1.3.9, we have $\mathfrak{m} \supseteq \mathfrak{m}_i$ for some i, and thus $\mathfrak{m} = \mathfrak{m}_i$. ∎

Define nilpoten

***Proposition* I.5.4.4.** — *In an artinian ring, the nilradical is nilpotent.*

First proof. — Let $\mathfrak{n}$ be the nilradical. We have $\mathfrak{n}^k = \mathfrak{n}^{k+1} = \cdots$ for some k; by way of contradiction, suppose that $\mathfrak{n}^k \neq (0)$. Consider the set of all ideals $\mathfrak{a}$ such that $\mathfrak{a}\mathfrak{n}^k \neq (0)$, which is non-empty since it contains $\mathfrak{n}$. Since the ring is artinian, let $\mathfrak{b}$ be a minimal element of this set. Then there is some $x \in \mathfrak{b}$ such that $x\mathfrak{n}^k \neq (0)$. As $(x) \subseteq \mathfrak{b}$, we have $(x) = \mathfrak{b}$ by minimality of $\mathfrak{b}$. However, $(x\mathfrak{n}^k)\mathfrak{n}^k = x\mathfrak{n}^{2k} = x\mathfrak{n}^k \neq (0)$ and $x\mathfrak{n}^k \subseteq (x)$, so $x\mathfrak{n}^k = (x)$ again by minimality, and hence $x = xy$ for some $y \in \mathfrak{n}^k$. However, $y \in \mathfrak{n}^k \subseteq \mathfrak{n}$ is nilpotent, whence $x = xy^n = 0$ for some n. This contradicts $x\mathfrak{n}^k \neq (0)$, and thus $\mathfrak{n}^k = (0)$.

Second proof. — By corollary I.5.4.2, we show that the Jacobson radical $\mathfrak{r}$ is nilpotent. We have $\mathfrak{r}^k = \mathfrak{r}^{k+1}$ for some k, so set $\mathfrak{a} := \operatorname{ann} \mathfrak{r}^k$. By way of contradiction, suppose that $\mathfrak{a} \neq (1)$. Then there is an ideal $\mathfrak{b}$ that is minimal among all ideals that strictly contain $\mathfrak{a}$. Since $\mathfrak{b} = \mathfrak{a} + (x)$ for some arbitrary $x \in \mathfrak{b} \setminus \mathfrak{a}$, the quotient module $\mathfrak{b}/\mathfrak{a} = (\bar{x})$ is finitely generated. Then corollary I.2.3.9 (Nakayama's lemma) says that $\mathfrak{b} \neq x\mathfrak{r} + \mathfrak{a}$. It follows that $x\mathfrak{r} + \mathfrak{a} = \mathfrak{a}$ by minimality of $\mathfrak{b}$, and hence $x \in (\mathfrak{a} : \mathfrak{r}) = ((0 : \mathfrak{r}^k) : \mathfrak{r}) = (0 : \mathfrak{r}^{k+1}) = \mathfrak{a}$, a contradiction. Thus, $\mathfrak{a} = \operatorname{ann} \mathfrak{r}^k = (1)$ and $\mathfrak{r}^k = (0)$.

***Lemma* I.5.4.5.** — *Let A be a ring in which $(0) = \mathfrak{m}_1 \cdots \mathfrak{m}_n$ is a product of (not necessarily distinct) maximal ideals of A. Then A is noetherian iff it is artinian.*

Proof. — By proposition I.5.2.7, it suffices to show that A is of finite length if it is noetherian or artinian. Consider the chain $A \supsetneq \mathfrak{m}_1 \supseteq \mathfrak{m}_1\mathfrak{m}_2 \supseteq \cdots \supseteq \mathfrak{m}_1 \cdots \mathfrak{m}_n = (0)$. For each $i = 1, \dots, n$, the quotient A-module $V_i := \mathfrak{m}_1 \cdots \mathfrak{m}_{i-1}/\mathfrak{m}_1 \cdots \mathfrak{m}_i$ is also an $A/\mathfrak{m}_i$-vector space, and its submodules and subspaces coincide. Each V_i is artinian or noetherian since A is, so proposition I.5.2.9 implies that $\operatorname{len} V_i < \infty$. By additivity of len, the short exact sequences $0 \to \mathfrak{m}_1 \cdots \mathfrak{m}_i \to \mathfrak{m}_1 \cdots \mathfrak{m}_{i-1} \to V_i \to 0$ give $\operatorname{len} A = \sum_{i=1}^n \operatorname{len} V_i < \infty$.

Complete converse with [AM16, Thm. 8.5].

***Hopkins's theorem* I.5.4.6** (1939, Akizuki 1935). — *Any artinian ring is noetherian.*

Proof. — Let A be an artinian ring with distinct maximal ideals $\mathfrak{m}_1, \dots, \mathfrak{m}_n$ (proposition I.5.4.3). Then $\prod_{i=1}^n \mathfrak{m}_i^k \subseteq (\bigcap_{i=1}^n \mathfrak{m}_i)^k = (\operatorname{nil} A)^k = (0)$ for some k by proposition I.5.4.4. Thus, A is noetherian by lemma I.5.4.5.

***Proposition* I.5.4.7.** — *Let $(A, \mathfrak{m})$ be a noetherian local ring. Then exactly one of the following is true:*

(i) $\mathfrak{m}^n \neq \mathfrak{m}^{n+1}$ *for all* n*;*
(ii) $\mathfrak{m}^n = (0)$ *for some* n*, in which case* A *is an artinian local ring.*

Proof. — Suppose that $\mathfrak{m}^n = \mathfrak{m}^{n+1}$ for some n. Applying Nakayama's lemma I.2.3.8 to the A-module $\mathfrak{m}^n$, which is finitely generated by proposition I.5.1.4, we obtain $\mathfrak{m}^n = 0$.

Complete [AM16, Prop. 8.6].

***Theorem* I.5.4.8** (structure theorem for artinian rings). — *An artinian ring is, up to isomorphism, a uniquely finite direct product of artinian local rings.*

Proof. — Suppose that $\mathfrak{m}_1, \dots, \mathfrak{m}_n$ are the distinct maximal ideals of an artinian ring A (proposition I.5.4.3). From the proof of Hopkins's theorem I.5.4.6 we know that $\prod_{i=1}^n \mathfrak{m}_i^k = (0)$ for some k. Since the $\mathfrak{m}_i$ are pairwise coprime, corollary I.1.3.21 implies that the ideals $\mathfrak{m}_i^k$ are pairwise coprime, so $A \cong \prod_{i=1}^n (A/\mathfrak{m}_i^k)$ by the Chinese remainder theorem I.1.2.16. Each factor $A/\mathfrak{m}_i^k$ is artinian and local (if $\mathfrak{m}_i^k \subseteq \mathfrak{m}_j$, then $\mathfrak{m}_i = \sqrt{\mathfrak{m}_i^k} \subseteq \sqrt{\mathfrak{m}_j} = \mathfrak{m}_j$ and $i = j$), whence A is a direct product of artinian local rings.

Conversely, suppose that $A \cong \prod_{i=1}^n A_i$ where the $(A_i, \mathfrak{q}_i)$ are artinian local rings. Setting $\mathfrak{a}_i := \ker(A \twoheadrightarrow A_i)$ so that $A_i \cong A/\mathfrak{a}_i$, the Chinese remainder theorem I.1.2.16 implies that the $\mathfrak{a}_i$ are pairwise coprime and that $\bigcap_{i=1}^n \mathfrak{a}_i = (0)$. If $\mathfrak{p}_i$ is the contraction of $\mathfrak{q}_i$ in A

[AM16, ch. 8,

CONTINUE TODO!!!

CHAPTER II

DIMENSION THEORY

***Definition* II.0.0.1.** — *Let A be a non-trivial ring. A **chain** of prime ideals of A is defined as a strictly increasing sequence $\mathfrak{p}_0 \subsetneq \cdots \subsetneq \mathfrak{p}_n$ of prime ideals; its **length** is n. Note that the length is the number of strict inclusions, not the number of ideals in this chain.*

*The **Krull dimension** or **dimension** $\dim A$ of A is defined as the supremum of the lengths of all chains of prime ideals of A. Hence $\dim A \in \mathbb{N} \cup \{\infty\}$.*

*Similarly, the **height** $\operatorname{ht}\mathfrak{p}$ of a prime ideal $\mathfrak{p}$ is the supremum of the lengths of all chains of prime ideals contained in $\mathfrak{p}$, i. e., of chains $\mathfrak{p}_0 \subsetneq \cdots \subsetneq \mathfrak{p}_n = \mathfrak{p}$. Dually, the **coheight** $\operatorname{coht}\mathfrak{p}$ of $\mathfrak{p}$ is the supremum of the lengths of all chains of prime ideals contained in $\mathfrak{p}$, i. e., of chains $\mathfrak{p} = \mathfrak{p}_0 \subsetneq \cdots \subsetneq \mathfrak{p}_n$.*[II.1]

***Fact* II.0.0.2.** — *By definition, we have $\operatorname{ht}\mathfrak{p} = \dim A_{\mathfrak{p}}$, $\operatorname{coht}\mathfrak{p} = \dim A/\mathfrak{p}$ and $\operatorname{ht}\mathfrak{p} + \operatorname{coht}\mathfrak{p} \leq \dim A$.*

***Example* II.0.0.3.** — (i) A field is the same as a zero-dimensional integral domain.

(ii) Since the prime ideals of $\mathbb{Z}$ are (0) and (p) for $p = 2, 3, 5, \ldots$ a prime number, we have $\dim \mathbb{Z} = 1$. More generally, a principal ideal domain which is not a field is one-dimensional.

(iii) An artinian Ring is zero-dimensional since every prime ideal is maximal (proposition I.5.4.1).

(iv) Consider the polynomial ring $k[x_1, \ldots, x_n]$ over a field k. Then $(0) \subsetneq (x_1) \subsetneq (x_1, x_2) \subsetneq \cdots \subsetneq (x_1, \ldots, x_n)$ is a chain of prime ideals, so $\dim k[x_1, \ldots, x_n] \geq n$. In fact, the dimension is equal to n, as we will show.

***Definition* II.0.0.4.** — (i) *Let $\mathfrak{a}$ be an ideal of a non-trivial ring A. Then the **height** $\operatorname{ht}\mathfrak{a}$ of $\mathfrak{a}$ is the infimum of the heights of all prime ideals containing $\mathfrak{a}$.*

(ii) *Let M be an A-module. Then the **Krull dimension** or **dimension** $\dim_A M = \dim M$ of M is $\dim(A/\operatorname{ann} M)$.*

***Fact* II.0.0.5.** — *We again have $\operatorname{ht}\mathfrak{a} + \dim A/\mathfrak{a} \leq \dim A$ for any ideal $\mathfrak{a}$ of A.*

***Definition* II.0.0.6.** — *Let A be a non-trivial ring. A chain of prime ideals is said to be **saturated** if it cannot be further refined, i. e., there are no prime ideals strictly contained between two consecutive terms. Then A is a so-called **catenary ring** if for any prime ideals $\mathfrak{p} \subseteq \mathfrak{q}$ of A, there exists a saturated chain of prime ideals starting from $\mathfrak{p}$ and ending at $\mathfrak{q}$, and all such chains have the same (finite) length.*

***Fact* II.0.0.7.** — *In a catenary ring A, we have $\operatorname{ht}\mathfrak{p} + \operatorname{coht}\mathfrak{p} = \dim A$ for any prime ideal $\mathfrak{p}$ of A.*

***Remark* II.0.0.8.** — If A is a noetherian local integral domain, then the converse is also true. The proof is difficult, so we postpone it (Ratliff 1972). In practice, all important noetherian rings are catenary, but there are non-catenary noetherian rings (Nagata 1956).

Insert reference.

(II.1) Traditionally, the height of a prime ideal was also called *rank*, and the coheight was also called *dimension*. Additionally, Nagata uses *altitude* for the Krull dimension.

§ II.1. DIMENSIONS OF FINITELY GENERATED ALGEBRAS OVER A FIELD

***Lemma* II.1.0.1.** — *Let K/k be an algebraic field extension, and let $\alpha_1, \dots, \alpha_n \in K$. Consider $\phi = \mathrm{ev}_{(\alpha_1,\dots,\alpha_n)} : k[x_1, \dots, x_n] \to k[\alpha_1, \dots, \alpha_n] = k(\alpha_1, \dots, \alpha_n)$ mapping $x_i \mapsto \alpha_i$. Then $\ker\phi$ is the maximal ideal generated by n elements of the form $f_1(x_1)$, $f_2(x_1, x_2)$, ..., $f_n(x_1, \dots, x_n)$ with each f_i monic in x_i.*

Proof. — Let f_1 be the minimal polynomial of α_1 over k. Then $k[\alpha_1] = k(\alpha_1) \cong k[x_1]/(f_1(x_1))$ is a field. Let g_2 be the minimal polynomial of α_2 over $k(\alpha_1)$. Then the coefficients of g_2 are polynomials in α_1, whence there is some $f_2 \in k[x_1, x_2]$ monic in x_2 such that $g_2(x_2) = f_2(\alpha_1, x_2)$. Thus

$$k[\alpha_1, \alpha_2] = k(\alpha_1)[\alpha_2] = k(\alpha_1, \alpha_2) \cong k(\alpha_1)[x_2]/(f_2(\alpha_1, x_2)).$$

Proceeding inductively, there is some $f_i(x_1, \dots, x_i) \in k[x_1, \dots, x_i]$ monic in x_i such that

$$k[\alpha_1, \dots, \alpha_i] = k(\alpha_1, \dots, \alpha_i) \cong k(\alpha_1, \dots, \alpha_{i-1})[x_i]/(f_i(\alpha_1, \dots, \alpha_{i-1}, x_i)).$$

Now, let $p \in \ker\phi$, that is, $p(\alpha_1, \dots, \alpha_n) = 0$. Then $p(\alpha_1, \dots, \alpha_{n-1}, x_n)$ is divisible by the minimal polynomial $f_n(\alpha_1, \dots, \alpha_{n-1}, x_n)$ of α_n over $k(\alpha_1, \dots, \alpha_{n-1})$. As polynomials in x_n, we can divide p by f_n to obtain $p = q_n f_n + r_n$ where the remainder satisfies $r_n(\alpha_1, \dots, \alpha_{n-1}, x_n) = 0$. Again, as polynomials in x_{n-1}, dividing r_n by f_{n-1} gives $r_n = q_{n-1} f_{n-1} + r_{n-1}$ where the remainder satisfies $r_{n-1}(\alpha_1, \dots, \alpha_{n-2}, x_{n-1}, x_n) = 0$. Proceeding inductively, we finally obtain $p = \sum_{i=1}^n q_i f_i + r_1$ where $r_1(x_1, \dots, x_n) = 0$. Thus, $p \in (f_1, \dots, f_n)$. The converse that $f_i \in \ker\phi$ for all i is clear. □

***Proposition* II.1.0.2.** — *Let A be a finitely generated k-algebra which is an integral domain. If the transcendence degree* $\operatorname{tr.deg}_k \operatorname{Quot} A$ *is positive, then A is not a field.*

Proof. — Write $A = k[\alpha_1, \dots, \alpha_n]$. Suppose that $\alpha_1, \dots, \alpha_r$ be a transcendence basis of $\operatorname{Quot} A$ over k where $r = \operatorname{tr.deg}_k \operatorname{Quot} A < n$, and set $K = k(\alpha_1, \dots, \alpha_r) \subseteq \operatorname{Quot} A$. Then $\alpha_{r+1}, \dots, \alpha_n$ are algebraic over K. By lemma II.1.0.1, there are polynomials $f_i \in K[x_{r+1}, \dots, x_i]$ monic in x_i for $i = r+1, \dots, n$ such that

$$K[\alpha_{r+1}, \dots, \alpha_n] \cong K[x_{r+1}, \dots, x_n]/(f_{r+1}, \dots, f_n)$$

with $\deg f_i = (K(\alpha_{r+1}, \dots, \alpha_i) : K(\alpha_{r+1}, \dots, \alpha_{i-1}))$ (recall that $f_i(\alpha_{r+1}, \dots, \alpha_{i-1}, x_i)$ is the minimal polynomial of α_i). Since the coefficients of the f_i are in K, let $g \in k[\alpha_1, \dots, \alpha_n] \setminus \{0\}$ be the product of all their denominators. Then each f_i is a polynomial with coefficients in $B = k[\alpha_1, \dots, \alpha_r, g^{-1}]$.

We claim that $A[g^{-1}] = B[\alpha_{r+1}, \dots, \alpha_n]$ is a free B-module with basis $\{\prod_{i=r+1}^n \alpha_i^{e_i} \mid 0 \le e_i < \deg f_i\}$. Let $p(\alpha_{r+1}, \dots, \alpha_n)$ with $p \in B[x_{r+1}, \dots, x_n]$ be any element of $B[\alpha_{r+1}, \dots, \alpha_n]$. Reducing p modulo f_n as polynomials in x_n, we may assume that $\deg p < \deg f_n$ in x_n. Further reducing p modulo f_{n-1} as polynomials in x_{n-1}, we may assume that $\deg p < \deg f_{n-1}$ in x_{n-1}. By induction, we may assume that $\deg p < \deg f_i$ in x_i for $i = r+1, \dots, n$. In other words, any element $p(\alpha_{r+1}, \dots, \alpha_n)$ is a linear combination of monomials $\prod_{i=r+1}^n \alpha_i^{e_i}$ with $0 \le e_i < \deg f_i$. □

CHAPTER III

INTEGRALITY AND VALUATION RINGS

***Motivation* III.0.0.1.** — A frequent technique to study curves in classical algebraic geometry was to project a curve onto a line so that the curve can be considered as a *ramified* covering of the line. A similar ramification behaviour is exhibited in algebraic number theory between the rings of integers $\mathcal{O}_K$ and $\mathcal{O}_{\mathbb{Q}}$ of a number field K and of the rational field $\mathbb{Q}$. These two notions are algebraically expressed through the notion of *integrality*.

§ III.1. INTEGRALITY

III.1.1. Definitions and basic properties.

Definition* III.1.1.1.** — *Let B be an extension ring of a ring A (i. e., $1 \in A$). An element $x \in B$ is said to be* ***integral *over A if x is a root of a* monic *polynomial with coefficients in A, i. e., if x satisfies an equation of the form $x^n + a_1 x^{n-1} + \dots + a_n = 0$ where $a_i \in A$.*

III.1.1.2. — Recall the *determinant trick* from lemma I.2.3.5: let $\mathfrak{a}$ be an ideal of a ring A, and let M be a finitely generated A-module. If $\phi\colon M \to M$ is an A-linear map such that $\operatorname{im}\phi \subseteq \mathfrak{a}M$, then we have a relation of the form $\phi^n + a_1\phi^{n-1} + \dots + a_n = 0$ where $a_i \in A$.

***Proposition* III.1.1.3** (characterisation of integrality)**.** — *Let B be an extension ring of a ring A. Then the following are equivalent:*

(i) *$x \in B$ is integral over A;*
(ii) *$A[x]$ is a finitely generated A-module.*
(iii) *$A[x]$ is contained in a finite A-subalgebra of B;*
(iv) *there exists a faithful $A[x]$-module which is finitely generated over A.*

Proof. — Suppose (i). Then $x^n = -(a_1x^{n-1} + \dots + a_n)$, and by induction all positive powers of x lie in the A-module generated by $1, x, \dots, x^{n-1}$. Hence, $A[x] = (1, x, \dots, x^{n-1})$ as an A-module, which implies (ii). (ii) clearly implies (iii).

Suppose (iii) and let $A[x] \subseteq C \subseteq B$ be the finite A-subalgebra. Then C is a faithful $A[x]$-module (we have $\operatorname{ann}_{A[x]} C = 0$ because $1 \in C$), giving (iv).

Finally, let M be a faithful $A[x]$-module of finite type over A, as in (iv), and let $x \in B$. We now apply the determinant trick (lemma I.2.3.5): let $\phi\colon M \to M$ be multiplication by x and let $\mathfrak{a} = A$ (we have $\operatorname{im}\phi \subseteq \mathfrak{a}M$ since M is an $A[x]$-module). By the determinant trick, there are suitable $a_i \in A$ so that multiplication by $x^n + a_1x^{n-1} + \dots + a_n$ kills M. Since M is faithful, we have $x^n + a_1x^{n-1} + \dots + a_n = 0$, which is (i). ∎

***Corollary* III.1.1.4.** — *Let B be an extension ring of a ring A. If $x_1, \dots, x_n \in B$ are integral over A, then $A[x_1, \dots, x_n]$ is a finitely generated A-module.*

Proof. — We perform induction on n. The base case $n = 1$ is proposition III.1.1.3. For the induction step, write $A_r := A[x_1, \dots, x_r]$. By induction hypothesis, A_{n-1} is of finite type over A and $A_n = A_{n-1}[x_n]$ is of finite type over A_{n-1} (since x_n is integral over A_{n-1}). Thus, A_n is of finite type over A. ∎

***Corollary* III.1.1.5.** — *Let B be an extension ring of a ring A. The subset $\tilde{A}$ of all elements in B that are integral over A is a subring of B containing A.*

Proof. — If $x, y \in \tilde{A}$, then $A[x, y]$ is a finite A-subalgebra of B by corollary III.1.1.4. Since $A[x + y]$, $A[x - y]$ and $A[xy]$ are contained in $A[x, y]$, we have $x \pm y, xy \in \tilde{A}$ by proposition III.1.1.3. ∎

***Definition* III.1.1.6.** — *Let B be an extension ring of a ring A. The ring $\tilde{A}$ of all elements in B that are integral over A is called the **integral closure** of A in B.*

*If $\tilde{A} = A$, then A is said to be **integrally closed** in B. If $\tilde{A} = B$, then B is said to be **integral** over A, or B is called an **integral extension** of A.*

More generally, we can generalise ring extensions to algebras, which in theory have little differences. In practice, however, we prefer extension rings since they are easier to work with.

***Definition* III.1.1.7.** — *Let $f\colon A \to B$ be an algebra. Then f is said to be **integral** and B is called an **integral** A-algebra if B is integral over the subring $f(A)$.*

***Corollary* III.1.1.8.** — *In this terminology, an A-algebra is finite iff it is of finite type and integral.*

Proof. — Finiteness implies integral dependence by proposition III.1.1.3, and it trivially implies finite type. The converse follows from corollary III.1.1.4. ∎

***Corollary* III.1.1.9** (transitivity of integral extensions). — *If $A \subseteq B \subseteq C$ are rings such that C is integral over B and B is integral over A, then C is integral over A.*

Proof. — Let $x \in C$. Then $x^n + b_1 x^{n-1} + \dots + b_n = 0$ with $b_i \in B$. Then $B' := A[b_1, \dots, b_n]$ is a finitely generated A-module by corollary III.1.1.4, and $B'[x]$ is a finitely generated B'-module by proposition III.1.1.3. Hence, $B'[x]$ is finitely generated over A and therefore, x is integral over A by proposition III.1.1.3. ∎

***Corollary* III.1.1.10.** — *Let B be an extension ring of A, and let $\tilde{A}$ be the integral closure of A in B. Then $\tilde{A}$ is integrally closed in B.*

Proof. — If $x \in B$ is integral over $\tilde{A}$, then x is integral over A by transitivity, whence $x \in \tilde{A}$. ∎

The following says that integral extensions are compatible with passing to quotients or localisations.

***Lemma* III.1.1.11.** — *Let B be an integral extension ring of A.*

(i) *If $\mathfrak{b}$ is an ideal of B, then $B/\mathfrak{b}$ is integral over $A/(\mathfrak{b} \cap A)$.*
(ii) *If S is multiplicative subset of A, then $S^{-1}B$ is integral over $S^{-1}A$.*
(iii) *More generally, suppose that A and B are both R-algebras. If C is another R-algebra, then $B \otimes_R C$ is integral over $A \otimes_R C$.*

Proof. — For (i), let $\bar{x} \in B/\mathfrak{b}$. Then any lift $x \in B$ of $\bar{x}$ satisfies an equation of the form $x^n + a_1 x^{n-1} + \dots + a_n = 0$ with $a_i \in A$. Reducing modulo $\mathfrak{b}$ yields $\bar{x}^n + \bar{a}_1 \bar{x}^{n-1} + \dots + \bar{a}_n = 0$ with $\bar{a}_i \in A/(\mathfrak{b} \cap A)$. Hence $\bar{x}$ is integral over $A/(\mathfrak{b} \cap A)$.

For (ii), let $x/s \in S^{-1}B$. Then $x \in B$ satisfies an equation $x^n + a_1 x^{n-1} + \dots + a_n = 0$ as before. Dividing by s^n yields $(x/s)^n + (a_1/s)(x/s)^{n-1} + \dots + a_n/s^n = 0$ with $a_i/s^i \in S^{-1}A$. Hence x/s is integral over $S^{-1}A$.

For (iii), let $x \otimes c \in B \otimes_R C$. Then $x \in B$ satisfies an equation $x^n + a_1 x^{n-1} + \dots + a_n = 0$ with $a_i \in A$ as before. Tensoring with c^n yields $(x \otimes c)^n + (a_1 \otimes c)(x \otimes c)^{n-1} + \dots + a_n \otimes c^n = 0$ with $a_i \otimes c^i \in A \otimes_R C$. Hence $x \otimes c$ is integral over $A \otimes_R C$, and so is any sum of elementary tensors.

***Warning* III.1.1.12.** — In particular, if $\mathfrak{p}$ is a prime ideal of A, then B being integral over A implies $B_\mathfrak{p} = (A \setminus \mathfrak{p})^{-1} B$ being integral over $A_\mathfrak{p}$. However, if $\mathfrak{q}$ is prime ideal of B lying over $\mathfrak{p}$, then $B_\mathfrak{q}$ is not necessarily integral over $A_\mathfrak{p}$.

Indeed, take $B = k[x]$ and $A = k[x^2 - 1]$ for k a field, and take $\mathfrak{q} = (x-1)$ so that $\mathfrak{p} = \mathfrak{q} \cap A = (x^2-1)$ (in fact, $\mathfrak{p}$ and $\mathfrak{q}$ are maximal). Suppose that $f = 1/(x+1) \in B_\mathfrak{q}$ is integral over $A_\mathfrak{p}$. Then there is an equation $f^n + a_1/b_1 f^{n-1} + \dots + a_n/b_n = 0$ where the polynomials b_i have no root in ± 1. Dividing by f^n gives $1 + a_1/b_1(x+1) + \dots + a_n/b_n(x+1)^n = 0$ in $k[x]$, and evaluating at -1 gives a contradiction.

***Corollary* III.1.1.13.** — *Let B be an integral extension ring of A, and let $\tilde{A}$ be the integral closure of A in B. If S is a multiplicative set of A, then $S^{-1}\tilde{A}$ is the integral closure of $S^{-1}A$ in $S^{-1}B$.*

Proof. — By lemma III.1.1.11, $S^{-1}\tilde{A}$ is integral over $S^{-1}A$. Conversely, suppose that $b/s \in S^{-1}B$ is integral over $S^{-1}A$. Then we have $(b/s)^n + (a_1/s_1)(b/s)^{n-1} + \dots + a_n/s_n = 0$ with suitable $a_i \in A$ and $s_i \in S$. Let $t := s_1 \cdots s_n$. Killing the common denominator by multiplying this equation by $(str)^n$ with suitable $r \in S$ (r exists solely for the definition of localisation), we obtain an integrality relation over A demonstrating that $btr \in \tilde{A}$. Therefore, $b/s = btr/str \in S^{-1}C$.

***Fact* III.1.1.14.** — *Finte products of integral A-algebras are integral over A.*

Proof. — Let $B_1, \dots, B_n$ be integral A-algebras, and let $b = (b_1, \dots, b_n) \in \prod_{i=1}^n B_i$. Then each b_i satisfies a monic polynomial $f_i(t) \in A[t]$. If $f(t) = \prod_{i=1}^n f_i(t) \in A[t]$, then $f(b_i) = 0$ for all i, and thus $f(b) = (f(b_1), \dots, f(b_n)) = 0$.

***Lemma* III.1.1.15.** — *Let B be an extension ring of A, and let $\tilde{A}$ be the integral closure of A in B. Let $f, g \in B[x]$ be monic polynomials such that $fg \in \tilde{A}[x]$. Then $f, g \in \tilde{A}[x]$.*

Proof. — We first show the simpler case where A and B are integral domains. Then $\operatorname{Quot} B$ admits a splitting field of f and g. Since any root of f or g is also a root of fg, these roots are integral over $\tilde{A}$. Since the coefficients of f or g are sums of products of these roots, these are integral over $\tilde{A}$ as well.

For the general case where B is not an integral domain, note that the construction of an extension ring C of B such that f and g split over C works identically to the case of fields. Indeed, if $h \in B[x]$ is a polynomial, consider the ring extension $B \hookrightarrow B' := B[x]/(h(x))$. Then h has a root $\bar{x} \in B'$, so consider $h' = h(t)/(t - \bar{x}) \in B'[t]$ with $\deg h' < \deg h$. Now, proceed with h' inductively.

***Lemma* III.1.1.16.** — *Let B be an extension ring of A, and let $\tilde{A}$ be the integral closure of A in B. Then the integral closure of $A[x]$ in $B[x]$ is $\tilde{A}[x]$.*

Proof. — Let $f \in B[x]$ be integral over $A[x]$, say, $f^m + g_1 f^{m-1} + \dots + g_m = 0$ with $g_i \in A[x]$ and m odd. Choose $r > \deg f, \deg g_1, \dots, \deg g_m$, and set $f' := f - x^r \in B[x]$. Then $f' + x^r$ satisfies the integrality equation of f, and expanding that yields $(f')^m + h_1 (f')^{m-1} + \dots + h_m = 0$ for suitable $h_i \in A[x]$, for example $h_m = (x^r)^m + g_1 (x^r)^{m-1} + \dots + g_m$. Note that h_i is monic of degree ri by choice of r. Hence $-f'$ and $(f')^{m-1} + h_1 (f')^{m-2} + \dots + h_{m-1}$, whose product is $h_m \in A[x] \subseteq \tilde{A}[x]$, are monic ($m$ is odd!) polynomials in $B[x]$. By lemma III.1.1.15, we obtain $-f' \in \tilde{A}[x]$ and hence $f \in \tilde{A}[x]$.

Conversely, let $f \in \tilde{A}[x]$, and let $a_1, \dots, a_n \in \tilde{A}$ be the coefficients of f. Then the A-algebra $A[a_1, \dots, a_n]$ is finite by corollary III.1.1.4, so the $A[x]$-algebra $A[x, a_1, \dots, a_n]$, which contains $A[x, f]$, is finite as well. By proposition III.1.1.3, f is integral over $A[x]$.

Insert remark on normal rings on [Mat86, p. 64].

***Definition* III.1.1.17.** — *An **integrally closed domain** A is an integral domain that is integrally closed in* $\operatorname{Quot} A$.[III.1] *A **normal ring** A is a ring where for any prime ideal $\mathfrak{p}$ of A, the localisation $A_{\mathfrak{p}}$ at $\mathfrak{p}$ is an integrally closed domain.*

***Example* III.1.1.18.** — A finite product of integrally closed domains $A_1, \dots, A_n$ is normal. The reason is that the prime ideals of $\prod_{i=1}^n A_i$ are of the form $\mathfrak{p} = A_1 \times \dots \times A_{i-1} \times \mathfrak{p}_i \times A_{i+1} \times \dots \times A_n$ where $\mathfrak{p}_i$ is a prime ideal of A_i, see proposition I.1.3.8. So, $(\prod_{i=1}^n A_i)_{\mathfrak{p}} \cong A_{\mathfrak{p}_i}$ is an integrally closed domain.

***Proposition* III.1.1.19** (integral closure is local)**.** — *Let A be an integral domain. Then the following are equivalent:*

(i) *A is integrally closed;*
(ii) *$A_{\mathfrak{p}}$ is integrally closed for all $\mathfrak{p} \in \operatorname{Spec} A$;*
(iii) *$A_{\mathfrak{m}}$ is integrally closed for all $\mathfrak{m} \in \operatorname{MaxSpec} A$;*

Proof. — Let $\tilde{A}$ be the integral closure of A in $\operatorname{Quot} A$, and let $f\colon A \to \tilde{A}$ be the inclusion map. Then A (resp. $A_{\mathfrak{p}}$ or $A_{\mathfrak{m}}$) is integrally closed iff f (resp. $f_{\mathfrak{p}}$ or $f_{\mathfrak{m}}$) is surjective. The statement now follows from proposition I.3.2.17.

***Definition* III.1.1.20.** — *Let B be an extension ring of a ring A, and let $\mathfrak{a}$ be an ideal of A. An element of B is said to be **integral** over $\mathfrak{a}$ if it is integral over A and the coefficients in the integrality relation lie in $\mathfrak{a}$. We also define the **integral closure** of $\mathfrak{a}$ in B as before.*

***Lemma* III.1.1.21.** — *Let B be an extension ring of A, and let $\tilde{A}$ be the integral closure of A in B. Let $\mathfrak{a}$ be an ideal of A, and let $\mathfrak{a}^e$ be the extension of $\mathfrak{a}$ in $\tilde{A}$. Then the integral closure of $\mathfrak{a}$ in B is $\sqrt{\mathfrak{a}^e}$ (and in particular closed under addition and multiplication).*

Proof. — Let $x \in B$ be integral over $\mathfrak{a}$ so that we have an equation, say, $x^n + a_1 x^{n-1} + \dots + a_n = 0$ with $a_i \in \mathfrak{a}$. Then $x \in \tilde{A}$ and hence $x^n \in \mathfrak{a}^e$ (since $a_i \in \mathfrak{a}$), so $x \in \sqrt{\mathfrak{a}^e}$.

Conversely, let $x \in \sqrt{\mathfrak{a}^e}$. Then $x^n = \sum_{i=1}^k a_i x_i$ for some $n > 0$, $a_i \in \mathfrak{a}$ and $x_i \in \tilde{A}$. By corollary III.1.1.4, the A-module $M = A[x_1, \dots, x_k]$ is finitely generated, and we have $x^n M \subseteq \mathfrak{a} M$ (note that $x_i x_j \in M$ since M is a ring). By lemma I.2.3.5 (determinant trick with multiplication by x^n), we obtain that x^n is integral over $\mathfrak{a}$ and thus, x is integral over $\mathfrak{a}$.

***Proposition* III.1.1.22.** — *Let A be an integrally closed domain with field of fractions $K = \operatorname{Quot} A$, and let B be an extension domain of A. If an element $x \in B$ is integral over an ideal $\mathfrak{a}$ of A, then its minimal polynomial over K exists and has coefficients in $\sqrt{\mathfrak{a}}$. This applies in particular to $\mathfrak{a} = A = \sqrt{\mathfrak{a}}$.*

Proof. — Since x is integral over $\mathfrak{a}$, it is also algebraic over K and hence admits a minimal polynomial f over K. Let L be the splitting field of f, and let $x_1, \dots, x_n \in L$ be the conjugates of x. Under the K-isomorphism $K[x] \xrightarrow{\sim} K[x_i]$, we see that each x_i satisfies the same integrality equation as x, so each x_i is also integral over $\mathfrak{a}$. Expanding the splitting of f into linear factors, we see that the coefficients of f are sums of products of the x_i and with lemma III.1.1.21 therefore are integral over $\mathfrak{a}$. Since A is integrally closed, lemma III.1.1.21 implies that these coefficients are contained in $\sqrt{\mathfrak{a}}$.

***Proposition* III.1.1.23.** — *Let A be an integrally closed domain with field of fractions $K = \operatorname{Quot} A$. Let L be a finite separable extension of K, and let B be the integral closure of A in L. Then there exists a basis $v_1, \dots, v_n$ of L over K such that $B \subseteq \sum_{i=1}^n A v_i$.*

[III.1] Some authors use *normal ring* for integrally closed domains. Our naming convention follows Serre and Grothendieck. **bib:stacks** uses *normal domain* instead.

Proof. — Any $v \in L$ satisfies its own minimal polynomial over K, say, $v^r + (a_1/b_1)v^{r-1} + \dots + a_r/b_r = 0$ with $a_i/b_i \in K$. Multiplying with $(b_1 \cdots b_r)^r$ shows that $b_1 \cdots b_r v$ is integral over A and hence contained in B. Doing this for each basis vector of an arbitrary basis of L over K, we obtain a basis with vectors in B. So, let $u_1, \dots, u_n \in B$ be a basis of L over K.

Let $\mathrm{Tr}_{L/K} \colon L \to L$ be the trace of L/K. Since L/K is separable, the trace pairing $(x, y) \mapsto \mathrm{Tr}_{L/K}$ on L is non-degenerate. We obtain a dual basis $v_1, \dots, v_n$ of L over K w. r. t. this pairing, i. e., $\mathrm{Tr}_{L/K}(u_i v_j) = \delta_{ij}$.

This is standard field theory. Maybe add a reference.

Let $x \in B$, and write $x = \sum_{i=1}^n x_i v_i$ with $x_i \in K$. We have $xu_i \in B$ for all i since $u_i \in B$. Its trace $\mathrm{Tr}_{L/K}(xu_i)$ is an integer multiple of the coefficient of second highest degree in the minimal polynomial of xu_i over K,[III.2] so $\mathrm{Tr}_{L/K}(xu_i) \in A$ by proposition III.1.1.22. However, $\mathrm{Tr}_{L/K}(xu_i) = \sum_{j=1}^n x_j \mathrm{Tr}_{L/K}(u_i v_j) = \sum_{j=1}^n x_j \delta_{ij} = x_i$, so $x_i \in A$. Thus, $x \in \sum_{i=1}^n A v_i$. □

Lemma III.1.1.24. — *Any unique factorisation domain is an integrally closed, e. g., $\mathbb{Z}$ and $k[x_1, \dots, x_n]$ for k a field.*

Proof. — Let A be a unique factorisation domain. If $x/y \in \operatorname{Quot} A$ with x and y coprime is integral over A, then we have $x^n + a_1 x^{n-1} y + \dots + a_n y^n = 0$ for suitable $a_i \in A$. Hence, $y \mid x^n$ implying that $y = \pm 1$. □

***Example* III.1.1.25.** — Let k be a field, and let $A = k[t^2, t^3] \subseteq B = k[t]$ be polynomial rings, whose fraction field is $K = k(t)$. Since B is a unique factorisation domain, it is integrally closed (lemma III.1.1.24), so the integral closure of A in K is contained in B. On the other hand, $t \in B$ is integral over A, whence B is the integral closure of A in K.

In classical algebraic geometry, $A \cong k[x, y]/(y^2 - x^3)$ is the affine coordinate ring of the cuspidal curve $y^2 = x^3$, which has a singularity in the origin. The fact that A is not integrally closed is directly related to the singularity at the origin. Moreover, taking the integral closure B of A 'straightens out' the singularity.

***Proposition* III.1.1.26.** — *Let A be a unique factorisation domain in which 2 is a unit, and let $d \in A$ be square-free (i. e., $p^2 \nmid d$ for any prime $p \in A$). Then $A[\sqrt{d}]$ is an integral domain.*

Proof. — Let $\alpha = \sqrt{d}$ be a square root of d. By lemma III.1.1.24, A is integrally closed in $K = \operatorname{Quot} A$. Hence, if $\alpha \in K$ then $\alpha \in A$, and $A[\alpha] = A$ is integrally closed as desired. Otherwise, if $\alpha \notin K$ then $\operatorname{Quot} A[\alpha] = K(\alpha) = K + K\alpha$. Any element $\beta \in K(\alpha)$ is of the form $\beta = x + y\alpha$ for $x, y \in K$, and its minimal polynomial over K is $T^2 - 2xT + (x^2 - y^2 f)$. Assume that β is integral over A. By proposition III.1.1.22, we have $2x \in A$ and $x^2 - y^2 f \in A$. Since 2 is a unit, we obtain $x \in A$ and $y^2 f \in A$. If there exists a prime factor $p \in A$ of the denominator of y, then $y^2 f \in A$ requires $p^2 \mid f$, contradicting that f is square-free. Hence $y \in A$ and therefore $\beta \in A + A\alpha = A[\alpha]$. On the other hand, α is integral over A, so $A[\alpha]$ is the integral closure of A in $K(\alpha)$. Now, if $\beta \in K(\alpha)$ is integral over $A[\alpha]$, then it is integral over A by transitivity, and thus $\beta \in A[\alpha]$. In other words, $A[\alpha]$ is integrally closed. □

III.1.2. Cohen-Seidenberg theorems (going-up and going-down).

Lemma III.1.2.1. — *Let $A \subseteq B$ be an extension of integral domains such that B is integral over A. Then B is a field iff A is a field.*

Proof. — Firstly, assume that A is a field. For $y \in B \setminus \{0\}$, let $y^n + a_1 y^{n-1} + \dots + a_n = 0$ with $a_i \in A$ be an integrality relation. Since B is an integral domain, we may assume w. l. o. g. $a_n \neq 0$, so that $y^{-1} = -a_n^{-1}(y^{n-1} + a_1 y^{n-2} + \dots + a_{n-1}) \in B$. Thus, B is a field.

[III.2] If $t^d + a_1 t^{d-1} + \dots + a_d$ is the minimal polynomial of xu_i, then $\mathrm{Tr}_{L/K}(xu_i) = (L : K)/d \cdot a_1$.

Conversely, assume that B is a field. For $x \in A \setminus \{0\}$, we have $x^{-1} \in B$, so let $x^{-m} + b_1 x^{1-m} + \dots + b_m = 0$ be an integrality relation with $b_i \in A$. Then $x^{-1} = -(b_1 + b_2 x + \dots + b_m x^{m-1}) \in A$ and thus, A is a field. □

***Corollary* III.1.2.2.** — *Let B be an extension ring of A which is integral over A. Let $\mathfrak{q}$ be a prime ideal of B. Then $\mathfrak{q}$ is maximal iff its contraction $\mathfrak{q} \cap A$ is maximal.*

Proof. — Passing to quotients, $B/\mathfrak{q}$ is integral over $A/(\mathfrak{q} \cap A)$ by lemma III.1.1.11. Since these are both integral domains, we can now apply lemma III.1.2.1. □

***Corollary* III.1.2.3.** — *Let B be an extension ring of A which is integral over A. Let $\mathfrak{q} \subseteq \mathfrak{q}'$ be two ideals of B where $\mathfrak{q}$ is prime.*[III.3] *If they lie over the same prime ideal $\mathfrak{p} = \mathfrak{q} \cap A = \mathfrak{q}' \cap A$ of A, then $\mathfrak{q} = \mathfrak{q}'$.*

Proof. — Passing to localisations, $B_{\mathfrak{p}}$ is integral over $A_{\mathfrak{p}}$ by lemma III.1.1.11 (the notation regards B as an A-module so that $B_{\mathfrak{p}} = (A \setminus \mathfrak{p})^{-1} B$). Moreover, for the extensions we have $\mathfrak{q}B_{\mathfrak{p}} \subseteq \mathfrak{q}'B_{\mathfrak{p}}$ and $\mathfrak{q}B_{\mathfrak{p}} \cap A_{\mathfrak{p}} = \mathfrak{q}'B_{\mathfrak{p}} \cap A_{\mathfrak{p}} = \mathfrak{p}A_{\mathfrak{p}}$ (if $S \coloneqq A \setminus \mathfrak{p}$, then $\mathfrak{p}A_{\mathfrak{p}} = S^{-1}\mathfrak{p} = S^{-1}\mathfrak{q} \cap S^{-1}A = \mathfrak{q}B_{\mathfrak{p}} \cap A_{\mathfrak{p}}$). In particular, $\mathfrak{q}'B_{\mathfrak{p}}$ is a proper ideal. Since $\mathfrak{p}A_{\mathfrak{p}}$ is the unique maximal ideal of $A_{\mathfrak{p}}$, corollary III.1.2.2 implies that $\mathfrak{q}B_{\mathfrak{p}}$ is maximal, whence $\mathfrak{q}B_{\mathfrak{p}} = \mathfrak{q}'B_{\mathfrak{p}}$. By ideal correspondence (proposition I.3.1.17), we have $\mathfrak{q} = \mathfrak{q}'$. □

***Theorem* III.1.2.4.** — *Let B be an extension ring over A which is integral over A. Then every prime ideal $\mathfrak{p}$ of A lies under some prime ideal $\mathfrak{q}$ of B (i. e., $\mathfrak{q} \cap A = \mathfrak{p}$).*

Proof. — By lemma III.1.1.11, $B_{\mathfrak{p}} = (A \setminus \mathfrak{p})^{-1} B$ is integral over $A_{\mathfrak{p}}$. Let $\mathfrak{m}$ be any maximal ideal of $B_{\mathfrak{p}}$. Then $\mathfrak{m} \cap A_{\mathfrak{p}}$ is maximal in $A_{\mathfrak{p}}$ by corollary III.1.2.2 and hence the unique maximal ideal $\mathfrak{p}A_{\mathfrak{p}}$. Let $\mathfrak{q}$ be the contraction of $\mathfrak{m}$ in B, which is prime. Since the diagram

$$\begin{tikzcd} A \arrow[r, hook] \arrow[d, "\alpha"'] & B \arrow[d, "\beta"] \\ A_{\mathfrak{p}} \arrow[r, hook] & B_{\mathfrak{p}} \end{tikzcd}$$

commutes, we have $\mathfrak{q} \cap A = \beta^{-1}(\mathfrak{m}) \cap A = \alpha^{-1}(\mathfrak{m} \cap A_{\mathfrak{p}}) = \alpha^{-1}(\mathfrak{p}A_{\mathfrak{p}}) = \mathfrak{p}$. In fact, the maximal ideals of $B_{\mathfrak{p}}$ corresponds to the prime ideals of B lying over $\mathfrak{p}$, cf. proposition I.3.1.17. □

Alternative proof for maximal ideals. — There is a weaker version of theorem III.1.2.4: every maximal ideal $\mathfrak{m}$ of A lies under some prime ideal $\mathfrak{n}$ of B. This follows from theorem III.1.2.4 and corollary III.1.2.2. However, this weaker version also has another proof.

It suffices to show $\mathfrak{m}B \neq B$. In this case, any maximal ideal $\mathfrak{n}$ of B containing $\mathfrak{m}B$ satisfies $\mathfrak{m} \subseteq \mathfrak{n} \cap A \neq (1)$, whence $\mathfrak{n} \cap A = \mathfrak{m}$. By way of contradiction, assume that $\mathfrak{m}B = B$. Then we can write $1 = \sum_{i=1}^{n} m_i x_i$ with $x_i \in B$ and $m_i \in \mathfrak{m}$. Setting $C = A[x_1, \dots, x_n]$, we have $\mathfrak{m}C = C$, and C is finite over A by corollary III.1.1.4, say, $C = (u_1, \dots, u_r)$ as an A-module. Write $u_i = \sum_{j \neq i} a_{ij} u_j \in \mathfrak{m}C$ with $a_{ij} \in \mathfrak{m}$. If $\Delta = \det(\delta_{ij} - a_{ij})_{ij}$ where δ_{ij} denotes the Kronecker delta, then $\Delta u_j = 0$ for all $j = 1, \dots, r$, and hence $\Delta C = 0$ (cf. lemma I.2.3.5). Because $1 \in C$, we obtain $\Delta = 0$. However, $\Delta \equiv 1 \mod \mathfrak{m}$. It follows that $1 \in \mathfrak{m}$, a contradiction. □

***Corollary* III.1.2.5.** — *Let $f\colon A \to \Omega$ be a ring map into an algebraically closed field Ω, and let B be an extension ring of A which is integral over A. Then f can be extended to a map $B \to \Omega$.*

(III.3) It is commonly assumed that $\mathfrak{q}'$ should be prime as well. This is, however, unnecessary.

Proof. — Let $\mathfrak{p} = \ker f$, which is a prime ideal of A. Then f factors as $A \twoheadrightarrow A/\mathfrak{p} \to K = \operatorname{Quot} A/\mathfrak{p} \hookrightarrow \Omega$. Theorem III.1.2.4 gives a prime ideal $\mathfrak{q}$ of B lying over A. Since $B/\mathfrak{q}$ is integral over $A/\mathfrak{p}$ (lemma III.1.1.11), we see that $L = \operatorname{Quot} B/\mathfrak{q}$ is algebraic over K. Hence, the question of extending $f\colon A \to \Omega$ to B where B is algebraic over A reduces to extending $K \hookrightarrow \Omega$ to L where L/K is algebraic. This is possible by classic field theory.

***Fact* III.1.2.6.** — *Let B be an extension ring of A which is finitely generated and integral over A. If $\mathfrak{p}$ is a prime ideal of A, then there are only finitely many prime ideals lying over $\mathfrak{p}$.*

Proof. — By corollary III.1.1.8, B is a finite A-algebra. As seen in the proof of theorem III.1.2.4, the prime ideals of B lying above $\mathfrak{p}$ correspond to maximal ideals of $C = B_{\mathfrak{p}}/\mathfrak{p}B_{\mathfrak{p}}$. If $k = A_{\mathfrak{p}}/\mathfrak{p}A_{\mathfrak{p}}$, then C is a finite k-algebra and in particular an artinian ring (proposition I.5.2.9). By proposition I.5.4.3, C has only finitely many maximal ideals.

***Going-up theorem* III.1.2.7.** — *Let B be an extension ring over A which is integral over A. Let $\mathfrak{p}_1 \subseteq \cdots \subseteq \mathfrak{p}_n$ be a chain of prime ideals of A, and let $\mathfrak{q}_1 \subseteq \cdots \subseteq \mathfrak{q}_m$ be a shorter ($m < n$) chain of prime ideals of B such that $\mathfrak{q}_i$ lies over $\mathfrak{p}_i$ for all $i = 1, \dots, m$. Then the chain in B can be extended to a chain $\mathfrak{q}_1 \subseteq \cdots \subseteq \mathfrak{q}_n$ of prime ideals of B such that $\mathfrak{q}_i$ lies over $\mathfrak{p}_i$ for all $i = 1, \dots, n$.*

Proof. — By induction, it suffices to consider the case $m = 1$ and $n = 2$. Let $\bar{A} = A/\mathfrak{p}_1$ and $\bar{B} = B/\mathfrak{q}_1$ with $\bar{\mathfrak{p}}_1$ the image of $\mathfrak{p}_1$ in $\bar{A}$. Then $\bar{B}$ is an extension ring of $\bar{A}$ which is integral over $\bar{A}$ by lemma III.1.1.11. By theorem III.1.2.4, there exists a prime ideal $\bar{\mathfrak{q}}_2$ of $\bar{B}$ lying over $\bar{\mathfrak{p}}_2$. Lifting back $\bar{\mathfrak{q}}_2$ to B, we obtain a prime ideal $\mathfrak{q}_2$ containing $\mathfrak{q}_1$. Since the diagram

$$\begin{tikzcd} A \arrow[r, hook] \arrow[d] & B \arrow[d] \\ \bar{A} \arrow[r, hook] & \bar{B} \end{tikzcd}$$

commutes, we see that $\mathfrak{q}_2$ lies over $\mathfrak{p}_2$, that is, $\mathfrak{q}_2 \cap A = \mathfrak{p}_2$.

We can also study the Galois action on extensions of integral domains.

III.1.2.8. — Let G be a finite group of automorphisms of a ring A. Then $A^G = \{x \in A \mid \sigma(x) = x \text{ for all } \sigma \in G\}$ is the so-called ring of *G-invariants*.

***Proposition* III.1.2.9.** — *Let G be a finite group of automorphisms of a ring A.*

(i) *Then A is integral over A^G.*

(ii) *Let $\mathfrak{p}$ be a prime ideal of A^G, and let Σ be the set of all prime ideals lying above $\mathfrak{p}$. Then G acts transitively on Σ (i. e., for any $\mathfrak{p}_1, \mathfrak{p}_2 \in \Sigma$ there exists $\sigma \in G$ such that $\mathfrak{p}_1 = \sigma(\mathfrak{p}_2)$). In particular, Σ is* finite.

Proof. — (i) is rather trivial. Any $x \in A$ is the root of the monic polynomial $\prod_{\sigma \in G}(t - \sigma(x))$ in t. Since the coefficients of this polynomial are elementary, they are G-invariant. Hence x is integral over A^G.

For (ii), let $\mathfrak{p}_1, \mathfrak{p}_2 \in \Sigma$. If $x \in \mathfrak{p}_1$, then $\prod_{\sigma \in G} \sigma(x) \in \mathfrak{p}_1 \cap A^G = \mathfrak{p} \subseteq \mathfrak{p}_2$, so $\sigma(x) \in \mathfrak{p}_2$ for some $\sigma \in G$. It follows that $\mathfrak{p}_1 \subseteq \bigcup_{\sigma \in G} \sigma(\mathfrak{p}_2)$. Of course, the $\sigma(\mathfrak{p}_2)$ are prime ideals, so the prime avoidance lemma I.1.3.9 says that $\mathfrak{p}_1 \subseteq \sigma(\mathfrak{p}_2)$ for some $\sigma \in G$.

***Fact* III.1.2.10.** — *Let A be an integrally closed domain with field of fractions $K = \operatorname{Quot} A$. Let L/K be a Galois extension with (infinite) Galois group $G = \operatorname{Gal}(L/K)$, and let B be the integral closure of A in L. Then $\sigma(B) = B$ for all $\sigma \in G$, and $A = B^G$.*

Proof. — Let $x \in B$ with an integrality equation with coefficients in A. Since $A^G = A$, applying some $\sigma \in G$ to this equation shows that $\sigma(x)$ is integral over A. Hence $\sigma(B) \subseteq B$ for all $\sigma \in G$. Applying σ^{-1} gives $B = \sigma^{-1}(\sigma(B)) \subseteq \sigma^{-1}(B)$ for all $\sigma^{-1} \in G$. Thus $\sigma(B) = B$ for all $\sigma \in G$.

For the second statement, recall from Galois theory that $L^G = K$. Then $B^G = B \cap L^G = B \cap K = A$ because A is integrally closed in K. ◻

There is also a generalisation of proposition III.1.2.9 to infinite groups in the context of field automorphisms.

***Proposition* III.1.2.11.** — *Let A be an integrally closed domain with field of fractions $K = \operatorname{Quot} A$. Let L/K be an (infinite) normal algebraic extension with automorphism group $G = \operatorname{Aut}(L/K)$, and let B be the integral closure of A in L. Let $\mathfrak{p}$ be a prime ideal of A, and let Σ be the set of all prime ideals of B lying over $\mathfrak{p}$. Then G acts transitively on Σ.*

Insert reference to field theory, e. g., Lang's algebra, p. 251, Prop. 6.11.

Proof. — Even for infinite extensions L/K, we can split it into a Galois extension L/L^G and a purely inseparable extension L^G/K. If $L^G \neq K$, then K is necessarily of positive characteristic $\operatorname{char} K = p$. Recall that the elements of L^G have minimal polynomials over K of the form $t^{p^n} - a \in K[t]$. Let $\tilde{A}$ be the integral closure of A in L^G. Then one easily checks that $\mathfrak{q} = \{x \in \tilde{A} \mid x^{p^n} \in \mathfrak{p} \text{ for some } n\}$ is a prime ideal of $\tilde{A}$ lying over $\mathfrak{p}$. Moreover, any other prime ideal lying over $\mathfrak{p}$ must be contained in $\mathfrak{q}$, whence $\mathfrak{q}$ is the only one by theorem III.1.2.4. By transitivity of integrality, B is the integral closure of $\tilde{A}$ in L. Hence, through replacing K by L^G we can assume w. l. o. g. that L/K is Galois.

The case where L/K is finite was proven in proposition III.1.2.9 with G acting on B. For the infinite case, we need infinite Galois theory. Recall that G is endowed with the Krull topology, i. e., the subspace topology $G \hookrightarrow \prod_{F/K \text{ finite}} \operatorname{Gal}(F/K)$, $\sigma \mapsto (\sigma|_F)_F$ where $\operatorname{Gal}(F/K)$ are finite discrete. Then G is profinite and hence compact.

Let $\mathfrak{p}_1, \mathfrak{p}_2 \in \Sigma$. For any finite Galois subextension F/K of L/K, the set

$$S(F) := \{\sigma \in G \mid \sigma(\mathfrak{p}_1 \cap F) = \mathfrak{p}_2 \cap F\}$$

is non-empty by the finite case (note that $\mathfrak{p}_i \cap F$ is prime in the integral closure of A in F). Now, the family of all $S(F)$ has the finite intersection property, i. e., $\bigcap_{i=1}^n S(F_i) \neq \varnothing$ for any finite Galois extensions $F_1, \dots, F_n$ of K. Indeed, if F is the composite of $F_1, \dots, F_n$, then it is also finite Galois, and $\bigcap_{i=1}^n S(F_i) \supseteq S(F) \neq \varnothing$. Moreover, all sets $S(F) \subseteq \operatorname{Gal}(F/K)$ are closed in G, so G being compact implies that $\bigcap_{F/K \text{ finite}} S(F)$ is non-empty. If $\sigma \in \bigcap_{F/K \text{ finite}} S(F)$, then $\sigma(\mathfrak{p}_1) = \mathfrak{p}_2$ globally since $L = \bigcup_{F/K \text{ finite}} F$. ◻

The following is the counterpart to the going-up theorem III.1.2.7, which requires integral domains.

***Going-down theorem* III.1.2.12.** — *Let A be an integrally closed domain, and let B be an extension domain of A which is integral over A. Let $\mathfrak{p}_1 \supseteq \dots \supseteq \mathfrak{p}_n$ be a chain of prime ideals of A, and let $\mathfrak{q}_1 \supseteq \dots \supseteq \mathfrak{q}_m$ be a shorter $(m < n)$ chain of prime ideals of B such that $\mathfrak{q}_i$ lies over $\mathfrak{p}_i$ for all $i = 1, \dots, m$. Then the chain in B can be extended to a chain $\mathfrak{q}_1 \supseteq \dots \supseteq \mathfrak{q}_n$ such that $\mathfrak{q}_i$ lies over $\mathfrak{p}_i$ for all $i = 1, \dots, n$.*

Proof. — By induction, it suffices to consider the case $m = 1$ and $n = 2$. The goal is to show that $\mathfrak{p}_2$ is a contraction of some prime ideal $\mathfrak{q}_2 B_{\mathfrak{q}_1}$ of $B_{\mathfrak{q}_1}$. By corollary I.3.1.18, this is equivalent to $\mathfrak{p}_2 B_{\mathfrak{q}_1} \cap A = \mathfrak{p}_2$.

Every element of $\mathfrak{p}_2 B_{\mathfrak{q}_1}$ is of the form x/s with $x \in \mathfrak{p}_2 B$ and $s \in B \setminus \mathfrak{q}_1$. Since B is integral over A, lemma III.1.1.21 says that $x \in \sqrt{\mathfrak{p}_2 B}$ is integral over $\mathfrak{p}_2$. Due to proposition III.1.1.22, the minimal polynomial of x over $K = \operatorname{Quot} A$ is of the form $f(t) = t^r + u_1 t^{r-1} + \dots + u_r$ with $u_1, \dots, u_r \in \mathfrak{p}_2$. Now, additionally assume that $x/s \in A$. Then the minimal polynomial of s over K

is $g(t) = t^r + v_1 t^{r-1} + \dots + v_r$ with $v_i = u_i(s/r)^i$; indeed, $g(s) = (s/x)^r f(x) = 0$. However, $s \in B$ is integral over A, so proposition III.1.1.22 again shows that $v_i \in A$ for all $i = 1, \dots, r$.

Suppose that $x/s \notin \mathfrak{p}_2$. Then $v_i(x/s)^i = u_i \in \mathfrak{p}_2$ implies $v_i \in \mathfrak{p}_2$ because $\mathfrak{p}_2$ is prime. Furthermore, $g(s) = 0$ implies $s^r \in \mathfrak{p}_2 B \subseteq \mathfrak{p}_1 B \subseteq \mathfrak{q}_1$. It follows that $s \in \mathfrak{q}_1$, a contradiction. Hence $x/s \in \mathfrak{p}_2$. This shows $\mathfrak{p}_2 B_{\mathfrak{q}_1} \cap A = \mathfrak{p}_2$, as desired.

Alternative proof. — Let $K = \operatorname{Quot} A$, let L be a normal algebraic field extension of K containing B, and let $\tilde{A} \supseteq B$ be the integral closure of A in L. By induction, it suffices to consider the case $m = 1$ and $n = 2$. By theorem III.1.2.4, $\mathfrak{q}_1$ (resp. $\mathfrak{p}_2$) lies under some prime ideal $\mathfrak{r}_1$ (resp. $\mathfrak{r}_2$) of $\tilde{A}$. The going-up theorem III.1.2.7 gives another prime $\mathfrak{r}_1'$ of $\tilde{A}$ lying over $\mathfrak{p}_1$ and containing $\mathfrak{r}_2$. By proposition III.1.2.11, there exists a K-automorphism σ on L such that $\sigma(\mathfrak{r}_1') = \mathfrak{r}_1$. It follows that $\sigma(\mathfrak{r}_2) \subseteq \sigma(\mathfrak{r}_1') = \mathfrak{r}_1$ and $\sigma(\mathfrak{r}_2) \cap A = \mathfrak{r}_2 \cap A = \mathfrak{p}_2$. Taking $\mathfrak{q}_2 = \sigma(\mathfrak{r}_2) \cap B$, we obtain $\mathfrak{q}_2 \cap A = \mathfrak{p}_2$ and $\mathfrak{q}_2 \subseteq \mathfrak{r}_1 \cap B = \mathfrak{q}_1$, as desired.

Insert [Mat86, Thm. 9.5] on going-down for flat algebras.

[Mat86, Thm. 9.6] missing.

Insert [Mat86, Exer. 9.1, 9.2] on going-down for flat algebras.

III.1.3. Noether's normalisation lemma and Jacobson rings.

***Noether's normalisation lemma* III.1.3.1** (1926). — *Let k be a field, and let $A \neq 0$ be a finitely generated k-algebra. Then there exist $y_1, \dots, y_r \in A$ which are algebraically independent over k and such that A is integral over $k[y_1, \dots, y_r]$.*

Proof for infinite fields k. — There is a less technical proof if k is infinite. Write $A = k[x_1, \dots, x_n]$ as a k-algebra. We proceed by induction on n. If $x_1, \dots, x_n$ are algebraically independent, then we are done; this case includes the base case $A = k$ or $A = k[x_1]$. So, suppose that $x_1, \dots, x_n$ are algebraically independent, say, there is a non-zero polynomial $f \in k[x_1, \dots, x_n]$ with $f(x_1, \dots, x_{n-1}, x_n) = 0$. Put $y_i \coloneqq x_i - \lambda_i x_n$ for $i = 1, \dots, n-1$ with $\lambda_i \in k$. Then we obtain

$$0 = f(y_1 + \lambda_1 x_n, \dots, x_{n-1} + \lambda_{n-1} x_n, x_n) = F(\lambda_1, \dots, \lambda_{n-1}, 1) x_n^d + g_1 x_n^{d-1} + \dots + g_d$$

where $g_i \in A' \coloneqq k[y_1, \dots, y_{n-1}]$ and where F is the homogeneous part of maximal degree $d = \deg f$ of f. Since k is infinite, we can choose $\lambda_1, \dots, \lambda_{n-1} \in k$ such that $F(\lambda_1, \dots, \lambda_{n-1}, 1) \neq 0$, and hence x_n is integral over A'. It follows that $x_i = y_i + \lambda_i x_n$ for all $i = 1, \dots, n-1$ are also integral over A', and thus A is integral over A'. Applying the induction hypothesis to A', we are done.

***Remark* III.1.3.2.** — The proof in the case where k is infinite, we see that we can choose $y_1, \dots, y_r$ to be linear combinations of $x_1, \dots, x_n$. From a geometrical standpoint, this means the following: suppose that k is an algebraically closed field (in particular infinite). If X is an affine variety of $\mathbb{A}^n(k)$ with coordinate ring $\Gamma(X) = A \neq 0$, then there exists an r-dimensional linear subspace L of $\mathbb{A}^n(k)$ and a linear map $\mathbb{A}^n(k) \to L$ mapping X *onto* L. This map gives a branched covering of L.

General proof. — The general case uses a similar variable shift, albeit a polynomial instead of a linear one. As in the previous proof, let $A = k[x_1, \dots, x_n]$ as a k-algebra and we proceed by induction on n. Suppose that $x_1, \dots, x_n$ are algebraically independent, so $f(x_1, \dots, x_n) = 0$.

We use the following technique. Write $f(t_1, \dots, t_n) = \sum_{i=1}^r \lambda_i M_i$ where each monomial M_i is of the form $\prod_{j=1}^n t_j^{a_{ij}}$. For positive integers $d_1, \dots, d_{n-1}$ and $d_n = 1$, let $\sum_{j=1}^n a_{ij} d_j$ be the *weight* associated to $M_i = \prod_{j=1}^n t_j^{a_{ij}}$. By choosing $d_1, \dots, d_{n-1}$ appropriately, we can assume that the monomials in f have pairwise different weights. Namely, for a large enough prime number p, let $d_i = p^i$ so that the weights are represented by unique numbers of base p.

Now, put $y_i \coloneqq x_i - x_n^{d_i}$ for $i = 1, \dots, n-1$. Then we obtain

$$0 = f(y_1 + x_n^{d_1}, \dots, y_{n-1} + x_n^{d_{n-1}}, x_n) = \lambda x_n^d + g_1 x_n^{d-1} + \dots + g_d$$

where $\lambda, g_i \in A' := k[y_1, \dots, y_{n-1}]$. Note that d is the highest weight of monomials of f, so by choice of the $d_1, \dots, d_{n-1}$ only one monomial of f contributes a term of degree d (all other having degree less than d), whence $\lambda \in k \setminus \{0\}$. Hence, x_n is integral over A', and it follows that $x_i = y_i + x_n^{d_i}$ for all $i = 1, \dots, n-1$ are also integral over A'. Thus, A is integral over A'. Applying the induction hypothesis to A', we are done. ∎

***Zariski's lemma* III.1.3.3.** — *Let A/k be a field extension where A is also a finitely generated k-algebra. Then A/k is a finite.*

First proof. — This is an elementary proof: let $A = k[x_1, \dots, x_n]$, say. We proceed by induction on n. The statement is clearly true in the case $n = 1$. For $n > 1$, let $B = k[x_1]$ and let $K = k(x_1)$ be its field of fractions. By induction hypothesis, A/K is finite, so each of $x_2, \dots, x_n$ satisfies a monic polynomial with coefficients in K. If f is the product of the denominators of all these coefficients, then $x_2, \dots, x_n$ are integral over B_f. Hence A and therefore $K \subseteq A$ is integral over B_f.

Suppose that x_i is transcendental over k so that B is a unique factorisation domain. Then B is integrally closed (lemma III.1.1.24), and by corollary III.1.1.13 B_f is as well. But K is integral over B_f, so $B_f = K$, clearly a contradiction (e. g., $1/(xf) \notin B_f$). Hence, x_i is algebraic over k and K/k is finite. Thus, $A/K/k$ is finite. ∎

Second proof. — This proof uses Noether's normalisation lemma III.1.3.1, which says that there exist algebraically independent $y_1, \dots, y_r \in A$ such that A is integral over $B = k[y_1, \dots, y_r]$. By corollary III.1.1.8, A is finite over B. Now, since A is a field, B is also a field by lemma III.1.2.1, whence $r = 0$ and $B = k$. Thus, A/k is finite. ∎

***Weak Nullstellensatz* III.1.3.4.** — *Let k be an algebraically closed field. Then each maximal ideal in the polynomial ring $k[t_1, \dots, t_n]$ is of the form $(t_1 - a_1, \dots, t_n - a_n)$ with $a_i \in k$.*

Proof. — Clearly, each ideal of the form $(t_1 - a_1, \dots, t_n - a_n)$ is maximal in $A = k[t_1, \dots, t_n]$. Conversely, let $\mathfrak{m}$ be a maximal ideal of A. By Zariski's lemma VIII.1.1.9, $A/\mathfrak{m}$ is a finite field extension of k. Since k is algebraically closed, we in fact have $A/\mathfrak{m} \cong k$. Now, consider the canonical ring map $\phi\colon A \twoheadrightarrow A/\mathfrak{m} \xrightarrow{\sim} k$. Setting $a_i = \phi(t_i)$ for $i = 1, \dots, n$, we see that $(t_1 - a_1, \dots, t_n - a_n) \subseteq \ker\phi = \mathfrak{m}$, which must be equal. ∎

***Lemma* III.1.3.5.** — *Let B be an extension domain of A which is finitely generated over A. Then there exist $s \in A \setminus \{0\}$ and $y_1, \dots, y_n \in B$ that are algebraically independent over A, such that B_s is integral over B'_s where $B' = A[y_1, \dots, y_n]$.*

Proof. — Let $S = A \setminus \{0\}$ so that $K = S^{-1}A = \operatorname{Quot} A$. Then $S^{-1}B$ is a finitely generated K-algebra, so by Noether's normalisation lemma III.1.3.1 there are $x_1, \dots, x_n \in S^{-1}B$ algebraically independent over K such that $S^{-1}B$ is integral over $K[x_1, \dots, x_n]$. Let $B = A[z_1, \dots, z_m]$ as an A-algebra. Then each z_i (as an element of $S^{-1}B$) is integral over $K[x_1, \dots, x_n]$. We can assume that the x_i are of the form $x_i = y_i/s$ with $y_1, \dots, y_n \in B$, which are algebraically independent over A, and $s \in S$. By multiplying the integrality equation of each z_i by a sufficiently large power s^r of s, we see that $s^r z_i$ is integral over $B' = A[y_1, \dots, y_n]$ for all i. Hence the z_i and thus B_s is integral over B'_s. ∎

***Lemma* III.1.3.6.** — *Let B be an extension domain of A which is finitely generated over A. Then there exists $s \in A \setminus \{0\}$ such that any ring map $f\colon A \to \Omega$ into an algebraically closed field Ω with $f(s) \neq 0$ extends to a map $B \to \Omega$.*

Proof. — Let $s \in A \setminus \{0\}$ and let $y_1, \dots, y_n \in B$ that are algebraically independent over A as in lemma III.1.3.5, such that B_s is integral over B'_s where $B' = A[y_1, \dots, y_n]$. We extend f to B' by, e. g., $y_i \mapsto 0$. Then we further extend to B'_s, which is possible since $f(s) \neq 0$. Finally, we extend to B_s by corollary III.1.2.5, which restricts to $B \to \Omega$. ∎

***Proposition* III.1.3.7.** — *Let B be an extension domain of A which is finitely generated over A. If* $\operatorname{rad} A = (0)$*, then* $\operatorname{rad} B = (0)$*.*

Proof. — We have to show that for $v \in B \setminus \{0\}$, there is a maximal ideal of B not containing v. Lemma III.1.3.6 applied to $A \subseteq B_v$ gives some $s \in A \setminus \{0\}$. Let $\mathfrak{m}$ be a maximal ideal of A not containing s, which exists since $\operatorname{rad} A = 0$. Then the canonical map $A \twoheadrightarrow A/\mathfrak{m} =: k$, which does not map s to zero, extends to a ring map $\phi\colon B_v \to \bar{k}$ by properties of s. Note that $\phi(v) \in \bar{k}^\times$ since $v \in B_v^\times$, so $\phi(v) \neq 0$. Moreover, we have $k \subseteq B/\mathfrak{n} \subseteq \bar{k}$ where $\mathfrak{n} = \ker \phi \cap B$. Since $\bar{k}$ is algebraic and hence integral over k, the integral domain $B/\mathfrak{n}$ is integral over k and hence a field by lemma III.1.2.1. Thus, $\mathfrak{n}$ is a maximal ideal in B not containing v. ∎

***Proposition* III.1.3.8** (characterisation of jacobson rings)**.** — *Let A be a ring. Then the following are equivalent:*

(i) *every prime ideal of A is an intersection of maximal ideals;*
(ii) *in every quotient ring of A, the nilradical equals the Jacobson radical;*
(iii) *every non-maximal prime ideal of A is the intersection of all prime ideals which contain it properly.*

A ring satisfying one of the above conditions is called a ***Jacobson ring*** *(or* ***Hilbert ring****).*

Proof. — Suppose (i), and let $\mathfrak{a}$ be an ideal of A. We of course have $\operatorname{nil} A/\mathfrak{a} \subseteq \operatorname{rad} A/\mathfrak{a}$. Conversely, for any prime ideal $\bar{\mathfrak{p}}$ of $A/\mathfrak{a}$ the lift $\mathfrak{p}$ to A is also a prime ideal containing $\mathfrak{a}$. Since $\mathfrak{p}$ is an intersection of maximal ideals containing $\mathfrak{a}$, which correspond to maximal ideals of $A/\mathfrak{a}$, we see that $\bar{\mathfrak{p}}$ is also an intersection of maximal ideals of $A/\mathfrak{a}$. Thus, $\operatorname{rad} A/\mathfrak{a} \subseteq \operatorname{nil} A/\mathfrak{a}$, and we obtain (ii).

Next, assume (ii), and let $\mathfrak{p}$ be a non-maximal prime ideal of A. By assumption, $\operatorname{rad} A/\mathfrak{p} = \operatorname{nil} A/\mathfrak{p} = (0)$. Since the maximal ideals of $A/\mathfrak{p}$ are not (0) and correspond to maximal ideals of A which contain $\mathfrak{p}$ properly, we can lift back $\operatorname{rad} A/\mathfrak{p} = (0)$ to see that $\mathfrak{p}$ is the intersection of all *maximal* ideals of A which contain $\mathfrak{p}$ properly. As the intersection of all *prime* ideals of A which contain $\mathfrak{p}$ strictly also contains $\mathfrak{p}$, this intersection equals $\mathfrak{p}$. This proves (iii).

Finally, that (iii) implies (i) is the most difficult. Suppose that (i) is false, i. e., there is a prime ideal which is not an intersection of maximal ideals. By passing to the quotient ring of A, we may assume w. l. o. g. that A is an integral domain with $\operatorname{rad} A \neq (0)$. So, let $f \in \operatorname{rad} A \setminus \{0\}$ so that $A_f \neq 0$. Then A_f has a maximal ideal, and its contraction in A is a prime ideal $\mathfrak{p}$ with $f \notin \mathfrak{p}$ which is maximal w. r. t. this property. Hence $\mathfrak{p}$ is not maximal (since $f \in \operatorname{rad} A$) and is not the intersection of prime ideals containing $\mathfrak{p}$ properly (prime ideals that properly contain $\mathfrak{p}$ must contain f, so their intersection contains f). Thus, (iii) is false. ∎

***Proposition* III.1.3.9.** — *Let B be an A-algebra where A is a Jacobson ring. If B is*

(i) *integral over A or*
(ii) *finitely generated as an A-algebra,*

then B is Jacobson. This is in particular true if $A = \mathbb{Z}$ (i. e., B is a finitely generated ring) or A a field.

Proof. — W. l. o. g., we assume that A is a subring of B.

For the first part, suppose that B is integral over A, and let $\mathfrak{q}$ be a prime ideal of B. Then the contraction $\mathfrak{q} \cap A$ in A is an intersection $\bigcap_i \mathfrak{m}_i$ of maximal ideals of A by proposition III.1.3.8. By the going-up theorem III.1.2.7 and corollary III.1.2.2, each $\mathfrak{m}_i$ lies under a maximal ideal $\mathfrak{n}_i$

of B which contains $\mathfrak{q}$. Since $\mathfrak{q} \subseteq \bigcap_i \mathfrak{n}_i$ with $\mathfrak{q} \cap A = \bigcap_i \mathfrak{n}_i \cap A = \bigcap_i \mathfrak{m}_i$, corollary III.1.2.3 implies $\mathfrak{q} = \bigcap_i \mathfrak{n}_i$.

For the second part, suppose that B is finitely generated over A, and let $\mathfrak{q}$ be a prime ideal of B. Then $B/\mathfrak{q}$ is an integral domain and finitely generated over $A/(\mathfrak{q} \cap A)$. Since $\operatorname{rad} A/(\mathfrak{q} \cap A) = \operatorname{nil} A/(\mathfrak{q} \cap A) = (0)$ by proposition III.1.3.8, we have $\operatorname{rad} B/\mathfrak{q} = (0)$ by proposition III.1.3.7. Thus, lifting back to B, we see that $\mathfrak{q}$ is the intersection of all maximal ideals of B which contain $\mathfrak{q}$. ⟤

***Proposition* III.1.3.10.** — *Let A be a ring. Then A is Jacobson iff any finitely generated A-algebra B which is also a field is finite.*

Proof. — Suppose that A is Jacobson, and let B be a finitely generated A-algebra which is also a field; w. l. o. g., suppose that A is a subdomain of B. Note that $\operatorname{rad} A = \operatorname{nil} A = (0)$ by proposition III.1.3.8. Then lemma III.1.3.6 gives an $s \in A \setminus \{0\}$, and let $\mathfrak{m}$ be a maximal ideal of A not containing s (since $s \notin \operatorname{rad} A$). By properties of s, the canonical map $A \twoheadrightarrow A/\mathfrak{m} = k$ extends to a map $\phi\colon B \to \bar{k}$. Because B is a field, ϕ is injective, and $\phi(B) \cong B$ is algebraic, finitely generated and hence finite over k (cf. corollary III.1.1.8). Thus, B is also finite over A by scalar restriction.

Conversely, let $\mathfrak{p}$ be a non-maximal prime ideal of A. Let $B = A/\mathfrak{p}$, and let $f \in B \setminus \{0\}$. Then B_f is a finitely generated A-algebra, e. g., by $1/f$. If B_f were a field, then it is finite over A by hypothesis. In particular, B_f is finite over B, so integral over B by corollary III.1.1.8. Then lemma III.1.2.1 gives the contradiction that B is a field. Thus, B_f is not a field and has a non-trivial prime ideal, whose contraction $\mathfrak{p}'$ in B is non-trivial with $f \notin \mathfrak{p}'$. In other words, the intersection of all non-trivial prime ideals of B is the zero ideal. Thus, A is Jacobson by proposition III.1.3.8. ⟤

§ III.2. VALUATION RINGS

CHAPTER IV

FLATNESS

PART II

ALGEBRAIC NUMBER THEORY

CHAPTER CONTENTS

CHAPTER V

INTRODUCTION

***Reference* V.0.0.1.** — Mainly [Neu99].

***Notation* V.0.0.2.** — – Despite that [Neu99] uses *replete divisor*, we will use the more common term *Arakelov divisor* instead.

CHAPTER VI

PART III

HOMOLOGICAL ALGEBRA

CHAPTER CONTENTS

VI.0.0.1. — We almost exclusively follow [Wei94].

CHAPTER VII

STANDARD TOOLS

PART IV

ALGEBRAIC GEOMETRY

CHAPTER CONTENTS

INTRODUCTION

***Reference* VII.0.0.1.** — Mainly [GW20].

***Notation* VII.0.0.2.** — – We use the convention that the empty topological space is *not* connected.

CHAPTER VIII

PRELIMINARIES

§ VIII.1. PREVARIETIES

Although algebraic geometry studies systems of polynomial equations in several variables over arbitrary rings, we will first go the classical route by considering polynomials over algebraically closed fields, originally over $\mathbb{C}$. This will serve as a motivation and guideline for what is to come.

***Convention* VIII.1.0.1.** — *Throughout, let k be an algebraically closed field. It is, in particular, infinite.*

***Motivation* VIII.1.0.2.** — Consider $f = T_2^2 - T_1^2(T_1 + 1) \in k[T_1, T_2]$. Although the real picture as in figure VIII.1 can be misleading as $\mathbb{R}$ is not algebraically closed, it is still often helpful. We can see that it is a 'one-dimensional' object. In fact, except for the origin $(0, 0)$, it locally looks like a line, i. e., there is a unique tangent line. In other words, the origin is 'singular' and all other points are 'smooth'. By the theorem of implicit functions from analysis, the set of roots of f is diffeomorphic to $\mathbb{R}$ at point (x_1, x_2) if the Jacobian $(\partial f/\partial T_1 \quad \partial f/\partial T_2)$ has full rank 1. Since partial derivatives can be done over arbitrary base fields, we will later on take this as the definition of algebraic smoothness.

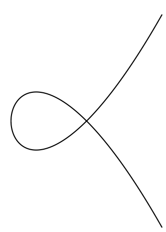

Figure VIII.1: Solutions of $T_2^2 - T_1^2(T_1 + 1) = 0$ in $\mathbb{R}^2$.

VIII.1.1. Affine algebraic sets.

***Definition* VIII.1.1.1.** — *Let $M \subseteq k[T_1, \dots, T_n]$ be a subset of polynomials. The **vanishing set** of M is the set*

$$\mathrm{V}(M) := \{(t_1, \dots, t_n) \in k^n \mid f(t_1, \dots, t_n) = 0 \text{ for all } f \in M\}.$$

If $M = \{f_i \mid i \in I\}$, then we also write $\mathrm{V}(f_i, i \in I)$ instead of $\mathrm{V}(M)$.

***Remark* VIII.1.1.2.** — (i) We have $\mathrm{V}(M) = \mathrm{V}((M))$.

(ii) By Hilbert's basis theorem I.5.3.4, $k[T_1, \dots, T_n]$ is noetherian, so the ideal (M) is generated by finitely many elements of M. Hence, for any $M \subseteq k[T_1, \dots, T_n]$ there are finitely many $f_1, \dots, f_r \in M$ such that $\mathrm{V}(M) = \mathrm{V}(f_1, \dots, f_r)$.

(iii) $\mathrm{V}(-)$ is inclusion-reversing: $M' \subseteq M$ implies $\mathrm{V}(M') \supseteq \mathrm{V}(M)$.

***Proposition* VIII.1.1.3** (Zariski topology)**.** — *The sets* $\mathrm{V}(\mathfrak{a})$ *where* $\mathfrak{a}$ *runs through all ideals of* $k[T_1, \dots, T_n]$ *define a topology on* k^n *by being closed, the so-called **Zariski topology**.*

Proof. — We need to check that this is a topology.

(i) We have $\varnothing = \mathrm{V}(1)$ and $k^n = \mathrm{V}(0)$.

(ii) For every family $(\mathfrak{a}_i)_{i \in I}$ of ideals $\mathfrak{a}_i$, we have

$$\bigcap_{i \in I} \mathrm{V}(\mathfrak{a}_i) = \{x \in k^n \mid f(x) = 0 \text{ for all } i \in I \text{ and } f \in \mathfrak{a}_i\} = \mathrm{V}\left(\bigcup_{i \in I} \mathfrak{a}_i\right) = \mathrm{V}\left(\sum_{i \in I} \mathfrak{a}_i\right) \quad \text{(VIII.1.1.3.1)}$$

because $(\bigcup_{i \in I} \mathfrak{a}_i) = \sum_{i \in I} \mathfrak{a}_i$ as ideals.

(iii) Since $\mathfrak{a}\mathfrak{b} \subseteq \mathfrak{a} \cap \mathfrak{b} \subseteq \mathfrak{a}, \mathfrak{b}$, we have $\mathrm{V}(\mathfrak{a}) \cup \mathrm{V}(\mathfrak{b}) \subseteq \mathrm{V}(\mathfrak{a} \cap \mathfrak{b}) \subseteq \mathrm{V}(\mathfrak{a}\mathfrak{b})$. Conversely, if $x \in \mathrm{V}(\mathfrak{a}\mathfrak{b}) \setminus \mathrm{V}(\mathfrak{a})$, then $f(x) \neq 0$ for some $f \in \mathfrak{a}$, but $(fg)(x) = 0$ for all $g \in \mathfrak{b}$ (since $fg \in \mathfrak{a}\mathfrak{b}$). Hence, $g(x) = 0$ for all $g \in \mathfrak{b}$ and $x \in \mathrm{V}(\mathfrak{b})$. Thus, we have equality:

$$\mathrm{V}(\mathfrak{a}) \cup \mathrm{V}(\mathfrak{b}) = \mathrm{V}(\mathfrak{a} \cap \mathfrak{b}) = \mathrm{V}(\mathfrak{a}\mathfrak{b}). \quad \text{(VIII.1.1.3.2)}$$

***Definition* VIII.1.1.4.** — *Define* $\mathbb{A}^n(k)$ *to be the topological space* k^n *with the Zariski topology and call it the **affine space** of dimension* n *over* k. *A closed subsets of* $\mathbb{A}^n(k)$ *is called an **affine algebraic set**.*

***Fact* VIII.1.1.5.** — *Finite subsets of* $\mathbb{A}^n(k)$ *are closed.*

Proof. — Single points are closed since for $x = (x_1, \dots, x_n) \in k^n$, we have $\{x\} = \mathrm{V}(\mathfrak{m}_x)$ where $\mathfrak{m}_x := (T_1 - x_1, \dots, T_n - x_n)$ is the kernel of the evaluation map $\mathrm{ev}_x \colon k[T_1, \dots, T_n] \to k$.

***Warning* VIII.1.1.6.** — The advantage of the Zariski topology is that it can be defined for any k. However, the drawback is that it is very coarse: for $n > 0$, it is never Hausdorff.

Insert reference [GW20, Prop. 1.20].

***Example* VIII.1.1.7.** — For $n = 1$, the ring $k[T]$ is a principal ideal domain, so the affine algebraic sets of $\mathbb{A}^1(k)$ are of the form $\mathrm{V}(f)$ with $f \in k[T]$. Since every non-zero polynomial has at most finitely many roots, the affine algebraic sets of $\mathbb{A}^1(k)$ are $\mathbb{A}^1(k)$ itself and finite subsets of it. In particular, if k is infinite then $\mathbb{A}^1(k)$ is not Hausdorff.

***Example* VIII.1.1.8.** — For $n = 2$, the affine algebraic sets of $\mathbb{A}^2(k)$ are more difficult to describe. The obvious ones are:

- $\mathbb{A}^2(k)$;
- single points $\mathrm{V}(\mathfrak{m}_x) = \{x\}$;
- $\mathrm{V}(f)$ for irreducible polynomials $f \in k[T_1, T_2]$.

In fact, these sets are *irreducible* in the sense that every other affine algebraic set is a finite union of these. The reason is that all of these sets are of the form $\mathrm{V}(\mathfrak{p})$ for $\mathfrak{p}$ a prime ideal in $k[T_1, T_2]$. We will show all of this later. As a consequence, $\mathbb{A}^2(k)$ is not Hausdorff if k is infinite.

Insert reference [GW20, §(1.5), Prop. 5.31].

In the case of algebraically closed fields, the deep relation between algebra and geometry is established by *Hilbert's Nullstellensatz*.

in ComAlg
er Jacobson is
l.

***Zariski's lemma* VIII.1.1.9.** — *Let K be a (not necessarily algebraically closed) field, and let A be a finitely generated K-algebra. Then A is Jacobson, i. e., for every prime ideal $\mathfrak{p}$ of A we have $\mathfrak{p} = \bigcap_{\mathfrak{m} \supseteq \mathfrak{p}} \mathfrak{m}$ where $\mathfrak{m}$ runs through all maximal ideals of A. If $\mathfrak{m}$ is a maximal ideal of A, then the field extension $(A/\mathfrak{m})/K$ is finite.*

Replace all of this with appropriate references in ComAlg.

VIII.1.1.10. — The proof is based on Noether's normalisation theorem. Recall that a ring map $R \to R'$ is *integral* if each element of R' is a root of a monic polynomial over R. It is *finite* if R' is additionally a finitely generated R-algebra, i. e., of finite type. Finiteness is equivalent to R' being a finite R-module.

Note that finite algebras and finitely generated algebras are quite different, and that the notion integral has nothing to do with integral domain. However, we will call an algebra *integral* if it is an integral domain. !!!

Replace by reference in ComAlg.

***Theorem* VIII.1.1.11** (Noether's normalisation theorem). — *Let K be a field, and let $A \neq 0$ be a finitely generated K-algebra. Then there exists some $n \geq 0$ and $t_1, \dots, t_n \in A$ such that the K-algebra map $K[T_1, \dots, T_n] \to A$, $T_i \mapsto t_i$ is injective and finite.* $a_{\mapsto}$

Replace by reference in ComAlg going up.

***Lemma* VIII.1.1.12.** — *Let $A \to B$ be an injective integral ring map of integral domains. Then A is a field iff B is a field.*

Maybe, this lemma needs to be replaced by something from ComAlg.

***Lemma* VIII.1.1.13.** — *Let L/K be a field extension where L is a finitely generated K-algebra. Then L is a finite extension of K.*

Proof. — Applying Noether normalisation to L, we obtain a finite inclusion $K[T_1, \dots, T_n] \hookrightarrow L$ of K-algebras. By the previous lemma, we must have $n = 0$, so $K \hookrightarrow L$ is a finite field extension. ∎

Edit references.

Proof of Zariski's lemma VIII.1.1.9. — The second statement follows from the previous lemma: if $\mathfrak{m}$ is a maximal ideal of A, then $A/\mathfrak{m}$ is a field extension of K which is finitely generated over K, whence finite over K.

Insert reference.

For the first statement, we firstly make an observation. Let L be a finite field extension of K and let $\phi\colon A \to K$ be a K-algebra map. Then $\operatorname{im}\phi$ is an integral K-algebra that is finite over K. By going up, $\operatorname{im}\phi$ is a field, and hence $\ker\phi$ is a maximal ideal of A.

Insert reference.

To see that A is Jacobson, let $\mathfrak{p}$ be a prime ideal of A. By passing to $A/\mathfrak{p}$, we may assume that A is an integral domain and it suffices to show that the Jacobson radical $\operatorname{rad} A$ is trivial. Here comes the trick: for any $x \in A \setminus \{0\}$, the K-algebra $A[x^{-1}]$ is finitely generated (and non-trivial since A is integral). If $\mathfrak{m}$ is a maximal ideal of $A[x^{-1}]$, then $A[x^{-1}]/\mathfrak{m}$ is a finite field extension of K by the second statement proven above. By the previous observation, the kernel of $A \to A[x^{-1}] \to A[x^{-1}]/\mathfrak{m}$ is a maximal ideal which does not contain x. Thus, $x \notin \operatorname{rad} A$. ∎

As an immediate corollary:

***Weak Nullstellensatz* VIII.1.1.14.** — *Let $K = k$ be algebraically closed.*

(i) *Let A be a finitely generated k-algebra, and let $\mathfrak{m}$ be a maximal ideal of A. Then $A/\mathfrak{m} = k$.*

(ii) *Let $\mathfrak{m}$ be a maximal ideal in $k[T_1, \dots, T_n]$. Then there exists a point $x = (x_1, \dots, x_n) \in \mathbb{A}^n(k)$ such that $\mathfrak{m} = \mathfrak{m}_x := (T_1 - x_1, \dots, T_n - x_n)$. In other words, the (geometric) points of $\mathbb{A}^n(k)$ are precisely the (algebraic) maximal ideals of $k[T_1, \dots, T_n]$.*

Proof. — For (i), we know that $(A/\mathfrak{m})/k$ is a finite field extension by Zariski's lemma VIII.1.1.9. Since k is algebraically closed, we must have $A/\mathfrak{m} = k$.

For (ii), let x_i be the image of T_i under $\phi\colon k[T_1, \dots, T_n] \to k[T_1, \dots, T_n]/\mathfrak{m} = k$. Then $\ker\phi = \mathfrak{m}$ contains the maximal ideal $\mathfrak{m}_x$, so both are equal. ∎

VIII.1.1.15. — Recall the definition I.1.3.16 of a radical of an ideal $\mathfrak{a}$ in a ring A:

$$\sqrt{\mathfrak{a}} := \{f \in A \mid f^r \in \mathfrak{a} \text{ for some } r > 0\}.$$

We say that an ideal $\mathfrak{a}$ is radical if $\mathfrak{a} = \sqrt{\mathfrak{a}}$.

***Notation* VIII.1.1.16.** — For a subset $Z \subseteq \mathbb{A}^n(k)$, we denote the ideal of functions vanishing on Z by

$$\mathrm{I}(Z) := \{f \in k[T_1, \dots, T_n] \mid f(x) = 0 \text{ for all } x \in Z\}.$$

***Hilbert's Nullstellensatz* VIII.1.1.17.** — (i) $\mathrm{I}(\mathrm{V}(\mathfrak{a})) = \sqrt{\mathfrak{a}}$ *for all ideals* $\mathfrak{a}$ *in* $k[T_1, \dots, T_n]$.
(ii) $\mathrm{V}(\mathrm{I}(Z)) = \overline{Z}$ *for all subsets* $Z \subseteq \mathbb{A}^n(k)$.

Thus, there is a bijection

$$\{\textit{radical ideals of } k[T_1, \dots, T_n]\} \underset{\mathrm{I}}{\overset{\mathrm{V}}{\rightleftarrows}} \{\textit{affine algebraic sets of } \mathbb{A}^n(k)\}$$

given by mutually inverse maps $\mathrm{V}(-)$ *and* $\mathrm{I}(-)$*, whose restrictions give another bijection*

$$\{\textit{maximal ideals of } k[T_1, \dots, T_n]\} \longleftrightarrow \{\textit{points of } \mathbb{A}^n(k)\}.$$

Proof. — To prove (i), observe the following general fact: for any $f \in k[T_1, \dots, T_n]$ and any $x \in \mathbb{A}^n(k)$ we have $f(x) = 0$ iff $f \in \mathfrak{m}_x := \ker \mathrm{ev}_x$. This implies that $\mathrm{I}(Z) = \bigcap_{x \in Z} \mathfrak{m}_x$ for arbitrary subsets $Z \subseteq \mathbb{A}^n(k)$, and that $x \in \mathrm{V}(\mathfrak{a})$ iff $\mathfrak{a} \subseteq \mathfrak{m}_x$. Hence, $\mathrm{I}(\mathrm{V}(\mathfrak{a})) = \bigcap_{x \in \mathrm{V}(\mathfrak{a})} \mathfrak{m}_x = \bigcap_{\mathfrak{m}_x \supseteq \mathfrak{a}} \mathfrak{m}_x$. By the Weak Nullstellensatz VIII.1.1.14, Zariski's lemma VIII.1.1.9, and proposition I.1.3.17, we finally obtain $\bigcap_{\mathfrak{m}_x \supseteq \mathfrak{a}} \mathfrak{m}_x = \bigcap_{\mathfrak{m} \supseteq \mathfrak{a}} \mathfrak{m} = \bigcap_{\mathfrak{p} \supseteq \mathfrak{a}} \mathfrak{p} = \sqrt{\mathfrak{a}}$ where $\mathfrak{m}$ runs through all maximal ideals and where $\mathfrak{p}$ runs through all prime ideals of $k[T_1, \dots, T_n]$.

The proof of (ii) does not need the Weak Nullstellensatz VIII.1.1.14. Obviously, $Z \subseteq \mathrm{V}(\mathrm{I}(Z))$ and $\mathrm{V}(\mathrm{I}(Z))$ is closed, which is why $\overline{Z} \subseteq \mathrm{V}(\mathrm{I}(Z))$. Conversely, let $\mathrm{V}(\mathfrak{a}) \subseteq \mathbb{A}^n(k)$ be a closed subset containing Z. Then $f(x) = 0$ for all $x \in Z$ and all $f \in \mathfrak{a}$, so $\mathfrak{a} \subseteq \mathrm{I}(Z)$ and hence $\mathrm{V}(\mathrm{I}(Z)) \subseteq \mathrm{V}(\mathfrak{a})$. ◻

VIII.1.2. Topological properties of affine spaces.

Definition* VIII.1.2.1.** — *A non-empty topological space X is called* ***irreducible *if X is not the union of two proper closed subsets. The same notion applies to subsets of X under the induced subspace topology. A maximal irreducible subset of X is called an* ***irreducible component*** *of X.*

***Proposition* VIII.1.2.2** (characterisation of irreducible spaces)**.** — *Let X be a non-empty topological space. Then the following are equivalent.*

(i) *X is irreducible.*
(ii) *Any two non-empty open subsets of X meet.*
(iii) *Every non-empty open subset is dense in X.*
(iv) *Every non-empty open subset is connected.*
(v) *Every non-empty open subset is irreducible.*

Proof. — The equivalence of (i) and (ii) follows by taking complements. Recall from point-set topology that a set is dense iff it meets every open subset; this is the equivalence of (ii) and (iii). The equivalence of (ii) and (iv) follows easily by definition.

Now assume (iii). Let $U \subseteq X$ be non-empty and open. Every non-empty open subset of U is dense in X, by assumption, and likewise dense in U. Hence, as we have already shown, U is irreducible and thus, (v) holds. On the other hand, (v) obviously implies (i). ◻

***Fact* VIII.1.2.3.** — *Let $f\colon X \to Y$ be a continuous map of topological space. If $Z \subseteq X$ is irreducible, then the image $f(Z)$ is irreducible.*

Proof. — By proposition VIII.1.2.2, any two non-empty open subsets of $f(Z)$ do not meet since their preimages do not meet. ∎

***Lemma* VIII.1.2.4.** — *Let X be a topological space. Then the subspace $Y \subseteq X$ is irreducible iff $\overline{Y}$ is irreducible. In particular, irreducible components are closed.*

Proof. — A non-empty open subset of Y is an open subset of X that meets Y. Moreover, an open subset of X meets Y iff it meets $\overline{Y}$. Through this reduction, we see that two non-empty open subsets of Y meet if they meet in $\overline{Y}$ and vice versa. The statement follows from proposition VIII.1.2.2. ∎

***Observation* VIII.1.2.5.** — Let X be a topological space, and fix an open subset U. Suppose that $Z \subseteq X$ is an irreducible closed subset with $Z \cap U \neq \varnothing$. Then $Z \cap U$ is an irreducible open subset of Z by proposition VIII.1.2.2 and hence an irreducible closed subset of U. Proposition VIII.1.2.2 also implies that $Z \cap U$ is dense in Z, so $\overline{Z \cap U} = Z$. With lemma VIII.1.2.4, this establishes a bijection

$$\begin{aligned}\{Y \subseteq U \text{ irreducible closed}\} &\longleftrightarrow \{Z \subseteq X \text{ irreducible closed with } Z \cap U \neq \varnothing\}\\ Y &\longmapsto \overline{Y} \text{ (in } X)\\ Z \cap U &\longleftarrow\!\shortmid Z.\end{aligned}$$

***Proposition* VIII.1.2.6.** — *Let X be a non-empty topological space. Every irreducible subset of X is contained in an irreducible component. In particular, X is the union of its irreducible components.*

Proof. — The second statement follows from the first since singleton sets are irreducible. The first statement is an application of Zorn's lemma. Consider all irreducible subsets of X containing a distinguished irreducible subset $X' \subseteq X$. Any chain $\{Y_\alpha\}_\alpha$ of irreducible subsets has their union $\bigcup_\alpha Y_\alpha$ as an upper bound. Indeed, if $\bigcup_\alpha Y_\alpha = Z_1 \cup Z_2$ for proper closed subsets Z_1 and Z_2, then any Y_α must be contained in Z_1 or Z_2 because Y_α is irreducible. In particular, there is some $Y_\beta \subseteq Z_1$ meeting $Z_1 \setminus Z_2$ and some $Y_\gamma \subseteq Z_2$ meeting $Z_2 \setminus Z_1$, contradicting the chain condition for Y_β and Y_γ. Thus, Zorn's lemma shows that X' is contained in an irreducible component. ∎

***Lemma* VIII.1.2.7.** — *Let X be a topological space and let $X = \bigcup_{i\in I} U_i$ be an open covering by connected open subsets U_i.*

(i) *If X is not connected, then there exists a subset $\varnothing \neq J \subsetneq I$ such that $U_j \cap U_i = \varnothing$ for all $j \in J$ and all $i \in I \setminus J$.*

(ii) *If X is connected and if all U_i are irreducible, then X is irreducible.*

Proof. — By assumption of (i), we can write $X = V_1 \amalg V_2$ where V_1 and V_2 are non-empty, open and closed. It follows that each U_i is either contained in V_1 or V_2, and we may set $J := \{i \in I \mid U_i \subseteq V_1\}$. This proves (i).

To show (ii), we will heavily employ proposition VIII.1.2.2. Let $V \subseteq X$ be a non-empty open subset; our goal is to show that V is dense in X. It suffices to show that V meets all U_i because then the closure $\overline{V}$ in X contains each irreducible U_i and hence, $\overline{V} = X$. So, consider $J = \{i \in I \mid V \cap U_i = \varnothing\}$. Suppose that there are $j \in J$ and $i \in I \setminus J$ such that U_j meets U_i. Since V meets U_i, irreducibility of U_i tells us that V and U_j meet inside U_i, contradicting $j \in J$. Hence, $X = \bigcup_{j\in J} U_j \amalg \bigcup_{i \in I\setminus J} U_i$ is a disjoint union. But X is connected, so J must be empty. ∎

***Proposition* VIII.1.2.8.** — *Let $Z \subseteq \mathbb{A}^n(k)$ be an affine algebraic set. Then Z is irreducible iff $\mathrm{I}(Z)$ is a prime ideal. In particular, $\mathbb{A}^n(k)$ is irreducible.*

Proof. — By definition, Z is irreducible iff it is not a union of two proper closed subsets. We know from equation (VIII.1.1.3.1) that closed subsets can be expressed as an intersection of sets $\mathrm{V}(f)$. Hence, the condition of irreducibility is equivalent to the property that if $\mathrm{V}(fg) = \mathrm{V}(f) \cup \mathrm{V}(g) \supseteq Z$ (equality follows from equation (VIII.1.1.3.2)) then $\mathrm{V}(f) \supseteq Z$ or $\mathrm{V}(g) \supseteq Z$. In other words, Z is irreducible iff $fg \in \mathrm{I}(Z)$ implies $f \in \mathrm{I}(Z)$ or $g \in \mathrm{I}(Z)$, that is, $\mathrm{I}(Z)$ is a prime ideal.

***Remark* VIII.1.2.9.** — The bijection in Hilbert's Nullstellensatz VIII.1.1.17 restricts to

$$\{\text{prime ideals in } k[T_1, \dots, T_n]\} \longleftrightarrow \{\text{irreducible closed subsets of } \mathbb{A}^n(k)\}.$$

Definition* VIII.1.2.10.** — *A topological space X is called* ***quasi-compact *if every open covering of X has a finite subcovering.*

***Warning* VIII.1.2.11.** — A closed subspace of a quasi-compact space is again quasi-compact. This need not hold for open subspaces, however.

Definition* VIII.1.2.12.** — *A topological space X is called* ***noetherian *if every descending chain $X \supseteq Z_1 \supseteq Z_2 \supseteq \cdots$ of closed subsets of X becomes stationary. This property is equivalent to the property that every non-empty set of closed subsets of X has a minimal element w. r. t. inclusion.*

***Lemma* VIII.1.2.13.** — *Let X be a topological space with a finite covering $X = \bigcup_{i=1}^r X_i$ by noetherian subspaces. Then X is noetherian.*

Proof. — Let $X \supseteq Z_1 \supseteq Z_2 \supseteq \cdots$ be a descending chain of closed subsets. Then $(Z_j \cap X_i)_j$ is a descending chain of closed subsets of X_i, so it stabilises starting at, say, index $N_i \geq 1$. So, starting at $N := \max\{N_1, \dots, N_r\}$ the descending chain of X stabilises to $\bigcup_{i=1}^r (Z_N \cap X_i) = Z_n$.

***Lemma* VIII.1.2.14.** — *Let X be a noetherian topological space.*

(i) *Every subspace is noetherian.*
(ii) *Every subset of X is quasi-compact (including X itself).*
(iii) *Every subset of X has finitely many irreducible components.*

Proof. — Let $Y \subseteq X$ be a subspace, and consider a descending chain $(Z_i)_i$ of closed subsets of Y. Then $(\overline{Z}_i)_i$ is a descending chain of closed subsets of X, which becomes stationary. Since $\overline{Z}_i \cap Y = Z_i$, the chain $(Z_i)_i$ becomes stationary as well. This proves (i).

To show (ii), by (i), it suffices to show that X is quasi-compact. Let $X = \bigcup_i U_i$ be an open covering, and let $\mathcal{U}$ be the set of all finite unions of the U_i. Since X is noetherian, $\mathcal{U}$ must have a maximal element V. If $V \neq X$, then there would be some U_i such that $V \subsetneq V \cup U_i \in \mathcal{U}$, a contradiction. Thus, $V = X$ is a finite union of certain U_i, i. e., the covering has a finite subcover.

Again by (i), it suffices to show that X has finitely many irreducible components in order to prove (iii). We claim that X is the union of finitely many irreducible subsets. Then, due to proposition VIII.1.2.6, we can write $X = \bigcup_{i=1}^r X_i$ where the X_i are (closed) irreducible components of X. If $Y \subseteq X$ is any irreducible component, then any $Y \cap X_i$ is a closed subset of Y by lemma VIII.1.2.4, so $Y = \bigcup_{i=1}^r (Y \cap X_i)$ implies $Y = Y \cap X_i$ for some i, whence $Y = X_i$.

To show the claim, let $\mathcal{Z}$ be the set of closed subsets of X that are not a finite union of irreducible subsets. If $\mathcal{Z}$ is non-empty, then there would be a minimal element $Z \in \mathcal{Z}$ since X is noetherian. In particular, Z would not be irreducible and hence, a union of two proper closed subsets not lying in $\mathcal{Z}$. These two subsets are a finite union of irreducible subsets and therefore Z as well, a contradiction. Thus, $\mathcal{Z}$ is empty and, in particular, X is a finite union of irreducible subsets.

***Proposition* VIII.1.2.15.** — *Any subspace of $\mathbb{A}^n(k)$ is noetherian.*

Proof. — By lemma VIII.1.2.14, it suffices to show that $\mathbb{A}^n(k)$ is noetherian. Hilbert's Nullstellensatz VIII.1.1.17 says that descending chains of closed subsets of $\mathbb{A}^n(k)$ correspond to ascending chains of radical ideals in $k[T_1, \dots, T_n]$. But we know from Hilbert's basis theorem I.5.3.4 that the latter is noetherian. ∎

With the same trick, we can obtain a weak version of the *primary composition in noetherian rings*.

[GW20] mentions [AM16, ch. 4, ch. 7].

***Corollary* VIII.1.2.16.** — *Let $\mathfrak{a}$ be a radical ideal of $k[T_1, \dots, T_n]$. Then $\mathfrak{a}$ is the intersection of finitely many prime ideals that do not contain each other. These prime ideals are uniquely determined by $\mathfrak{a}$.*

VIII.1.3. Category of affine algebraic sets.

Naturally, morphisms between affine algebraic sets should be polynomial.

Definition* VIII.1.3.1.** — *Let $X \subseteq \mathbb{A}^m(k)$ and $Y \subseteq \mathbb{A}^n(k)$ be affine algebraic sets. A* ***morphism of affine algebraic sets *$X \to Y$ is a map $f\colon X \to Y$ of sets such that there are $f_1, \dots, f_n \in k[T_1, \dots, T_m]$ with $f(x) = (f_1(x), \dots, f_n(x))$ for all $x \in X$. We denote the set of morphisms $X \to Y$ with $\operatorname{Hom}(X, Y)$.*

***Observation* VIII.1.3.2.** — Morphisms of algebraic sets can be always extended to morphisms $\mathbb{A}^m(k) \to \mathbb{A}^n(k)$, although not uniquely unless $X = \mathbb{A}^m(k)$. So, if $f = (f_1, \dots, f_n)\colon \mathbb{A}^m(k) \to \mathbb{A}^n(k)$ is a morphism, then we obtain a k-algebra map $\mathrm{ev}_f\colon k[S_1, \dots, S_n] \to k[T_1, \dots, T_m]$, $S_i \mapsto f_i$. For any closed subset $\mathrm{V}(\mathfrak{a}) \subseteq \mathbb{A}^n(k)$, the preimage $f^{-1}(\mathrm{V}(\mathfrak{a})) = \mathrm{V}(\mathrm{ev}_f(\mathfrak{a})) \subseteq \mathbb{A}^m(k)$ is again closed. Thus, morphisms of affine algebraic sets are continuous.

***Observation* VIII.1.3.3.** — Let $X \subseteq \mathbb{A}^m(k)$, $Y \subseteq \mathbb{A}^n(k)$ and $Z \subseteq \mathbb{A}^r(k)$ be affine algebraic sets, and let $f = (f_1, \dots, f_n)\colon X \to Y$ and $g = (g_1, \dots, g_r)\colon Y \to Z$ be morphisms. Then $g \circ f = (h_1, \dots, h_r)$ where $h_i \in k[T_1, \dots, T_m]$ is the image of g_i under the map $\mathrm{ev}_f\colon k[S_1, \dots, S_n] \to k[T_1, \dots, T_m]$, $S_i \mapsto f_i$. Thus, $g \circ f$ is again a morphism and we obtain the category of affine algebraic sets.

***Example* VIII.1.3.4.** — (i) The map $\mathbb{A}^1(k) \to \mathrm{V}(T_1 - T_2^2)$, $x \mapsto (x^2, x)$ is a morphism. It is even an isomorphism with inverse $(x, y) \mapsto y$.

(ii) The map $\mathbb{A}^1(k) \to \mathrm{V}(T_2^2 - T_1^2(T_1 + 1))$, $x \mapsto (x^2 - 1, x(x^2 - 1))$ is a morphism. If $\operatorname{char} k \neq 2$, then it is not bijective because -1 and 1 are both mapped to the origin $(0, 0)$. If $\operatorname{char} k = 2$, then the map is bijective; however, it is still not an isomorphism.

Show this. It is not possible to use degrees due to characteristic.

(iii) We can identify $\mathrm{Mat}_n(k)$ with $\mathbb{A}^{n^2}(k)$ so that it has the structure of an affine algebraic set. Then the map $\mathrm{Mat}_n(k) \to \mathbb{A}^1(k)$, $A \mapsto \det A$ is a morphism.

(iv) The complex exponential function $\exp\colon \mathbb{A}^1(\mathbb{C}) \to \mathbb{A}^1(\mathbb{C})$ is not a morphism as it is not continuous w. r. t. the Zariski topology: the preimage $\exp^{-1}(0) = \{2\pi k\mathrm{i} \mid k \in \mathbb{Z}\}$ is not closed by example VIII.1.1.7.

***Warning* VIII.1.3.5.** — Bijective morphisms are not necessarily isomorphisms. Another example includes $\mathbb{A}^1(k) \to \mathrm{V}(T_2^2 - T_1^3)$, $t \mapsto (t^2, t^3)$, which is a bijective morphism, but not an isomorphism. Any inverse morphism would be a polynomial $f \in k[T_1, T_2]$ with $f(t^2, t^3) = t$. However, the possible degrees of the monomials of f are $2n + 3m$ for $m, n \geq 0$, and $1 \neq 2n + 3m$.

***Motivation* VIII.1.3.6.** — The notion of affine algebraic sets still has some caveats:

(i) Open subsets of affine algebraic sets are not naturally affine algebraic sets. This prevents us from gluing affine algebraic sets along open subsets, which is desirable for geometric objects.

(ii) We cannot distinguish between $\mathrm{V}(Y) \cap \mathrm{V}(X) \subseteq \mathbb{A}^2(k)$ (one point) and $\mathrm{V}(Y) \cap \mathrm{V}(X^2 - Y) \subseteq \mathbb{A}^2(k)$ ('two' points), although the geometric situation is a bit different. We will later see that this also happens with fibres of morphisms.

(iii) Affine algebraic sets require algebraically closed fields and do not allow polynomial equations over general rings.

We will solve (i) in the following; the problem arises as affine algebraic sets are embedded into affine space. Problems (ii) and (iii) will be solved by scheme theory.

***Definition* VIII.1.3.7.** — *Let $X \subseteq \mathbb{A}^n(k)$ be an affine algebraic set. The set of functions $\mathrm{Hom}(X, \mathbb{A}^1(k))$ is naturally endowed with a k-algebra structure by pointwise addition and multiplication, where k corresponds to constant functions. Every $f \in k[T_1, \dots, T_n]$ induces a morphism (function) $X \to \mathbb{A}^1(k)$, $x \mapsto f(x)$, so we obtain a k-algebra map $k[T_1, \dots, T_n] \twoheadrightarrow \mathrm{Hom}(X, \mathbb{A}^1(k))$ with kernel $\mathrm{I}(X)$. We call the k-algebra*

$$\Gamma(X) := k[T_1, \dots, T_n]/\mathrm{I}(X) \cong \mathrm{Hom}(X, \mathbb{A}^1(k))$$

*the **affine coordinate ring** of X.*

***Notation* VIII.1.3.8.** — Let $X \subseteq \mathbb{A}^n(k)$ be an affine algebraic set. For $x \in X$, let

$$\mathfrak{m}_x := \{f \in \Gamma(X) \mid f(x) = 0\}$$

be the kernel of the evaluation map $\mathrm{ev}_x \colon \Gamma(X) \to k$, $f \mapsto f(x)$. In other words, it is the image of the maximal ideal $(T_1 - x_1, \dots, T_n - x_n)$ under $\Gamma(\mathbb{A}^n(k)) = k[T_1, \dots, T_n] \twoheadrightarrow \Gamma(X)$. As ev_x is surjective, $\mathfrak{m}_x$ is a maximal ideal with $\Gamma(X)/\mathfrak{m}_x = k$.

***Definition* VIII.1.3.9.** — *Let X be an affine algebraic set. For any ideal $\mathfrak{a}$ of $\Gamma(X)$, define the **vanishing set** of $\mathfrak{a}$ as*

$$\mathrm{V}(\mathfrak{a}) := \{x \in X \mid f(x) = 0 \textit{ for all } f \in \mathfrak{a}\} = \mathrm{V}(\pi^{-1}(\mathfrak{a})) \cap X$$

*where $\pi \colon k[T_1, \dots, T_n] \twoheadrightarrow \Gamma(X)$. The vanishing sets over all ideals $\mathfrak{a}$ define the closed sets of the **Zariski topology** on X, which agree with the closed sets of X w.r.t. the subspace topology of $X \subseteq \mathbb{A}^n(k)$. For any $f \in \Gamma(X)$, further define the **principal open subset** of f as*

$$\mathrm{D}(f) := \{x \in X \mid f(x) \neq 0\} = X \setminus \mathrm{V}(f).$$

***Lemma* VIII.1.3.10.** — *The principal open subsets $\mathrm{D}(f)$ with $f \in \Gamma(X)$ form a basis of the Zariski topology. Finite intersections of principal open subsets are again principal open.*

Proof. — Since $\mathrm{D}(f) \cap \mathrm{D}(g) = \mathrm{D}(fg)$ for all $f, g \in \Gamma(X)$, the second statement is clear. For the first statement, let $U = X \setminus \mathrm{V}(\mathfrak{a})$ be an open subset for some ideal $\mathfrak{a}$. Since $\Gamma(X)$ is noetherian, we can write $\mathfrak{a} = (f_1, \dots, f_n)$ so that $\mathrm{V}(\mathfrak{a}) = \bigcap_{i=1}^n \mathrm{V}(f_i)$, whence $U = \bigcup_{i=1}^n \mathrm{D}(f_i)$. ◻

***Proposition* VIII.1.3.11.** — *For an affine algebraic set X, the affine coordinate ring $\Gamma(X)$ is a reduced finite type k-algebra. Moreover, X is irreducible iff $\Gamma(X)$ is an integral domain.*

Proof. — $\Gamma(X) = k[T_1, \dots, T_n]/\mathrm{I}(X)$ is obviously finitely generated. It is also reduced since $\mathrm{I}(X)$ is radical by Hilbert's Nullstellensatz VIII.1.1.17. Finally, proposition VIII.1.2.8 tells us that X is irreducible iff $\mathrm{I}(X)$ is prime, which is the second statement. ◻

***Theorem* VIII.1.3.12.** — *Associating to a morphism $f \colon X \to Y$ of affine algebraic sets the k-algebra map*

$$\Gamma(f) = f^* \colon \mathrm{Hom}(Y, \mathbb{A}^1(k)) \cong \Gamma(Y) \longrightarrow \mathrm{Hom}(X, \mathbb{A}^1(k)) \cong \Gamma(X), \qquad g \longmapsto g \circ f$$

induces an equivalence of categories

$$\Gamma \colon \{\textit{affine algebraic sets}\}^{\mathrm{op}} \longrightarrow \{\textit{reduced finitely generated } k\textit{-algebras}\}.$$

This restricts to an equivalence of categories

$$\Gamma \colon \{\textit{irreducible affine algebraic sets}\}^{\mathrm{op}} \longrightarrow \{\textit{integral finitely generated } k\textit{-algebras}\}.$$

Proof. — We first check that Γ is fully faithful, i. e., for affine algebraic sets $X \subseteq \mathbb{A}^m(k)$ and $Y \subseteq \mathbb{A}^n(k)$ we construct an inverse to the map $\Gamma\colon \operatorname{Hom}(X,Y) \to \operatorname{Hom}(\Gamma(Y),\Gamma(X))$. Given any $\phi\colon \Gamma(Y) \to \Gamma(X)$, there exists a lift[VIII.1]

$$\begin{tikzcd} k[S_1,\dots,S_n] \arrow[r, dashed, "\tilde\phi"] \arrow[d] & k[T_1,\dots,T_m] \arrow[d] \\ \Gamma(Y) \arrow[r, "\phi"] & \Gamma(X). \end{tikzcd}$$

Define $f := (\tilde\phi(S_1),\dots,\tilde\phi(S_n))\colon X \to Y$ as the image of ϕ. This is indeed the inverse of Γ: given $f = (f_1,\dots,f_n)\colon X \to Y$, let $\phi = \Gamma(f)$. Then $\tilde\phi(g) = g(f_1,\dots,f_n)$ for all $g \in k[S_1,\dots,S_n]$, so setting $g = S_i$ gives $\tilde\phi(S_i) = f_i$, as desired. Conversely, given $\phi\colon \Gamma(Y) \to \Gamma(X)$, we obtain $f = (\tilde\phi(S_1),\dots,\tilde\phi(S_n))\colon X \to Y$. So, for $\bar g \in \Gamma(Y)$ with $g \in k[S_1,\dots,S_n]$ this yields

$$\Gamma(f)(\bar g) = g(\tilde\phi(S_1),\dots,\tilde\phi(S_n)) \bmod \mathrm{I}(X) = \tilde\phi(g(S_1,\dots,S_n)) \bmod \mathrm{I}(X) = \phi(\bar g).$$

We now show that Γ is essentially surjective, i. e., for every reduced finite type k-algebra A there exists an affine algebraic set X with $A \cong \Gamma(X)$. Since $A \cong k[T_1,\dots,T_n]/\mathfrak{a}$ for a radical ideal $\mathfrak{a}$, we can choose $X = \mathrm{V}(\mathfrak{a}) \subseteq \mathbb{A}^n(k)$ so that $\Gamma(X) = k[T_1,\dots,T_n]/\mathfrak{a}$ by Hilbert's Nullstellensatz VIII.1.1.17.

Finally, the restriction follows from proposition VIII.1.3.11. ∎

***Proposition* VIII.1.3.13.** — *Let $f\colon X \to Y$ be a morphism of affine algebraic sets, and let $\Gamma(f)\colon \Gamma(Y) \to \Gamma(X)$ be the corresponding map of affine coordinate rings. Then $\Gamma(f)^{-1}(\mathfrak{m}_x) = \mathfrak{m}_{f(x)}$ for all $x \in X$.*

Proof. — We have $g \in \mathfrak{m}_{f(x)} \subseteq \Gamma(Y)$ iff $g(f(x)) = \Gamma(f)(g)(x) = 0$ iff $g \in \Gamma(f)^{-1}(\mathfrak{m}_x)$. ∎

VIII.1.4. Prevarieties.

We will now introduce a notion due to *Séminaire de Chevalley*, a precursor to ringed spaces and alluding to how real smooth manifolds can be described as a topological space with sets of smooth functions on each open subset.

***Definition* VIII.1.4.1.** — *Let K be a field.*

(i) *A **space with functions** over K is defined as a topological space X with a family $\mathcal{O}_X$ of K-subalgebras $\mathcal{O}_X(U) \subseteq \operatorname{map}(U,K)$ for every open subset U of X satisfying the following properties:*

(a) Restriction axiom: for $U' \subseteq U \subseteq X$ open and $f \in \mathcal{O}_X(U)$, the restriction $f|_{U'} \in \operatorname{map}(U',K)$ lies again in $\mathcal{O}_X(U')$.

(b) Gluing axiom: for $U_i \subseteq X$ open and $f_i \in \mathcal{O}_X(U_i)$ where $i \in I$ $f_i|_{U_i\cap U_j} = f_j|_{U_i\cap U_j}$ for all $i,j \in I$, the unique function $f\colon \bigcup_{i\in I} U_i \to K$ determined by $f|_{U_i} = f_i$ for all $i \in I$ lies again in $\mathcal{O}_X(\bigcup_{i\in I} U_i)$.

We will often write X instead of $(X,\mathcal{O}_X)$.

(ii) *A **morphism of spaces with functions** $g\colon (X,\mathcal{O}_X) \to (Y,\mathcal{O}_Y)$ is a continuous map $g\colon X \to Y$ such that for all $V \subseteq Y$ open and $f \in \mathcal{O}_Y(V)$ the function $f \circ g|_{g^{-1}(V)}\colon g^{-1}(V) \to K$ lies in $\mathcal{O}_X(g^{-1}(V))$.*

Clearly, spaces with functions over K form a category.

***Definition* VIII.1.4.2.** — *Let X be a space with functions, and let $U \subseteq X$ be an open subspace. We then denote by $(U,\mathcal{O}_X|_U)$ the space U with functions $\mathcal{O}_X|_U(V) = \mathcal{O}_X(V)$ for $V \subseteq U$ open.*

(VIII.1) $k[S_1,\dots,S_n]$ is a free k-algebra.

We now specialise to affine algebraic sets.

***Definition* VIII.1.4.3.** — *Let X be an irreducible affine algebraic set, in which case $\Gamma(X)$ is an integral domain. We call $\mathrm{K}(X) := \operatorname{Quot}\Gamma(X)$ the **function field** of X.*

***Observation* VIII.1.4.4.** — We want to turn an affine algebraic set $X \subseteq \mathbb{A}^n(k)$ into a space of functions over k. To this end, the k-subalgebras $\mathcal{O}_X(U)$ will be contained in $\mathrm{K}(X)$. If we understand $\Gamma(X)$ as $\operatorname{Hom}(X, \mathbb{A}^1(k))$, then each element of $\mathrm{K}(X)$ will be of the form f/g for $f, g \in \Gamma(X)$ with $g \neq 0$. However, f/g usually does not define a function on X because g may have roots on X, but it certainly is a function on $\mathrm{D}(g)$ (or an even bigger open subset of X depending on whether f/g is completely reduced).

***Lemma* VIII.1.4.5.** — *Let X be an irreducible affine algebraic set, and consider $f_1/g_1, f_2/g_2 \in \mathrm{K}(X)$ such that there is a non-empty open subset $U \subseteq \mathrm{D}(g_1g_2)$ with $f_1(x)/g_1(x) = f_2(x)/g_2(x)$ for all $x \in U$. Then $f_1/g_1 = f_2/g_2$ in $\mathrm{K}(X)$.*

Proof. — From proposition VIII.1.2.2, we know that U is dense in X. Since $\mathrm{V}(f_1g_2 - f_2g_1)$ contains U, it must equal X and hence, $f_1g_2 - f_2g_1 = 0$ in $\Gamma(X)$. ∎

***Construction* VIII.1.4.6.** — Let X be an irreducible affine algebraic set. In order to construct the space $(X, \mathcal{O}_X)$ with functions, we define for each non-empty open subset $U \subseteq X$ the set

$$\mathcal{O}_X(U) := \bigcap_{x\in U} \Gamma(X)_{\mathfrak{m}_x} \subseteq \mathrm{K}(X)$$

where $\Gamma(X)_{\mathfrak{m}_x} = \bigcup_{f\in\Gamma(X)\setminus\mathfrak{m}_x} \Gamma(X)_f$ is the localisation of $\Gamma(X)$ at $\mathfrak{m}_x$. If $U = \varnothing$, we define $\mathcal{O}_X(\varnothing) = \{*\}$ as a singleton set.

To see that $f \in \mathcal{O}_X(U)$ indeed defines a function $U \to k$, note that $f \in \Gamma(X)_{\mathfrak{m}_x}$. So, we may write $f = g/h$ with $g, h \in \Gamma(X)$ and $h \notin \mathfrak{m}_x$. This means that $h(x) \neq 0$ and that $f(x) = g(x)/h(x) \in k$ is well-defined. In this way, we obtain a map $\mathcal{O}_X(U) \to \operatorname{map}(U, k)$, which is injective by lemma VIII.1.4.5, i. e., $\mathcal{O}_X(U)$ can be identified with a k-subalgebra of $\operatorname{map}(U, k)$.

For non-empty open subsets $V \subseteq U \subset X$, we have $\mathcal{O}_X(U) \subseteq \mathcal{O}_X(V)$ by definition, and this inclusion is precisely restriction of functions by the above identification. Finally, the gluing of functions is also possible by definition of $\mathcal{O}_X(U)$. Thus, $(X, \mathcal{O}_X)$ is the **space with functions** (over k) associated to X.

***Proposition* VIII.1.4.7.** — *Let $(X, \mathcal{O}_X)$ be a space with functions associated to an irreducible affine algebraic set X. For all $f \in \Gamma(X)$, we have*

$$\mathcal{O}_X(\mathrm{D}(f)) = \Gamma(X)_f$$

(as subsets of $\mathrm{K}(X)$). In particular, $\mathcal{O}_X(X) = \Gamma(X)$ when taking $f = 1$.

Proof. — We obviously have $\Gamma(X)_f \subseteq \mathcal{O}_X(\mathrm{D}(f))$. Conversely, let $g \in \mathcal{O}_X(\mathrm{D}(f))$ and set $\mathfrak{a} := (\Gamma(X) : g) = \{h \in \Gamma(X) \mid hg \in \Gamma(X)\}$. We need to show that $f \in \sqrt{\mathfrak{a}}$, in which case $g = h/f^n$ for some $h \in \mathfrak{a}$ and some $n \geq 0$. Since $\sqrt{\mathfrak{a}} = \mathrm{I}(\mathrm{V}(\mathfrak{a}))$ by Hilbert's Nullstellensatz VIII.1.1.17, we only need to check that $\mathrm{V}(\mathfrak{a}) \subseteq \mathrm{V}(f)$. Let $x \in \mathrm{D}(f)$. As $g \in \mathcal{O}_X(\mathrm{D}(f)) = X \setminus \mathrm{V}(f)$, we can write $g = g_1/g_2$ with $g_1, g_2 \in \Gamma(X)$ and $g_2 \notin \mathfrak{m}_x$. Of course, $g_2 \in \mathfrak{a}$ and $g_2(x) \neq 0$, so necessarily $x \notin \mathrm{V}(\mathfrak{a})$. ∎

***Warning* VIII.1.4.8.** — For $f \in \mathcal{O}_X(U)$ with $U \subseteq X$ open, we cannot expect f to be *globally* of the form g/h with $g, h \in \Gamma(X)$ and $h(x) \neq 0$ for all $x \in U$. Such a representation is only guaranteed *locally* on U: for any other point $x' \in U \cap \mathrm{D}(h)$, we have $h \notin \mathfrak{m}_{x'}$, so any other representation $f = g'/h'$ at y is equal to g/h in $\mathrm{K}(X)$ and hence in $\Gamma(X)_{\mathfrak{m}_{x'}}$.

[GW20] gives Ex. 5.36.

However, if $\Gamma(X)$ is a unique factorisation domain, e. g., $X = \mathbb{A}^n(k)$, then the global representation is possible. For any two points $x_1, x_2 \in U$, suppose that we have representations f_1/g_1 and f_2/g_2 of f, resp. They are equal to f in $\mathrm{K}(X)$, so $f_1 g_2 = f_2 g_1$ in $\Gamma(X)$. Assuming that f_1/g_1 and f_2/g_2 are completely reduced, we can deduce that $f_1 = f_2$ and $g_1 = g_2$.

***Remark* VIII.1.4.9.** — We could have defined $(X, \mathcal{O}_X)$ also in a different way by hypothesising

$$\mathcal{O}_X(\mathrm{D}(f)) = \Gamma(X)_f$$

for all $f \in \Gamma(X)$ proposition VIII.1.4.7 and recovering construction VIII.1.4.6 with the gluing axiom. This is possible because the principal open subsets $\mathrm{D}(f)$ form a basis of the Zariski topology (lemma VIII.1.3.10). We would be only left to show that such a space indeed exists, i. e., that it is well-defined ($\mathrm{D}(f) = \mathrm{D}(g)$ implies $\Gamma(X)_f = \Gamma(X)_g$) and that gluing is possible. This would be similar to what we have argued in construction VIII.1.4.6 and proposition VIII.1.4.7. This approach will be used to define affine schemes.

***Remark* VIII.1.4.10.** — Given an integral finite type k-algebra A, we can recover a space with functions $(X, \mathcal{O}_X)$ of the corresponding irreducible affine algebraic set X (note that X is unique up to isomorphism by theorem VIII.1.3.12). To do this, set $X := \operatorname{MaxSpec} A$ with closed sets

$$\mathrm{V}(\mathfrak{a}) := \{\mathfrak{m} \in X \mid \mathfrak{m} \supseteq \mathfrak{a}\}$$

where $\mathfrak{a}$ runs through all ideals of A. For open subsets $U \subseteq X$, we define

$$\mathcal{O}_X(U) := \bigcap_{\mathfrak{m} \in U} A_{\mathfrak{m}} \subseteq \operatorname{Quot} A.$$

This approach will be used to define schemes.

***Proposition* VIII.1.4.11.** — *Let X and Y be irreducible affine algebraic sets, and let $f\colon X \to Y$ be a map of sets. Then the following are equivalent:*

(i) *The map f is a morphism of affine algebraic sets.*
(ii) *If $g \in \Gamma(Y)$, then $g \circ f \in \Gamma(X)$.*
(iii) *The map f is a morphism of spaces with functions, i. e., f is continuous and for $U \subseteq Y$ open and $g \in \mathcal{O}_Y(U)$, we have $g \circ f|_{f^{-1}(U)} \in \mathcal{O}_X(f^{-1}(U))$.*

Proof. — We know from theorem VIII.1.3.12 that (i) and (ii) are equivalent. Moreover, (iii) implies (ii) by proposition VIII.1.4.7 if we take $U = Y$.

We need to show that (ii) implies (iii). Let $f^*\colon \Gamma(Y) \to \Gamma(X)$, $h \mapsto h \circ f$, which is well-defined by assumption. For $g \in \Gamma(Y)$, we have

$$f^{-1}(\mathrm{D}(g)) = \{x \in X \mid g(f(x)) \neq 0\} = \mathrm{D}(f^* g).$$

Since the principal open subsets form a topology basis (lemma VIII.1.3.10), this shows that f is continuous. Next, f^* induces $\Gamma(Y)_g = \mathcal{O}_Y(\mathrm{D}(g)) \to \Gamma(X)_{f^* g} = \mathcal{O}_X(\mathrm{D}(f^* g))$, $h \mapsto h \circ f|_{\mathrm{D}(f^* g)}$ by proposition VIII.1.4.7. This shows the second claim for principal open subsets, and by gluing, the claim for all open subsets $U \subseteq Y$. ◻

***Corollary* VIII.1.4.12.** — *Construction VIII.1.4.6 defines a fully faithful functor*

$$\{\textit{irreducible affine algebraic sets}\} \longrightarrow \{\textit{spaces with functions over } k\}.$$

Of course, not all spaces with functions can be obtained from irreducible affine algebraic sets, but we can obtain a larger class of them through gluing finitely many affine algebraic sets. This is similar to the definition of differentiable manifolds which can be glued from open subsets of $\mathbb{R}^n$ endowed with their differentiable structure.

***Definition* VIII.1.4.13.** — *A space with functions $(X, \mathcal{O}_X)$ is called* **connected** *if the underlying topological space is connected.*

(i) *An* ***affine variety*** *is a space with functions $(X, \mathcal{O}_X)$ that is isomorphic to a space with functions associated to an irreducible affine algebraic set. We will often write $\Gamma(X) \coloneqq \mathcal{O}_X(X)$.*

(ii) *A* ***prevariety*** *is a connected space with functions $(X, \mathcal{O}_X)$ such that there exists a finite open covering $X = \bigcup_{i=1}^n U_i$ where $(U_i, \mathcal{O}_X|_{U_i})$ are affine varieties for all $i = 1, \dots, n$.*

(iii) *A* ***morphism of prevarieties*** *is a morphism of spaces with functions.*

Recall the convention that the empty space with functions is not *a prevariety. This defines the category of prevarieties.*

***Remark* VIII.1.4.14.** — The reason why we call affine varieties not 'affine prevarieties' is that we will later define varieties as *separated* prevarieties, and that we will see that affine varieties are always separated.(VIII.2)

[GW20] mentions Ch. 9.

***Remark* VIII.1.4.15.** — That affine varieties are *irreducible* and that prevarieties are glued from a *finite* number of affine varieties are for purely technical reasons. With some more effort, we could drop both properties to obtain more general affine varieties and prevarieties.

***Corollary* VIII.1.4.16.** — *We have the following equivalences of categories:*

$$\{\textit{integral finite type } k\textit{-algebras}\}^{\mathrm{op}} \longleftrightarrow \{\textit{irreducible affine algebraic sets}\} \longleftrightarrow \{\textit{affine varieties}\}.$$

Proof. — Combine theorem VIII.1.3.12 and corollary VIII.1.4.12. ∎

***Proposition* VIII.1.4.17.** — *Let $(X, \mathcal{O}_X)$ be a prevariety. Then the topological space X is noetherian (in particular quasi-compact) and irreducible.*

Proof. — The first part follows from lemmas VIII.1.2.13 and VIII.1.2.14 and proposition VIII.1.2.15, the second one from lemma VIII.1.2.7. ∎

***Remark* VIII.1.4.18.** — The standard definition of a differentiable manifold uses charts with differentiable transition maps. This is in our case, however, impossible as open subsets of affine algebraic sets are generally not affine algebraic sets as well. Instead, our definition via spaces of functions is equivalent to the definition in differentiable geometry if applied to differentiable manifolds.

Given a differentiable manifold X, we define $\mathcal{O}_X$ as $\mathcal{O}_X(U) \coloneqq C^\infty(U)$ for all $U \subseteq X$ open. Then $X \mapsto (X, \mathcal{O}_X)$ is a fully faithful functor

$$\{\text{differentiable manifolds}\} \hookrightarrow \{\text{spaces with functions over } \mathbb{R}\}.$$

The essential image are those spaces with functions whose underlying topological space is Hausdorff and that have open coverings of spaces with functions $(U, \mathcal{O}_{\mathbb{R}^n}|_U)$ for $U \subseteq \mathbb{R}^n$ open. Thus, the category of differentiable manifolds is equivalent to this essential image, and we may take it as the definition. One can proceed similarly with complex manifolds with holomorphic transition maps.

***Lemma* VIII.1.4.19.** — *Let X be an affine variety, and consider the corresponding principal open subset $\mathrm{D}(f) \subseteq X$ of some $f \in \Gamma(X) = \mathcal{O}_X(X)$. Let Y be the affine variety corresponding to the integral finite type k-algebra $\Gamma(X)_f$. Then $(\mathrm{D}(f), \mathcal{O}_X|_{\mathrm{D}(f)})$ and $(Y, \mathcal{O}_Y)$ are isomorphic spaces with functions. In particular, $(\mathrm{D}(f), \mathcal{O}_X|_{\mathrm{D}(f)})$ is an affine variety.*

(VIII.2) This terminology was also Grothendieck's old distinction between preschemes and schemes, which are now schemes and separated schemes, resp.

Proof. — W. l. o. g., let $X \subseteq \mathbb{A}^n(k)$, and let $\mathfrak{a} := \mathrm{I}(X) \subseteq k[T_1, \dots, T_n]$. We naturally consider $k[T_1, \dots, T_n] \subseteq k[T_1, \dots, T_{n+1}]$ and define $\mathfrak{a}' := \mathfrak{a} + (fT_{n+1} - 1)$. Then we have $\Gamma(Y) = \Gamma(X)_f \cong k[T_1, \dots, T_{n+1}]/\mathfrak{a}'$ and we can identify Y with $\mathrm{V}(\mathfrak{a}') \subseteq \mathbb{A}^{n+1}(k)$.

The projection $\mathbb{A}^{n+1}(k) \twoheadrightarrow \mathbb{A}^n(k)$ by forgetting the last coordinate induces a bijection

$$\phi\colon Y = \{(x, x_{n+1}) \in X \times \mathbb{A}^1(k) \mid x_{n+1}f(x) = 1\} \longrightarrow \mathrm{D}(f) = \{x \in X \mid f(x) \neq 0\}.$$

We want to show that $\phi\colon (Y, \mathcal{O}_Y) \to (\mathrm{D}(f), \mathcal{O}_X|_{\mathrm{D}(f)})$ is an isomorphism of spaces of functions. It is continuous since the projection was continuous. It is open, too, because for all $g/f^r \in \Gamma(Y)$ with $g \in \Gamma(X)$ we have $\phi(\mathrm{D}(g/f^r)) = \phi(\mathrm{D}(gf)) = \mathrm{D}(gf)$. Thus, ϕ is a homeomorphism.

It remains to see that for all $U \subseteq \mathrm{D}(f)$, the map $\phi^*\colon \mathcal{O}_X(U) \to \mathcal{O}_Y(\phi^{-1}(U))$, $s \mapsto s \circ \phi$ is an isomorphism. Note that the principal open subsets of $\mathrm{D}(f)$ are $\mathrm{D}(g) \cap \mathrm{D}(f) = \mathrm{D}(fg)$ where $g \in \Gamma(X)$, and that $\phi^{-1}(\mathrm{D}(fg)) = \{(x, x_{n+1}) \in Y \mid g(x) \neq 0\}$. So, by proposition VIII.1.4.7, we have $\mathcal{O}_X(\mathrm{D}(fg)) = \Gamma(X)_{fg} = \Gamma(Y)_g = \mathcal{O}_Y(\phi^{-1}(\mathrm{D}(fg)))$, and the middle equality is precisely ϕ^*. ◻

***Proposition* VIII.1.4.20.** — *Let X be a prevariety and let $U \subseteq X$ be a non-empty open subset. Then $(U, \mathcal{O}_X|_U)$ is an (open) prevariety and the inclusion $U \hookrightarrow X$ is a morphism of prevarieties.*

Proof. — Since X is irreducible, U is connected (propositions VIII.1.4.17 and VIII.1.2.2). Next, lemma VIII.1.4.19 says that $(\mathrm{D}(f), \mathcal{O}_X|_{\mathrm{D}(f)})$ for $f \in \Gamma(X)$ are open affine subvarieties of X. Since X is a union of affine varieties and due to lemma VIII.1.3.10, we can cover U by open affine subvarieties. Lastly, since X is noetherian, U is quasi-compact (proposition VIII.1.4.17 and lemma VIII.1.2.14), so a finite subcover exists. ◻

***Warning* VIII.1.4.21.** — Open subprevarieties of affine varieties are generally not affine, e. g., $U := \mathbb{A}^n(k) \setminus \{0\} \subseteq \mathbb{A}^n(k)$ for $n \geq 2$. To see this, suppose that U is affine and consider the morphism $i\colon U \hookrightarrow \mathbb{A}^n(k)$. On affine coordinate rings, we obtain

$$i^*\colon k[T_1, \dots, T_n] \longrightarrow \Gamma(U) = \mathcal{O}_{\mathbb{A}^n(k)}(U) = \bigcap_{x \in U} k[T_1, \dots, T_n]_{\mathfrak{m}_x} \subseteq k(T_1, \dots, T_n).$$

For any $f = g/h \in \Gamma(U)$ with $g, h \in k[T_1, \dots, T_n]$ and $h \neq 0$, we have $h(x) \neq 0$ for all $x \in U$. This is only possible, however, if $h \in k$ since $n \geq 2$ (the polynomial $h(1, \dots, 1, T_i, 1, \dots, 1)$ in one variable T_i must not have a root, so T_i cannot appear in h). It follows that $\Gamma(U) = k[T_1, \dots, T_n]$, so i^* is the identity. By corollary VIII.1.4.16, it follows that i is an isomorphism of affine varieties, contradicting that i is not bijective on the underlying topological spaces.(VIII.3)

***Construction* VIII.1.4.22.** — Given a prevariety X, we want to compare the function fields of any two non-empty open affine subvarieties $U, V \subseteq X$, e. g., those in the finite open covering of X of affine varieties. Since X is connected, we have $U \cap V \neq \varnothing$. By definition of $\mathcal{O}_X|_U$, we have $\Gamma(U) = \mathcal{O}_X(U) \subseteq \mathcal{O}_X(U \cap V) \subseteq \mathrm{K}(U)$, whence $\operatorname{Quot} \mathcal{O}_X(U \cap V) = \mathrm{K}(U)$. A similar argument shows that $\mathrm{K}(U) = \mathrm{K}(V)$. Thus, we define the **function field** of X as $\mathrm{K}(X) := \mathrm{K}(U) = \mathrm{K}(V)$.

***Warning* VIII.1.4.23.** — Let $f\colon X \to Y$ be a morphism of affine varieties. Then there is generally no induced map of function fields $\mathrm{K}(Y) \to \mathrm{K}(X)$ because the map $f^*\colon \Gamma(Y) \to \Gamma(X)$ is generally not injective. In other words, $\mathrm{K}(-)$ is not functorial.

However, if $f\colon X \to Y$ is a morphism of prevarieties whose image contains a non-empty open (hence dense by propositions VIII.1.4.17 and VIII.1.2.2) subset of Y, then it will induce a map $\mathrm{K}(Y) \to \mathrm{K}(X)$. This will follow from a theorem we will prove later (keyword: *dominant maps*).

[GW20] mentions Thm. 10.19 and Exer. 10.1.

***Proposition* VIII.1.4.24.** — *Let X be prevariety and let $U \subseteq X$ be a non-empty open subset. Then $\mathcal{O}_X(U)$ is a k-subalgebra of $\mathrm{K}(X)$. If $U' \subseteq U$ is another subset, then the restriction $\mathcal{O}_X(U) \to \mathcal{O}_X(U')$ is the inclusion of subalgebras in $\mathrm{K}(X)$. If $U, V \subseteq X$ are arbitrary non-empty open subsets, then $\mathcal{O}_X(U \cup V) = \mathcal{O}_X(U) \cap \mathcal{O}_X(V)$.*

(VIII.3) In fact, $\mathbb{A}^1(k) \setminus \{0\} \subseteq \mathbb{A}^1(k)$ is isomorphic to $\mathrm{V}(XY - 1) \subseteq \mathbb{A}^2(k)$, hence an open affine variety.

Proof. — It is clear by definition that $\mathcal{O}_X(U)$ is a k-subalgebra. Consider any $f \in \mathcal{O}_X(U) \subseteq \operatorname{Hom}(U, \mathbb{A}^1(k))$ (note that the k-algebra structure is defined pointwise, in particular on each open affine subvarieties of U). Then its vanishing set $f^{-1}(0) \subseteq U$ is closed because f is continuous and $\{0\} \subseteq \mathbb{A}^1(k)$ is closed. Therefore, if $f|_{U'} = 0$, that is, $U' \subseteq f^{-1}(0)$, then $f = 0$ since U' is dense in U (propositions VIII.1.4.17 and VIII.1.2.2). Thus, the restriction map $f \mapsto f|_{U'}$ is injective.

For the last statement, the above shows that $\mathcal{O}_X(U \cup V) \subseteq \mathcal{O}_X(U) \cap \mathcal{O}_X(V)$ for $U, V \subseteq X$ non-empty open. On the other hand, $\mathcal{O}_X(U) \cap \mathcal{O}_X(V) \subseteq \mathcal{O}_X(U \cup V)$, where we consider $\mathcal{O}_X(U)$ and $\mathcal{O}_X(V)$ as subsets of $\mathcal{O}_X(U \cap V)$, by gluing. ◻

***Construction* VIII.1.4.25.** — Let X be a prevariety, and let $Z \subseteq X$ be an irreducible closed subset. We wish to define a prevariety structure on Z: for each $U \subseteq Z$ open, define

$$\mathcal{O}'_Z(U) := \left\{ f \in \operatorname{map}(U, k) \;\middle|\; \begin{array}{l} \text{for each } x \in U, f|_{U \cap V} = g|_{U \cap V} \text{ for some } g \in \mathcal{O}_X(V) \\ \text{where } V \subseteq X \text{ is an open neighbourhood of } x \end{array} \right\}.$$

In other words, $\mathcal{O}'_Z(U)$ consists of those functions which, at each point in U, locally agree to some function of $\mathcal{O}_X$. Hence, $(Z, \mathcal{O}'_Z)$ is a space with functions.

***Lemma* VIII.1.4.26.** — *Let X be an affine variety and let $Z \subseteq X$ be an irreducible closed subset. Then the affine varieties $(Z, \mathcal{O}_Z)$ and $(Z, \mathcal{O}'_Z)$ coincide.*

Proof. — We consider $(Z, \mathcal{O}_Z)$ as an affine subvariety of $(X, \mathcal{O}_X)$ by virtue of the morphism $Z \hookrightarrow X$. Let $f \in \mathcal{O}'_Z(U)$. For each $x \in U$, we have $f|_{U \cap V} = g|_{U \cap V}$ for some $g \in \mathcal{O}_X(V)$ where V is an open neighbourhood of x. Since $g \in \mathcal{O}_X(V) \subseteq \Gamma(X)_{\mathfrak{m}_x}$, we also have $f \in \Gamma(X)_{\mathfrak{m}_x}$. Hence, $f \in \bigcap_{x \in U} \Gamma(X)_{\mathfrak{m}_x} = \mathcal{O}_Z(U)$.

Conversely, let $f \in \mathcal{O}_Z(U)$. For each $x \in U$ there is some $h \in \Gamma(Z)$ such that $\mathrm{D}(h) \subseteq U$ is an open neighbourhood of x. Then $f|_{\mathrm{D}(h)} \in \mathcal{O}_Z(\mathrm{D}(h)) = \Gamma(Z)_h$ has the form g/h^n for some $n \geq 0$ and $g \in \Gamma(Z)$. Lifting g and h to $\tilde{g}, \tilde{h} \in \Gamma(X)$ and setting $V := \mathrm{D}(\tilde{h}) \subseteq X$, which is an open neighbourhood of x, we obtain $\tilde{g}/\tilde{h}^n \in \mathcal{O}_X(V)$ with $f|_{U \cap V} = f|_{\mathrm{D}(h)} = g/h^n = (\tilde{g}/\tilde{h}^n)|_{U \cap V}$, that is, $f \in \mathcal{O}'_Z(U)$. ◻

***Proposition* VIII.1.4.27.** — *Let X be a prevariety and let $Z \subseteq X$ be an irreducible closed subset. Then $(Z, \mathcal{O}'_Z)$ is a prevariety and the inclusion $Z \hookrightarrow X$ is a morphism of prevarieties. From now on, we will write always write $(Z, \mathcal{O}_Z)$ instead of $(Z, \mathcal{O}'_Z)$.*

Proof. — Let $X = \bigcup_{i=1}^n U_i$ be an open covering of affine varieties. Then $Z \cap U_i$ is an irreducible closed subset of U_i (cf. observation VIII.1.2.5), so an affine variety by lemma VIII.1.4.26. Thus, $Z = \bigcup_{i=1}^n (Z \cap U_i)$ is a prevariety. ◻

Insert subsection on projective varieties, [GW20, p. 26].

§ VIII.2. SHEAVES

VIII.2.0.1. — Recall that a space with functions $(X, \mathcal{O}_X)$ over a field K equips a topological space X with a system $\mathcal{O}_X$ of functions $U \to K$ where $U \subseteq X$ is an open subset. In order to define schemes later, it is paramount to generalise the notion of systems of functions since we will not deal with functions proper. As it turns out, the essential operations on functions are restricting and gluing, rather than evaluating. This leads to the abstract notion of a *sheaf* on X, although it is still instructive to think of sections of this sheaf as functions on open subsets of X. Pairing a sheaf of rings on X with the topological space X, we obtain the notion of *ringed spaces* replacing spaces with functions.

VIII.2.1. Presheaves and sheaves.

Definition* VIII.2.1.1.** — *Let X be a topological space. A* ***presheaf $\mathcal{F}$ *on X consists of the data*

(i) *a set $\mathcal{F}(U)$ for each open subset $U \subseteq X$;*
(ii) *a* ***restriction map*** $\operatorname{res}^V_U\colon \mathcal{F}(V) \to \mathcal{F}(U)$ *for each pair of open subsets $U \subseteq V \subseteq X$;*

such that the following condition hold:

(i) $\operatorname{res}^U_U = \operatorname{id}_{\mathcal{F}(U)}$ *for every open subset U;*
(ii) $\operatorname{res}^W_U = \operatorname{res}^V_U \circ \operatorname{res}^W_V$ *for every open subsets $U \subseteq V \subseteq W$.*

Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be presheaves on X. A ***morphism of presheaves*** $\phi\colon \mathcal{F}_1 \to \mathcal{F}_2$ *is a family of maps $\phi_U\colon \mathcal{F}_1(U) \to \mathcal{F}_2(U)$ for all $U \subseteq X$ open, such that for all open subsets $U \subseteq V$ the following diagram commutes:*

$$\begin{tikzcd}
\mathcal{F}_1(V) \arrow[r, "\phi_V"] \arrow[d, "\operatorname{res}^V_U"'] & \mathcal{F}_2(V) \arrow[d, "\operatorname{res}^V_U"] \\
\mathcal{F}_1(U) \arrow[r, "\phi_U"] & \mathcal{F}_2(U).
\end{tikzcd}$$

We obtain the ***category of presheaves*** *on X, denoted $\mathsf{PSh}(X)$.*

***Notation* VIII.2.1.2.** — Let $\mathcal{F}$ be a presheaf on X. If $U \subseteq V$ are open subsets, then we write $s|_U \coloneqq \operatorname{res}^V_U(s)$ for $s \in \mathcal{F}(V)$. We call elements $s \in \mathcal{F}(U)$ **sections** of $\mathcal{F}$ over U. Moreover, we also write $\Gamma(U, \mathcal{F}) \coloneqq \mathcal{F}(U)$ for the local sections over U.

***Notation* VIII.2.1.3.** — Let X be a topological space. We define the category of open subsets Ouv_X of X as follows: for open subsets $U, V \subseteq X$, the morphisms are

$$\operatorname{Hom}(U, V) = \begin{cases} \varnothing & \text{if } U \nsubseteq V, \\ \{U \hookrightarrow V\} & \text{if } U \subseteq V. \end{cases}$$

The composition of morphisms is defined by composition of inclusions.

***Remark* VIII.2.1.4.** — Alternatively, a presheaf is the same as a functor $\mathcal{F} : \mathsf{Ouv}_X^{\mathrm{op}} \to \mathsf{Set}$. By replacing Set with arbitrary categories $\mathcal{C}$, we obtain a **presheaf** $\mathcal{F}$ on X with values in $\mathcal{C}$. In this case, $\mathcal{F}(U)$ are objects in $\mathcal{C}$ and res^V_U are morphisms in $\mathcal{C}$. For instance, $\mathcal{C} = \mathsf{AbGrp}, \mathsf{Ring}, R\text{-}\mathsf{Mod}, R\text{-}\mathsf{Alg}$, etc., give *presheaves of abelian groups*, *presheaves of rings*, etc., resp. A **morphism of presheaves** $\mathcal{F}_1 \to \mathcal{F}_2$ with values in $\mathcal{C}$ then is simply a natural transformation between morphisms.

***Definition* VIII.2.1.5.** — *Let $\mathcal{F}$ be a presheaf on a topological space X, and let $U \subseteq X$ be an open subset with an open covering $U = \bigcup_{i \in I} U_i$. Depending on this covering, we define*

$$\begin{aligned}
\rho\colon \mathcal{F}(U) &\longrightarrow \prod_{i \in I} \mathcal{F}(U_i), & s &\longmapsto (s|_{U_i})_i; \\
\sigma\colon \prod_{i \in I} \mathcal{F}(U_i) &\longrightarrow \prod_{i,j \in I} \mathcal{F}(U_i \cap U_j), & (s_i)_i &\longmapsto (s_i|_{U_i \cap U_j})_{i,j}; \\
\sigma'\colon \prod_{i \in I} \mathcal{F}(U_i) &\longrightarrow \prod_{i,j \in I} \mathcal{F}(U_i \cap U_j), & (s_i)_i &\longmapsto (s_j|_{U_i \cap U_j})_{i,j}.
\end{aligned}$$

The presheaf $\mathcal{F}$ is called a ***sheaf*** *if it satisfies the* ***sheaf condition****: for every open covering $U = \bigcup_{i \in I} U_i$ for an open set U, the diagram*

$$\mathcal{F}(U) \xrightarrow{\rho} \prod_{i \in I} \mathcal{F}(U_i) \underset{\sigma'}{\overset{\sigma}{\rightrightarrows}} \prod_{i,j \in I} \mathcal{F}(U_i \cap U_j)$$

is an equaliser.[VIII.4] *This means that ρ is a bijection onto the set of $(s_i)_i \in \prod_{i\in I}\mathcal{F}(U_i)$ such that $\sigma((s_i)_i) = \sigma'((s_i)_i)$.*

In other words, $\mathcal{F}$ is a sheaf if for every open covering $U = \bigcup_{i\in I} U_i$ for an open set U, the following hold:

(i) *Locality axiom: if $s, s' \in \mathcal{F}(U)$ with $s|_{U_i} = s'|_{U_i}$ for all i, then $s = s'$;*

(ii) *Gluing axiom: if $s_i \in \mathcal{F}(U_i)$ for all i with $s_i|_{U_i\cap U_j} = s_j|_{U_i\cap U_j}$ for all i, j, then there is some $s \in \mathcal{F}(U)$ such that $s|_{U_i} = s_i$ (this s is unique by locality).*

Essentially, sheaves are determined by local data and functions can be glued.

***Remark* VIII.2.1.6.** — Let $\mathcal{F}$ be a sheaf on X. Then $\mathcal{F}(\varnothing)$ is a singleton set since the empty covering of the empty set gives the singleton set (terminal object) $\prod_{i\in\varnothing}\mathcal{F}(U_i)$. Hence if X is a single point, then $\mathcal{F}$ is completely determined by $\mathcal{F}(X)$; we will sometimes identify $\mathcal{F}$ with $\mathcal{F}(X)$.

Definition* VIII.2.1.7.** — *A* ***morphism of sheaves *is a morphism of presheaves. We obtain the* ***category of sheaves*** *on a topological space X, denoted $\mathsf{Sh}(X)$. Similarly, we can obtain* sheaf of abelian groups*,* sheaf of rings*, etc.*

***Remark* VIII.2.1.8.** — Alternatively, we can define sheaves with values in an abelian category, e. g., sheaves of abelian groups, differently. A presheaf $\mathcal{F}$ with values in an abelian category is a sheaf if for all open coverings $U = \bigcup_{i\in I} U_i$ of an open subset U, the sequence

$$0 \longrightarrow \mathcal{F}(U) \xrightarrow{\rho} \prod_{i\in I}\mathcal{F}(U_i) \xrightarrow{\sigma} \prod_{i,j\in I}\mathcal{F}(U_i\cap U_j)$$

is exact where

$$\begin{aligned} \rho\colon \mathcal{F}(U) &\longrightarrow \prod_{i\in I}\mathcal{F}(U_i), & s &\longmapsto (s|_{U_i})_i;\\ \sigma\colon \prod_{i\in I}\mathcal{F}(U_i) &\longrightarrow \prod_{i,j\in I}\mathcal{F}(U_i\cap U_j), & (s_i)_i &\longmapsto (s_i|_{U_i\cap U_j} - s_j|_{U_i\cap U_j})_{i,j}. \end{aligned}$$

Definition* VIII.2.1.9.** — *Let $\mathcal{F}$ be a presheaf on a topological space X, and let $U \subseteq X$ be an open subspace. The* ***restriction *of $\mathcal{F}$ to U is the presheaf $\mathcal{F}|_U$ defined by $\mathcal{F}|_U(V) := \mathcal{F}(V)$ for every open subset $V \subseteq U$. If $\mathcal{F}$ is a sheaf, then so is $\mathcal{F}|_U$.*

***Example* VIII.2.1.10.** — (i) Let X and Y be topological spaces. For every open subsets $U \subseteq V \subseteq X$, define $\mathcal{F}(U)$ as the set of continuous maps $U \to Y$ and define res^V_U to be the usual restriction of maps $V \to Y$ to U. Then $\mathcal{F}$ is a sheaf.

(ii) Let $(X, \mathcal{O}_X)$ be a space with functions over a field K. Then $\mathcal{O}_X$ is a sheaf of K-algebras on X. In particular, if X is a real C^r-manifold and if $\mathcal{C}^r_X(U)$ is the set of C^r-functions $U \to \mathbb{R}$, then $\mathcal{C}^r_X$ is a sheaf of $\mathbb{R}$-algebras on X.

(iii) Let X be a topological space. For every open $U \subseteq X$, define $\mathcal{F}(U)$ as the set of bounded continuous maps $U \to \mathbb{R}$. Then $\mathcal{F}$, with the usual restriction, is a presheaf, but generally not a sheaf.

By locality and gluing, it actually suffices to define sheaves on a basis $\mathcal{B}$ of a topology space X.

[VIII.4] This definition works also for $\mathcal{C}$-valued presheaves, assuming that $\mathcal{C}$ admits equalisers.

***Definition* VIII.2.1.11.** — *Let $\mathcal{B}$ be a basis of a topological space X. Considering $\mathcal{B}$ as a full subcategory of Ouv_X, a **presheaf** on $\mathcal{B}$ is a functor $\mathcal{F}\colon \mathcal{B}^{\mathrm{op}} \to \mathsf{Set}$. Every such presheaf can be extended to a presheaf $\mathcal{F}'$ on X by defining*

$$\mathcal{F}'(V) := \left\{ (s_U)_U \in \prod_{\substack{U\in\mathcal{B}\\ U\subseteq V}} \mathcal{F}(U) \;\middle|\; s_U|_{U'} = s_{U'} \text{ for all } U' \subseteq U \text{ in } \mathcal{B} \right\} = \varprojlim_{\substack{U\in\mathcal{B}\\ U\subseteq V}} \mathcal{F}(U)$$

*for all open subsets $V \subseteq X$ (the inverse system is ordered by inclusion with restriction maps). A **morphism of presheaves** on $\mathcal{B}$ is simply a natural transformation of functors.*

To define the sheaf property for presheaves on $\mathcal{B}$, we encounter the problem that $\mathcal{B}$ might not be stable under finite intersections, i. e., for an open covering $U = \bigcup_{i\in I} U_i$ with elements in $\mathcal{B}$, the intersections $U_i \cap U_j$ might not be in $\mathcal{B}$. However, we can simply cover these intersections.

***Proposition* VIII.2.1.12.** — *Let $\mathcal{B}$ be a basis of a topological space X. Let $\mathcal{F}$ be a presheaf on $\mathcal{B}$ with associated presheaf $\mathcal{F}'$ on X. Then $\mathcal{F}'$ is a sheaf iff $\mathcal{F}$ satisfies the following: for every $U \in \mathcal{B}$, for every open covering $U = \bigcup_{i\in I} U_i$ with $U_i \in \mathcal{B}$, and for every open covering $U_i \cap U_j = \bigcup_k U_{ijk}$ with $U_{ijk} \in \mathcal{B}$, the diagram*

$$\mathcal{F}(U) \xrightarrow{\rho} \prod_{i\in I} \mathcal{F}(U_i) \underset{\sigma'}{\overset{\sigma}{\rightrightarrows}} \prod_{i,j,k} \mathcal{F}(U_{ijk})$$

*is an equaliser. Here, σ and σ' are defined as in definition VIII.2.1.5. In this case, $\mathcal{F}$ is called a **sheaf** on $\mathcal{B}$.*

Proof. — Suppose that $\mathcal{F}'$ is a sheaf, and let $U, U_i, U_{ijk} \in \mathcal{B}$ be as in the property. In particular, we have $\mathcal{F}'(U) = \mathcal{F}(U)$ etc. Then the sheaf condition of $\mathcal{F}'$ states that $\mathcal{F}(U)$ is in bijection with elements $(s_i)_i \in \prod_{i\in I} \mathcal{F}(U_i)$ such that $s_i|_{U_i\cap U_j} = s_j|_{U_i\cap U_j}$ for all i, j (as sections in $\mathcal{F}'(U_i \cap U_j)$). Restricting this condition further to U_{ijk}, we obtain $s_i|_{U_{ijk}} = s_j|_{U_{ijk}}$ for all i, j, k, as desired.

Conversely, suppose that the stated property for $\mathcal{F}$ is satisfied. Then $\mathcal{F}$ satisfies the locality axiom, as before, and the following modified gluing axiom: if $s_i \in \mathcal{F}(U_i)$ for all i with $s_i|_{U_{ijk}} = s_j|_{U_{ijk}}$ for all i, j, k where $U_{ijk} \in \mathcal{B}$ cover $U_i \cap U_j$, then there is some $s \in \mathcal{F}(U)$ with $s|_{U_i} = s_i$ for all i. Let $V \subseteq X$ be open with an open covering $V = \bigcup_{i\in I} V_i$. Then $\mathcal{F}'(V)$ consists of elements $(s_U)_U \in \prod_{U\subseteq V} \mathcal{F}(U)$ such that $s_U|_{U'} = s_{U'}$ for all $U' \subseteq U$ (here, $U, U' \in \mathcal{B}$), and we similarly denote elements of $\mathcal{F}'(V_i)$ (resp. $\mathcal{F}'(V_i \cap V_j)$) by $(s_{U_i})_{U_i} \in \prod_{U_i\subseteq V_i} \mathcal{F}(U_i)$ (resp. $(s_{U_{ij}})_{U_{ij}} \in \prod_{U_{ij}\subseteq V_i\cap V_j} \mathcal{F}(U_{ij})$).

We will check locality and gluing. On the one hand, let $(s_U)_U, (s'_U)_U \in \mathcal{F}'(V)$ with $(s_U)_U|_{V_i} = (s'_U)_U|_{V_i}$ for all i. Then $s_U|_{U_i} = s'_U|_{U_i}$ for all i, so $s_U = s'_U$ by locality axiom of $\mathcal{F}$. On the other hand, let $(s_{U_i})_{U_i} \in \mathcal{F}'(V_i)$ with $(s_{U_i})_{U_i}|_{V_i\cap V_j} = (s_{U_j})_{U_j}|_{V_i\cap V_j}$ for all i, j. Then $s_{U_i}|_{U_{ijk}} = s_{U_j}|_{U_{ijk}}$ for all i, j where U_{ijk} iterates through all U_{ij} with $U_{ij} \subseteq U_i \cap U_j$. So, there is some $s_U \in \mathcal{F}(U)$ with $s_U|_{U_i} = s_{U_i}$ for all i by gluing axiom of $\mathcal{F}$. Hence we have $(s_U)_U \in \mathcal{F}'(V)$ with $(s_U)_U|_{V_i} = (s_U|_{U_i})_{U_i} = (s_{U_i})_{U_i}$ for all i. ∎

***Corollary* VIII.2.1.13.** — *The construction $\mathcal{F} \mapsto \mathcal{F}'$ is clearly functorial. We obtain an equivalence of categories*

$$\{\textit{sheaves on } \mathcal{B}\} \longrightarrow \mathsf{Sh}(X).$$

***Remark* VIII.2.1.14.** — Analogous definitions and proposition VIII.2.1.12 also holds for sheaves with values in a category $\mathcal{C}$ admitting inverse limits.

VIII.2.2. Gluing sheaves.

***Proposition* VIII.2.2.1** (gluing of sheaf morphisms)**.** — *Let $X = \bigcup_{i \in I} U_i$ be an open covering of a topological space X. Let $\mathcal{F}$ and $\mathcal{G}$ be two sheaves on X. If $\phi_i \colon \mathcal{F}|_{U_i} \to \mathcal{G}|_{U_i}$ is a family of morphisms of sheaves for all $i \in I$ such that ϕ_i and ϕ_j agree on $U_i \cap U_j$ for all $i, j \in I$, then they glue to a unique morphism $\phi \colon \mathcal{F} \to \mathcal{G}$, i. e., ϕ and ϕ_i agree on U_i for all $i \in I$.*

Proof. — If such a morphism ϕ exists, then for all $U \subseteq X$ open and all $s \in \mathcal{F}(U)$, we must have

$$\phi_U(s)|_{U \cap U_i} = \phi_{i, U \cap U_i}(s|_{U \cap U_i}) \tag{VIII.2.2.1.1}$$

for all i. In fact, this condition uniquely defines ϕ: since

$$\begin{aligned}\phi_{i, U \cap U_i}(s|_{U \cap U_i})|_{U \cap U_i \cap U_j} &= \phi_{i, U \cap U_i \cap U_j}(s|_{U \cap U_i \cap U_j}) \\ &= \phi_{j, U \cap U_i \cap U_j}(s|_{U \cap U_i \cap U_j}) = \phi_{j, U \cap U_j}(s|_{U \cap U_j})|_{U \cap U_i \cap U_j}\end{aligned}$$

for all i, j, the sheaf property of $\mathcal{G}$ gives a unique section which we denote by $\phi_U(s) \in \mathcal{G}(U)$ satisfying equation (VIII.2.2.1.1) for all i. In particular, equation (VIII.2.2.1.1) gives $\phi_U = \phi_{i,U}$ for all $U \subseteq U_i$ open.

It remains to check that ϕ is compatible with restrictions. For all $U \subseteq V \subseteq X$ open and all $s \in \mathcal{F}(V)$, by equation (VIII.2.2.1.1) we have

$$\phi_V(s)|_{U \cap U_i} = \phi_V(s)|_{V \cap U_i}|_{U \cap U_i} = \phi_{i, V \cap U_i}(s|_{V \cap U_i})|_{U \cap U_i} = \phi_{i, U \cap U_i}(s|_{U \cap U_i}) = \phi_U(s|_U)|_{U \cap U_i}$$

for all i. Hence by locality, we obtain $\phi_V(s)|_U = \phi_U(s|_U)$. ◻

***Definition* VIII.2.2.2.** — *Let $X = \bigcup_{i \in I} U_i$ be an open covering of a topological space X. A **gluing datum** of sheaves relative to $(U_i)_{i \in I}$ consists of:*

- *a sheaf $\mathcal{F}_i$ on U_i for each $i \in I$;*
- *an isomorphism $\phi_{ij} \colon \mathcal{F}_i|_{U_i \cap U_j} \xrightarrow{\sim} \mathcal{F}_j|_{U_i \cap U_j}$ for each pair $(i, j) \in I \times I$;*

*such that the **cocycle condition** holds: the diagram*

$$\begin{tikzcd} \mathcal{F}_i|_{U_i \cap U_j \cap U_k} \arrow[rr, "\phi_{ik}"] \arrow[dr, "\phi_{ij}"'] & & \mathcal{F}_k|_{U_i \cap U_j \cap U_k} \\ & \mathcal{F}_j|_{U_i \cap U_j \cap U_k} \arrow[ur, "\phi_{jk}"'] & \end{tikzcd}$$

commutes for all $i, j, k \in I$.

***Proposition* VIII.2.2.3** (gluing of sheaves)**.** — *Let $X = \bigcup_{i \in I} U_i$ be an open covering of a topological space X, and let $(\mathcal{F}_i, \phi_{ij})$ be a gluing datum of sheaves relative to $(U_i)_{i \in I}$. Then there exists a sheaf $\mathcal{F}$ on X and isomorphisms $\psi_i \colon \mathcal{F}_i \xrightarrow{\sim} \mathcal{F}|_{U_i}$ for $i \in I$ such that the diagram*

$$\begin{tikzcd} \mathcal{F}_i|_{U_i \cap U_j} \arrow[rr, "\phi_{ij}"] \arrow[dr, "\psi_i"'] & & \mathcal{F}_j|_{U_i \cap U_j} \arrow[dl, "\psi_j"] \\ & \mathcal{F}|_{U_i \cap U_j} & \end{tikzcd} \tag{VIII.2.2.3.1}$$

*commutes for all $i, j \in I$. We say that $\mathcal{F}$ is obtained by **gluing the** $\mathcal{F}_i$ via the ϕ_{ij}.*

Proof. — To simplify notation, we will drop the subscript open set from a morphism of sheaves if it is clear from context. Define the presheaf $\mathcal{F}$ by

$$\mathcal{F}(U) := \left\{ (s_i)_{i\in I} \in \prod_{i\in I} \mathcal{F}_i(U \cap U_i) \;\middle|\; \phi_{ij}(s_i|_{U\cap U_i\cap U_j}) = s_j|_{U\cap U_i \cap U_j} \text{ for all } i,j \in I \right\}$$

for all $U \subseteq X$ open, with the restriction maps induced by those of the $\mathcal{F}_i$. We can easily check locality and gluing axioms for $\mathcal{F}$ by restricting to U_i and using the sheaf properties of $\mathcal{F}_i$ to make statements about the entry s_i in $(s_i)_{i\in I} \in \mathcal{F}(U)$. Thus, $\mathcal{F}$ is a sheaf.

Next, define ψ_i by

$$\psi_{i,U} \colon \mathcal{F}_i(U) \longrightarrow \mathcal{F}(U), \qquad s \longmapsto (\phi_{ij}(s|_{U\cap U_j}))_{j\in I}$$

for all $U \subseteq U_i$ open. We indeed have $\psi_{i,U}(s) \in \mathcal{F}(U)$ since, by the cocycle condition,

$$\phi_{jk}(\phi_{ij}(s|_{U\cap U_j})|_{U\cap U_j\cap U_k}) = \phi_{jk}(\phi_{ij}(s|_{U\cap U_j \cap U_k})) = \phi_{ik}(s|_{U\cap U_j\cap U_k}) = \phi_{ik}(s|_{U\cap U_k})|_{U\cap U_j \cap U_k}.$$

Moreover, for all $U \subseteq U_i \cap U_j$ open we have, again by the cocycle condition,

$$\psi_{i,U}(s) = (\phi_{ik}(s|_{U\cap U_k}))_{k\in I} = (\phi_{jk}(\phi_{ij}(s|_{U\cap U_k})))_{k\in I} = (\phi_{jk}(\phi_{ij}(s)|_{U\cap U_k}))_{k\in I} = \psi_{j,U}(\phi_{ij}(s)),$$

which is the commutativity of diagram (VIII.2.2.3.1). To show that $\psi_{i,U}$ is bijective for $U \subseteq U_i$ open, we construct the inverse map. If $(s_j)_{j\in I} \in \mathcal{F}(U)$, then we glue the sections $\phi_{ji}(s_j) \in \mathcal{F}_i(U \cap U_j)$ for all j to a unique section $s \in \mathcal{F}_i(U)$. This is possible since, by definition of $\mathcal{F}(U)$,

$$\begin{aligned} \phi_{ji}(s_j)|_{U\cap U_j \cap U_k} = \phi_{ji}(s_j|_{U\cap U_j})|_{U\cap U_j\cap U_k} &= s_i|_{U\cap U_j}|_{U\cap U_j \cap U_k} \\ &= s_i|_{U\cap U_k}|_{U\cap U_j\cap U_k} = \phi_{ki}(s_k|_{U\cap U_k})|_{U\cap U_j\cap U_k} = \phi_{ki}(s_k)|_{U\cap U_j \cap U_k}. \end{aligned}$$

Now, the inverse map is given by $(s_j)_{j\in I} \mapsto s$. ◻

We obtain a universal property of gluing sheaves.

***Corollary* VIII.2.2.4** (universal property of gluing sheaves)**.** — *Let $X = \bigcup_{i\in I} U_i$ be an open covering of a topological space X, and let $(\mathcal{F}_i, \phi_{ij})$ be a gluing datum of sheaves relative to $(U_i)_{i\in I}$. Let $\mathcal{F}$ be a sheaf with isomorphisms $\psi_i \colon \mathcal{F}_i \xrightarrow{\sim} \mathcal{F}|_{U_i}$ obtained by gluing the $\mathcal{F}_i$ via ϕ_{ij}. If $\mathcal{G}$ is a sheaf on X and if $\xi_i \colon \mathcal{F}_i \to \mathcal{G}|_{U_i}$ for all $i \in I$ are morphisms, such that $\xi_j \circ \phi_{ij} = \xi_i$ on U_i for all $i, j \in I$, then there exists a unique morphism $\xi \colon \mathcal{F} \to \mathcal{G}$ with $\xi \circ \psi_i = \xi_i$ for all $i \in I$. In particular, $\mathcal{F}$ is unique up to unique isomorphism.*

In other words: a morphism $(\mathcal{F}_i, \phi_{ij}) \to (\mathcal{G}_i, \psi_{ij})$ of gluing data is defined as a family of morphisms $\mathcal{F}_i \to \mathcal{G}_i$ of sheaves for $i \in I$ such that the diagram

$$\begin{tikzcd} \mathcal{F}_i|_{U_i\cap U_j} \arrow[r, "\phi_{ij}"] \arrow[d] & \mathcal{F}_j|_{U_i\cap U_j} \arrow[d] \\ \mathcal{G}_i|_{U_i\cap U_j} \arrow[r, "\psi_{ij}"] & \mathcal{G}_j|_{U_i \cap U_j} \end{tikzcd}$$

commutes for all $i, j \in I$. Then there is an equivalence of categories

$$\begin{gathered} \mathsf{Sh}(X) \longrightarrow \{\textit{gluing data of sheaves relative to } (U_i)_{i\in I}\}, \\ \mathcal{F} \longmapsto (\mathcal{F}|_{U_i}, \mathcal{F}|_{U_i}|_{U_i\cap U_j} \xrightarrow{\text{id}} \mathcal{F}|_{U_j}|_{U_i\cap U_j}). \end{gathered}$$

Proof. — The universal property follows from proposition VIII.2.2.1 by gluing the morphisms $\xi \circ \psi_i^{-1}$, giving ξ. The functor is fully faithful by proposition VIII.2.2.1 and essentially surjective by proposition VIII.2.2.3. ◻

VIII.2.3. Stalks and sheafification.

***Definition* VIII.2.3.1.** — *Let $\mathcal{F}$ be a presheaf on a topological space X, and let $x \in X$ be a point. Then*

$$\mathcal{F}_x := \varinjlim_{U \ni x} \mathcal{F}(U),$$

where U runs through all open subsets of X, is called the ***stalk*** *of $\mathcal{F}$ in x (the inverse system is ordered by reverse inclusion with restriction maps). Explicitly, $\mathcal{F}_x$ consists of equivalence classes of pairs (U, s) where $U \subseteq X$ is an open neighbourhood of x and where $s \in \mathcal{F}(U)$. Here, (U_1, s_1) and (U_2, s_2) are equivalent if there exists an open neighbourhood $V \subseteq U_1 \cap U_2$ of x such that $s_1|_V = s_2|_V$.*

For each open neighbourhood U of x, there is a canonical map

$$\mathcal{F}(U) \longrightarrow \mathcal{F}_x, \qquad s \longmapsto s_x = [(U, s)].$$

The element s_x is called the ***germ*** *of s in x.*

***Observation* VIII.2.3.2.** — A morphism $\phi \colon \mathcal{F} \to \mathcal{G}$ of presheaves on X induces a map

$$\phi_x := \varinjlim_{U \ni x} \phi_U \colon \mathcal{F}_x \longrightarrow \mathcal{G}_x$$

of stalks in x. We obtain a functor $\mathsf{PSh}(X) \to \mathsf{Set}$, $\mathcal{F} \mapsto \mathcal{F}_x$. In other words, for every open subset $U \subseteq X$ and every point $x \in U$ we have $\phi_U(s)_x = \phi_x(s_x)$ for all $s \in \mathcal{F}(U)$.

Now, consider presheaves on X with values in a category $\mathcal{C}$. If $\mathcal{C}$ admits direct limits, then the stalk $\mathcal{F}_x$ is an object in $\mathcal{C}$, and we obtain a functor $\mathcal{F} \mapsto \mathcal{F}_x$ from the category of presheaves on X with values in $\mathcal{C}$ to the category $\mathcal{C}$.

***Example* VIII.2.3.3.** — Let $X = \mathbb{C}$ and let $\mathcal{O}_{\mathbb{C}}$ be the sheaf of holomorphic functions on $\mathbb{C}$, i. e., $\mathcal{O}_{\mathbb{C}}(U)$ is the set of holomorphic functions $U \to \mathbb{C}$. This sheaf is a sheaf of $\mathbb{C}$-algebras. Fix $z_0 \in \mathbb{C}$. Two holomorphic functions $f_1 \in \mathcal{O}_{\mathbb{C}}(U_1)$ and $f_2 \in \mathcal{O}_{\mathbb{C}}(U_2)$ defined on an open neighbourhood of z_0 agree iff they have the same analytic (Taylor) expansion at z_0. Hence, $\mathcal{O}_{\mathbb{C},z_0}$ is the set of power series $\sum_{n \geq 0} a_n (z - z_0)^n$ with positive radius of convergence, and the identity theorem states that only for connected open neighbourhoods U of z_0 we have a natural inclusion $\mathcal{O}_{\mathbb{C}}(U) \hookrightarrow \mathcal{O}_{\mathbb{C},z_0}$.

Definition* VIII.2.3.4.** — *A morphism $\phi \colon \mathcal{F} \to \mathcal{G}$ of sheaves is called* ***injective *(resp.* ***surjective*** *resp.* ***bijective****) if $\phi_x \colon \mathcal{F}_x \to \mathcal{G}_x$ is injective (resp. surjective resp. bijective) for all $x \in X$.*[VIII.5]

***Proposition* VIII.2.3.5.** — *Let $\phi, \psi \colon \mathcal{F} \to \mathcal{G}$ be morphisms of sheaves on X. Then we have the following:*

(i) *ϕ is injective iff $\phi_U \colon \mathcal{F}(U) \to \mathcal{G}(U)$ is injective for all $U \subseteq X$ open. This is also generally true if $\mathcal{G}$ is only a presheaf on X.*
(ii) *ϕ is bijective iff ϕ_U is bijective for all $U \subseteq X$ open.*
(iii) *$\phi = \psi$ iff $\phi_x = \psi_x$ for all $x \in X$.*

Proof. — We claim that for $U \subseteq X$ open the map

$$\mathcal{F}(U) \longrightarrow \prod_{x \in U} \mathcal{F}_x, \qquad s \longmapsto (s_x)_{x \in U}$$

is injective if $\mathcal{F}$ is a sheaf. Indeed, given $s, t \in \mathcal{F}(U)$ with $s_x = t_x$ for all $x \in U$, then there are open neighbourhood $V_x \subseteq U$ of each x such that $s|_{V_x} = t|_{V_x}$. Since $U = \bigcup_{x \in U} V_x$, we have $s = t$ by locality.

[VIII.5] In fact, $\phi \colon \mathcal{F} \to \mathcal{G}$ is a monomorphism (resp. epimorphism resp. isomorphism) in the categorical sense iff ϕ_x for any $x \in X$ are. This is mainly proven with proposition VIII.2.3.5 (iii).

Since the diagram

$$\begin{tikzcd} \mathcal{F}(U) \arrow[r, hook] \arrow[d, "\phi_U"'] & \prod_{x\in U}\mathcal{F}_x \arrow[d, "\prod_{x\in U}\phi_x"] \\ \mathcal{G}(U) \arrow[r, hook] & \prod_{x\in U}\mathcal{G}_x \end{tikzcd}$$

commutes, we already have (iii) (that $\phi = \psi$ implies $\phi_x = \psi_x$ for all x is trivial) and the 'only if' direction of (i). On the other hand, since the direct limit of injective (resp. surjective) maps is injective (resp. surjective) (this can be checked instantly, also cf. proposition I.2.9.9), the 'if' directions of (i) and (ii) are true, too.

In order to show (ii), it remains to see that if ϕ is bijective, then ϕ_U is surjective. To this end, let $t \in \mathcal{G}(U)$. Since ϕ_x is surjective, there is an open neighbourhood $U^x \subseteq U$ of x and $s^x \in \mathcal{F}(U^x)$ such that $\phi_{U^x}(s^x)_x = \phi_x((s^x)_x) = t_x$. Then there exists an open neighbourhood $V^x \subseteq U^x$ of x such that $\phi_{V^x}(s^x|_{V^x}) = \phi_{U^x}(s^x)|_{V^x} = t|_{V^x}$. In particular, $U = \bigcup_{x\in U} V^x$ and $\phi_{V^x\cap V^y}(s^x|_{V^x\cap V^y}) = t|_{V^x\cap V^y} = \phi_{V^x\cap V^y}(s^y|_{V^x\cap V^y})$ for all $x, y \in U$. Since $\phi_{V^x\cap V^y}$ is injective by (i), we obtain $s^x|_{V^x\cap V^y} = s^y|_{V^x\cap V^y}$. By the gluing axiom, there is some $s \in \mathcal{F}(U)$ such that $s|_{V^x} = s^x|_{V^x}$ for all $x \in U$. Hence, $\phi_U(s)_x = \phi_{U^x}(s^x)_x = t_x$ for all $x \in U$ and thus, $\phi_U(s) = t$ by the above claim.

***Remark* VIII.2.3.6.** — We have also shown in the proof of proposition VIII.2.3.5 that ϕ is surjective iff for every $U \subseteq X$ open and every $t \in \mathcal{G}(U)$ there exists an open covering $U = \bigcup_i U_i$ depending on t and sections $s_i \in \mathcal{F}(U_i)$ such that $\phi_{U_i}(s_i) = t|_{U_i}$, that is, ϕ_U is locally surjective.

***Warning* VIII.2.3.7.** — However, if ϕ is surjective, then we *do not* have that ϕ_U is surjective for all $U \subseteq X$ open (although the converse is true). For example, let $X = \mathbb{C} \setminus \{0\}$ and let $\mathcal{O}_X$ (resp. $\mathcal{O}_X^*$) be the sheaf of holomorphic (resp. non-vanishing holomorphic) functions on X. Then $\phi\colon \mathcal{O}_X \to \mathcal{O}_X^*$, $f \mapsto \exp\circ f$ is certainly surjective on stalks since each non-vanishing holomorphic function has a logarithm *locally* (on any branch), but not globally on X: there is no function $f \in \mathcal{O}_X(X)$ such that $\exp\circ f = \mathrm{id}_X$. Otherwise, $f'(z) = 1/z$ on X, so $0 = \oint_C f'(z)\,\mathrm{d}z = 2\pi\mathrm{i}$ where $C \subseteq X$ is the unit circle.

Another example is $D\colon \mathcal{O}_X \to \mathcal{O}_X$, $f \mapsto f'$ which takes derivatives. This morphism is surjective on stalks since for any $z \in X$, a holomorphic function on a simply connected neighbourhood of z has an anti-derivative (since any contour integral on this neighbourhood vanishes). However, the function $1/z \in \mathcal{O}_X(X)$ has no anti-derivative on X.

Definition* VIII.2.3.8.** — *Let $\mathcal{F}$ and $\mathcal{G}$ be presheaves (resp. sheaves) on X such that $\mathcal{F}(U) \subseteq \mathcal{G}(U)$ for all open subsets $U \subseteq X$, and such that the restriction maps of $\mathcal{F}$ are induced by those of $\mathcal{G}$. Then $\mathcal{F}$ is called a* ***subpresheaf *(resp.* ***subsheaf****) of $\mathcal{G}$.*

***Proposition* VIII.2.3.9** (universal property of sheafification)**.** — *Let $\mathcal{F}$ be a presheaf on a topological space X. Then there exists a sheaf $\tilde{\mathcal{F}}$ on X together with a morphism $\iota_{\mathcal{F}}\colon \mathcal{F} \to \tilde{\mathcal{F}}$ such that: if $\mathcal{G}$ is a sheaf on X and if $\phi\colon \mathcal{F} \to \mathcal{G}$ is a morphism, then there is a unique morphism $\tilde{\phi}\colon \tilde{\mathcal{F}} \to \mathcal{G}$ such that the following diagram commutes:*

$$\begin{tikzcd} \mathcal{F} \arrow[r, "\iota_{\mathcal{F}}"] \arrow[d, "\phi"'] & \tilde{\mathcal{F}} \arrow[dl, dashed, "\tilde{\phi}"] \\ \mathcal{G}. & \end{tikzcd}$$

The sheaf $\tilde{\mathcal{F}}$ is called the ***sheafification*** *of $\mathcal{F}$. Moreover, we have the following:*

(i) *The map* $\iota_{\mathcal{F},x}\colon \mathcal{F}_x \to \tilde{\mathcal{F}}_x$ *on stalks is bijective for all* $x \in X$.

(ii) *For every presheaf* $\mathcal{G}$ *on* X *and every morphism* $\phi\colon \mathcal{F} \to \mathcal{G}$, *there exists a unique morphism* $\tilde{\phi}\colon \tilde{\mathcal{F}} \to \tilde{\mathcal{G}}$ *such that the diagram*

$$\begin{tikzcd} \mathcal{F} \arrow[r, "\iota_{\mathcal{F}}"] \arrow[d, "\phi"'] & \tilde{\mathcal{F}} \arrow[d, dashed, "\tilde{\phi}"] \\ \mathcal{G} \arrow[r, "\iota_{\mathcal{G}}"] & \tilde{\mathcal{G}} \end{tikzcd} \tag{VIII.2.3.9.1}$$

commutes. In particular, $\mathsf{PSh}(X) \to \mathsf{Sh}(X)$, $\mathcal{F} \mapsto \tilde{\mathcal{F}}$ *is a functor.*

Proof. — For $U \subseteq X$ open, we define

$$\tilde{\mathcal{F}}(U) := \left\{ (s_x)_{x\in U} \in \prod_{x\in U} \mathcal{F}_x \;\middle|\; \begin{array}{l} \text{for all } x \in U, \text{ there is an open neighbourhood } W \subseteq U \text{ of } x \\ \text{and some } t \in \mathcal{F}(W) \text{ such that } s_w = t_w \text{ for all } w \in W \end{array} \right\}.$$

Conceptually, the elements in $\tilde{\mathcal{F}}(U)$ are those families $(s_x)_x$ of germs which locally (on W) give rise to a section t of $\mathcal{F}$. The restriction maps $\tilde{\mathcal{F}}(V) \to \tilde{\mathcal{F}}(U)$ where $U \subseteq V \subseteq X$ open are induced by the projection $\prod_{x\in V} \mathcal{F}_x \twoheadrightarrow \prod_{x\in U} \mathcal{F}_x$. This turns $\tilde{\mathcal{F}}$ into a presheaf.

To show that $\tilde{\mathcal{F}}$ is a sheaf, suppose that $U = \bigcup_{i\in I} U_i$ is an open covering and that $(s^i_x)_x \in \tilde{\mathcal{F}}(U_i)$ with $s^i_x = s^j_x$ for all $x \in U_i \cap U_j$ and all i, j. Defining $s_x := s^i_x$ for all $x \in U_i$ and all i, we see that $(s_x)_{x\in U}$ is the unique section of $\tilde{\mathcal{F}}$ over U such that $(s_x)_x|_{U_i} = (s^i_x)_x$ for all i.

For $U \subseteq X$ open, we define $\iota_{\mathcal{F},U}\colon \mathcal{F}(U) \to \tilde{\mathcal{F}}(U)$, $s \mapsto (s_x)_{x\in U}$. By definition of $\tilde{\mathcal{F}}$, we have $\tilde{\mathcal{F}}_x = \mathcal{F}_x$ and $\iota_{\mathcal{F},x} = \mathrm{id}$ for all $x \in X$. This shows (i).

To show (ii), let $\mathcal{G}$ be a presheaf and let $\phi\colon \mathcal{F} \to \mathcal{G}$ be a morphism. We define the morphism $\tilde{\phi}$ by $\tilde{\phi}_U\colon \tilde{\mathcal{F}}(U) \to \tilde{\mathcal{G}}(U)$, $(s_x)_x \mapsto (\phi_x(s_x))_x$ for every $U \subseteq X$ open so that diagram (VIII.2.3.9.1) commutes. Passing to stalks, we have $\phi_x = \tilde{\phi}_x$ for all $x \in X$ by (i). Hence, proposition VIII.2.3.5 shows that $\tilde{\phi}$ is unique.

By (i), $\iota_{\mathcal{G}}$ is bijective (on stalks). If we additionally assume that $\mathcal{G}$ is a sheaf in diagram (VIII.2.3.9.1), then $\iota_{\mathcal{G}}$ is an isomorphism by proposition VIII.2.3.5. Thus, $\iota_{\mathcal{G}}^{-1} \circ \tilde{\phi}$ is the desired unique morphism in the universal property. ∎

***Remark* VIII.2.3.10.** — In other words, the universal property give the adjoint pair $\tilde{-} \dashv \mathrm{For}$ where $\mathrm{For}\colon \mathsf{Sh}(X) \hookrightarrow \mathsf{PSh}(X)$ is the forgetful functor. In particular, as always with universal properties, the sheafification of a presheaf is unique up to unique isomorphism.

***Corollary* VIII.2.3.11.** — *Let* $\mathcal{F}$ *be a presheaf and let* $\mathcal{G}$ *be a sheaf, both on* X. *Then* $\mathcal{G} \cong \tilde{\mathcal{F}}$ *iff there is some morphism* $\iota\colon \mathcal{F} \to \mathcal{G}$ *such that* ι_x *is bijective for all* $x \in X$.

Proof. — Suppose that there is a morphism $\iota\colon \mathcal{F} \to \mathcal{G}$ as stated. By the universal property of sheafification, there is a unique morphism $\tilde{\phi}\colon \tilde{\mathcal{F}} \to \mathcal{G}$ of sheaves such that $\tilde{\phi} \circ \iota_{\mathcal{F}} = \iota$. Passing to stalks, we see that $\tilde{\phi}_x$ is bijective for all $x \in X$, so $\tilde{\phi}$ is an isomorphism by proposition VIII.2.3.5. The converse follows from proposition VIII.2.3.5, too. ∎

Definition* VIII.2.3.12.** — *Let* E *be a set and let* $\mathcal{F}$ *be the presheaf on* X *defined by* $\mathcal{F}(U) := E$ *for any open subset* $U \subseteq X$ *(the restriction maps being the identity). Then the sheafification* $\underline{E} := \underline{E}_X$ *of* $\mathcal{F}$ *is called the* ***constant sheaf *with value* E. *The set* $\underline{E}(U)$ *consists of locally constant functions* $U \to E$.

***Remark* VIII.2.3.13.** — Proposition VIII.2.3.9 and corollary VIII.2.3.11 also hold for presheaves $\mathcal{F}$ with values in a category $\mathcal{C}$. It is then clear that the sheafification $\tilde{\mathcal{F}}$ will also be a sheaf with values in $\mathcal{C}$.

VIII.2.4. Direct and inverse images.

If $f\colon X \to Y$ is a continuous map of topological space, it is natural to ask how we can transport sheaves from X to Y, or vice versa, along f.

***Construction* VIII.2.4.1.** — Let $f\colon X \to Y$ be a continuous map of topological spaces. If $\mathcal{F}$ is a presheaf on X, then define a presheaf on Y by

$$f_*\mathcal{F}(V) := \mathcal{F}(f^{-1}(V))$$

for all $V \subseteq Y$ open and the restriction maps are given by those of $\mathcal{F}$. We call $f_*\mathcal{F}$ the **direct image** of $\mathcal{F}$ under f.

If $\phi\colon \mathcal{F} \to \mathcal{G}$ is a morphism of presheaves, the maps $(f_*\phi)_V := \phi_{f^{-1}(V)}$ for $V \subseteq Y$ open define a morphism $f_*\phi\colon f_*\mathcal{F} \to f_*\mathcal{G}$. Thus, we obtain a functor $f_*\colon \mathsf{PSh}(X) \to \mathsf{PSh}(Y)$.

***Fact* VIII.2.4.2.** — *Let $f\colon X \to Y$ be a continuous map.*

(i) *If $\mathcal{F}$ is a sheaf on X, then $f_*\mathcal{F}$ is a sheaf on Y, too. Thus, we obtain a functor $f_*\colon \mathsf{Sh}(X) \to \mathsf{Sh}(Y)$.*

(ii) *Let $g\colon Y \to Z$ be a continuous map. Then $g_*(f_*\mathcal{F}) = (g \circ f)_*\mathcal{F}$ which is functorial in $\mathcal{F}$.*

Proof. — (i) follows from the facts that if $U = \bigcup_{i\in I} U_i$ is an open covering, then $f^{-1}(U) = \bigcup_{i\in I} f^{-1}(U_i)$ is an open covering as well, and that $f^{-1}(U_i \cap U_j) = f^{-1}(U_i) \cap f^{-1}(U_j)$. For (ii), use the fact that $f^{-1} \circ g^{-1} = (g \circ f)^{-1}$. ◻

***Fact* VIII.2.4.3.** — *Let $f\colon X \to Y$ be a continuous map, and let $\mathcal{F}$ be a presheaf on X. Then for all $x \in X$, we have*

$$(f_*\mathcal{F})_{f(x)} = \mathcal{F}_x$$

which is functorial in $\mathcal{F}$.

Proof. — By definition,

$$(f_*\mathcal{F})_{f(x)} = \varinjlim_{V \ni f(x)} \mathcal{F}(f^{-1}(V)) = \varinjlim_{U \ni x} \mathcal{F}(U) = \mathcal{F}_x. \qquad ◻$$

***Construction* VIII.2.4.4.** — Let $f\colon X \to Y$ be a continuous map of topological spaces. If $\mathcal{G}$ is a presheaf on Y, then define the presheaf $f^+\mathcal{G}$ on X by

$$f^+\mathcal{G}(U) := \varinjlim_{V \supseteq f(U)} \mathcal{G}(V)$$

for all $U \subseteq X$ open, where V runs through all open subsets $V \subseteq Y$, and the restriction maps are induced by those of $\mathcal{G}$. The idea is that, since we cannot evaluate $\mathcal{G}(f(U))$ as $f(U)$ is not necessarily open, we approximate $f(U)$ by open sets (cf. construction VIII.1.4.25 on closed subprevarieties).

Note that $f^+\mathcal{G}$ is generally not a sheaf even if $\mathcal{G}$ is. Hence, we call the sheafification $f^{-1}\mathcal{G}$ of $f^+\mathcal{G}$ the **inverse image** of $\mathcal{G}$ under f.

Certainly, $f^+\mathcal{G}$ and hence $f^{-1}\mathcal{G}$ is functorial in $\mathcal{G}$, i. e., we obtain a functor $f^{-1}\colon \mathsf{PSh}(Y) \to \mathsf{Sh}(X)$.

***Notation* VIII.2.4.5.** — If $f\colon X \hookrightarrow Y$ is an inclusion of a subspace X and if $\mathcal{G}$ is a sheaf on Y, then we also write $\mathcal{G}|_X := f^{-1}\mathcal{G}$. If X is an open subspace, then this coincides with the restriction of $\mathcal{G}$ to X (see definition VIII.2.1.9, the direct limit becomes trivial).

***Lemma* VIII.2.4.6.** — *Let $f\colon X \to Y$ be a continuous map, and let $\mathcal{G}$ be a presheaf on Y. Then for all $x \in X$, we have*

$$(f^{-1}\mathcal{G})_x \cong \mathcal{G}_{f(x)},$$

which is functorial in $\mathcal{G}$.

Proof. — With proposition VIII.2.3.9, we have

$$(f^{-1}\mathcal{G})_x \cong (f^+\mathcal{G})_x = \varinjlim_{U \ni x} f^+\mathcal{G}(U) = \varinjlim_{U\ni x} \varinjlim_{V \supseteq f(U)} \mathcal{G}(V) = \varinjlim_{V \ni f(x)} \mathcal{G}(V) = \mathcal{G}_{f(x)}.$$

***Fact* VIII.2.4.7.** — *Let $f\colon X \to Y$ and $g\colon Y \to Z$ be two continuous maps, and let $\mathcal{H}$ be a presheaf on Z. Then*

$$f^{-1}(g^{-1}\mathcal{H}) \cong (g \circ f)^{-1}\mathcal{H},$$

which is functorial in $\mathcal{H}$.

Proof. — Fix an open subset $U \subseteq X$. For an open subset $W \subseteq Z$, we have $W \supseteq g(f(U))$ iff $W \supseteq g(V)$ for some open subset $V \subseteq Y$ with $V \supseteq f(U)$: indeed, set $V := g^{-1}(W)$. We thus obtain

$$f^+(g^+\mathcal{H}(U)) = \varinjlim_{V\supseteq f(U)} \varinjlim_{W \supseteq g(V)} \mathcal{H}(W) = \varinjlim_{W \supseteq g(f(U))} \mathcal{H}(W) = (g\circ f)^+\mathcal{H}(U),$$

i. e., $f^+(g^+\mathcal{H}) = (g \circ f)^+\mathcal{H}$. Sheafifying yields $f^{-1}(g^+\mathcal{H}) = (g\circ f)^{-1}\mathcal{H}$.

The canonical map $f^{-1}\iota_{g^+\mathcal{H}}\colon f^{-1}(g^+\mathcal{H}) \to f^{-1}(g^{-1}\mathcal{H})$ induces $(g^+\mathcal{H})_{f(x)} \to (g^{-1}\mathcal{H})_{f(x)}$ on stalks for any $x \in X$ by virtue of lemma VIII.2.4.6. The map on stalks is bijective by proposition VIII.2.3.9, so $f^{-1}\iota_{g^+\mathcal{H}}$ is an isomorphism by proposition VIII.2.3.5. Thus, $(g\circ f)^{-1}\mathcal{H} \xrightarrow{\sim} f^{-1}(g^{-1}\mathcal{H})$.

***Proposition* VIII.2.4.8** (direct image-inverse image adjunction)**.** — *Let $f\colon X \to Y$ be a continuous map. Let $\mathcal{F}$ be a sheaf on X and let $\mathcal{G}$ be a presheaf on Y. Then there is a bijection*

$$\begin{aligned} \mathrm{Hom}_{\mathsf{Sh}(X)}(f^{-1}\mathcal{G}, \mathcal{F}) &\longleftrightarrow \mathrm{Hom}_{\mathsf{PSh}(Y)}(\mathcal{G}, f_*\mathcal{F}), \\ \phi &\longmapsto \phi^\flat, \\ \psi^\sharp &\longleftarrow\!\shortmid \psi, \end{aligned}$$

functorial in $\mathcal{F}$ and $\mathcal{G}$. In other words, we have the adjoint pair $f^{-1} \dashv f_*$.

Proof. — We first construct the two maps. Let $\phi\colon f^{-1}\mathcal{G} \to \mathcal{F}$ be a morphism in $\mathsf{Sh}(X)$, and let $V \subseteq Y$ be open. Since $V \supseteq f(f^{-1}(V))$, there is a canonical map $\mathcal{G}(V) \to f_*(f^+\mathcal{G})(V)$ given by the direct limit. We define $\phi^\flat$ by

$$\phi^\flat_V\colon \mathcal{G}(V) \longrightarrow f_*(f^+\mathcal{G})(V) \longrightarrow f_*(f^{-1}\mathcal{G})(V) \xrightarrow{(f_*\phi)_V} f_*\mathcal{F}(V).$$

Conversely, let $\psi\colon \mathcal{G} \to f_*\mathcal{F}$ be a morphism in $\mathsf{PSh}(Y)$. To define $\psi^\sharp$, it suffices to define a morphism of presheaves $f^+\mathcal{G} \to \mathcal{F}$, which we will also denote by $\psi^\sharp$, and then sheafify. Let $U \subseteq X$ be open and let $s \in f^+\mathcal{G}(U)$. Let $V \supseteq f(U)$ be open such that s is represented by some $s_V \in \mathcal{G}(V)$ in the direct limit. Note that $\psi_V(s_V) \in f_*\mathcal{F}(V)$ and $f^{-1}(V) \supseteq U$. Hence, define $\psi^\sharp_U(s) := \psi_V(s_V)|_U \in \mathcal{F}(U)$, which clearly does not depend on the choice of V.

To check that they are mutually inverse is straightforward and hence omitted. In accordance with the construction of $\psi^\sharp$, one should instead consider $\phi\colon f^+\mathcal{G} \to \mathcal{F}$ by pre-composing $f^{-1}\mathcal{G} \to \mathcal{F}$ with the canonical map $\iota_{f^+\mathcal{G}}\colon f^+\mathcal{G} \to f^{-1}\mathcal{G}$. Then $\phi^\flat$ is given more succinctly by $\mathcal{G} \to f_*(f^+\mathcal{G}) \to f_*\mathcal{F}$.

It remains to check that both maps are functorial in $\mathcal{F}$ and $\mathcal{G}$. This is quite cumbersome and hence omitted.

Maybe fill in the details.

Alternative proof of fact VIII.2.4.7. — Let $f\colon X \to Y$ and $g\colon Y \to Z$ be continuous maps. For a presheaf $\mathcal{H}$ on Z, proposition VIII.2.4.8 and $(g\circ f)_* = g_* \circ f_*$ give the bijection

$$\mathrm{Hom}_{\mathsf{Sh}(X)}((g\circ f)^{-1}\mathcal{H}, \mathcal{F}) \cong \mathrm{Hom}_{\mathsf{Sh}(X)}(f^{-1}(g^{-1}\mathcal{H}), \mathcal{F})$$

functorial in $\mathcal{F}$. By the Yoneda lemma, we obtain the isomorphism $f^{-1}(g^{-1}\mathcal{H}) \cong (g\circ f)^{-1}\mathcal{H}$.

Insert reference. This is the contravariant version.

Alternative proof of lemma VIII.2.4.6. — Let $f\colon X \to Y$ be a continuous map, and let $\mathcal{G}$ be a presheaf on Y. Observe that if $i\colon \{x\} \hookrightarrow X$ is the inclusion of a point $x \in X$, then $i^{-1}\mathcal{F} = \underline{\mathcal{F}_x}$ for any presheaf $\mathcal{F}$ on X by definition. Since the diagram

$$\begin{tikzcd} \{x\} \arrow[r, "i"] \arrow[d, "f|_x"'] & X \arrow[d, "f"] \\ \{f(x)\} \arrow[r, "j"] & Y \end{tikzcd}$$

commutes and since $f|_x$ is a homeomorphism, we have

$$(f^{-1}\mathcal{G})_x = i^{-1}(f^{-1}\mathcal{G}) \cong (f \circ i)^{-1}\mathcal{G} = (j \circ f|_x)^{-1}\mathcal{G} \cong j^{-1}\mathcal{G} = \mathcal{G}_{f(x)}.$$

***Remark* VIII.2.4.9.** — We will almost never use the concrete construction of $f^{-1}\mathcal{G}$. The typical setting is that of a continuous map $f\colon X \to Y$, a sheaf $\mathcal{F}$ on X and a presheaf $\mathcal{G}$ on Y. Given a morphism of presheaves $\psi\colon \mathcal{G} \to f_*\mathcal{F}$, we have the morphism $\psi^\sharp\colon f^{-1}\mathcal{G} \to \mathcal{F}$ of sheaves and it is usually sufficient to understand the maps

$$\psi^\sharp_x\colon \mathcal{G}_{f(x)} = (f^{-1}\mathcal{G})_x \longrightarrow \mathcal{F}_x$$

on stalks for each $x \in X$ (the identification is lemma VIII.2.4.6). This is constructed as follows by the proof of proposition VIII.2.4.8: for every open neighbourhood $V \subseteq Y$ of $f(x)$, we have

$$\mathcal{G}(V) \xrightarrow{\psi_V} f_*\mathcal{F}(V) = \mathcal{F}(f^{-1}(V)) \longrightarrow \mathcal{F}_x,$$

and taking the direct limit over all such V yields $\psi^\sharp_x\colon \mathcal{G}_{f(x)} = \varinjlim_{V \ni f(x)} \mathcal{G}(V) \to \mathcal{F}_x$.

***Remark* VIII.2.4.10.** — It is clear that if $\mathcal{F}$ is a sheaf with values in a category $\mathcal{C}$, then $f_*\mathcal{F}$ is a sheaf with values in $\mathcal{C}$ as well. The same is true for $f^{-1}\mathcal{G}$. Moreover, proposition VIII.2.4.8 is also true in this setting (with the same proof) if $\mathcal{C}$ is concrete.

***Definition* VIII.2.4.11.** — *Let X be a topological space and let $i\colon \{x\} \hookrightarrow X$ be the inclusion of a point. For any set E, we call $i_*\underline{E}_{\{x\}}$ the **skyscraper sheaf** in x with value E. In other words, for every $U \subseteq X$ open we have*

$$i_*\underline{E}(U) = \begin{cases} E & \text{if } x \in U, \\ \{*\} & \text{if } x \notin U. \end{cases}$$

***Fact* VIII.2.4.12.** — *Let $i\colon \{x\} \hookrightarrow X$ be the inclusion of a point into a topological space, and let E be a set. Then for all $y \in X$, we have*

$$(i_*\underline{E})_y = \begin{cases} E & \text{if } y \in \overline{\{x\}}, \\ \{*\} & \text{if } y \notin \overline{\{x\}}. \end{cases}$$

Proof. — There exists an open neighbourhood, say, U of y not containing x iff $y \notin \overline{\{x\}}$. In this case, $i_*\underline{E}(U) = \{*\}$ and the statement follows from the definitions.

***Corollary* VIII.2.4.13** (stalk-skyscraper adjunction). — *Let $i\colon \{x\} \hookrightarrow X$ be the inclusion of a point into a topological space. Let E be a set and let $\mathcal{F}$ be a sheaf on x. Then there is a bijection*

$$\mathrm{Hom}_{\mathsf{Set}}(\mathcal{F}_x, E) \longleftrightarrow \mathrm{Hom}_{\mathsf{Sh}(X)}(\mathcal{F}, i_*\underline{E}).$$

In other words, taking stalks is the left adjoint of constructing skyscraper sheaves.

Proof. — Since $\{x\}$ is a singleton set, we have the obvious bijection

$$\mathrm{Hom}_{\mathsf{Set}}(\mathcal{F}_x, E) \longleftrightarrow \mathrm{Hom}_{\mathsf{Sh}(\{x\})}(\underline{\mathcal{F}_x}, \underline{E}).$$

Observe that $i^{-1}\mathcal{F} = \underline{\mathcal{F}_x}$ by definition, so the statement is a consequence of proposition VIII.2.4.8.

VIII.2.5. Locally ringed spaces.

***Definition* VIII.2.5.1.** — *A **ringed space** is a pair $(X, \mathcal{O}_X)$ of a topological space X and a sheaf $\mathcal{O}_X$ of (commutative) rings on X. We call $\mathcal{O}_X$ the **structure sheaf** of $(X, \mathcal{O}_X)$ and often write X instead of $(X, \mathcal{O}_X)$.*

*If $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$ are ringed spaces, then a **morphism of ringed spaces** is a pair $(f, f^\flat) \colon (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ consisting of a continuous map $f \colon X \to Y$ and a morphism $f^\flat \colon \mathcal{O}_Y \to f_*\mathcal{O}_X$ of sheaf of rings on Y. Instead of $f^\flat$, the same datum can be given by the morphism $f^\sharp \colon f^{-1}\mathcal{O}_Y \to \mathcal{O}_X$ of sheaf of rings on X by proposition VIII.2.4.8. We will often write f instead of $(f, f^\sharp)$ or $(f, f^\flat)$.*

The composition of morphisms of ringed spaces is defined as usual (using $(g \circ f)_* = g_* \circ f_*$ or $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$). Thus, we obtain the category of ringed spaces.

***Motivation* VIII.2.5.2.** — Recall that we think of a sheaf on X as a system of functions on subsets of X. The structure sheaf $\mathcal{O}_X$, in particular, should be thought of as the system of all 'permissible' functions within the current context, e. g., of all continuous, differentiable, holomorphic, polynomial, etc. functions. Moreover, a morphism $(f, f^\flat)$ should transport a permissible function on any open subset $V \subseteq Y$ to a permissible function on $f^{-1}(V) \subseteq X$ by pulling back along f. Since the sections are not actual (permissible) functions, we instead have maps $f^\flat_V \colon \mathcal{O}_Y \to f_*\mathcal{O}_X(V)$ for every open subset $V \subseteq Y$.

Define this term.

VIII.2.5.3. — Recall that a map $\phi \colon (A, \mathfrak{m}_A) \to (B, \mathfrak{m}_B)$ of local rings is *local* if $\phi(\mathfrak{m}_A) \subseteq \mathfrak{m}_B$.

***Definition* VIII.2.5.4.** — *A **locally ringed space** is a ringed space $(X, \mathcal{O}_X)$ such that the stalk $\mathcal{O}_{X,x}$ is a local ring for all $x \in X$. We call $\mathcal{O}_{X,x}$ the **local ring** of X in x, and denote by $\mathfrak{m}_x$ the **maximal ideal** and by $\kappa(x) = \mathcal{O}_{X,x}/\mathfrak{m}_x$ the **residue field** of $\mathcal{O}_{X,x}$. For an open neighbourhood $U \subseteq X$ of x and for a section $f \in \mathcal{O}_X(U)$, the value $f(x)$ denotes the image of f under the canonical map $\mathcal{O}_X(U) \to \mathcal{O}_{X,x} \to \kappa(x)$.*

*A **morphism of locally ringed spaces** is a morphism $(f, f^\flat) \colon (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ of ringed spaces such that the map $f^\sharp_x \colon (f^{-1}\mathcal{O}_Y)_x = \mathcal{O}_{Y,f(x)} \to \mathcal{O}_{X,x}$ on stalks (see remark VIII.2.4.9) is a local ring map for all $x \in X$. Composition is defined as usual, giving rise to the category of locally ringed spaces.*

***Warning* VIII.2.5.5.** — If $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$ are locally ringed spaces, then a morphism $(X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ of ringed spaces might not be a morphism of locally ringed spaces, i. e., the map on stalks might not be local. In other words, the category of locally ringed spaces is not a *full* subcategory of the category of ringed spaces. A counterexample is postponed when affine schemes are introduced.

Insert [GW20, Exer. 2.18].

***Motivation* VIII.2.5.6.** — Why *locally* ringed spaces $(X, \mathcal{O}_X)$? Recall that we think of sections of $\mathcal{O}_X$ as functions on an open subset of X, and of germs in the stalk $\mathcal{O}_{X,x}$ as (equivalence classes of) functions defined on an open neighbourhood of $x \in X$. In classical settings, we want that functions which do not vanish at x to be locally invertible. In other words, all germs not contained in the ideal of functions vanishing at x are units in $\mathcal{O}_{X,x}$. Hence, $\mathcal{O}_{X,x}$ is a local ring and this ideal is the maximal ideal $\mathfrak{m}_x$.

Now, consider a morphism $(f, f^\flat)\colon (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ of ringed spaces where $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$ are locally ringed spaces. If a function on an open subset of Y vanishes at $f(x)$ for some $x \in X$, then the pullback along f must vanish at x. In other words, $f_x^\sharp\colon \mathcal{O}_{Y,f(x)} \to \mathcal{O}_{X,x}$ must map $\mathfrak{m}_{f(x)}$ into $\mathfrak{m}_x$. Since sections are not functions in general, we must put locality in our definition explicitly.

***Example* VIII.2.5.7.** — Let $(X, \mathcal{C}_X)$ be the ringed space of $\mathbb{R}$-valued continuous functions on a topological space X, i. e., for $U \subseteq X$ open, $\mathcal{C}_X(U)$ is the $\mathbb{R}$-algebra of continuous functions $s\colon U \to \mathbb{R}$.

Fix a point $x \in X$. The stalk $\mathcal{C}_{X,x}$ is the ring of germs $[s]$ of functions s defined on a neighbourhood of x. Let $\mathfrak{m}_x = \{[s] \in \mathcal{C}_{X,x} \mid s(x) = 0\}$, which is obviously a proper ideal. Let $[s] \in \mathcal{C}_{X,x} \setminus \mathfrak{m}_x$ be arbitrary with representative s. Since s is continuous with $s(x) \neq 0$, there is a (small) neighbourhood U of x such that $s(u) \neq 0$ for all $u \in U$. Hence, the inverse $[1/(s|_U)]$ of $[s]$ exists and therefore, $\mathcal{C}_{X,x}$ is a local ring, whence $(X, \mathcal{C}_X)$ is a locally ringed space. Moreover, the evaluation map $\mathcal{C}_{X,x} \to \mathbb{R}$, $[s] \mapsto s(x)$ is obviously surjective with kernel $\mathfrak{m}_x$, whence $\kappa(x) \cong \mathbb{R}$.

If $f\colon X \to Y$ is a continuous map, then we obtain a $\mathbb{R}$-algebra maps $f_V^\flat\colon \mathcal{C}_Y(V) \to \mathcal{C}_X(f^{-1}(V))$, $t \mapsto t \circ f$ for every $V \subseteq Y$ open. This defines the morphism $f^\flat\colon \mathcal{C}_Y \to f_*\mathcal{C}_X$ of sheaves. The induced map of stalks is $f_x^\sharp\colon \mathcal{C}_{Y,f(x)} \to \mathcal{C}_{X,x}$, $[t] \mapsto [t \circ f]$. Obviously, $f_x^\sharp(\mathfrak{m}_{f(x)}) \subseteq \mathfrak{m}_x$, whence $(f, f^\flat)$ is a morphism of locally ringed spaces.

***Example* VIII.2.5.8.** — Let $(X, \mathcal{O}_X)$ be a prevariety over an algebraically closed field k, i. e., for $U \subseteq X$ open, $\mathcal{O}_X(U)$ is a k-algebra of certain continuous functions $U \to \mathbb{A}^1(k)$ w. r. t. the Zariski topology. As in example VIII.2.5.7, we show that $\mathfrak{m}_x = \{[f] \in \mathcal{O}_{X,x} \mid f(x) = 0\}$ is the unique maximal ideal of $\mathcal{O}_{X,x}$ (use $f^{-1}(\mathbb{A}^1(k) \setminus \{0\})$ in place of U). Hence, $(X, \mathcal{O}_X)$ is a locally ringed space. Moreover, the evaluation map $\mathcal{O}_{X,x} \to k$, $[f] \mapsto f(x)$ is surjective, since $(X, \mathcal{O}_X)$ is locally an affine variety $(U, \mathcal{O}_X|_U)$ with $x \in U$ and since $\mathcal{O}_X(U)$ contains all constant functions. The kernel of the map is $\mathfrak{m}_x$, so we obtain $\kappa(x) \cong k$ for all $x \in X$.

We now study a prime example of the algebro-geometric perspective algebraic geometry offers: the connection between connected components and idempotents. However, we first make the following observation.

***Lemma* VIII.2.5.9.** — *Let $\mathcal{F}$ be a sheaf on a topological space X. If $s, t \in \mathcal{F}(U)$ are two sections with $U \subseteq X$ open, then $\{x \in X \mid s_x = t_x\}$ is open in U.*

Proof. — If $s_x = t_x$ for some $x \in X$, then there is some open neighbourhood $V \subseteq U$ of x such that $s|_V = t|_V$. Hence $s_y = t_y$ for all $y \in V$, that is, V is an open neighbourhood of x in $\{x \in X \mid s_x = t_x\}$. □

***Proposition* VIII.2.5.10.** — *Let $(X, \mathcal{O}_X)$ be a locally ringed space. Then there is a bijection*

$$\{\textit{open and closed subsets of } X\} \longrightarrow \{\textit{idempotents of } \Gamma(X, \mathcal{O}_X)\},$$
$$U \longmapsto e_U$$

where $e_U \in \Gamma(X, \mathcal{O}_X)$ is the unique global section such that $e_U|_V = 1$ for all $V \subseteq U$ open and $e_U|_V = 0$ for all $V \subseteq X \setminus U$ open.

Proof. — Note that e_U exists since the local sections $1 \in \Gamma(U, \mathcal{O}_X)$ and $0 \in \Gamma(X \setminus U, \mathcal{O}_X)$ glue to a unique global section e_U. Moreover, for $U, U' \subseteq X$ both open and closed, we have $e_U e_{U'} = e_{U \cap U'}$ on the open subsets $X \setminus U$, $X \setminus U'$ and $U \cap U'$, and hence we have $e_U e_{U'} = e_{U \cap U'}$ in $\Gamma(X, \mathcal{O}_X)$. In particular, for $U = U'$, we see that e_U is idempotent.

As for whether the inverse map, consider

$$U_e := \{x \in X \mid e_x = 1_x\} \longleftarrow e.$$

Note that U_e is open by lemma VIII.2.5.9. On the other hand, we have $e_x \in \{0_x, 1_x\}$ since the image of an idempotent is again idempotent and since the idempotents of the local ring $\mathcal{O}_{X,x}$ are trivial. Hence, $X \setminus U_e = \{x \in X \mid e_x = 0_x\}$ which is also open by lemma VIII.2.5.9, i. e., U_e is closed.

One easily verifies that both maps are mutually inverse map. ∎

***Corollary* VIII.2.5.11.** — *Let $(X, \mathcal{O}_X)$ be a non-empty locally ringed space. Then the following are equivalent:*

(i) *X is connected;*
(ii) *the idempotents of $\Gamma(X, \mathcal{O}_X)$ are trivial;*
(iii) *there is no decomposition $\Gamma(X, \mathcal{O}_X) = A_1 \times A_2$ into non-trivial rings A_1 and A_2.*

Proof. — The equivalence of (i) and (ii) is proposition VIII.2.5.10. On the other hand, the equivalence of (ii) and (iii) is purely algebraic.

Write $A := \Gamma(X, \mathcal{O}_X)$. If $e \in A$ is a non-trivial idempotent, then we have a non-trivial decomposition $A = eA \times (1-e)A$ as an A-module. Since e and $1-e$ are idempotents, eA and $(1-e)A$ are, in fact, rings. Conversely, if $A = A_1 \times A_2$ is a non-trivial decomposition as a ring, then $(1,0)$ and $(0,1)$ are non-trivial idempotents in A. ∎

CHAPTER IX

SCHEMES

§ IX.1. SPECTRUM OF A RING

***Motivation* IX.1.0.1.** — Let k be an algebraically closed field. Recall that we have attached for any subset $M \subseteq k[T_1, \dots, T_n]$ the affine algebraic set $\mathrm{V}(M) \subseteq \mathbb{A}^n(k) = k^n$. If this set $X = \mathrm{V}(M)$ is irreducible, then we attach to it the space of functions $(X, \mathcal{O}_X)$ called affine varieties. Gluing finitely many of them, we obtain prevarieties.

In light of motivation VIII.1.3.6, we will generalise this by constructing, given any ring A, the topological space $\operatorname{Spec} A$ whose underlying set consists of the prime ideals of A. This is a generalisation insofar that an affine variety X corresponds to a topological space $\operatorname{MaxSpec} A$ whose underlying set consists of the maximal ideals of a ring A (more precisely, A is an integral finitely generated k-algebra, see remark VIII.1.4.10). In short, $\operatorname{Spec} A$ has more points. The advantage is that Spec gives a functorial construction, since prime ideals are stable under contractions, while maximal ideals are generally not.

However, we cannot recover the ring A from $\operatorname{Spec} A$ alone, as opposed to the correspondence between coordinate rings $\Gamma(X)$ and affine varieties X (cf. theorem VIII.1.3.12). Indeed, $\operatorname{Spec} K$ is the singleton for any field K, but if K is a finite field extension of an algebraically closed field k then $K \cong k$. This is why we endow $\operatorname{Spec} A$ with 'functions' via a sheaf $\mathcal{O}_{\operatorname{Spec} A}$ of rings on $\operatorname{Spec} A$. Together, $(\operatorname{Spec} A, \mathcal{O}_{\operatorname{Spec} A})$ is an example of a locally ringed space which we call *affine schemes*. Gluing them will give the basic object in modern algebraic geometry: *schemes*.

IX.1.1. Spectrum of a ring as a topological space.

***Definition* IX.1.1.1.** — *Let A be a ring. Recall that* $\operatorname{Spec} A$ *is the set of all prime ideals $\mathfrak{p}$ of A. For every subset $M \subseteq A$, we denote*

$$\mathrm{V}(M) := \{\mathfrak{p} \in \operatorname{Spec} A \mid \mathfrak{p} \supseteq M\}.$$

Obviously, $\mathrm{V}(M) = \mathrm{V}((M))$*. Moreover, we will write* $\mathrm{V}(f)$ *instead of* $\mathrm{V}(\{f\})$ *for any $f \in A$.*

***Lemma* IX.1.1.2.** — *The map $\mathfrak{a} \mapsto \mathrm{V}(\mathfrak{a})$ on the set of ideals of A is inclusion reversing. Moreover, we have the following:*

(i) $\mathrm{V}(0) = \operatorname{Spec} A$ *and* $\mathrm{V}(1) = \varnothing$*;*
(ii) *for every family $(\mathfrak{a}_i)_{i\in I}$ of ideals*

$$\mathrm{V}\left(\bigcup_{i\in I} \mathfrak{a}_i\right) = \mathrm{V}\left(\sum_{i\in I} \mathfrak{a}_i\right) = \bigcap_{i\in I} \mathrm{V}(\mathfrak{a}_I);$$

(iii) *for two ideals $\mathfrak{a}$ and $\mathfrak{a}'$*

$$\mathrm{V}(\mathfrak{a} \cap \mathfrak{a}') = \mathrm{V}(\mathfrak{a}\mathfrak{a}') = \mathrm{V}(\mathfrak{a}) \cup \mathrm{V}(\mathfrak{a}').$$

Proof. — (i) is obvious. For (ii), note that $(\bigcup_{i\in I}\mathfrak{a}_i) = \sum_{i\in I}\mathfrak{a}_i$. For (iii), we have $\mathrm{V}(\mathfrak{a}\cap\mathfrak{a}') \subseteq \mathrm{V}(\mathfrak{a}\mathfrak{a}') \subseteq \mathrm{V}(\mathfrak{a})\cup\mathrm{V}(\mathfrak{a}')$ by the prime avoidance lemma I.1.3.9.

Definition* IX.1.1.3.** — *Let A be a ring. The set* $\operatorname{Spec} A$ *with the topology whose closed sets are* $\mathrm{V}(\mathfrak{a})$*, where* $\mathfrak{a}$ *runs through all ideals of* A*, is called the* ***prime spectrum *or simply* ***spectrum*** *of* A*. The topology is called the* ***Zariski topology*** *on* $\operatorname{Spec} A$.

***Notation* IX.1.1.4.** — For a point $x \in \operatorname{Spec} A$, we will also write $\mathfrak{p}_x$ instead of x to emphasize that x is a prime ideal of A.

***Remark* IX.1.1.5.** — This is analogous to definition VIII.1.1.1 where $A = k[T_1, \dots, T_n]$ for an algebraically closed field k and where $\mathrm{V}(\mathfrak{a})$ corresponds to the *maximal* ideals of A containing $\mathfrak{a}$. We will later see that maximal ideals suffice when considering finitely generated algebras over a field.

[GW20] gives §3.13 as a reference.

***Definition* IX.1.1.6.** — *For a subset Y of* $\operatorname{Spec} A$*, we define*

$$\mathrm{I}(Y) := \bigcap_{\mathfrak{p}\in Y}\mathfrak{p}.$$

By definition, $\mathrm{I}(\varnothing) = A$.

***Fact* IX.1.1.7.** — *The map* $Y \mapsto \mathrm{I}(Y)$ *on the power set of* $\operatorname{Spec} A$ *is inclusion reversing. Moreover, we have* $\sqrt{\mathrm{I}(Y)} = \mathrm{I}(Y)$.

Proof. — Prime ideals are radical, and taking radicals commutes with arbitrary intersections.

***Proposition* IX.1.1.8** (cf. Hilbert's Nullstellensatz VIII.1.1.17)**.** — *Let A be a ring. Then we have the following:*

(i) $\mathrm{I}(\mathrm{V}(\mathfrak{a})) = \sqrt{\mathfrak{a}}$ *for all ideals* $\mathfrak{a}$ *of* A*;*
(ii) $\mathrm{V}(\mathrm{I}(Y)) = \overline{Y}$ *for all subsets Y of* $\operatorname{Spec} A$.

Thus, there is the bijection

$$\{\textit{radical ideals of } A\} \underset{\mathrm{I}}{\overset{\mathrm{V}}{\rightleftarrows}} \{\textit{closed subsets of } \operatorname{Spec} A\}.$$

In particular, for $x \in \operatorname{Spec} A$ *we have* $\overline{\{x\}} = \mathrm{V}(\mathfrak{p}_x)$.

Proof. — (i) is proposition I.1.3.17. For (ii), observe that a closed set $\mathrm{V}(\mathfrak{b})$, with $\mathfrak{b}$ some ideal, contains Y iff $\mathfrak{b} \subseteq \mathfrak{p}$ for every $\mathfrak{p} \in Y$, that is, $\mathfrak{b} \subseteq \mathrm{I}(Y)$. Hence, $\mathrm{V}(\mathrm{I}(Y))$ is the smallest closed subset of $\operatorname{Spec} A$ containing Y (any even smaller closed subset $\mathrm{V}(\mathfrak{b})$ would satisfy $\mathfrak{b} \not\subseteq \mathrm{I}(Y)$).

***Definition* IX.1.1.9.** — *Let A be a ring. For $f \in A$,*

$$\mathrm{D}(f) := \mathrm{D}_A(f) := \operatorname{Spec} A \setminus \mathrm{V}(f)$$

is the open set of prime ideals not containing f and is called a ***principal open set*** *of* $\operatorname{Spec} A$.

***Fact* IX.1.1.10.** — *We have* $\mathrm{D}(0) = \varnothing$ *and* $\mathrm{D}(u) = \operatorname{Spec} A$ *for all units* $u \in A^\times$*. By definition of prime ideals, we have*

$$\mathrm{D}(f) \cap \mathrm{D}(g) = \mathrm{D}(fg)$$

for all $f, g \in A$.

***Lemma* IX.1.1.11.** — *Let* $(f_i)_{i\in I}$ *be a family of elements in A and let $g \in A$. Then* $\mathrm{D}(g) \subseteq \bigcup_{i\in I}\mathrm{D}(f_i)$ *iff* $g^n \in (f_i, i \in I)$ *for some* $n \geq 1$*. In particular for* $g = 1$*, it follows that* $\bigcup_{i\in I}\mathrm{D}(f_i)$ *is an open covering of* $\operatorname{Spec} A$ *iff* $(f_i, i \in I)$ *is the unit ideal.*

Proof. — Let $\mathfrak{a} = (f_i, i \in I)$. We have $\mathrm{D}(g) \subseteq \bigcup_{i\in I} \mathrm{D}(f_i)$ iff $\mathrm{V}(g) \supseteq \bigcap_{i\in I} \mathrm{V}(f_i) = \mathrm{V}(\mathfrak{a})$ iff $g \in \mathrm{I}(\mathrm{V}(\mathfrak{a})) = \sqrt{\mathfrak{a}}$ by proposition IX.1.1.8.

***Proposition* IX.1.1.12.** — *Let A be a ring. Then the principal open subsets $\mathrm{D}(f)$ for $f \in A$ form a basis of the Zariski topology of $\operatorname{Spec} A$. Moreover, $\mathrm{D}(f)$ is quasi-compact, and in particular, $\operatorname{Spec} A$ is quasi-compact.*

Proof. — By lemma IX.1.1.2, every closed subset $\mathrm{V}(\mathfrak{a})$ can be written as $\bigcup_{f\in\mathfrak{a}} \mathrm{V}(f)$. Taking complements yields the first statement.

For the second statement, let $\mathrm{D}(f) \subseteq \bigcup_{i\in I} \mathrm{D}(g_i)$ with $g_i \in A$ be an open covering. By lemma IX.1.1.11, we have a finite sum $f^n = \sum_{i\in I} a_i g_i$ for some $n \geq 1$ and $a_i \in A$ where almost all a_i are zero. If $J \subseteq I$ is the finite subset of indices for which $a_j \neq 0$ for all $j \in J$, then we have $\mathrm{D}(f) \subseteq \bigcup_{j\in J} \mathrm{D}(g_j)$, again by lemma IX.1.1.11.

***Proposition* IX.1.1.13.** — *Let A be a ring. A subset $Y \subseteq \operatorname{Spec} A$ is irreducible iff $\mathfrak{p} := \mathrm{I}(Y)$ is prime. In this case, $\{\mathfrak{p}\}$ is dense in $\overline{Y}$.*

Proof. — Suppose that Y is irreducible. If $fg \in \mathfrak{p} = \mathrm{I}(Y)$, then $Y \subseteq \mathrm{V}(fg) = \mathrm{V}(f) \cup \mathrm{V}(g)$. Since Y is irreducible, we have $Y \subseteq \mathrm{V}(f)$ or $Y \subseteq \mathrm{V}(g)$, so $f \in \mathfrak{p}$ or $g \in \mathfrak{p}$.

Conversely, suppose that $\mathfrak{p}$ be prime. Then proposition IX.1.1.8 implies that $\overline{Y} = \mathrm{V}(\mathrm{I}(Y)) = \mathrm{V}(\mathrm{I}(\{\mathfrak{p}\})) = \overline{\{\mathfrak{p}\}}$. Since $\{\mathfrak{p}\}$ is irreducible, lemma VIII.1.2.4 says that $\overline{Y}$ as its closure is irreducible and that Y as a dense subset of $\overline{Y}$ is irreducible as well.

***Warning* IX.1.1.14.** — Note that for irreducible subsets Y the prime ideal $\mathrm{I}(Y)$ is certainly a point in $\overline{Y}$, but not necessarily in Y. This is clearly true if Y is closed or, more generally, if Y is locally closed: by definition, Y being locally closed means that $\overline{Y} \setminus Y$ is a closed subset of $\operatorname{Spec} A$, so $\mathrm{I}(Y) \in \overline{Y} \setminus Y$ would give the contradiction $\overline{Y} = \overline{\{\mathrm{I}(Y)\}} = \overline{Y} \setminus Y$.

***Corollary* IX.1.1.15.** — *The map*

$$\begin{aligned} \operatorname{Spec} A &\longrightarrow \{\textit{irreducible closed subsets of } \operatorname{Spec} A\}, \\ \mathfrak{p} &\longmapsto \mathrm{V}(\mathfrak{p}) = \overline{\{\mathfrak{p}\}} \end{aligned}$$

is a bijection. This further restricts to the bijection

$$\{\textit{minimal prime ideals of } A\} \longleftrightarrow \{\textit{irreducible components of } \operatorname{Spec} A\}.$$

Proof. — This map is the restriction of the bijection in proposition IX.1.1.8 by virtue of proposition IX.1.1.13.

***Definition* IX.1.1.16.** — *Let X be an arbitrary topological space.*

(i) *A point $x \in X$ is called a* ***closed point*** *if $\{x\}$ is closed.*
(ii) *A point $\eta \in X$ is called a* ***generic point*** *if $\{\eta\}$ is dense in X.*
(iii) *Let $x, x' \in X$. We say that x is a* ***generisation*** *of x' or that x' is a* ***specialisation*** *of x if $x' \in \overline{\{x\}}$.*
(iv) *A point $x \in X$ is called a* ***maximal point*** *if $\overline{\{x\}}$ is an irreducible component of X.*

Thus, $\eta \in X$ is generic iff it is the generisation of every point of X. By lemma VIII.1.2.4, X is irreducible if it has a generic point.

***Example* IX.1.1.17.** — Let $X = \operatorname{Spec} A$ for a ring A. Recall that the closure of a point $x \in X$ is $\mathrm{V}(\mathfrak{p}_x)$.

(i) A point $x \in X$ is closed iff $\mathfrak{p}_x$ is a maximal ideal.

(ii) A point $\eta \in X$ is generic iff $\mathfrak{p}_\eta$ is the unique minimal prime ideal. Hence a generic point exists iff $\operatorname{nil} A = \bigcap_{\mathfrak{p} \in \operatorname{Spec} A} \mathfrak{p}$ is prime iff, by proposition IX.1.1.13, X is irreducible.

(iii) A point x is a generisation of a point x' iff $\mathfrak{p}_x \subseteq \mathfrak{p}_{x'}$.

(iv) A point $x \in X$ is maximal iff $\mathfrak{p}_x$ is a minimal prime ideal by corollary IX.1.1.15.

IX.1.1.18. — Recall that if $\phi\colon A \to B$ is a ring map and if $\mathfrak{q}$ is a prime ideal of B, then the contraction $\phi^{-1}(\mathfrak{q})$ is a prime ideal of A. We obtain an induced map $\operatorname{Spec}\phi\colon \operatorname{Spec} B \to \operatorname{Spec} A$, $\mathfrak{q} \mapsto \phi^{-1}(\mathfrak{q})$.

***Proposition* IX.1.1.19.** — *Let $\phi\colon A \to B$ be a ring map.*

(i) *For every subset $M \subseteq A$, we have*

$$(\operatorname{Spec}\phi)^{-1}(\mathrm{V}(M)) = \mathrm{V}(\phi(M)).$$

In particular, $(\operatorname{Spec}\phi)^{-1}(\mathrm{D}(f)) = \mathrm{D}(\phi(f))$ *for all $f \in A$. Hence,* $\operatorname{Spec}\phi$ *is continuous and we obtain a functor* $\operatorname{Spec}\colon \mathsf{Ring}^{\mathrm{op}} \to \mathsf{Top}$.

(ii) *For every ideal $\mathfrak{b}$ of B, we have*

$$\mathrm{V}(\phi^{-1}(\mathfrak{b})) = \overline{(\operatorname{Spec}\phi)(\mathrm{V}(\mathfrak{b}))}.$$

In particular, $\mathrm{V}(\ker\phi) = \overline{\operatorname{im}\operatorname{Spec}\phi}$ *for $\mathfrak{b} = 0$. In other words,* $\operatorname{Spec}\phi$ *is* ***dominant*** *(i. e., has dense image in* $\operatorname{Spec} A$*) iff* $\ker\phi \subseteq \operatorname{nil} A$.

Proof. — (i) follows from the observation that a prime ideal $\mathfrak{q} \in \operatorname{Spec} B$ contains $\phi(M)$ iff $(\operatorname{Spec}\phi)(\mathfrak{q}) = \phi^{-1}(\mathfrak{q})$ contains $M \subseteq \phi^{-1}(\phi(M))$.

To show (ii), we can write $\overline{(\operatorname{Spec}\phi)(\mathrm{V}(\mathfrak{b}))} = \mathrm{V}(\mathrm{I}((\operatorname{Spec}\phi)(\mathrm{V}(\mathfrak{b}))))$ by proposition IX.1.1.8. However,

$$\mathrm{I}((\operatorname{Spec}\phi)(\mathrm{V}(\mathfrak{b}))) = \bigcap_{\mathfrak{q} \in \mathrm{V}(\mathfrak{b})} \phi^{-1}(\mathfrak{q}) = \phi^{-1}(\sqrt{\mathfrak{b}}) = \sqrt{\phi^{-1}(\mathfrak{b})}$$

by laws for contractions. Since $\mathrm{V}(\sqrt{\phi^{-1}(\mathfrak{b})}) = \mathrm{V}(\phi^{-1}(\mathfrak{b}))$ by definition, (ii) follows. ∎

***Proposition* IX.1.1.20.** — *Let A be a ring.*

(i) *Let $\mathfrak{a}$ be an ideal of A and let $\varphi\colon A \twoheadrightarrow A/\mathfrak{a}$ be the canonical projection. Then* $\operatorname{Spec}\phi$ *induces a homeomorphism* $\operatorname{Spec}(A/\mathfrak{a}) \xrightarrow{\sim} \mathrm{V}(\mathfrak{a})$.

(ii) *Let S be a multiplicative subset of A and let $\phi\colon A \to S^{-1}A$ be the canonical map. Then* $\operatorname{Spec}\phi$ *induces a homeomorphism* $\operatorname{Spec}(S^{-1}A) \xrightarrow{\sim} \{\mathfrak{p} \in \operatorname{Spec} A \mid S \cap \mathfrak{p} = \varnothing\} = \bigcap_{f \in S} \mathrm{D}(f)$.

Proof. — Let ϕ be any of the two maps. We already know that $\operatorname{Spec}\phi$ induces the stated bijections, so we only need to check that these are closed maps. If $\mathfrak{q}$ is a prime ideal of B and if $\mathfrak{b}$ is an ideal of B, then $\mathfrak{q} \supseteq \mathfrak{b}$ iff $\phi^{-1}(\mathfrak{q}) \supseteq \phi^{-1}(\mathfrak{b})$. Hence $(\operatorname{Spec}\phi)(\mathrm{V}(\mathfrak{b})) = \mathrm{V}(\phi^{-1}(\mathfrak{b})) \cap \operatorname{im}\operatorname{Spec}\phi$ and the induced maps are closed. ∎

***Example* IX.1.1.21.** — (i) If $A = \{0\}$, then $\operatorname{Spec} A = \varnothing$.

(ii) If A is a field or any ring with a single prime ideal, then $\operatorname{Spec} A$ is a singleton set.

(iii) If A is an artinian ring, then $\operatorname{Spec} A$ is finite and discrete (see propositions I.5.4.1 and I.5.4.3).

***Example* IX.1.1.22.** — Let A be a principal ideal domain, e. g., $A = \mathbb{Z}$ or $A = k[T]$ for k a field. The prime ideals are (0) and the maximal ideals (p) for $p \in A$ a prime element. Hence, the closed points of $\operatorname{Spec} A$ are all (p) (where $(p) = (p')$ if $p' = up$ for some $u \in A^\times$), and the generic point is $\eta = (0)$.

Moreover, every closed subset of $\operatorname{Spec} A$ is of the form $\mathrm{V}(f)$ for some $f \in A$. If f is non-zero, i.e., $\mathrm{V}(f)$ is proper, write $f = p_1^{e_1} \cdots p_r^{e_r}$ for non-associate prime elements $p_1, \dots, p_r$ (with $r \geq 0$) and

$e_i \geq 1$. Then $\mathrm{V}(f) = \{(p_1), \ldots, (p_r)\}$. Hence, the closed subsets of $\operatorname{Spec} A$ are the finite sets of closed points and $\operatorname{Spec} A$ itself.

If $g \in A$ is another non-zero element, then $\mathrm{V}(f) \cap \mathrm{V}(g) = \mathrm{V}(f, g) = \mathrm{V}(\gcd(f, g))$ and $\mathrm{V}(f) \cup \mathrm{V}(g) = \mathrm{V}((f) \cap (g)) = \mathrm{V}(\operatorname{lcm}(f, g))$.

If $(A, \mathfrak{m})$ is a local principal ideal domain but not a field, i. e., if A is a discrete valuation ring, then $\operatorname{Spec} A$ consists of two points: the generic point η with $\mathfrak{p}_\eta = (0)$ and the unique closed point x with $\mathfrak{p}_x = \mathfrak{m}$. The only non-trivial open subset of $\operatorname{Spec} A$ is $\{\eta\}$.

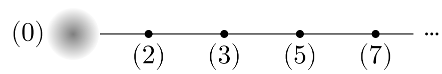

Figure IX.1: $\operatorname{Spec} \mathbb{Z}$ with the generic (dense) point (0).

***Example* IX.1.1.23.** — Let k be an algebraically closed field. Recall from theorem VIII.1.3.12 that there is a contravariant equivalence between the category of integral finitely generated k-algebras and the category of affine varieties. If V is an affine variety with corresponding k-algebra $\Gamma(V)$, then the maximal ideals of A correspond to the points of V. We can therefore consider V as a subset of $\operatorname{Spec} \Gamma(V)$. The definition of the Zariski topology of V shows that it coincides with the subspace topology of V in $\operatorname{Spec} \Gamma(V)$.

***Example* IX.1.1.24.** — Let $A = R[T]$ for R a principal ideal domain which is not a field; otherwise see example IX.1.1.22. Let $X = \operatorname{Spec} R[T]$. Then $R[T]$, like R, is a unique factorisation domain. The prime elements of $R[T]$ are of the prime elements $p \in R$ or the primitive polynomials $f \in R[T]$ which are irreducible in $(\operatorname{Quot} R)[T]$ (such f are also irreducible in $R[T]$).

Insert reference, Gauß' theorem.

Let $p \in R$ be a prime element. By proposition IX.1.1.20, the closure of $\{pR[T]\}$ is $\mathrm{V}(pR[T]) \cong \operatorname{Spec}(R/pR)[T]$ where R/pR is a field so that $(R/pR)[T]$ is a principal ideal domain with infinitely many non-associate prime elements (cf. example IX.1.1.22). In particular, $pR[T]$ is not maximal and $\mathrm{V}(pR[T])$ consists of $pR[T]$ and all maximal ideals (p, f) where $f \in R[T]$ which are irreducible modulo p.

The case of primitive irreducible $f \in R[T]$ is more complicated. If f is monic, then the Euclidean algorithm for division by f is possible, and therefore $R[T]/fR[T]$ is a free R-module of rank $\deg f$. In particular, $fR[T]$ is not maximal since R is not a field. On the other hand, $fR[T]$ might also be maximal, e. g., if R has finitely many prime elements. If $a \in R$ is a non-zero element divisible by all prime elements of R, then for $f := aT - 1$ we have $R[T]/fR[T] \cong R[a^{-1}] = \operatorname{Quot} R$, so in this case $fR[T]$ is a maximal ideal.

[GW20] mentions Lem. 1.9, which is going up ($A \to B$ then A field iff B field).

Weird example. Maybe add more stuff, i. e., complete description of $\operatorname{Spec} R[T]$.

IX.1.2. Spectrum of a ring as a locally ringed space.

***Motivation* IX.1.2.1.** — Like in the case of prevarieties, the topological space $\operatorname{Spec} A$ is not sufficient to describe geometric objects. Recall that we defined prevarieties by attaching a system of functions to it. The entire functions on the prevariety constitute the affine coordinate ring.

We now want to work backwards, in a certain sense: we start with a ring A and associate to it a topological space $\operatorname{Spec} A$. *This means that we should interpret elements of A as functions on* $\operatorname{Spec} A$.

Strictly speaking, this is not possible. However, if $f \in A$ and $x \in \operatorname{Spec} A$, then we defined $f(x) \in \kappa(x) := A_{\mathfrak{p}_x}/\mathfrak{p}_x A_{\mathfrak{p}_x}$ as the image of f in $\kappa(x)$ via $A \to A_{\mathfrak{p}_x} \to \kappa(x)$, see definition VIII.2.5.4. This is exactly what happens in the case of prevarieties, cf. construction VIII.1.4.6. The caveat is that f is not a function on $\operatorname{Spec} A$ proper with well-defined target, but a function $f\colon \operatorname{Spec} A \to \coprod_{x \in \operatorname{Spec} A} \kappa(x)$ with $f(x) \in \kappa(x)$. Nevertheless, this geometric point of view is still useful, e. g., $\mathrm{D}(f)$ is the set of points with $f(x) \neq 0$ (if $f \notin \mathfrak{p}_x$, then $f \in A_{\mathfrak{p}_x}^\times$ and $f(x) = \bar{f} \in \kappa(x) \setminus \{0\}$).

To circumvent this problem, we will replace A by the structure sheaf $\mathcal{O}_{\operatorname{Spec} A}$ of $\operatorname{Spec} A$. How should it be defined? We certainly want $\mathcal{O}_{\operatorname{Spec} A}(\operatorname{Spec} A) = A$. Moreover, if $\iota_f\colon A \to A_f$ denotes the canonical morphism of localisation for $f \in A$, then $\operatorname{Spec} \iota_f$ is a homeomorphism $\operatorname{Spec} A_f \xrightarrow{\sim} \mathrm{D}(f)$ by proposition IX.1.1.20, so it seems natural to define $\mathcal{O}_{\operatorname{Spec} A}(\mathrm{D}(f)) = A_f$ (also cf. proposition VIII.1.4.7 in the prevariety case). We should think of the elements A_f as functions with poles along $\mathrm{V}(f)$. Since the principal open subsets $\mathrm{D}(f)$ form a basis of the topology of $\operatorname{Spec} A$ by proposition IX.1.1.12, this is already sufficient to define the structure sheaf.

IX.1.2.2. — Recall that a *locally ringed space* $(X, \mathcal{O}_X)$ consists of a topological space X together with a structure sheaf $\mathcal{O}_X$ of rings on X with local rings $\mathcal{O}_{X,x}$ for all $x \in X$. A *morphism of locally ringed spaces* $(f, f^\flat)\colon (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ consists of a continuous map $f\colon X \to Y$ and a morphism $f^\flat\colon \mathcal{O}_Y \to f_*\mathcal{O}_X$ of sheaves on Y such that the map $f^\sharp_x\colon \mathcal{O}_{Y,f(x)} \to \mathcal{O}_{X,x}$ on stalks is local for all $x \in X$.

***Construction* IX.1.2.3.** — Let $X = \operatorname{Spec} A$ for a ring A. Let $\mathcal{B} = \{\mathrm{D}(f) \mid f \in A\}$, which is a basis of the Zariski topology of X by proposition IX.1.1.12.

We make the following observations. Suppose that $\mathrm{D}(f) \subseteq \mathrm{D}(g)$ for $f, g \in A$. By lemma IX.1.1.11, this is equivalent to $f^n \in (g)$ for some $n \geq 1$, that is, $g/1 \in A_f^\times$. By the universal property, the canonical map $A \to A_f$ factors through the unique ring map $\rho_{f,g}\colon A_g \to A_f$. Again by the universal property, we have $\rho_{f,g} \circ \rho_{g,h} = \rho_{f,h}$ whenever $\mathrm{D}(f) \subseteq \mathrm{D}(g) \subseteq \mathrm{D}(h)$. In particular, if $\mathrm{D}(f) = \mathrm{D}(g)$ then $\rho_{f,g}$ is an isomorphism and we identify $A_g = A_f$.

Hence, define the presheaf $\mathcal{O}_X$ of rings on $\mathcal{B}$ by

$$\mathcal{O}_X(\mathrm{D}(f)) \coloneqq A_f$$

for all $f \in A$, which is well-defined. The restriction maps are given by the $\rho_{f,g}$. In particular, $\operatorname{res}^X_{\mathrm{D}(f)}\colon \mathcal{O}_X(X) = A \to A_f = \mathcal{O}_X(\mathrm{D}(f))$ is the canonical map. For each $x \in X$, the stalk at x is the local ring

$$\mathcal{O}_{X,x} = \varinjlim_{\mathrm{D}(f) \ni x} \mathcal{O}_X(\mathrm{D}(f)) = \varinjlim_{f \notin \mathfrak{p}_x} A_f = A_{\mathfrak{p}_x}.$$

***Theorem* IX.1.2.4.** — *The presheaf $\mathcal{O}_X$ is a sheaf on $\mathcal{B}$, which extends to a sheaf on X by proposition VIII.2.1.12, also denoted by $\mathcal{O}_X$. Thus, $(X, \mathcal{O}_X) = (\operatorname{Spec} A, \mathcal{O}_{\operatorname{Spec} A})$ is a locally ringed space, which we will often abbreviate with* $\operatorname{Spec} A$.

Proof. — Let $\mathrm{D}(f) = \bigcup_{i \in I} \mathrm{D}(f_i)$ be a covering of principal open sets. Luckily, $\mathcal{O}_X$ has values in the category of rings and the intersection of principal open sets is principal open. Hence, it suffices to check these two sheaf properties:

(i) locality: if $s \in \mathcal{O}_X(\mathrm{D}(f))$ with $s|_{\mathrm{D}(f_i)} = 0$ for all $i \in I$, then $s = 0$;

(ii) gluing: if $s_i \in \mathcal{O}_X(\mathrm{D}(f_i))$ for $i \in I$ with $s_i|_{\mathrm{D}(f_i f_j)} = s_j|_{\mathrm{D}(f_i f_j)}$ for all $i, j \in I$, then there is some $s \in \mathcal{O}_X(\mathrm{D}(f))$ such that $s|_{\mathrm{D}(f_i)} = s$ for all $i \in I$.

Since $\mathrm{D}(f)$ is quasi-compact by proposition IX.1.1.12, we may show the properties for a finite subcover instead. Then (i) is also true in its generality. For (ii), we show this only for a finite subcover and we then obtain $s|_{\mathrm{D}(f_i)} = s$ for *all* $i \in I$ by (i). Thus, w. l. o. g., I is finite.

By restricting $\mathcal{O}_X$ to $\mathcal{O}_X|_{\mathrm{D}(f)}$, w. l. o. g. we may assume that $f = 1$ so that $A_f = A$ and $\mathrm{D}(f) = X$. Let $n \geq 1$. Since $\mathrm{D}(f_i) = \mathrm{D}(f_i^n)$, we have $X = \bigcup_{i \in I} \mathrm{D}(f_i^n)$ and hence $(f_i^n, i \in I) = A$. In other words, there are $b_i \in A$ such that

$$\sum_{i \in I} b_i f_i^n = 1. \tag{IX.1.2.4.1}$$

To show (i), let $s \in A$ such that $s/1 = 0$ in A_{f_i} for all i. Since I is finite, there exists $n \geq 1$ such that $f_i^n s = 0$ for all i. By equation (IX.1.2.4.1), we obtain $s = (\sum_{i \in I} b_i f_i^n) s = 0$.

To show (ii), let $s_i \in A_{f_i}$ with $s_i = s_j$ in $A_{f_i f_j}$ for all i, j. Since I is finite, there exists $n \geq 1$ and suitable $a_i \in A$ such that $s_i = a_i / f_i^n$ for all i. By assumption, there exists $m \geq 1$ such that $(f_i f_j)^m (f_j^n a_i - f_i^n a_j) = 0$ for all i, j. Passing to $s_i = (f_i^m a_i)/f_i^{m+n}$ for all i instead, we may replace $f_i^m a_i$ by a_i and $m+n$ by n to obtain $f_j^n a_i = f_i^n a_j$ for all i, j. Set $s \coloneqq \sum_{j \in I} b_j a_j \in A$ where the b_j are those in equation (IX.1.2.4.1). We obtain $f_i^n s = \sum_{j \in I} b_j f_i^n a_j = (\sum_{j \in I} b_j f_j^n) a_i = a_i$. Thus, $s/1 = a_i / f_i^n = s_i$ in A_{f_i} for all i, as desired. □

Definition* IX.1.2.5.** — *A locally ringed space $(X, \mathcal{O}_X)$ is called an* ***affine scheme *if there exists a ring A such that $(X, \mathcal{O}_X)$ is isomorphic to $(\operatorname{Spec} A, \mathcal{O}_{\operatorname{Spec} A})$. A* ***morphism of affine schemes*** *is a morphism of locally ringed spaces. We obtain the* ***category of affine schemes*** *Aff.*

***Theorem* IX.1.2.6.** — *If $\phi\colon A \to B$ is a ring map, define a morphism $\operatorname{Spec}\phi\colon \operatorname{Spec} B \to \operatorname{Spec} A$ of affine schemes by*

$$(\operatorname{Spec}\phi)^\flat_{\mathrm{D}(s)}\colon \mathcal{O}_{\operatorname{Spec} A}(\mathrm{D}(s)) = A_s \xrightarrow{\phi_s} B_{\phi(s)} = f_* \mathcal{O}_{\operatorname{Spec} B}(\mathrm{D}(s))$$

for all $s \in A$. Conversely, if $(f, f^\flat)\colon (X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ is a morphism of ringed spaces, define a ring map

$$\Gamma(f)\colon \Gamma(Y, \mathcal{O}_Y) = \mathcal{O}_Y(Y) \xrightarrow{f^\flat_Y} f_* \mathcal{O}_X(Y) = \Gamma(X, \mathcal{O}_X)$$

on global sections. These functors give an equivalence of categories

$$\mathsf{Ring}^{\mathrm{op}} \underset{\Gamma}{\overset{\operatorname{Spec}}{\rightleftarrows}} \mathsf{Aff}.$$

Proof. — We should note that $(\operatorname{Spec}\phi)^\flat$ for a ring map $\phi\colon A \to B$ is obviously compatible with restriction maps. Moreover, for every $x \in \operatorname{Spec} B$ the map

$$(\operatorname{Spec}\phi)^\sharp_x = \phi_{\phi^{-1}(\mathfrak{p}_x)}\colon \mathcal{O}_{\operatorname{Spec} A, (\operatorname{Spec}\phi)(x)} = A_{\phi^{-1}(\mathfrak{p}_x)} \longrightarrow B_{\mathfrak{p}_x} = \mathcal{O}_{\operatorname{Spec} B, x} \qquad \text{(IX.1.2.6.1)}$$

on stalks is obviously local, whence $\operatorname{Spec}\phi$ is indeed a morphism of locally ringed spaces.

The functor Spec is essentially surjective by definition. Hence, to show the equivalence it suffices to show that for any two rings A and B we have mutually inverse bijections

$$\operatorname{Hom}_{\mathsf{Ring}}(A, B) \underset{\Gamma}{\overset{\operatorname{Spec}}{\rightleftarrows}} \operatorname{Hom}_{\mathsf{Aff}}(\operatorname{Spec} B, \operatorname{Spec} A).$$

Let $\phi\colon A \to B$ be a ring map. Then $\Gamma(\operatorname{Spec}\phi) = (\operatorname{Spec}\phi)^\flat_{\mathrm{D}(1)} = \phi_1 = \phi$. Conversely, let $f\colon X \coloneqq \operatorname{Spec} B \to Y \coloneqq \operatorname{Spec} A$ be a morphism of affine schemes. For every point $x \in X$, the universal property of the stalk $\mathcal{O}_{Y, f(x)} = A_{\mathfrak{p}_{f(x)}}$ shows that $f^\sharp_x$ is the unique map such that the diagram

$$\begin{tikzcd} A \arrow[r, "\Gamma(f) = f^\sharp_X"] \arrow[d] & B \arrow[d] \\ A_{\mathfrak{p}_{f(x)}} \arrow[r, "f^\sharp_x"] & B_{\mathfrak{p}_x} \end{tikzcd} \qquad \text{(IX.1.2.6.2)}$$

commutes. By the universal property of the localisation $A_{\mathfrak{p}_{f(x)}}$, the image of $\Gamma(f)(A \setminus \mathfrak{p}_{f(x)})$ in $B_{\mathfrak{p}_x}$ is contained in $B^\times_{\mathfrak{p}_x} = B_{\mathfrak{p}_x} \setminus \mathfrak{p}_x B_{\mathfrak{p}_x}$. In other words, $\Gamma(f)(A \setminus \mathfrak{p}_{f(x)}) \subseteq B \setminus \mathfrak{p}_x$ and hence $\Gamma(f)^{-1}(\mathfrak{p}_x) \subseteq \mathfrak{p}_{f(x)}$. On the other hand, $f^\sharp_x$ is local, so $\Gamma(f)(\mathfrak{p}_{f(x)}) \subseteq \mathfrak{p}_x$ and hence $\Gamma(f)^{-1}(\mathfrak{p}_x) = \mathfrak{p}_{f(x)}$. It follows that $\operatorname{Spec}\Gamma(f) = f$ as continuous maps. Next, by equation (IX.1.2.6.1), $(\operatorname{Spec}\Gamma(f))^\sharp_x$ also makes diagram (IX.1.2.6.2) commute, so $(\operatorname{Spec}\Gamma(f))^\sharp_x = f^\sharp_x$ for all $x \in X$. We deduce $(\operatorname{Spec}\Gamma(f))^\sharp = f^\sharp$ by proposition VIII.2.3.5. Thus, $\operatorname{Spec}\Gamma(f) = f$ as morphisms of affine schemes. □

IX.1.3. Applications.

Definition* IX.1.3.1.** — *Let R be a ring. Then $\mathbb{A}^n_R := \operatorname{Spec} R[T_1, \dots, T_n]$ is called the* ***affine space *of relative dimension n over R.*

***Example* IX.1.3.2** (integral domains). — Let A be an integral domain, let $K = \operatorname{Quot} A$ and let $X = \operatorname{Spec} A$. The zero ideal of A is prime, so denote $\eta := (0) \in X$. Then $\overline{\{\eta\}} = X$ and hence η is a generic point. The local ring in η is $\mathcal{O}_{X,\eta} = A_{(0)} = K$.

We identify any localisation of A as a subring of K. Recall that $\mathcal{O}_X(\mathrm{D}(f)) = A_f$ for all $f \in A$. For all other open subsets $U \subseteq X$, we have $\mathcal{O}_X(U) = \varinjlim_{\mathrm{D}(f)\subseteq U} \mathcal{O}_X(\mathrm{D}(f)) = \bigcap_{\mathrm{D}(f)\subseteq U} A_f$ by definition.

On the other hand, we claim that $A_f = \bigcap_{\mathfrak{p}\in\mathrm{D}(f)} A_\mathfrak{p}$. Indeed: let $g \in \bigcap_{\mathfrak{p}\in\mathrm{D}(f)} A_\mathfrak{p}$ and $\mathfrak{a} = (A : g)$. For all $\mathfrak{p} \in \mathrm{D}(f)$, we have $\mathfrak{a} \not\subseteq \mathfrak{p}$ as $g \in A_\mathfrak{p}$. In other words, $f \in \mathfrak{p}$ for all prime ideals $\mathfrak{p} \supseteq \mathfrak{a}$, so $f \in \sqrt{\mathfrak{a}}$. Thus, $g \in A_f$, as desired. (Cf. the proof of proposition VIII.1.4.7, but we do not use Hilbert's Nullstellensatz VIII.1.1.17.)

As $A_{\mathfrak{p}_x} = \mathcal{O}_{X,x}$ for all $x \in X$, we obtain

$$\mathcal{O}_X(U) = \bigcap_{x\in U} \mathcal{O}_{X,x}$$

for all non-empty open subsets $U \subseteq X$.

***Construction* IX.1.3.3** (principal open subschemes of affine schemes). — Consider the affine scheme $\operatorname{Spec} A$. For any $f \in A$, let $j\colon \operatorname{Spec} A_f \to \operatorname{Spec} A$ be the morphism of affine schemes corresponding to the canonical map $A \to A_f$. Then j induces a homeomorphism $\operatorname{Spec} A_f \xrightarrow{\sim} \mathrm{D}(f)$ by proposition IX.1.1.20. Under this identification, we have $j^\flat_x\colon A_{\mathfrak{p}_x} \xrightarrow{\sim} (A_f)_{\mathfrak{p}_x}$ on stalks for all $x \in \mathrm{D}(f)$, so $j^\flat$ is an isomorphism by proposition VIII.2.3.5. Hence, $(j, j^\flat)$ induces an isomorphism $\operatorname{Spec} A_f \xrightarrow{\sim} (\mathrm{D}(f), \mathcal{O}_{\operatorname{Spec} A}|_{\mathrm{D}(f)})$.

***Construction* IX.1.3.4** (closed subschemes of affine schemes). — Consider the affine scheme $\operatorname{Spec} A$. For any ideal $\mathfrak{a}$ of A, let $i\colon \operatorname{Spec} A/\mathfrak{a} \to \operatorname{Spec} A$ be the morphism of affine schemes corresponding to the canonical map $A \twoheadrightarrow A/\mathfrak{a}$. Then i induces a homeomorphism $\operatorname{Spec} A/\mathfrak{a} \xrightarrow{\sim} \mathrm{V}(\mathfrak{a})$ by proposition IX.1.1.20 again. Under this identification, we have $i^\flat_x\colon A_{\mathfrak{p}_x} \to (A/\mathfrak{a})_{\bar{\mathfrak{p}}_x}$ on stalks for all $x \in \mathrm{V}(\mathfrak{a})$. Using the induced homeomorphism $i\colon \operatorname{Spec} A/\mathfrak{a} \xrightarrow{\sim} \mathrm{V}(\mathfrak{a})$, we endow $\mathrm{V}(\mathfrak{a})$ with the structure of a locally ringed space given by $\operatorname{Spec} A/\mathfrak{a}$.

Insert chapter where closed subscheme is defined.

We will later define the notion of a *closed subscheme* such that the notion coincides with $\mathrm{V}(\mathfrak{a})$.

***Example* IX.1.3.5.** — Let B be a ring and let $\mathfrak{b}$ be an ideal of B. Since a prime ideal of B contains $\mathfrak{b}$ iff it contains $\mathfrak{b}^n$ for some $n \geq 1$, the closed subset $\mathrm{V}(\mathfrak{b}^n) \subseteq \operatorname{Spec} B$ is independent of n. But as an affine scheme, $\mathrm{V}(\mathfrak{b}^n) = \operatorname{Spec} B/\mathfrak{b}^n$ depends on n.

We explain this by the example $B = k[T]$ and $\mathfrak{b} = (T)$ for k an algebraically closed field. Then the closed points (i.e., maximal ideals) of $\mathbb{A}^1_k = \operatorname{Spec} k[T]$ correspond to elements of k with $\mathfrak{b}$ corresponding to 0. Set $A = k[T]/(T^n)$ and $X = \operatorname{Spec} A$. Then X is a single point x, and we have $\mathcal{O}_X(X) = \mathcal{O}_{X,x} = A$, $\mathfrak{m}_x = (\bar{T})$ (which is non-zero if $n > 1$) and $\kappa(x) = k$. We think of X as a 'closed subscheme' of $\mathbb{A}^1_k$ which is the single point (T) or, under the correspondence with k, 'concentrated in 0'.

We identify $\mathbb{A}^1_k$ with the affine variety $\mathbb{A}^1(k)$ by virtue of example IX.1.1.23, so that X is a closed subvariety (point in 0). The affine coordinate ring is $\Gamma(\mathbb{A}^1(k)) = k[T] = B$. The restriction of functions on $\mathbb{A}^1(k)$ to X is given by $k[T] \twoheadrightarrow \Gamma(X) = k[T]/(T^n) = A$. If $n = 1$, then this restriction is $\mathrm{ev}_0\colon k[T] \to k$, $f \mapsto f(0)$, but for $n > 1$, we keep the higher order terms of the 'Taylor expansion' of f in 0. Through this picture, we should imagine X as a slightly fuzzy or fat point within $\mathbb{A}^1_k$ whose infinitesimal 'radius' is divided into $n - 1$ segments.

***Example* IX.1.3.6.** — Here another picture of this kind. Consider the affine space $\mathbb{A}^2_k = \operatorname{Spec} k[T,U]$ for k an affine algebraic field, which we visualise as the plane k^2, corresponding to the maximal ideals of $k[T,U]$. The ideals $\mathfrak{a}_1 := (U)$ and $\mathfrak{a}_2 := (U - T^n)$ define closed subvarieties $X_i = \operatorname{Spec} k[T,U]/\mathfrak{a}_i$, which are $X_1 = \{(u,t) \in k^2 \mid u = 0\}$ (constant line) and $X_2 = \{(u,t) \in k^2 \mid u = t^n\}$ (nth power curve). We have $X_1 \cap X_2 = \{(0,0)\}$ as a set, but it should make a difference that X_1 and X_2 intersect tangentially if $n > 1$.

We will later define $X_1 \cap X_2 = \operatorname{Spec} k[T,U]/(\mathfrak{a}_1 + \mathfrak{a}_2)$ as affine schemes. Under the identification $k[T,U]/\mathfrak{a}_1 \cong k[T]$, we have $k[T,U]/(\mathfrak{a}_1 + \mathfrak{a}_2) \cong k[T]/(T^n)$. So, from the point of view of affine schemes, we retain more information in the intersection $X_1 \cap X_2$.

Insert chapter where intersections are defined.

§ IX.2. SCHEMES

IX.2.1. Definitions.

Like prevarieties are obtained by gluing affine varieties, schemes are obtained by gluing affine schemes.

***Definition* IX.2.1.1.** — *A **scheme**[(IX.1)] is a locally ringed space $(X, \mathcal{O}_X)$ which admits an open covering $X = \bigcup_{i\in I} U_i$ such that the locally ringed spaces $(U_i, \mathcal{O}_X|_{U_i})$ are affine schemes. Such a covering is called an **affine open covering** of X. We will usually denote $(X, \mathcal{O}_X)$ simply by X. A **morphism of schemes** is a morphism of locally ringed spaces. Thus, we obtain the **category of schemes** denoted by* Sch. *Clearly, affine schemes are schemes.*

***Definition* IX.2.1.2.** — *Fix a scheme S. An **S-scheme** or a **scheme over** S is a scheme X together with a morphism $X \to S$ of schemes. This morphism $X \to S$ is called the **structural morphism** of X and often dropped from notation, and S is called the **base scheme** of X.*

The morphism $X \to Y$ of S-schemes is a morphism $X \to Y$ of schemes such that the following diagram commutes:

$$\begin{tikzcd} X \arrow[rr] \arrow[dr] & & Y \arrow[dl] \\ & S. & \end{tikzcd}$$

Thus, we obtain the category Sch/S *of schemes over S.*

*If the base scheme $S = \operatorname{Spec} R$ is affine, then we say **R-schemes** or **schemes over** R.*

IX.2.1.3. — Recall from construction IX.1.3.3 that if $X = \operatorname{Spec} A$ is an affine scheme and if $f \in A$, then $(\mathrm{D}(f), \mathcal{O}_X|_{\mathrm{D}(f)}) \cong \operatorname{Spec} A_f$ is a *principal open* affine scheme.

***Proposition* IX.2.1.4.** — *Let X be a scheme.*

(i) *Let $U \subseteq X$ be an open subset. Then the locally ringed space $(U, \mathcal{O}_X|_U)$ is a scheme called an **open subscheme** of X. If U is affine, then it is called an **affine open subscheme**.*

(ii) *The underlying sets of all affine open subschemes of X form a basis of the topology of X.*

Proof. — The scheme X can be covered by open affine subschemes by definition, and each of these affine schemes has a basis of its topology consisting of principal open affine subschemes by proposition IX.1.1.12 and construction IX.1.3.3. Hence, the underlying principal open sets form a basis of X, and U can be covered by principal open affine subschemes. ∎

(IX.1) Originally, Grothendieck coined the terms *scheme* and *prescheme*, which we nowadays call *separated schemes* and *schemes*, resp.

***Construction* IX.2.1.5.** — Let X be a scheme, and let U be an open subscheme of X. Let $j\colon U \hookrightarrow X$ be the inclusion as topological spaces. The restriction maps of the structure sheaf $\mathcal{O}_X$ induce ring maps

$$j^\flat_V\colon \Gamma(V,\mathcal{O}_X) \longrightarrow \Gamma(V\cap U,\mathcal{O}_X) = \Gamma(j^{-1}(V),\mathcal{O}_X|_U) = \Gamma(V, j_*\mathcal{O}_X|_U)$$

for all $V \subseteq X$ open. Hence, if we speak about the inclusion morphism of an open subscheme $U \to X$, then we mean $(j, j^\flat)\colon U \to X$.

***Lemma* IX.2.1.6.** — *Let X be a scheme, and let U and V be affine open subschemes of X. Then for all $x \in U \cap V$, there exists an open subscheme W of $U \cap V$ with $W \ni x$ such that W is principal open in U and in V.*

Proof. — W. l. o. g., by replacing V with a principal open subset of U containing x, we may assume that $V \subseteq U$. Choose $f \in \Gamma(U,\mathcal{O}_X)$ such that $x \in \mathrm{D}(f) \subseteq V$, and set $W := \mathrm{D}(f)$. Then $W = \mathrm{D}_U(f) = \mathrm{D}_V(f|_V)$ (since $W \subseteq V \subseteq U$) is principal open in U and in V. (By definition of affine schemes, we even have $\Gamma(U,\mathcal{O}_X)_f \cong \Gamma(W,\mathcal{O}_X) \cong \Gamma(V,\mathcal{O}_X)_{f|_V}$.)

IX.2.2. Gluing schemes.

***Proposition* IX.2.2.1** (gluing of scheme morphisms). — *Let X and Y be locally ringed spaces. Then assigning to open subsets $U \subseteq X$ the set of morphisms $\mathrm{Hom}_{\mathsf{LRS}}((U,\mathcal{O}_X|_U), Y)$ of locally ringed spaces defines a sheaf of sets on X.*

In other words, if $X = \bigcup_{i\in I} U_i$ is an open covering, then a family $(U_i \to Y)_{i\in I}$ of morphisms such that any two morphisms agree on intersections $U_i \cap U_j$ glues to a unique morphism $X \to Y$.

Proof. — We know that the continuous maps on topological spaces can be glued. Moreover, the morphisms $(f^{-1}\mathcal{O}_Y)|_{U_i} \to \mathcal{O}_X|_{U_i}$ of sheaves also glue by proposition VIII.2.2.1.

***Definition* IX.2.2.2.** — *A **gluing datum** of schemes consists of the following data:*

- *a family $(U_i)_{i\in I}$ of schemes;*
- *an open subscheme $U_{ij} \subseteq U_i$ for all $i, j \in I$;*
- *an isomorphism $\phi_{ij}\colon U_{ij} \xrightarrow{\sim} U_{ji}$ of schemes for all $i, j \in I$;*

such that

(i) *$U_{ii} = U_i$ for all $i \in I$ (for simplicity);*
(ii) *$\phi_{ij}(U_{ij} \cap U_{ik}) \subseteq U_{jk}$ for all $i, j, k \in I$;*
(iii) *the **cocycle condition** holds: the diagram*

$$\begin{tikzcd} U_{ij}\cap U_{ik} \arrow[rr, "\phi_{ik}"] \arrow[dr, "\phi_{ij}"'] & & U_{ki}\cap U_{kj} \\ & U_{ji}\cap U_{jk} \arrow[ur, "\phi_{jk}"'] & \end{tikzcd}$$

commutes for all $i, j, k \in I$.

In particular, for $i = j = k$ the cocycle condition implies $\phi_{ii} = \phi_{ii} \circ \phi_{ii}$ and hence $\phi_{ii} = \mathrm{id}_{U_i}$. Moreover, for $i = k$ the cocycle condition implies $\phi_{ji}^{-1} = \phi_{ij}$ and hence ϕ_{ij} induces an isomorphism $U_{ij} \cap U_{ik} \xrightarrow{\sim} U_{ji} \cap U_{jk}$.

***Proposition* IX.2.2.3** (gluing of schemes). — *Let (U_i, U_{ij}, ϕ_{ij}) be a gluing datum of schemes. Then there exists a scheme X together with morphisms $\psi_i\colon U_i \to X$ such that:*

- *ψ_i yields isomorphisms onto open subschemes of X for all $i \in I$;*

– $\psi_j \circ \phi_{ij} = \psi_i$ *on* U_{ij} *for all* $i, j \in I$;
– $X = \bigcup_{i \in I} \psi_i(U_i)$;
– $\psi_i(U_i) \cap \psi_j(U_j) = \psi_i(U_{ij}) = \psi_j(U_{ji})$ *for all* $i, j \in I$.

We say that X *is obtained by* ***gluing the*** U_i *via* ϕ_{ij}.

Proof. — To define the topological space X, set

$$X := \coprod_{i \in I} U_i / \sim$$

as a set. Here, for $x_i \in U_i$ and $x_j \in U_j$ we define $x_i \sim x_j$ iff $x_i \in U_{ij}$, $x_j \in U_{ji}$ and $x_j = \phi_{ij}(x_i)$. This is an equivalence relation: reflexivity and symmetry follow from $\phi_{ii} = \mathrm{id}_{U_i}$ and $\phi_{ji} = \phi_{ij}^{-1}$. For transitivity, suppose that $x_i \in U_i$, $x_j \in U_j$, $x_k \in U_k$ and $x_i \sim x_j$, $x_j \sim x_k$. Then $x_j \in U_{ji} \cap U_{jk}$, so $x_i = \phi_{ij}^{-1}(x_j) \in U_{ik}$ and $x_k = \phi_{jk}(x_j) \in U_{ki}$ since ϕ_{ij} and ϕ_{jk} are isomorphisms $U_{ij} \cap U_{ik} \xrightarrow{\sim} U_{ji} \cap U_{jk} \xrightarrow{\sim} U_{ki} \cap U_{kj}$. Finally, $x_k = \phi_{ik}(x_i)$ by the cocycle condition, whence $x_i \sim x_k$.

The set X comes with natural injections $\psi_i \colon U_i \hookrightarrow X$ satisfying $\psi_i(U_{ij}) = \psi_i(U_i) \cap \psi_j(U_j)$ and $\psi_j \circ \phi_{ij} = \psi_i$ for all $i, j \in I$ by definition of $\sim$. Endow X with the quotient topology, i. e., a subset $U \subseteq X$ is open iff $\psi_i^{-1}(U)$ is open in U_i for all i. In particular, the ψ_i are continuous and open, since for all $V \subseteq U_i$ open we have $\psi_j^{-1}(\psi_i(V)) = \phi_{ij}(V \cap U_{ij})$ for all $j \in I$, so $\psi_i(V)$ is open in X. Thus, the ψ_i are homeomorphisms onto open subsets of X.

To turn X into a locally ringed space, we need to glue the structure sheaves on U_i to a sheaf $\mathcal{O}_X$ on X. The isomorphisms ϕ_{ij} induce isomorphisms $\psi_{i,*}\phi_{ij} \colon (\psi_{i,*}\mathcal{O}_{U_i})|_{\psi_i(U_{ij})} \xrightarrow{\sim} (\psi_{i,*}\mathcal{O}_{U_j})|_{\psi_j(U_{ji})}$ of sheaves on $\psi_i(U_i) \cap \psi_j(U_j)$. Like the ϕ_{ij}, these satisfy the cocycle condition. By proposition VIII.2.2.3, the sheaves $\psi_{i,*}\mathcal{O}_{U_i}$ glue to a sheaf $\mathcal{O}_X$ on X with isomorphisms $\psi_{i,*}\mathcal{O}_{U_i} \xrightarrow{\sim} \mathcal{O}_X|_{\psi_i(U_i)}$ for all $i \in I$ which are compatible with the gluing morphisms $\psi_{i,*}\phi_{ij}$.

We obtain a locally ringed space $(X, \mathcal{O}_X)$. Moreover, the construction of $\mathcal{O}_X$ also turns $\psi_i \colon U_i \xrightarrow{\sim} (\psi_i(U_i), \mathcal{O}_X|_{\psi_i(U_i)}) \hookrightarrow X$ into a morphism of locally ringed spaces for all $i \in I$. Since X is covered by schemes U_i, each of which admits an affine open covering, X is a scheme as well. ∎

Corollary **IX.2.2.4** (universal property of gluing schemes)**.** — *Let* (U_i, U_{ij}, ϕ_{ij}) *be a gluing datum of schemes, and let* X *be the scheme with morphisms* $\psi_i \colon U_i \to X$ *obtained by gluing the* U_i *via* ϕ_{ij}. *If* T *is a scheme and if* $\xi_i \colon U_i \to T$ *for* $i \in I$ *are morphisms that induce isomorphisms onto open subschemes of* T, *such that* $\xi_j \circ \phi_{ij} = \xi_i$ *on* U_{ij} *for all* $i, j \in I$, *then there exists a unique morphism* $\xi \colon X \to T$ *with* $\xi \circ \psi_i = \xi_i$ *for all* $i \in I$. *In particular,* X *(together with* $(\psi_i)_{i \in I}$*) is unique up to unique isomorphism.*

Proof. — This follows by gluing the morphisms $\xi_i \circ \psi_i^{-1}$ on $\psi_i(U_i) \subseteq X$ (proposition IX.2.2.1). ∎

Remark **IX.2.2.5.** — The notion of gluing data can be defined for topological spaces or (locally) ringed spaces, too. Gluing this data yields a new object in the respective category, which satisfies a universal property as above.

Definition **IX.2.2.6.** — *The* ***disjoint union*** *of schemes (or locally ringed spaces)* $(U_i)_{i \in I}$ *is given by the topological space* $\coprod_{i \in I} U_i$ *with the 'obvious' structure sheaf* $U \mapsto \prod_{i \in I} \mathcal{O}_{U_i}(U \cap U_i)$.

We can obtain the disjoint union from proposition IX.2.2.3 as well by setting $U_{ij} = \varnothing$ *for all* $i, j \in I$.

Example **IX.2.2.7.** — If $X_i = \operatorname{Spec} A_i$ for $i = 1, \dots, n$ are affine schemes, then $\coprod_{i=1}^n X_i = \operatorname{Spec} \prod_{i=1}^n A_i =: X$ is again affine. Indeed, this is true on the level of topological spaces, see proposition I.1.3.8. Moreover, for all i and all $f \in A_i$, we have $\mathcal{O}_{\coprod_{i=1}^n X_i}(\mathrm{D}_{X_i}(f)) = \mathcal{O}_{X_i}(\mathrm{D}(f)) = \mathcal{O}_X(\mathrm{D}_X(\hat{f}))$ where $\hat{f} = (0, \dots, 0, f, 0, \dots, 0) \in \prod_{j=1}^n A_j$. Since the principal open subsets $\mathrm{D}_{X_i}(f)$ form a basis of X, we have the claimed equality on the level of sheaves.

***Warning* IX.2.2.8.** — For an infinite family $(X_i)_{i\in I}$ of affine schemes $X_i = \operatorname{Spec} A_i$, the disjoint union $X = \coprod_{i\in I} X_i$ is not affine anymore. The reason is that X is not quasi-compact, but affine schemes are quasi-compact by proposition IX.1.1.12.

***Example* IX.2.2.9.** — In the special case of two schemes U_1 and U_2, open subschemes $U_{12} \subseteq U_1$ and $U_{21} \subseteq U_2$ with an isomorphism $\phi\colon U_{12} \xrightarrow{\sim} U_{21}$ yield a gluing datum. Let X be the scheme obtained by gluing U_1 and U_2 along ϕ, and we consider U_1 and U_2 as open subschemes of X. The structure sheaf is then given by

$$\Gamma(V, \mathcal{O}_X) = \{(s_1, s_2) \in \Gamma(V \cap U_1, \mathcal{O}_{U_1}) \times \Gamma(V \cap U_2, \mathcal{O}_{U_2}) \mid \phi^\flat(s_2|_{U_{21}\cap V}) = s_1|_{U_{12}\cap V}\}$$

for all $V \subseteq X$ open, cf. proposition VIII.2.2.3.

***Example* IX.2.2.10** (affine line with two origins)**.** — As an example for example IX.2.2.9, let $U_1 = U_2 = \mathbb{A}^1_k = \operatorname{Spec} k[T]$ for k a field. Fix a closed point $x \in \mathbb{A}^1_k$, e.g., $x = (T)$, and set $U_{12} = U_1 \setminus \{x\}$ and $U_{21} = U_2 \setminus \{x\}$. The gluing isomorphism $\phi\colon U_{12} \xrightarrow{\sim} U_{21}$ is the identity. We should think of the scheme X obtained by gluing as an affine line with the point x doubled.

[GW20] mentions Exer. 3.26 for showing that X is not affine.

IX.2.3. Morphisms into affine schemes, morphisms from field spectra.

Morphisms from schemes to an affine scheme are easy to understand. We will use the following lemma shortly.

***Lemma* IX.2.3.1.** — *Let $(X, \mathcal{O}_X)$ be a locally ringed space. For any $f \in \mathcal{O}_X(X)$, the subset* $\mathrm{D}(f) := \{x \in X \mid f(x) \neq 0\}$ *is open in X. Moreover, $f|_{\mathrm{D}(f)}$ is a unit in $\mathcal{O}_X(\mathrm{D}(f))$.*

Proof. — Note that $\mathrm{D}(f) = \{x \in X \mid f_x \in \mathcal{O}_{X,x}^\times\}$ by definition. So, for $x \in X$ there is some inverse $g_x \in \mathcal{O}_{X,x}$ of f_x. Hence, there exists a representative $g \in \mathcal{O}_X(U)$ of g_x for a suitable open subset $U \subseteq X$ such that $f|_U g = 1$ in $\mathcal{O}_X(U)$. It follows that $f_y g_y = 1$ in $\mathcal{O}_{X,y}$ for all $y \in U$ and therefore, $U \subseteq \mathrm{D}(f)$ is an open neighbourhood of x. Thus, $\mathrm{D}(f)$ is open.

Next, suppose that there is an open cover $\mathrm{D}(f) = \bigcup_i U_i$ such that $f|_{U_i}$ has inverse $g_i \in \mathcal{O}_X(U_i)$. Since inverses are unique, we obtain $g_i|_{U_i\cap U_j} = g_j|_{U_i\cap U_j}$ for all i, j, whence we can glue the g_i to a unique $g \in \mathcal{O}_X(\mathrm{D}(f))$. Similarly, the sections $f|_{U_i} g_i = 1$ must glue to $f|_{\mathrm{D}(f)} g = 1$. ∎

***Proposition* IX.2.3.2** (global sections-Spec adjunction)**.** — *Let $(X, \mathcal{O}_X)$ be a locally ringed space and let $Y = \operatorname{Spec} A$ be an affine scheme. Then we have a canonical bijection*

$$\operatorname{Hom}_{\mathsf{LRS}}(X, Y) \xrightarrow{\sim} \operatorname{Hom}_{\mathsf{Ring}}(A, \Gamma(X, \mathcal{O}_X)), \qquad (f, f^\flat) \longmapsto f^\flat_Y$$

functorial in X and A. In other words, we have the adjunction $\Gamma \dashv \operatorname{Spec}$ (note that $\operatorname{Spec}\colon \mathsf{Ring}^{\mathrm{op}} \to \mathsf{Aff}$ and Γ are contravariant).

Easy proof if X is a scheme. — There is an easier proof if we assume that X is a scheme. By definition, there is an affine open covering $X = \bigcup_i U_i$. We know from theorem IX.1.2.6 that

$$\operatorname{Hom}_{\mathsf{Aff}}(U_i, Y) \xrightarrow{\sim} \operatorname{Hom}_{\mathsf{Ring}}(A, \Gamma(U_i, \mathcal{O}_X))$$

is a bijection for all i. By functoriality of Γ, we have the following commutative diagram for all affine open subschemes V of $U_i \cap U_j$:

$$\begin{tikzcd}
\operatorname{Hom}_{\mathsf{Aff}}(U_i, Y) \arrow[r, "\sim"] \arrow[d] & \operatorname{Hom}_{\mathsf{Ring}}(A, \Gamma(U_i, \mathcal{O}_X)) \arrow[d] \\
\operatorname{Hom}_{\mathsf{Aff}}(V, Y) \arrow[r, "\sim"] & \operatorname{Hom}_{\mathsf{Ring}}(A, \Gamma(V, \mathcal{O}_X)).
\end{tikzcd}$$

By gluing the sheaves $U \mapsto \operatorname{Hom}_{\mathsf{Aff}}(U_i, Y)$, see proposition IX.2.2.1, we obtain the statement. ∎

General proof. — For the general case, we first construct an inverse map $\mathrm{Hom}_{\mathsf{Ring}}(A, \Gamma(X, \mathcal{O}_X)) \to \mathrm{Hom}_{\mathsf{LRS}}(X, Y)$. Given a ring map $\phi\colon A \to \Gamma(X, \mathcal{O}_X)$, we need to construct a morphism $(f, f^\flat)\colon X \to Y$. We first define the map $f\colon X \to Y$ of topological spaces. For any $x \in X$, let $\mathfrak{p} := \{a \in A \mid \phi(a)(x) = 0 \text{ in } \kappa(x)\}$. Then $A/\mathfrak{p}$ is a subring of the field $\kappa(x)$, so $\mathfrak{p}$ is a prime ideal of A. Set $f(x) := \mathfrak{p}$. Furthermore, for any $s \in A$ we have $f^{-1}(\mathrm{D}(s)) = \{x \in X \mid \phi(s)(x) \neq 0\} =: \mathrm{D}(\phi(s))$. By lemma IX.2.3.1, $\mathrm{D}(\phi(s))$ is open in X, whence f is continuous, and $\phi(s)|_{\mathrm{D}(\phi(s))}$ is a unit in $\Gamma(\mathrm{D}(\phi(s)), \mathcal{O}_X)$.

We next define the morphism $f^\flat\colon \mathcal{O}_Y \to f_*\mathcal{O}_X$ of sheaves agreeing with ϕ on global sections. For any $s \in A$, we hence define $f^\flat_{\mathrm{D}(s)}$ as the unique map such that the diagram

$$\begin{tikzcd}
\Gamma(Y, \mathcal{O}_Y) = A \arrow[r] \arrow[d, "\phi"'] & \Gamma(\mathrm{D}(s), \mathcal{O}_Y) = A_s \arrow[d, dashed, "f^\flat_{\mathrm{D}(s)}"] \\
\Gamma(X, \mathcal{O}_X) \arrow[r] & \Gamma(\mathrm{D}(\phi(s)), \mathcal{O}_X)
\end{tikzcd}$$

commutes by the universal of localisation, since $s \in A$ is mapped to $\phi(s)|_{\mathrm{D}(\phi(s))} \in \Gamma(\mathrm{D}(\phi(s)), \mathcal{O}_X)^\times$. Furthermore, for any $x \in X$ we have $f^\sharp_x\colon \mathcal{O}_{Y,f(x)} = A_\mathfrak{p} \to \mathcal{O}_{X,x}$ on stalks where $\mathfrak{p} = f(x)$. By definition of f, we have $\phi(a) \in \mathfrak{m}_x$ for all $a \in \mathfrak{p}$ where $\mathfrak{m}_x$ is the maximal ideal of $\mathcal{O}_{X,x}$. Hence, $\phi(\mathfrak{p}A_\mathfrak{p}) \subseteq \mathfrak{m}_x$ and $f^\sharp_x$ is local for all $x \in X$. Thus, $(f, f^\flat)$ is a morphism of locally ringed spaces.

It remains to show that the two constructions are mutually inverse. If $\phi \in \mathrm{Hom}_{\mathsf{Ring}}(A, \Gamma(X, \mathcal{O}_X))$ and if we construct the associated $(f, f^\flat) \in \mathrm{Hom}_{\mathsf{LRS}}(X, Y)$, then $f^\flat_Y = \phi$ by definition. Conversely, let $f \in \mathrm{Hom}_{\mathsf{LRS}}(X, Y)$, let $\phi = f^\flat_Y$, and let $g\colon X \to Y$ be the morphisms associated to ϕ. Since $f^\sharp_x$ is local, we have a map

$$\Gamma(Y, \mathcal{O}_Y) = A \longrightarrow \mathcal{O}_{Y,f(x)} \twoheadrightarrow \kappa(f(x)) \xhookrightarrow{\bar{f}^\sharp_x} \kappa(x)$$

for all $x \in X$. Note that $f(x) \in Y$, a prime ideal, is the kernel of $A \to \kappa(f(x))$, Hence, $f(x)$ agrees with the definition of $g(x)$ above, and we have $f = g$ as continuous maps. Moreover, for any $x \in X$ both $f^\flat$ and $g^\flat$ induce $A_\mathfrak{p} \to \mathcal{O}_{X,x}$ on stalks where $\mathfrak{p} = f(x) = g(x)$. Since both $f^\sharp_x$ and $g^\sharp_x$ satisfy the commutative diagram

$$\begin{tikzcd}
A \arrow[r] \arrow[d, "\phi"'] & A_\mathfrak{p} \arrow[d, dashed] \\
\Gamma(X, \mathcal{O}_X) \arrow[r] & \mathcal{O}_{X,x},
\end{tikzcd}$$

we must have $f^\sharp_x = g^\sharp_x$ by the universal property of localisation (since $\mathfrak{p} = f(x) = g(x)$, we have $\phi(A \setminus \mathfrak{p}) \subseteq \mathcal{O}_{X,x}^\times$). Thus, we have $f^\sharp = g^\sharp$ by proposition VIII.2.3.5.

***Corollary* IX.2.3.3.** — *If X is a locally ringed space, then there exists a unique morphism $X \to \operatorname{Spec}\mathbb{Z}$ of locally ringed spaces. In particular, $\operatorname{Spec}\mathbb{Z}$ is the final object in* Sch.

Proof. — This follows from proposition IX.2.3.2 and the fact that $\mathbb{Z}$ is an initial object in Ring.

***Observation* IX.2.3.4.** — Proposition IX.2.3.2 also says that Γ is a representable functor with representation $\operatorname{Spec}\mathbb{Z}[T]$:

$$\mathrm{Hom}_{\mathsf{LRS}}(X, \operatorname{Spec}\mathbb{Z}[T]) = \mathrm{Hom}_{\mathsf{Ring}}(\mathbb{Z}[T], \Gamma(X, \mathcal{O}_X)) = \Gamma(X, \mathcal{O}_X).$$

In particular, for each R-scheme X we have

$$\mathrm{Hom}_{\mathsf{Sch}/R}(X, \mathbb{A}^1_R) = \mathrm{Hom}_{R\text{-}\mathsf{Alg}}(R[T], \Gamma(X, \mathcal{O}_X)) = \Gamma(X, \mathcal{O}_X).$$

Thus, global sections of $\mathcal{O}_X$ can be identified with morphisms from X to the affine line $\mathbb{A}^1_R$.

Maybe explain more category theory.

***Construction* IX.2.3.5.** — Proposition IX.2.3.2 also gives the unit $c\colon \mathrm{id}_{\mathsf{LRS}} \to \operatorname{Spec} \circ \Gamma$ of the adjunction. For every locally ringed space X, we call the morphism

$$c_X\colon X \longrightarrow \operatorname{Spec} \Gamma(X, \mathcal{O}_X)$$

of locally ringed spaces **canonical**.

***Construction* IX.2.3.6.** — Let X be a scheme. Let $x \in X$ and let $U \subseteq X$ be an affine open neighbourhood of x, say, $U = \operatorname{Spec} A$. Let $\mathfrak{p}$ be the prime ideal of A corresponding to x. Then $\mathcal{O}_{X,x} = \mathcal{O}_{U,x} = A_{\mathfrak{p}}$ and the canonical map $A \to A_{\mathfrak{p}}$ induces the morphism

$$j_x\colon\ \operatorname{Spec} \mathcal{O}_{X,x} = \operatorname{Spec} A_{\mathfrak{p}} \hookrightarrow \operatorname{Spec} A = U \hookrightarrow X$$

of schemes. This is independent of the choice of U: for arbitrary $f \in A \setminus \mathfrak{p}$, the map $A \to A_{\mathfrak{p}}$ factors through A_f, so $\operatorname{Spec} \mathcal{O}_{X,x} \to \operatorname{Spec} A$ factors through $\operatorname{Spec} A_f$, which is homeomorphic to $\mathrm{D}(f) \subseteq U$ by proposition IX.1.1.20. Hence, $\operatorname{Spec} \mathcal{O}_{X,x} \to \operatorname{Spec} A$ agrees with $\operatorname{Spec} \mathcal{O}_{X,x} \to \operatorname{Spec} A_f$. As a consequence, if j_x and j'_x are defined with affine open neighbourhoods U and U', resp., then they agree with j''_x defined with an affine open neighbourhood $U'' \subseteq U \cap U'$ of x (see proposition IX.2.1.4). We remark that j_x induces a homeomorphism onto the subspace $Z = \{\mathfrak{p}' \in U \mid \mathfrak{p} \in \mathrm{V}(\mathfrak{p}')\} = \bigcap_{V \ni x} V$, where V runs through all open subsets of X, by proposition IX.1.1.20.

Hence, the canonical map $\mathcal{O}_{X,x} \to \kappa(x)$ induces a morphism

$$i_x\colon\ \operatorname{Spec} \kappa(x) \longrightarrow \operatorname{Spec} \mathcal{O}_{X,x} \xrightarrow{j_x} X$$

called **canonical**. By definition of j_x, we have $i_x((0)) = x$ where (0) is the unique point in $\kappa(x)$.

***Proposition* IX.2.3.7.** — *Let $\operatorname{Spec} K = \{p\}$ and X be schemes for K a field. There is a bijection*

$$\begin{aligned}\operatorname{Hom}_{\mathsf{Sch}}(\operatorname{Spec} K, X) &\xrightarrow{\sim} \{(x, \iota) \mid x \in X, \iota\colon \kappa(x) \to \kappa(p) = K\},\\ f &\longmapsto (f(p), \bar{f}^\sharp_p)\end{aligned}$$

where the map $\bar{f}^\sharp_p\colon \kappa(f(p)) \to \kappa(p)$ on residue fields is induced by the local map $f^\sharp_p\colon \mathcal{O}_{X,f(p)} \to \mathcal{O}_{\operatorname{Spec} K,p} = K$ on stalks. The same is more generally true for local rings (R, p):

$$\begin{aligned}\operatorname{Hom}_{\mathsf{Sch}}(\operatorname{Spec} R, X) &\xrightarrow{\sim} \{(x, \phi) \mid x \in X, \textit{local ring map } \phi\colon \mathcal{O}_{X,x} \to R_{\mathfrak{p}_p} = R\},\\ f &\longmapsto (f(p), f^\sharp_p).\end{aligned}$$

Proof. — Note that a morphism $f\colon \operatorname{Spec} K \to X$ is the same as

$$\operatorname{Spec} K \xrightarrow{\operatorname{Spec} \bar{f}^\sharp_p} \operatorname{Spec} \kappa(x) \xrightarrow{i_x} X$$

where $x = f(p)$. For the converse map, we map $(x, \iota\colon \kappa(x) \to K)$ to the morphism

$$\operatorname{Spec} K \xrightarrow{\operatorname{Spec} \iota} \operatorname{Spec} \kappa(x) \xrightarrow{i_x} X.$$

These two maps are obviously mutually inverse.

The general case is proven similarly. Suppose that $f\colon \operatorname{Spec} R \to X$ is given. Let $U \subseteq X$ be an affine open neighbourhood of $x = f(p)$. Then the image of f is contained in U because $f^{-1}(U)$ is an open neighbourhood of p in $\operatorname{Spec} R$ and because p is the *unique* closed point in $\operatorname{Spec} R$. Hence, f factors through $\operatorname{Spec} R \to U$. Furthermore, this $\operatorname{Spec} R \to U$ factors as by the equivalence between affine schemes and rings, see . Under the equivalence $\mathsf{Ring}^{\mathrm{op}} \rightleftarrows \mathsf{Aff}$ (theorem IX.1.2.6), this $\operatorname{Spec} R \to U$ corresponds to the ring map $A := \Gamma(U, \mathcal{O}_U) \to R$. If the prime ideal $\mathfrak{p}$ of A denotes the contraction of p in R, where $\mathfrak{p}$ also corresponds to $x \in U$, this ring map factors as $A \to A_{\mathfrak{p}} \to R$ where $A_{\mathfrak{p}} \to R$ is a local map of local rings. Passing to affine schemes gives the factorisation

$$\operatorname{Spec} R \xrightarrow{\operatorname{Spec} f_p^\sharp} \operatorname{Spec} \mathcal{O}_{X,x} \longrightarrow U$$

of $\operatorname{Spec} R \to U$. Thus, f factors as

$$\operatorname{Spec} R \xrightarrow{\operatorname{Spec} f_p^\sharp} \operatorname{Spec} \mathcal{O}_{X,x} \xrightarrow{j_x} X.$$

The inverse map is constructed as in the case of fields K.

IX.2.4. Basic properties of schemes.

IX.2.4.1. — A topological space is *quasi-compact* if every open covering admits a finite subcovering. A topological space is *irreducible* if it is non-empty and if it is not the union of two proper closed subsets.

***Definition* IX.2.4.2.** — *A scheme is called **connected**, **quasi-compact** or **irreducible**, resp., if the underlying topological space has this property.*

***Example* IX.2.4.3.** — For example, affine schemes are quasi-compact by proposition IX.1.1.12. On the other hand, infinite disjoint unions of non-empty schemes are not quasi-compact. This example is rather trivial; there are connected and even irreducible schemes which are not quasi-compact.

[GW20, p. 76] references Exer. 3.2 and 3.26.

***Definition* IX.2.4.4.** — *A morphism $f\colon X \to Y$ of schemes is called **injective**, **surjective** or **bijective**, resp., if the continuous map $X \to Y$ of the underlying topological spaces has this property. We similarly call f **open**, **closed**, or a **homeomorphism**, resp., if the continuous map $X \to Y$ has this property.*

*A morphism $f\colon X \to Y$ of schemes is called **dominant** if the image $f(X)$ is dense in Y.*

***Warning* IX.2.4.5.** — Homeomorphisms are not isomorphisms in general.

See [GW20, Exer. 3.6 and 12.21].

We will now generalise the notion of noetherian rings to schemes, which is an important finiteness condition.

***Definition* IX.2.4.6.** — *A scheme X is called **locally noetherian** if it admits an affine open covering $X = \bigcup_i U_i$ such that the rings of sections $\Gamma(U_i, \mathcal{O}_X)$ are noetherian. If X is additionally quasi-compact, then X is called **noetherian**.*

***Lemma* IX.2.4.7.** — *Every locally noetherian scheme admits a basis for its topology consisting of noetherian affine open subschemes. The local rings (stalks) of a locally noetherian scheme are all noetherian.*

Proof. — Let X be a locally noetherian scheme, and let $X = \bigcup_i U_i$ be an affine open covering such that $A_i := \Gamma(U_i, \mathcal{O}_X)$ is noetherian. For each U_i, the principal open subschemes $\mathrm{D}(f)$ of U_i with $f \in A_i$ have ring of sections $\Gamma(\mathrm{D}(f), \mathcal{O}_X) = (A_i)_f$, which is noetherian as a localisations of a noetherian ring. The principal open subschemes are, furthermore, noetherian since they are quasi-compact by proposition IX.1.1.12. They form the desired basis for the topology of X.

For the second statement, note that for $x \in U_i$, the local ring $\mathcal{O}_{X,x} = (A_i)_{\mathfrak{p}_x}$ is again noetherian as a localisation of a noetherian ring.

We need the following lemma for the proof after it.

***Lemma* IX.2.4.8.** — *Let M be an A-module. Suppose that $A = (f_1, \dots, f_n)$, i. e., the unit ideal is finitely generated. Then M is finitely generated iff M_{f_i} is a finitely generated A_{f_i}-module for all i.*

Proof. — Suppose for all i that M_{f_i} is finitely generated over A_{f_i}, say by $m_{ij}/f_i^{n_{ij}}$ where $m_{ij} \in M$, $n_{ij} \geq 0$ and $j = 1, \dots, r_i$. Then $N := (m_{ij}; i, j)$ is a finitely generated A-submodule of M with $N_{f_i} = M_{f_i}$ for all i. Now, since for any prime ideal $\mathfrak{p}$ of A there is some f_i not contained in $\mathfrak{p}$, we have $(M/N)_\mathfrak{p} = ((M/N)_f)_\mathfrak{p} = 0$. By proposition I.3.2.13, we have $M/N = 0$, i. e., $M = N$ is finitely generated. The converse is clear.

***Proposition* IX.2.4.9.** — *An affine scheme $X = \operatorname{Spec} A$ is noetherian iff A is a noetherian ring.*

Proof. — By definition, X is noetherian if A is. Conversely, suppose that X is noetherian, and let $\mathfrak{a} \subseteq A$ be an ideal. We claim that $\mathfrak{a}$ is finitely generated. By assumption, there is a finite affine open covering $X = \bigcup_{i=1}^n \mathrm{D}(f_i)$ of principal open subschemes with $f_i \in A$ such that all $\Gamma(\mathrm{D}(f_i), \mathcal{O}_X) = A_{f_i}$ are noetherian. Then the extended ideals $\mathfrak{a}_{f_i} = A_{f_i}\mathfrak{a}$ in A_{f_i} are finitely generated, and so the statement follows from lemma IX.2.4.8.

***Lemma* IX.2.4.10.** — *For any noetherian scheme, the underlying topological space is noetherian and thus has finitely many irreducible components.*

Proof. — A noetherian scheme X can be covered by finitely many noetherian affine open subschemes, whose underlying topological space is noetherian by proposition IX.2.4.9. By lemma VIII.1.2.13, the underlying topological space of X is noetherian and, by lemma VIII.1.2.14, has finitely many irreducible components.

***Warning* IX.2.4.11.** — In general, a scheme whose underlying topological space is noetherian is not noetherian. For example, let $A = k[x_1, x_2, \dots]/(x_1^2, x_2^2, \dots)$ for k a field. Then $\mathfrak{m} = (\bar{x}_1, \bar{x}_2, \dots)$ is a maximal ideal of A, but also contained in $\operatorname{rad} A$ since each $\bar{x}_i$ is nilpotent. Hence $\operatorname{Spec} A$ as a topological space consists of a single point and is, in particular, noetherian. However, by proposition IX.2.4.9, $\operatorname{Spec} A$ as a scheme is not noetherian because the ring A is not noetherian.

***Corollary* IX.2.4.12.** — *An open subscheme of a (locally) noetherian scheme is (locally) noetherian.*

Proof. — The locally noetherian case follows from (the proof of) lemma IX.2.4.7. Hence, we only need to show that an open subscheme of a noetherian scheme is quasi-compact. Since the underlying topological space of a noetherian scheme is noetherian, this follows from lemma VIII.1.2.14.

***Motivation* IX.2.4.13.** — Usually, schemes have many points that are not closed; in fact, schemes are far from being Hausdorff. Instead of seeing this as a pathology, we should think of this as an advantage. Since the closure $\overline{\{x\}}$ of non-closed points x are quite large, the study of some properties can be reduced to considering the 'generisation' x.

IX.2.4.14. — Let Z be a subset of a topological space X. Recall that a point $z \in Z$ is a *generic point* of Z if $\{z\}$ is dense in Z. Moreover, since the closure of irreducible sets is irreducible (lemma VIII.1.2.4), Z is irreducible if it admits a generic point. We will now show that the converse is true for schemes X as well.

***Proposition* IX.2.4.15.** — *For a scheme X, the map*

$$X \longrightarrow \{\textit{irreducible closed subsets of } X\},$$
$$x \longmapsto \overline{\{x\}},$$

is a bijection, i. e., every irreducible closed subset of X has a unique generic point.

Proof. — We already know from corollary IX.1.1.15 that this is true for affine schemes. Let $Z \subseteq X$ be irreducible closed and let $U \subseteq X$ be affine open such that $Z \cap U$ is non-empty. Then $\overline{Z \cap U} = Z$ because Z is irreducible; otherwise, we would have $Z = \overline{Z \cap U} \cup (Z \setminus U)$. By lemma VIII.1.2.4, $Z \cap U$ is irreducible. Restricting to the affine subscheme U, we see that $Z \cap U$ admits a *unique* generic point, which is also a generic point of Z.

On the other hand, any generic point of Z must be contained in any open subset of X which meets Z. So, if Z has two generic points, then they are both generic points of $Z \cap U$ and hence coincide. ∎

***Observation* IX.2.4.16.** — Let X be a scheme. Recall that a maximal point of X is the generic point of an irreducible component of X. For every point $x \in X$ there exists a maximal point $\eta \in X$ such that η is a generalisation of x, that is, $x \in \overline{\{\eta\}}$.

The question whether points have closed points as specialisations is more subtle.

***Warning* IX.2.4.17.** — There are non-empty irreducible schemes which do not contain closed points.

***Lemma* IX.2.4.18.** — *Let X be a quasi-compact scheme. Then any point $x \in X$ has a closed point $\xi \in X$ such that ξ is a specialisation of x, that is, $\xi \in \overline{\{x\}}$.*

Proof. — We first show that X contains a closed point if X is non-empty. Consider a finite affine open covering $X = \bigcup_{i=1}^n U_i$ such that $U_i \nsubseteq \bigcup_{j \neq i} U_j$ for all i (we need quasi-compactness only for this irredundancy condition). Since $X \setminus \bigcup_{i=2}^n U_i$ is a non-empty closed subset of U_1, it contains a closed point in U_1 by Krull's theorem I.1.3.12. This point is also closed in X because $X \setminus \bigcup_{i=2}^n U_i$ is a closed subset of X.

Finally, if $Z \subseteq X$ is a non-empty closed subset, then the above shows that Z contains a point closed in Z, which is also closed in X because Z is closed. This applies in particular to sets $Z = \overline{\{x\}}$ for $x \in X$ a point. ∎

***Remark* IX.2.4.19.** — This lemma IX.2.4.18 is also true for so-called schemes *locally of finite type* and, more generally, for arbitrary locally noetherian schemes.

[GW20, §3.12].

[GW20, Exer. 5.5].

PART V

REPRESENTATION THEORY

CHAPTER CONTENTS

***Reference* IX.2.4.20.** — Serre on finite groups. Mainly Duistermaat and Kolk on Lie groups and Humphreys on Lie algebras and their representations. Kirillov introduction to Lie groups and Lie algebras seems to be popular as well, maybe compare later. Maybe Knapp Lie groups: beyond an introduction for more general results. Then Milne algebraic groups.

***Notation* IX.2.4.21.** — – The normal tensor product $\otimes$ is assumed to be over a field k, i. e., $V \otimes W = V \otimes_k W$ for k-vector spaces V and W.
– For a group G, the conjugacy class of $g \in G$ is $\mathrm{Cl}(g)$.

The goal of representation theory is to study group actions on vector spaces, which are ubiquitous

CHAPTER X

INTRODUCTION

§ X.1. REPRESENTATION THEORY OF ALGEBRAS

X.1.1. Algebras.

We will recall basic notions for algebras, but now in a generally non-commutative setting and over a field.

Definition* X.1.1.1.** — (i) *An (associative)* ***ring *(with unit) is a ring A in the sense of definition I.1.1.1, but not necessarily commutative.*

(ii) *The* ***opposite ring*** *of a ring A is the ring A^{op} where multiplication $*\colon A^{\mathrm{op}} \times A^{\mathrm{op}} \to A^{\mathrm{op}}$ is defined by $a * b := b \cdot a$.*

(iii) *Let k be a commutative ring. An (associative)* ***k-algebra*** *(with unit) is a ring A together with a (unital) ring map $\phi\colon k \to A$ whose image lies in the centre of A w. r. t. multiplication, i. e., $\phi(\lambda)a = a\phi(\lambda)$ for all $\lambda \in k$ and all $a \in A$. In other words, it is a k-module A together with a multiplication operation $\cdot\colon A \times A \to A$ such that $(ab)c = a(bc)$ for all $a, b, c \in A$ and such that $\lambda a = a\lambda$ for all $\lambda \in k$ and all $a \in A$.*

***Convention* X.1.1.2.** — *In the context of representation theory, we almost always assume k to be a field. The reason is that we want to study the action of the k-algebra on k-vector spaces, and the theory of linear algebra is well-understood.*

***Remark* X.1.1.3.** — From a categorical standpoint, we can equivalently define a k-algebra as a *k-algebra object* in the category of k-vector spaces k-Vect. More precisely, a vector space $A \in k$-Vect is a *k-algebra* if there exist morphisms *multiplication* $m\colon A \otimes A \to A$ and *unit* $u\colon k \to A$ such that:

(i) associativity holds, i. e., the following diagram commutes:

$$\begin{tikzcd} A \otimes A \otimes A \arrow[r, "m\otimes\mathrm{id}"] \arrow[d, "\mathrm{id}\otimes m"'] & A \otimes A \arrow[d, "m"] \\ A \otimes A \arrow[r, "m"] & A; \end{tikzcd}$$

(ii) unit laws hold, i. e., the diagram

$$\begin{tikzcd} k \otimes A \arrow[r, "u\otimes\mathrm{id}"] \arrow[dr, "\sim"'] & A \otimes A \arrow[d, "m"] & A \otimes k \arrow[l, "\mathrm{id}\otimes u"'] \arrow[dl, "\sim"] \\ & A & \end{tikzcd}$$

commutes where the isomorphisms are the canonical ones: $k \otimes A \xrightarrow{\sim} A$, $\lambda \otimes a \mapsto \lambda a$ and $A \otimes k \xrightarrow{\sim} A$, $a \otimes \lambda \mapsto \lambda a$.

By convention, we write $1 := u(1)$ and $ab := m(a \otimes b)$ for all $a, b \in A$, and bilinearity of m is automatic. Moreover, *commutativity* is expressed as $m = m \circ \tau$ where $\tau\colon A \otimes A \to A \otimes A$ is the twisting morphism $\tau(a \otimes b) = b \otimes a$.

***Definition* X.1.1.4.** — *Let A and B be two k-algebras. A k-linear map $f\colon A \to B$ is a **homomorphism of k-algebras** or simply k-*algebra map *if it is also a ring homomorphism, i. e., $f(1) = 1$ and $f(ab) = f(a)f(b)$ for all $a, b \in A$. The notions **isomorphism**, **endomorphism** and **automorphism** are inherited from the notions for k-vector spaces. We denote the* Hom*-set of all k-algebra homomorphisms between A and B by $\mathrm{Hom}_{k\text{-}\mathsf{Alg}}(A, B)$, and the **category of k-algebras** by k-*Alg.

***Remark* X.1.1.5.** — Again, the categorical view defines a *homomorphism of k-algebras* as a morphism $f\colon A \to B$ in k-Vect such that the following diagrams commutes:

$$\begin{tikzcd} A \otimes A \arrow[r, "f \otimes f"] \arrow[d, "m_A"'] & B \otimes B \arrow[d, "m_B"] \\ A \arrow[r, "f"] & B, \end{tikzcd} \qquad \begin{tikzcd} A \arrow[rr, "f"] & & B \\ & k. \arrow[ul, "u_B"] \arrow[ur, "u_A"'] & \end{tikzcd}$$

***Remark* X.1.1.6.** — Compare these definitions with definitions I.1.1.1, I.2.8.1 and I.2.8.3 from commutative algebra. The only difference is that algebras are now allowed to be non-commutative. In particular, constructions like the following work identically to the commutative case.

(i) The *tensor product* of k-algebras A and B is the k-vector space $A \otimes B$ with multiplication $(a \otimes b)(a' \otimes b') = aa' \otimes bb'$ for all $a, a' \in A$ and $b, b' \in B$. This is again a k-algebra via $k \cong k \otimes k \to A \otimes B$ and the unit is $1 \otimes 1 \in A \otimes B$. Note that $A \otimes B \cong B \otimes A$ as k-algebras.

(ii) A field extension K/k can be naturally interpreted as a k-algebra via the inclusion $k \hookrightarrow K$. The *extension of scalars* or *base change* of a k-algebra by K/k is the tensor product $K \otimes A$. This becomes a K-algebra via $K \cong K \otimes k \to K \otimes A$. Thus, we obtain a functor $K \otimes -\colon k\text{-}\mathsf{Alg} \to K\text{-}\mathsf{Alg}$ for any field extension K/k.

***Example* X.1.1.7.** — (i) The archetypical example is the k-algebra $\mathrm{End}\, V$ of all k-linear endomorphisms of a k-vector space V. Multiplication is given by composition, and k acts on $\mathrm{End}\, V$ by $\lambda \mapsto \lambda \mathrm{id}_V$. If V is finite-dimensional, any choice of basis gives an isomorphism $\mathrm{End}\, V \cong \mathrm{Mat}_n(k)$ of k-algebras.

(ii) Of course, any k-algebra from commutative algebra is a k-algebra, e. g., the polynomial k-algebra $k[x_i, i \in I]$.

***Definition* X.1.1.8.** — *The **free k-algebra** $k\langle x_i, i \in I\rangle$ over the variables (letters) $(x_i)_{i \in I}$ has the monomials (finite words) $x_{i_1} \cdots x_{i_k}$ as a basis. Note that the variables x_i are non-commutative, and the empty monomial is $1 \in k\langle x_i, i \in I\rangle$. Multiplication of monomials is given by concatenation of words. We will also just write $k\langle x_1, \dots, x_n\rangle$ in case of finitely many variables.*

The free k-algebra has the universal property that any k-algebra map $f\colon k\langle x_i, i \in I\rangle \to A$ is already defined by the values $f(x_i) \in A$ for all $i \in I$.

***Definition* X.1.1.9.** — *A **subalgebra** of a k-algebra A is a linear subspace B of A that is also a subring of A, which means that $1 \in B$ and $ab \in B$ for all $a, b \in B$. Then B is again a k-subalgebra and the canonical inclusion $\iota\colon B \hookrightarrow A$ is a k-algebra map.*

***Definition* X.1.1.10.** — *Let A be a k-algebra. A **left ideal** in A is a linear subspace $\mathfrak{a}$ in A such that $A\mathfrak{a} \subseteq \mathfrak{a}$. Similarly, a **right ideal** in A is a linear subspace $\mathfrak{a}$ in A such that $\mathfrak{a}A \subseteq \mathfrak{a}$. A **two-sided ideal** is a linear subspace $\mathfrak{a}$ in A that is both a left and right ideal.*

Let S be a subset in A. Then

$$\langle S\rangle := \bigcap_{\textit{two-sided ideals } \mathfrak{a} \supseteq S} \mathfrak{a} = \mathrm{span}(ASB)$$

is the two-sided ideal generated *by S, which is the smallest two-sided ideal containing S. Similarly, $\langle S\rangle_l := \operatorname{span}(AS)$ and $\langle S\rangle_r := \operatorname{span}(SB)$ are the left and right ideals* generated *by S, resp.*

***Example* X.1.1.11.** — The trivial (two-sided) ideals in a k-algebra A are 0 and A.

***Remark* X.1.1.12.** — Like in the commutative case, we can define left, right and two-sided **maximal** ideals. In the same way as Krull's theorem I.1.3.12, every non-zero algebra has a maximal ideal of any kind.

***Construction* X.1.1.13.** — Let A be a k-algebra, and let $\mathfrak{a}$ be a two-sided ideal in A. Then the additive quotient group $A/\mathfrak{a}$ inherits a k-algebra structure by virtue of the canonical projection $\pi\colon A \twoheadrightarrow A/\mathfrak{a}$, $a \mapsto a+\mathfrak{a}$, i. e., $(a+\mathfrak{a})(b+\mathfrak{a}) = ab+\mathfrak{a}$. By using that $\mathfrak{a}$ is a two-sided ideal, one easily checks that this multiplication is well-defined (it does not depend on the chosen representative). The k-algebra $A/\mathfrak{a}$ is the **quotient algebra** of A by $\mathfrak{a}$.

***Example* X.1.1.14.** — If $f_1, \dots, f_m \in k\langle x_1, \dots, x_n\rangle$ are free polynomials, then the k-algebra $k\langle x_1, \dots, x_n\rangle/\langle f_1, \dots, f_m\rangle$ is the algebra generated by $x_1, \dots, x_n$ subject to defining relations $f_1 = \dots = f_m = 0$. A specific instance is the polynomial k-algebra

$$k[x_i, i \in I] = k\langle x_i, i \in I\rangle/\langle x_i x_j - x_j x_i; i, j \in I\rangle.$$

Another one is the first **Weyl algebra** $k\langle x, y\rangle/(yx - xy - 1)$.

Maybe remove example of Weyl algebra.

***Homomorphism theorem* X.1.1.15.** — *Let $f\colon A \to B$ be a k-algebra map. Then the kernel $\ker f$ is a two-sided ideal of A and the image $\operatorname{im} f$ is a subalgebra of B. Moreover, there exists a canonical isomorphism*

$$A/\ker f \xrightarrow{\sim} \operatorname{im} f$$

induced by f.

X.1.2. Tensor algebra.

***Construction* X.1.2.1** (universal property of the tensor algebra)**.** — Let V be a k-vector space. We construct a k-algebra $\operatorname{Ten} V$ with a k-linear map $V \to \operatorname{Ten} V$ such that the following universal property is satisfied: for any k-linear map $f\colon V \to A$ into a k-algebra A, there exists a unique k-algebra map $\operatorname{Ten} V \to A$ such that the following diagram commutes:

$$\begin{tikzcd} V \arrow[r] \arrow[dr, "f"'] & \operatorname{Ten} V \arrow[d, dashed, "\exists!"] \\ & A. \end{tikzcd}$$

Indeed, we first define the nth **tensor power** $V^{\otimes n}$ of V, recalling that a basis is given by elementary tensors $v_1 \otimes \dots \otimes v_n$ for $v_1, \dots, v_n \in V$. These tensor powers give rise to the **tensor algebra**

$$\operatorname{Ten} V := \bigoplus_{n\geq 0} V^{\otimes n}$$

of V where multiplication of elementary tensors is given by taking their tensor product. Observe that $\operatorname{Ten} V$ is, in fact, a graded ring, i. e., for $v \in V^{\otimes n}$ and $w \in V^{\otimes m}$, we have $v \otimes w \in V^{\otimes(n+m)}$. The elements in $V^{\otimes 0} = k$ of degree zero are the scalars and the elements in $V^{\otimes 1} = V$ of degree one generate $\operatorname{Ten} V$. In fact, any basis of V already generates $\operatorname{Ten} V$.

To check the universal property, let $f\colon V \to A$ be a linear map into a k-algebra A. Then $f^n\colon V^n \to A$, $(v_1, \dots, v_n) \mapsto f(v_1)\cdots f(v_n)$ is multilinear and hence induces $f^{\otimes n}\colon V^{\otimes n} \to A$. Taking the direct sum gives $\bigoplus_{n\geq 0} f^{\otimes n}\colon \operatorname{Ten} V \to A$, as desired. Uniqueness follows from $(\bigoplus_{n\geq 0} f^{\otimes n})|_V = f$ and the fact that V generates $\operatorname{Ten} V$.

***Example* X.1.2.2.** — The free k-algebra is an incarnation of a tensor product: we have a k-algebra isomorphism $k\langle x_i, i \in I\rangle \cong \operatorname{Ten}\operatorname{span}(x_i, i \in I)$.

***Construction* X.1.2.3** (universal property of symmetric algebras)**.** — Like for tensor products, we can construct a k-algebra in the commutative case. Let V be a k-vector space, and consider the two-sided ideal

$$\mathfrak{a} = \langle v \otimes w - w \otimes v \in \operatorname{Ten} V; v, w \in V\rangle$$

Insert reference on homogeneous polynomials giving graded quotients.

in $\operatorname{Ten} V$. Since $\mathfrak{a}$ is homogeneous, we obtain the graded **symmetric algebra**

$$\operatorname{Sym} V \coloneqq \operatorname{Ten} V/\mathfrak{a} = \bigoplus_{n\geq 0} \operatorname{Sym}^n V$$

of V where

$$\operatorname{Sym}^n V = V^{\otimes n}/\mathfrak{a} \cap V^{\otimes n} = V^{\otimes n}/\operatorname{span}\{v_1 \otimes \cdots \otimes v_n - \sigma(v_1 \otimes \cdots \otimes v_n) \mid \sigma \in \mathfrak{S}_n\}$$

is the nth **symmetric power** of V. Note that $\mathfrak{a} \cap V^{\otimes 1} = \{0\}$ by degree reasons, so the canonical map $V \hookrightarrow \operatorname{Ten} V \twoheadrightarrow \operatorname{Sym} V$ is an inclusion. The symmetric algebra $\operatorname{Sym} V$ is clearly commutative. We often write $v_1 \cdots v_n \in \operatorname{Sym}^n V$ for the image of $v_1 \otimes \cdots \otimes v_n \in V^{\otimes n}$ in $\operatorname{Sym}^n V$.

Symmetric algebras satisfy the following universal property: for any k-linear map $f\colon V \to A$ into a commutative k-algebra A, there exists a unique k-algebra map $\operatorname{Sym} V \to A$ such that the following diagram commutes:

$$\begin{tikzcd}
V \arrow[r] \arrow[dr, "f"'] & \operatorname{Ten} V \arrow[r, two heads] \arrow[d, dashed] & \operatorname{Sym} V \arrow[dl, dashed, "\exists!"] \\
& A. &
\end{tikzcd}$$

This universal property follows from combining the universal property of tensor algebras (construction X.1.2.1) and quotients (we used that $\mathfrak{a}$ is in the kernel of the induced map $\operatorname{Ten} V \to A$ because A is commutative).

***Example* X.1.2.4.** — The polynomial k-algebra is an example of a symmetric algebra: we have a k-algebra isomorphism $k[x_i, i \in I] \cong \operatorname{Sym}\operatorname{span}(x_i, i \in I)$.

***Construction* X.1.2.5.** — We can also construct an algebra in the same way as symmetric algebras, but for anti-commutative k-algebras. Let V be a k-vector space, and consider the two-sided ideal

$$\mathfrak{a} = \langle v \otimes v \in \operatorname{Ten} V, v \in V\rangle$$

in $\operatorname{Ten} V$. We obtain the graded **alternating algebra** or **exterior algebra**

$$\operatorname{Alt} V \coloneqq \wedge V \coloneqq \operatorname{Ten} V/\mathfrak{a} = \bigoplus_{n\geq 0} \operatorname{Alt}^n V$$

of V where

$$\operatorname{Sym}^n V = \wedge^n V = V^{\otimes n}/\mathfrak{a} \cap V^{\otimes n} = V^{\otimes n}/\operatorname{span}\{v_1 \otimes \cdots \otimes v_n \mid v_i = v_j \text{ for some } i \neq j\}$$

is the nth **alternating power** or **exterior power** of V. We often write $v_1 \wedge \cdots \wedge v_n \in \operatorname{Alt}^n V$ for the image of $v_1 \otimes \cdots \otimes v_n \in V^{\otimes n}$ in $\operatorname{Alt}^n V$.

Maybe simplify this part with reference to commutative algebra on graded rings.

The alternating property $v \wedge v$ for $v \in V$ implies that $v \wedge w = -w \wedge v$ for all $v, w \in V$. More generally, for homogeneous elements $a, b \in \operatorname{Alt} V$, we have $ab = (-1)^{\deg a \deg b} ba$, which is anti-commutativity (hence the name alternating).

***Observation* X.1.2.6.** — All three constructions X.1.2.1, X.1.2.3 and X.1.2.5 are functorial. Let $f\colon V \to W$ be a linear map. Then $f^{\otimes n}\colon V^{\otimes n} \to W^{\otimes n}$ is given by $f(v_1 \otimes \cdots \otimes v_n) = f(v_1) \otimes \cdots \otimes f(v_n)$. Taking the direct sum gives the graded k-algebra map $\operatorname{Ten} f\colon \operatorname{Ten} V \to \operatorname{Ten} W$. We thus obtained a functor

$$\operatorname{Ten}\colon k\text{-}\mathsf{Vect} \longrightarrow k\text{-}\mathsf{Alg}.$$

Moreover, let $\mathfrak{a}$ and $\mathfrak{b}$ be the defining ideals of the symmetric algebras of V and W, resp., so that $\operatorname{Sym} V = \operatorname{Ten} V/\mathfrak{a}$ and $\operatorname{Sym} W = \operatorname{Ten} W/\mathfrak{b}$. Then $\operatorname{Ten} f\colon \operatorname{Ten} V \to \operatorname{Ten} W$ sends $\mathfrak{a}$ to $\mathfrak{b}$, so it descends to the k-algebra map $\operatorname{Sym} f\colon \operatorname{Sym} V \to \operatorname{Sym} W$. In fact, $\operatorname{Sym} f$ is graded since it restricts to maps $\operatorname{Sym}^n V \to \operatorname{Sym}^n W$ for any $n \geq 0$. We thus obtained a functor

$$\operatorname{Sym}\colon k\text{-}\mathsf{Vect} \longrightarrow k\text{-}\mathsf{CommAlg}.$$

Similarly, we can obtain a functor

$$\operatorname{Alt}\colon k\text{-}\mathsf{Vect} \longrightarrow k\text{-}\mathsf{Alg}.$$

***Remark* X.1.2.7.** — Let V be an n-dimensional vector space with basis $(e_1, \ldots, e_n)$. Then the various powers are also finite-dimensional.

(i) The standard basis of $V^{\otimes m}$ is given by elementary tensors $e_{i_1} \otimes \cdots \otimes e_{i_m}$ for indices $1 \leq i_1, \ldots, i_m \leq n$. Thus, $\dim V^{\otimes m} = n^m$.

(ii) The standard basis of $\operatorname{Sym}^m V$ is given by monomials $e_1^{m_1} \cdots e_n^{m_n}$ with $m_1 + \cdots + m_n = m$. Thus, $\dim \operatorname{Sym}^m V = \binom{m+n-1}{n-1}$ using stars and bars.[X.1]

(iii) The standard basis of $\operatorname{Alt}^m V$ is given by $e_{i_1} \wedge \cdots \wedge e_{i_m}$ for indices $i_1 < \cdots < i_m$. Thus, $\dim \operatorname{Alt}^m V = \binom{n}{m}$. In particular, note that $\dim \operatorname{Alt}^m V = 0$ if $m > n = \dim V$.

(iv) By (iii), the exterior algebra $\operatorname{Alt} V$ has finite dimension, which is different from $\operatorname{Ten} V$ and $\operatorname{Sym} V$. Indeed, $\dim \operatorname{Alt} V = \sum_{m=1}^{n} \binom{n}{m} = 2^n$.

X.1.3. Modules.

***Definition* X.1.3.1.** — *Let A be a (not necessarily commutative) ring.*

(i) *A **left A-module** M is defined verbatim as in definition I.2.1.1, in particular with a left A-action $\cdot\colon A \times M \to M$. Equivalently, a left A-module is a (unital) ring map $A \to \operatorname{End}_{\mathsf{AbGrp}}(M)$ with some abelian group $(M, +)$.*

(ii) *For completeness sake, a **right A-module** M is defined like in definition I.2.1.1, but with a right A-action $\cdot\colon M \times A \to M$. In particular,*

$$m(ba) = (mb)a, \qquad m1 = 1.$$

Equivalently, a right A-module is a (unital) ring map $A \to \operatorname{End}_{\mathsf{AbGrp}}(M)^{\mathrm{op}}$.

***Definition* X.1.3.2.** — *Let M and N be two left A-modules. A **left A-module homomorphism** or a **left A-linear** map $f\colon M \to N$ is defined as in definition I.2.1.3. In particular, $f(av) = af(v)$ for all $a \in A$ and all $v \in V$, or as a diagram:*

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ {\scriptstyle a}\downarrow & & \downarrow{\scriptstyle a} \\ M & \xrightarrow{f} & N. \end{array}$$

[X.1] We choose a multiset of m vectors (stars) from $\{e_1, \ldots, e_n\}$. Since ordering does not matter, we can group the vectors into powers and insert bars between each power. This gives a sequence of $m + n - 1$ stars and bars, and we need to choose the position of the $n - 1$ bars.

We write $\operatorname{Hom}_A(M,N)$ *for the set of all* A*-linear maps* $M \to N$*, and* $\operatorname{End}_A M := \operatorname{Hom}_A(M,M)$*. Hence, we obtain the* ***category of left*** A***-modules*** A-LMod.

Likewise, given right A*-modules* M *and* N*, we define a* ***right*** A***-module homomorphisms*** *or a* ***right*** A***-linear maps*** $f\colon M \to N$*, which satisfies* $f(ma) = f(m)a$ *for all* $a \in A$ *and all* $m \in M$*. Hence, we obtain the* ***category of right*** A***-modules*** A-RMod.

***Fact* X.1.3.3.** — *There is an equivalence of categories* A-$\mathsf{RMod} \cong A^{\mathrm{op}}$-$\mathsf{LMod}$.

Proof. — If M is a right A-module, then it can be endowed with a left A^{op}-module structure via $A \times M \to M$, $am := ma$, and vice versa. Alternatively, if $A \to \operatorname{End}_{\mathsf{AbGrp}}(M)^{\mathrm{op}}$ is a right A-module, then $A^{\mathrm{op}} \to (\operatorname{End}_{\mathsf{AbGrp}}(M)^{\mathrm{op}})^{\mathrm{op}} \cong \operatorname{End}_{\mathsf{AbGrp}}(M)$ is a left A^{op}-module. Hence, the right A-modules are in bijection with the left A^{op}-modules. This extends to hom-sets, so we obtain an equivalence of categories ◻

***Convention* X.1.3.4.** — *Because of fact X.1.3.3,* A-modules *are always assumed to be* left A-modules*, unless otherwise specified. Furthermore, we will always say* A-module map *instead of* (left) A-module homomorphism*.*

***Remark* X.1.3.5.** — All constructions and notions like submodules, quotients, finitely generated, noetherian, artinian, etc. from commutative algebra can be defined for this more general notion of left or right modules.

***Homomorphism theorem* X.1.3.6.** — *Let* $f\colon M \to N$ *be an* A*-module map. Then* $\ker f$ *and* $\operatorname{im} f$ *are submodules of* M *and* N*, resp., and* f *induces an isomorphism*

$$M/\ker f \xrightarrow{\sim} \operatorname{im} f.$$

***Remark* X.1.3.7.** — If A is a k-algebra, then a left A-module V is automatically a k-vector space, and both addition and scalar multiplication are k-linear. In this case, $\operatorname{End}_A V$ is a k-subalgebra of $\operatorname{End}_k V$.

***Convention* X.1.3.8.** — *Henceforth, we will always work with modules over* k*-algebras, unless otherwise specified.*

***Remark* X.1.3.9.** — We can equivalently define an A-module as an *A-module object* in k-Vect. Let A be a k-algebra together with multiplication $m\colon A \otimes A \to A$ and unit $u\colon k \to A$ maps. A vector space $V \in k$-Vect is an *A-module* if there is a morphism $\mu\colon A \otimes V \to V$ such that the diagrams

$$\begin{tikzcd} A \otimes A \otimes V \arrow[r, "m\otimes \mathrm{id}_V"] \arrow[d, "\mathrm{id}_A\otimes\mu"'] & A \otimes V \arrow[d, "\mu"] \\ A \otimes V \arrow[r, "\mu"] & V, \end{tikzcd} \qquad \begin{tikzcd} k \otimes V \arrow[r, "u\otimes \mathrm{id}_V"] \arrow[dr, "\sim"'] & A \otimes V \arrow[d, "\mu"] \\ & V \end{tikzcd}$$

commute. By convention, we write $av := \mu(a \otimes v)$ for all $a \in A$ and $v \in V$. Observe that from the tensor-hom adjunction $\operatorname{Hom}_k(A \otimes V, V) \cong \operatorname{Hom}_k(A, \operatorname{End}_k V)$, we obtain the equivalent definition $\rho\colon V \to \operatorname{End}_k V$ from μ.

Similarly, a k-linear morphism $f\colon V \to W$ of A-modules is *A-linear* if the following diagram commutes:

$$\begin{tikzcd} A \otimes V \arrow[r, "\mathrm{id}_A\otimes f"] \arrow[d, "\mu_V"'] & A \otimes W \arrow[d, "\mu_W"] \\ V \arrow[r, "f"] & W. \end{tikzcd}$$

***Definition* X.1.3.10.** — *Let A and B be (not necessarily commutative) rings. An (A, B)-**bimodule** is an abelian group M which is both a left A-module and a right B-module such that $(am)b = a(mb)$ for all $a \in A$, $b \in B$ and all $m \in M$.*

*A map $f\colon M \to N$ between (A, B)-modules is an (A, B)-**bimodule homomorphism** if it is both left A-linear and right B-linear. We denote the **category of** (A, B)-**bimodules** by (A, B)-BMod.*

***Fact* X.1.3.11.** — *There is an equivalence of categories (A, B)-$\mathsf{BMod} \cong (A \otimes B^{\mathrm{op}})$-$\mathsf{LMod}$.*

Proof. — An (A, B)-bimodule M is completely described by its left and right module structures $A \to \mathrm{End}_{\mathsf{AbGrp}}(M)$ and $B^{\mathrm{op}} \to \mathrm{End}_{\mathsf{AbGrp}}(M)$. By the universal property of tensor products of $\mathbb{Z}$-algebras, these ring maps are in bijection with ring maps $A \otimes_{\mathbb{Z}} B^{\mathrm{op}} \to \mathrm{End}_{\mathsf{AbGrp}}(M)$. Hence, we have the bijection on the level of objects. Functoriality of tensor products shows that this bijection extends to hom-sets. ∎

***Remark* X.1.3.12.** — Henceforth, let us work over k-algebras. Observe that A-modules are (A, k)-bimodules and right A-modules are (k, A)-bimodules.

Bimodules allow us to endow tensor products and hom-spaces with an A-module structure.

***Construction* X.1.3.13.** — Let V be a (B, A)-bimodule and let W be an (A, C)-bimodule. We define the **tensor product** of M and N over A as

$$V \otimes_A W := (V \otimes_k W)/\langle va \otimes w - v \otimes aw \mid v \in V, w \in W, a \in A\rangle.$$

This is a (B, C)-bimodule. We denote the image of the elementary tensor $v \otimes w \in V \otimes_k W$ by $v \otimes_A w$.

***Construction* X.1.3.14.** — Let V be an (A, B)-bimodule and let W be an (A, C)-bimodule. We can endow the **hom-space** $\mathrm{Hom}_A(V, W)$ with a (B, C)-bimodule structure given by $(bfc)(v) := f(vb) \cdot c$ for all $f \in \mathrm{Hom}_A(V, W)$, $b \in B$ and $c \in C$.

As a special case, to simply notation, let V be an A-module V and let W be a B-module. Then the hom-space $\mathrm{Hom}(V, W) = \mathrm{Hom}_k(V, W)$ is a (B, A)-bimodules defined by $(bfa)(v) = bf(av)$ for all $f \in \mathrm{Hom}(V, W)$, $b \in B$ and $a \in A$.

***Tensor-hom adjunction* X.1.3.15.** — *Let U be an (A, B)-bimodule, let V be a (B, C)-bimodule, and let W be a (A, D)-bimodule. Then we have an isomorphism*

$$\begin{aligned}\mathrm{Hom}_A(U \otimes_B V, W) &\xrightarrow{\sim} \mathrm{Hom}_B(V, \mathrm{Hom}_A(U, W)),\\ f &\longmapsto (v \longmapsto f(- \otimes_B v)),\\ (u \otimes_B v \longmapsto f(v)(u)) &\longleftarrow f\end{aligned}$$

of (C, D)-bimodules.

Proof. — Check directly. ∎

Representation theory roughly studies symmetries of vector spaces by letting an algebraic object act on it. It all started with the representation theory of finite groups, expounded by Frobenius in 1896. The general flavour of representation theory is *non-commutative*, a prime example for non-commutative structures being endomorphism rings or algebras. The main objective is to study when a given algebra (resp. representation) can be decomposed into simpler algebras (resp. representations), and to classify alls of these simpler algebras (resp. representations).

X.1.4. Representations.

***Definition* X.1.4.1.** — *Let A be an algebra over a field k.*[(X.2)]

(i) *A **representation** of A or a **left** A**-module** is a k-vector space V equipped with a (unital) k-algebra map $\rho\colon A \to \operatorname{End}_k(V) = \operatorname{End}_{k\text{-}\mathsf{Vect}}(V)$. We write $av := \rho(a)v$ for all $a \in A$ and all $v \in V$. By abuse of language, we will often not mention the map ρ although different choices of ρ yield different representations.*

(ii) *For completeness sake, a **right** A**-module** is a k-vector space V equipped with a (unital) k-algebra map $\rho\colon A^{\mathrm{op}} \to \operatorname{End}_k(V)$. We write $va := \rho(a)v$ for all $a \in A$ and all $v \in V$. In particular, $v(ab) = (va)b$ for all $a, b \in A$ and all $v \in V$.*

***Remark* X.1.4.2.** — Whether scalars in A are written on the left or right of V is just a matter of convention. For example, a right A-module satisfies $(ab)v = b(av)$ for all $a, b \in A$ and all $v \in V$ if scalar multiplication is written on the left.

***Remark* X.1.4.3.** — If A is a commutative k-algebra, then the notions of left and right A-modules agree: if V is a left A-module, define $va := av$ to turn it into a right A-module, and vice versa. This is true even more more general commutative $\mathbb{Z}$-algebras A, i. e., commutative rings.

***Definition* X.1.4.4.** — *A representation $\rho\colon A \to \operatorname{End}_k(V)$ of a k-algebra A is **faithful** if ρ is injective.*

***Convention* X.1.4.5.** — *If not explicitly mentioned, we will always assume* left *A-modules, i. e., representations.*

***Example* X.1.4.6.** — (i) The trivial k-vector space $V = 0$ is a representation for any k-algebra.

(ii) For any k-algebra A, the **trivial representation** is any k-vector space V with $av = v$ for all $a \in A$ and $v \in V$, i. e., A acts by the identity on V.

(iii) For any k-algebra A, the algebra $V = A$ itself defines the **regular representation** $\rho\colon A \to \operatorname{End}_k A$ given by the usual multiplication $\rho(a)v = av$ in A.

(iv) Let $A = k\langle x_1, \dots, x_n\rangle$ be a finitely generated free k-algebra. Then any representation $\rho\colon A \to \operatorname{End}_k V$ of A is given by a k-vector space V together with an arbitrary collection of k-linear maps $\rho(x_1), \dots, \rho(x_n)\colon V \to V$.

***Definition* X.1.4.7.** — *Let A be a k-algebra. A **homomorphism of representations** or A**-linear map** between two representations V and W of A is a k-linear map $\phi\colon V \to W$ such that $\phi(av) = a\phi(v)$ for all $a \in A$ and all $v \in V$. In other words, the following diagram commutes:*

$$\begin{tikzcd} V \arrow[r, "\phi"] \arrow[d, "a"'] & W \arrow[d, "a"] \\ V \arrow[r, "\phi"] & W. \end{tikzcd}$$

We denote the space of homomorphisms between two representations V and W as $\operatorname{Hom}_A(V, W)$. This is again a k-vector space by pointwise operations. If $V = W$, then we denote the space of endomorphisms by $\operatorname{End}_A(V)$.

*If $\phi\colon V \to W$ is an isomorphism of k-vector spaces, then it is an **isomorphism of representations**. In this case, ϕ^{-1} is an isomorphism as well and we say that V and W are **isomorphic**, written $V \cong W$.*

***Fact* X.1.4.8.** — *For a k-algebra A, we have the isomorphism $A^{\mathrm{op}} \xrightarrow{\sim} \operatorname{End}_A(A)$ mapping $a \in A$ to right-multiplication by a.*

(X.2) The definitions works verbatim also for any algebra over a commutative ring R. Just replace k-vector spaces V by R-modules M.

Proof. — Let ϕ be this map. Note that it is well-defined: for any $a \in A$, the map $\phi(a)$ is left A-linear because $\phi(a)(bv) = bva = b\phi(a)(v)$ for all $b, v \in A$. Moreover, ϕ is clearly k-linear, and ϕ is also a ring homomorphism since $\phi(a * b)(v) = v(a * b) = v(ba) = (vb)a = \phi(a) \circ \phi(b)(v)$ for all $v \in A$. Finally, the inverse map of ϕ is given by evaluating $f \in \mathrm{End}_A(A)$ at 1 to give $f(1) \in A$. This is clearly k-linear, and it is a ring homomorphism since $g \circ f(1) = g(f(1)) = f(1)g(1) = g(1) * f(1)$. ∎

***Definition* X.1.4.9.** — *Let V be a representation of a k-algebra A. A **subrepresentation** of V is a k-linear subspace $W \subseteq V$ which is **stable** or **invariant** under A, i. e., $aw \in W$ for all $a \in A$ and all $w \in W$. Then W is again a representation of A, and the canonical inclusion $\iota\colon W \hookrightarrow V$ is a homomorphism of representations.*

***Example* X.1.4.10.** — If V is a representation, then 0 and V are the trivial subrepresentations.

***Construction* X.1.4.11.** — Let V be a representation of a k-algebra A, and let W be a subrepresentation of V. Then the quotient space V/W inherits a left A-module structure by $a(v+W) = av+W$ for all $a \in A$ and all $v \in V$. The canonical projection $\pi\colon V \twoheadrightarrow V/W$, $v \mapsto v + W$ is a homomorphism of representations, and V/W is the **quotient representation** of V by W.

***Homomorphism theorem* X.1.4.12.** — *Let $\phi\colon V \to W$ be a homomorphism of representations of a k-algebra. Then the kernel $\ker\phi$ is a subrepresentation of V and the image $\operatorname{im}\phi$ is a subrepresentation of W. Moreover, there exists a canonical isomorphism*

$$V/\ker\phi \xrightarrow{\sim} \operatorname{im}\phi$$

induced by ϕ.

***Definition* X.1.4.13.** — *A non-zero representation V is **irreducible** or **simple** if the only subrepresentations are the trivial ones: 0 and V.*

***Remark* X.1.4.14.** — Any non-zero finite-dimensional representation V has an irreducible subrepresentation. For example, if V is not irreducible, already, then there is a proper subrepresentation W of V with $\dim W < \dim V$. By induction, W and hence V has an irreducible subrepresentation.

***Warning* X.1.4.15.** — If V is infinite-dimensional, irreducible subrepresentations might not exist. For example, take $V = k[x]$ to be the regular representation of $k[x]$. Then subrepresentations of V are just ideals of the principal ideal domain $k[x]$, and any non-zero ideal (f) of $k[x]$ properly contains the ideal (xf).

***Definition* X.1.4.16.** — *Let V and W be two representations over a k-algebra A. The **direct sum** of V and W is the k-linear direct sum $V \oplus W$ endowed with the obvious left A-module structure given by $a(v, w) = (av, aw)$ for all $a \in A$ and all $v \in V$ and $w \in W$.*

***Definition* X.1.4.17.** — *Let V be a non-zero representation of a k-algebra A. If V is isomorphic to a direct sum $W \oplus W'$ of two non-zero representations of A, then $W \oplus W'$ is a **decomposition** of V. Otherwise, V is **indecomposable**.*

***Warning* X.1.4.18.** — Irreducible representations are obviously indecomposable, but indecomposable representations are generally not irreducible, as example X.1.4.24 shows.

***Definition* X.1.4.19.** — *The **degree** or **dimension** of a representation V is its k-dimension $\dim V$.*

***Example* X.1.4.20.** — One-dimensional representations are automatically irreducible.

***Schur's lemma* X.1.4.21.** — *Let $\phi\colon V \to W$ be a homomorphism between representations of a k-algebra.*

(i) *If V is irreducible, then ϕ is either injective or zero. If W is irreducible, then ϕ is either surjective or zero. Thus, if V and W are both irreducible, then ϕ is either an isomorphism or zero*

(ii) *Suppose that k is an algebraically closed field, and suppose that $V = W$ is a finite-dimensional irreducible representation. Then $\phi = \lambda \mathrm{id}_V$ is homothety by some $\lambda \in k$.*

Proof. — For $v \in \ker\phi$, we have $\phi(av) = a\phi(v) = 0$, so $av \in \ker\phi$ and $\ker f$ is stable under G. If V is irreducible, then we either have $\ker\phi = 0$ or $\ker\phi = V$. The latter case implies $\phi = 0$, the form case implies that ϕ is injective. A similar argument for $\operatorname{im}\phi$ shows that ϕ is surjective, proving (i).

Now, suppose that $V = W$ is irreducible of finite dimension and that k is algebraically closed. Let λ be an eigenvalue of ϕ, which must exist since the characteristic polynomial $\det(\phi - \lambda\mathrm{id}_V)$ has a root over k. Consider the homomorphism $f - \lambda\mathrm{id}_V$ of representations. Since its determinant is vanishes, it cannot be isomorphic, and (i) shows that $f - \lambda\mathrm{id}_V = 0$, as desired for (ii). ◻

***Corollary* X.1.4.22.** — *Let A be a commutative k-algebra with k algebraically closed. Then every finite-dimensional irreducible representation of A is one-dimensional.*

Proof. — Let V be a finite-dimensional irreducible representation of A. Any $a \in A$ defines a homomorphism $V \to V$ of representations, because $a(bv) = (ab)v = (ba)v = b(a(v))$ for all $b \in A$ and all $v \in V$. Then Schur's lemma X.1.4.21 says that any $a \in A$ is a homothety on V, and hence any linear subspace of V is already stable under A. By irreducibility, V must be one-dimensional (note that $V \neq 0$). ◻

***Example* X.1.4.23.** — Let $A = k$. Then the representations of A are precisely the k-vector spaces V. The only irreducible or indecomposable representation is $V = A$.

***Example* X.1.4.24.** — Let $A = \mathbb{C}[x]$ (or $A = k[x]$ for any algebraically closed field k), which is a commutative $\mathbb{C}$-algebra. Recall that it suffices to give a linear operator $x\colon V \to V$ in order to define a representation V of $\mathbb{C}[x]$.

By corollary X.1.4.22, the finite-dimensional irreducible representations V of $\mathbb{C}[x]$ are one-dimensional, say, $V = \mathbb{C}$, and a linear operator $x\colon \mathbb{C} \to \mathbb{C}$ is a number $x \in \mathbb{C}$. Thus, the finite-dimensional irreducible representations of $\mathbb{C}[x]$ are $V_\lambda = \mathbb{C}$ with $x = \lambda\mathrm{id}_{\mathbb{C}}$ for any $\lambda \in \mathbb{C}$ (up to isomorphism).

To classify all finite-dimensional indecomposable representations V of $\mathbb{C}[x]$, recall that we can bring a linear operator $x\colon V \to V$ into Jordan normal form by some base change. More specifically, x is equivalent to a diagonal matrix consisting of Jordan blocks

$$J_{\lambda,n} = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}$$

acting on $\mathbb{C}^n$ where $\lambda \in \mathbb{C}$. So, if V is finite-dimensional and indecomposable, then x can only consist of a single Jordan block. Thus, the finite-dimensional indecomposable representations of $\mathbb{C}[x]$ are $V_{\lambda,n} = \mathbb{C}^n$ with $x = J_{\lambda,n}$ for any $\lambda \in \mathbb{C}$. These are pairwise non-isomorphic since the Jordan normal form is unique up to permutation of Jordan blocks.

We even have the following generalisation of Schur's lemma X.1.4.21.

***Corollary* X.1.4.25** (Dixmier). — *Let A be an algebra over an uncountable algebraically closed field k, e. g., $k = \mathbb{C}$, and let V be an irreducible representation of A with countable basis. Then any homomorphism $\phi\colon V \to V$ is a homothety.*

Proof. — By Schur's lemma X.1.4.21, $D = \mathrm{End}_A(V)$ is a division algebra, i. e., $D^\times = D \setminus \{0\}$, and it is of countable dimension because V is. By way of contradiction, suppose that ϕ is not a homothety, i. e., $\phi \notin k$. Since k is algebraically closed and since D has no zero-divisors, ϕ must be transcendental over k. Hence, $\{1/(\phi - \lambda) \mid \lambda \in k\}$ is an uncountable linearly independent set over k and the transcendental extension $k(\phi)/k$ is of uncountable degree. However, D was of only countable dimension. ∎

X.1.5. Ideals, quotients and tensor products of algebras.

***Remark* X.1.5.1.** — Left ideals in A are precisely the subrepresentations of the regular representation of A. Right ideals in A are precisely the subrepresentations of the regular representation of A^{op}.

Definition* X.1.5.2.** — *A non-zero k-algebra A is* ***simple *if the trivial* 0 *and A are the only two-sided ideals.*

***Example* X.1.5.3.** — If $\mathfrak{a}$ is a left ideal of a k-algebra A, then it is also a subrepresentation of the regular representation A, so $A/\mathfrak{a}$ is the quotient representation of A by $\mathfrak{a}$. This is usually not an algebra since $\mathfrak{a}$ is not two-sided.

Definition* X.1.5.4.** — *Let V be a non-zero representation of a k-algebra A. If $v \in V$ alone generates V, i. e., $Av = V$, then v and V are called* ***cyclic*.*

***Fact* X.1.5.5** (properties of cyclic representations)**.** — *Let V be a non-zero representation of a k-algebra A.*

(i) *V is irreducible if and only if all non-zero vectors of V are cyclic.*
(ii) *V is cyclic if and only if $V \cong A/\mathfrak{a}$ where $\mathfrak{a}$ is a left ideal in A.*

Proof. — (i) follows easily from the definitions. For (ii), let V be cyclic, say, $V = Av$ for some $v \in V$. Then $\phi\colon A \twoheadrightarrow V$, $a \mapsto av$ is a surjective homomorphism of representations, so we obtain an isomorphism $A/\mathfrak{a} \xrightarrow{\sim} V$ where $\mathfrak{a} = \ker\phi$ is a left ideal. Conversely, let $\phi\colon A/\mathfrak{a} \xrightarrow{\sim} V$ be an isomorphism of representations with $\mathfrak{a}$ a left ideal in A. Then $\phi(1) \in V$ is cyclic since ϕ is A-linear and surjective. ∎

***Counterexample* X.1.5.6.** — Let $A = \mathbb{C}[x,y]/(x^2, y^2, xy)$, so that A has a basis $(1, x, y)$. Consider the three-dimensional representation $V = A^* = \mathrm{Hom}(A, \mathbb{C})$ of A given by $af(b) = f(ba)$ for all $b \in A$. Then $(1^*, x^*, y^*)$ is a basis of V with action $xx^* = yy^* = 1^*$ and $x1^* = xy^* = y1^* = yx^* = 0$. So for a general element $f = \lambda 1^* + \mu x^* + \nu y^* \in V$, we have $xf = \mu 1^*$ and $yf = \nu 1^*$. Hence, the subrepresentation Af is at most two-dimensional and thus, V is not cyclic.

On the other hand, suppose that $V = W \oplus W'$ is a decomposition. Write $1^* = w + w'$ with $w \in W$ and $w' \in W'$. For any $f \in W$ in the form above, we have $xf = \mu(w + w') \in W$ and $yf = \nu(w+w') \in W$, so $\mu = \nu = 0$. Since $W \neq 0$, we must have $W = \mathrm{span}(1^*)$. The same argument yields $W' = \mathrm{span}(1^*)$, a contradiction. Thus, V is indecomposable.

X.1.6. Examples.

The representation theory of algebras can be applied to other, more specialised representation theories. We will give examples how representations of other objects are defined in terms of algebras.

***Example* X.1.6.1.** — For a group G, define the *group algebra* $A = k[G]$ over k as the k-vector space with basis $(a_g)_{g \in G}$ and multiplication given by $a_g a_h = a_{gh}$. Then a *group representation* of G is a representation $\rho\colon k[G] \to \mathrm{End}_k V$ of the group algebra $k[g]$. Note that, since elements of G are invertible, the representation is a k-algebra map $\rho\colon k[G] \to \mathrm{GL}_k V$.

***Example* X.1.6.2.** — A *quiver* is just a directed graph $Q = (I, E)$, allowing loops and multiple edges, where I is the set of vertices and where E is the set of edges. For example,

$$\begin{tikzcd} \bullet \arrow[r] & \bullet \arrow[loop above] & \bullet \arrow[l] \\ & \bullet \arrow[u] & \end{tikzcd}$$

A *quiver representation* of Q assigns to each vertex $i \in I$ a k-vector space V_i and to each edge $e\colon e' \to e''$ a k-linear map $f_e\colon V_{e'} \to V_{e''}$.

To pass to algebras, define the *path algebra* of a quiver Q as the k-algebra P_Q whose basis is given by $(a_p)_p$ where p runs through all oriented paths in Q, including the trivial paths p_i at each vertex $i \in I$, and multiplication of paths is given by concatenation. By convention, the product ab of two paths a and b in Q is read right-to-left: first traversing b, then a. If two paths cannot be connected, then their product vanishes. Observe that $\sum_{i\in I} p_i = 1$ is the unit in P_Q if Q is finite. In fact, P_Q is generated by trivial paths p_i for $i \in I$ and edge paths a_e for $e \in E$ with defining relations:

(i) $\sum_{i\in I} p_i = 1$;
(ii) $p_i^2 = p_i$ and $p_i p_j = 0$ for all $i \neq j$;
(iii) $a_e p_{e'} = a_e$ and $a_e p_i = 0$ for all $i \neq e'$ and all $e\colon e' \to e''$;
(iv) $p_{e''} a_e = a_e$ and $p_j a_e = 0$ for all $j \neq e''$ and all $e\colon e' \to e''$.

One can show that there is a correspondence between representations of a finite quiver Q and representations of the path algebra P_Q. Let V be a representation of the path algebra P_Q. Define $V_i := p_i V$ for $i \in I$ and $f_e := a_e|_{p_{e'}V}\colon p_{e'}V \to p_{e''}V$ where $a_e \in P_Q$ is the path corresponding to $e \in E$. Then $(V_i, f_e)_{i\in I, e\in E}$ is a representation of Q. Conversely, let $(V_i, f_e)_{i\in I,e\in E}$ be a representation of the quiver Q. Define $V := \bigoplus_{i\in I} V_i$, define $p_i\colon V \twoheadrightarrow V_i \hookrightarrow V$ for $i \in I$ and define $a_e\colon V \twoheadrightarrow V_{e'} \to V_{e''} \hookrightarrow V$ for $e \in E$ induced by f_e. One easily checks that these constructions are mutually inverse.

We can further define *homomorphisms* $\phi\colon (V_i, f_e)_{i\in I,e\in E} \to (W_i, g_e)_{i\in I,e\in E}$ of representations of quiver Q by a family of maps $(\phi_i\colon V_i \to W_i)_{i\in I}$ such that $g_e \circ \phi_{e'} = \phi_{e''} \circ f_e$ for all edges $e \in E$, i. e., such that the ϕ_i are compatible with the linear maps. The correspondence between representations of Q and of P_Q is even an *equivalence of categories*.

***Example* X.1.6.3.** — A *Lie algebra* over a field k is a k-algebra $\mathfrak{g}$ whose multiplication is given by a *Lie bracket* $[-,-]\colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, which satisfies anti-symmetry $[a, b] = -[b, a]$ and the *Jacobi identity* $[[a, b], c] + [[b, c], a] + [[c, a], b] = 0$. The prototypical example is the endomorphism algebra $\mathfrak{g} = \mathfrak{gl}(V) = \operatorname{End} V$ of a k-vector space V. Its Lie bracket is given by the commutator $[a, b] = ab - ba$. Another important example is the space $\operatorname{Der} A$ of *derivations* of a k-algebra A, i. e., k-linear maps satisfying the Leibniz rule $D(ab) = D(a)b + aD(b)$. A *Lie algebra representation* of a Lie algebra $\mathfrak{g}$ over k is a representation $\rho\colon \mathfrak{g} \to \operatorname{End}_k V$.

The theory of Lie algebras is closely related with the study of *Lie groups* (historically, Lie algebras were just a device to study Lie groups), which are smooth manifolds with smooth group operations. One can associate to a Lie group G a Lie algebra $\mathfrak{g}$, which is the tangent space of G at $1 \in G$ and whose elements are derivations that are invariant under the action of G. This association is, in fact, bijective for isomorphism classes of simply connected Lie groups and finite-dimensional real Lie algebras over. We obtain a similar correspondence between Lie algebra and Lie group representations.

X.1.7. Semisimple algebras.

Definition* X.1.7.1.** — *A representation of a k-algebra is* ***semisimple *or* ***completely reducible*** *if it is a direct sum of irreducible subrepresentations.*

***Example* X.1.7.2.** — Let V be an n-dimensional irreducible representation of a k-algebra A. Then $\operatorname{End} V$ is a representation of A by left multiplication (post-composition). If $(e_1, \dots, e_n)$ is a basis of V, then we obtain an isomorphism $\operatorname{End} V \xrightarrow{\sim} V^{\oplus n}$, $f \mapsto (f(e_1), \dots, f(e_n))$ of representations, so $\operatorname{End} V$ is semisimple.

***Remark* X.1.7.3.** — Let A be an algebra over an algebraically closed field k. Any finite-dimensional semisimple representation V of A admits a canonical isomorphism

$$\bigoplus_{[X]} \operatorname{Hom}_A(X, V) \otimes X \xrightarrow{\sim} V, \qquad f \otimes x \longmapsto f(x)$$

of representations where $[X]$ runs through all isomorphism classes of irreducible representations of A. Here, A acts on X in $\operatorname{Hom}_A(X, V) \otimes X$.

Indeed, if this is an isomorphism for finitely many representations V_i, then it is also an isomorphism for the finite direct sum $V = \bigoplus_i V_i$. Hence, w. l. o. g., let V be irreducible. By Schur's lemma X.1.4.21, we have $\operatorname{Hom}_A(X, V) \cong k$ if $X \cong V$ and $\operatorname{Hom}_A(X, V) = 0$ else. This shows that the domain of the map is isomorphic to $k \otimes V \cong V$, as desired.

***Proposition* X.1.7.4.** — *Let A be an algebra over an algebraically closed field, and let V be a finite-dimensional semisimple representation of A, say $V \cong \bigoplus_{i=1}^m V_i^{\oplus n_i}$ where $V_1, \dots, V_m$ are pairwise non-isomorphic irreducible representations of A. Then a subrepresentation W of V is isomorphic to $\bigoplus_{i=1}^m V_i^{\oplus r_i}$ for some $r_i \le n_i$, and the inclusion $\iota\colon W \hookrightarrow V$ is a direct sum of inclusions $\iota_i\colon V_i^{\oplus r_i} \hookrightarrow V_i^{\oplus n_i}$ given by $\iota_i \begin{pmatrix} v_1 & \cdots & v_{r_i} \end{pmatrix}^\top = X_i \begin{pmatrix} v_1 & \cdots & v_{r_i} \end{pmatrix}^\top$ where the $(n_i \times r_i)$-matrix X_i has full column rank r_i.*

Proof. — We prove this by induction on $n = \sum_{i=1}^m n_i$. The case $n = 1$ is clear, in which case V is irreducible. For the induction step, let $W \neq 0$ be a subrepresentation of V, and let P be an irreducible subrepresentation of W (which exists by remark X.1.4.14). Then the inclusion $\iota|_P\colon P \hookrightarrow V$ can be composed with projections $V \twoheadrightarrow V_i^{\oplus n_i}$. By Schur's lemma X.1.4.21, there is some i such that $P \to V \twoheadrightarrow V_i^{\oplus n_i}$ is injective, whence $\iota|_P$ factors through $V_i^{\oplus n_i}$. By the same argument with a projection $V_i^{\oplus n_i} \twoheadrightarrow V_i$, we may identify $P \cong V_i$ in which case $\iota|_P\colon P = V_i \hookrightarrow V_i^{\oplus n_i}$ is given by $v \mapsto \begin{pmatrix} q_1 v & \cdots & q_{n_i} v \end{pmatrix}^\top$ with $q_l \in k$ not all zero. ◻

X.1.8. Definitions.

X.1.8.1. — Let V be a complex vector space and let $\mathrm{GL}(V)$ be the group of automorphisms of V called the *general linear group*. If $(e_1, \dots, e_n)$ is a finite basis of V, then any linear map $a\colon V \to V$ can be expressed by the square matrix $(a_{ij})_{ij} \in \mathrm{Mat}_n(\mathbb{C})$ where $a(e_j) = \sum_{i=1}^n a_{ij} e_i$. Then $a \in \mathrm{GL}(V)$ if and only if $\det a = \det(a_{ij})_{ij} \neq 0$. Thus, we obtain an isomorphism $\mathrm{GL}(V) \cong \mathrm{GL}_n(\mathbb{C})$ of groups.

Definition* X.1.8.2.** — *Let G be a finite group and let V be a complex vector space. A* ***representation *of G in V is a group homomorphism $\rho\colon G \to \mathrm{GL}(V)$, i. e., we have $\rho(st) = \rho(s) \circ \rho(t)$ for all $s, t \in G$.*

Then V obtains a G-module structure by virtue of ρ: by abuse of notation, $gv := \rho_g(v) := \rho(g)(v)$ for all $g \in G$ and all $v \in V$. In particular, $1 \in G$ is id_V and $g(hv) = (gh)v$ for all $g, h \in G$ and all $v \in V$. By abuse of language, V is called the ***representation space*** *or just* ***representation*** *of G.*

The ***dimension*** *(or* ***degree****) of a representation V of G is the complex dimension $\dim_{\mathbb{C}} V$.*

***Remark* X.1.8.3.** — We mostly deal with *finite-dimensional* representations V. This is, however, not a serve restriction. Most applications involve only finitely many elements $v_1, \dots, v_n \in V$, and we can restrict to a finite-dimensional subrepresentation $\operatorname{span}_{g \in G}(gv_1, \dots, gv_n)$.

***Remark* X.1.8.4.** — Let G be a finite group, and let V be a finite-dimensional representation of G. By choice of a basis $(e_1, \dots, e_n)$ of V, we can associate a matrix $R_g \in \mathrm{GL}_n(\mathbb{C})$ to $g\colon V \to V$. Then $R_{gh} = R_g R_h$ for all $g, h \in G$. Conversely, given matrices $R_g \in \mathrm{GL}_n(\mathbb{C})$ for each $g \in G$ such that $R_{gh} = R_g R_h$, then these matrices define a representation $G \to \mathrm{GL}_n(\mathbb{C})$. This is what we mean we give a representation 'in matrix form'.

***Definition* X.1.8.5.** — *An **equivariant map** or G-**linear map** $\phi\colon V \to W$ between two representations $\rho\colon G \to \mathrm{GL}(V)$ and $\sigma\colon G \to \mathrm{GL}(W)$ is a linear map $\phi\colon V \to W$ such that $\rho(s) = \phi^{-1} \circ \sigma(s) \circ \phi$ for all $s \in G$. In other words, the linear map ϕ is compatible with the G-module structures, and the following diagram commutes:*

$$\begin{array}{ccc} V & \xrightarrow{\phi} & W \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle g} \\ V & \xrightarrow{\phi} & W. \end{array}$$

*If ϕ is a linear isomorphism, then we say that V and W are **isomorphic** representations.*

***Example* X.1.8.6.** — A one-dimensional representation of a finite group G is a group homomorphism $\rho\colon G \to \mathbb{C}^\times$. Since the order of any $g \in G$ is finite, $\rho(g) \in \mathbb{C}^\times$ is a root of unity. In particular, $|\rho(s)| = 1$.

***Example* X.1.8.7.** — A representation V of a finite group G with $g = \mathrm{id}_V$ for all $g \in G$ is the **trivial** (or **unit**) representation.

***Example* X.1.8.8.** — Let $n = \#G$ be the order of a group G, and let V be an n-dimensional vector space with basis $(e_g)_{g\in G}$ indexed by the elements of G. For each $g \in G$, define $ge_h := e_{gh}$. This gives the **regular representation** of G, whose dimension is the order of G.

Since $e_g = ge_1$, the images of e_1 form a basis of V. Conversely, if W is a representation of G with a vector $w \in W$ such that $(gw)_{g\in G}$ is a basis of W, then W is isomorphic to the regular representation, namely via $\phi\colon V \to W$, $e_g \mapsto gw$.

***Example* X.1.8.9.** — More generally, suppose that a group G acts on a finite set X. This means that each $g \in G$ defines a permutation of X, satisfying $1x = x$ and $g(hx) = (gh)x$ for all $x \in X$ and all $g, h \in G$. If V is a vector space with basis $(e_x)_{x\in X}$ indexed by the elements of X, then it is also a representation of G via $ge_x = e_{gx}$ for all $g \in G$. This is the **permutation representation** of X.

***Definition* X.1.8.10.** — *Let $\rho\colon G \to \mathrm{GL}(V)$ be a representation of G, and let W be a linear subspace of V. Then W is **stable** or **invariant** under G if $gw \in W$ for all $w \in W$ and all $g \in G$. Then the restrictions $\rho_g|_W \in \mathrm{GL}(W)$ again define a representation $\rho|_W\colon G \to \mathrm{GL}(W)$ of G. We call $\rho|_W$ or W a **subrepresentation** of V.*

***Example* X.1.8.11.** — Let V be the regular representation of G (example X.1.8.8), and let $W = \mathrm{span}(\sum_{g\in G} e_g)$ which is a one-dimensional subspace of V. Since $h\sum_{g\in G} e_g = \sum_{g\in G} e_g$, we see that W is stable under G and hence defines a subrepresentation of V. Note that W is isomorphic to the trivial representation.

X.1.8.12. — Let V be a vector space with two subspaces W and W'. Then V is the *direct sum* $W \oplus W'$ of W and W' if any $v \in V$ is uniquely of the form $v = w + w'$ with $w \in W$ and $w' \in W'$. This is the same as saying that $W \cap W' = \{0\}$ and that $V = W + W'$. In this case, W' is the *complement* of W in V.

The direct sum comes with a *projection* $\pi\colon W \oplus W' \to W$, $w + w' \mapsto w$. Note that π is surjective and that $\pi|_W = \mathrm{id}_W$. Conversely, if $\pi\colon V \twoheadrightarrow W$ is a surjective linear map with $\pi|_W = \mathrm{id}_W$, then $V = W \oplus \ker\pi$ is a direct sum decomposition of V. We obtain a correspondence between projections $V \to W$ and complements of W in V.

***Maschke's theorem* X.1.8.13.** — *Let V be a representation of a finite group G, and let W be a subspace of V that is stable under G. Then there exists a complement W' of W in V that is stable under G.*

Proof. — For an arbitrary complement of W in V, let $\pi\colon V = W \oplus W' \to W$ be the corresponding projection. Consider the average

$$\pi'(v) = \frac{1}{\#G}\sum_{g\in G} g\pi(g^{-1}v)$$

of all conjugations of π by elements of G. Since W is stable under G and since π is a map $V \to W$, we see that π' is a linear map $V \to W$ as well. Moreover, since $\pi|_W = \mathrm{id}_W$, we have $\pi'|_W = \mathrm{id}_W$ again. Thus, π' is a projection and defines a complement W' of W. A direct computation shows that π' is even G-linear:

$$g\pi'(g^{-1}v) = \frac{1}{\#G}\sum_{h\in G} gh\pi(h^{-1}g^{-1}v) = \frac{1}{\#G}\sum_{h\in G} h\pi(h^{-1}v) = \pi'(v).$$

Therefore, for any $w' \in W' = \ker \pi'$ and any $g \in G$, we have $\pi'(gw') = g\pi'(w') = 0$, i. e., $gw' \in W'$ and W' is stable under G. ◻

Alternative proof. — Suppose that V comes with a complex inner product $\langle -, - \rangle$ on V, i. e., a positive definite hermitian form. For example, pick any basis of V and define the inner product by asserting that this basis is orthonormal. We can then replace the inner product by a G-invariant one by considering $(v, w) = \sum_{g\in G}\langle gv, gw\rangle$. This means that $(gv, gw) = (v, w)$ for all $v, w \in V$ and all $g \in G$, or in other words, the matrix corresponding to g is unitary. Then the *orthogonal complement* $W^\perp$ of W in V is stable under G. ◻

***Definition* X.1.8.14.** — *Let V be a representation of a finite group G. For a subrepresentation W of V, let W' be the complementary subrepresentation of W in V. We call $V = W \oplus W'$ the **direct sum** of W and W' or a **decomposition** of V into W and W'.*

***Remark* X.1.8.15.** — The G-module structure is compatible with this decomposition: writing $v \in V$ uniquely as $v = w + w'$ with $w \in W$ and $w' \in W'$, we can uniquely write $gv = gw + gw'$ with $gw \in W$ and $gw' \in W'$. Alternatively, if the action of g on W and W' is given by matrices R_g and R'_g, resp., then the action of g on V is the matrix $\mathrm{diag}(R_g, R'_g)$.

***Definition* X.1.8.16.** — *A representation V of a finite group G is **irreducible** or **simple** if $V \neq 0$ and if there are no non-trivial subrepresentations of V (the trivial ones being 0 and V itself). In other words, V is cannot be non-trivially decomposed into a direct sum of two representations.*

***Corollary* X.1.8.17.** — *Any representation is a direct sum of irreducible representations.*

Proof. — Let V be a representation of a finite group G. We perform induction on $\dim V$. If $\dim V = 0$, the $V = 0$ is the empty direct sum. Suppose that $\dim V \geq 1$. If V is irreducible, then we are done. Otherwise, by Maschke's theorem X.1.8.13 write $V = W \oplus W'$ with $\dim W, \dim W' < \dim V$, and W and W' are a direct sum of irreducible representations by induction hypothesis. ◻

***Motivation* X.1.8.18.** — Given a representation V with its decomposition $V = W_1 \oplus \cdots \oplus W_n$ into irreducible representations, we may ask if this decomposition is unique. This is generally not true: if V is the trivial representation, then there are many decompositions into lines W_i. However, the decomposition is unique up to isomorphism, i. e., any decomposition of V is essentially a permutation of the direct summands.

X.1.8.19. — Let V and W be two vector spaces. Then a bilinear map $V \times W \to V \otimes W$ is a *tensor product* if it satisfies the following universal property: if $\phi\colon V \times W \to T$ is a bilinear map, then there is a unique factorisation

$$\begin{tikzcd} V \times W \arrow[r] \arrow[dr, "\phi"'] & V \otimes W \arrow[d, "\exists!"] \\ & T. \end{tikzcd}$$

One can show that tensor products exist. If $(e_i)_i$ and $(f_j)_j$ are bases of V and W, resp., then $e_i \otimes f_j$ is a basis of $V \otimes W$, and $\dim(V \otimes W) = \dim V \dim W$.

***Construction* X.1.8.20.** — Let V and W be two representations of a finite group G. We want to endow the tensor product $V \otimes W$ of vector spaces with a G-module structure such that the canonical map $\phi\colon V \times W \to V \otimes W$ is also G-linear. This means that $g(v \otimes w) = g\phi(v, w) = \phi(gv, gw) = gv \otimes gw$ for all $v \in V$ and $w \in W$. The representation $V \otimes W$ of G thus obtained is the **tensor product** of the representations V and W.

***Remark* X.1.8.21.** — Let V and W be two representations of the group G, and let $(e_i)_i$ and $(f_k)_k$ be some bases of them, resp. Then for any $g \in G$, its matrix form on V is $(r_{ij})_{ij}$ and its matrix form on W is $(s_{kl})_{kl}$. In other words, $ge_j = \sum_i r_{ij} e_i$ and $gf_l = \sum_k s_{kl} f_k$. This implies $g(e_j \otimes f_l) = ge_j \otimes gf_l = \sum_{i,k} r_{ij} s_{kl} e_i \otimes f_k$. W. r. t. the basis $(e_i \otimes f_k)_{ik}$ of $V \otimes W$, the matrix form of g is the *Kronecker product* $(r_{ij})_{ij} \otimes (s_{kl})_{kl}$.

***Warning* X.1.8.22.** — The tensor product of two irreducible representations is in general not irreducible. One can determine the decomposition into irreducible representations by character theory.

***Example* X.1.8.23.** — We specialise to the tensor product $V \otimes V$. Let $\theta\colon V \otimes V \to V \otimes V$ be the automorphism defined by $\theta(v \otimes w) = w \otimes v$ (this can be constructed by choosing a basis of V and defining θ on this basis accordingly), which is an involution, i. e., $\theta^2 = 1$. From linear algebra, we obtain the decomposition $V \otimes V = \ker(\theta - \mathrm{id}) \oplus \mathrm{im}(\theta - \mathrm{id})$. If $(e_i)_i$ is a basis of V, then the symmetric square $\mathrm{Sym}^2 V := \ker(\theta - \mathrm{id})$ has the basis $(e_i \otimes e_j + e_j \otimes e_i)_{i \le j}$ and the alternating square $\mathrm{Alt}^2 V := \mathrm{im}(\theta - \mathrm{id})$ has the basis $(e_i \otimes e_j - e_j \otimes e_i)_{i<j}$. Their dimensions are $\dim \mathrm{Sym}^2 V = \frac{1}{2}n(n+1)$ and $\dim \mathrm{Alt}^2 V = \frac{1}{2}n(n-1)$ where $n = \dim V$.

Both $\mathrm{Sym}^2 V$ and $\mathrm{Alt}^2 V$ are clearly stable under G, and thus define the **symmetric square** and **alternating square** representations.

X.1.8.24. — Let $\mathrm{Hom}(V, W)$ be the hom space between two vector spaces V and W. In particular, this is again a vector space and satisfies the tensor-hom adjunction(X.3)

$$\mathrm{Hom}(U, \mathrm{Hom}(V, W)) \xrightarrow{\sim} \mathrm{Hom}(U \otimes V, W), \qquad \phi \longmapsto (u \otimes v \longmapsto \phi(u)(v)).$$

***Construction* X.1.8.25.** — For representations U, V and W of a finite group G, we again want the tensor-hom adjunction to be a bijection between Hom-sets of equivariant maps. The map $u \otimes v \mapsto \phi(u)(v)$ being G-linear means that $g\phi(u)(v) = \phi(gu)(gv)$. On the other hand, ϕ being G-linear means that $(g\phi(u))(v) = \phi(gu)(v)$ where $g\phi(u)$ is still undefined. Replacing v by $g^{-1}v$ in the first equation, we obtain $(g\phi(u))(v) = \phi(gu)(v) = g\phi(u)(g^{-1}v)$.

(X.3) Categorically, we want $\mathrm{Hom}(V, W)$ to still be an *internal hom* in the category of representations of G.

Thus, for representations V and W of a finite group G, we endow the Hom-space $\mathrm{Hom}(V,W)$ of vector spaces with a G-module structure via $(gf)(v) = gf(g^{-1}v)$ for all $f \in \mathrm{Hom}(V,W)$ and all $v \in V$. In other words, the following diagram commutes:

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "g"'] & W \arrow[d, "g"] \\ V \arrow[r, "gf"] & W. \end{tikzcd}$$

The representation $\mathrm{Hom}(V,W)$ of G thus obtained is the Hom-**space** between the representations V and W.

***Warning* X.1.8.26.** — The space $\mathrm{Hom}_G(V,W)$ of G-linear maps is not the same as the space $\mathrm{Hom}(V,W)$ of complex linear maps, which caries a G-module structure.

***Example* X.1.8.27.** — If we specialise to the case where $W = \mathbb{C}$ is the trivial representation of G, then we obtain the **dual representation** $V^* = \mathrm{Hom}(V,\mathbb{C})$ of the representation V of G. Its G-module structure is given by $(gf)(v) = f(g^{-1}v)$ for all $f \in V^*$ and $v \in V$.

Recall that the dual space V^* admits a natural pairing $\langle -,- \rangle\colon V^* \times V \to \mathbb{C}$, $\langle f, v\rangle = f(v)$. Then the G-action preserves this pairing: $\langle gf, gv\rangle = \langle f, v\rangle$ for all $f \in V^*$ and all $v \in V$.

***Remark* X.1.8.28.** — Let V be a representation of G with a basis $(e_i)_i$. Then $e_i^*\colon e_j \mapsto \delta_{ij}$ is a basis of V^*. For any $g \in G$, suppose that the matrix from of g^{-1} on V is $(r_{ij})_{ij}$, i.e., $g^{-1}e_j = \sum_k r_{kj}e_k$. Then $ge_i^*(e_j) = e_i^*(g^{-1}e_j) = \sum_k r_{kj}e_i^*(e_k) = r_{ij}$, so $ge_i^* = \sum_j r_{ij}e_j^*$. Thus, the matrix form of g on V^* w.r.t. the basis $(e_i^*)_i$ is given by the transpose $(r_{ji})_{ij}$.

§ X.2. CHARACTER THEORY

X.2.1. Basic properties.

X.2.1.1. — Let V be a finite-dimensional vector space with a (in fact, any) basis $(e_1, \dots, e_n)$. For an endomorphism $a\colon V \to V$, the *trade* of a is given by $\operatorname{tr} a = \operatorname{tr}(a_{ij})_{ij} = \sum_{i=1}^n a_{ii}$ where $(a_{ij})_{ij}$ is the matrix representation of a w.r.t. the basis. The trace does not depend on the chosen basis and is the sum of the eigenvalues of a (counted with multiplicities, look at the characteristic polynomial).

***Definition* X.2.1.2.** — *Let $\rho\colon G \to \mathrm{GL}(V)$ be a representation of a finite group G in a vector space V. The **character** of ρ is $\chi\colon G \to \mathbb{C}$, $\chi(g) = \operatorname{tr}\rho_g$, i.e., its trace.*

***Fact* X.2.1.3** (properties of characters)**.** — *Let χ be a character of a representation V. Then we have the following:*

(i) $\chi(1) = \dim V$;
(ii) $\chi(g^{-1}) = \overline{\chi(g)}$ *for all* $g \in G$;
(iii) $\chi(hgh^{-1}) = \chi(g)$ *for all* $g, h \in G$.

Proof. — (i) holds since $\chi(1) = \operatorname{tr}\mathrm{id}_V = \dim V$. For (ii), let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of $g \in G$, counted with multiplicities. Since $g^{\#G} = 1$, we also have $\lambda_i^{\#G} = 1$, so $|\lambda_i| = 1$. Thus, $\overline{\chi(g)} = \overline{\operatorname{tr} g} = \sum_{i=1}^n \overline{\lambda_i} = \sum_{i=1}^n \lambda_i^{-1} = \operatorname{tr} g^{-1} = \chi(g^{-1})$. Finally, the trace satisfies $\operatorname{tr}(ab) = \operatorname{tr}(ba)$, whence $\chi(hgh^{-1}) = \operatorname{tr}(hgh^{-1}) = \operatorname{tr} g = \chi(g)$, which is (iii). ∎

***Fact* X.2.1.4.** — *Let V and W be two representations of a finite group G, and let χ and θ be their characters. Then the following hold:*

(i) *the character of the direct sum representation $V \oplus W$ is $\chi + \theta$;*

(ii) *the character of the tensor product representation* $V \otimes W$ *is* $\chi\theta$.
(iii) *the character of the dual representation* V^* *is* $\overline{\chi}$.
(iv) *the character of the representation* $\operatorname{Hom}(V, W)$ *is* $\overline{\chi}\theta$.

Proof. — Let $g \in G$, and let R_g and S_g be the matrix form of g on V and W, resp. Then the representation in $V \oplus W$ is given by $\operatorname{diag}(R_g, S_g)$, whence the character is $g \mapsto \operatorname{tr} R_g + \operatorname{tr} S_g = \chi(g) + \theta(g)$. This shows (i).

Likewise for (ii), if $R_g = (r_{ij})_{ij}$ and $S_g = (s_{kl})_{kl}$, then the action of g on $V \otimes W$ is given by the Kronecker product $R_g \otimes S_g$. Then the character is $g \mapsto \operatorname{tr}(R_g \otimes S_g) = \sum_{i,k} r_{ii} s_{kk} = \chi(g)\theta(g)$.

For (iii), the action of g on V^* is given by $(R_g^{-1})^\top$. Hence, the character is $g \mapsto \operatorname{tr}(R_g^{-1})^\top = \chi(g^{-1}) = \overline{\chi(g)}$.

Finally, (iv) follows from the isomorphism $V^* \otimes W \xrightarrow{\sim} \operatorname{Hom}(V, W)$, $f \otimes w \mapsto (v \mapsto f(v)w)$ for finite-dimensional vector spaces. ◻

***Proposition* X.2.1.5.** — *Let V be a representation of a finite group G and let χ be its character. Let χ_s^2 and χ_a^2 be the characters of $\operatorname{Sym}^2 V$ and $\operatorname{Alt}^2 V$, resp. Then for each $g \in G$, we have*

$$\chi_s^2(g) = \frac{\chi(g)^2 + \chi(g^2)}{2}, \qquad \chi_a^2(g) = \frac{\chi(g)^2 - \chi(g^2)}{2}, \qquad \chi_s^2 + \chi_a^2 = \chi^2.$$

Proof. — Let $g \in G$. Recalling the alternative proof of Maschke's theorem X.1.8.13, g is unitary, and the spectral theorem says that g is diagonalisable with eigenvalues $(\lambda_1, \dots, \lambda_n)$ (counted with multiplicities) and the corresponding eigenvectors $(e_1, \dots, e_n)$ is a basis of V. This means that $ge_i = \lambda_i e_i$, so $\chi(g) = \sum_{i=1}^n \lambda_i$ and $\chi(g^2) = \sum_{i=1}^n \lambda_i^2$. Recall that $(e_i \otimes e_j + e_j \otimes e_i)_{i \le j}$ is a basis of $\operatorname{Sym}^2 V$ and that $(e_i \otimes e_j - e_j \otimes e_i)_{i<j}$ is a basis of $\operatorname{Alt}^2 V$. The eigenvalues of these vectors is $g(e_i \otimes e_j \pm e_j \otimes e_i) = \lambda_i\lambda_j(e_i \otimes e_j \pm e_j \otimes e_i)$, and hence

$$\chi_s^2(g) = \sum_{i \le j} \lambda_i\lambda_j = \sum_i \lambda_i^2 + \sum_{i<j} \lambda_i\lambda_j = \frac{(\sum_i \lambda_i)^2 + \sum_i \lambda_i^2}{2} = \frac{\chi(g)^2 + \chi(g^2)}{2},$$
$$\chi_a^2(g) = \sum_{i<j} \lambda_i\lambda_j = \frac{(\sum_i \lambda_i)^2 - \sum_i \lambda_i^2}{2} = \frac{\chi(g)^2 - \chi(g^2)}{2}.$$ ◻

***Remark* X.2.1.6.** — The formula $\chi_s^2 + \chi_a^2 = \chi^2$ mirrors the decomposition $V \otimes V = \operatorname{Sym}^2 V \oplus \operatorname{Alt}^2 V$.

***Corollary* X.2.1.7.** — *Let χ and θ be the characters of two representations. Then we have*

$$(\chi + \theta)_s^2 = \chi_s^2 + \theta_s^2 + \chi\theta, \qquad (\chi + \theta)_a^2 = \chi_a^2 + \theta_a^2 + \chi\theta.$$

Proof. — Let $g \in G$. We calculate

$$\begin{aligned}(\chi + \theta)_s^2(g) = \frac{(\chi + \theta)(g)^2 + (\chi + \theta)(g^2)}{2} &= \frac{\chi(g)^2 + \theta(g)^2 + 2\chi\theta(g) \pm \chi(g^2) \pm \theta(g^2)}{2} \\ &= \chi_s^2(g) + \theta_s^2(g) + \chi\theta(g)\end{aligned}$$

and similarly for $(\chi + \theta)_a^2(g)$. ◻

Alternative proof. — Let χ and θ be the characters of the representations V and W, resp. It suffices to show that $\operatorname{Sym}^2(V \oplus W) \cong \operatorname{Sym}^2 V \oplus \operatorname{Sym}^2 W \oplus V \otimes W$. Let $(e_i)_i$ and $(f_k)_k$ be bases of V and W, resp. Then a basis of the right-hand side consists of the bases of $\operatorname{Sym}^2 V$ and $\operatorname{Sym}^2 W$, and additionally $(e_i \otimes f_k)_{ik}$. A basis of the left-hand side consists of the bases of $\operatorname{Sym}^2 V$ and $\operatorname{Sym}^2 W$, but additionally $(e_i \otimes f_k + f_k \otimes e_i)_{ik}$. Hence, we obtain the isomorphism of vector spaces. Moreover, this isomorphism is equivariant, so we have an isomorphism of representations and the formula for characters follows. A similar argument works for $\operatorname{Alt}^2(V \oplus W)$. ◻

***Corollary* X.2.1.8.** — *Let V and W be two irreducible representations of a finite group G. Let $f\colon V \to W$ be a linear map of vector spaces, and put*

$$f'\colon V \longrightarrow W, \qquad f'(v) = \frac{1}{\#G}\sum_{g\in G} g^{-1}f(gv).$$

Then the following hold:

(i) *If $V \not\cong W$, then $f' = 0$.*
(ii) *If $V = W$, then f' is a homothety of ratio $\operatorname{tr} f/\dim V$.*

Proof. — Firstly, f' is equivariant since

$$h^{-1}f'(hv) = \frac{1}{\#G}\sum_{g\in G} h^{-1}g^{-1}f(ghv) = f'(v)$$

for all $h \in G$ and all $v \in V$. By Schur's lemma X.1.4.21, we have $f' = 0$ in (i), and $f' = \lambda \mathrm{id}_V$ in (ii). Moreover, in the latter case where $f' = \lambda \mathrm{id}_V$, we have

$$\lambda \dim V = \operatorname{tr} f' = \frac{1}{\#G}\sum_{g\in G} \operatorname{tr}(g^{-1}\circ f \circ g) = \operatorname{tr} f,$$

whence $\lambda = \operatorname{tr} f/\dim V$. ◻

***Corollary* X.2.1.9.** — *Let V and W be two irreducible representations of a finite group G. Suppose that the representations are given in matrix forms $(r_{ij}(g))_{ij}$ on V and $(s_{kl}(g))_{kl}$ on W for each $g \in G$. Then we have the following:*

(i) *If $V \not\cong W$, then $(1/\#G)\sum_{g\in G} s_{kl}(g^{-1})r_{ji}(g) = 0$ for any indices i, j, k, l.*
(ii) *If $V = W$, then $(1/\#G)\sum_{g\in G} s_{kl}(g^{-1})r_{ji}(g) = \delta_{ki}\delta_{lj}/\dim V$.*

Proof. — In order to apply corollary X.2.1.8, suppose that a linear map $f\colon V \to W$ is in matrix form $(x_{ki})_{ki}$ and write f' in matrix form $(x'_{ki})_{ki}$ such that

$$x'_{ki} = \frac{1}{\#G}\sum_{\substack{g\in G\\ l,j}} s_{kl}(g^{-1})x_{lj}r_{ji}(g).$$

In (i), we have $x'_{ki} = 0$ for any given f, which is only possible if (i) holds. In (ii), we have $x'_{ki} = \delta_{ki}\operatorname{tr} f/\dim V$ with $\operatorname{tr} f = \sum_{l,j}\delta_{lj}x_{lj}$. Since this hold for any given f, both expressions for x'_{ki} must have the same coefficients in x_{lj}, i. e., (ii) holds. ◻

X.2.2. Inner product and uniqueness of decompositions.

***Notation* X.2.2.1.** — For functions $\phi, \psi\colon G \to \mathbb{C}$, we define

$$\langle \phi, \psi\rangle = \frac{1}{\#G}\sum_{g\in G}\phi(g)\overline{\psi(g)},$$

which is an inner product: it is linear in ϕ, anti-linear in ψ, hermitian and positive definite. Moreover, we define

$$(\phi, \psi) = \frac{1}{\#G}\sum_{g\in G}\phi(g^{-1})\psi(g) = \frac{1}{\#G}\sum_{g\in G}\phi(g)\psi(g^{-1}),$$

which is a symmetric positive-definite bilinear form.

***Remark* X.2.2.2.** — If $\chi\colon G \to \mathbb{C}$ is a character so that $\chi(g^{-1}) = \overline{\chi(g)}$, then $\langle \phi, \chi\rangle = (\phi, \chi)$ for all functions $\phi\colon G \to \mathbb{C}$. In this case, both products agree. If θ is also a character, then $\overline{\langle \chi, \theta\rangle} = \langle \theta, \chi\rangle = (\theta, \chi) = \langle \chi, \theta\rangle$ and the product $\langle \chi, \theta\rangle$ is real. In other words, the inner product of two characters is symmetric.

***Orthogonality relations* X.2.2.3.** — (i) *Two characters χ and θ of non-isomorphic irreducible representations are orthogonal, i. e., $\langle \chi, \theta\rangle = 0$.*

(ii) *Any character χ of an irreducible representation has norm one, i. e., $\langle \chi, \chi\rangle = 1$.*

Proof. — Let $(r_{ij}(g))_{ij}$ and $(s_{kl}(g))_{kl}$ be the irreducible representations V and W in matrix form corresponding to the characters χ and θ, resp. Then corollary X.2.1.9 becomes $(s_{kl}, r_{ji}) = 0$ if $V \ncong W$ and $(s_{kl}, r_{ji}) = \delta_{ki}\delta_{lj}/\dim V$ if $V = W$. Since $\chi(g) = \sum_i r_{ii}(g)$, this implies that

$$\langle \chi, \theta\rangle = (\chi, \theta) = \sum_{i,k}(r_{ii}, s_{kk}) = 0, \qquad \langle \chi, \chi\rangle = (\chi, \chi) = \sum_{i,j}(r_{ii}, r_{ji}) = \sum_{i,j}\frac{\delta_{ij}}{\dim V} = 1.$$

The following statement proves the uniqueness of the decomposition into a direct sum of irreducible representations, cf. Maschke's theorem (motivation X.1.8.18).

***Theorem* X.2.2.4** (uniqueness of decompositions)**.** — *Let V be a representation of a finite group G with character χ, and let $V = W_1 \oplus \cdots \oplus W_k$ be a decomposition into irreducible representations. If W is an irreducible representation of G with character θ, then the number of W_i isomorphic to W is equal to $\langle \chi, \theta\rangle = (\chi, \theta)$. This number is the same for any decomposition of V.*

Proof. — Let χ_i be the character of W_i. Then $\langle \chi, \theta\rangle = \langle \chi_1, \theta\rangle + \cdots + \langle \chi_k, \theta\rangle$, and $\langle \chi_i, \theta\rangle \in \{0, 1\}$ depending on whether $W_i \cong W$ by the orthogonality relations X.2.2.3.

***Corollary* X.2.2.5.** — *Representations of G with the same character are isomorphic.*

Proof. — Given two representations V and W with the same character, their irreducible subrepresentations are in one-to-one correspondence to each other. Thus, up to isomorphism of the direct summands, the decompositions of V and W are just permutations of the same direct summands.

Definition* X.2.2.6.** — *A character χ of a representation V of a finite group G is* ***irreducible *if V is irreducible.*

***Irreducibility criterion* X.2.2.7.** — *Let χ be the character of a representation V of G. Then $\langle \chi, \chi\rangle \in \mathbb{N}$. Moreover, $\langle \chi, \chi\rangle = 1$ if and only if V is irreducible.*

Proof. — Write $V \cong W_1^{\oplus n_1} \oplus \cdots \oplus W_k^{\oplus n_k}$ where $W_1, \dots, W_k$ are mutually non-isomorphic irreducible representations of G and where $n_1, \dots, n_k \in \mathbb{N}$. If χ_i is the character of W_i, then $\langle \chi, \chi_i\rangle = n_i$. The orthogonality relations X.2.2.3 imply that $\langle \chi, \chi\rangle = \sum_{i=1}^k n_i^2 \in \mathbb{N}$. Moreover, $\langle \chi, \chi\rangle = 1$ if and only if we have $n_i = 1$ for exactly one i and $n_j = 0$ for all other j. This means that $V \cong W_i$ is irreducible by theorem X.2.2.4.

***Lemma* X.2.2.8.** — *The character ρ of the regular representation of G is $\rho(1) = \#G$ and $\rho(g) = 0$ for all $g \in G \setminus \{1\}$.*

Proof. — Recall that R admits a basis $(e_g)_{g\in G}$ such that $ge_h = e_{gh}$. If $g \in G \setminus \{1\}$, then $\rho(g) = \operatorname{tr} g = 0$. If $g = 1$, then $\rho(1) = \dim R = \#G$.

***Corollary* X.2.2.9.** — *Every irreducible representation of G is isomorphic to a subrepresentation of the regular representation of G, whose multiplicity equals its dimension. In other words, the regular representation R has the decomposition $R \cong W_1^{\oplus \dim W_1} \oplus \cdots \oplus W_k^{\oplus \dim W_k}$ where the W_i represent all mutually non-isomorphic irreducible representations of G up to isomorphism. In particular, there are only finitely many irreducible representations.*

Proof. — Let ρ be the character of R. Let $R \cong W_1^{\oplus n_1} \oplus \cdots \oplus W_k^{\oplus n_k}$ be the decomposition of R into mutually non-isomorphic irreducible representations W_i. Let W be an irreducible representation with character χ. By lemma X.2.2.8, we have

$$\langle \rho, \chi \rangle = (\rho, \chi) = \frac{1}{\#G} \sum_{g \in G} \rho(g^{-1})\chi(g) = \chi(1) = \dim W,$$

so there is some i with $W_i \cong W$ and $n_i = \dim W$ by theorem X.2.2.4. ∎

***Corollary* X.2.2.10.** — *Let $W_1, \dots, W_k$ be the mutually non-isomorphic irreducible representations of G with characters $\chi_1, \dots, \chi_k$. Then we have the following:*

(i) $\sum_{i=1}^k (\dim W_i)^2 = \#G$;

(ii) $\sum_{i=1}^k (\dim W_i)\chi_i(g) = 0$ *for all* $g \in G \setminus \{1\}$.

Proof. — Firstly, the regular representation R has the decomposition $R \cong W_1^{\oplus \dim W_1} \oplus \cdots \oplus W_k^{\oplus \dim W_k}$. Then the character ρ of R is $\rho = \sum_{i=1}^k (\dim W_i)\chi_i$. Then $\#G = \rho(1) = \sum_{i=1}^k (\dim W_i)^2$ and $0 = \rho(g) = \sum_{i=1}^k (\dim W_i)\chi_i(g)$ for all $g \neq 1$. ∎

***Remark* X.2.2.11.** — This can be used to make sure that we found all irreducible representations of a finite group G. Suppose that we have constructed k mutually non-isomorphic irreducible representations of dimensions $n_1, \dots, n_k$. Then these are *all* irreducible representations if and only if $n_1^2 + \cdots + n_k^2 = \#G$.

X.2.3. Class functions.

***Definition* X.2.3.1.** — *A function $\phi \colon G \to \mathbb{C}$ is a **class function** of G if $f(hgh^{-1}) = f(g)$ for all $g, h \in G$. The set of all class functions of G is denoted by* $\operatorname{Class} G$*; it forms a complex vector space.*

***Example* X.2.3.2.** — Characters are class functions.

***Lemma* X.2.3.3.** — *Let $\phi \colon G \to \mathbb{C}$ be a class function, and let V be an irreducible representation of G with character χ. Then the endomorphism $\sum_{g \in G} \phi(g) g$ on V is a homothety of ratio*

$$\frac{1}{\dim V} \sum_{g \in G} \phi(g)\chi(g) = \frac{\#G}{\dim V} \langle \phi, \overline{\chi} \rangle.$$

Proof. — Let $\rho = \sum_{g \in G} \phi(g) g$. Then

$$h^{-1}\rho(hv) = \sum_{g \in G} \phi(g) h^{-1}ghv = \sum_{g \in h^{-1}Gh} \phi(hgh^{-1}) gv = \sum_{g \in G} \phi(g) gv = \rho(v)$$

for all $h \in G$ and all $v \in V$ since ϕ is a class function, so ρ is equivariant. By Schur's lemma X.1.4.21, we see that ρ is a homothety of ratio $\lambda \in \mathbb{C}$. Taking traces gives

$$\lambda \dim V = \operatorname{tr}(\lambda \operatorname{id}_V) = \operatorname{tr} \rho = \sum_{g \in G} \phi(g)\chi(g) = \#G \langle \phi, \overline{\chi} \rangle.$$ ∎

***Proposition* X.2.3.4.** — *The irreducible characters $\chi_1, \dots, \chi_k$ form an orthonormal basis of* $\operatorname{Class} G$.

Proof. — By the orthogonality relations X.2.2.3, we know that the characters $(\chi_1, \dots, \chi_k)$ form an orthonormal system. Then it is automatically linearly independent, e. g., $\chi_k = \lambda_1\chi_1 + \dots + \lambda_k\chi_k$ with $\lambda_i \in \mathbb{C}$ implies $\lambda_i = \langle \chi_k, \chi_i \rangle = 0$ for all i. It remains to see that the characters generate all of $\operatorname{Class} G$. To this end, let $\phi \in \operatorname{Class} G$ such that $\langle \phi, \overline{\chi_i} \rangle = 0$ for all i, and we need to show that $\phi = 0$.

Let V be a representation of G, and define $\rho = \sum_{g\in G} \phi(g)g$. Since ϕ is orthogonal to all $\overline{\chi_i}$ and since ρ restricts to any irreducible subrepresentation of V by the direct sum decomposition, we see that $\rho = 0$ by lemma X.2.3.3. In particular, if V is the regular representation with standard basis $(e_g)_{g\in G}$, then $\rho(e_1) = \sum_{g\in G} \phi(g)e_g = 0$. It follows that $\phi(g) = 0$ for all $g \in G$.

X.2.3.5. — Recall that two elements $g, g' \in G$ are *conjugate* if $g' = hgh^{-1}$ for some $h \in G$. This defines an equivalence relation, and the equivalence classes are called *conjugacy classes*. These partition G into disjoint sets.

***Theorem* X.2.3.6.** — *The number of irreducible representations of G is the number of conjugacy classes of G.*

Proof. — Let $C_1, \dots, C_k$ be the distinct conjugacy classes of G. Then the class functions $\operatorname{Class} G$ are precisely the functions $f\colon G \to \mathbb{C}$ that are constant on each C_i. Hence, $\operatorname{Class} G \cong \mathbb{C}^k$. On the other hand by proposition X.2.3.4, $\dim \operatorname{Class} G$ is the number of irreducible representations of G.

The following generalises corollary X.2.2.10 beyond $g = 1$.

***Proposition* X.2.3.7.** — *Let $\chi_1, \dots, \chi_k$ be the irreducible representations of G. Then we have the following:*

(i) $\sum_{i=1}^k \chi_i(g)\overline{\chi_i(g)} = \#G/\#\operatorname{Cl}(g)$ *for all* $g \in G$*;*

(ii) $\sum_{i=1}^k \chi_i(g)\overline{\chi_i(h)} = 0$ *whenever* $g, h \in G$ *are not conjugate.*

Proof. — Let $\mathbf{1}_{\operatorname{Cl}(g)} \in \operatorname{Class} G$ be 1 on $\operatorname{Cl}(g)$ and 0 anywhere else. By proposition X.2.3.4, we can write

$$\mathbf{1}_{\operatorname{Cl}(g)} = \sum_{i=1}^k \langle \mathbf{1}_{\operatorname{Cl}(g)}, \chi_i \rangle \chi_i = \frac{1}{\#G} \sum_{i=1}^k \sum_{h\in\operatorname{Cl}(g)} \overline{\chi_i(h)}\chi_i = \frac{\#\operatorname{Cl}(g)}{\#G} \sum_{i=1}^k \overline{\chi_i(g)}\chi_i$$

since χ_i is also constant on conjugacy classes. We obtain (i) from $\mathbf{1}_{\operatorname{Cl}(g)}(g) = 1$ and (ii) from $\mathbf{1}_{\operatorname{Cl}(g)}(h) = 0$ for all $h \notin \operatorname{Cl}(g)$.

***Example* X.2.3.8** (character tables)**.** — Recall that $G = S_3$ is the permutation group of three objects. It is generated by a rotation σ and a transposition τ, i. e., $\tau^2 = \sigma^3 = 1$ and $\tau\sigma = \sigma^2\tau$. Thus, $S_3 = \{1, \sigma, \sigma^2, \tau, \sigma\tau, \sigma^2\tau\}$ is of order 6.

Since one-dimensional (hence irreducible) representations agree with their characters, the one-dimensional (irreducible) characters χ are group homomorphisms $G \to \mathbb{C}^\times$. The relations imply $\chi(\tau) \in \{\pm 1\}$ and $\chi(\sigma) \in \{1, \mathrm{e}^{2\pi\mathrm{i}/3}, \mathrm{e}^{4\pi\mathrm{i}/3}\} \cap \{0, 1\} = \{1\}$. Thus, the only one-dimensional characters are the unit character $\chi_1 = 1$ and the sign $\chi_2 = \operatorname{sgn}$ with $\chi_2(\tau) = -1$.

Since there are three conjugacy classes $\operatorname{Cl}(1) = \{1\}$, $\operatorname{Cl}(\tau) = \{\tau, \sigma\tau, \sigma^2\tau\}$, $\operatorname{Cl}(\sigma) = \{\sigma, \sigma^2\}$, there is a third irreducible character θ by theorem X.2.3.6. If its dimension is n, then $1 + 1 + n^2 = 6$ by corollary X.2.2.10, so $n = 2$. By corollary X.2.2.9, the character of the regular representation is $\chi_1 + \chi_2 + 2\theta$, which is explicitly described by lemma X.2.2.8. Thus, table X.1 shows the **character table** of S_3.

For an explicit representation realising θ, consider the action of G on $\{(x, y, z) \in \mathbb{C}^3 \mid x+y+z = 0\}$ permuting the variables.

	1	τ	σ
χ_1	1	1	1
χ_2	1	-1	1
θ	2	0	-1

Table X.1: Character table of S_3.

***Proposition* X.2.3.9.** — *Let $W_1, \dots, W_k$ be the mutually non-isomorphic irreducible representations of G with characters $\chi_1, \dots, \chi_k$. Let $V \cong W_1^{\oplus n_1} \oplus \dots \oplus W_k^{\oplus n_k}$ be the decomposition of a representation of G into irreducible representations. Then the decomposition $V = V_1 \oplus \dots \oplus V_k$ with $V_i \cong W_i^{\oplus n_i}$ is unique up to permutation, and the projections $\pi_i \colon V \twoheadrightarrow V_i$ are given by*

$$\pi_i = \frac{\dim W_i}{\# G} \sum_{g \in G} \overline{\chi_i(g)} g.$$

This decomposition $V = V_1 \oplus \dots \oplus V_k$ is called the ***canonical decomposition****.*

Proof. — Let $p_i = \dim W_i / \# G \sum_{g \in G} \overline{\chi_i(g)} g$. Let W be an irreducible subrepresentation of V with character χ. By lemma X.2.3.3, we have

$$p_i|_W = \frac{\dim W_i}{\# G \dim W} \sum_{g \in G} \overline{\chi_i(g)} \chi(g) \mathrm{id}_W = \frac{\dim W_i}{\dim W} \langle \chi, \chi_i \rangle \mathrm{id}_W = \begin{cases} \mathrm{id}_W & \text{if } \chi = \chi_i, \\ 0 & \text{if } \chi \neq \chi_i. \end{cases}$$

If $W \cong W_i$, then we must have $W \subseteq V_i$; otherwise, W_i would be isomorphic to an irreducible subrepresentation of $W_j^{\oplus n_j}$ for some $j \neq i$. It follows that $p_i|_{V_j} = \delta_{ij} \mathrm{id}_{V_i}$ and thus, $p_i = \pi_i$. This also shows that the decomposition $V = V_1 \oplus \dots \oplus V_k$ is unique since p_i does not involve any choice of decompositions of V, i. e., it is intrinsic to V. ◻

***Observation* X.2.3.10.** — To determine a decomposition of V into irreducible representations, we first determine the canonical decomposition $V = V_1 \oplus \dots \oplus V_n$, e. g., via the formula in proposition X.2.3.9. If $W_1, \dots, W_k$ are the mutually non-isomorphic irreducible representations, then $V_i \cong W_i^{\oplus n_i}$. We can then further decompose each V_i into irreducible subrepresentations $V_i = U_{i,1} \oplus \dots \oplus U_{i,n_i}$ with $U_{i,j} \cong W_i$. The decomposition of V_i is, however, not unique, akin to an arbitrary choice of basis for a vector space.

To see this, note that any choice of isomorphism $W_i^{\oplus n_i} \xrightarrow{\sim} V_i$ gives maps $f_j \colon W_i \hookrightarrow W_i^{\oplus n_i} \xrightarrow{\sim} V_i$ where $\iota_j \colon W_i \hookrightarrow W_i^{\oplus n_i}$ is the inclusion of the jth summand. The maps $(f_1, \dots, f_{n_i})$ are linearly independent and form a basis of $\mathrm{Hom}_G(W_i, V_i)$ since $\mathrm{Hom}_G(W_i, V_i) \cong \mathrm{Hom}_G(W_i, W_i)^{\oplus n_i} \cong \mathbb{C}^{n_i}$ as vector spaces by Schur's lemma X.1.4.21. Conversely, any choice of basis $(f_1, \dots, f_{n_i})$ of $\mathrm{Hom}_G(W_i, V_i)$ defines an isomorphism $\mathbb{C}^{n_i} \xrightarrow{\sim} \mathrm{Hom}_G(W_i, V_i)$, where $\mathrm{Hom}_G(W_i, V_i)$ is a trivial representation of G. This gives rise to an isomorphism

$$\begin{aligned} W_i^{\oplus n_i} &\xrightarrow{\sim} \mathbb{C}^{n_i} \otimes W_i \xrightarrow{\sim} \mathrm{Hom}_G(W_i, V_i) \otimes W_i \xrightarrow{\sim} V_i, \\ (w_1, \dots, w_{n_i}) &\longmapsto \sum_{j=1}^{n_i} 1 \otimes w_j \longmapsto \sum_{j=1}^{n_i} f_j \otimes w_j \longmapsto \sum_{j=1}^{n_i} f_j(w_j). \end{aligned}$$

Thus, the possible decompositions of V_i correspond to the choices of bases of $\mathrm{Hom}_G(W_i, V_i)$.

One can make this explicit, see §2.7 Serre.

CHAPTER XI

LIE ALGEBRAS

§ XI.1. SEMISIMPLE LIE ALGEBRAS

XI.1.1. Definitions.

***Definition* XI.1.1.1.** — *A **Lie algebra** over a field k is a k-vector space $\mathfrak{g}$*[(XI.1)] *together with a bilinear operation called **Lie bracket** $[-,-]\colon \mathfrak{g}\times\mathfrak{g}\to\mathfrak{g}$ such that:*

(i) $[x,x]=0$ *for all* $x\in\mathfrak{g}$*;*
(ii) ***Jacobi identity***: $[x,[y,z]]+[y,[z,x]]+[z,[x,y]]=0$ *for all* $x,y,z\in\mathfrak{g}$.

*The **dimension** of $\mathfrak{g}$ is the dimension of the underlying vector space.*

***Fact* XI.1.1.2.** — *The Lie bracket is anti-commutative: $[x,y]=-[y,x]$ for all $x,y\in\mathfrak{g}$. If* $\operatorname{char} k\neq 2$*, then anti-commutativity implies $[x,x]=0$.*

Proof. — If $[x,x]=0$, then $0=[x+y,x+y]=[x,y]+[y,x]$. ∎

***Remark* XI.1.1.3.** — Any finite-dimensional Lie algebra $\mathfrak{g}$ with basis $\{e_1,\dots,e_n\}$ is completely determined by the k-vector space $\mathfrak{g}$ and the constants $a_{ij}^m\in k$ appearing $[e_i,e_j]=\sum_{m=1}^n a_{ij}^m e_m$, called the *structure constants* of $\mathfrak{g}$. By anti-commutativity, it even suffices to give the constants for $i<j$.

Conversely, we can endow a Lie algebra structure on a k-vector space $\mathfrak{g}$ by giving structure constants $(a_{ij}^m)_{ijm}$ subject to the conditions $a_{ii}^m=0=a_{ij}^m+a_{ji}^m$ and $\sum_m(a_{ij}^m a_{kl}^p+a_{jl}^m a_{ki}^p+a_{li}^m a_{kj}^p)=0$. We will, however, almost never use this approach to define Lie algebras.

***Definition* XI.1.1.4.** — *A Lie algebra $\mathfrak{g}$ is **abelian** if the Lie bracket $[-,-]$ is trivial, i. e., $[x,y]=0$ for all $x,y\in\mathfrak{g}$.*

***Definition* XI.1.1.5.** — *A **Lie algebra homomorphism** over k is a k-linear map $\phi\colon\mathfrak{g}\to\mathfrak{h}$ between Lie algebras over k such that $\phi([x,y])=[\phi(x),\phi(y)]$ for all $x,y\in\mathfrak{g}$. If ϕ is an isomorphism of vector spaces (i. e., bijective), then ϕ is called a **Lie algebra isomorphism**.*

***Convention* XI.1.1.6.** — *For brevity, we will always say* Lie algebra map *instead of Lie algebra homomorphism.*

***Convention* XI.1.1.7.** — *All Lie algebras are* finite-dimensional*, unless otherwise stated.*

(XI.1) A more general definition considers A-modules $\mathfrak{g}$ instead of k-vector spaces.

Definition* XI.1.1.8.** — *A linear subspace $\mathfrak{s}$ of a Lie algebra $\mathfrak{g}$ over k is a* ***Lie subalgebra *over k if it is closed under the bracket of $\mathfrak{g}$. Then $\mathfrak{s}$ is again a Lie algebra and the inclusion $\iota\colon \mathfrak{s} \hookrightarrow \mathfrak{g}$ is a Lie algebra map.*

If $\mathfrak{s}$ and $\mathfrak{t}$ are subalgebras of $\mathfrak{g}$, then $[\mathfrak{s}, \mathfrak{t}]$ denotes the subalgebra generated by brackets $[s,t]$ for all $s \in \mathfrak{s}$ and all $t \in \mathfrak{t}$.

***Example* XI.1.1.9.** — For a Lie algebra $\mathfrak{g}$, any $x \in \mathfrak{g}$ defines a one-dimensional Lie subalgebra $\operatorname{span} x$ with the trivial bracket.

Maybe this should be a convention.

We are mainly concerned with Lie algebras whose underlying vector space is finite-dimensional.

***Example* XI.1.1.10.** — For a k-vector space V, recall that $\operatorname{End} V$ is the k-algebra of all linear maps $V \to V$. We can endow $\operatorname{End} V$ with a Lie algebra structure by virtue of the *commutator* $[x,y] = x \circ y - y \circ x$. A pure computation shows that the commutator satisfies the Jacobi identity.

This turns $\mathfrak{gl}(V) := \operatorname{End} V$ into a Lie algebra over k called the **general linear Lie algebra** of V (closely associated to $\mathfrak{gl}(V)$, the k-algebra of all automorphisms of V). By extension, any Lie subalgebra of $\mathfrak{gl}(V)$ is a **linear Lie algebra**.

***Example* XI.1.1.11.** — If V is an n-dimensional k-vector space, any choice of basis of V gives an isomorphism $\mathfrak{gl}(V) \cong \operatorname{Mat}_n(k)$, and we will denote it by $\mathfrak{gl}_n(k) := \operatorname{Mat}_n(k)$. It has dimension n^2. If e_{ij} is the standard basis element of $\mathfrak{gl}_n(k)$ with a single 1 in the ith row and jth column, then $e_{ij}e_{kl} = \delta_{jk}e_{il}$, so $[e_{ij}, e_{kl}] = \delta_{jk}e_{il} - \delta_{li}e_{kj} \in \{-1, 0, 1\}$.

***Example* XI.1.1.12.** — We now describe four types A_n, B_n, C_n and D_n, where $n \leq 1$, of *classical linear Lie algebras*, which are subalgebras of the general linear Lie algebra and correspond to classical linear Lie groups. They will play a central role later.

(A_n) Let V be a $(n+1)$-dimensional k-vector space. Then the **special linear Lie algebra** is

$$\mathfrak{sl}(V) := \{x \in \mathfrak{gl}(V) \mid \operatorname{tr} x = 0\} \cong \mathfrak{sl}_{n+1}(k),$$

which is closely related to the special linear group $\mathfrak{sl}(V)$ consisting of all $x \in \mathfrak{gl}(V)$ with $\det x = 1$. Indeed, $\mathfrak{sl}(V)$ is a subalgebra of $\mathfrak{gl}(V)$ since $\operatorname{tr}[x,y] = \operatorname{tr}(xy) - \operatorname{tr}(yx) = 0$. Moreover, since $\mathfrak{sl}(V)$ is a proper subalgebra, its dimension is at most $\dim \mathfrak{gl}(V) - 1 = (n+1)^2 - 1$. On the other hand, the set of elements e_{ij} where $i \neq j$ together with $e_{ii} - e_{i+1,i+1}$ is linearly independent, which forms a basis of $\mathfrak{sl}_{n+1}(k) \cong \mathfrak{sl}(V)$. In particular, $\mathfrak{sl}(V)$ has dimension $(n+1)^2 - 1$.

(C_n) Let V be a $2n$-dimensional k-vector space, and let $f\colon V \times V \to k$ be the non-degenerate skew-symmetric, i. e., $f(v,w) = -f(w,v)$, bilinear form given by

$$s = \begin{pmatrix} & \mathbf{1}_n \\ -\mathbf{1}_n & \end{pmatrix}$$

after a choice of basis for V. Note that $\dim V$ must be even in order for such a bilinear form to exist.(XI.2) Then the **symplectic Lie algebra** is

$$\mathfrak{sp}(V) := \{x \in \mathfrak{gl}(V) \mid f(x(v), w) = -f(v, x(w)) \text{ for all } v, w \in V\} \cong \mathfrak{sp}_{2n}(k).$$

Under the chosen basis, the symplectic condition for $x \in \mathfrak{sp}_{2n}(k)$ is equivalent to $sx = -x^\top s$. This verifies that $\mathfrak{sp}(V)$ is a subalgebra of $\mathfrak{gl}(V)$ since $s[x,y] = x^\top y^\top s - y^\top x^\top s = -[x,y]^\top s$.

Writing $x = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a, b, c, d \in \mathfrak{gl}_n(k)$, we obtain $b^\top = b$, $c^\top = c$ and $a^\top = -d$. The latter means that $\operatorname{tr} x = 0$, i. e., $\mathfrak{sp}(V)$ is a subalgebra of $\mathfrak{sl}(V)$. With these relations, a basis is given by:

(XI.2) Let $f\colon V \times V \to k$ be a non-degenerate skew-symmetric bilinear form. Note that by skew-symmetry, $f(v,v) = 0$ for all $v \in V$ if $\operatorname{char} k \neq 2$. Let $v \in V$ be non-zero. Then $f(v,w) \neq 0$ for some $w \in V$ by non-degeneracy, and f is non-degenerate if restricted to $W = \operatorname{span}(v,w)$. Hence, we have the orthogonal decomposition $V = W \oplus W^\perp$ and $f|_{W^\perp}$ is non-degenerate. By induction, V is has even dimension.

- $e_{ij} - e_{n+j,n+i}$ for $i, j \leq n$, generating all x with $b = c = 0$ and $a^\top = -d$;
- $e_{i,n+i}$ for $i \leq n$ and $e_{i,n+j} + e_{j,n+i}$ for $i < j \leq n$, generating all symmetric b; and similarly
- $e_{n+i,i}$ for $i \leq n$ and $e_{n+j,i} + e_{n+i,j}$ for $i < j \leq n$, generating all symmetric c.

Hence, the dimension of $\mathfrak{sp}(V)$ is $n^2 + 2(n + (n^2 - n)/2) = 2n^2 + n$.

(B_n) Similarly, let V be a $(2n+1)$-dimensional k-vector space, and let $f\colon V \times V \to k$ be the non-degenerate symmetric, i. e., $f(v, w) = f(w, v)$, bilinear form given by

$$s = \begin{pmatrix} 1 & & \\ & & \mathbf{1}_n \\ & \mathbf{1}_n & \end{pmatrix}.$$

Then the **orthogonal Lie algebra** is

$$\mathfrak{o}(V) := \{x \in \mathfrak{gl}(V) \mid f(x(v), w) = -f(v, x(w)) \text{ for all } v, w \in V\} \cong \mathfrak{o}_{2n+1}(k),$$

which is the same condition as in the symplectic case. Writing

$$x = \begin{pmatrix} p & q_1 & q_2 \\ r_1 & a & b \\ r_2 & c & d \end{pmatrix}$$

in the same way as s, the symplectic relation $sx = -x^\top s$ gives $p = 0$, $q_1^\top = -r_2$, $q_2^\top = -r_1$, $a^\top = -d$, $b^\top = -b$ and $c^\top = -c$. In particular, $\mathfrak{o}(V) \subseteq \mathfrak{sl}(V)$. To determine a basis, we first consider elements with $\begin{pmatrix} r_1 & f_2 \end{pmatrix}^\top = \begin{pmatrix} q_1 & q_2 \end{pmatrix} = 0$. Since the submatrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ of x is subject to similar relations as in the symplectic case, we can adapt the symplectic basis (note that b and c are skew-symmetric now):

- $e_{ij} - e_{n+j,n+i}$ for $2 \leq i, j \leq n+1$;
- $e_{i,n+j} - e_{j,n+i}$ for $2 \leq i < j \leq n+1$;
- $e_{n+j,i} - e_{n+i,j}$ for $2 \leq i < j \leq n+1$.

Adding $e_{1,n+i} - e_{i1}$ and $e_{1i} - e_{n+i,1}$ for $2 \leq i \leq n+1$, which are the matrices with $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = 0$, we arrive at a basis for $\mathfrak{o}(V)$. Its dimension is $2n^2 + n$.

(D_n) Let V be a $2n$-dimensional k-vector space, and let $f\colon V \times V \to k$ be the non-degenerate symmetric bilinear form given by

$$s = \begin{pmatrix} & \mathbf{1}_n \\ \mathbf{1}_n & \end{pmatrix}.$$

The Lie algebra of this type is again the *orthogonal Lie algebra* $\mathfrak{o}(V) \cong \mathfrak{o}_{2n}(k)$. Writing $x = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathfrak{o}(V)$, so that $sx = -x^\top s$, we similarly obtain $a^\top = -d$, $b^\top = -b$, $c^\top = -c$. Similar to the orthogonal case, the dimension of $\mathfrak{o}(V)$ is $2n^2 - n$.

***Example* XI.1.1.13.** — The **upper triangular matrices**, **strictly upper triangular matrices** and **diagonal matrices**

$$\begin{aligned} \mathfrak{b}_n(k) &:= \{(a_{ij})_{ij} \in \mathfrak{gl}_n(k) \mid a_{ij} = 0 \text{ if } i > j\}, \\ \mathfrak{n}_n(k) &:= \{(a_{ij})_{ij} \in \mathfrak{gl}_n(k) \mid a_{ij} = 0 \text{ if } i \geq j\}, \\ \mathfrak{t}_n(k) &:= \{(a_{ij})_{ij} \in \mathfrak{gl}_n(k) \mid a_{ij} = 0 \text{ if } i \neq j\}, \end{aligned}$$

resp., are Lie subalgebras of $\mathfrak{gl}_n(k)$, which can be easily checked. Note that $\mathfrak{b}_n(k) = \mathfrak{t}_n(k) \oplus \mathfrak{n}_n(k)$ as k-vector spaces. Moreover, $[\mathfrak{t}_n(k), \mathfrak{n}_n(k)] \subseteq \mathfrak{n}_n(k)$ and $[e_{ii}, e_{ij}] = e_{ij}$ for $i < j$, so $[\mathfrak{t}_n(k), \mathfrak{n}_n(k)] = \mathfrak{n}_n(k)$. Hence,

$$\mathfrak{n}_n(k) \subseteq [\mathfrak{b}_n(k), \mathfrak{b}_n(k)] \subseteq [\mathfrak{t}_n(k), \mathfrak{n}_n(k)] + [\mathfrak{n}_n(k), \mathfrak{t}_n(k)] + \mathfrak{n}_n(k) \subseteq \mathfrak{n}_n(k),$$

so we have equality everywhere.

XI.1.1.14. — Recall that a (not necessarily associative) k-*algebra* is a k-vector space A with a bilinear (multiplication) map $A \times A \to A$.

Definition* XI.1.1.15.** — *A* ***derivation *of a k-algebra A is a linear map $D\colon A \to A$ such that the* ***Leibniz rule*** *$D(ab) = aD(b) + D(a)b$ for all $a, b \in A$ is satisfied. We denote the k-linear subspace of k-derivations of A by $\operatorname{Der} A \subseteq \mathfrak{gl}(A)$.*

***Fact* XI.1.1.16.** — *The derivations $\operatorname{Der} A$ for a k-algebra A is a Lie subalgebra of $\mathfrak{gl}(A)$.*

Proof. — Let $D, E \in \operatorname{Der} A$ be two derivations. Then

$$\begin{aligned} D(E(ab)) &= D(a)E(b) + E(a)D(b) + aD(E(b)) + D(E(a))b, \\ E(D(ab)) &= D(a)E(b) + E(a)D(b) + aE(D(b)) + E(D(a))b, \end{aligned}$$

so

$$[D, E](ab) = a(D(E(b)) - E(D(b))) + (D(E(a)) - E(D(a)))b = a[D, E](b) + [D, E](a)b$$

and $[D, E] \in \operatorname{Der} A$. ∎

***Warning* XI.1.1.17.** — In general, $\operatorname{Der} A$ for a k-algebra A is not a subalgebra of $\mathfrak{gl}(A)$ under composition. If $D \circ E$ were a derivation, then $D(E(ab)) = aD(E(b)) + D(E(a))b$, whence $D(a)E(b) + E(a)D(b) = 0$, which is generally not true. For instance, if $A = k[x]$ and $D = E = \partial/\partial x$, then $D(x)E(x) + E(x)D(x) = 2$.

Definition* XI.1.1.18.** — *Let $\mathfrak{g}$ be a Lie algebra. Any $x \in \mathfrak{g}$ defines a linear map $\operatorname{ad}_{\mathfrak{g}} x := [x, -]\colon \mathfrak{g} \to \mathfrak{g}$, $y \mapsto [x, y]$ called the* ***adjoint map *of x.*

***Fact* XI.1.1.19.** — *Let $\mathfrak{g}$ be a Lie algebra. Then $\operatorname{ad}_{\mathfrak{g}} x \in \operatorname{Der} \mathfrak{g}$, so the collection of all adjoint maps defines a linear map $\mathfrak{g} \to \operatorname{Der} \mathfrak{g}$.*

Proof. — The Jacobi identity can be rewritten as $[x, [y, z]] = [[x, y], z] + [y, [x, z]]$, so $\operatorname{ad}_{\mathfrak{g}} x$ is a derivation.

∎

Definition* XI.1.1.20.** — *Let $\mathfrak{g}$ be a Lie algebra. We call the linear map $\operatorname{ad} := \operatorname{ad}_{\mathfrak{g}}\colon \mathfrak{g} \to \operatorname{Der} \mathfrak{g}$ the* ***adjoint representation *of $\mathfrak{g}$. Moreover, we call the derivations in its image* ***inner****, otherwise* ***outer****.*

***Warning* XI.1.1.21.** — The adjoint representation is not faithful, i. e., injective. If $\mathfrak{g} = k$ is one-dimensional, then $\operatorname{Der} \mathfrak{g} = 0$ since $D(1 \cdot 1) = 2D(1)$ and so $D(1) = 0$ for any $D \in \operatorname{Der} \mathfrak{g}$. In particular, $\operatorname{ad} 1 = 0$ although $1 \neq 0$.

***Fact* XI.1.1.22.** — *Up to isomorphism, the only one-dimensional Lie algebra over k is the abelian algebra k. Up to isomorphism, there are only two two-dimensional Lie algebras over k: the abelian algebra $k\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right) \oplus k\left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) \cong k^2$ and the non-abelian algebra $k\left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right) \oplus k\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$.*

Proof. — Let $\mathfrak{g}$ be a Lie algebra. If $\dim \mathfrak{g} = 1$, then any $x \in \mathfrak{g} \setminus \{0\}$ is a basis vector and $[a_1 x, a_2 x] = a_1 a_2 [x, x] = 0$ for all $a_1, a_2 \in k$. Clearly, all one-dimensional abelian Lie algebras are isomorphic to k.

If $\dim \mathfrak{g} = 2$, let $\{x, y\}$ be a basis of $\mathfrak{g}$. Then the Lie bracket is completely determined by $x' = [x, y]$ (due to anti-commutativity), whence $[\mathfrak{g}, \mathfrak{g}] = \operatorname{span} x'$. If $x' = 0$, then $\mathfrak{g}$ is abelian and isomorphic to k^2. Otherwise, take any $y' \notin \operatorname{span} x'$, so that $\{x', y'\}$ is a basis of $\mathfrak{g}$. Then $[x', y'] \in \operatorname{span} x'$, say, $[x', y'] = ax'$ for some $a \in k^\times$, whence $[x', a^{-1}y'] = x'$. In other words, any non-abelian two-dimensional Lie algebra $\mathfrak{g}$ exhibits a basis $\{x, y\}$ such that the bracket is $[x, y] = x$, and this is unique up to isomorphism.

To construct an explicit example, consider the adjoint representation $\operatorname{ad}\colon \mathfrak{g} \to \operatorname{Der}\mathfrak{g} \subseteq \operatorname{End}\mathfrak{g}$. Given a basis $\{x, y\}$ such that $[x, y] = x$, then $\operatorname{ad} x = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ and $\operatorname{ad} y = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ under the identification $\operatorname{End}\mathfrak{g} \cong \mathfrak{g}_2(k)$.

XI.1.2. Ideals.

***Definition* XI.1.2.1.** — *A subspace $\mathfrak{a}$ of a Lie algebra $\mathfrak{g}$ is an **ideal** of $\mathfrak{g}$ if $[x, y] \in \mathfrak{a}$ (or equivalently $[y, x] \in \mathfrak{a}$) for all $x \in \mathfrak{g}$ and all $y \in \mathfrak{a}$.*

***Example* XI.1.2.2.** — 0 and $\mathcal{G}$ are ideals of $\mathcal{G}$

PART VI

COMPLEX ANALYSIS

CHAPTER CONTENTS

§ XI.2. INTRODUCTION

***Reference* XI.2.0.1.** — Primarily [SS03], sometimes enriched by [FB09]; [Ahl21].

***Notation* XI.2.0.2.** — – A sequence $z_1, z_2, ...$ is denoted as $(z_n)_{n\geq 1}$ or just $(z_n)_n$. Note that it is ordered, so we avoid the notation $\{z_n\}_n$ common in [SS03].

– We avoid the notation $\mathbb{D}$ for the unit disc in $\mathbb{C}$, although ubiquitous in [SS03]. For clarity, the notation $D_1(0)$ is preferred. There is one exception: the section on conformal maps.

– Contrary to standard topology, we define a homotopy between curves with common endpoints to be *pointed*. For closed curve which do not share a common base point, however, the homotopy need not be pointed.

CHAPTER XII

BASIC THEORY

§ XII.1. PRELIMINARIES

XII.1.1. Complex numbers.

Definition* XII.1.1.1.** — *The field of* ***complex numbers *is the set*

$$\mathbb{C} := \{z = x + \mathrm{i}y \mid x, y \in \mathbb{R}\},$$

where i *is the* ***imaginary number*** *satisfying* $\mathrm{i}^2 = -1$*, together with addition and multiplication:*

$$(x_1 + \mathrm{i}y_1)(x_2 + \mathrm{i}y_2) = (x_1 + x_2) + \mathrm{i}(y_1 + y_2),$$
$$(x_1 + \mathrm{i}y_1)(x_2 + \mathrm{i}y_2) = (x_1x_2 - y_1y_2) + \mathrm{i}(x_1y_2 + y_1x_2),$$

following usual associativity and distributivity subject under $\mathrm{i}^2 = -1$*. The complex numbers being a field means that it is, in particular, commutative, associative and distributive.*[(XII.1)]

For $z = x + \mathrm{i}y \in \mathbb{C}$*, we call* $\operatorname{Re} z =: x$ *and* $\operatorname{Im} z := y$ *the* ***real part*** *and* ***imaginary part*** *of* z*, resp. So, the real numbers are those* $z \in \mathbb{C}$ *with* $\operatorname{Im} z = 0$*, and the* ***purely imaginary numbers*** *are those* $z \in \mathbb{C}$ *with* $\operatorname{Re} z = 0$*.*

Definition* XII.1.1.2.** — *The* ***absolute value *of* $z = x + \mathrm{i}y \in \mathbb{C}$ *is given by* $|z| := \sqrt{x^2 + y^2}$*.*

***Remark* XII.1.1.3.** — There is a natural $\mathbb{R}$-linear isomorphism between the complex numbers and the Euclidean plane

$$\mathbb{C} \xrightarrow{\sim} \mathbb{R}^2, \qquad z = x + \mathrm{i}y \longmapsto (x, y).$$

In fact, this is an isometry, meaning that $|z| = \|(x, y)\|_2$.

Under this identification, we also call $\mathbb{C}$ the **complex plane**. For example, $0 \in \mathbb{C}$ is the origin and $\mathrm{i} \in \mathbb{C}$ is $(0, 1)$. The x- and y-axis are called the **real axis** and **imaginary axis**, resp.

***Fact* XII.1.1.4** (properties of the absolute value)**.** — *The absolute value satisfies the following:*

(i) $|z| \geq 0$ *with equality if and only if* $z = 0$*;*
(ii) $|zw| = |z||w|$*;*
(iii) *triangle inequality:* $|z + w| \leq |z| + |w|$*;*
(iv) *reverse triangle inequality:* $|z| - |w| \leq ||z| - |w|| \leq |z - w|$*;*
(v) $|\operatorname{Re} z| \leq |z|$ *and* $|\operatorname{Im} z| \leq |z|$*.*

Definition* XII.1.1.5.** — *The* ***complex conjugate *of* $z = x + \mathrm{i}y \in \mathbb{C}$ *is defined as* $\overline{z} := x - \mathrm{i}y$*, which corresponds to reflection along the real axis.*

[(XII.1)] Algebraically, $\mathbb{C}$ is the splitting field of the polynomial x^2+1 over $\mathbb{R}$, whence is unique up to isomorphism. If i is a solution of $x^2 + 1 = 0$, then $\mathbb{C} = \mathbb{R}(\mathrm{i}) \cong \mathbb{R}[x]/(x^2 + 1)$.

***Fact* XII.1.1.6** (properties of the complex conjugate). — *The complex conjugate satisfies the following:*

(i) $\overline{\overline{z}} = z$, *i. e., complex conjugation is involutory;*
(ii) $\overline{z+w} = \overline{z} + \overline{w}$ *and* $\overline{zw} = \overline{z}\overline{w}$*;*
(iii) z *is real if and only if* $z = \overline{z}$*;* z *is purely imaginary if and only if* $z = -\overline{z}$*;*
(iv) $\operatorname{Re} z = \frac{1}{2}(z + \overline{z})$ *and* $\operatorname{Im} z = \frac{1}{2}(z - \overline{z})$*;*
(v) $|z|^2 = z\overline{z}$*, so* $1/z = \overline{z}/|z|^2$ *for* $z \neq 0$*.*

XII.1.2. Topological properties and continuous functions.

Definition* XII.1.2.1.** — (i) *A sequence* $(z_n)_{n\geq 1} \subseteq \mathbb{C}$ *of complex numbers* ***converges *to* $w \in \mathbb{C}$ *if* $\lim_{n\to\infty}|z_n - w| = 0$*, in which case we write* $w = \lim_{n\to\infty} z_n$*.*

(ii) *A sequence* $(z_n)_{n\geq 1} \subseteq \mathbb{C}$ *is a* ***Cauchy sequence*** *(or simply* Cauchy*) if* $|z_n - z_m| \to 0$ *as* $n, m \to \infty$*, i. e., for each* $\varepsilon > 0$ *there is* $N \in \mathbb{N}$ *such that* $|z_n - z_m| < \varepsilon$ *whenever* $n, m \geq N$*.*

***Fact* XII.1.2.2.** — *A sequence* $(z_n)_n \subseteq \mathbb{C}$ *converges if and only if the real sequences* $(\operatorname{Re} z_n)_n$ *and* $(\operatorname{Im} z_n)_n$ *converges. In this case,* $\lim_{n\to\infty} z_n = \lim_{n\to\infty} \operatorname{Re} z_n + \mathrm{i} \lim_{n\to\infty} \operatorname{Im} z_n$*.*

Proof. — If $(z_n)_n$ converges, then the real and imaginary parts converge since $|\operatorname{Re} z|, |\operatorname{Im} z| \leq |z|$ for all $z \in \mathbb{C}$. Conversely, if the real and imaginary parts converge, then $(z_n)_n$ converges since $z_n = \operatorname{Re} z_n + \mathrm{i} \operatorname{Im} z_n$.

***Proposition* XII.1.2.3.** — *The complex numbers* $\mathbb{C}$ *are complete, i. e., Cauchy sequences in* $\mathbb{C}$ *converge with limit point in* $\mathbb{C}$*.*

Proof. — Recall that the Bolzano-Weierstraß theorem(XII.2) implies that $\mathbb{R}$ is complete. Since $(z_n)_n$ is Cauchy if and only if $(\operatorname{Re} z_n)_n$ and $(\operatorname{Im} z_n)_n$ are Cauchy, the statement follows.

The following definitions are standard and agree with the definitions for $\mathbb{R}^2$.

Definition* XII.1.2.4.** — *Let* $z_0 \in \mathbb{C}$ *and let* $r > 0$*. The* ***open disc*,* ***closed disc*** *and* ***circle*** *of radius* r *centred at* z_0 *are*

$$
\begin{aligned}
\mathrm{D}_r(z_0) &= \{z \in \mathbb{C} \mid |z - z_0| < r\},\\
\bar{\mathrm{D}}_r(z_0) &= \{z \in \mathbb{C} \mid |z - z_0| \leq r\},\\
\mathrm{C}_r(z_0) &= \{z \in \mathbb{C} \mid |z - z_0| = r\},
\end{aligned}
$$

resp. We call $\mathrm{D}_1(0) = \{z \in \mathbb{C} \mid |z| < r\}$ *the* ***unit disc****.*

Definition* XII.1.2.5.** — *Let* $\Omega \subseteq \mathbb{C}$ *be a set. A point* $z_0 \in \Omega$ *is an* ***interior point *of* Ω *if* $\mathrm{D}_r(z_0) \subseteq \Omega$ *for some* $r > 0$*. The* ***interior*** $\mathring{\Omega}$ *of* Ω *is the set of all interior points, and* Ω *is* ***open*** *if* $\Omega = \mathring{\Omega}$*.*

A point $z \in \mathbb{C}$ *is a* ***limit point*** *of* Ω *if there exists a sequence* $(z_n)_n \subseteq \Omega$ *with* $z \neq z_n$ *and* $z = \lim_{n\to\infty} z_n$*. The* ***closure*** $\overline{\Omega}$ *of* Ω *is the union of* Ω *and its limit points, and* Ω *is* ***closed*** *if* $\Omega = \overline{\Omega}$*. Equivalently,* Ω *is closed if the complement* $\mathbb{C} \setminus \Omega$ *is open.*

Finally, the ***boundary*** *of* Ω *is* $\partial\Omega = \overline{\Omega} \setminus \mathring{\Omega}$*.*

Definition* XII.1.2.6.** — *A set* $\Omega \subseteq \mathbb{C}$ *is* ***bounded *if there exists* $M > 0$ *such that* $|z| < M$ *for all* $z \in \Omega$*; in other words,* $\Omega \subseteq \mathrm{D}_r(0)$ *for some large* r*. If* Ω *is bounded, its* ***diameter*** *is* $\operatorname{diam} \Omega := \sup_{z,w\in\Omega}|z - w|$*.*

A set $\Omega \subseteq \mathbb{C}$ *is* ***compact*** *if it is closed and bounded.*

(XII.2) The *Bolzano-Weierstraß theorem* states that any bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

***Proposition* XII.1.2.7** (characterisation of compactness in metric spaces)**.** — *For a set $\Omega \subseteq \mathbb{C}$, the following are equivalent:*

(i) *Ω is compact, i. e., closed and bounded;*
(ii) *every sequence $(z_n)_n \subseteq \Omega$ has a convergent subsequence with limit in Ω;*
(iii) *every open cover $\Omega \subseteq \bigcup_\alpha U_\alpha$ of Ω by an family of open sets $(U_\alpha)_\alpha$ has a finite subcover.*

Proof. — The proof is identical to the real statement. Firstly, suppose that Ω is compact, i. e., closed and bounded. Let $(z_n)_n \subseteq \Omega$ be a sequence. Since $|\operatorname{Re} z|, |\operatorname{Im} z| \leq |z|$, the sequences of real and imaginary parts are also bounded. By Bolzano-Weierstraß in $\mathbb{R}$, the sequence $(\operatorname{Re} z_n)_n$ has a convergent subsequence $(\operatorname{Re} z_{n_i})_i$. Again by Bolzano-Weierstraß, the sequence $(\operatorname{Im} z_{n_i})_i$ has a convergent subsequence $(\operatorname{Im} z_{n_{i,j}})_j$. Hence, the real and imaginary parts of $(z_{n_{i,j}})_j$ converge and thus, the sequence itself converges. Since Ω is closed, the limit lies in Ω.

Next, suppose that every sequence in Ω has a convergent subsequence in Ω. Let $\Omega \subseteq \bigcup_\alpha U_\alpha$ be an open cover that has no finite subcover. Since Ω is sequentially compact, this cover has a Lebesgue number $\varepsilon > 0$ such that each disc $\mathrm{D}_\varepsilon(z)$ with centre $z \in \Omega$ is contained in some U_α. We claim that there is a finite cover of ε-discs with centres in Ω. If not, start with $z_1 \in \Omega$ and iteratively pick $z_n \in \Omega \setminus \bigcup_{i=1}^{n-1} \mathrm{D}_\varepsilon(z_i)$, so that $|z_n - z_m| \geq \varepsilon$ for any $n \neq m$. By assumption, $(z_n)_n$ has a convergent subsequence, which is absurd since it cannot be Cauchy. Thus, Ω is covered by finitely many such discs, each contained in some U_α, giving rise to a finite subcover.

Finally, we reproduce the proof of Heine-Borel. Suppose that every open cover of Ω has a finite subcover. Since $\mathbb{C}$ is Hausdorff, the topologically compact Ω must be closed. To show that Ω is bounded, consider the open cover $\Omega \subseteq \bigcup_{z\in\Omega} \mathrm{D}_1(z)$, and let $\Omega \subseteq \bigcup_{n=1}^N \mathrm{D}_1(z_n)$ be a finite subcover. Then for all $w_i \in \mathrm{D}_1(z_i)$ and all $w_j \in \mathrm{D}_1(z_j)$, we have

$$|w_i - w_j| \leq |w_i - z_i| + |z_i - z_j| + |w_j - z_j| \leq 2 + \operatorname{diam}(\{z_1, \dots, z_N\}).$$

***Proposition* XII.1.2.8.** — *If $\Omega_1 \supseteq \Omega_2 \supseteq \cdots$ is a descending sequence of non-empty compact sets in $\mathbb{C}$ such that $\operatorname{diam} \Omega_n \to 0$ as $n \to \infty$, then there exists a unique point $w \in \mathbb{C}$ such that $w \in \Omega_n$ for all n.*

Proof. — Choose $z_n \in \Omega_n$ for each n. The condition $\operatorname{diam} \Omega_n \to 0$ means that $(z_n)_n$ is a Cauchy sequence, so let $w = \lim_{n\to\infty} z_n \in \mathbb{C}$. Since each Ω_n is closed, we have $w \in \Omega_n$ for all n. Any other w' with $w' \in \Omega_n$ for all n satisfies $|w - w'| \leq \operatorname{diam} \Omega_n \to 0$, so $w = w'$.

***Definition* XII.1.2.9.** — *A set $\Omega \subseteq \mathbb{C}$ is **connected** if there are no two disjoint non-empty open sets Ω_1 and Ω_2 such that $\Omega = \Omega_1 \cup \Omega_2$. We call an open connected set a **region**.*

***Remark* XII.1.2.10.** — Similarly, a set $F \subseteq \Omega$ is closed if there are no two disjoint non-empty closed sets F_1 and F_2 such that $F = F_1 \cup F_2$.

***Definition* XII.1.2.11.** — *A set $\Omega \subseteq \mathbb{C}$ is **path-connected** if any two points in Ω can be joined by a continuous curve contained in Ω.*

***Proposition* XII.1.2.12** (characterisation of connectedness in metric spaces)**.** — *An open set $\Omega \subseteq \mathbb{C}$ is path-connected if and only if it is connected.*

Proof. — The proof proceeds like the real case. Suppose that Ω is open path-connected, and suppose that $\Omega = \Omega_1 \cup \Omega_2$ where Ω_1 and Ω_2 are disjoint non-empty open. For any $w_1 \in \Omega_2$ and any $w_2 \in \Omega_2$, let $\gamma\colon [0,1] \to \Omega$ be a curve from w_1 to w_2. Let $t^* = \sup\{t \in [0,1] \mid \gamma(s) \in \Omega_1 \text{ for all } 0 \leq s < t\}$. If $\gamma(t^*) \in \Omega_1$, then there is an open neighbourhood of $\gamma(t^*)$ in Ω_1, and t^* was not the supremum yet. If $\gamma(t^*) \in \Omega_2$, then there is an open neighbourhood of $\gamma(t^*)$ in $\Omega_2 = \Omega \setminus \Omega_1$, and t^* was larger than the supremum. We obtain a contradiction in either case.

The converse requires that Ω is a metric space, i. e., that open sets are defined in terms of discs. Suppose that Ω open connected. Fixing a point $w \in \Omega$, let $\Omega_1 \subseteq \Omega$ be the maximal path-connected component of w, i. e., it is the set of all points in w which can be connected to w via a continuous curve in Ω. Let $\Omega_2 = \Omega \setminus \Omega_1$. Then for any $z \in \Omega_1$, we have $\mathrm{D}_r(z) \subseteq \Omega$ for some $r > 0$, so any point in $\mathrm{D}_r(z)$ is connected to z by a straight line. Hence, Ω_1 is open. Similarly, for any $z \in \Omega_2$, we have $\mathrm{D}_r(z) \subseteq \Omega$ for some $r > 0$. If some point in $z' \in \mathrm{D}_r(z)$ were in Ω_1, then $z \in \Omega_1$ by extending the connecting curve from w to z' by a straight line from z' to z, a contradiction. Thus, Ω_1 and Ω_2 are both open. Since $w \in \Omega_1 \neq \varnothing$, we have $\Omega = \Omega_1$. ▢

Definition* XII.1.2.13.** — *Let $\Omega \subseteq \mathbb{C}$ be an open set. The* ***connected component *of Ω is the subset of all points in Ω such that any two can be joined by a continuous curve contained in Ω.*

***Fact* XII.1.2.14** (chacterisation of connected components)**.** — *Let $\Omega \subseteq \mathbb{C}$ be an open set.*

(i) *Connected components of Ω are open and connected. They define an equivalence relation of points in Ω. Therefore, Ω is the disjoint union of all its connected components.*

(ii) *There are only countably many connected components of Ω.*

(iii) *If $\Omega = \mathbb{C} \setminus K$ for K compact, then Ω has only one unbounded connected component.*

Proof. — Connected components are, by definition, connected. If C_z is the connected component of Ω containing $z \in \Omega$, then there is a disc $D \subseteq \Omega$ centred at z since Ω is open. As any point of D is connected to z, we have $D \subseteq C_z$ and C_z is open. Finally, the equivalence relation can be easily checked, and hence we proved (i).

For (ii), suppose that there are uncountably many connected components. This allows us to decompose Ω into an uncountable disjoint union of open subsets, giving an uncountable number of disjoint open discs. However, $\mathbb{Q}^2 \subseteq \mathbb{R}^2 \cong \mathbb{C}$ is a dense subset of $\mathbb{C}$, i. e., any complex number is an accumulation point of the countable set $\mathbb{Q}^2$. Then there is an open disc which does not contain rational points, a contradiction.

Lastly, if $\Omega = \mathbb{C} \setminus K$ for K compact, let D be a large disc containing K. Then $\mathbb{C} \setminus D$ is a connected subset of Ω, and there is exactly one connected component that contains $\mathbb{C} \setminus D$ and is hence unbounded. All other connected components are contained in $\Omega \cap D$ and are hence bounded. This establishes (iii). ▢

Definition* XII.1.2.15.** — *A function $f\colon \Omega \to \mathbb{C}$ is* ***continuous *at a point $z_0 \in \Omega$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(z) - f(z_0)| < \varepsilon$ whenever $z \in \Omega$ with $|z - z_0| < \delta$. Equivalently, f is continuous if for any converging sequence $(z_n)_n \subseteq \Omega$ with limit in Ω, we have $\lim_{n\to\infty} f(z_n) = f(\lim_{n\to\infty} z_n)$. Moreover, f is* ***continuous*** *on Ω if it is continuous at every point of Ω.*

***Fact* XII.1.2.16.** — (i) *Sums and products of continuous functions are continuous.*

(ii) *A function $f\colon \Omega \to \mathbb{C}$ is continuous if and only if it is continuous as a function $U \to \mathbb{C}$, $(x, y) \mapsto f(x + \mathrm{i}y)$ where $U \subseteq \mathbb{R}^2$ is the set corresponding to $\Omega \subseteq \mathbb{C}$.*

(iii) *If $f\colon \Omega \to \mathbb{C}$ is continuous, then $|f|\colon \Omega \to \mathbb{R}$, $z \mapsto |f(z)|$ is continuous.*

Definition* XII.1.2.17.** — *A function $f\colon \Omega \to \mathbb{C}$ attains a* ***maximum *at $z_0 \in \Omega$ if $|f(z)| \leq |f(z_0)|$ for all $z \in \Omega$. We similarly define a* ***minimum*** *of f.*

***Proposition* XII.1.2.18** (extreme value theorem)**.** — *A continuous function on a compact set is bounded and attains a maximum and minimum.*

Proof. — The proof is identical to the real case. Let $f\colon \Omega \to \mathbb{C}$ be a continuous function on an open set Ω. Suppose that there is a sequence $(z_n)_n \subseteq \Omega$ such that $|f(z_n)| \to \infty$ as $n \to \infty$. Since Ω is compact, w. l. o. g., we may assume that $z_n \to x \in \Omega$ converges by restricting to subsequences. Then, however, $|f(z_n)| \to |f(z)| = \infty$, a contradiction. Thus, f is bounded.

So, let $M = \sup_{z\in\Omega} f(z)$ be the supremum of f. There is a sequence $(z_n)_n \subseteq \Omega$ such that $f(z_n) \to M$. Again by compactness, $z_n \to z \in \Omega$ converges w. l. o. g. Then $f(z) = M$ is a maximum. The minimum follows by noting that any minimum is a maximum of $-f$. ▢

XII.1.3. Holomorphic functions.

Definition* XII.1.3.1.** — *Let $f\colon \Omega \to \mathbb{C}$ be a function on an open set $\Omega \subseteq \mathbb{C}$. Then f is* ***holomorphic *at a point $z_0 \in \Omega$ if the limit*

$$\lim_{\substack{h\in\mathbb{C}\setminus\{0\}\\ z_0+h\in\Omega}} \frac{f(z_0+h)-f(z_0)}{h} =: f'(z_0)$$

exists, called the (complex) ***derivative*** *of f at z_0. Note that $h \to 0$ may approach 0 from any direction.*

Moreover, f is ***holomorphic*** *on Ω if f is holomorphic at every point of Ω. If $C \subseteq \mathbb{C}$ is closed, then $f\colon C \to \mathbb{C}$ is* ***holomorphic*** *on C if it is holomorphic on some open neighbourhood of C. Finally, f is* ***entire*** *if f is holomorphic on all of $\mathbb{C}$.*

***Warning* XII.1.3.2.** — Sometimes, *regular* and *complex differentiable* are used interchangeably with *holomorphic*, but holomorphic functions have much stronger properties then real-differentiable functions. On the other hand, we will use the term *analytic* and *holomorphic* synonymously, see next section. Insert reference?

***Remark* XII.1.3.3.** — Alternatively, $f\colon \Omega \to \mathbb{C}$ is holomorphic at $z_0 \in \Omega$ if and only if there exists $a \in \mathbb{C}$ such that

$$f(z_0+h) - f(z_0) - ah = h\psi(h)$$

where $\psi(h)$ is a function for small h with $\lim_{h\to 0}\psi(h) = 0$. In this case, $a = f'(z_0)$. In particular, holomorphy implies continuity.

***Proposition* XII.1.3.4** (differentiation rules). — *If $f, g\colon \Omega \to \mathbb{C}$ are holomorphic at z_0, then the following hold:*

(i) *$f+g$ is holomorphic at z_0 with $(f+g)' = f' + g'$;*
(ii) *fg is holomorphic at z_0 with $(fg)' = f'g + fg'$;*
(iii) *if $g(z_0) \neq 0$, then f/g is holomorphic at z_0 with $(f/g)' = (f'g - fg')/g^2$.*

Moreover, if $f\colon \Omega \to U \subseteq \mathbb{C}$ is holomorphic at z_0 and if $g\colon U \to \mathbb{C}$ is holomorphic at $f(z_0)$, then $g \circ f$ is holomorphic at z_0 and the chain rule $(g\circ f)' = (g' \circ f)f'$ holds.

Proof. — Use remark XII.1.3.3 and proceed as in the real case. We will give a few key points here. Firstly, it is necessary to use continuity of f and g at z_0. Moreover, $g(z_0) \neq 0$ implies $g(z) \neq 0$ in a small neighbourhood of z_0, so f/g is well-defined.

The sum, product and quotient rule follow by direct computation. For the chain rule, we have

$$f(z_0+h) - f(z_0) - f'(z_0)h = h\psi(h), \qquad g(f(z_0)+h) - g(f(z_0)) - g'(f(z_0))h = h\phi(h)$$

where $\psi(h), \phi(h) \to 0$ as $h \to 0$. By taking $f(z_0+h) - f(z_0) = f'(z_0)h + h\psi(h)$ for h, the second equation becomes

$$g(f(z_0+h)) - g(f(z_0)) = (g'(f(z_0)) + \phi(h))(f'(z_0) + \psi(h))h = g'(f(z_0))f'(z_0)h + h\chi(h)$$

where $\chi(h) = g'(f(z_0))\psi(h) + f'(z_0)\phi(h) + \psi(h)\phi(h) \to 0$ as $h \to 0$. Hence, $g \circ f$ is holomorphic at z_0 with $(g\circ f)' = (g' \circ f)f'$. ∎

***Example* XII.1.3.5.** — The function $f(z) = z$ is entire (holomorphic on $\mathbb{C}$) with $f'(z) = 1$. Hence, polynomials $p(z) = a_0 + a_1 z + \cdots + a_n z^n$ are entire with derivative $p'(z) = a_1 + \cdots + na_n z^{n-1}$.

***Example* XII.1.3.6.** — The function $f(z) = 1/z$ is holomorphic on any open set in $\mathbb{C} \setminus \{0\}$ with $f'(z) = -1/z^2$.

***Counterexample* XII.1.3.7.** — The function $f(z) = \overline{z}$ is not holomorphic. Indeed, $(f(z_0 + h) - f(z_0))/h = \overline{h}/h$ does not converge as $h \to 0$ since $\overline{h}/h \to 1$ if $h \in \mathbb{R}$ is real, and $\overline{h}/h \to -1$ if $h \in \mathrm{i}\mathbb{R}$ is purely imaginary.

***Warning* XII.1.3.8.** — Counterexample XII.1.3.7 also shows the difference to real-differentiability. Interpreting f as a function $F\colon \mathbb{R}^2 \to \mathbb{R}^2$, we have $F(x, y) = (x, -y)$. Since it is linear, the Jacobian (derivative) of F at any point is F itself, and F is smooth. Hence, real-differentiability does not imply holomorphy.

XII.1.3.9. — Let $F\colon U \to \mathbb{R}^2$, $F(x, y) = (u(x, y), v(x, y))$ with $U \subseteq \mathbb{R}^2$. Recall that F is *differentiable* at $P_0 \in U$ if there exists a linear transformation $J\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$\lim_{\|H\| \to 0} \frac{\|F(P_0 + H) - F(P_0) - J(H)\|}{\|H\|} = 0,$$

or equivalently,

$$F(P_0 + H) - F(P_0) = J(H) + |H|\Psi(H)$$

where $\|\Psi(H)\| \to 0$ as $\|H\| \to 0$. In this case, J is unique, the partial derivatives of u and v exist, and J is the *Jacobian* (derivative)

$$J = \mathrm{J}_F(x, y) := \begin{pmatrix} \partial u/\partial x(x, y) & \partial u/\partial y(x, y) \\ \partial v/\partial x(x, y) & \partial v/\partial y(x, y) \end{pmatrix}$$

of F.

***Observation* XII.1.3.10.** — More generally, we want to study the relation between the complex and real derivative. Let $f\colon \Omega \to \mathbb{C}$, which we will also write as $f(x, y) = f(x + \mathrm{i}y)$ on $\mathbb{R}^2$ by abuse of notation. Suppose that f is holomorphic at $z_0 = x_0 + \mathrm{i}y_0$. By first taking the limit of the difference quotient over real h, say, $h = h_1 \in \mathbb{R}$, we have

$$f'(z_0) = \lim_{h_1 \to 0} \frac{f(x_0 + h_1, y_0) - f(x_0, y_0)}{h_1} = \frac{\partial f}{\partial x}(z_0)$$

where $\partial/\partial x$ is the partial derivative of $f(x, y)$ in x. Now taking the limit over purely imaginary h, say, $h = \mathrm{i}h_2$ with $h_2 \in \mathbb{R}$, we similarly have

$$f'(z_0) = \lim_{h_2 \to 0} \frac{f(x_0, y_0 + h_2) - f(x_0, y_0)}{\mathrm{i}h_2} = \frac{1}{\mathrm{i}} \frac{\partial f}{\partial y}(z_0)$$

where $\partial/\partial y$ is the partial derivative in y. Thus, if f is holomorphic then

$$\frac{\partial f}{\partial x} = \frac{1}{\mathrm{i}} \frac{\partial f}{\partial y}.$$

By taking the real and imaginary parts of this equation (and using $1/\mathrm{i} = -\mathrm{i}$), if we write $f = u + \mathrm{i}v$ then we equivalently obtain the **Cauchy-Riemann equations**

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{XII.1.3.10.1}$$

***Observation* XII.1.3.11.** — For a (holomorphic) function $f\colon \Omega \to \mathbb{C}$, we write $f(z) = f(x + \mathrm{i}y) = f(x, y)$. The chain rule gives

$$\frac{\partial f}{\partial z} = \frac{\partial x}{\partial z}\frac{\partial f}{\partial x} + \frac{\partial y}{\partial z}\frac{\partial f}{\partial y} = \frac{\partial}{\partial z}\left(\frac{z + \overline{z}}{2}\right)\frac{\partial f}{\partial x} + \frac{\partial}{\partial z}\left(\frac{z - \overline{z}}{2\mathrm{i}}\right)\frac{\partial f}{\partial y} = \frac{1}{2}\frac{\partial f}{\partial x} + \frac{1}{2\mathrm{i}}\frac{\partial f}{\partial y}$$

by treating $\overline{z}$ constant, and similarly for $\partial f/\partial\overline{z}$. Hence, we define the **Wirtinger derivatives**

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} + \frac{1}{\mathrm{i}}\frac{\partial}{\partial y}\right), \qquad \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \frac{1}{\mathrm{i}}\frac{\partial}{\partial y}\right).$$

***Proposition* XII.1.3.12.** — *Let $f = u + \mathrm{i}v\colon \Omega \to \mathbb{C}$ is holomorphic at $z_0 \in \Omega$. Then u and v satisfy the Cauchy-Riemann eqs. (XII.1.3.10.1). With Wirtinger derivatives, we have*

$$\frac{\partial f}{\partial \overline{z}}(z_0) = 0, \qquad f'(z_0) = \frac{\partial f}{\partial z}(z_0) = 2\frac{\partial u}{\partial z}(z_0).$$

Moreover, if we write $F(x, y) = f(x + \mathrm{i}y)$ as a function in $\mathbb{R}^2$ and $z_0 = x_0 + \mathrm{i}y_0$, then F is real-differentiable and

$$\det \mathrm{J}_F(x_0, y_0) = |f'(z_0)|^2.$$

Proof. — We have shown $\partial f/\partial\overline{z}(z_0) = 0$ in observation XII.1.3.10. This observation also implies that

$$f'(z_0) = \frac{1}{2}\left(\frac{\partial f}{\partial x}(z_0) + \frac{1}{\mathrm{i}}\frac{\partial f}{\partial y}(z_0)\right) = \frac{\partial f}{\partial z}(z_0),$$

and the Cauchy-Riemann eqs. (XII.1.3.10.1) yield

$$\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial u}{\partial x} + \mathrm{i}\frac{\partial v}{\partial x} - \mathrm{i}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}\right) = \frac{\partial u}{\partial x} - \mathrm{i}\frac{\partial u}{\partial y} = 2\frac{\partial u}{\partial z}.$$

To show differentiability of F, if $H = (h_1, h_2)$ and $h = h_1 + \mathrm{i}h_2$ then the Cauchy-Riemann eqs. (XII.1.3.10.1) give

$$\mathrm{J}_F(x_0, y_0)(H) = \begin{pmatrix} \partial u/\partial x & \partial u/\partial y \\ -\partial u/\partial y & \partial u/\partial x \end{pmatrix}(x_0, y_0)\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \left(\frac{\partial u}{\partial x}(z_0) - \mathrm{i}\frac{\partial u}{\partial y}(z_0)\right)(h_1 + \mathrm{i}h_2) = f'(z_0)h$$

(here, we identified $\mathbb{C} \cong \mathbb{R}^2$), so differentiability of f at z_0 directly translates to differentiability of F at (x_0, y_0). Lastly,

$$\det \mathrm{J}_F(x_0, y_0) = \left(\frac{\partial u}{\partial x}(z_0)\right)^2 + \left(\frac{\partial u}{\partial y}(z_0)\right)^2 = \left|2\frac{\partial u}{\partial z}\right|^2 = |f'(z_0)|^2.$$

There is also a sort of converse.

***Theorem* XII.1.3.13.** — *Let $f = u + \mathrm{i}v\colon \Omega \to \mathbb{C}$ be a function on an open set Ω. If u and v are continuously differentiable and satisfy the Cauchy-Riemann eqs. (XII.1.3.10.1) on Ω, then f is holomorphic on Ω with $f' = \partial f/\partial z$.*

Proof. — We freely identify $\mathbb{C} \cong \mathbb{R}^2$ and write $f(x, y) = f(x + \mathrm{i}y)$. For $h = h_1 + \mathrm{i}h_2$, differentiability of u and v give

$$u(x + h_1, y + h_2) - u(x, y) = \frac{\partial u}{\partial x}(x, y)h_1 + \frac{\partial u}{\partial y}(x, y)h_2 + |h|\psi_1(h),$$
$$v(x + h_1, y + h_2) - v(x, y) = \frac{\partial v}{\partial x}(x, y)h_1 + \frac{\partial v}{\partial y}(x, y)h_2 + |h|\psi_2(h),$$

where $\psi_j(h) \to 0$ as $|h| \to 0$. The Cauchy-Riemann eqs. (XII.1.3.10.1) then imply that

$$f(z + h) - f(z) = \left(\frac{\partial u}{\partial x}(z) - \mathrm{i}\frac{\partial u}{\partial y}(z)\right)(h_1 + \mathrm{i}h_2) + |h|\psi(h)$$

where $\psi(h) = \psi_1(h) + \psi_2(h) \to 0$ as $|h| \to 0$. Therefore, f is holomorphic by remark XII.1.3.3 with $f' = 2\partial u/\partial z = \partial f/\partial z$.

***Warning* XII.1.3.14.** — Theorem XII.1.3.13 does not work pointwise any more, but requires at least continuous differentiability on an open neighbourhood Ω. Consider $f(x+\mathrm{i}y) = \sqrt{|xy|}$ for $x, y \in \mathbb{R}$. Writing $f = u + \mathrm{i}v$, we have $u = f$ and $v = 0$. For the partial derivatives of u, we explicitly compute

$$\frac{\partial u}{\partial x}(0,0) = \lim_{h\to 0} \frac{u(h,0) - u(0,0)}{h} = 0, \qquad \frac{\partial u}{\partial y}(0,0) = \lim_{h\to 0} \frac{u(0,h) - u(0,0)}{h} = 0.$$

Thus, f satisfies the Cauchy-Riemann eqs. (XII.1.3.10.1) at the origin. If it were holomorphic at the origin, then $f'(0) = 0$ by proposition XII.1.3.12. However,

$$\lim_{\substack{h\to 0\\ h>0}} \frac{f(h+\mathrm{i}h) - f(0)}{h+\mathrm{i}h} = \frac{1}{1+\mathrm{i}}, \qquad \lim_{\substack{h\to 0\\ h<0}} \frac{f(h+\mathrm{i}h) - f(0)}{h+\mathrm{i}h} = -\frac{1}{1+\mathrm{i}},$$

so f is not holomorphic at the origin.

XII.1.4. Power series.

***Definition* XII.1.4.1.** — *A **power series** is a series of the form $\sum_{n=0}^{\infty} a_n z^n$ with $a_n \in \mathbb{C}$.*

***Fact* XII.1.4.2.** — *If a power series $\sum_{n=0}^{\infty} a_n z^n$ converges absolutely for some $z_0 \in \mathbb{C}$, then it converges absolutely for all $z \in \bar{\mathrm{D}}_{|z_0|}(0)$.*

Proof. — This follows from the direct comparison test. ∎

***Hadamard's formula* XII.1.4.3.** — *For any power series $\sum_{n=0}^{\infty} a_n z^n$, there exists a number $0 \le R \le \infty$, called the **radius of convergence**, such that:*

(i) *the series converges absolutely for z in the **disc of convergence** $\mathrm{D}_R(0)$;*
(ii) *the series diverges for $|z| > R$.*

With the conventions $1/0 = \infty$ and $1/\infty = 0$, the radius of convergence is given by

$$1/R = \limsup_{n\to\infty} |a_n|^{1/n}.$$

Proof. — Define $L := 1/R := \limsup_{n\to\infty} |a_n|^{1/n}$. If $L = \infty$, then $|a_n| = \infty$ for infinitely many n. So for $|z| > R = 0$, the sequence $(a_n z^n)_n$ cannot converge to 0, whence $\sum_{n=0}^{\infty} a_n z^n$ diverges.

Hence, assume that $L < \infty$. If $|z| < R = 1/L$, then choose $\varepsilon > 0$ so small that $r := (L+\varepsilon)|z| < 1$. By the definition of L, we have $|a_n|^{1/n} \le L + \varepsilon$ for all large n, whence $|a_n||z|^n \le (L+\varepsilon)^n |z|^n = r^n$. Comparing with the geometric series $\sum_n r^n$ shows that the power series $\sum_{n=0}^{\infty} a_n z^n$ converges.

On the other hand, if $|z| > R$, then a similar argument shows that there is $r > 1$ such that $|a_n||z|^n \ge (L-\varepsilon)^n |z|^n > r^n$ for infinitely many n. Therefore, $(|a_n z^n|)_n$ diverges and the power series diverges. ∎

***Warning* XII.1.4.4.** — The behaviour is more unpredictable on the radius of convergence. For all three series $\sum_{n=0}^{\infty} nz^n$, $\sum_{n=1}^{\infty} z^n/n^2$ and $\sum_{n=1}^{\infty} z^n/n$, the radius of convergence is $R = 1$ since $\lim_{n\to\infty} n^{1/n} = 1$.

(i) If $|z| = 1$, then $\sum_{n=0}^{\infty} nz^n$ diverges as the sequence of terms in this power series does not converge to 0.

(ii) If $|z| = 1$, then $\sum_{n=1}^{\infty} z^n/n^2$ converges absolutely as it can be compared to $\sum_{n=1}^{\infty} 1/n^2 = \zeta(2) = \pi^2/6$.

(iii) If $z = 1$, then $\sum_{n=1}^{\infty} z^n/n$ is the harmonic series and hence diverges. If $z \in \mathrm{C}_1(0) \setminus \{1\}$, however, the series converges. Summation by parts[(XII.3)] gives

$$\sum_{n=1}^{N} \frac{z^n}{n} = \frac{B_N}{N} + \sum_{n=1}^{N-1} \Big(\frac{1}{n} - \frac{1}{n+1}\Big) B_n$$

where $B_k = \sum_{n=1}^{k} z^n$. Since B_k is a geometric sum, we have $|B_k| = |1 - z^k|/|1 - z| \le 2/|1 - z|$, so B_k is uniformly bounded for all k. Hence, $B_N/N \to 0$ as $N \to \infty$ and the remaining series is absolutely convergent (as a telescoping series).

***Theorem* XII.1.4.5** (analytic implies holomorphic). — *The power series $f(z) = \sum_{n=0}^{\infty} a_n z^n$ is a holomorphic function on its disc of convergence. There, the derivative is $f'(z) = \sum_{n=0}^{\infty} n a_n z^{n-1}$, i. e., it is formed term-wise. Moreover, f' has the same radius of convergence as f.*

Proof. — Let $g(z) = \sum_{n=0}^{\infty} n a_n z^{n-1}$. We first show the statement about the radius of convergence. Then $f(z)$ and $\sum_{n=0}^{\infty} n a_n z^n$ have the same radius of convergence by Hadamard's formula XII.1.4.3 and since $\lim_{n\to\infty} n^{1/n} = 1$. So, $\sum_{n=0}^{\infty} n a_n z^n$ and $g(z)$ have the same radius of convergence since $z g(z) = \sum_{n=0}^{\infty} n a_n z^n$.

For the first statement, let R be the radius of convergence of f, and we want to show holomorphy at z_0 with $|z_0| < r < R$. Write $f(z) = S_N(z) + E_N(z)$ with $S_N(z) = \sum_{n=0}^{N} a_n z^n$ and $E_N(z) = \sum_{n=N+1}^{\infty} a_n z^n$. For h with $|z_0 + h| < r$, we have

$$\begin{aligned}\frac{f(z_0+h) - f(z_0)}{h} - g(z_0) = \Big(\frac{S_N(z_0+h) - S_N(z_0)}{h} - S_N'(z_0)\Big)& \\ + (S_N'(z_0) - g(z_0)) + \Big(\frac{E_N(z_0+h) - E_N(z_0)}{h}\Big)&.\end{aligned}$$

We now bound all three terms on the right. Recall that $a^n - b^n = (a - b)(a^{n-1} + a^{n-2}b + \dots + b^{n-1})$. Taking $a = z_0 + h$ and $b = z_0$ and using $|z_0 + h| < r$ and $|z_0| < r$, we obtain

$$\Big|\frac{E_N(z_0+h) - E_N(z_0)}{h}\Big| \le \sum_{n=N+1}^{\infty} |a_n| \Big|\frac{(z_0+h)^n - z_0^n}{h}\Big| \le \sum_{n=N+1}^{\infty} |a_n| n r^{n+1} \longrightarrow 0$$

as $N \to \infty$ since $g(r)$ converges absolutely for $r < R$. Next, $S_N'(z_0) \to g(z_0)$ as $N \to \infty$ implies $|S_N'(z_0) - g(z_0)| \to 0$. Lastly, for any fixed N we have

$$\lim_{h\to 0} \Big|\frac{S_N(z_0+h) - S_N(z_0)}{h} - S_N'(z_0)\Big| = 0$$

since S_N is a polynomial.

Hence, given $\varepsilon > 0$, we choose N large enough so that the second and third term are estimated as ε. Having N, we choose $\delta > 0$ such that the first term is estimated as ε for all $|h| < \delta$. Thus,

$$\Big|\frac{f(z_0+h) - f(z_0)}{h} - g(z_0)\Big| < 3\varepsilon$$

for all $|h| < \delta$, and f is holomorphic at z_0 with derivative g. ∎

(XII.3) *Summation by parts* is the discrete analogue of integration by parts. For finite sequences $(a_n)_{n=1}^N, (b_n)_{n=1}^N \subseteq \mathbb{C}$, define partial sums $B_k = \sum_{n=1}^{k} b_n$ with $B_0 = 0$. Then

$$\sum_{n=M}^{N} a_n b_n = a_N B_N - a_M B_{M-1} - \sum_{n=M}^{N-1} (a_{n+1} - a_n) B_n.$$

***Corollary* XII.1.4.6.** — *A power series is infinitely complex differentiable (smooth) in its disc of convergence. The higher derivatives are also calculated by term-wise differentiation.*

***Example* XII.1.4.7.** — The geometric series $\sum_{n=0}^{\infty} z^n$ converges absolutely only on the unit disc. Like in the real case, there this series computes to $1/(1-z)$, which is holomorphic on $\mathbb{C} \setminus \{1\}$.

***Remark* XII.1.4.8.** — More generally, we can also consider power series centred at $z_0 \in \mathbb{C}$ of the form $f(z) = \sum_{n=0}^{\infty} a_n (z-z_0)^n$. Note that if $g(z) = \sum_{n=0}^{\infty} a_n z^n$, then $f(z) = g(z-z_0)$ is g but translated by z_0. Thus, everything about g also holds for f after translating by z_0. In particular, the radius R of convergence of f is the same as the one for g, hence given by Hadamard's formula XII.1.4.3, and the disc of convergence is $\bar{\mathrm{D}}_R(z_0)$. Moreover, by the chain rule, we have $f'(z) = g'(z-z_0) = \sum_{n=0}^{\infty} n a_n (z-z_0)^{n-1}$.

Definition* XII.1.4.9.** — *A function $f\colon \Omega \to \mathbb{C}$ on an open set Ω is* ***analytic *(or has a* power series expansion*) at $z_0 \in \Omega$ if there exists a power series $\sum_{n=0}^{\infty} a_n (z-z_0)^n$ with non-empty disc of convergence such that $f(z) = \sum_{n=0}^{\infty} (z-z_0)^n$ in an open neighbourhood of z_0. Moreover, f is* ***analytic*** *on Ω if it is analytic at every point of Ω.*

***Example* XII.1.4.10.** — The main example is the complex **exponential** function

$$\exp\colon \mathbb{C} \longrightarrow \mathbb{C}, \qquad \exp z := \mathrm{e}^z := \sum_{n=0}^{\infty} \frac{z^n}{n!},$$

which agrees with the usual definition on $\mathbb{R}$. This is absolutely convergent everywhere since $\sum_{n=0}^{\infty} |z^n/n!| = \sum_{n=0}^{\infty} |z|^n/n! = \mathrm{e}^{|z|} < \infty$. In fact, $|\sum_{n=N}^{\infty} z^n/n!| \leq \sum_{n=N}^{\infty} |z|^n/n! \to 0$ as $N \to \infty$ shows that e^z converges uniformly on any disc in $\mathbb{C}$. Thus, the exponential function is entire (holomorphic on all of $\mathbb{C}$).

***Example* XII.1.4.11.** — Similarly, the **sine** and **cosine** functions

$$\sin z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}, \qquad \cos z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}$$

agree with the usual definitions on $\mathbb{R}$, and are absolutely convergent everywhere. Thus, they are entire.

***Fact* XII.1.4.12.** — *The exponential, sine and cosine functions enjoy the following properties:*

(i) $\exp' = \exp$, $\sin' = \cos$ *and* $\cos' = -\sin$*;*
(ii) ***Euler's formulas****:*

$$\sin z = \frac{\mathrm{e}^{\mathrm{i}z} - \mathrm{e}^{-\mathrm{i}z}}{2\mathrm{i}}, \qquad \cos z = \frac{\mathrm{e}^{\mathrm{i}z} + \mathrm{e}^{-\mathrm{i}z}}{2}, \qquad \mathrm{e}^{\mathrm{i}z} = \cos z + \mathrm{i} \sin z;$$

in particular, $\exp$ *is* 2π*-periodic;*
(iii) $\mathrm{e}^{z+w} = \mathrm{e}^z \mathrm{e}^w$*;*
(iv) ***addition theorems****:*

$$\sin(z+w) = \sin z \cos w + \cos z \sin w, \qquad \cos(z+w) = \cos z \cos w - \sin z \sin w$$

(v) $\overline{\mathrm{e}^z} = \mathrm{e}^{\bar{z}}$ *and* $|\mathrm{e}^{\mathrm{i}\theta}| = 1$ *for* $\theta \in \mathbb{R}$.

Proof. — (i) and (ii) follow from direct computation with power series. (iv) follows from direct computation by invoking (ii) and (iii). The first part of (v) also follows from the representation by power series, and the second part follows from (iii) by $|\mathrm{e}^{\mathrm{i}\theta}|^2 = \mathrm{e}^{\mathrm{i}\theta}\mathrm{e}^{-\mathrm{i}\theta} = \mathrm{e}^0 = 1$. Hence, we only show (iii) from $(\mathrm{e}^z)' = \mathrm{e}^z$.

We first claim that if $f' = 0$ for an entire function f, then f is constant. Indeed, as we have seen in observation XII.1.3.10, we have $\frac{\partial f}{\partial x} = 0 = -\mathrm{i}\frac{\partial f}{\partial y}$. This shows that the real and imaginary parts of f have zero partial derivatives and are hence constant. Thus, f is constant. Also, cf. corollary XII.1.5.17 for a generalisation.

Now, we have $(\mathrm{e}^z\mathrm{e}^{w-z})' = \mathrm{e}^z\mathrm{e}^{w-z} - \mathrm{e}^z\mathrm{e}^{w-z} = 0$ by differentiating in z. By the claim above, $\mathrm{e}^z\mathrm{e}^{w-z}$ is constant, and setting $z = 0$ gives $\mathrm{e}^z\mathrm{e}^{w-z} = \mathrm{e}^w$. A shift of variables yields $\mathrm{e}^z\mathrm{e}^w = \mathrm{e}^{z+w}$. □

***Warning* XII.1.4.13.** — In contrast to the real exponential function, the complex exponential function is not injective. For example, let us solve the equation $\mathrm{e}^z = 1$. Writing $z = x + \mathrm{i}y$, we obtain $1 = \mathrm{e}^z = \mathrm{e}^x\mathrm{e}^{\mathrm{i}y}$, so $1 = |\mathrm{e}^z| = \mathrm{e}^x$ and $x = 0$. Hence, $z = \mathrm{i}y$ is purely imaginary. By Euler's formulas, $1 = \mathrm{e}^{\mathrm{i}y} = \cos y + \mathrm{i}\sin y$, so $\cos y = 1$ and $\sin y = 0$. Since the exponential function is 2π-periodic and $\mathrm{e}^0 = 1$, we see that $z = 2\pi\mathrm{i}k$ for an integer k.

***Observation* XII.1.4.14.** — Any $z \in \mathbb{C} \setminus \{0\}$ can be written in **polar form** $z = r\mathrm{e}^{\mathrm{i}\theta}$ where $r > 0$ and $\theta \in [0, 2\pi)$. Indeed, take $r = |z|$ and write $z = |z|(x + \mathrm{i}y)$. Then $x + \mathrm{i}y$ lies on the unit circle, and there is a unique $\theta \in [0, 2\pi)$ such that $x = \cos\theta$ and $y = \sin\theta$. By Euler's formulas, we obtain $z = r\mathrm{e}^{\mathrm{i}\theta}$.

We call $\arg z := \theta$ the **argument** of z. By contrast, $z = x + \mathrm{i}y$ is called the **Cartesian form**.

***Remark* XII.1.4.15.** — On the complex plane, addition is simply vector addition, so addition by a number is translation.

The polar form $z = r\mathrm{e}^{\mathrm{i}\theta}$ can be understood as a point z which has distance $|z| = r$ from the origin and has angle $\arg z = \theta$ from the real axis to the ray starting at 0 and passing through z. Under this point of view, the product of $z = r\mathrm{e}^{\mathrm{i}\theta}$ and $w = s\mathrm{e}^{\mathrm{i}\phi}$ is $zw = rs\mathrm{e}^{\mathrm{i}(\theta+\phi)}$, so multiplication by a complex number is a homothety (rotation followed by dilation).

XII.1.5. Contour integrals.

Definition* XII.1.5.1.** — *A* ***parametrised curve *is a continuous function $z(t)\colon [a, b] \to \mathbb{C}$. It is* ***smooth*** *if z is continuously differentiable on $[a, b]$ and if $z' \neq 0$ on $[a, b]$. Note that the derivative z' at $t = a$ and $t = b$ is given by the right-hand and left-hand derivatives*

$$z'(a) = \lim_{\substack{h\to 0\\ h>0}} \frac{z(a+h) - z(a)}{h}, \qquad z'(b) = \lim_{\substack{h\to 0\\ h<0}} \frac{z(b+h) - z(b)}{h},$$

resp. Similarly, z is ***piecewise-smooth*** *if there are points $a = a_0 < \cdots < a_n = b$ such that z is smooth on $[a_k, a_{k+1}]$. In particular, the right-hand and left-hand derivative at a_k might differ.*

Definition* XII.1.5.2.** — *Two parametrised curves $z\colon [a, b] \to \mathbb{C}$ and $\tilde{z}\colon [c, d] \to \mathbb{C}$ are* ***equivalent *if there is a continuously differentiable bijection $t\colon [c, d] \to [a, b]$ such that $t'(s) > 0$ and $\tilde{z}(s) = z(t(s))$. The condition $t'(s) > 0$ ensures that $t(s)$ goes from c to d as s goes from a to b, i. e., t preserves the orientation. Every equivalence class of parametrised curves defines a* ***curve*** *$\gamma \subseteq \mathbb{C}$, which is the image of any representative $z(t)\colon [a, b] \to \mathbb{C}$ together with the orientation given as t goes from a to b. The points $z(a)$ and $z(b)$ are called* ***endpoints****, and we say that γ starts at $z(a)$ and ends at $z(b)$.*

Hence, for each curve γ there is a curve γ^- with the reversed orientation (the underlying sets of γ and γ^- are the same). If γ has a parametrisation $z\colon [a, b] \to \mathbb{C}$, then a parametrisation of γ^- is, e. g., $z^-\colon [a, b] \to \mathbb{C}$, $z^-(t) = z(b + a - t)$.

***Definition* XII.1.5.3.** — *A curve is **smooth** or **piecewise-smooth** if it has such a parametrisation (in our applications, the parametrisation only needs to be (piecewise) continuously differentiable, not (piecewise) smooth). A curve is **closed** if the endpoints match for any of its parametrisations, i. e., $z(a) = z(b)$. A piecewise-smooth curve is **simple** if it does not intersect itself, i. e., it is injective except possibly for endpoints.*

***Convention* XII.1.5.4.** — *All our curves will be piecewise-smooth, so we will not mention this explicitly. Such a curve is also called **contour**.*

***Example* XII.1.5.5.** — The arguably easiest of all contours is the circle $\mathrm{C}_r(z_0)$. The **positive orientation** (counter-clockwise) of the circle is given by $z(t) = z_0 + r\mathrm{e}^{\mathrm{i}t}$ where $t \in [0, 2\pi]$, and the **negative orientation** (clockwise) is given by $z(t) = z_0 + r\mathrm{e}^{-\mathrm{i}t}$ where $t \in [0, 2\pi]$. In practice, we generally consider contours with *positive* orientation.

***Definition* XII.1.5.6.** — *Let γ be a smooth curve parametrised by $z\colon [a, b] \to \mathbb{C}$, and let $f\colon \gamma \to \mathbb{C}$ be a function on γ. The **contour integral** of f along γ is*

$$\int_\gamma f(z)\,\mathrm{d}z := \int_a^b f(z(t))z'(t)\,\mathrm{d}t.$$

If γ is piecewise-smooth, then the contour integral of f along γ is the sum of integrals along the smooth pieces, i. e., if there are $a = a_0 < \cdots < a_n = b$ such that z is smooth on $[a_k, a_{k+1}]$, then

$$\int_\gamma f(z)\,\mathrm{d}z = \sum_{k=0}^{n-1} \int_{a_k}^{a_{k+1}} f(z(t))z'(t)\,\mathrm{d}t.$$

***Definition* XII.1.5.7.** — *The **length** of a smooth curve γ is*

$$\operatorname{len}\gamma := \int_a^b |z'(t)|\,\mathrm{d}t,$$

and the length of a piecewise-smooth curve γ is the sum of lengths of the smooth pieces.

***Remark* XII.1.5.8.** — Definitions XII.1.5.6 and XII.1.5.7 are independent of their parametrisation. If $z\colon [a, b] \to \mathbb{C}$ and $\tilde{z}\colon [c, d] \to \mathbb{C}$ are equivalent parametrisations of a smooth curve, then a change of variables and the chain rule give

$$\int_a^b f(z(t))z'(t)\,\mathrm{d}t = \int_c^d f(z(t(s)))z'(t(s))t'(s)\,\mathrm{d}s = \int_c^d f(\tilde{z}(s))\tilde{z}'(s)\,\mathrm{d}s.$$

Moreover for a piecewise-smooth curve, given two partitions $a = a_0 < \cdots < a_n = b$ and $a = b_0 < \cdots < b_m = b$ for a parametrisation $z\colon [a, b] \to \mathbb{C}$, then the contour integral over the common refinement is the same since the integral is additive over intervals. Hence, the contour integral is well-defined. For the same reasons, the length of a contour is well-defined.

***Notation* XII.1.5.9.** — By abuse of notation, we denote any (all) parametrisations of a contour γ by $\gamma(t)\colon [a, b] \to \mathbb{C}$ and implicitly choose an parametrisations of γ.

***Fact* XII.1.5.10** (properties of contour integrals)**.** — *Contour integration enjoys the following properties:*

(i) *linearity:* $\int_\gamma (\alpha f(z) + \beta g(z))\,\mathrm{d}z = \alpha \int_\gamma f(z)\,\mathrm{d}z + \beta \int_\gamma g(z)\,\mathrm{d}z$ *for all* $\alpha, \beta \in \mathbb{C}$;

(ii) $\int_\gamma f(z)\,\mathrm{d}z = -\int_{\gamma^-} f(z)\,\mathrm{d}z$ *where γ^- is the contour γ with reverse orientation;*

(iii) $|\int_\gamma f(z)\,\mathrm{d}z| \le \sup_{z\in\gamma}|f(z)| \operatorname{len}\gamma$*; note that* $\sup_\gamma f$ *is finite by proposition XII.1.2.18.*

Proof. — (i) is obvious from the linearity of the Riemann integral. For (ii), let $\gamma^-(t) = \gamma(b+a-t)\colon [a,b] \to \mathbb{C}$ be a parametrisation of γ^-. Then by the chain rule and change of variables,

$$\int_a^b f(\gamma^-(t))(\gamma^-)'(t)\,\mathrm{d}t = -\int_a^b f(\gamma(b+a-t))\gamma'(b+a-t)\,\mathrm{d}t$$
$$= \int_b^a f(\gamma(s))\gamma'(s)\,\mathrm{d}s = -\int_a^b f(\gamma(s))\gamma'(s)\,\mathrm{d}s.$$

For (iii), we have

$$\left|\int_\gamma f(z)\,\mathrm{d}z\right| \leq \int_a^b |f(\gamma(t))||\gamma'(t)|\,\mathrm{d}t \leq \sup_{t\in[a,b]}|f(\gamma(t))|\int_a^b |\gamma'(t)|\,\mathrm{d}t = \sup_{z\in\gamma}|f(z)|\operatorname{len}\gamma.$$

Definition* XII.1.5.11.** — *Let $f\colon \Omega \to \mathbb{C}$ be a function on an open set Ω. A* ***primitive *for f on Ω is a function $F\colon \Omega \to \mathbb{C}$ holomorphic on Ω such that $F' = f$.*

***Fundamental theorem for contour integrals* XII.1.5.12.** — *Let $f\colon \Omega \to \mathbb{C}$ be a continuous function on an open set Ω. If $F\colon \Omega \to \mathbb{C}$ is a primitive for f and if $\gamma\colon [a,b] \to \Omega$ is a contour, then*

$$\int_\gamma f(z)\,\mathrm{d}z = F(\gamma(a)) - F(\gamma(b)).$$

Proof. — If γ is smooth, then the chain rule and the fundamental theorem of calculus yield

$$\int_\gamma f(z)\,\mathrm{d}z = \int_a^b f(\gamma(t))\gamma'(t)\,\mathrm{d}t = \int_a^b F'(\gamma(t))\gamma'(t)\,\mathrm{d}t = \int_a^b \frac{\mathrm{d}}{\mathrm{d}t}F(\gamma(t))\,\mathrm{d}t = F(\gamma(b)) - F(\gamma(a)).$$

If γ is only piecewise-smooth, then from the above we obtain the telescoping sum $\int_\gamma f(z)\,\mathrm{d}z = \sum_{k=0}^{n-1}(F(\gamma(a_{k+1})) - F(\gamma(a_k))) = F(\gamma(b)) - F(\gamma(a))$ where $a_0 = a$ and $a_n = b$.

***Corollary* XII.1.5.13.** — *Let $f\colon \Omega \to \mathbb{C}$ be a continuous function on a region Ω. Then any two primitives for f differ by a constant.*

Proof. — Let F and $\tilde{F}$ by two primitives for f. Then $\hat{F} = F - \tilde{F}$ is a primitive for 0. Fixing $w_0 \in \Omega$, we need to show that $\hat{F}(w) = \hat{F}(w_0)$ for all $w \in \Omega$. Since Ω is connected, let γ be a contour connecting w_0 to w. Then $0 = \int_\gamma 0\,\mathrm{d}z = \hat{F}(w) - \hat{F}(w_0)$.

***Notation* XII.1.5.14.** — By convention, we write the integral of f along a closed contour γ as

$$\oint_\gamma f(z)\,\mathrm{d}z := \int_\gamma f(z)\,\mathrm{d}z.$$

Such integrals play an important role in complex analysis.

***Corollary* XII.1.5.15.** — *Let $f\colon \Omega \to \mathbb{C}$ be a continuous function on an open set Ω. If f has a primitive on Ω and if γ is a closed contour in Ω, then*

$$\oint_\gamma f(z)\,\mathrm{d}z = 0.$$

***Example* XII.1.5.16.** — The function $f(z) = 1/z$ is continuous on the punctured plane $\mathbb{C} \setminus \{0\}$. If C is the unit circle parametrised by $C(t) = \mathrm{e}^{\mathrm{i}t}$ on $[0, 2\pi]$, then

$$\oint_C f(z)\,\mathrm{d}z = \int_0^{2\pi} \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}t}}{\mathrm{e}^{\mathrm{i}t}}\,\mathrm{d}t = 2\pi\mathrm{i} \neq 0,$$

whence f has no primitive on $\mathbb{C} \setminus \{0\}$.

***Corollary* XII.1.5.17.** — *Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function on a region Ω. If $f' = 0$, then f is constant.*

Proof. — For a fixed $w_0 \in \Omega$, we need to show $f(w) = f(w_0)$ for all $w \in \Omega$. Since Ω is connected, it is path-connected, meaning for any $w \in \Omega$ there is a contour γ connecting w_0 to w. Since f is a primitive for f', we obtain $0 = \int_\gamma f'(z)\,\mathrm{d}z = f(w) - f(w_0)$. ∎

§ XII.2. CAUCHY'S INTEGRAL THEOREM

april

XII.2.1. Cauchy's integral theorem.

***Goursat's theorem* XII.2.1.1.** — *Let $\Omega \subseteq \mathbb{C}$ open and let $T \subseteq \Omega$ be a triangle whose interior is also contained in Ω. If $f\colon \Omega \to \mathbb{C}$ is a holomorphic function, then*

$$\oint_T f(z)\,\mathrm{d}z = 0.$$

Proof. — Let $T^{(0)} := T$ with positive orientation, and let $d^{(0)}$ and $p^{(0)}$ be the diameter and perimeter of $T^{(0)}$, resp. By bisecting all sides, we obtain four new triangles $T_1^{(1)}, \dots, T_4^{(1)}$ with positive orientation that are similar to $T^{(0)}$, see figure XII.1. Since the orientations of sides of $T_2^{(1)}$ cancels with the orientation of the outer triangles, we obtain

$$\oint_{T^{(0)}} f(z)\,\mathrm{d}z = \oint_{T_1^{(1)}} f(z)\,\mathrm{d}z + \oint_{T_2^{(1)}} f(z)\,\mathrm{d}z + \oint_{T_3^{(1)}} f(z)\,\mathrm{d}z + \oint_{T_4^{(1)}} f(z)\,\mathrm{d}z.$$

Let j be so that the absolute value of the integral over $T_j^{(1)}$ is the largest, and set $T^{(1)} := T_j^{(1)}$. Then $|\oint_{T^{(0)}} f(z)\,\mathrm{d}z| \leq 4|\oint_{T^{(1)}} f(z)\,\mathrm{d}z|$, $d^{(1)} = \frac{1}{2}d^{(0)}$ and $p^{(1)} = \frac{1}{2}p^{(0)}$.

By iteration, we obtain triangles $T^{(0)}, T^{(1)}, \dots$ such that $|\oint_{T^{(0)}} f(z)\,\mathrm{d}z| \leq 4^n|\oint_{T^{(n)}} f(z)\,\mathrm{d}z|$, $d^{(n)} = 2^{-n}d^{(0)}$ and $p^{(n)} = 2^{-n}p^{(0)}$. If $\mathcal{T}^{(n)}$ denotes the triangle $T^{(n)}$ *together* with its interior, then $\mathcal{T}^{(0)} \supseteq \mathcal{T}^{(1)} \supseteq \cdots$ with $\operatorname{diam} \mathcal{T}^{(n)} \to 0$. By proposition XII.1.2.8, there exists a unique point $z_0 \in \bigcap_{n=0}^\infty \mathcal{T}^{(n)}$. Since f is holomorphic at z_0, write $f(z) = f(z_0) + f'(z_0)(z - z_0) + \psi(z)(z - z_0)$ where $\psi(z) \to 0$ as $z \to z_0$. Since $f(z_0)$ and $f'(z_0)(z - z_0)$ have primitives and since $|z - z_0| \leq d^{(n)}$ if $z \in T^{(n)}$, integration over $T^{(n)}$ gives

$$\left|\oint_{T^{(n)}} f(z)\,\mathrm{d}z\right| = \left|\oint_{T^{(n)}} \psi(z)(z - z_0)\,\mathrm{d}z\right| \leq \sup_{z \in T^{(n)}} |\psi(z)(z - z_0)| \operatorname{len} T^{(n)} = \varepsilon_n d^{(n)} p^{(n)}$$

by corollary XII.1.5.15 where $\varepsilon_n := \sup_{z \in T^{(n)}} |\psi(z)|$. Thus,

$$\left|\oint_{T^{(0)}} f(z)\,\mathrm{d}z\right| \leq 4^n \left|\oint_{T^{(n)}} f(z)\,\mathrm{d}z\right| \leq 4^n 4^{-n} \varepsilon_n d^{(0)} p^{(0)}$$

and $\varepsilon_n \to 0$ as $n \to 0$ finishes. ∎

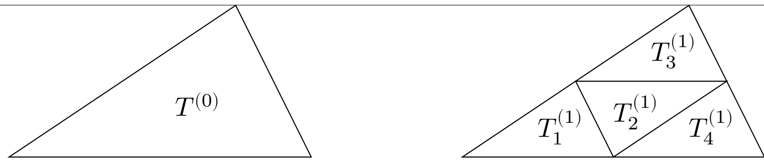

Figure XII.1: Bisection of triangles.

***Proposition* XII.2.1.2.** — *A holomorphic function $f\colon D \to \mathbb{C}$ on an open disc D has a primitive on D.*

Proof. — After translation, w. l. o. g., we assume that D is centred at the origin. For $z \in D$, let γ_z be the contour that starts at 0, goes to $\tilde{z} = \operatorname{Re} z$ horizontally, then goes to z vertically, with this very orientation. Define $F(z) = \int_{\gamma_z} f(w)\,\mathrm{d}w$.

We contend that F is holomorphic on D with $F' = f$. To see this, fix $z \in D$ and let $h \in \mathbb{C}$ be so small that $z + h \in D$. We need to investigate $F(z+h) - F(z)$. After cancellation of line segments that are integrated in both orientations, $F(z+h) - F(z)$ is the integral of f along the solid contour from z to $z+h$ in figure XII.2. By completing the trapezoid with triangles of positive orientation, Goursat's theorem XII.2.1.1 shows that $F(z+h) - F(z) = \int_\eta f(w)\,\mathrm{d}w$ where η is the straight line from z to $z+h$. Since f is continuous at z, write $f(w) = f(z) + \psi(w)$ with $\psi(w) \to 0$ as $w \to z$. Then

$$F(z+h) - F(z) = \int_\eta f(z)\,\mathrm{d}w + \int_\eta \psi(w)\,\mathrm{d}w = f(z)h + \int_\eta \psi(w)\,\mathrm{d}w.$$

Note that $|\int_\eta \psi(w)\,\mathrm{d}w| \leq \sup_{w\in\eta}|\psi(w)||h|$ and the supremum converges to 0 as $h \to 0$, so $\int_\eta \psi(w)\,\mathrm{d}w \to 0$ as $h \to 0$. Thus, F is holomorphic on D with derivative $F'(z) = f(z)$. ∎

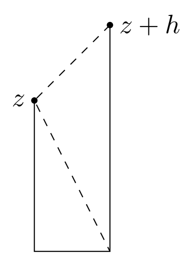

Figure XII.2: Contour of integration of $F(z+h) - F(z)$.

To generalise this statement, we need to understand on which kind of regions do primitives exist.

***Definition* XII.2.1.3.** — *Let Ω be an open set, and let $\gamma_0, \gamma_1\colon [a,b] \to \Omega$ be two curves with common endpoints $\gamma_0(a) = \gamma_1(a) = \alpha$ and $\gamma_0(b) = \gamma_1(b) = \beta$. A family of curves $\gamma_s(t)\colon [a,b] \to \Omega$ for each $s \in [0,1]$ is a **homotopy** in Ω from γ_0 to γ_1 if:*

(i) *$\gamma_s(t)|_{s=0} = \gamma_0(t)$ and $\gamma_s(t)|_{s=1} = \gamma_1(t)$ for all $t \in [a,b]$, i. e., the homotopy starts with γ_0 and ends with γ_1;*

(ii) *$\gamma_s(a) = \alpha$ and $\gamma_s(b) = \beta$ for all $s \in [0,1]$, i. e., each curve γ_s has the same endpoints as γ_0 and γ_1;*

(iii) *the function $F(s,t) = \gamma_s(t)\colon [0,1] \times [a,b] \to \Omega$ is continuous.*

We say that γ_0 and γ_1 are ***homotopic*** *in Ω if there is a homotopy in Ω from γ_0 to γ_1.*[(XII.4)]

In other words, two curves are homotopic if one can be deformed into the other by a continuous transformation inside Ω.

***Fact* XII.2.1.4.** — *Let $\gamma_1 \simeq \gamma_2$ denote that two curves γ_1 and γ_2 are homotopic in Ω. For two curves γ_1 and γ_2 in Ω where the endpoint of γ_1 is the starting point of γ_2, let $\gamma_1 \cup \gamma_2$ denote the concatenated curve.*

(i) *For $z, w \in \Omega$, the relation $\simeq$ is an equivalence relation on all curves from z to w.*
(ii) *By re-parametrisation, $(\gamma_1 \cup \gamma_2) \cup \gamma_3 \simeq \gamma_1 \cup (\gamma_2 \cup \gamma_3)$ for curves $\gamma_1, \gamma_2, \gamma_3$ in Ω which can be concatenated.*
(iii) *$z \cup \gamma \cup w \simeq \gamma$ for any curve γ in Ω starting at z and ending at w.*
(iv) *$\gamma \cup \gamma^- \simeq z$ and $\gamma^- \cup \gamma \simeq w$ for any curve γ in Ω starting at z and ending at w.*

Proof. — A rather tedious re-parametrisation of homotopies. ◻

***Proposition* XII.2.1.5.** — *Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function on an open set Ω. If γ_0 and γ_1 are homotopic contours in Ω, then*

$$\int_{\gamma_0} f(z)\,\mathrm{d}z = \int_{\gamma_1} f(z)\,\mathrm{d}z.$$

Proof. — Let $F(s,t) = \gamma_s(t)\colon [0,1] \times [a,b] \to \Omega$ be a homotopy from γ_0 to γ_1. Since the image $K \subseteq \Omega$ of F is compact, we claim that there exists $\varepsilon > 0$ such that $\mathrm{D}_{3\varepsilon}(z) \subseteq \Omega$ for all $z \in K$. Otherwise, for each $k \geq 1$ there are $z_k \in K$ and $w_k \in \mathbb{C} \setminus \Omega$ such that $|z_k - w_k| < 1/k$. By compactness in K, we can restrict to a convergent subsequence, so w. l. o. g., $z_k \to z \in K$ converges. Then $w_k \to z$, too, and $z \in \mathbb{C} \setminus \Omega$ because $\mathbb{C} \setminus \Omega$ is closed, which is absurd.

Since F is uniformly continuous (continuous on a compact set), given $\varepsilon > 0$ above there is $\delta > 0$ such that $\sup_{t\in[a,b]} |\gamma_{s_1}(t) - \gamma_{s_2}(t)| < \varepsilon$ whenever $|s_1 - s_2| < \delta$. So, fixing s_1 and s_2 with $|s_1 - s_2| < \delta$, the two contours γ_{s_0} and γ_{s_1} are ε-close. We choose discs $D_0, \dots, D_n \subseteq \Omega$ of radius 2ε such that they cover both γ_{s_1} and γ_{s_2}. Moreover, we choose points $z_0, \dots, z_{n+1} \in \gamma_{s_1}$ and $w_0, \dots, w_{n+1} \in \gamma_{s_2}$ such that $z_i, w_i \in D_{i-1} \cap D_i$ and such that $z_0 = w_0$ are the common starting points and $z_{n+1} = w_{n+1}$ are the common endpoints of γ_{s_1} and γ_{s_2}.

On each D_i, let F_i be the primitive of f by proposition XII.2.1.2. Since F_i and F_{i+1} are two primitives of f on $D_i \cap D_{i+1}$, they differ by a constant (corollary XII.1.5.13), so $F_{i+1}(z_{i+1}) - F_{i+1}(w_{i+1}) = F_i(z_{i+1}) - F_i(w_{i+1})$. By the fundamental theorem for contour integrals XII.1.5.12,

$$\begin{aligned}
\int_{\gamma_{s_1}} f(z)\,\mathrm{d}z - \int_{\gamma_{s_2}} f(z)\,\mathrm{d}z &= \sum_{i=0}^{n} (F_i(z_{i+1}) - F_i(z_i)) - \sum_{i=0}^{n} (F_i(w_{i+1}) - F_i(w_i)) \\
&= \sum_{i=0}^{n} (F_i(z_{i+1}) - F_i(w_{i+1})) - (F_i(z_i) - F_i(w_i)) \\
&= (F_n(z_{n+1}) - F_n(w_{n+1})) - (F_0(z_0) - F_0(w_0)) = 0.
\end{aligned}$$

By subdividing $[0,1]$ into finitely many intervals $[s_i, s_{i+1}]$ of length less than δ, we obtain $\int_{\gamma_0} f(z)\,\mathrm{d}z = \int_{\gamma_{s_1}} f(z)\,\mathrm{d}z = \cdots = \int_{\gamma_1} f(z)\,\mathrm{d}z$. ◻

***Remark* XII.2.1.6.** — Observe that in the prove above, we only required the homotopic contours γ_{s_1} and γ_{s_2} to have common endpoints $z_0 = w_0$ and $z_{n+1} = w_{n+1}$ in order for $F_n(z_{n+1}) - F_n(w_{n+1})$ and $F_0(z_0) - F_0(w_0)$ to cancel. Instead, if γ_{s_1} and γ_{s_2} are *closed* contours with $z_0 = z_{n+1} \in \gamma_{s_1}$ and

(XII.4) Usually, a homotopy with property (ii) is called a *pointed homotopy*.

$w_0 = w_{n+1} \in \gamma_{s_2}$, then the cancellation still happens. Therefore, we may replace property (ii) in definition XII.2.1.3 that the homotopy $\gamma_s(t)$ between γ_1 and γ_2 must be pointed.

Instead for two *closed* curves $\gamma_0, \gamma_1 \colon [a,b] \to \Omega$, we define a **homotopy** in Ω as a family of curves $\gamma_s(t) \colon [a,b] \to \Omega$ for each $s \in [0,1]$ such that:

(i) $\gamma_s(t)|_{s=0} = \gamma_0(t)$ and $\gamma_s(t)|_{s=1} = \gamma_1(t)$ for all $t \in [a,b]$, i.e., the homotopy starts with γ_0 and ends with γ_1;

(ii*) $\gamma_s(a) = \gamma_s(b)$ for all $s \in [0,1]$, i.e., each curve γ_s is closed;

(ii) the function $F(s,t) = \gamma_s(t) \colon [0,1] \times [a,b] \to \Omega$ is continuous.

***Definition* XII.2.1.7.** — *A region $\Omega \subseteq \mathbb{C}$ is **simply connected** if any two curves in Ω with common endpoints are homotopic.*

***Remark* XII.2.1.8.** — Equivalently, a region $\Omega \subseteq \mathbb{C}$ is simply connected if any closed curve in Ω is homotopic to any point (which might not lie on the closed curve).(XII.5) Indeed, suppose that any closed curve is homotopic to any point. Given two curves $\gamma_1, \gamma_2 \colon [a,b] \to \Omega$ with $\gamma_1(a) = \gamma_2(a) = \alpha$ and $\gamma_1(b) = \gamma_2(b) = \beta$, we then have the following chain of homotopies: $\gamma_1 \simeq \gamma_1 \cup \beta \simeq \gamma_1 \cup \gamma_2^- \cup \gamma_2 \simeq \alpha \cup \gamma_2 \simeq \gamma_2$.

***Proposition* XII.2.1.9.** — *A holomorphic function $f \colon \Omega \to \mathbb{C}$ on a simply connected region Ω has a primitive on Ω.*

Proof. — Fix $z_0 \in \Omega$ and define $F(z) = \int_\gamma f(w)\,\mathrm{d}w$ where γ is any contour in Ω connecting z_0 to z. This is well-defined by proposition XII.2.1.5 and since Ω is simply connected. For h small enough such that z and $z+h$ are contained in an open disc, we can write $F(z+h) - F(z) = \int_\eta f(w)\,\mathrm{d}w$ where η is the line segment from z to $z+h$. As in proposition XII.2.1.2, we show that F is holomorphic on Ω with derivative f. ◻

***Cauchy's integral theorem* XII.2.1.10.** — *Let $f \colon \Omega \to \mathbb{C}$ be a holomorphic function on a simply connected region Ω. Then*

$$\oint_\gamma f(z)\,\mathrm{d}z = 0$$

for any closed contour γ in Ω.

Proof. — Since f has a primitive on Ω by proposition XII.2.1.9, corollary XII.1.5.15 implies the statement. ◻

***Example* XII.2.1.11.** — Recall from example XII.1.5.16 that $\oint_C 1/z\,\mathrm{d}z = 2\pi\mathrm{i} \neq 0$ where C is the unit circle. Hence, the punctured plane $\mathbb{C} \setminus \{0\}$ is not simply connected.

***Definition* XII.2.1.12.** — (i) *A set $\Omega \subseteq \mathbb{C}$ is **convex** if for any $z_1, z_2 \in \Omega$, the straight line segment connecting z_1 and z_2 is contained in Ω.*

(ii) *A set $\Omega \subseteq \mathbb{C}$ is **star-shaped** if there exists $z_0 \in \Omega$ such that for any $z \in \Omega$, the straight line segment connecting z_0 and z is contained in Ω.*

***Fact* XII.2.1.13.** — *Star-shaped open sets in $\mathbb{C}$ are simply connected. In particular, convex open sets in $\mathbb{C}$ are simply connected.*

(XII.5) Topologically, this means that the first fundamental group $\pi_1(\Omega, z)$ for $z \in \Omega$ is trivial, i.e., any loop in Ω through z is null-homotopic to z.

Proof. — Let $\Omega \subseteq \mathbb{C}$ be star-shaped w. r. t. $z \in \Omega$. Then Ω is obviously path-connected. Any closed curve $\gamma\colon [a,b] \to \Omega$ is homotopic to the constant curve z by a linear homotopy $\gamma_s(t) = (1-s)\gamma(t)+sz$. Note that $\gamma_s(t) \in \Omega$ because it lies on the line segment joining $\gamma(t)$ and z, and that $\gamma_s(0) = \gamma_s(1)$ for all s.

***Example* XII.2.1.14.** — Important examples of simply connected regions are discs and the slit plane $\mathbb{C} \setminus \{(-\infty, 0]\}$. We present more direct arguments.

(i) Discs D, and in general convex sets, are simply connected since for two curves γ_0 and γ_1 in D with common endpoints, a homotopy is given by $\gamma_s(t) = (1-s)\gamma_0(t) + s\gamma_1(t)$. Note that $\gamma_s(t) \in D$ since D is convex.

(ii) The slit plane $\Omega = \mathbb{C} \setminus \{[-\infty, 0)\}$ is simply connected. For two curves γ_0 and γ_1 in Ω, write $\gamma_j(t) = r_j(t)\mathrm{e}^{\mathrm{i}\theta_j(t)}$ with $r_j(t) > 0$ continuous and $-\pi < \theta_j(t) < \pi$ continuous. Then a homotopy $\gamma_s(t) = r_s(t)\mathrm{e}^{\mathrm{i}\theta_s(t)}$ is given by $r_s = (1-s)r_0(t) + sr_1(t)$ and $\theta_s(t) = (1-s)\theta_0(t) + s\theta_1(t)$. Note that $\gamma_s(t) \in \Omega$.

***Lemma* XII.2.1.15.** — *Let $\Omega \subseteq \mathbb{C}$ be a bounded region, and let L be a straight line intersecting Ω. Suppose that $\Omega \cap L = I$ is an open line segment. As L subdivides the plane into two open half-planes, it subdivides Ω into regions Ω_1 and Ω_2. Then $\Omega = \Omega_1 \amalg I \amalg \Omega_2$.*

If Ω_1 and Ω_2 are simply connected, then Ω is simply connected.

Proof. — Let γ be a closed curve in Ω with $\gamma \cap I \neq \varnothing$. Since γ is compact, $\gamma \cap I$ is contained in a closed line segment $J \subseteq I$. Hence, there is a convex open set $U \subseteq \Omega$ containing J. In particular, U is simply connected.

Suppose that there is an arc of γ contained in Ω_1, i. e., there are $z_1, z_2 \in \gamma \cap I$ such that the open segment of γ between z_1 and z_2 is completely contained in Ω_1. Let $w_1, w_2 \in \gamma \cap U$ be as in figure XII.3. Since Ω_1 and U are simply connected, we find that the arc of γ is homotopic to the line segment between z_1 and z_2. The same goes for arcs of γ contained in Ω_2.

Since γ is compact, there can only be finitely many such arcs. Thus, γ is homotopic to a closed curve in I, which is again homotopic to a constant curve.

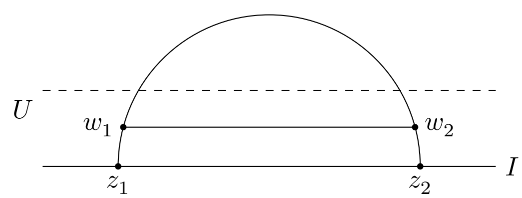

Figure XII.3: An arc of γ.

***Example* XII.2.1.16.** — The following closed contours ('toy contours') are important in practical applications. Their interiors are all simply connected by lemma XII.2.1.15.

- Circles, triangles, rectangles.
- Semicircles, parallelograms, sectors of a circle (piece of a pie).
- The contours **keyhole**, **multiple keyhole**, **rectangular keyhole**, **indented semicircle**, see figure XII.4.

So, if f is a holomorphic function on an open set containing such a contour γ with its 'interior', then $\oint_\gamma f(z)\,\mathrm{d}z = 0$.

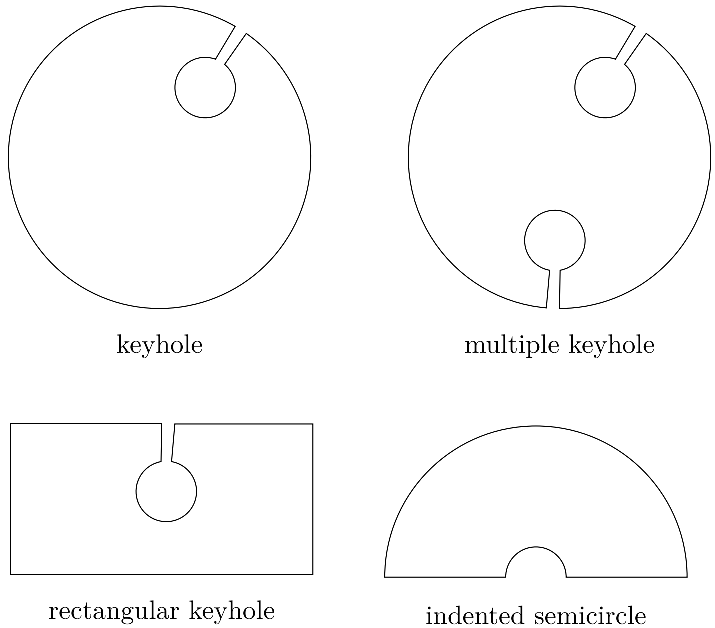

Figure XII.4: More exotic 'toy contours'.

XII.2.2. Example calculations.

The real power of complex analysis is to evaluate real integrals by means of Cauchy's integral theorem XII.2.1.10.

***Example* XII.2.2.1.** — We show that

$$\mathrm{e}^{-\pi\xi^2} = \int_{-\infty}^{\infty} \mathrm{e}^{-\pi x^2} \mathrm{e}^{-2\pi \mathrm{i} x\xi} \, \mathrm{d}x$$

for $\xi \in \mathbb{R}$, i. e., that $\mathrm{e}^{-\pi x^2}$ is its own Fourier transform.

We assume that we already know the case $\xi = 0$:

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\pi x^2} \, \mathrm{d}x = 1.$$

Alternatively, this follows from the fact that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ where Γ is the gamma function.

Insert reference for the Γ fact.

We now suppose that $\xi > 0$. Consider the toy contour γ_R as depicted in figure XII.5. Integrating the entire function $f(z) = \mathrm{e}^{-\pi z^2}$ over γ_R, Cauchy's integral theorem XII.2.1.10 gives

$$\oint_{\gamma_R} \mathrm{e}^{-\pi z^2} \, \mathrm{d}z = 0.$$

This contour integral consists of four contour integrals of each of the sides. The contour integral over the real side is

$$\int_{-R}^{R} \mathrm{e}^{-\pi x^2} \, \mathrm{d}x \longrightarrow \int_{-\infty}^{\infty} \mathrm{e}^{-\pi x^2} \, \mathrm{d}x = 1$$

as $R \to \infty$. The contour integral over the right side is estimated by

$$\left|\int_0^{\xi} \mathrm{e}^{-\pi(R+\mathrm{i}y)^2}\mathrm{i}\,\mathrm{d}y\right| \leq \int_0^{\xi} |\mathrm{e}^{-\pi(R^2+2\mathrm{i}Ry-y^2)}\mathrm{i}|\,\mathrm{d}y = \mathrm{e}^{-\pi R^2}\int_0^{\xi} \mathrm{e}^{\pi y^2}\,\mathrm{d}y$$

where the integral of the right-hand side is constant and independent of R, so the contour integral over the right side of γ_R vanishes as $R \to \infty$. The same goes for the contour integral over the left side. Finally, the contour integral over the top side is

$$\int_R^{-R} \mathrm{e}^{-\pi(x+\mathrm{i}\xi)^2}\,\mathrm{d}x = -\mathrm{e}^{\pi\xi^2}\int_{-R}^{R} \mathrm{e}^{-\pi x^2}\mathrm{e}^{-2\pi\mathrm{i}x\xi}\,\mathrm{d}x.$$

Thus, as $R \to \infty$ we have

$$0 = 1 - \mathrm{e}^{\pi\xi^2}\int_{-\infty}^{\infty} \mathrm{e}^{-\pi x^2}\mathrm{e}^{-2\pi\mathrm{i}x\xi}\,\mathrm{d}x,$$

as desired. For $\xi < 0$, we consider the contour which is the reflection of γ_R at the real axis with negative orientation.

This choice of contour is a common technique: the original contour was the real line, so we shift it vertically to apply Cauchy's integral theorem XII.2.1.10.

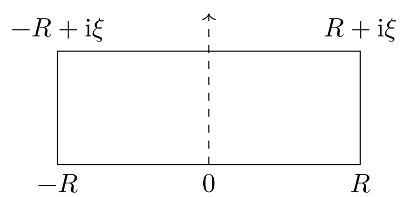

Figure XII.5: Rectangle contour γ_R of example XII.2.2.1.

***Example* XII.2.2.2.** — We want to show that

$$\int_0^{\infty} \frac{1-\cos x}{x^2}\,\mathrm{d}x = \frac{\pi}{2}.$$

Consider the contour as depicted in figure XII.6 and the function $f(z) = (1-\mathrm{e}^{\mathrm{i}z})/z^2$ holomorphic on the punctured plane $\mathbb{C}\setminus\{0\}$. By Cauchy's integral theorem XII.2.1.10, we have

$$\int_{-R}^{-\varepsilon} \frac{1-\mathrm{e}^{\mathrm{i}x}}{x^2}\,\mathrm{d}x + \int_{\varepsilon}^{R} \frac{1-\mathrm{e}^{\mathrm{i}x}}{x^2}\,\mathrm{d}x + \int_{\gamma_\varepsilon^+} \frac{1-\mathrm{e}^{\mathrm{i}z}}{z^2}\,\mathrm{d}z + \int_{\gamma_R^-} \frac{1-\mathrm{e}^{\mathrm{i}z}}{z^2}\,\mathrm{d}z = 0.$$

Since

$$\left|\int_{\gamma_R^-} \frac{1-\mathrm{e}^{\mathrm{i}z}}{z^2}\,\mathrm{d}z\right| \leq \int_{\gamma_R^-} \frac{1+|\mathrm{e}^{\mathrm{i}z}|}{|z|^2}\,\mathrm{d}z = \int_{\gamma_R^-} \frac{1+\mathrm{e}^{-\operatorname{Im} z}}{R^2}\,\mathrm{d}z \leq \frac{2}{R^2}\pi R,$$

the contour integral over γ_R^- vanishes as $R \to \infty$. Next, the power series representation of the exponential function gives

$$\frac{1-\mathrm{e}^{\mathrm{i}z}}{z} = -\mathrm{i} + \left(\frac{z}{2!} + \frac{\mathrm{i}z^2}{3!} - \frac{z^3}{4!} - \cdots\right) \longrightarrow -\mathrm{i}$$

as $z \to 0$. Since $z = \varepsilon \mathrm{e}^{\mathrm{i}\theta}$ on γ_ε^+, we hence obtain

$$\int_{\gamma_\varepsilon^+} \frac{1-\mathrm{e}^{\mathrm{i}z}}{z^2}\,\mathrm{d}z = \int_\pi^0 \frac{1-\mathrm{e}^{\mathrm{i}\varepsilon\mathrm{e}^{\mathrm{i}\theta}}}{\varepsilon\mathrm{e}^{\mathrm{i}\theta}}\mathrm{i}\,\mathrm{d}\theta \longrightarrow \int_\pi^0 (-\mathrm{i})\mathrm{i}\,\mathrm{d}\theta = -\pi$$

as $\varepsilon \to 0$. Thus,

$$\int_{-\infty}^{\infty} \frac{1-\mathrm{e}^{\mathrm{i}x}}{x^2}\,\mathrm{d}x = \pi.$$

Finally, recall that $\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x$. Taking real parts and noting that the cosine is an even function gives the result.

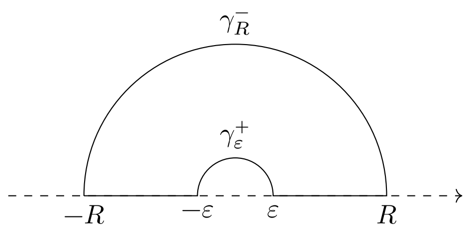

Figure XII.6: Indented semicircle contour of example XII.2.2.2.

XII.2.3. Complex logarithm.

***Motivation* XII.2.3.1.** — We want to furnish a complex logarithmic function on $\mathbb{C} \setminus \{0\}$ which shall be the inverse of the complex exponential function. If $z = r\mathrm{e}^{\mathrm{i}\theta}$, then the logarithm must naturally be

$$\log z = \log r + \mathrm{i}\theta$$

so that $\mathrm{e}^{\log z} = r\mathrm{e}^{\mathrm{i}\theta}$. Here, $\log r$ is the (real) natural logarithm for positive real numbers, and this definition agrees with the real natural logarithm.

The issue with this definition arises from warning XII.1.4.13, namely that the choice for θ is only unique up to adding $2\pi\mathbb{Z}$. Fixing such a choice θ for a given z and requiring that the argument $\theta = \arg z$ is continuous in z, this defines the logarithm *locally* at z. Such a definition cannot work *globally*, however. For instance, if z starts at 1 and winds around the origin in positive orientation until it arrives at 1 again, local continuity of $\theta = \arg z$ shows that $\log 1$ gains integer multiples of $2\pi\mathrm{i}$. As such, the complex logarithm is not a function in the ordinary sense but is 'multi-valued'. Restricting to a single-valued function is a choice of a so-called **branch** or **sheet** of the logarithm.

***Theorem* XII.2.3.2** (properties of the logarithm)**.** — *Let Ω be a simply connected set with $1 \in \Omega$ and $0 \notin \Omega$. Then there is a branch of the logarithm $F(z) =: \log_\Omega z$ in Ω such that:*

(i) *F is holomorphic in Ω;*
(ii) *$\mathrm{e}^{F(z)} = z$ for all $z \in \Omega$;*
(iii) *$F(r) = \log r$ for all $r \in \mathbb{R}$ near 1, i. e., F extends the real logarithm.*

Proof. — We shall construct F as a primitive of $1/z$. Since $0 \notin \Omega$, the function $1/z$ is holomorphic on Ω. Define $F(z) = \int_\gamma 1/w\,\mathrm{d}w$ for any contour γ in Ω connecting 1 to z. As in the proof of proposition XII.2.1.9, this is well-defined and holomorphic on Ω with $F'(z) = 1/z$. Moreover,

$$\frac{\mathrm{d}}{\mathrm{d}z}(z\mathrm{e}^{-F(z)}) = \mathrm{e}^{-F(z)} - zF'(z)\mathrm{e}^{-F(z)} = (1 - zF'(z))\mathrm{e}^{-F(z)} = 0.$$

Since Ω is connected, corollary XII.1.5.17 implies that $z\mathrm{e}^{-F(z)}$ is constant. Plugging in $z = 1$ and using $F(1) = 0$, we hence obtain $z\mathrm{e}^{-F(z)} = 1$ or, equivalently, $\mathrm{e}^{F(z)} = z$. Finally, if $r \in \mathbb{R}$ is close to 1, take the line segment from 1 to r for γ so that $F(r) = \int_1^r 1/x \,\mathrm{d}x = \log r$.

Corollary* XII.2.3.3.** — *Let $\Omega = \mathbb{C} \setminus \{(-\infty, 0]\}$ be the slit plane. For $z = r\mathrm{e}^{\mathrm{i}\theta}$ with $|\theta| < \pi$, the* ***principal branch *of the logarithm is*

$$\log z := \log_{\Omega} r\mathrm{e}^{\mathrm{i}\theta} = \log r + \mathrm{i}\theta.$$

Proof. — Let γ be the contour depicted in figure XII.7. Then for $z = r\mathrm{e}^{\mathrm{i}\theta}$ with $|\theta| < \pi$, we have

$$\log z = \int_\gamma \frac{1}{z}\,\mathrm{d}z = \int_1^r \frac{1}{x}\,\mathrm{d}x + \int_0^\theta \frac{\mathrm{i}r\mathrm{e}^{\mathrm{i}t}}{r\mathrm{e}^{\mathrm{i}t}}\,\mathrm{d}t = \log r + \mathrm{i}\theta.$$

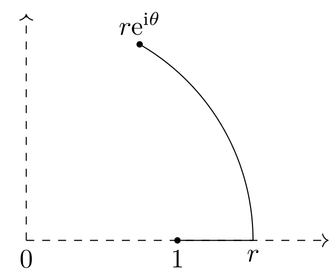

Figure XII.7: Contour for the principal branch of the logarithm.

***Warning* XII.2.3.4.** — In general, we have

$$\log(z_1 z_2) \neq \log z_1 + \log z_2$$

as opposed to $\mathrm{e}^{z_1+z_2} = \mathrm{e}^{z_1}\mathrm{e}^{z_2}$. For example, take $z_1 = z_2 = \mathrm{e}^{2\pi\mathrm{i}/3}$. On the principal branch, we have $\log z_1 = \log z_2 = 2\pi\mathrm{i}/3$, but $z_1 z_2 = \mathrm{e}^{4\pi\mathrm{i}/3} = \mathrm{e}^{-2\pi\mathrm{i}/3}$ and hence $\log(z_1 z_2) = -2\pi\mathrm{i}/3$.

***Fact* XII.2.3.5.** — *The principal branch of the logarithm admits the Taylor expansion*

$$\log(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \cdots = -\sum_{n=1}^{\infty}(-1)^n \frac{z^n}{n}$$

for $|z| < 1$.

Proof. — The derivative of the right-hand side is $\sum_{n=0}^{\infty}(-z)^n = 1/(1+z)$ and agrees with the derivative of $\log(1+z)$. By corollary XII.1.5.13, the equation holds up to a constant. Plugging in $z = 0$, both sides of the equation become zero and the constant must be zero in the first place.

***Definition* XII.2.3.6.** — *Let Ω be a simply connected set with $1 \in \Omega$ and $0 \notin \Omega$, and choose the branch of the logarithm with $\log 1 = 0$. For $\alpha \in \mathbb{C}$ and $z \in \Omega$, define the power $z^\alpha := \mathrm{e}^{\alpha \log z}$. Then $1^\alpha = 1$ and $(z^{1/n})^n = (\mathrm{e}^{\log z/n})^n = z$.*

***Proposition* XII.2.3.7** (logarithm of functions)**.** — *Let $f\colon \Omega \to \mathbb{C}$ be a nowhere vanishing holomorphic function on a simply connected region Ω. Then there exists a holomorphic function $g\colon \Omega \to \mathbb{C}$ (which is not unique!) such that $f(z) = \mathrm{e}^{g(z)}$.*

Proof. — By the chain rule, we would have $g' = (\log f)' = f'/f$, so a candidate would be the contour integral of f'/f. The proof is similar to theorem XII.2.3.2. Fix $z_0 \in \Omega$ and define $g(z) = \int_\gamma f'(w)/f(w)\,\mathrm{d}w + c_0$ where γ is a contour in Ω connecting z_0 to z and where $c_0 \in \mathbb{C}$ such that $\mathrm{e}^{c_0} = f(z_0)$. As in the proof of proposition XII.2.1.9, g is well-defined and holomorphic with $g' = f'/f$. Moreover,

$$(f\mathrm{e}^{-g})' = f'\mathrm{e}^{-g} - fg'\mathrm{e}^{-g} = (f' - fg')\mathrm{e}^{-g} = 0,$$

so $f(z)\mathrm{e}^{-g(z)}$ is constant. Plugging in $z = z_0$ yields $f(z_0)\mathrm{e}^{-g(z_0)} = \mathrm{e}^{c_0}\mathrm{e}^{-c_0} = 1$, whence $f(z) = \mathrm{e}^{g(z)}$ for all $z \in \Omega$. ∎

XII.2.4. Cauchy's integral formula.

Integral representation theorems are powerful since they allow us to express a function by an integral over a smaller set. The following result should not be surprising since holomorphic functions are harmonic. Insert reference.

***Cauchy's integral formula* XII.2.4.1.** — *Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function on an open set Ω. Suppose that Ω contains the closure of a disc D. Then*

$$f(z) = \frac{1}{2\pi\mathrm{i}} \oint_{\partial D} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta$$

for all $z \in D$ where ∂D is positively oriented.

Proof. — Let $z \in D$. We consider the keyhole contour $\Gamma_{\delta,\varepsilon} \subseteq \overline{D}$ (see figure XII.4) such that z is the centre of the keyhole and such that the outer circle is ∂D. Here, δ is the width of the corridor connecting the inner and outer circle, and ε is the radius of keyhole, the inner circle. Since $F(\zeta) = f(\zeta)/(\zeta - z)$ is holomorphic on $\Omega \setminus \{z\}$, we have $\oint_{\Gamma_{\delta,\varepsilon}} F(\zeta)\,\mathrm{d}\zeta = 0$ by Cauchy's integral theorem XII.2.1.10. We now take the limit as $\delta \to 0$. Since $\overline{D}$ is compact, F is uniformly continuous and we may pull the limit inside the integral. Using the continuity of F, the two integrals over the corridor have the same integrand in the limit and perfectly cancel. Hence, $\oint_{\Gamma_{0,\varepsilon}} F(\zeta)\,\mathrm{d}\zeta = 0$ consists of an integral over the larger outer circle ∂D with positive orientation, and an integral over the inner keyhole circle $\mathrm{C}_\varepsilon(z)$ with negative orientation. For the integral over $\mathrm{C}_\varepsilon(z)$, we have

$$\oint_{\mathrm{C}_\varepsilon(z)} F(\zeta)\,\mathrm{d}\zeta = \oint_{\mathrm{C}_\varepsilon(z)} \frac{f(\zeta) - f(z)}{\zeta - z}\,\mathrm{d}\zeta + \oint_{\mathrm{C}_\varepsilon(z)} \frac{f(z)}{\zeta - z}\,\mathrm{d}\zeta.$$

The second integral on the right-hand side is

$$\oint_{\mathrm{C}_\varepsilon(z)} \frac{f(z)}{\zeta - z}\,\mathrm{d}\zeta = f(z) \oint_0^{2\pi} \frac{-\mathrm{i}\varepsilon\mathrm{e}^{-\mathrm{i}t}}{\varepsilon\mathrm{e}^{-\mathrm{i}t}}\,\mathrm{d}t = -f(z)2\pi\mathrm{i},$$

and the first integral on the right-hand side converges to 0 as $\varepsilon \to 0$ because

$$\left| \oint_{\mathrm{C}_\varepsilon(z)} \frac{f(\zeta) - f(z)}{\zeta - z}\,\mathrm{d}\zeta \right| \leq 2\pi\varepsilon \sup_{\zeta \in \mathrm{C}_\varepsilon(z)} \frac{|f(\zeta) - f(z)|}{\varepsilon}$$

and f is continuous at z. Thus, $0 = \oint_{\partial D} F(\zeta)\,\mathrm{d}\zeta - f(z)2\pi\mathrm{i}$ as $\varepsilon \to 0$. ∎

Alternative proof. — Let $z \in D$. By Cauchy's integral theorem XII.2.1.10, we can deform ∂D to a small circle C centred at z without changing the integral $\oint_{\partial D} f(\zeta)/(\zeta - z)\,\mathrm{d}\zeta$. Now, we can write

$$\oint_{\partial D} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta = \oint_C \frac{f(\zeta) - f(z)}{\zeta - z}\,\mathrm{d}\zeta + \oint_C \frac{f(z)}{\zeta - z}\,\mathrm{d}\zeta.$$

As in the proof above, we calculate the second integral on the right-hand sice to have value $f(z)2\pi\mathrm{i}$. As for the first integral on the right-hand side, the integrand $(f(\zeta)-f(z))/(\zeta-z)$ is bounded since $|\zeta-z|$ is constant for $\zeta\in C$ and $|f(\zeta)-f(z)|$ is bounded on the compact set C. So, Riemann's theorem on removable singularities XII.3.1.5 (which has a proof not relying on Cauchy's integral theorem XII.2.1.10) says that the singularity ζ is removable and the first integral vanishes. Thus, $\oint_{\partial D} f(\zeta)/(\zeta-z)\,\mathrm{d}\zeta = f(z)2\pi\mathrm{i}$.

***Remark* XII.2.4.2.** — (i) If $z\notin D$, then the integral $\oint_{\partial D} f(\zeta)/(\zeta-z)\,\mathrm{d}\zeta$ vanishes as the integrand is holomorphic on a simply connected region containing $\overline{D}$.

(ii) Instead of a circle ∂D, Cauchy's integral formula XII.2.4.1 also works for any closed contour γ in Ω that is homotopic to ∂D in $\Omega\setminus\{z\}$ (proposition XII.2.1.5), e. g., simple closed contours.

***Cauchy's derivation formula* XII.2.4.3.** — *A holomorphic function $f\colon \Omega\to\mathbb{C}$ on an open set Ω has infinitely complex differentiable (smooth) in Ω. Moreover, if Ω contains the closure of a disc D, then*

$$f^{(n)}(z) = \frac{n!}{2\pi\mathrm{i}}\oint_{\partial D}\frac{f(\zeta)}{(\zeta-z)^{n+1}}\,\mathrm{d}\zeta$$

for all $z\in D$ where ∂D is positively oriented.

Proof. — We perform induction on n, where the initial case $n=0$ is Cauchy's integral formula XII.2.4.1. Suppose that f is $(n-1)$-times complex differentiable with the claimed integral representation for $f^{(n-1)}$. If h is so small that $z+h\in D$, the difference quotient for $f^{(n-1)}$ take the form

$$\frac{f^{(n-1)}(z+h)-f^{(n-1)}(z)}{h} = \frac{(n-1)!}{2\pi\mathrm{i}}\oint_{\partial D}\frac{f(\zeta)}{h}\left(\frac{1}{(\zeta-z-h)^n}-\frac{1}{(\zeta-z)^n}\right)\mathrm{d}\zeta.$$

Recall that $A^n-B^n=(A-B)(A^{n-1}+A^{n-2}B+\dots+B^{n-1})$. Taking $A=1/(\zeta-z-h)$ and $B=1/(\zeta-z)$, the integrand is then

$$\frac{f(\zeta)}{(\zeta-z-h)(\zeta-z)}(A^{n-1}+A^{n-2}B+\dots+AB^{n-2}+B^{n-1})\longrightarrow\frac{f(\zeta)}{(\zeta-z)^2}\frac{n}{(\zeta-z)^{n-1}}$$

as $h\to 0$. Note that this convergence is uniform in $\zeta\in\partial D$ for small h (uniform continuity of the integrand in $(\zeta,h)\in\partial D\times\overline{D}$ implies uniform convergence in ζ as $h\to 0$). Thus, we may interchange the limit $h\to 0$ and the integral to see that the difference quotient converges to

$$\frac{n!}{2\pi\mathrm{i}}\oint_{\partial D}\frac{f(\zeta)}{(\zeta-z)^{n+1}}\,\mathrm{d}\zeta = f^{(n)}(z).$$

***Cauchy's inequality* XII.2.4.4.** — *Let $f\colon\Omega\to\mathbb{C}$ be a holomorphic function on an open set Ω. If $\bar{\mathrm{D}}_R(z_0)\subseteq\Omega$, then*

$$|f^{(n)}(z_0)|\le\frac{n!\|f\|_{\mathrm{C}_R(z_0)}}{R^n}$$

where $\|f\|_{\mathrm{C}_R(z_0)}=\sup_{z\in\mathrm{C}_R(z_0)}|f(z)|$.

Proof. — Let $C=\mathrm{C}_R(z_0)$. Cauchy's derivation formula XII.2.4.3 gives

$$|f^{(n)}(z_0)| = \left|\frac{n!}{2\pi\mathrm{i}}\oint_C\frac{f(\zeta)}{(\zeta-z_0)^{n+1}}\,\mathrm{d}\zeta\right| = \frac{n!}{2\pi}\left|\int_0^{2\pi}\frac{f(z_0+R\mathrm{e}^{\mathrm{i}\theta})}{(R\mathrm{e}^{\mathrm{i}\theta})^{n+1}}R\mathrm{i}\mathrm{e}^{\mathrm{i}\theta}\,\mathrm{d}\theta\right|\le\frac{n!}{2\pi}\frac{\|f\|_C}{R^n}2\pi.$$

We saw in theorem XII.1.4.5 that analytic functions are holomorphic. Here is the converse.

***Theorem* XII.2.4.5** (holomorphic implies analytic). — *Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function on an open set Ω. If D is a disc centred at z_0 with $\overline{D} \subseteq \Omega$, then f admits a power series expansion $f(z) = \sum_{n=0}^{\infty} a_n(z-z_0)^n$ at z_0 for all $z \in D$ with coefficients*

$$a_n = \frac{f^{(n)}(z_0)}{n!}.$$

Proof. — Fix $z \in D$. By Cauchy's integral formula XII.2.4.1, we have

$$f(z) = \frac{1}{2\pi \mathrm{i}} \oint_{\partial D} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta.$$

Our idea is to use the geometric series in order to write

$$\frac{1}{\zeta - z} = \frac{1}{\zeta - z_0} \frac{1}{1 - \frac{z-z_0}{\zeta - z_0}} = \frac{1}{\zeta - z_0} \sum_{n=0}^{\infty} \left(\frac{z - z_0}{\zeta - z_0} \right)^n,$$

which is possible since there exists $0 < r < 1$ such that $|z - z_0|/|\zeta - z_0| < r$ for all $\zeta \in \partial D$. In particular, the series converges uniformly for $\zeta \in \partial D$ and we may interchange the infinite sum with the integral. Hence,

$$f(z) = \sum_{n=0}^{\infty} \frac{1}{2\pi \mathrm{i}} \oint_{\partial D} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} \, \mathrm{d}\zeta \cdot (z - z_0)^n,$$

and Cauchy's derivation formula XII.2.4.3 says that the coefficients are $f^{(n)}(z_0)/n!$. ∎

***Remark* XII.2.4.6.** — Instead of determining the coefficients by means of Cauchy's derivation formula XII.2.4.3, we could also use theorem XII.1.4.5. This gives an alternative proof of the fact that holomorphic functions are analytic on discs and of Cauchy's derivation formula XII.2.4.3.

***Remark* XII.2.4.7.** — If $f\colon \mathbb{C} \to \mathbb{C}$ is entire, then theorem XII.2.4.5 implies that f admits a *global* power series expansion at 0. In other words, $f(z) = \sum_{n=0}^{\infty} f^{(n)}/n! z^n$ converges on all of $\mathbb{C}$.

***Identity theorem* XII.2.4.8.** — *Let $f, g\colon \Omega \to \mathbb{C}$ be two holomorphic functions on a region Ω (connected open). If $f = g$ on a sequence of distinct points with limit point in Ω, e. g., on a non-empty open subset of Ω, then $f = g$ on Ω. In other words, if coincidence set $\{z \in \Omega \mid f(z) = g(z)\}$ accumulates in Ω, then $f = 0$.*[XII.6]

Proof. — By considering $f - g$ and 0 instead, w. l. o. g., let $g = 0$. Let $(w_k)_{k \geq 1}$ be a sequence of distinct points with $f(w_k) = 0$ and let $z_0 \in \Omega$ be its limit point. W. l. o. g., we may assume that $w_k \neq z_0$ for all k. We first show that $f = 0$ locally at z_0. Indeed, let $D \subseteq \Omega$ be a disc centred at z_0, and consider the power series expansion $f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n$ of f on D (theorem XII.2.4.5). If $f \neq 0$ on D, then there exists a smallest index m such that $a_m \neq 0$. We then write $f(z) = a_m(z - z_0)^m(1 + g(z - z_0))$ where g is a polynomial with no constant term. In particular, $g(z - z_0) \to 0$ as $z \to z_0$. Taking $z = w_k \to z_0$, however, $f(w_k) = 0$ and $a_m(w_k - z_0)^m \neq 0$ implies $0 = 1 + g(w_k - z_0) \to 1$, a contradiction. Thus, $f = 0$ on D.

To extend $f = 0$ to all of Ω, let U be the interior of the zero set $\{z \in \Omega \mid f(z) = 0\}$. Then U is open, and U is non-empty since $z_0 \in D \subseteq U$. Moreover, U is closed since if $(z_n)_n \subseteq U$ is a sequence with $z_n \to z$, then $f(z) = 0$ by continuity. By the above argument, which only requires that $z_n \neq z$ for all n, we have $z \in U$ already or f vanishes on a disc centred at z, in which case $z \in U$ too. Hence, $\Omega = U \amalg (\Omega \setminus U)$ is a disjoint union of two open sets, and connectedness of Ω implies $\Omega = U$. ∎

[XII.6] An **accumulation point** of a subset $\Omega \subseteq \mathbb{C}$ is a point $z \in \mathbb{C}$ such that every open neighbourhood of z meets Ω in a point other than z itself. Note that for sequences $\Omega = (w_n)_n$, a limit point is not necessarily.

Definition* XII.2.4.9.** — *Let $f\colon \Omega \to \mathbb{C}$ and $F\colon \Omega' \to \mathbb{C}$ be holomorphic (analytic) functions on regions $\Omega \subseteq \Omega'$. Then F is the* ***analytic continuation *of f into Ω' if $F = f$ on the smaller set Ω. By identity theorem XII.2.4.8, the analytic continuation is unique.*

XII.2.5. Liouville's theorem and holomorphy criteria.

***Liouville's theorem* XII.2.5.1.** — *Bounded entire functions are constant.*

Proof. — By corollary XII.1.5.17 and since $\mathbb{C}$ is connected, it suffices to show that $f' = 0$. For any $z \in \mathbb{C}$ and any $R > 0$, we have $|f'(z)| \leq \|f\|_{C_R(z)}/R \leq \|f\|_{\mathbb{C}}/R$. As $R \to \infty$, we have $f'(z) = 0$. ◻

***Fundamental theorem of algebra* XII.2.5.2.** — *Every non-constant polynomial $p(z) = a_n z^n + \dots + a_0 \in \mathbb{C}[z]$ has a root in $\mathbb{C}$. Therefore, if $p(z)$ has degree $n \geq 1$, then it has precisely n roots in $\mathbb{C}$, say, $w_1, \dots, w_n$, and we can write $p(z) = a_n(z - w_1)\cdots(z - w_n)$.*

Proof. — W. l. o. g., $a_n \neq 0$. By way of contradiction, suppose that $p(z)$ has no roots. For $z \neq 0$, we have $p(z)/z^n = a_n + a_{n-1}/z + \dots + a_0/z^n \to a_n$ as $|z| \to \infty$. Hence, there exists $R > 0$ such that $|p(z)/z^n - a_n| \leq \frac{1}{2}|a_n|$ whenever $|z| > R$, or by the reverse triangle inequality, $|p(z)| \geq \frac{1}{2}|a_n||z|^n$ whenever $|z| > R$. In particular, $p(z)$ is bounded from below when $|z| > R$. On the other hand, $p(z)$ is bounded from below by a positive constant on the compact disc $|z| \leq R$ since $p(z)$ has no roots. Thus, $1/p(z)$ is a bounded entire function, and Liouville's theorem XII.2.5.1 implies that it is constant. This contradicts the hypothesis that $p(z)$ is non-constant.

Thus, $p(z)$ has a root, say, $w_1 \in \mathbb{C}$. By long division, we can write $p(z) = (z - w_1)q(z)$ where $q(z)$ is a polynomial of degree $n - 1$ with leading coefficient a_n. By induction, we obtain $p(z) = a_n(z - w_1)\cdots(z - w_n)$. ◻

***Morera's theorem* XII.2.5.3.** — *Let $f\colon \Omega \to \mathbb{C}$ be a continuous function on an open set Ω. If $\oint_T f(z)\,\mathrm{d}z = 0$ for any triangle T contained in Ω, then f is holomorphic.*

Proof. — As in the proof of proposition XII.2.1.2, which only requires f to be continuous and the assertion of Goursat's theorem XII.2.1.1, we obtain a primitive of f in each open disc contained in Ω. Since the primitives are regular (Cauchy's derivation formula XII.2.4.3), they are twice complex differentiable and f is holomorphic everywhere. ◻

***Proposition* XII.2.5.4.** — *Let $(f_n\colon \Omega \to \mathbb{C})_{n\geq 1}$ be a sequence of holomorphic function on an open set Ω. If $f_n \to f$ uniformly on every compact subset of Ω, then $f\colon \Omega \to \mathbb{C}$ is holomorphic.*

Proof. — Let D be a disc with $\overline{D} \subseteq \Omega$ and let $T \subseteq D$ be a triangle. Since each f_n is holomorphic, Goursat's theorem XII.2.1.1 implies that $\oint_T f_n(z)\,\mathrm{d}z = 0$. By uniform convergence $f_n \to f$ on $\overline{D}$, we see that f is continuous on D and that $\oint_T f_n(z)\,\mathrm{d}z \to \oint_T f(z)\,\mathrm{d}z$. Hence, $\oint_T f(z)\,\mathrm{d}z = 0$ on D and Morera's theorem XII.2.5.3 shows that f is holomorphic on D. Thus, f is holomorphic on Ω. ◻

Alternative proof. — Let D be a disc with $\overline{D} \subseteq \Omega$. By Cauchy's integral formula XII.2.4.1 and by uniform convergence $f_n \to f$ on $\overline{D}$, we have

$$f_n(z) = \frac{1}{2\pi\mathrm{i}} \oint_{\partial D} \frac{f_n(\zeta)}{\zeta - z}\,\mathrm{d}\zeta \longrightarrow \frac{1}{2\pi\mathrm{i}} \oint_{\partial D} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta = f(z)$$

for all $z \in D$ (the right integral involving f is well-defined since f is continuous on $\overline{D}$). The proof of Cauchy's derivation formula XII.2.4.3 shows that f is holomorphic. ◻

***Counterexample* XII.2.5.5.** — This is not true in the real case. Every continuous function $f\colon [0,1] \to \mathbb{R}$ can be uniformly approximated by polynomials due to Weierstrass's approximation theorem, but not every such f is differentiable.

***Proposition* XII.2.5.6.** — *Let $(f_n\colon \Omega \to \mathbb{C})_{n\geq 1}$ be a sequence of holomorphic function on an open set Ω. If $f_n \to f$ uniformly on every compact subset of Ω, then $f_n^{(k)} \to f^{(k)}$ uniformly on every compact subset of Ω and for all $k \geq 0$.*

Proof. — We show this for $k=1$, as all higher k follow by induction. W. l. o. g., we assume that $f_n \to f$ uniformly on all of Ω. For $\delta > 0$, let $\Omega_\delta := \{z \in \Omega \mid \bar{\mathrm{D}}_\delta(z) \subseteq \Omega\}$, which is the set of all points which have distance more than δ from $\partial\Omega$. Since any compact subset of Ω is contained in some Ω_δ, it suffices to show that $f_n' \to f'$ uniformly on Ω_δ for each δ.

To see this, we claim for each holomorphic function $F\colon \Omega \to \mathbb{C}$ that

$$\sup_{z\in\Omega_\delta} |F'(z)| \leq 1/\delta \sup_{\zeta\in\Omega} |F(\zeta)|.$$

Applying this to $F = f_n - f$, then $f_n' - f' \to 0$ uniformly since $f_n - f \to 0$ uniformly, as desired. The claim follows at once from Cauchy's derivation formula XII.2.4.3 since for each $z \in \Omega_\delta$ we have $\bar{\mathrm{D}}_\delta(z) \subseteq \Omega$, whence

$$|F'(z)| = \left| \frac{1}{2\pi \mathrm{i}} \oint_{\mathrm{C}_\delta(z)} \frac{F(\zeta)}{(\zeta - z)^2}\, \mathrm{d}\zeta \right| \leq \frac{1}{2\pi} \sup_{\zeta\in\Omega} \frac{|F(\zeta)|}{\delta^2} 2\pi\delta = \frac{1}{\delta} \sup_{\zeta\in\Omega} |F(\zeta)|.$$

***Remark* XII.2.5.7.** — A common application due to Weierstrass (1841) is to consider a series $F(z) = \sum_{n=1}^\infty f_n(z)$ of holomorphic functions $f_n\colon \Omega \to \mathbb{C}$. If the series converges uniformly on compact subsets of Ω, then F is holomorphic on Ω by proposition XII.2.5.4. A specific example is the Riemann zeta function. Insert reference.

***Proposition* XII.2.5.8.** — *Let $F(z,s)\colon \Omega \times [a,b] \to \mathbb{C}$ be a function with $\Omega \subseteq \mathbb{C}$ open such that:*

(i) *$F(z,s)$ is holomorphic in z for each s;*
(ii) *F is continuous on $\Omega \times [a,b]$.*

Then

$$f\colon \Omega \longrightarrow \mathbb{C}, \qquad f(z) = \int_a^b F(z,s)\, \mathrm{d}s$$

is holomorphic.

Proof. — By a linear change of variables, w. l. o. g., $a = 0$ and $b = 1$. For $n \geq 1$, consider the Riemann sum $f_n(z) = 1/n \sum_{k=1}^n F(z, k/n)$, which is holomorphic on Ω by property (i). We claim that $f_n \to f$ uniformly on each disc D with $\overline{D} \subseteq \Omega$, so that f is holomorphic on D by proposition XII.2.5.4. Thus, f is holomorphic on Ω.

To prove the claim, recall that continuous functions on compact sets are uniformly continuous. So by property (ii), for each $\varepsilon > 0$ there exists $\delta > 0$ such that $\sup_{z\in D} |F(z,s_1) - F(z,s_2)| \leq \varepsilon$ whenever $|s_1 - s_2| < \delta$ (whenever $|s_1 - s_2| < \delta$, we have $|F(z,s_1) - F(z,s_2)| < \varepsilon$ for all $z \in D$ and then take the supremum). Then, if $1/n < \delta$, we obtain

$$|f_n(z) - f(z)| = \left| \sum_{k=1}^n \int_{(k-1)/n}^{k/n} F(z,k/n) - F(z,s)\, \mathrm{d}s \right|$$
$$\leq \sum_{k=1}^n \int_{(k-1)/n}^{k/n} |F(z,k/n) - F(z,s)|\, \mathrm{d}s \leq \sum_{k=1}^n \frac{\varepsilon}{n} = \varepsilon$$

for all $z \in D$.

XII.2.6. Schwarz reflection principle.

***Motivation* XII.2.6.1.** — There are several techniques in real analysis to extend a continuous (or a more regular) function to a larger domain. This becomes more difficult if the extended function must satisfy more conditions.

In complex analysis, this problem becomes much more difficult since holomorphic functions are characteristically rigid. For instance, recall that holomorphic functions are complex smooth (Cauchy's derivation formula XII.2.4.3), and that if a function vanishes on a non-trivial line segment, then it must be zero by the identity theorem XII.2.4.8.

***Counterexample* XII.2.6.2.** — Let $f(z) = \sum_{n=0}^{\infty} z^{2^n}/2^n$ whose disc of convergence is the unit disc (by Hadamard's formula XII.1.4.3, evaluate $\lim_{n\to\infty}(1/2^n)^{1/2^n} = 1$). For $|z| = 1$, note that f converges absolutely and, by the Weierstrass M-test, uniformly on the closed unit disc, so f can be extended continuously into the closed unit disc.

However, f cannot be extended analytically past the closed unit disc. Otherwise, f would be analytic in a small neighbourhood of a point on the unit circle, and this neighbourhood contains a root of unity of the form $\zeta = \mathrm{e}^{2\pi \mathrm{i}\cdot p/2^k}$. Then $f'(z) = \sum_{n=0}^{\infty} z^{2^n-1}$, and for $0 < r < 1$ we have

$$|f'(r\zeta)| = \left|\sum_{n=0}^{k-1}(r\zeta)^{2^n-1} + \sum_{n\geq k}\frac{r^{2^n}}{r\zeta}\right| \geq \sum_{n\geq k} r^{2^n-1} - \left|\sum_{n=0}^{k-1}(r\zeta)^{2^n-1}\right|.$$

Since absolute value of the finite sum on the right is a real constant, this blows up as $r \to 1$, a contradiction.

***Symmetry principle* XII.2.6.3.** — *Let $\Omega \subseteq \mathbb{C}$ be open and symmetric w. r. t. the real axis, i. e., $\overline{z} \in \Omega$ if and only if $z \in \Omega$. Let $\Omega^+ = \{z \in \Omega \mid \operatorname{Im} z > 0\}$ and $\Omega^- = \{z \in \Omega \mid \operatorname{Im} z < 0\}$ be the points of Ω on the upper and lower half-plane, resp., and let $I = \Omega \cap \mathbb{R}$, so $\Omega = \Omega^+ \amalg I \amalg \Omega^-$.*

Let $f^+ : \Omega^+ \cup I \to \mathbb{C}$ and $f^- : I \cup \Omega^- \to \mathbb{C}$ be continuous functions that are holomorphic on Ω^+ and Ω^-, resp., and that agree on I. Then the function

$$f : \Omega \longrightarrow \mathbb{C}, \qquad f(z) = \begin{cases} f^+(z) & \text{if } z \in \Omega^+, \\ f^+(z) = f^-(z) & \text{if } z \in I, \\ f^-(z) & \text{if } z \in \Omega^- \end{cases}$$

is holomorphic on Ω.

Proof. — Obviously, f is continuous on Ω, and we only need to show that f is holomorphic on I. Let $D \subseteq \Omega$ be disc centred at a point in I, and let $T \subseteq D$ be a triangle. We distinguish three cases.

(i) If T does not intersect I, then $\oint_T f(z)\,\mathrm{d}z = 0$ by holomorphy of f^+ and f^-.

(ii) If one vertex or side of T lies in I, the rest of T is completely contained in, say, Ω^+. By slightly moving the vertex or side of T on I away from I, we obtain a triangle T_ε where ε is the moving distance, see figure XII.8. Then $\oint_{T_\varepsilon} f(z)\,\mathrm{d}z = 0$ by (i), and continuity of f implies $\oint_T f(z)\,\mathrm{d}z = 0$ as $\varepsilon \to 0$.

(iii) If the interior of I intersects I, then we can triangulate T, as in figure XII.8, into subtriangles such that each of them has a vertex or side in I. By (ii), the contour integral over each of the subtriangle vanishes, so $\oint_T f(z)\,\mathrm{d}z = 0$.

By Morera's theorem XII.2.5.3, we conclude that f is holomorphic on D. ◻

***Schwarz reflection principle* XII.2.6.4.** — *Let $\Omega \subseteq \mathbb{C}$ be open and symmetric w. r. t. the real axis, and let Ω^+, I and Ω^- be as in the symmetry principle XII.2.6.3.*

Let $f : \Omega^+ \cup I \to \mathbb{C}$ be a continuous function that is holomorphic on Ω^+ and real-valued on I. Then there exists a holomorphic function $F : \Omega \to \mathbb{C}$ such that $F = f$ on Ω^+.

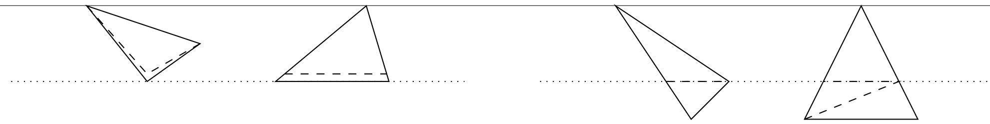

Figure XII.8: Triangle constructions for the symmetry principle XII.2.6.3. Left: (ii). Right: (iii).

Proof. — Define

$$F(z) = \begin{cases} f(z) & \text{if } z \in \Omega^+, \\ f(z) = \overline{f(\overline{z})} & \text{if } z \in I, \\ \overline{f(\overline{z})} & \text{if } z \in \Omega^-. \end{cases}$$

To show that F is holomorphic at $z_0 \in \Omega^-$, note that $\overline{z_0} \in \Omega^+$. If $f(\overline{z}) = \sum_{n=0}^{\infty} a_n (\overline{z} - \overline{z_0})^n$ is the power series expansion of f near $\overline{z_0}$, we have $F(z) = \sum_{n=0}^{\infty} \overline{a_n}(z - z_0)^n$ and F is analytic at $z_0 \in \Omega^-$. Moreover, F is continuous on $I \cup \Omega^-$. Thus, F is holomorphic by the symmetry principle XII.2.6.3.

XII.2.7. Harmonic functions and Fourier series.

The strong regularity of complex functions alludes to the theory of harmonic functions.

***Definition* XII.2.7.1.** — *For twice continuously differentiable functions $U \to \mathbb{R}$ on an open set $U \subseteq \mathbb{R}^2$, the **Laplace operator** or* Laplacian *is the differential operator*

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

*A function $F\colon U \to \mathbb{R}$ is **harmonic** if $\Delta F = 0$.*

***Lemma* XII.2.7.2.** — *In Wirtinger derivatives XII.1.3.11, we have*

$$\Delta = 4\frac{\partial}{\partial \overline{z}}\frac{\partial}{\partial z} = 4\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}}.$$

If $f\colon \Omega \to \mathbb{C}$ is a holomorphic function on an open set Ω, then $\operatorname{Re} f$ *and* $\operatorname{Im} f$ *are harmonic, and $\Delta f = 0$.*

Proof. — We have

$$\Delta = \frac{\partial}{\partial x}\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\frac{\partial}{\partial y} = \left(\frac{\partial}{\partial x} + \frac{1}{\mathrm{i}}\frac{\partial}{\partial y}\right)\left(\frac{\partial}{\partial x} - \frac{1}{\mathrm{i}}\frac{\partial}{\partial y}\right) = 4\frac{\partial}{\partial z}\frac{\partial}{\partial \overline{z}}.$$

Recall from Schwarz's theorem that partial differential operators commute. This proves the second representation for Δ.

Now, let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function. Recall from proposition XII.1.3.12 that $\partial f/\partial \overline{z} = 0$ and from observation XII.1.3.10 that $f' = \partial f/\partial z = -\mathrm{i}\partial f/\partial y$. Since f is infinitely complex differentiable, $\operatorname{Re} f$ and $\operatorname{Im} f$ are twice continuously differentiable as functions $\mathbb{R}^2 \to \mathbb{R}$. We then have $\Delta f = 0$, and taking real and imaginary parts implies $\Delta(\operatorname{Re} f) = 0$ and $\Delta(\operatorname{Im} f) = 0$. Thus, $\Delta f = 0$ and f is harmonic.

***Lemma* XII.2.7.3.** — *Let $u\colon \mathrm{D}_1(0) \to \mathbb{R}$ be a twice continuously differentiable and harmonic function. Then there exists a holomorphic function $f\colon \mathrm{D}_1(0) \to \mathbb{C}$ such that* $\operatorname{Re} f = u$*, and* $\operatorname{Im} f$ *is unique up to an additive real constant.*

Proof. — The function $g = 2\partial u/\partial z$ (which is the desired f') is holomorphic because $\partial g/\partial \overline{z} = \frac{1}{2}\Delta u = 0$ by proposition XII.1.3.12 and lemma XII.2.7.2. So, proposition XII.2.1.2 gives a primitive F of g on $\mathrm{D}_1(0)$. Then $2\partial(\operatorname{Re} F)/\partial z = F' = g = 2\partial u/\partial z$, whence $u - (\operatorname{Re} F)$ has zero partial derivatives and thus must be a real constant C. Now, take $f = F + C$ so that $\operatorname{Re} f = u$.

For the second statement, let $\tilde{f}$ be another holomorphic function such that $\operatorname{Re} \tilde{f} = u$. Then $f - \tilde{f}$ is purely imaginary, so $f - \tilde{f} = \mathrm{i} \operatorname{Im}(f - \tilde{f})$. Since $\operatorname{Im}(f - \tilde{f})$ is holomorphic, we have $\partial \operatorname{Im}(f - \tilde{f})/\partial \overline{z} = 0$, whence $\operatorname{Im}(f - \tilde{f})$ is a real constant. ∎

Complex analysis also has obvious connections to Fourier theory.

***Theorem* XII.2.7.4.** — *Let $f\colon \mathrm{D}_R(z_0) \to \mathbb{C}$ be a holomorphic function, so that it has a power series expansion $f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$ there. Then the coefficients are given by*

$$a_n = \frac{1}{2\pi r^n} \int_0^{2\pi} f(z_0 + r\mathrm{e}^{\mathrm{i}\theta})\mathrm{e}^{-\mathrm{i}n\theta}\,\mathrm{d}\theta$$

for all $0 < r < R$, which are the Fourier coefficients of $f(z_0 + r\mathrm{e}^{\mathrm{i}\theta})$ up to a factor $1/r^n$. Moreover, these Fourier coefficients vanish if $n < 0$.

Proof. — By the power series expansion of f, we have $f^{(n)}(z_0) = a_n n!$. Then fundamental theorem for contour integrals XII.1.5.12 gives

$$a_n = \frac{1}{2\pi\mathrm{i}} \oint_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\,\mathrm{d}\zeta$$

where $C = \mathrm{C}_r(z_0)$ with $0 < r < R$ and with positive orientation. Choosing the parametrisation $\zeta = z_0 + r\mathrm{e}^{\mathrm{i}\theta}$ for C, we have

$$\frac{1}{2\pi\mathrm{i}} \oint_C \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\,\mathrm{d}\zeta = \frac{1}{2\pi\mathrm{i}} \int_0^{2\pi} \frac{f(z_0 + r\mathrm{e}^{\mathrm{i}\theta})}{(r\mathrm{e}^{\mathrm{i}\theta})^{n+1}} r\mathrm{i}\mathrm{e}^{\mathrm{i}\theta}\,\mathrm{d}\theta = \frac{1}{2\pi r^n} \int_0^{2\pi} f(z_0 + r\mathrm{e}^{\mathrm{i}\theta})\mathrm{e}^{-\mathrm{i}n\theta}\,\mathrm{d}\theta,$$

proving the description for a_n. The last equations also holds if $n < 0$. Since $-n > 0$, the function $f(\zeta)(\zeta - z_0)^{-n-1}$ is holomorphic on $\mathrm{D}_r(z_0)$, so the contour integral over C vanishes by Cauchy's integral theorem XII.2.1.10. ∎

***Mean-value property* XII.2.7.5.** — *Let $f\colon \mathrm{D}_R(z_0) \to \mathbb{C}$ be a holomorphic function. Then*

$$f(z_0) = \frac{1}{2\pi} \int_0^{2\pi} f(z_0 + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta$$

for all $0 < r < R$. If $u = \operatorname{Re} f$, then also

$$u(z_0) = \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta$$

for all $0 < r < R$.

Proof. — For the first equation, take $n = 0$ in theorem XII.2.7.4 and note that $a_0 = f(z_0)$. For the second equation, we simply take the real parts of both sides. ∎

***Remark* XII.2.7.6.** — The mean-value property XII.2.7.5 is actually from harmonic analysis and is enjoyed by all harmonic functions. This follows at once from lemma XII.2.7.2.

§ XII.3. MEROMORPHIC FUNCTIONS

XII.3.1. Singularities.

***Definition* XII.3.1.1.** — *Let $\Omega \subseteq \mathbb{C}$ be an open set and let $z_0 \in \Omega$ be a point. A **deleted neighbourhood** of z_0 in Ω is a set $U \setminus \{z_0\} \subseteq \Omega \setminus \{z_0\}$ where U is an full open neighbourhood of z_0. Equivalently, a deleted neighbourhood of z_0 in Ω is a punctured disc $\mathrm{D}_r(z_0) \setminus \{z_0\} \subseteq \Omega \setminus \{z_0\}$ of some radius $r > 0$.*

***Definition* XII.3.1.2.** — *Let $\Omega \subseteq \mathbb{C}$ be an open set and let $z_0 \in \Omega$ be a point. A function $f\colon \Omega \setminus \{z_0\} \to \mathbb{C}$ has an (isolated or point) **singularity** at z_0 if f is defined on a deleted neighbourhood of z_0 in Ω, but f is not defined at z_0.*

(i) *If f can be defined at z_0 such that f becomes holomorphic on all of Ω, then z_0 is a **removable singularity** for f.*
(ii) *If $1/f$, defined by $(1/f)(z_0) = 0$ at z_0, is holomorphic in a full neighbourhood of z_0, then z_0 is a **pole** for f.*
(iii) *Otherwise, z_0 is an **essential singularity** for f.*

***Example* XII.3.1.3.** — (i) The function $f(z) = z$ on the punctured plane $\mathbb{C} \setminus \{0\}$ has a removable singularity at the origin. Indeed, we simply extend f by defining $f(0) = 0$, and f is entire.
(ii) The function $f(z) = 1/z$ on the punctured plane has a pole at the origin because $(1/f)(z) = z$ can be extended to a holomorphic function by $(1/f)(0) = 0$.
(iii) The function $f(z) = \mathrm{e}^{1/z}$ on the punctured plane has an essential singularity at the origin. Indeed, we have $f(z) \to \infty$ as $z \to 0$ on the positive real axis, but $f(z) \to 0$ as $z \to 0$ on the negative real axis, whence neither f nor $1/f$ can be analytically (or even continuously) extended to the origin.

***Fact* XII.3.1.4.** — *Any set S of isolated singularities for a meromorphic function is countable. In particular, a meromorphic function has countably many poles.*

Proof. — Each singularity can be separated from the rest of S by a disc centred at that singularity. By halving the radii, this gives us a set of disjoint discs in $\mathbb{C}$, and these must be countably many as in the proof of fact XII.1.2.14. □

***Riemann's theorem on removable singularities* XII.3.1.5.** — *Let $f\colon \Omega \setminus \{z_0\} \to \mathbb{C}$ be a holomorphic function on an open set Ω except at a singularity $z_0 \in \Omega$. If f is bounded on $\Omega \setminus \{z_0\}$, then z_0 is a removable singularity.*

Proof. — Define $h(z) = (z - z_0)^2 f(z)$ on $\Omega \setminus \{z_0\}$ and define $h(z_0) = 0$ at z_0. Then h is holomorphic on $\Omega \setminus \{z_0\}$ and also holomorphic at z_0 because

$$h'(z_0) = \lim_{z \to z_0} \frac{h(z) - h(z_0)}{z - z_0} = \lim_{z \to z_0} (z - z_0) f(z) = 0$$

where we used that f is bounded on $\Omega \setminus \{z_0\}$. Hence, h admits a power series expansion $h(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$ near z_0. However, $a_0 = h(z_0) = 0$ and $a_1 = h'(z_0) = 0$, so h factors as $h(z) = (z - z_0)^2 \sum_{n=0}^{\infty} a_{n+2} (z - z_0)^n$. Therefore, $f(z) = \sum_{n=0}^{\infty} a_{n+2} (z - z_0)^n$ is analytic near z_0. □

Alternative proof. — Since extending f to z_0 is a local problem, we only consider f on a small disc D centred at z_0 with $\overline{D} \subseteq \Omega$. If we can show that

$$f(z) = \frac{1}{2\pi \mathrm{i}} \oint_{\partial D} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta$$

for all $z \in D \setminus \{z_0\}$ where ∂D is positively oriented, then proposition XII.2.5.8 says that the integral defines a holomorphic function on *all* of D which agrees with f on $D \setminus \{z_0\}$. This gives the desired extension.

To show this, consider the multiple keyhole contour where the outer circle is ∂D and with two negatively oriented keyholes: one centred at a fixed $z \in D \setminus \{z_0\}$ and one centred at z_0. Making the corridors narrower so that they cancel in the limit, we obtain

$$\oint_{\partial D} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}z + \oint_{\gamma_\varepsilon} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}z + \oint_{\gamma'_\varepsilon} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}z = 0$$

by Cauchy's integral formula XII.2.4.1 where γ_ε and γ'_ε are small circles of radius ε centred at z and z_0, resp. By the same argument as in the proof of Cauchy's integral formula, we have

$$\oint_{\gamma_\varepsilon} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta = -2\pi\mathrm{i} f(z).$$

Since f is bounded by, say, $C > 0$, we have

$$\left| \oint_{\gamma'_\varepsilon} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta \right| \leq \sup_{\zeta \in \gamma'_\varepsilon} \frac{C}{|\zeta - z|} 2\pi\varepsilon.$$

Since $\sup_{\zeta \in \gamma'_\varepsilon} 1/|\zeta - z|$ is bounded from above for all small ε, the integral estimate tends to zero as $\varepsilon \to 0$. This proves the claim. □

***Corollary* XII.3.1.6.** — *Let $f\colon \Omega \setminus \{z_0\} \to \mathbb{C}$ be a holomorphic function on an open set Ω except at a singularity $z_0 \in \Omega$. Then z_0 is a pole of f if and only if $f(z) \to \infty$ as $z \to z_0$.*

Proof. — Let z_0 be a pole of f. Then $1/|f(z)| = |(1/f)(z)| \to 0$ as $z \to z_0$, so $|f(z)| \to \infty$. Conversely, suppose that $|f(z)| \to \infty$ as $z \to z_0$. Then $1/|f(z)| \to 0$ as $z \to z_0$ and $1/f$ is bounded near z_0. Hence, z_0 is a removable singularity of $1/f$ and it must be defined by $(1/f)(z_0) = 0$ at z_0. In other words, z_0 is a pole of f. □

***Casorati-Weierstrass theorem* XII.3.1.7.** — *Let $f\colon \Omega \setminus \{z_0\} \to \mathbb{C}$ be a holomorphic function on an open set Ω except at a singularity $z_0 \in \Omega$. If z_0 is an essential singularity for f, then the image $f(\Omega \setminus \{z_0\})$ is dense in $\mathbb{C}$.*

Proof. — By way of contradiction, assume that there exists some $w \in \mathbb{C}$ that is not approximated by points in the image of f. That is to say there is $\delta > 0$ such that $|f(z) - w| > \delta$ for all $z \in \Omega \setminus \{z_0\}$. Define $g(z) = 1/(f(z) - w)$ on $\Omega \setminus \{z_0\}$ which is bounded by $1/\delta$. Due to Riemann's theorem on removable singularities XII.3.1.5, the singularity z_0 is removable for g. If $g(z_0) \neq 0$, then $f(z) - w = 1/g(z)$ is holomorphic at z_0, contradicting that z_0 is not a removable singularity for f. If $g(z_0) = 0$, then $f(z) - w$ has a pole at z_0 by corollary XII.3.1.6, again contradicting that z_0 is not a pole for f. □

***Remark* XII.3.1.8.** — Picard proved a much stronger result that a function $f\colon \Omega \setminus \{z_0\} \to \mathbb{C}$ with an essential singularity at z_0 takes every value in $\mathbb{C}$ infinitely many times, except for one exception.

***Proposition* XII.3.1.9.** — *Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function on an open set Ω. If Ω contains an annulus $A = \{z \in \mathbb{C} \,|\, r_1 \leq |z - z_0| \leq r_2\}$ centred at z_0 where $0 \leq r_1 < r_2$, then f admits a **Laurent series** expansion $f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$ at z_0 for all $z \in \mathring{A}$.*

Proof. — Fix $z \in \mathring{A}$. We consider a keyhole contour containing z whose outer circle has radius r_2 and whose keyhole has radius r_1. By narrowing the corridor so that the integrals over the corridor cancel in the limit, Cauchy's integral formula XII.2.4.1 gives

$$f(z) = \frac{1}{2\pi\mathrm{i}} \oint_{C_2} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta - \frac{1}{2\pi\mathrm{i}} \oint_{C_1} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta$$

where C_1 and C_2 are positively oriented circles of radius r_1 and r_2, resp., bounding A. By the same argument as in the proof of theorem XII.2.4.5 (note that $|z - z_0|/|\zeta - z_0| < 1$ for $\zeta \in C_2$ and $|\zeta - z_0|/|z - z_0| < 1$ for $\zeta \in C_1$), we obtain

$$\begin{aligned} f(z) &= \frac{1}{2\pi\mathrm{i}} \oint_{C_2} \frac{f(\zeta)}{\zeta - z_0} \sum_{n=0}^{\infty} \left(\frac{z - z_0}{\zeta - z_0}\right)^n \mathrm{d}\zeta - \frac{1}{2\pi\mathrm{i}} \oint_{C_1} \frac{f(\zeta)}{z - z_0} \sum_{n=0}^{\infty} \left(\frac{\zeta - z_0}{z - z_0}\right)^n \mathrm{d}\zeta \\ &= \frac{1}{2\pi\mathrm{i}} \sum_{n=0}^{\infty} \oint_{C_2} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\,\mathrm{d}\zeta \cdot (z - z_0)^n - \frac{1}{2\pi\mathrm{i}} \sum_{n=0}^{\infty} \oint_{C_1} f(\zeta)(\zeta - z_0)^n\,\mathrm{d}\zeta \cdot (z - z_0)^{-n-1}. \end{aligned}$$

***Definition* XII.3.1.10.** — *Let $f\colon \Omega \setminus \{z_0\} \to \mathbb{C}$ be a holomorphic function on an open set Ω except at a singularity $z_0 \in \Omega$. Then f admits a Laurent series expansion $f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n$ on a small deleted disc centred at z_0. We call the series $\sum_{n=1}^{\infty} a_{-n}/(z - z_0)^n$ the **principal part** of f at z_0, and we call $\operatorname{res}_{z_0} f = a_{-1}$ the **residue** of f at z_0.*

XII.3.2. The residue theorem.

***Motivation* XII.3.2.1.** — Suppose that γ is a closed contour in $\mathbb{C}$ and that $z \in \mathbb{C}$ is a point not on γ. The intuitive idea to count how often γ winds around z is to consider how $\arg(\zeta - z)$ changes as ζ travels on γ. For every full loop around z, the argument $\arg(\zeta - z)$ changes by 2π. Since $\log w = \log|w| + \mathrm{i}\arg w$, we can capture this change by the integer

$$\frac{1}{2\pi\mathrm{i}}(\log(\zeta_1 - z) - \log(\zeta_2 - z)) \mathrel{\hat{=}} \frac{1}{2\pi\mathrm{i}} \oint_{\gamma} \frac{1}{\zeta - z}\,\mathrm{d}\zeta$$

where ζ_1 and ζ_2 are the endpoints of γ (note that the branch choice does not matter because the branch contribution cancels in the difference).

***Definition* XII.3.2.2.** — *Let γ be a closed contour in $\mathbb{C}$. The **winding number** of γ around a point $z \notin \gamma$ is*

$$\mathrm{W}_{\gamma}(z) = \frac{1}{2\pi\mathrm{i}} \oint_{\gamma} \frac{1}{\zeta - z}\,\mathrm{d}\zeta.$$

*This is also called the **index** of z w.r.t. γ.*

***Example* XII.3.2.3.** — For $k \in \mathbb{N}$, consider the unit circle $\gamma\colon [0, 2\pi] \to \mathbb{C}$, $\gamma(t) = \mathrm{e}^{\mathrm{i}kt}$ traversed k times in the positive direction. Then

$$\mathrm{W}_{\gamma}(z) = \frac{k}{2\pi\mathrm{i}} \int_{\mathrm{C}_1(0)} \frac{1}{\zeta - z}\,\mathrm{d}\zeta = \begin{cases} k & \text{if } |z| < 1, \\ 0 & \text{if } |z| > 1. \end{cases}$$

Similarly, if $\gamma(t) = \mathrm{e}^{-\mathrm{i}kt}$ is the unit circle traversed k times in the negative direction, then

$$\mathrm{W}_{\gamma}(z) = \begin{cases} -k & \text{if } |z| < 1, \\ 0 & \text{if } |z| > 1. \end{cases}$$

If γ is a toy contour, then

$$\mathrm{W}_\gamma(z) = \begin{cases} 1 & \text{if } z \text{ is in the interior of } \gamma, \\ 0 & \text{if } z \text{ is in the exterior of } \gamma. \end{cases}$$

***Fact* XII.3.2.4** (properties of winding numbers). — *Let γ be a closed curve in $\mathbb{C}$. Then the following hold about winding numbers:*

(i) W_γ *is integer-valued on the complement of γ;*
(ii) W_γ *is constant on connected components of the complement of γ;*
(iii) W_γ *vanishes on the unbounded connected component of the complement of γ.*

Proof. — For (i), let $\gamma\colon [0,1] \to \mathbb{C}$ be a parametrisation of γ, and define $G(t) = \int_0^t \gamma'(s)/(\gamma(s)-z)\,\mathrm{d}s$. By the fundamental theorem of calculus, G is continuous and differentiable with $G'(t) = \gamma'(t)/(\gamma(t)-z)$ whenever $\gamma'(t)$ is defined (γ is piecewise-smooth). Then $H(t) = (\gamma(t)-z)\mathrm{e}^{-G(t)}$ is continuous and piecewise-smooth with derivative $H'(t) = 0$ almost everywhere, and continuity implies that $H(t) = c$ is constant. With $\gamma(0) = \gamma(1)$ (the contour is closed), we obtain

$$1 = \mathrm{e}^{G(0)} = \frac{\gamma(0)-z}{c} = \frac{\gamma(1)-z}{c} = \mathrm{e}^{G(1)}.$$

Thus, $G(1) \in 2\pi\mathrm{i}\mathbb{Z}$. Since $G(1) = 2\pi\mathrm{i}\,\mathrm{W}_\gamma(z)$, we are done with (i).

As for (ii), note that $\mathrm{W}_\gamma\colon \mathbb{C}\setminus\gamma \to \mathbb{Z}$, $z \mapsto (1/2\pi\mathrm{i})\oint_\gamma 1/(\zeta - z)\,\mathrm{d}\zeta$ is a continuous function, which can be justified by the same argument as in Cauchy's derivation formula XII.2.4.3. Hence, it must be constant on connected components of $\mathbb{C}\setminus\gamma$.

Finally, we have

$$|\mathrm{W}_\gamma(z)| \le \frac{1}{2\pi}\sup_{\zeta\in\gamma}\frac{1}{|\zeta - z|}\operatorname{len}\gamma$$

and the right-hand side converges to 0 as $|z| \to \infty$. Therefore, $\lim_{|z|\to\infty}\mathrm{W}_\gamma(z) = 0$ and (ii) shows that $\mathrm{W}_\gamma = 0$ on the unbounded connected component. This proves (iii).

The celebrated residue theorem provides a strong generalisation of Cauchy's integral formula XII.2.4.1.

***Residue theorem* XII.3.2.5.** — *Let $f\colon \Omega\setminus S \to \mathbb{C}$ be a holomorphic function on a simply connected open set Ω except for a set $S \subseteq \Omega$ of isolated singularities. Then*

$$\oint_\gamma f(z)\,\mathrm{d}z = 2\pi\mathrm{i}\sum_{z\in S}\mathrm{W}_\gamma(z)\operatorname{res}_z f$$

for any closed contour γ in $\Omega\setminus S$ and this sum is finite.

Proof. — We first show that the sum is finite. Since γ is compact, it is contained in a large closed disc K. Then $K\setminus S$ together with the separating neighbourhoods of points in S cover the compact K, whence K can only contain finitely many points of S. All other points of S must have zero winding number. Therefore, w. l. o. g., we assume that $S = \{z_1, \dots, z_N\}$ is finite.

For each z_j, let $P_j(z) = \sum_{n=1}^\infty a_{-n}^{(j)}/(z-z_j)^n$ be the principal part of f at z_j. Since γ is compact, there exists $\zeta_0 \in \gamma$ with $|\zeta - z_j| \ge |\zeta_0 - z_j| =: r$ for all ζ. Then $P_j(\zeta)$ can be compared to $\sum_{n=1}^\infty |a_{-n}^{(j)}|/r^n$, which converges since $P_j(\zeta_0)$ is absolute convergent. Hence, P_j is uniformly convergent on γ by the Weierstrass M-test.

The function $f - \sum_{j=1}^{N} P_j$ has only removable singularities by Riemann's theorem on removable singularities XII.3.1.5, so Cauchy's integral theorem XII.2.1.10 gives

$$\oint_\gamma f(\zeta)\,\mathrm{d}\zeta = \sum_{j=1}^{N} \oint_\gamma P_j(\zeta)\,\mathrm{d}\zeta = \sum_{j=1}^{N} \sum_{n=1}^{\infty} \oint_\gamma \frac{a_{-n}^{(j)}}{(\zeta - z_j)^n}\,\mathrm{d}\zeta$$

where we were permitted to interchange the infinite sum and integral because γ is compact and because P_j converges uniformly on γ. Now,

$$\oint_\gamma \frac{a_{-n}^{(j)}}{(\zeta - z_j)^n}\,\mathrm{d}\zeta = \begin{cases} a_{-1}^{(j)} & \text{if } n = 1, \\ 0 & \text{otherwise} \end{cases}$$

due to Cauchy's integral formula XII.2.4.1 and Cauchy's derivation formula XII.2.4.3 applied to the constant function $a_{-n}^{(j)}$. Thus,

$$f(z) = \sum_{j=1}^{N} a_{-1}^{(j)} \oint_\gamma \frac{1}{\zeta - z_j}\,\mathrm{d}\zeta = \sum_{j=1}^{N} \operatorname{res}_{z_j} f \cdot 2\pi\mathrm{i}\, \mathrm{W}_\gamma(z_j).$$

***Observation* XII.3.2.6.** — With the residue theorem XII.3.2.5, we are able to generalise Cauchy's integral formula XII.2.4.1 to arbitrary contours by introducing winding numbers. Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function on a simply connected set Ω. Let $z \in \Omega$ and let γ be a closed contour in Ω such that $\mathrm{W}_\gamma(z) \neq 0$. Then $f(\zeta)/(\zeta - z)$ is holomorphic in ζ except at the pole z, so

$$\frac{1}{2\pi\mathrm{i}} \oint_\gamma \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta = \mathrm{W}_\gamma(z) f(z)$$

since $\operatorname{res}_{\zeta=z} 1/(\zeta - z) = -1$.

XII.3.3. Poles and meromorphic functions.

Definition* XII.3.3.1.** — *Let $f\colon \Omega \to \mathbb{C}$ be a function. A point $z_0 \in \Omega$ is a* ***zero *of f if $f(z_0) = 0$. It is an* ***isolated zero*** *if $f(z_0) = 0$, but $f \neq 0$ on a deleted neighbourhood of z_0 in Ω.*

***Remark* XII.3.3.2.** — Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function on a connected open set Ω. By the identity theorem XII.2.4.8, the zeroes of f are all isolated, except when $f = 0$.

***Proposition* XII.3.3.3** (order of a zero)**.** — *Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function on a region Ω. Suppose that f is not not the zero function on Ω. If $z_0 \in \Omega$ is a zero of f, then there exists an open neighbourhood U of z_0 in Ω, a non-vanishing holomorphic function $g\colon U \to \mathbb{C}$, and a unique integer $n \geq 1$ such that*

$$f(z) = (z - z_0)^n g(z)$$

for all $z \in U$. The number n is called the ***order*** *(or* ***multiplicity****) of the zero z_0, sometimes written* $\operatorname{ord}_f z_0$*, which describes the rate of vanishing. If $n = 1$, then the zero is* ***simple****.*

Proof. — By the identity theorem XII.2.4.8, the zero z_0 is isolated. Let $U \subseteq \Omega$ be a small disc centred at z_0 such that f is non-vanishing on $U \setminus \{z_0\}$. Then f admits a power series expansion $f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$ on U. Since f is not zero in a deleted neighbourhood of z_0, there exists a smallest n with $a_n \neq 0$. If $g(z) = \sum_{k=0}^{\infty} a_{n+k}(z - z_0)^k$, then $f(z) = (z - z_0)^n g(z)$. Clearly, g is holomorphic and non-vanishing on U (since $a_n \neq 0$).

For the uniqueness of n, suppose that $f(z) = (z - z_0)^n g(z) = (z - z_0)^m h(z)$ with $h(z_0) \neq 0$ and, w. l. o. g., $m > n$. Then $g(z) = (z - z_0)^{m-n} h(z)$ and $g(z_0) = \lim_{z \to z_0} g(z) = 0$, a contradiction. Thus, $m = n$ and $h = g$ (use the limit $z \to z_0$ for $h(z_0) = g(z_0)$), as desired.

***Corollary* XII.3.3.4** (order of a pole). — *Let $f\colon \Omega \setminus \{z_0\} \to \mathbb{C}$ be a holomorphic function on an open set Ω except for a pole at $z_0 \in \Omega$. Then there exists a neighbourhood U of z_0 in Ω, a non-vanishing holomorphic function $h\colon U \to \mathbb{C}$ and a unique integer $n \geq 1$ such that*

$$f(z) = (z - z_0)^{-n} h(z)$$

*for all $z \in U$. The number n is called the **order** (or **multiplicity**) of the pole z_0, sometimes written $\operatorname{ord}_f z_0$, which describes the rate of growth. If $n = 1$, then the pole is **simple**.*

Proof. — We may restrict the domain of $1/f$ to $U \subseteq \Omega$ such that it has an isolated zero at $z_0 \in U$. Then $1/f(z) = (z - z_0)^n g(z)$ on U for $g\colon U \to \mathbb{C}$ a holomorphic non-vanishing function. The statement follows by setting $h = 1/g$.

***Corollary* XII.3.3.5.** — *Let $f\colon \Omega \setminus \{z_0\} \to \mathbb{C}$ be a holomorphic function on an open set Ω except for a pole at $z_0 \in \Omega$ of order n. Then the Laurent expansion of f in a neighbourhood of z_0 is*

$$f(z) = \frac{a_{-n}}{(z - z_0)^n} + \frac{a_{-n+1}}{(z - z_0)^{n-1}} + \cdots + \frac{a_{-1}}{z - z_0} + G(z),$$

i. e., the principal part of f is a finite.

Proof. — Take the presentation $f(z) = (z - z_0)^{-n} h(z)$, and substitute the power series expansion $h(z) = A_0 + A_1(z - z_0) + \cdots$ of h. Then

$$f(z) = \frac{A_0}{(z - z_0)^n} + \frac{A_1}{(z - z_0)^{n-1}} + \cdots + \frac{A_{n-1}}{z - z_0} + G(z).$$

***Proposition* XII.3.3.6.** — *Let $f\colon \Omega \setminus \{z_0\} \to \mathbb{C}$ be a holomorphic function on an open set Ω except for a pole at $z_0 \in \Omega$ of order n. Then*

$$\operatorname{res}_{z_0} f = \lim_{z \to z_0} \frac{1}{(n-1)!} \frac{\mathrm{d}^{n-1}}{\mathrm{d}z^{n-1}} ((z - z_0)^n f(z)).$$

Proof. — From corollary XII.3.3.5, we have

$$(z - z_0)^n f(z) = a_{-n} + a_{-n+1}(z - z_0) + \cdots + a_{-1}(z - z_0)^{n-1} + G(z)(z - z_0)^n.$$

Differentiating $n - 1$ times leaves $(n-1)!\, a_{-1}$ plus terms with factor $(z - z_0)$. The latter terms vanish as $z \to z_0$.

***Definition* XII.3.3.7.** — *Let $\Omega \subseteq \mathbb{C}$ be an open set and let $S \subseteq \Omega$. A function $f\colon \Omega \setminus S \to \mathbb{C}$ is **meromorphic** if:*

(i) *f is holomorphic on $\Omega \setminus S$;*
(ii) *S is a set of isolated poles for f.*

Since S consists of isolated points, it is always countable.

The notion of meromorphic functions allows us to capture functions which are infinity-valued. So, it is natural to ask for the behaviour of holomorphic functions at infinity.

***Definition* XII.3.3.8.** — *The **extended complex plane** is the set $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ with the usual operations:*

(i) *$\infty + z = \infty$ for all $z \in \mathbb{C}$;*
(ii) *$\infty z = \infty$ for all $z \in \mathbb{C} \setminus \{0\}$;*

(iii) $1/0 = \infty$ *and* $1/\infty = 0$.

Definition* XII.3.3.9.** — *Let $f : \Omega \to \mathbb{C}$ be a holomorphic function on an open set Ω that contains the complement of a large disc, i. e., f is holomorphic for all large $z \in \Omega$. Then $F(z) = f(1/z)$ is holomorphic in a deleted neighbourhood of the origin. We say that f has a* ***removable singularity at infinity *(hence* ***holomorphic at infinity****),* ***pole at infinity****, or* ***essential singularity at infinity*** *if F has the property at the origin.*

A meromorphic function on the complex plane that is holomorphic at infinity or that has a pole at infinity is said to be ***meromorphic on the extended complex plane****.*

***Theorem* XII.3.3.10.** — *The meromorphic functions $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$ are the rational functions.*

Proof. — Let $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ be meromorphic on the extended complex plane. Then $f(1/z)$ has either a pole or a removable singularity at 0, and in either case $f(1/z)$ is holomorphic in a deleted disc of 0. Returning to f, this deleted disc becomes the complement of a large disc in $\mathbb{C}$ centred at 0, whence f has only finitely many poles, say, at $z_1, \dots, z_n \in \mathbb{C}$ in a bounded disc in $\mathbb{C}$.

Near each pole z_k, decompose $f(z) = f_k(z) + g_k(z)$ into the principal part f_k and the remaining holomorphic tail g_k of f at z_k. Note that f_k is a polynomial in $1/(z - z_k)$ (cf. corollary XII.3.3.5). Similarly, decompose $f(1/z) = \tilde{f}_\infty(z) + \tilde{g}_\infty(z)$ into the principal part $\tilde{f}_\infty$ (which might be zero) and the remaining holomorphic tail $\tilde{g}_\infty$ of $f(1/z)$ at 0. Finally, set $f_\infty(z) = \tilde{f}_\infty(1/z)$, which is a polynomial in z.

We claim that $C = f - f_\infty - f_1 - \dots - f_n$ is entire and bounded on $\mathbb{C}$. Indeed, all singularities z_k of C are now removable by Riemann's theorem on removable singularities XII.3.1.5, and $C(1/z)$ is now bounded near 0 since we removed the singularity at ∞ of C. This means that C is bounded outside of a large disc, and of course C is bounded on that disc, too. Thus, the claim is true and Liouville's theorem XII.2.5.1 says that C is constant on $\mathbb{C}$. Since $f_1, \dots, f_n, f_\infty$ are rational, f is rational. Conversely, every rational function is obviously meromorphic. □

***Partial fraction decomposition* XII.3.3.11.** — *Any rational function can be decomposed as a sum of a polynomial and finitely many (partial) functions of the form $z \mapsto (z - z_k)^{-n}$ where $n \in \mathbb{N}$ and $z_k \in \mathbb{C}$. Conversely, up to a multiplicative constant, a rational function f is uniquely determined by its zeroes $w_1, \dots, w_M$ and poles $z_1, \dots, z_N$ including their orders $m_1, \dots, m_M$ and $n_1, \dots, n_N$, resp., namely*

$$f(z) = \frac{(z - w_1)^{m_1} \cdots (z - w_M)^{m_M}}{(z - z_1)^{n_1} \cdots (z - z_N)^{n_N}}.$$

Proof. — We have shown the decomposition in the proof of theorem XII.3.3.10. For the converse, let f and g be rational functions with the same zeroes and poles of the same orders, and consider f/g. If z_k is a pole of order $n_k \in \mathbb{N}$, then we can write $f(z) = (z - z_k)^{-n_k} h_f(z)$ and $g(z) = (z - z_k)^{-n_k} h_g(z)$ in a neighbourhood of z_k where h_f and h_g are holomorphic there. Hence, f/g has a removable pole at z_k. Similarly for zeroes, if w_k is a zero, then f/g is holomorphic at z_k and has no zero at w_k. Since f/g has no poles, as we have see in the proof of theorem XII.3.3.10, f/g is a polynomial. Since f/g has no zeroes, f/g is constant by the fundamental theorem for contour integrals XII.1.5.12. □

We have seen that the behaviour of meromorphic functions on the extended complex plane $\hat{\mathbb{C}}$ required special handling at infinity. Instead, there is a more natural and geometric view of the extended complex plane, thereby treating all numbers in $\hat{\mathbb{C}}$ equally.

***Construction* XII.3.3.12.** — Consider the Euclidean space $\mathbb{R}^3$ in three coordinates (X, Y, Z). We identify $\mathbb{C}$ with the XY-plane, and we consider the sphere $\mathbb{S}$ of radius $\frac{1}{2}$ centred at $(0, 0, \frac{1}{2})$, i. e., such that the south pole of $\mathbb{S}$ touches the origin in $\mathbb{C}$. Let $\mathcal{N} = (0, 0, 1)$ be the north pole of $\mathbb{S}$.

There is now a homeomorphism

$$\mathbb{S} \setminus \{\mathcal{N}\} \xrightarrow{\sim} \mathbb{C}, \qquad (X, Y, Z) \longmapsto \frac{X + \mathrm{i}Y}{1 - Z}, \qquad \frac{(x, y, x^2 + y^2)}{x^2 + y^2 + 1} \longleftarrow x + \mathrm{i}y,$$

and the map $S \setminus \{\mathcal{N}\} \to \mathbb{C}$ is called the **stereographic projection**. Geometrically, for any $W \in \mathbb{S} \setminus \{\mathcal{N}\}$ there is a line through W and $\mathcal{N}$, which intersects $\mathbb{C}$ in the image point $w \in \mathbb{C}$, see figure XII.9. Conversely, for any $w \in \mathbb{C}$ there is a line through w and $\mathcal{N}$, which intersects $\mathbb{S}$ in the image point $W \in \mathbb{S}$. This intuitively 'wraps' $\mathbb{C}$ onto $\mathbb{S} \setminus \{\mathcal{N}\}$.

As $w \to \infty \in \hat{\mathbb{C}}$ (in the sense that $|w| \to \infty$), the corresponding point $W \in \mathbb{S} \setminus \{\mathcal{N}\}$ converges to $\mathcal{N}$. Hence, $\mathcal{N} \in \mathbb{S}$ is the natural candidate for $\infty \in \hat{\mathbb{C}}$. By identifying $\hat{\mathbb{C}}$ with $\mathbb{S}$, the extended complex plane $\hat{\mathbb{C}}$ is also called the **Riemann sphere**. In particular, meromorphic function are maps $\mathbb{S} \to \mathbb{S}$. An important topological property is that $\hat{\mathbb{C}} \cong \mathbb{S}$ is compact.(XII.7)

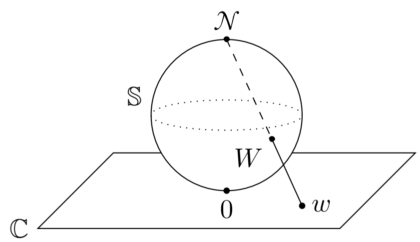

Figure XII.9: Stereographic projection of the Riemann sphere.

XII.3.4. Example calculations.

***Motivation* XII.3.4.1.** — Suppose that our goal is to evaluate the improper Riemann integral $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$. The approach of complex analysis is to find a family of toy contours $(\gamma_R)_{R\in\mathbb{R}}$, which are simple closed, such that

$$\lim_{R\to\infty} \oint_{\gamma_R} f(z)\,\mathrm{d}z = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x.$$

The contour integrals $\oint_{\gamma_R} f(z)\,\mathrm{d}z$ can then be computed by means of residue theorem XII.3.2.5. We usually assume that f has only poles, whose residues are easier to determine. Then for a pole z_0 for f, the winding number is $\mathrm{W}_\gamma(z_0) = 1$ if z_0 is in the 'interior' of γ_R and $\mathrm{W}_\gamma(z_0) = 0$ if not. The main obstacle is the choice of $(\gamma_R)_R$, which usually depends on the decay behaviour of f.

***Example* XII.3.4.2.** — We wish to show

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,\mathrm{d}x = \pi.$$

In real analysis, this could be done by knowing that $\arctan x$ is the anti-derivative of $1/(1+x^2)$. Note that by the change of variables $x \mapsto x/y$, we have $1/\pi \int_{-\infty}^{\infty} y/(x^2+y^2)\,\mathrm{d}x$, which is the integral of the Poisson kernel.

Consider the function

$$f(z) = \frac{1}{1+z^2} = \frac{1}{(z-\mathrm{i})(z+\mathrm{i})},$$

which is holomorphic everywhere except for poles at i and $-$i, and take the contour γ_R as in figure XII.10. Note that $f(z) = (z-\mathrm{i})^{-1}h(z)$ where $h(z) = 1/(z+\mathrm{i})$ is holomorphic near i. Hence, i is a simple pole with residue $\mathrm{res}_{\mathrm{i}} f = \lim_{z\to\mathrm{i}}(z-\mathrm{i})f(z) = 1/(2\mathrm{i})$. Therefore, for very large R we have

$$\oint_{\gamma_R} f(z)\,\mathrm{d}z = \frac{2\pi\mathrm{i}}{2\mathrm{i}} = \pi$$

(XII.7) This is called a *one-point compactification* since we added a point to $\mathbb{C}$ in order to make it compact.

by residue theorem XII.3.2.5. Also for sufficiently large R, we have $|1+z^2| \geq |z|^2 - 1 \geq |z|^2/2$ for all $z \in C_R^+$, whence $|f(z)| \leq 2/|z|^2$ and

$$\left| \int_{C_R^+} f(z)\,\mathrm{d}z \right| \leq \frac{2}{R^2}\pi R = \frac{2\pi}{R}.$$

This shows that the integral along C_R^+ converges to 0 as $R \to \infty$. Thus,

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \oint_{\gamma_R} f(z)\,\mathrm{d}z = \pi,$$

as desired. Of course, a similar calculation would work if we instead choose a semicircle contour on the lower half-plane with the pole $-\mathrm{i}$.

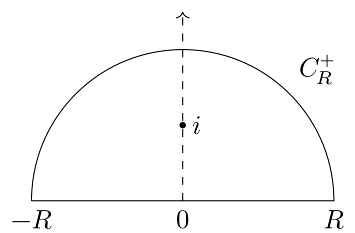

Figure XII.10: Semicircle contour γ_R for example XII.3.4.2.

***Example* XII.3.4.3.** — An integral we later need is

$$\int_{-\infty}^{\infty} \frac{\mathrm{e}^{ax}}{1+\mathrm{e}^x}\,\mathrm{d}x = \frac{\pi}{\sin \pi a}$$

for $0 < a < 1$.

Let $f(z) = \mathrm{e}^{az}/(1+\mathrm{e}^z)$ and consider the toy contour γ_R as in figure XII.11. Since the denominator $1+\mathrm{e}^z$ vanishes if and only if $z = (2k+1)\pi\mathrm{i}$ for any integer k, the only pole inside γ_R is $z = \pi\mathrm{i}$. To determine the residue, observe that

$$(z - \pi\mathrm{i})f(z) = \mathrm{e}^{az}\frac{z - \pi\mathrm{i}}{\mathrm{e}^z - \mathrm{e}^{\pi\mathrm{i}}}.$$

The quotient on the right-hand side is the reciprocal of a difference quotient and

$$\lim_{z\to\pi\mathrm{i}} \frac{\mathrm{e}^z - \mathrm{e}^{\pi\mathrm{i}}}{z - \pi\mathrm{i}} = \exp'(\pi\mathrm{i}) = \mathrm{e}^{\pi\mathrm{i}} = -1.$$

We conclude that

$$\lim_{z\to\pi\mathrm{i}} (z - \pi\mathrm{i})f(z) = -\mathrm{e}^{a\pi\mathrm{i}}$$

and $(z - \pi\mathrm{i})f(z)$ has a removable singularity in $\pi\mathrm{i}$ by Riemann's theorem on removable singularities XII.3.1.5. In other words, $\pi\mathrm{i}$ is a simple pole with residue $\operatorname{res}_{\pi\mathrm{i}} f = -\mathrm{e}^{a\pi\mathrm{i}}$. Therefore by residue theorem XII.3.2.5,

$$\oint_{\gamma_R} f(z)\,\mathrm{d}z = -2\pi\mathrm{i}\mathrm{e}^{a\pi\mathrm{i}}.$$

We now investigate the contour integrals over each side of the rectangle. Let $I_R = \int_{-R}^{R} f(x)\,\mathrm{d}x$ be the integral on the real axis. Then the integral along the top side is simply I_R but the integrand is shifted by $2\pi\mathrm{i}$:

$$-\int_{-R}^{R} f(x+2\pi\mathrm{i})\,\mathrm{d}x = -\mathrm{e}^{2\pi\mathrm{i}a} I_R.$$

We can estimate the integral over the right side by

$$\left|\int_0^{2\pi} f(R+\mathrm{i}t)\mathrm{i}\,\mathrm{d}t\right| \le \int_0^{2\pi}\left|\frac{\mathrm{e}^{a(R+\mathrm{i}t)}}{1+\mathrm{e}^{R+\mathrm{i}t}}\right|\mathrm{d}t \le \frac{\mathrm{e}^{aR}}{\mathrm{e}^R/2}2\pi$$

since $|1+\mathrm{e}^{R+\mathrm{i}t}| \ge \mathrm{e}^R - 1 \ge \mathrm{e}^R/2$ for R very large. So, the integral over the right side converges to 0 as $R \to \infty$ because $a < 1$. Similarly, the integral over the left side can be estimated by $\mathrm{e}^{-aR}/\frac{1}{2}2\pi$, which converges to 0 as $R \to \infty$ because $a > 0$. In the limit $R \to \infty$, this finally yields $I_\infty - \mathrm{e}^{2\pi\mathrm{i}a} I_\infty = -2\pi\mathrm{i}\mathrm{e}^{a\pi\mathrm{i}}$ and rearranging gives

$$I_\infty = -2\pi\mathrm{i}\frac{\mathrm{e}^{a\pi\mathrm{i}}}{1-\mathrm{e}^{2\pi\mathrm{i}a}} = \frac{2\pi\mathrm{i}}{\mathrm{e}^{\pi\mathrm{i}a}-\mathrm{e}^{-\pi\mathrm{i}a}} = \frac{\pi}{\sin\pi a}.$$

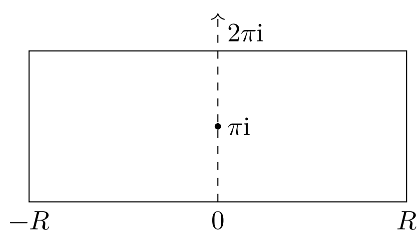

Figure XII.11: Rectangle contour γ_R of example XII.3.4.3.

***Example* XII.3.4.4.** — Let $\cosh z = \frac{1}{2}(\mathrm{e}^z + \mathrm{e}^{-z})$ be the hyperbolic cosine. We want to show that

$$\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-2\pi\mathrm{i}x\xi}}{\cosh\pi x}\,\mathrm{d}x = \frac{1}{\cosh\pi\xi},$$

i. e., that $1/\cosh\pi x$ is its own Fourier transform.

For a fixed $\xi \in \mathbb{R}$, let $f(z) = \mathrm{e}^{-2\pi\mathrm{i}z\xi}/\cosh\pi z$ and consider the toy contour γ_R as depicted in figure XII.12. Note that the denominator $\cosh\pi z$ of f vanishes if and only if $\mathrm{e}^{2\pi z} = -1$, the only poles for f in γ_R are $\alpha = \mathrm{i}/2$ and $\beta = 3\mathrm{i}/2$. For α, observe that

$$(z-\alpha)f(z) = \mathrm{e}^{-2\pi\mathrm{i}z\xi}\frac{2(z-\alpha)}{\mathrm{e}^{\pi z}+\mathrm{e}^{-\pi z}} = 2\mathrm{e}^{-2\pi\mathrm{i}z\xi}\mathrm{e}^{\pi z}\frac{z-\alpha}{\mathrm{e}^{2\pi z}-\mathrm{e}^{2\pi\alpha}}$$

since $\mathrm{e}^{2\pi\alpha} = -1$. Since the quotient on the right-hand side is the reciprocal of the difference quotient of $\mathrm{e}^{2\pi z}$ at $z = \alpha$, we obtain

$$\lim_{z\to\alpha}(z-\alpha)f(z) = 2\mathrm{e}^{-2\pi\mathrm{i}\alpha\xi}\mathrm{e}^{\pi\alpha}\frac{1}{2\pi\mathrm{e}^{2\pi\alpha}} = \frac{\mathrm{e}^{\pi\xi}}{\pi\mathrm{i}}.$$

So, $(z-\alpha)f(z)$ has a removable singularity at α by Riemann's theorem on removable singularities XII.3.1.5, and hence α is a simple pole with $\operatorname{res}_\alpha f = \mathrm{e}^{\pi\xi}/(\pi\mathrm{i})$. By a similar argument, β is a simple pole with

$$\operatorname{res}_\beta f = \lim_{z\to\beta}(z-\beta)f(z) = \lim_{z\to\beta} 2\mathrm{e}^{-2\pi\mathrm{i}z\xi}\mathrm{e}^{\pi z}\frac{z-\beta}{\mathrm{e}^{2\pi z}-\mathrm{e}^{2\pi\beta}} = \frac{2\mathrm{e}^{-2\pi\mathrm{i}\beta\xi}\mathrm{e}^{\pi\beta}}{2\pi\mathrm{e}^{2\pi\beta}} = \frac{\mathrm{e}^{3\pi\xi}}{-\pi\mathrm{i}}.$$

We now investigate the integrals over each of the sides of γ_R. For the right side, let $z = R + \mathrm{i}y$ with $0 \leq y \leq 2$. Then the numerator of f is estimated by $|\mathrm{e}^{-2\pi \mathrm{i} z\xi}| = \mathrm{e}^{2\pi y\xi} \leq \mathrm{e}^{4\pi|\xi|}$ and the denominator is estimated by

$$|\cosh \pi z| = \frac{|\mathrm{e}^{\pi z} + \mathrm{e}^{-\pi z}|}{2} \geq \frac{||\mathrm{e}^{\pi z}| - |\mathrm{e}^{-\pi z}||}{2} = \frac{\mathrm{e}^{\pi R} - \mathrm{e}^{-\pi R}}{2}.$$

Hence,

$$\left| \int_0^2 f(R + \mathrm{i}y)\mathrm{i}\,\mathrm{d}y \right| \leq \frac{2\mathrm{e}^{4\pi|\xi|}}{\mathrm{e}^{\pi R} - \mathrm{e}^{-\pi R}} 2 \longrightarrow 0$$

as $R \to \infty$. A similar argument shows the same vanishing of the integral over the left side as $R \to \infty$. If $I = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$, then the integral over the top side is $-\mathrm{e}^{4\pi\xi} I$ since $\cosh \pi z$ is 2i-periodic (recall that the integral goes in the reverse direction compared to I). The residue theorem XII.3.2.5 yields

$$I - \mathrm{e}^{4\pi\xi} I = 2\pi \mathrm{i} \left(\frac{\mathrm{e}^{\pi\xi}}{\pi \mathrm{i}} - \frac{\mathrm{e}^{3\pi\xi}}{\pi \mathrm{i}} \right) = 2\mathrm{e}^{2\pi\xi}(\mathrm{e}^{-\pi\xi} - \mathrm{e}^{\pi\xi}),$$

whence

$$I = 2\frac{\mathrm{e}^{2\pi\xi}(\mathrm{e}^{-\pi\xi} - \mathrm{e}^{\pi\xi})}{1 - \mathrm{e}^{4\pi\xi}} = 2\frac{\mathrm{e}^{-\pi\xi} - \mathrm{e}^{\pi\xi}}{\mathrm{e}^{-2\pi\xi} - \mathrm{e}^{2\pi\xi}} = 2\frac{\mathrm{e}^{-\pi\xi} - \mathrm{e}^{\pi\xi}}{(\mathrm{e}^{-\pi\xi} - \mathrm{e}^{\pi\xi})(\mathrm{e}^{-\pi\xi} + \mathrm{e}^{\pi\xi})} = \frac{2}{\mathrm{e}^{-\pi\xi} + \mathrm{e}^{\pi\xi}} = \frac{1}{\cosh \pi\xi}.$$

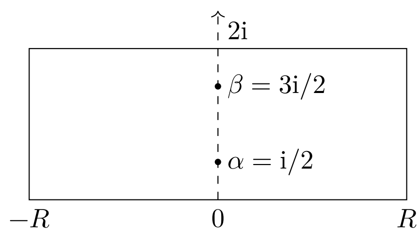

Figure XII.12: Rectangle contour γ_R of example XII.3.4.4.

***Example* XII.3.4.5.** — A similar argument to example XII.3.4.4 establishes

$$\int_{-\infty}^{\infty} \mathrm{e}^{2\pi \mathrm{i} x\xi} \frac{\sin \pi a}{\cosh \pi x + \cos \pi a}\,\mathrm{d}x = \frac{2\sinh 2\pi a\xi}{\sinh 2\pi\xi}$$

for $0 < a < 1$, where $\sinh z = \frac{1}{2}(\mathrm{e}^z - \mathrm{e}^{-z})$. In fact, example XII.3.4.4 is the case $a = \frac{1}{2}$.

XII.3.5. Argument principle.

***Argument principle* XII.3.5.1.** — *Let $f\colon \Omega \setminus S \to \mathbb{C}$ be a meromorphic function on a simply connected open set Ω. Then*

$$\frac{1}{2\pi \mathrm{i}} \oint_\gamma \frac{f'(z)}{f(z)}\,\mathrm{d}z = \sum_{\textit{zeroes } z \textit{ of } f} \mathrm{W}_\gamma(z)\,\mathrm{ord}_f\, z - \sum_{\textit{poles } p \textit{ of } f} \mathrm{W}_\gamma(p)\,\mathrm{ord}_f\, p$$

for any closed contour γ in Ω that avoids all zeroes and poles.

Proof. — Firstly, we observe that the product rule gives

$$\frac{(f_1 f_2)'}{f_1 f_2} = \frac{f_1' f_2 + f_1 f_2'}{f_1 f_2} = \frac{f_1'}{f_1} + \frac{f_2'}{f_2}.$$

If f has a zero of order n at z_0, then $f(z) = (z-z_0)^n g(z)$ where g is holomorphic and non-vanishing in a neighbourhood of z_0. Then

$$\frac{f'(z)}{f(z)} = \frac{n}{z-z_0} + \frac{g'(z)}{g(z)}$$

and f'/f has a simple pole at z_0 with residue n. Similarly, if f has a pole of order n at z_0, then $f(z) = (z-z_0)^{-n} h(z)$ where h is holomorphic and non-vanishing in a neighbourhood of z_0. Then

$$\frac{f'(z)}{f(z)} = \frac{-n}{z-z_0} + \frac{h'(z)}{h(z)}$$

and f'/f has a simple pole at z_0 with residue $-n$. These two cases describe all singularities (poles) of f'/f and the residue theorem XII.3.2.5 finishes.

***Corollary* XII.3.5.2.** — *Let $f\colon \Omega \setminus S \to \mathbb{C}$ be a meromorphic function on an open set Ω. If Ω contains the closure of a disc D whose boundary avoids zeroes and poles of f, then*

$$\frac{1}{2\pi \mathrm{i}} \oint_{\partial D} \frac{f'(z)}{f(z)}\,\mathrm{d}z = \#\{\textit{zeroes of } f \textit{ inside } D\} - \#\{\textit{poles of } f \textit{ inside } D\}$$

where zeroes and poles are counted with multiplicities.

***Rouché's theorem* XII.3.5.3** (number of zeroes is invariant under slight pertubations)**.** — *Let $f, g\colon \Omega \to \mathbb{C}$ be holomorphic functions on an open set Ω containing the closure of a disc D. If $|f(z)| > |g(z)|$ for all $z \in \partial D$, then f and $f+g$ have the same number of zeroes in D, counted with multiplicities.*

Proof. — Consider the family of functions $f_t(z) = f(z) + tg(z)$ for $t \in [0,1]$ so that $f_0 = f$ and $f_1 = f+g$. Let $n_t \in \mathbb{Z}$ be the number of zeroes of f_t in D counted with multiplicities. If $z \in \partial D$ were a zero of f_t, then $|g(z)| < |f(z)| = t|g(z)|$, a contradiction. Hence, f_t has no zeroes on the circle, and $2\pi \mathrm{i} n_t = \oint_{\partial D} f_t'(z)/f_t(z)\,\mathrm{d}z$ by the argument principle XII.3.5.1.

Observe that $f_t'(z)$ and $f_t(z)$ are continuous for $(t,z) \in [0,1] \times \partial D$ and $f_t(z)$ never vanishes on $[0,1] \times \partial D$. Hence, $f_t'(z)/f_t(z)$ is continuous on $[0,1] \times \partial D$, and so n_t is a continuous function in t because we integrate over a compact contour. But $n_t \in \mathbb{Z}$, which is why n_t is constant (use the intermediate value theorem). In particular, $n_0 = n_1$.

***Open mapping theorem* XII.3.5.4.** — *If $f\colon \Omega \to \mathbb{C}$ is a non-constant holomorphic function on a region Ω, then f is open, i. e., it maps open to open sets.*

Proof. — Let $w_0 = f(z_0)$. Choose $\delta > 0$ such that $\bar{\mathrm{D}}_\delta(z_0) \subseteq \Omega$ and such that $f(z) \neq w_0$ for all $z \in \mathrm{C}_\delta(z_0)$. This δ exists since otherwise, $f = w_0$ everywhere by identity theorem XII.2.4.8, contradicting the hypothesis. Then choose $\varepsilon > 0$ such that $|f(z) - w_0| \geq \varepsilon$ for all $z \in \mathrm{C}_\delta(z_0)$ by the extreme value theorem. Now, for $w \in \mathrm{D}_\varepsilon(w_0)$ we have $|w - w_0| < \varepsilon \leq |f(z) - w_0|$ for all $z \in \mathrm{C}_\delta(z_0)$. Hence, we may invoke Rouché's theorem XII.3.5.3 and see that $(f(z) - w_0) + (w_0 - w) = f(z) - w$ has a zero in $\mathrm{D}_\delta(z_0)$ since $f(z) - w_0$ has. Expressed differently, we have $\mathrm{D}_\varepsilon(w_0) \subseteq f(\mathrm{D}_\delta(z_0))$ and f is open.

The following statement is also enjoyed by harmonic functions.

***Definition* XII.3.5.5.** — *The **maximum** of a holomorphic function $f\colon \Omega \to \mathbb{C}$ is the maximum of the absolute value $|f|\colon \Omega \to \mathbb{R}_{\geq 0}$.*

***Maximum modulus principle* XII.3.5.6.** — *If $f\colon \Omega \to \mathbb{C}$ is a non-constant holomorphic function on a region Ω, then f cannot attain a maximum in Ω.*

Proof. — Suppose that f would attain a maximum at $z_0 \in \Omega$. By the open mapping theorem XII.3.5.4, if $D \subseteq \Omega$ is a small disc centred at z_0 then $f(D)$ is an open neighbourhood of $f(z_0)$. In the non-negative real numbers, this means that $|f(D)|$ is an open neighbourhood of $|f(z_0)|$, so there is $z \in D$ with $|f(z)| > |f(z_0)|$, contradicting maximality of z_0. ◻

***Corollary* XII.3.5.7.** — *Let Ω be a region with compact closure. If $f\colon \overline{\Omega} \to \mathbb{C}$ is continuous on $\overline{\Omega}$ and holomorphic on Ω, then*

$$\sup_{z\in\Omega}|f(z)| \le \sup_{z\in\partial\Omega}|f(z)|.$$

Proof. — By the extreme value theorem, $|f(z)|$ attains a maximum in $\overline{\Omega}$. But this cannot happen in Ω if f is non-constant. On the other hand, if f is constant, then the statement is trivial. ◻

***Warning* XII.3.5.8.** — The compactness (boundedness) of $\overline{\Omega}$ is essential. Let $\Omega = \{z \in \mathbb{C} \mid \operatorname{Re} z > 0, \operatorname{Im} z > 0\}$ be the open first quadrant and consider $f(z) = \mathrm{e}^{-\mathrm{i}z^2}$. Then f is entire and clearly continuous on $\overline{\Omega}$. Moreover, $|f(z)| = 1$ for all $z \in \partial\Omega$, which consists of the positive real and imaginary axes. However, f is unbounded on Ω, e. g., we have $f(z) = \mathrm{e}^{r^2}$ if $z = r\mathrm{e}^{\mathrm{i}\pi/4}$, which is the angle bisector between the two axes.

§ XII.4. CONFORMAL MAPS

XII.4.1. Conformality.

Definition* XII.4.1.1.** — *A holomorphic map $f\colon U \to V$ between open sets $U, V \subseteq \mathbb{C}$ is* ***locally bijective *on U if every $z \in U$ has an open neighbourhood $W \subseteq U$ such that $f\colon W \to f(W)$ is bijective. Instead of W, we could also only require an open disc $D \subseteq U$ centred at z such that $f\colon D \to f(D)$ is bijective.*

***Lemma* XII.4.1.2.** — *A holomorphic map $f\colon U \to V$ between open sets U and V is locally bijective on U if and only if $f'(z) \neq 0$ for all $z \in U$. Moreover, any local inverse of f is holomorphic.*

Proof. — Let $z_0 \in U$. If $f'(z_0) \neq 0$, then f is invertible on an open neighbourhood of z by the inverse function theorem on $\mathbb{R}^2$ (note that holomorphy implies continuous differentiability).

Conversely, let f be injective on an open neighbourhood of z_0; w. l. o. g., let U be this open neighbourhood. By way of contradiction, suppose that $f'(z_0) = 0$. Then $f(z) - f(z_0) = a(z-z_0)^k + G(z)$ near z_0 where $a \neq 0$, $k \ge 2$ and where G has a zero of order $k+1$ at z_0. For a small enough circle C centred at z_0 and for sufficiently small $w \neq 0$ (depending on C), we have $|G(z)| < |F(z)|$ for all $z \in C$ where $F(z) = a(z-z_0)^k - w$ because $G(z)/(z-z_0)^k \to 0$ as $z \to z_0$. Note that F has $k \ge 2$ distinct zeroes inside C, so $F(z) + G(z) = f(z) - f(z_0) - w$ has k zeroes inside C by Rouché's theorem XII.3.5.3, too. As its derivative $f'(z)$ has an isolated zero at z_0, since otherwise $f' = 0$ and f is not injective, by choosing an even smaller C, all roots of $f(z) - f(z_0) - w$ are simple. Hence, all $k \ge 2$ roots of $f(z) - f(z_0) - w$ are distinct, contradicting injectivity of f.

For the second part, w. l. o. g., let $f\colon U \to V$ be bijective and let $w_0 \in V$. For all $w \neq w_0$ close to w_0, we have

$$\frac{f^{-1}(w) - f^{-1}(w_0)}{w - w_0} = \left(\frac{w - w_0}{f^{-1}(w) - f^{-1}(w_0)}\right)^{-1} = \left(\frac{f(z) - f(z_0)}{z - z_0}\right)^{-1}$$

where $z = f^{-1}(w)$ and $z_0 = f^{-1}(w_0)$. As $z \to z_0$, we have $w \to w_0$ and so $(f^{-1})'(w_0) = 1/f'(z_0)$ exists (note that $f'(z_0) \neq 0$). ◻

XII.4.1.3. — For $z, w \in \mathbb{C} \setminus \{0\}$, the (oriented) angle $\alpha \in (-\pi, \pi]$ between z and w is the real angle between them when they are understood as vectors. If $\langle -, - \rangle$ is the Euclidean inner product on $\mathbb{R}^2$, given in complex form by $\langle z, w \rangle = \operatorname{Re}(z\overline{w})$, the angle α is uniquely determined by

$$\frac{\langle z, w \rangle}{|z||w|} = \cos \alpha \qquad \text{and} \qquad \frac{\langle z, -\mathrm{i}w \rangle}{|z||w|} = \sin \alpha.$$

By extension, let $\gamma, \eta \colon [a, b] \to \mathbb{C}$ be two smooth curves that intersect at z_0, say, $\gamma(t_0) = \eta(t_0) = z_0$ after reparametrisation. If $\gamma'(t_0)$ and $\eta'(t_0)$ is non-zero, then they represent the tangents of γ and η at z_0, resp., and the angle of intersection of γ and η at z_0 is the angle between $\gamma'(t_0)$ and $\eta'(t_0)$.

***Definition* XII.4.1.4.** — *A real-differentiable map $f \colon \Omega \to \mathbb{C}$ is **conformal** at a point $z_0 \in \Omega$ if for any two smooth curves γ and η intersecting at z_0, the angle between γ and η at z_0 equals the angle between $f \circ \gamma$ and $f \circ \eta$ at $f(z_0)$. In particular, we assume that the tangents of γ, η, $f \circ \gamma$ and $f \circ \eta$ at z_0 and $f(z_0)$ are non-zero.*

***Proposition* XII.4.1.5.** — *Let $f \colon \Omega \to \mathbb{C}$ be a complex-valued function on an open set Ω. If f is holomorphic at z_0 with $f'(z_0) \neq 0$, then f is conformal at z_0. Conversely, if f is real-differentiable at $z_0 \in \Omega$ with $\mathrm{J}_f(z_0) \neq 0$, then conformality at z_0 implies holomorphy at z_0 with $f'(z_0) \neq 0$.*

Proof. — Let γ and η be two smooth curves intersecting at $z_0 = \gamma(t_0) = \eta(t_0)$. Let f be holomorphic at z_0 with $f'(z_0) \neq 0$. Then

$$\langle f'(z_0)\gamma'(t_0), f'(z_0)\eta'(t_0) \rangle = |f'(z_0)|^2 \operatorname{Re}(\gamma'(t_0)\overline{\eta'(t_0)}) = |f'(z_0)|^2 \langle \gamma'(t_0), \eta'(t_0) \rangle,$$

so f preserves the cosine of the angle between γ and η at z_0. A similar argument shows the same for the sine, so f is conformal.

Conversely, let f be real-differentiable at z_0 with $\mathrm{J}_f(z_0) \neq 0$. If $\gamma'(t_0)$ and $\eta'(t_0)$ are orthogonal, i. e., $\langle \gamma'(t_0), \eta'(t_0) \rangle = 0$, then

$$0 = \cos \pm\pi = \langle \mathrm{J}_f(z_0)\gamma'(t_0), \mathrm{J}_f(z_0)\eta'(t_0) \rangle$$

by conformality. The says that $\mathrm{J}_f(z_0)$ preserves orthogonality, so $\mathrm{J}_f(z_0) = sO$ where $O \in \mathrm{Mat}_2(\mathbb{R})$ is an orthogonal matrix and where $s \in \mathbb{R} \setminus \{0\}$ is a scaling factor.[(XII.8)] If $\gamma'(t_0) = 1$ and $\eta'(t_0) = \mathrm{i}$, then $\cos \alpha = 0$ and $\sin \alpha = 1$ for the angle α between them. Then conformality yields (note that $-i(x, y) = (y, -x)$)

$$0 = \langle \mathrm{J}_f(z_0)\gamma'(t_0), \mathrm{J}_f(z_0)\eta'(t_0) \rangle, \qquad 0 < \langle \mathrm{J}_f(z_0)\gamma'(t_0), -\mathrm{i}\mathrm{J}_f(z_0)\eta'(t_0) \rangle = \det \mathrm{J}_f(z_0).$$

The second condition shows that $\det O = \det \mathrm{J}_f(z_0)/s^2$ is positive, so O is a rotation, not a reflection. Thus,

$$\mathrm{J}_f(z_0) = s \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

for some θ, and f satisfies the Cauchy-Riemann eqs. (XII.1.3.10.1) at z_0. Recall from theorem XII.1.3.13 that $|f'(z_0)|^2 = \det \mathrm{J}_f(z_0) > 0$, whence $f'(z_0) \neq 0$. ◻

Thus, it is only natural to study differentiable conformal maps $\mathbb{R}^2 \to \mathbb{R}^2$ by means of complex analysis.

***Definition* XII.4.1.6.** — *A bijective holomorphic map $f \colon U \to V$ between open sets is a **biholomorphism**. In this case, U and V are **biholomorphic**. This terminology is justified since, in particular, $f^{-1} \colon V \to U$ is again holomorphic by lemma XII.4.1.2.*

(XII.8) If $M \in \mathrm{Mat}_2(\mathbb{R})$ preserves orthogonality, then $z^\top M^\top M w = 0$ whenever $z^\top w = 0$ for all $z, w \in \mathbb{R}^2$. Note that $M^\top M$ is symmetric. Plugging in $z = \binom{1}{0}$ and $w = \binom{0}{1}$, we see that $M = \operatorname{diag}(s, t)$ is diagonal. Plugging in $z = \binom{1}{1}$ and $w = \binom{1}{-1}$, we obtain $s = t$. Thus, $M^\top M = s\mathrm{id}$, i. e., M is orthogonal up to scaling by s.

Corollary* XII.4.1.7** (characterisation of conformality). — *A holomorphic map $f\colon U \to V$ between open sets is (everywhere)* ***conformal *if and only if $f'(z) \neq 0$ for all $z \in U$. Equivalently, f is conformal if and only if it is locally a biholomorphism, i. e., for any $x \in U$ there exists an open neighbourhood W in U such that $f\colon W \to f(W)$ is a biholomorphism.*

Proof. — Combine lemma XII.4.1.2 and proposition XII.4.1.5.

***Convention* XII.4.1.8.** — *Henceforth, we will use the terms* biholomorphism *and* holomorphic conformal map *interchangeably, and we also say that two open sets are* conformally equivalent *if they are biholomorphic.*

***Warning* XII.4.1.9.** — Of course, for a holomorphic map being a biholomorphism is a much stronger property than being conformal: conformal maps are only locally bijective, but biholomorphisms are *globally* bijective. For instance, $f(z) = z^2$ on the punctured disc $\mathbb{C} \setminus \{0\}$ satisfies $f'(z) \neq 0$ everywhere, but f is cleary not injective. This convention is not universal in the literature, and the reason for our adoption is that we only consider conformal maps in the context of Riemann's mapping theorem. Insert reference.

Here a number of examples of conformal maps.

Definition* XII.4.1.10.** — *The complex* ***unit disc *and the complex* ***upper half-plane*** *are*

$$\mathcal{D} := \{z \in \mathbb{C} \mid |z| < 1\}, \qquad \mathcal{H} := \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}.$$

***Theorem* XII.4.1.11.** — *The unit disc and the upper half-plane are conformally equivalent, via the mutually inverse conformal maps*

$$F\colon \mathcal{H} \longrightarrow \mathcal{D}, \quad F(z) = \frac{\mathrm{i} - z}{\mathrm{i} + z} \qquad \text{and} \qquad G\colon \mathcal{D} \longrightarrow \mathcal{H}, \quad G(w) = \mathrm{i}\frac{1-w}{1+w}.$$

The map F is called the ***Cayley transform****.*

Proof. — Both maps are holomorphic on their respective domains. They are also well-defined. Any $z \in \mathcal{H}$ is closer to i than to $-\mathrm{i}$, so $|F(z)| < 1$ and F maps into $\mathcal{D}$. Next, for $w = u + \mathrm{i}v \in \mathcal{D}$ we have

$$\operatorname{Im} G(w) = \operatorname{Re} \frac{1-w}{1+w} = \operatorname{Re} \frac{(1-u-\mathrm{i}v)(1+u-\mathrm{i}v)}{(1+u)^2 + v^2} = \frac{1-u^2-v^2}{(1+u)^2+v^2} > 0$$

since $|w| = u^2 + v^2 < 1$, whence $G(w) \in \mathcal{H}$ and G maps into $\mathcal{H}$. Finally,

$$F(G(w)) = \frac{\mathrm{i} - \mathrm{i}\frac{1-w}{1+w}}{\mathrm{i} + \mathrm{i}\frac{1-w}{1+w}} = \frac{1+w-1+w}{1+w+1-w} = w,$$

and similarly $G(F(z)) = z$.

***Observation* XII.4.1.12.** — A general strategy in finding conformal maps is to study the boundary behaviour of these maps. We will exhibit this with theorem XII.4.1.11. The map F can be extended to $\mathbb{C} \setminus \{-\mathrm{i}\}$, so it is continuous on $\partial\mathcal{H} = \mathbb{R}$, the real line. Then $|F(x)| = 1$ for all $x \in \mathbb{R}$ since the distance of any real x to either i or to $-\mathrm{i}$ is the same, so F maps $\partial\mathcal{H}$ into the unit circle $\partial\mathcal{D}$. In fact, we can write

$$F(x) = \frac{\mathrm{i} - x}{\mathrm{i} + x} = \frac{1-x^2}{1+x^2} + \mathrm{i}\frac{2x}{1+x^2}$$

and parametrise $\mathbb{R}$ by $x = \tan t$ for $t \in (-\pi/2, \pi/2)$. Then

$$\sin 2t = \frac{2\tan t}{1+\tan^2 t}, \qquad \cos 2t = \frac{1-\tan^2 t}{1+\tan^2 t}$$

imply $F(\tan t) = \cos 2t + \mathrm{i}\sin 2t = \mathrm{e}^{\mathrm{i}2t}$. Therefore, the image $F(\mathbb{R})$ is $\partial\mathcal{D} \setminus \{-1\}$. Moreover, as x goes from $-\infty$ to ∞, the value $F(x)$ starts at -1 and travels counter-clockwise. The point -1 on the unit circle is the 'point at infinity' of the upper half-plane.

Example **XII.4.1.13** (translations, dilations and rotations)**.** — Let $h \in \mathbb{C}$. Then $z \mapsto z + h$ is a conformal map $\mathbb{C} \to \mathbb{C}$ with inverse $w \mapsto w - h$. If $h \in \mathbb{R}$, then these maps restrict to $\mathcal{H} \to \mathcal{H}$.

Let $c \in \mathbb{C} \setminus \{0\}$. Then $f\colon \mathbb{C} \to \mathbb{C}$, $z \mapsto cz$ is a conformal map with inverse $w \mapsto c^{-1}w$. If $|c| = 1$, say $c = \mathrm{e}^{\mathrm{i}\theta}$ for some $\theta \in \mathbb{R}$, then f is a rotation by θ. If $c > 0$, then f is a dilation by c. If $c < 0$, then f is a dilation by $|c|$ followed by a rotation by π.

Example **XII.4.1.14** (infinite sectors)**.** — Let $n \in \mathbb{N}$. Then $z \mapsto z^n$ is conformal map from the sector $S = \{z \in \mathbb{C} \mid 0 < \arg z < \pi/n\}$ to $\mathcal{H}$. The inverse map is $w \mapsto w^{1/n} = \mathrm{e}^{(\log w)/n}$ with the principal branch of the logarithm.

More generally, for $0 < \alpha < 2$ consider the map

$$f\colon \mathcal{H} \longrightarrow S = \{w \in \mathbb{C} : 0 < \arg w < \alpha\pi\}, \qquad f(z) = z^\alpha$$

Introduce the term branch cut. Also that it is possible to take branch cuts not containing 1.

from the upper half-plane to a sector. Indeed, if we take the branch cut $[0, \infty)$ for the logarithm, then for $z = r\mathrm{e}^{\mathrm{i}\theta} \in \mathcal{H}$ with $r > 0$ and $0 < \theta < \pi$ we have $f(z) = z^\alpha = r^\alpha \mathrm{e}^{\mathrm{i}\alpha\theta} \in S$. Its invers is $w \mapsto w^{1/\alpha}$, noting that S lies in the chosen branch of the logarithm. Therefore, if we compose f with translations and rotations from example XII.4.1.13, then any infinite sector in $\mathbb{C}$ is conformally equivalent to $\mathcal{H}$.

We now note the boundary behaviour of f. If x goes from $-\infty$ to 0, then $f(x)$ goes from $\infty\mathrm{e}^{\mathrm{i}\alpha\pi} = -\infty$ to 0. Similarly, if x goes from 0 to ∞, then $f(x)$ goes from 0 to ∞.

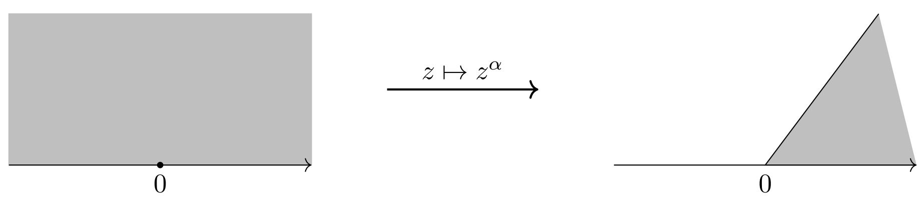

Figure XII.13: The conformal map $z \mapsto z^\alpha$.

Example **XII.4.1.15** (quadrants)**.** — Consider the map

$$f\colon \mathcal{D} \cap \mathcal{H} \longrightarrow Q = \{z \in \mathbb{C} \mid \operatorname{Re} z > 0, \operatorname{Im} z > 0\}, \qquad f(z) = \frac{1+z}{1-z}$$

from the upper half-disc to the first quadrant. This is indeed well-defined, since

$$f(z) = \frac{(1+z)(1-\overline{z})}{|1-z|^2} = \frac{1-|z|^2}{|1-z|^2} + \mathrm{i}\frac{2\operatorname{Im} z}{|1-z|^2} \in Q$$

for all $z \in \mathcal{D} \cap \mathcal{H}$. Moreover, the map $g(w) = (w-1)/(w+1)$ is holomorphic on Q. We have $|w+1| > |w-1|$ for all $w \in Q$ since w has a bigger distance to -1 than to 1, and a simple calculation shows that $\operatorname{Im} g(w) > 0$ for all $w \in Q$. Thus, $g\colon Q \to \mathcal{D} \cap \mathcal{H}$ is holomorphic, and a direct computation as in theorem XII.4.1.11 shows that it is the inverse of f. In other words, the upper half-disc is conformally equivalent to the first quadrant.

For the boundary behaviour of f, for $z = \mathrm{e}^{\mathrm{i}\theta} \in \partial\mathcal{D} \cap \mathcal{H}$ in the upper half-circle we have

$$f(z) = \frac{1+\mathrm{e}^{\mathrm{i}\theta}}{1-\mathrm{e}^{\mathrm{i}\theta}} = \frac{\mathrm{e}^{-\mathrm{i}\theta/2}+\mathrm{e}^{\mathrm{i}\theta/2}}{\mathrm{e}^{-\mathrm{i}\theta/2}-\mathrm{e}^{\mathrm{i}\theta/2}} = \frac{\mathrm{i}}{\tan(\theta/2)}.$$

So, as θ travels from 0 to π, the value $f(z)$ goes from $\mathrm{i}\infty$ to 0 along the imaginary axis. On the other hand, for $z = x \in (-1,1)$ real we have $f(z) = (1+x)/(1-x)$. In this case, as $f(x)$ goes from 0 to ∞ as x goes from -1 to 1. Note that $f(0) = 1$.

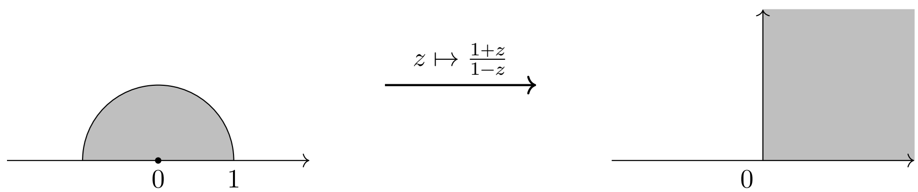

Figure XII.14: The conformal map $z \mapsto (1+z)/(1-z)$.

***Example* XII.4.1.16.** — Consider the map

$$\mathcal{H} \longrightarrow \{w \in \mathbb{C} \mid 0 < \operatorname{Im} w < \pi\}, \qquad z \longmapsto \log z,$$

defined by the branch cut $\mathrm{i}(-\infty,0]$, from the upper half-plane to a strip. Indeed, if $z = r\mathrm{e}^{\mathrm{i}\theta}$ with $-\pi/2 < \theta < 3\pi/2$, then $\log z = \log r + \mathrm{i}\theta$, and restricting to $0 < \theta < \pi$ gives the map. Its inverse is $w \mapsto \mathrm{e}^w$.

For the boundary behaviour, as x goes from $-\infty$ to 0, the value $\log x$ goes from $\infty + \mathrm{i}\pi$ to $-\infty + \mathrm{i}\pi$ on the top line of the strip. Furthermore, as x goes from 0 to ∞, then $\log x$ goes from $-\infty$ to ∞ along the real axis.

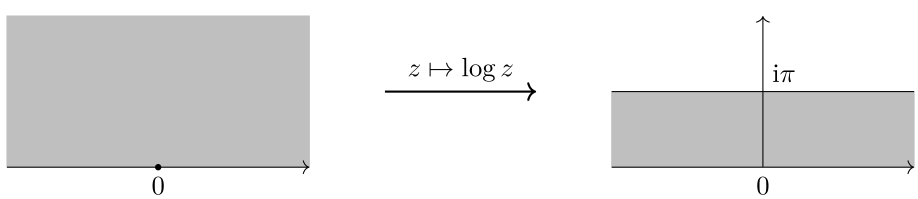

Figure XII.15: The conformal map $z \mapsto \log z$.

***Example* XII.4.1.17.** — Restricting $z \mapsto \log z = \log|z| + \mathrm{i}\arg z$ from example XII.4.1.16 to the upper half-disc $\mathcal{D} \cap \mathcal{H}$, we obtain a conformal map from the upper half-disc to a left half-strip:

$$\mathcal{D} \cap \mathcal{H} \longrightarrow T = \{w \in \mathbb{C} \mid \operatorname{Re} w < 0, 0 < \operatorname{Im} w < \pi\}.$$

If z starts at 0, goes to 1 on the real line, then to -1 along the upper half-circle, and finally back to 0, then $\log z$ goes from $-\infty$ to $-\infty + \mathrm{i}\pi$ by passing through 0 and then through $\mathrm{i}\pi$.

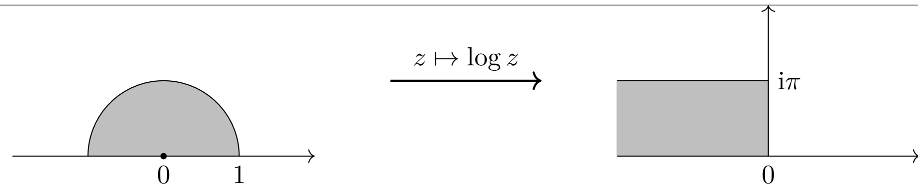

Figure XII.16: The conformal map $z \mapsto \log z$ restricted to the upper half-disc.

***Example* XII.4.1.18.** — The map

$$\{z \in \mathbb{C} \mid -\tfrac{\pi}{2} < \operatorname{Re} z < \tfrac{\pi}{2}, \operatorname{Im} z > 0\} \longrightarrow \{w \in \mathbb{C} \mid |w| < 1, \operatorname{Re} w > 0\}, \qquad z \longmapsto \mathrm{e}^{\mathrm{i}z}$$

from a half-strip to a half-circle is conformal since $\mathrm{e}^{\mathrm{i}z} = \mathrm{e}^{-\operatorname{Im} z}\mathrm{e}^{\mathrm{i}\operatorname{Re} z}$. As z goes from $\pi/2 + \mathrm{i}\infty$ to $-\pi/2 + \mathrm{i}\infty$ by passing through $\pi/2$ and $-\pi/2$, the value $\mathrm{e}^{\mathrm{i}z}$ goes from 0 to i, then to $-\mathrm{i}$ along the left half-circle, and finally back to 0. This map is closely related to the inverse map of the map in example XII.4.1.17.

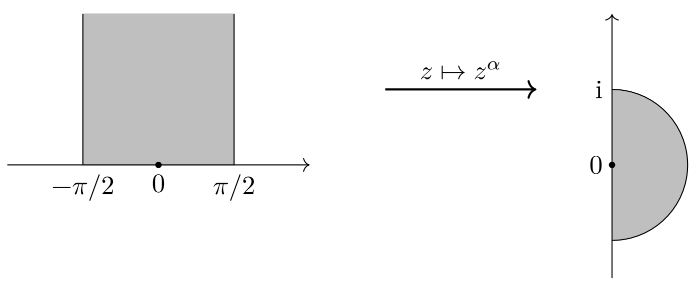

Figure XII.17: The conformal map $z \mapsto \mathrm{e}^{\mathrm{i}z}$.

***Example* XII.4.1.19.** — Consider the map

$$f\colon \mathcal{D} \cap \mathcal{H} \longrightarrow \mathcal{H}, \qquad z \longmapsto -\frac{1}{2}\left(z + \frac{1}{z}\right)$$

from the upper half-disc to the upper half-plane. This is well-defined: if $z = r\mathrm{e}^{\mathrm{i}\theta} \in \mathcal{D} \cap \mathcal{H}$ where $r < 1$ and $0 < \theta < \pi$, then

$$\operatorname{Im}\left(z + \frac{1}{z}\right) = r \sin\theta + \frac{1}{r}\sin(-\theta) = \left(r - \frac{1}{r}\right)\sin\theta < 0.$$

To show bijectivity of f, observe that solving $f(z) = w$ for z means solving the quadratic equation $z^2 + 2wz + 1 = 0$ where $w \in \mathcal{H}$. If z is a solution, then $1/z$ is as well, and these are distinct if and only if $w \neq \pm 1$. This is the case for $w \in \mathcal{H}$, so f is injective, and since the complex numbers are algebraically closed, f is surjective.

Let us examine the behaviour of f at the boundary. If $x \in \mathbb{R}$ goes from 0 to 1, then $f(x)$ goes from ∞ to 1. If $z = \mathrm{e}^{\mathrm{i}\theta}$ goes from 1 to -1, then $f(z) = \cos\theta$ goes from -1 to 1. Finally, if $x \in \mathbb{R}$ goes from -1 to 0, then $f(x)$ goes from -1 to $-\infty$ on the real axis.

***Example* XII.4.1.20.** — The sine

$$\sin z = \frac{\mathrm{e}^{\mathrm{i}z} - \mathrm{e}^{-\mathrm{i}z}}{2\mathrm{i}} = -\frac{1}{2}\left(\mathrm{i}\mathrm{e}^{\mathrm{i}z} + \frac{1}{\mathrm{i}\mathrm{e}^{\mathrm{i}z}}\right)$$

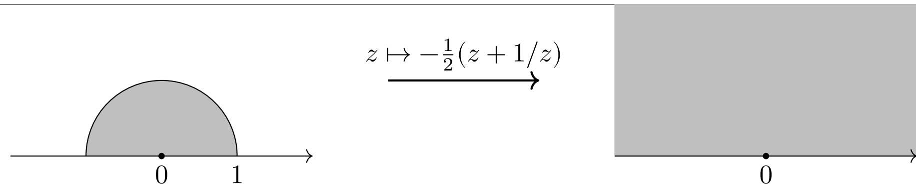

Figure XII.18: The conformal map $z \mapsto -\frac{1}{2}(z+1/z)$.

can be written as the composition of the map in example XII.4.1.18, followed by multiplication with i (rotation by $\pi/2$), followed by the map in example XII.4.1.19. Thus, we have a conformal map

$$\{z \in \mathbb{C} \mid -\tfrac{\pi}{2} < \operatorname{Re} z < \tfrac{\pi}{2}, \operatorname{Im} z > 0\} \longrightarrow \mathcal{H}, \qquad z \longmapsto \sin z$$

from the upper half-strip to the upper half-plane.

For the boundary behaviour, as z goes from $-\pi/2 + \mathrm{i}\infty$ over $-\pi/2$ and $\pi/2$ to $\pi/2 + \mathrm{i}\infty$, the value $f(z)$ goes from $-\infty$ over -1 and 1 to ∞.

XII.4.2. Möbius transformations.

There is a large class of relatively simple conformal maps called *Möbius transformation.* The most natural way to study them is, however, to extend them to the *Riemann sphere* $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$.

Definition* XII.4.2.1.** — *An* ***automorphism *of the Riemann sphere $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ is a bijective meromorphic map $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$. The set of all automorphisms of $\hat{\mathbb{C}}$ is denoted by $\operatorname{Aut}\hat{\mathbb{C}}$, which is a group under composition of maps. For $f \in \operatorname{Aut}\hat{\mathbb{C}}$, the inverse f^{-1} is again meromorphic (it has exactly one pole at $f(\infty)$).*

Definition* XII.4.2.2.** — *A* ***fractional linear transformation *or a* ***Möbius transformation*** *is a map*

$$f\colon \hat{\mathbb{C}} \longrightarrow \hat{\mathbb{C}}, \qquad z \longmapsto \frac{az+b}{cz+d}$$

for $a, b, c, d \in \mathbb{C}$ such that $ad - bc \neq 0$. The behaviour at infinity is as follows: $f(\infty) = a/c$ and $f(-d/c) = \infty$.

***Proposition* XII.4.2.3.** — *The Möbius transformations form a group under composition where the inverse of $(az+b)/(cz+d)$ is $(dz-b)/(a-cz)$. More precisely, there is a surjective group homomorphism*

$$\varGamma\colon\ \mathrm{GL}_2(\mathbb{C}) \longrightarrow\!\!\!\!\!\rightarrow \{\textit{Möbius transformations}\}, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \longmapsto \frac{az+b}{cz+d}$$

with kernel $\mathbb{C}^\times$. In particular, the Möbius transformations are isomorphic to the projective general linear group $\mathrm{PGL}_2(\mathbb{C}) \coloneqq \mathrm{GL}_2(\mathbb{C})/\mathbb{C}^\times$.

Proof. — The composition of two Möbius transformations is again a Möbius transformation:

$$\frac{a'\frac{az+b}{cz+d} + b'}{c'\frac{az+b}{cz+d} + d'} = \frac{a'az + a'b + b'cz + b'd}{c'az + c'b + d'cz + d'd} = \frac{(a'a + b'c)z + (a'b + b'd)}{(c'a + d'c)z + (c'b + d'd)}.$$

The coefficients in the new Möbius transformations take the form of the entries of a matrix product. Hence, $\varGamma$ is a surjective homomorphism from a group to a magma, so the magma of Möbius transformations is a group. The inverse of a Möbius transformation $(az+b)/(cz+d)$ is given by

$$\varGamma\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \varGamma\left(\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\right) = \frac{dz-b}{a-cz}.$$

For the kernel, suppose that $(az+b)/(cz+d) = z$. Taking $z = 0$ implies $b = 0$, and taking $z = \infty$ implies $c = 0$. Finally, taking $z = 1$ gives $a = d$, and we obtain $\ker \Gamma = \mathbb{C}^\times$. ◻

***Remark* XII.4.2.4.** — The condition $ad-bc \neq 0$ on Möbius transformations $f(z) = (az+b)/(cz+d)$ holds if and only if f is non-constant. Indeed, if $b/d = f(0) \neq f(\infty) = a/c$, then $ad - bc \neq 0$.

***Lemma* XII.4.2.5.** — *Let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function on open set Ω. If f is injective, then the zero is simple, if it exists.*

Proof. — Let $z_0 \in \Omega$ be the zero, and write $f(z) = (z-z_0)^n g(z)$ near z_0 with g holomorphic near z_0 and $g(z_0) \neq 0$. Then g is non-vanishing near z_0, so there is a function h such that $h(z)^n = g(z)$ (proposition XII.2.3.7). Then the derivative of $(z-z_0)h(z)$ at z_0 is $h(z_0) \neq 0$, so $(z-z_0)h(z)$ is locally bijective by lemma XII.4.1.2. Let us say that it restricts to a bijective holomorphic map $U \to D$ where U is a neighbourhood of z_0 and where D is a disc centred at 0. For $w \in D \setminus \{0\}$, there are distinct $z_1, \dots, z_n \in U$ with $(z_k - z_0)h(z_k) = \mathrm{e}^{2\pi \mathrm{i} k/n} w$. However, $f(z_k) = (\mathrm{e}^{2\pi \mathrm{i} k/n} w)^n = w^n$ for all k. Since f is injective, we must have $n = 1$. ◻

***Theorem* XII.4.2.6.** — *The automorphisms of the Riemann sphere $\hat{\mathbb{C}}$ are precisely the Möbius transformations*

$$f(z) = \frac{az+b}{cz+d}$$

where $a, b, c, d \in \mathbb{C}$ with $ad - bc \neq 0$.

Proof. — Möbius transformations are invertible and hence automorphisms of $\hat{\mathbb{C}}$. Conversely, let $f \in \operatorname{Aut}\hat{\mathbb{C}}$, which is a rational function by theorem XII.3.3.10. Since f is bijective, the partial fraction decomposition XII.3.3.11 says that f is of the form $f(z) = c(z-\alpha)^n/(z-\beta)^k$ for some $c, \alpha, \beta \in \mathbb{C}$ with $\alpha \neq \beta$ and $c \neq 0$. For any $z_0 \in \mathbb{C} \setminus \{\beta\}$, the injective function $f(z) - f(z_0)$ has a zero at z_0 of order n, so $n = 1$ by lemma XII.4.2.5. A similar argument applied to $1/f$ shows that $k = 1$. Thus, $f(z) = c(z-\alpha)/(z-\beta)$ is a Möbius transformation. Note that $c\beta - c\alpha \neq 0$. ◻

***Remark* XII.4.2.7.** — In particular, $\mathrm{PGL}_2(\mathbb{C}) \xrightarrow{\sim} \operatorname{Aut}\hat{\mathbb{C}}$ is the projective general linear group. Another point of view is to identify the projective line $\mathbb{P}^1_{\mathbb{C}} := \mathbb{C}^2/\mathbb{C}^\times$ with $\hat{\mathbb{C}}$ via $(z_1, z_2) \mapsto z_1/z_2$. Then the automorphisms of $\hat{\mathbb{C}}$ acting on $\hat{\mathbb{C}}$ is the same as the $\mathrm{PGL}_2(\mathbb{C})$-action on $\mathbb{P}^1_{\mathbb{C}}$.

***Fact* XII.4.2.8.** — *Given three different points $a, b, c \in \hat{\mathbb{C}}$, there is exactly one Möbius transformation f with $f(a) = 0$, $f(b) = 1$ and $f(c) = \infty$. In particular, for any three pairs of points (z_i, w_i) for $i = 1, 2, 3$, there is exactly one Möbius transformation f with $f(z_i) = w_i$.*

Proof. — Consider the **cross ratio**

$$f(z) = \frac{z-a}{z-c} \Big/ \frac{b-a}{b-c}$$

of a, b, c. Then $f(a) = 0$, $f(b) = 1$ and $f(c) = \infty$. Since f is non-constant, it satisfies the invertibility condition on its coefficients.

To show uniqueness, let f_1 and f_2 be two Möbius transformations mapping (a, b, c) to $(0, 1, \infty)$. Then $f_2^{-1} \circ f_1(z) = (pz+q)/(rz+s)$ fixes $0, 1, \infty$. This implies $q/s = 0$, $(p+q)/(r+s) = 1$ and $p/r = \infty$, whence $q = r = 0$ and $p = s$. Thus, $f_2^{-1} \circ f_1(z) = z$ is the identity. ◻

***Construction* XII.4.2.9.** — Visually, circles and straight lines in $\mathbb{C}$ can be generalised to a single object in the Riemann sphere $\hat{\mathbb{C}}$ since lines are 'circles through ∞'. The equation for a circle is $|z - z_0| = r$ with $z_0 \in \mathbb{C}$ and $r > 0$, i.e.,

$$(z-z_0)\overline{z-z_0} - r^2 = |z|^2 - z_0\overline{z} - \overline{z_0}z + (|z_0|^2 - r^2) = 0.$$

The equation for a line is $z_0 z + \overline{z_0 z} = d$ with $z_0 \in \mathbb{C}$ and $d \in \mathbb{R}$ (this is the vertical line through d, then rotated by $\arg z_0$). Both equations are of the form

$$\alpha|z|^2 + \beta z + \overline{\beta z} + \gamma = 0 \tag{XII.4.2.9.1}$$

where $\alpha, \gamma \in \mathbb{R}$ and $\beta \in \mathbb{C}$ satisfy $|\beta|^2 > \alpha\gamma$ (in the circle case, we have $r^2 = (|\beta|^2 - \alpha\gamma)/\alpha^2 > 0$; in the line case, we have $z_0 = \beta \neq 0$). This is a circle if $\alpha \neq 0$ and a straight line if $\alpha = 0$. The set of points satisfying equation (XII.4.2.9.1) is called a **generalised circle**.

***Fact* XII.4.2.10.** — *Möbius transformations preserve generalised circles.*

Proof. — Any Möbius transformation can be written as a composition of translations, rotations, dilations and inversions $z \mapsto 1/z$:

$$z \longmapsto cz + d \longmapsto \frac{1}{cz+d} \longmapsto \frac{a}{c} + \frac{bc - ad}{c} \cdot \frac{1}{cz+d} = \frac{az+b}{cz+d}.$$

Clearly, translations and rotations/dilations preserve the equation of generalised circles. As for inversions, if $w = 1/z$ is the image of the inversion, then we have

$$0 = (\alpha/|w|^2 + \beta/w + \overline{\beta/w} + \gamma)|w|^2 = \alpha + \beta\overline{w} + \overline{\beta}w + \gamma|w|^2$$

with $|\beta|^2 > \alpha\gamma$ still, so the inversion of a generalised circle is a generalised circle. ◻

Alternative proof. — Equation (XII.4.2.9.1) is equivalent to

$$\begin{pmatrix} \overline{z} & 1 \end{pmatrix} \begin{pmatrix} \alpha & \overline{\beta} \\ \beta & \gamma \end{pmatrix} \begin{pmatrix} z \\ 1 \end{pmatrix} = 0$$

where $\alpha, \gamma \in \mathbb{R}$ and $\beta \in \mathbb{C}$ such that $|\beta|^2 > \alpha\gamma$. This is equivalent to saying that the matrix C above is hermitian ($C^{\mathrm{H}} = C$) with $\det C < 0$. Since a Möbius transformation can be written as a matrix product (see remark XII.4.2.7), the preimages z of the generalised circle under the Möbius transformation A satisfy

$$\begin{pmatrix} \overline{z} & 1 \end{pmatrix} A^{\mathrm{H}} C A \begin{pmatrix} z \\ 1 \end{pmatrix} = 0.$$

We observe that $A^{\mathrm{H}}CA$ is hermitian and that $\det(A^{\mathrm{H}}CA) = |\det A|^2 \det C < 0$. Thus, Möbius transformations preserve generalised circles. ◻

XII.4.3. The Schwarz lemma.

***Schwarz lemma* XII.4.3.1.** — *Let $f \colon \mathcal{D} \to \mathcal{D}$ be holomorphic with $f(0) = 0$. Then the following hold:*

(i) *$|f(z)| \le |z|$ for all $z \in \mathcal{D}$;*
(ii) *if $|f(z_0)| = |z_0|$ for some $z_0 \neq 0$, then f is a rotation, i. e., $f(z) = cz$ with $|c| = 1$;*
(iii) *$|f'(0)| \le 1$, and if equality holds then f is a rotation.*

Proof. — Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be the power series expansion on $\mathcal{D}$ centred at 0. Since $f(0) = 0$, we have $a_0 = 0$, and so $f(z)/z$ has a removable singularity at 0. Let $r < 1$. If $|z| = r$, then $|f(z)/z| \le 1/r$, and the maximum modulus principle (corollary XII.3.5.7) says that this is true for all $|z| \le r$. Letting $r \to 1$ yields $|f(z)/z| \le 1$ on $\mathcal{D}$, which is (i).

For (ii), $|f(z_0)| = |z_0|$ states that $f(z)/z$ attains its maximum on $\mathcal{D}$. By the maximum modulus principle XII.3.5.6, it is constant, say, $f(z) = cz$. Hence, $|z_0| = |f(z_0)| = |cz_0|$ implies $|c| = 1$, and f is a rotation.

Finally for (iii), let $g(z) = f(z)/z$, defined on all of $\mathcal{D}$. Then

$$|f'(0)| = \lim_{z\to 0}\left|\frac{f(z) - f(0)}{z}\right| = |g(0)| \leq 1.$$

If we have equality, then g attains its maximum at 0, so it is constant by the maximum modulus principle XII.3.5.6, say, $f(z) = cz$ with $1 = |f'(0)| = |c|$.

***Definition* XII.4.3.2.** — *An **automorphism** of an open set Ω is a conformal map $\Omega \to \Omega$. The set of all automorphisms of Ω is denoted by $\operatorname{Aut}\Omega$. We endow it with a group structure by virtue of composition of maps. Then the identity is id_Ω, and the inverse of $f \in \operatorname{Aut}\Omega$ is f^{-1}.*

The following maps have many applications in complex analysis.

***Proposition* XII.4.3.3.** — *For a fixed $\alpha \in \mathcal{D}$, the Möbius transformation*

$$\psi_\alpha(z) = \frac{\alpha - z}{1 - \overline{\alpha} z}$$

*is an automorphism on the unit disc $\mathcal{D}$ and restricts to the unit circle $\partial\mathcal{D} \to \partial\mathcal{D}$. Moreover, ψ_α interchanges 0 and α. This transformation ψ_α is called the **Blaschke factor** of α.*

Proof. — Since $|\alpha| < 1$, we see that ψ_α is holomorphic on $\mathcal{D}$. If $|z| = 1$, then $1/z = \overline{z}$ and

$$\psi_\alpha(z) = \frac{\alpha - z}{z(\overline{z - \alpha})},$$

whence $|\psi_\alpha(z)| = 1$ and ψ_α restricts to $\partial\mathcal{D} \to \partial\mathcal{D}$. By maximum modulus principle XII.3.5.6 and corollary XII.3.5.7, we have $|\psi_\alpha(z)| < 1$ for all $z \in \mathcal{D}$, i. e., ψ_α gives a map $\mathcal{D} \to \mathcal{D}$.

Finally, we observe that ψ_α is an involution:

$$(\psi_\alpha \circ \psi_\alpha)(z) = \frac{\alpha - \frac{\alpha - z}{1 - \overline{\alpha} z}}{1 - \overline{\alpha}\frac{\alpha - z}{1 - \overline{\alpha} z}} = \frac{\alpha - |\alpha|^2 z - \alpha + z}{1 - \overline{\alpha} z - |\alpha|^2 + \overline{\alpha} z} = \frac{(1 - |\alpha|^2)z}{1 - |\alpha|^2} = z.$$

Thus, ψ_α is its own inverse and interchanges 0 and α.

Alternative proof. — Here a more direct way to check that ψ_α maps into $\overline{\mathcal{D}}$ and restricts to $\partial\mathcal{D}$. We first show that for $\overline{\alpha} z \neq 1$, we have

$$\left|\frac{\alpha - z}{1 - \overline{\alpha} z}\right| \leq 1$$

if $|z| \leq 1$ and $|\alpha| \leq 1$, with equality if and only if $|z| = 1$ or $|\alpha| = 1$. To prove this claim, expanding the products shows that the inequality

$$|\alpha - z| = (\alpha - z)\overline{\alpha - z} \leq (1 - \overline{\alpha} z)\overline{1 - \overline{\alpha} z} = |1 - \overline{\alpha} z|$$

holds if and only if $0 \leq (1 - |z|^2)(1 - |\alpha|^2)$, with equality if and only if $|z| = 1$ or $|\alpha| = 1$.

***Example* XII.4.3.4.** — Let $\theta \in \mathbb{R}$ be an angle. Then the rotation $r_\theta\colon z \mapsto \mathrm{e}^{\mathrm{i}\theta} z$ by θ is an automorphism of $\mathcal{D}$ whose inverse is the rotation $r_{-\theta}\colon z \mapsto \mathrm{e}^{\mathrm{i}\theta} z$ by $-\theta$.

***Schwarz-Pick lemma* XII.4.3.5.** — *If $f\colon \mathcal{D} \to \mathcal{D}$ is a holomorphic function on the unit disc, then*

$$\frac{|f'(z)|}{1 - |f(z)|^2} \leq \frac{1}{1 - |z|^2}$$

for all $z \in \mathcal{D}$, with equality if and only if f is an automorphism of $\mathcal{D}$.

Proof. — Firstly, we define the **pseudo-hyperbolic distance**

$$\rho(z,w) := \left|\frac{z-w}{1-\overline{w}z}\right| = |\psi_w(z)|$$

for $z,w\in\mathcal{D}$. We claim that $\rho(f(z),f(w))\leq\rho(z,w)$ with equality if and only if f is an automorphism of $\mathcal{D}$. Indeed, let $g=\psi_{f(w)}\circ f\circ\psi_w^{-1}$ where ψ_α is the Blaschke factor interchanging 0 and α. Since $g(0)=\psi_{f(w)}\circ f(w)=0$, the Schwarz lemma XII.4.3.1 says that $|g(z)|\leq|z|$. Plugging in $\psi_w(z)$ for z gives the inequality.

For the equality case, suppose that f is an automorphism. Then g^{-1} is a holomorphic map $\mathcal{D}\to\mathcal{D}$ with $g^{-1}(0)=0$, so applying the Schwarz lemma XII.4.3.1 again to g^{-1} gives $|g^{-1}(w)|\leq|w|$. Plugging in $w=g(z)$ yields $|g(z)|=|z|$. Conversely, suppose that $|g(z)|=|z|$. Then by the Schwarz lemma XII.4.3.1, g is a rotation and in particular an automorphism of $\mathcal{D}$. Thus, $f=\psi_{f(w)}^{-1}\circ g\circ\psi_w$ is an automorphism of $\mathcal{D}$.

So, the claim gives

$$\left|\frac{f(z)-f(w)}{z-w}\right|\frac{1}{1-\overline{f(w)}f(z)}\leq\frac{1}{1-\overline{w}z}.$$

Letting $w\to z$ gives the statement. ∎

***Theorem* XII.4.3.6.** — *The automorphisms of the unit disc $\mathcal{D}$ are precisely the products of a rotation and a Blaschke factor*

$$f(z)=\mathrm{e}^{\mathrm{i}\theta}\frac{\alpha-z}{1-\overline{\alpha}z}$$

where $\theta\in\mathbb{R}$ and $\alpha\in\mathcal{D}$.

Proof. — Let $\alpha\in\mathcal{D}$ be the unique element such that $f(\alpha)=0$, and consider $g=f\circ\psi_\alpha$. Then $g(0)=0$, and $|g(z)|\leq|z|$ for all $z\in\mathcal{D}$ by the Schwarz lemma XII.4.3.1. Similarly, $g^{-1}(0)=0$, and $|g^{-1}(w)|\leq|w|$ for all $w\in\mathcal{D}$ by the Schwarz lemma XII.4.3.1. Taking $w=g(z)$, we arrive at $|g(z)|=|z|$ for all $z\in\mathcal{D}$. Once again by the Schwarz lemma XII.4.3.1, $g(z)=\mathrm{e}^{\mathrm{i}\theta}z$ is a rotation or some $\theta\in\mathbb{R}$. Taking $z=\psi_\alpha(w)$ and exploiting that ψ_α is an involution, we obtain $f(w)=\mathrm{e}^{\mathrm{i}\theta}\psi_\alpha(w)$, as desired. ∎

***Corollary* XII.4.3.7.** — *The only automorphisms of the unit disc that fix the origin are rotations.*

Proof. — If $f(z)=\mathrm{e}^{\mathrm{i}\theta}(\alpha-z)/(1-\overline{\alpha}z)$, then $f(0)=0$ implies that $\alpha=0$, whence $f(z)=\mathrm{e}^{\mathrm{i}\theta}z$. ∎

***Remark* XII.4.3.8.** — The group $\operatorname{Aut}\mathcal{D}$ acts *transitively* on $\mathcal{D}$: given $\alpha,\beta\in\mathcal{D}$, the map $\psi_\beta\circ\psi_\alpha\in\operatorname{Aut}\mathcal{D}$ will take α to β.

***Fact* XII.4.3.9.** — *The special unitary group*

$$\mathrm{SU}_{1,1}=\{M\in\mathrm{SL}_2(\mathbb{C})\mid\langle Mz,Mw\rangle=\langle z,w\rangle \textit{ for all } z,w\in\mathbb{C}^2\},$$

where $\langle z,w\rangle:=w^{\mathrm{H}}Jz$ with $J=\operatorname{diag}(1,-1)$ is the hermitian form with signature $(1,-1)$, is

$$\mathrm{SU}_{1,1}=\left\{\begin{pmatrix}a&b\\\overline{b}&\overline{a}\end{pmatrix}\in\mathrm{Mat}_2(\mathbb{C})\;\middle|\;|a|^2-|b|^2=1\right\}.$$

Proof. — Let $M=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\mathrm{SU}_{1,1}$, so that $\det M=ad-bc=1$ and $w^{\mathrm{H}}M^{\mathrm{H}}JMz=w^{\mathrm{H}}Jz$. By plugging in $\left(\begin{smallmatrix}1&0\end{smallmatrix}\right)^\top$ and $\left(\begin{smallmatrix}0&1\end{smallmatrix}\right)^\top$ for z and w, we obtain

$$M^{\mathrm{H}}JM=\begin{pmatrix}\overline{a}&\overline{c}\\\overline{b}&\overline{d}\end{pmatrix}\begin{pmatrix}a&b\\-c&-d\end{pmatrix}=\begin{pmatrix}\overline{a}a-\overline{c}c&\overline{a}b-\overline{c}d\\\overline{b}a-\overline{d}c&\overline{b}b-\overline{d}d\end{pmatrix}=\begin{pmatrix}1&\\&-1\end{pmatrix}=J,$$

i. e., $|a|^2 - |c|^2 = |d|^2 - |b|^2 = 1$ and $\overline{a}b = \overline{c}d$. Then multiplying $\det M = 1$ with $\overline{c}$ gives $\overline{c} = a\overline{c}d - b\overline{c}c = |a|^2 b - b|c|^2 = b$, and plugging this in $\det M = 1$ gives $ad = 1 + |b|^2 = |d|^2$. If $d = 0$, then $a = 0$. If $d \neq 0$, then $a = \overline{d}$. In both cases, we obtain

$$M = \begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix}$$

with $|a|^2 - |b|^2 = 1$, and these matrices M constitute all of $\mathrm{SU}_{1,1}$.

***Remark* XII.4.3.10.** — In fact, $\operatorname{Aut} \mathcal{D}$ is almost isomorphic to the group $\mathrm{SU}_{1,1}$. Since Blaschke factors are Möbius transformations, the group homomorphism $\Gamma\colon \mathrm{GL}_2(\mathbb{C}) \to \operatorname{Aut} \hat{\mathbb{C}}$ from proposition XII.4.2.3 restricts to

$$\mathrm{SU}_{1,1} \longrightarrow \operatorname{Aut} \mathcal{D}, \qquad \begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix} \longmapsto \frac{az + b}{\overline{b}z + \overline{a}} = \frac{a}{\overline{a}} \cdot \frac{z + \alpha}{\overline{\alpha}z + 1}$$

where $\alpha = b/a$ and $|a/\overline{a}| = 1$. Conversely, any element of $\operatorname{Aut} \mathcal{D}$ has the association

$$\mathrm{e}^{2\mathrm{i}\theta} \frac{z - \alpha}{1 - \overline{\alpha}z} = \frac{\mathrm{e}^{\mathrm{i}\theta} z - \alpha \mathrm{e}^{\mathrm{i}\theta}}{-\overline{\alpha}\mathrm{e}^{-\mathrm{i}\theta} z + \mathrm{e}^{-\mathrm{i}\theta}} \longmapsto \frac{1}{\sqrt{1 - |\alpha|^2}} \begin{pmatrix} \mathrm{e}^{\mathrm{i}\theta} & -\alpha \mathrm{e}^{\mathrm{i}\theta} \\ -\overline{\alpha}\mathrm{e}^{-\mathrm{i}\theta} & \mathrm{e}^{-\mathrm{i}\theta} \end{pmatrix} \in \mathrm{SU}_{1,1},$$

which is the inverse up to scaling. Hence, we obtain a surjective group homomorphism $\mathrm{SU}_{1,1} \twoheadrightarrow \operatorname{Aut} \mathcal{D}$. For its kernel, we have $\mathbb{C}^\times \cap \mathrm{SU}_{1,1} = \{\pm 1\}$. Thus, we have a group isomorphism $\mathrm{PSU}_{1,1} := \mathrm{SU}_{1,1} / \{\pm 1\} \xrightarrow{\sim} \operatorname{Aut} \mathcal{D}$, which is the *projective special unitary group*.

***Remark* XII.4.3.11.** — Recall from theorem XII.4.1.11 that the Cayley transform $F\colon \mathcal{H} \to \mathcal{D}$ is conformal. Hence, we obtain a group isomorphism

$$\operatorname{Aut} \mathcal{D} \xrightarrow{\sim} \operatorname{Aut} \mathcal{H}, \qquad \phi \longmapsto F^{-1} \circ \phi \circ F$$

by the conjugation action of F.

***Theorem* XII.4.3.12.** — *The automorphisms of the upper half-plane $\mathcal{H}$ are precisely of the Möbius transformations induced by $\mathrm{SL}_2(\mathbb{R})$, i. e.,*

$$f(z) = \frac{az + b}{cz + d}$$

where $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$.

Proof. — First of all, the group homomorphism $\Gamma\colon \mathrm{GL}_2(\mathbb{C}) \to \operatorname{Aut} \hat{\mathbb{C}}$ from proposition XII.4.2.3 restricts to $\Gamma\colon \mathrm{SL}_2(\mathbb{R}) \to \operatorname{Aut} \mathcal{H}$. Indeed, for $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ we have

$$\operatorname{Im} \Gamma_M(z) = \frac{(ad - bc) \operatorname{Im} z}{|cz + d|^2} = \frac{\operatorname{Im} z}{|cz + d|^2} > 0$$

if $z \in \mathcal{H}$.

Next, we show that $\mathrm{SL}_2(\mathbb{R})$ acts transitively on $\mathcal{H}$. Given $z \in \mathcal{H}$, choose $c \in \mathbb{R}$ such that $(\operatorname{Im} z)/|cz|^2 = 1$. Let $M_1 = \left(\begin{smallmatrix} & -1/c \\ c & \end{smallmatrix}\right)$ so that $\operatorname{Im} \Gamma_{M_1}(z) = (\operatorname{Im} z)/|cz|^2 = 1$, meaning $\Gamma_{M_1}(z)$ lies on the horizontal line through i. Now, let $M_2 = \left(\begin{smallmatrix} 1 & b \\ & 1 \end{smallmatrix}\right)$ be the translation by $b = -\operatorname{Re} \Gamma_{M_1}(z)$ so that $\Gamma_{M_2}(\Gamma_{M_1}(z)) = \mathrm{i}$. Thus, $\Gamma_{M_2 M_1}$ takes z to i.

Finally, let $f \in \operatorname{Aut} \mathcal{H}$ with $f(\beta) = \mathrm{i}$. By transitivity of $\mathrm{SL}_2(\mathbb{R})$, there is $N \in \mathrm{SL}_2(\mathbb{R})$ such that $\Gamma_N(\mathrm{i}) = \beta$, and let $g = f \circ \Gamma_N$ with $g(\mathrm{i}) = \mathrm{i}$. Let $F\colon \mathcal{H} \to \mathcal{D}$ be the Cayley transform, and note that $F(0) = \mathrm{i}$. Then $F \circ g \circ F^{-1}$ is an automorphism of $\mathcal{D}$ fixing the origin, so the Schwarz lemma XII.4.3.1 says that it is a rotation, say, by the angle -2θ. If

$$M_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}),$$

which is a rotation matrix, then a direct computation shows that $F \circ \mathit{\Gamma}_{M_\theta} = \mathrm{e}^{-2\mathrm{i}\theta} F$. Hence, $F \circ g \circ F^{-1} = F \circ \mathit{\Gamma}_{M_\theta} \circ F^{-1}$ and $f \circ \mathit{\Gamma}_N = g = \mathit{\Gamma}_{M_\theta}$, and thus, $f = \mathit{\Gamma}_{M_\theta} \circ \mathit{\Gamma}_N^{-1} = \mathit{\Gamma}_{M_\theta N^{-1}}$. □

Remark **XII.4.3.13.** — Similar to remark XII.4.3.10, $\mathit{\Gamma}\colon \mathrm{GL}_2(\mathbb{C}) \to \operatorname{Aut}\hat{\mathbb{C}}$ induces a surjective group homomorphism $\mathrm{SL}_2(\mathbb{R}) \twoheadrightarrow \operatorname{Aut}\mathcal{H}$ with kernel $\mathbb{C}^\times \cap \mathrm{SL}_2(\mathbb{R}) = \{\pm 1\}$. Thus, $\mathrm{PSL}_2(\mathbb{R}) := \mathrm{SL}_2(\mathbb{R})/\{\pm 1\} \xrightarrow{\sim} \operatorname{Aut}\mathcal{H}$ is the *projective special linear group*.

XII.4.4. The Riemann mapping theorem.

Motivation **XII.4.4.1.** — We want to determine when there is a conformal map $F\colon \Omega \to \mathcal{D}$ from a non-empty open set Ω to the unit disc $\mathcal{D}$. Some necessary conditions on Ω are the following.

(i) If $\Omega = \mathbb{C}$, then any map $F\colon \Omega \to \mathcal{D}$ is constant by Liouville's theorem XII.2.5.1 and hence not conformal. Thus, Ω must be proper.

(ii) Since $\mathcal{D}$ is simply connected (in the topological sense) and since F is continuous and bijective, Ω must be simply connected as well. This also implies that Ω is connected.

Surprisingly, these necessary conditions are actually sufficient.

CHAPTER XIII

ANALYTIC NUMBER THEORY

§ XIII.1. ENTIRE FUNCTIONS

***Jensen's formula* XIII.1.0.1.** — *Let Ω be an open set containing $\bar{\mathrm{D}}_R(0)$, and let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function with $f(0) \neq 0$ and that does not vanish anywhere on $\mathrm{C}_R(0)$. If $z_1, \dots, z_n$ are the zeroes of f inside $\mathrm{D}_R(0)$, then*

$$\log|f(0)| = \sum_{k=1}^{N} \operatorname{ord}_f z_k \log(|z_k|/R) + \frac{1}{2\pi} \int_0^{2\pi} \log|f(R\mathrm{e}^{\mathrm{i}\theta})| \,\mathrm{d}\theta.$$

PART VII

MODULAR FORMS

CHAPTER CONTENTS

§ XIII.2. INTRODUCTION

***Reference* XIII.2.0.1.** — For now, [DS05] only.

***Motivation* XIII.2.0.2.** — The modularity theorem states: *all rational elliptic curves arise from modular forms.* This goes back to Taniyama, Shimura nad Weil in the 1950's and 60's. Later, Wiles and Taylor proved the modularity theorem for a large class of elliptic curves in 1959, giving rise to Fermat's last theorem. The modularity theorem was proved in full generality by Breuil, Conrad, Diamond and Taylor in 2001.

***Motivation* XIII.2.0.3.** — Consider the quadratic $Q\colon x^2 = d$ for $d \in \mathbb{Z}\setminus\{0\}$, and for each prime p, define $a_p(Q) + 1$ as the number of solutions of Q modulo p. We can extend this multiplicatively by $a_{nm}(Q) = a_n(Q)a_m(Q)$. Note for odd p that $a_p(Q) = (d/p)$ is the Legendre symbol, so quadratic reciprocity says that $a_p(Q)$ only depends on the value $p \mod 4|d|$.

We can reinterpret the system of solution counts $\{a_p(Q)\}_p$ as eigenvalues of a finite-dimensional complex vector space associated to Q. Let $N = 4|d|$, let $G = (\mathbb{Z}/N\mathbb{Z})^\times$ and let $V_N = \operatorname{map}(G, \mathbb{C})$, which is a complex vector space. For each prime p, define the linear operator

$$T_p\colon V_N \longrightarrow V_N, \qquad (T_pf)(n) = \begin{cases} f(pn) & \text{if } p \nmid N, \\ 0 & \text{otherwise.} \end{cases}$$

Consider the particular element $f_Q\colon G \to \mathbb{C}$, $f(n) = a_n(Q)$ in V_N, which is well-defined by quadratic reciprocity as noted above. Then $(T_pf_Q)(n) = f_Q(pn) = a_{pn}(Q) = a_p(Q)a_n(Q)$ if $p \nmid N$ and $(T_pf_Q)(n) = 0$ otherwise. Thus, $T_pf_Q = a_p(Q)f_Q$ for all p and f is an eigenvector of all T_p with eigenvalues $\{a_p(Q)\}_p$, resp.

***Motivation* XIII.2.0.4.** — The modularity theorem is an analogues result. Consider the cubic $E\colon y^2 = 4x^3g_2x - g_3$ for $g_2, g_3 \in \mathbb{Z}$ with $g_2^3 - 27g_3^2 \neq 0$, which defines objects called *elliptic curves*. For each prime p, define $p - a_p(E)$ as the number of solutions (x, y) of E modulo p. Then a part of the modularity theorem states that the solution-counts $\{a_p(E)\}_p$ arises as a system of eigenvalues.

***Motivation* XIII.2.0.5.** — A *modular form* is a function f on the complex upper half plane $\mathcal{H}$ satisfying certain transformation and holomorphy conditions. For $\tau \in \mathcal{H}$, there is a Fourier expansion $f(\tau) = \sum_{n=0}^\infty a_n(f)\mathrm{e}^{2\pi i n\tau}$ where $a_n(f) \in \mathbb{C}$ for all n. Each non-zero modular form has an associated *weight* $k \in \mathbb{Z}$ and *level* $N \in \mathbb{Z}$, and all modular forms of a given weight and level form a vector space. So-called *Hecke operators*, including an operator T_p for each prime p, act linearly on these vector spaces. An *eigenform* is a modular form that is a simultaneous eigenvector for all Hecke operators.

The modularity theorem associates to E an eigenform f_E in a vector space V_N of modular forms of weight 2 at level N; this eigenform is called the *conductor* of E. The eigenvalues of f_E are its Fourier coefficients determined by $T_p(f) = a_p(f)f$ for all primes p, and a version of modularity states that $a_p(f) = a_p(E)$ for all primes p, i. e., *the Fourier coefficients are the solution counts.*

CHAPTER XIV

MODULAR FORMS AND MODULAR CURVES

***Definition* XIV.0.0.1.** — *The **modular group** is*

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Mat}_2(\mathbb{Z}) \;\middle|\; \det\gamma = ad - bc = 1 \right\}.$$

***Fact* XIV.0.0.2.** — *The modular group* $\mathrm{SL}_2(\mathbb{Z})$ *is generated by*

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Proof. — Let $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Since we have

$$\alpha \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^n = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} * & * \\ c & nc + d \end{pmatrix}$$

for all $n \in \mathbb{Z}$, if $c \neq 0$ then we may choose a unique n such that $-c/2 < nc + d \leq c/2$. So, w. l. o. g., assume for α that $-c/2 < d \leq c/2$. Since

$$\alpha \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} * & * \\ d & -c \end{pmatrix},$$

which swaps entries c and d up to sign, if $d \neq 0$ then we can redo the previous step to obtain $-d/2 < c' \leq d/2$. By iterating, we eventually can assume, w. l. o. g., that $c = 0$. Because the choice of n is unique, this only happens when $d = \pm 1$. Since $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)^2 = -\mathrm{id}$, we may assume that $d = 1$. By hypothesis, $\det\alpha = a = 1$, whence $\alpha = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)^b$. ◻

***Definition* XIV.0.0.3.** — *The **upper half plane** is* $\mathscr{H} = \{\tau \in \mathbb{C} : \mathrm{Im}\,\tau > 0\}$.(XIV.1)

***Definition* XIV.0.0.4.** — *Each* $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ *of the modular group defines an automorphism, the fractional linear transformation*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : \hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\} \longrightarrow \hat{\mathbb{C}}, \qquad \tau \longmapsto \frac{a\tau + b}{c\tau + d}.$$

Note that $-d/c \mapsto \infty$ *and* $\infty \mapsto a/c$ *if* $c \neq 0$, *and* $\infty \mapsto \infty$ *if* $c = 0$.

***Remark* XIV.0.0.5.** — For any $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, both $\pm\gamma$ define the same transformation. Moreover, the group of transformations is generated by $\tau \mapsto \tau + 1$ and $\tau \mapsto -tau^{-1}$, corresponding to the generators of the modular group.

(XIV.1) From Riemann surface theory, the three simply connected Riemann surfaces are the plane $\mathbb{C}$, the Riemann sphere $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ and the upper half plane $\mathscr{H}$.

***Fact* XIV.0.0.6.** — *The transformation group $\mathrm{SL}_2(\mathbb{Z})$ on $\hat{\mathbb{C}}$ restricts to a group action on $\mathcal{H}$.*

Proof. — To show that $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ restricts to $\mathcal{H}$, observe that

$$\operatorname{Im}\gamma(\tau) = \frac{\operatorname{Im}((a\tau+b)(c\overline{\tau}+d))}{|c\tau+d|^2} = \frac{\operatorname{Im}(ad\tau+bc\overline{\tau})}{|c\tau+d|^2} = \frac{\operatorname{Im}\tau}{|c\tau+d|^2}$$

since $ad-bc=1$, so $\gamma(\tau)\in\mathcal{H}$ if $\tau\in\mathcal{H}$. That $\mathrm{SL}_2(\mathbb{Z})$ is a group action means that $\mathrm{id}(\tau)=\tau$ and that $(\gamma\gamma')(\tau)=\gamma(\gamma'(\tau))$ for all $\gamma,\gamma'\in\mathrm{SL}_2(\mathbb{Z})$ (for all $\tau\in\hat{\mathbb{C}}$). The second identity follows from associativity of matrix multiplication if we interpret τ as $\left(\begin{smallmatrix} \tau \\ 1 \end{smallmatrix}\right)$ and $(a\tau+b)/(c\tau+d)$ as $\left(\begin{smallmatrix} a\tau+b \\ c\tau+d \end{smallmatrix}\right)$. □

***Definition* XIV.0.0.7.** — *For $k\in\mathbb{Z}$, a meromorphic function $f\colon\mathcal{H}\to\mathbb{C}$ is **weakly modular** of weight k if for all $\gamma=\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\in\mathrm{SL}_2(\mathbb{Z})$ and all $\tau\in\mathcal{H}$,*

$$f(\gamma(\tau)) = (c\tau+d)^k f(\tau).$$

***Remark* XIV.0.0.8.** — We will later show that it suffices to check this condition on the generators of $\mathrm{SL}_2(\mathbb{Z})$, i. e., it suffices to check that $f(\tau+1)=f(\tau)$ and $f(-\tau^{-1})=\tau^k f(\tau)$.

Insert reference.

***Example* XIV.0.0.9.** — Weak modularity of weight 0 is simply $\mathrm{SL}_2(\mathbb{Z})$-invariance.

***Example* XIV.0.0.10.** — Weak modularity of weight 2 naturally occurs in complex analysis. If we compute the path integral of $f(\tau)$ over $\mathrm{SL}_2(\mathbb{Z})$-invariant paths in $\mathcal{H}$, then f must satisfy $f(\gamma(\tau))\frac{\mathrm{d}}{\mathrm{d}\tau}\gamma(\tau)=f(\tau)$ for all $\gamma\in\mathrm{SL}_2(\mathbb{Z})$ by change of variables. Since

$$\frac{\mathrm{d}\gamma(\tau)}{\mathrm{d}\tau} = (c\tau+d)^{-2}(a(c\tau+d)-(a\tau+b)c) = (c\tau+d)^{-2}(ad-bc) = (c\tau+d)^{-2},$$

this condition is precisely weak modularity with weight 2.

***Observation* XIV.0.0.11.** — (i) The product of two weakly modular functions of weight k and l has weight kl.

(ii) Taking $\gamma=-\mathrm{id}$, a weakly modular function f satisfies $f=(-1)^k f$, so the only weakly modular function of any odd weight k is $f=0$. However, we will later develop weak modularity in more general contexts.

Insert reference.

(iii) While weak modularity does not make a function $\mathrm{SL}_2(\mathbb{Z})$-invariant, at least f and $f\circ\gamma$ have the same roots and poles because $c\tau+d$ on $\mathcal{H}$ has neither.

***Construction* XIV.0.0.12.** — Let $f\colon\mathcal{H}\to\mathbb{C}$ be a $\mathbb{Z}$-periodic holomorphic function. For example, since $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)\in\mathrm{SL}_2(\mathbb{Z})$, we have $f(\tau+1)=f(\tau)$ for any weakly modular function $f\colon\mathcal{H}\to\mathbb{C}$, and assume that f is holomorphic on $\mathcal{H}$. Let $D=\{q\in\mathbb{C} : |q|<1\}$ be the open unit disk, and recall from complex analysis that the $\mathbb{Z}$-periodic surjective map $\mathcal{H}\to D\setminus\{0\}$, $\tau\mapsto \mathrm{e}^{2\pi\mathrm{i}\tau}$ is holomorphic. The function $g\colon D\setminus\{0\}\to\mathbb{C}$, $q\mapsto f((\log q)/(2\pi\mathrm{i}))$ is well-defined, even though $\log q$ is determined up to $2\pi\mathrm{i}\mathbb{Z}$, and $f(\tau)=g(e^{2\pi\mathrm{i}\tau})$. If f is holomorphic, then g is holomorphic since the logarithm is locally holomorphic on $D\setminus\{0\}$. Hence, we have a Laurent expansion $g(q)=\sum_{n\geq -N} a_n q^n$ for all $q\in D\setminus\{0\}$. Observe that $q=\mathrm{e}^{2\pi\mathrm{i}\tau}\to 0$ as $\operatorname{Im}\tau\to\infty$ since $|q|=\mathrm{e}^{-2\pi\operatorname{Im}\tau}$.

So, thinking of ∞ as lying on the imaginary line, we define f to be **holomorphic at ∞** if g extends holomorphically to $q=0$, i. e., $g(q)=\sum_{n=0}^{\infty} a_n q^n$ sums over the naturals. In other words, f admits a *Fourier expansion* $f(\tau)=g(\mathrm{e}^{2\pi\mathrm{i}\tau})=\sum_{n=0}^{\infty} a_n(f)\mathrm{e}^{2\pi\mathrm{i}\tau n}$.

***Remark* XIV.0.0.13.** — The classical definition says that a holomorphic function f is holomorphic at infinity if $f(1/\tau)$ has a removable singularity at 0. In the context of modular forms, however, we exploit the fact that weakly modular functions are $\mathbb{Z}$-periodic and use the transformation $\tau\mapsto\mathrm{e}^{2\pi\mathrm{i}\tau}$ instead of $\tau\mapsto 1/\tau$.

***Lemma* XIV.0.0.14.** — *A $\mathbb{Z}$-periodic holomorphic function $f\colon \mathcal{H} \to \mathbb{C}$ is holomorphic at ∞ if $f(\tau)$ is bounded as $\operatorname{Im}\tau \to \infty$.*

Proof. — Since $\operatorname{Im}\tau \to \infty$ if and only if $q = \mathrm{e}^{2\pi\mathrm{i}\tau} \to 0$ and since $f(\tau) = g(\mathrm{e}^{2\pi\mathrm{i}\tau})$, this follows from Riemann's theorem on removable singularities. ◻

***Definition* XIV.0.0.15.** — *For $k \in \mathbb{Z}$, a function $f\colon \mathcal{H} \to \mathbb{C}$ is a **modular form** of weight k if f is:*

(i) *holomorphic on $\mathcal{H}$;*
(ii) *weakly modular of weight k;*
(iii) *holomorphic at ∞.*

We write $\mathcal{M}_k(\mathrm{SL}_2(\mathbb{Z}))$ for the complex vector space of modular forms of weight k.

***Remark* XIV.0.0.16.** — (i) Holomorphy at ∞ makes $\mathcal{M}_k(\mathrm{SL}_2(\mathbb{Z}))$ finite-dimensional. This also holds more generally.

(ii) Since the product of two modular forms of weight k and l is a modular form of weight $k+l$, which is easy to check with lemma XIV.0.0.14. Thus, we have a (commutative) graded ring

$$\mathcal{M}(\mathrm{SL}_2(\mathbb{Z})) = \bigoplus_{k\in\mathbb{Z}} \mathcal{M}_k(\mathrm{SL}_2(\mathbb{Z})).$$

Insert reference or explanation. Maybe we need explicit dimension formulas.

***Example* XIV.0.0.17.** — The zero function on $\mathcal{H}$ is a modular form of any weight, and constant functions on $\mathcal{H}$ are modular forms of weight 0.

We define a two-dimensional version of the Riemann zeta function $\zeta(k) = \sum_{d=1}^{\infty} 1/d^k$.

***Definition* XIV.0.0.18.** — *For $k \geq 3$, the **Eisenstein series** of weight k is*

$$G_k\colon \mathcal{H} \longrightarrow \mathbb{C}, \qquad G_k(\tau) = \sum_{(c,d)\in\mathbb{Z}^2\setminus\{(0,0)\}} \frac{1}{(c\tau+d)^k}.$$

***Fact* XIV.0.0.19.** — *The Eisenstein series $G_k(\tau)$ converges absolutely and locally uniformly (i. e., uniformly on compact sets) on $\mathcal{H}$. Hence, G_k is holomorphic and the sum can be rearranged.*

Proof. — If we consider partial sums over $(c,d) \in [-n,n]^2 \setminus \{(0,0)\}$, the series

$$\sum_{(c,d)\in\mathbb{Z}^2\setminus\{(0,0)\}} \frac{1}{\max\{|c|,|d|\}^k} = \sum_{n=1}^{\infty} \sum_{\max\{|c|,|d|\}=n} \frac{1}{n^k} = \sum_{n=1}^{\infty} \frac{8n}{n^k} = 8\zeta(k-1) \tag{XIV.0.0.19.1}$$

converges and dominates $\sum_{(c,d)} |c\tau+d|^{-k}$, whence $G_k(\tau)$ is absolutely convergent.

Next, for $A, B > 0$ let $\Omega = \{\tau \in \mathcal{H} \mid |\operatorname{Re}\tau| \leq A, \operatorname{Im}\tau \geq B\}$. We claim that there is a constant $C > 0$ such that $|\tau+\delta| > C\max\{1,|\delta|\}$ for all $\tau \in \Omega$ and all $\delta \in \mathbb{R}$. By exhaustion:

- if $|\delta| < 1$, then $|\tau+\delta| \geq \operatorname{Im}\tau \geq B$;
- if $1 \leq |\delta| \leq 2A$ and $\operatorname{Im}\tau > A$, then $|\tau+\delta| \geq \operatorname{Im}\tau > A \geq |\delta|/2$;
- if $1 \leq |\delta| \leq 2A$ and $\operatorname{Im}\tau \leq A$, then $|\tau+\delta|/|\delta|$ attains a minimum m on a compact set, so $|\tau+\delta| \geq m|\delta|$;
- if $|\delta| > 2A$, then $|\tau+\delta| \geq |\operatorname{Re}\tau+\delta| \geq |\delta| - |\operatorname{Re}\tau| \geq |\delta| - A \geq |\delta|/2$;

so any $C < \min\{B, \frac{1}{2}, m\}$ does the trick.

Any compact subset of $\mathcal{H}$ is bounded and hence is contained in a suitable set Ω. Hence for τ in this compact subset,

$$\sum_{(c,d)\in\mathbb{Z}^2\setminus\{(0,0)\}} \frac{1}{|c\tau+d|^k} = 2\zeta(k) + \sum_{c\neq 0} \frac{1}{|c|^k|\tau+d/c|^k} \leq 2\zeta(k) + \sum_{c\neq 0} \frac{1}{C^k\max\{|c|,|d|\}^k}$$

converges uniformly by the dominating equation (XIV.0.0.19.1). This allows us to interchange contour integrals over compact paths with the uniform limit, so Morrera's theorem shows that G_k is holomorphic.

PART VIII

CATEGORY THEORY

CHAPTER CONTENTS

XIV.0.0.20. — Main references are [KS06]; [nLa26] with some helpful examples from [Rie16]. Subsection on monoidal categories is supplemented by [Eti+15]; [Kas95].

***Notation* XIV.0.0.21.** — – Different from [KS06], we say locally small categories instead of $\mathcal{U}$-categories, and we say small categories instead of $\mathcal{U}$-small categories.

– We try to use small Latin letters for morphisms, capital Latin letters for functors, and Greek letters for natural transformations.

– Different from [KS06] and in accordance to [Rie16]; [nLa26], whiskering is notated by juxtaposition.

– We define the conventional contravariant Yoneda embedding $\mathrm{h}^{\mathcal{C}} \colon \mathcal{C}^{\mathrm{op}} \to \mathrm{Fun}(\mathcal{C}, \mathsf{Set})$. In the notation of [KS06], one defines $\mathrm{k}_{\mathcal{C}} = (\mathrm{h}^{\mathcal{C}})^{\mathrm{op}}$.

– We write $F{\downarrow}G$ instead of $M[\mathcal{C} \to \mathcal{E} \leftarrow \mathcal{D}]$ for the comma category of $F \colon \mathcal{C} \to \mathcal{E}$ and $G \colon \mathcal{D} \to \mathcal{E}$.

CHAPTER XV

CATEGORIES AND FUNCTORS

§ XV.1. BASIC NOTIONS

XV.1.1. Foundational issues.

***Axiom* XV.1.1.1** (Zermelo-Fraenkel set theory with choice, ZFC)**.** — ZFC can be expressed in first-order logic where the only objects are sets and the only predicate is membership of sets, written $\in$. Except for the axiom of extensionality (equality of sets) and axiom of regularity (virtually only used for handling ordinals or transfinite induction), we require the following:

(i) axiom schema of specification: if S is a set, then $T = \{x \in S \mid \phi(x, S, w_1, \dots, w_n)\}$ is a set based on a formula (condition) $\phi(x, S, w_1, \dots, w_n)$ (we are only allowed to form *subsets*);
(ii) axiom of pairing: if S and T are sets, then $\{S, T\}$ is a set;
(iii) axiom of union: if S is a set (of sets), then $\bigcup_{x\in S} x$ is a set;
(iv) axiom schema of replacement: if $(S_i)_{i\in I}$ is a family of sets indexed by a set I (more formally, a function $i \mapsto S_i$ on a set I with sets S_i as values), then $\{S_i \mid i \in I\}$ is a set;
(v) axiom of infinity: given $\varnothing$, then $\mathbb{N}$ is a set, formed by von Neumann ordinals $0 = \varnothing$ and $n = \{\varnothing, n-1\}$;
(vi) axiom of power set: if S is a set, then $\mathcal{P}(S) = \{T \subseteq S\}$ is a set.;
(vii) axiom of choice: if $(S_i)_{i\in I}$ is a family of non-empty sets indexed by an (infinite) set I, then there exists a family $(x_i)_{i\in I}$ with $x_i \in S_i$ (more formally, if I is a set of non-empty sets, then there exists a function f on I with $f(i) \in i$ for each $i \in I$).

This allows us to perform the following basic operations:

– complements $S \setminus T = \{x \in S \mid x \notin T\}$;
– unions $\bigcup_{i\in I} S_i = \bigcup_{x\in T} x$ where $T = \{S_i \mid i \in I\}$ by the axiom of replacement;
– intersections $\bigcap_{i\in I} S_i = \{x \in \bigcup_{i\in I} S_i \mid x \in S_i \text{ for all } i\}$;
– ordered pairs $(S, T) = \{\{S\}, \{S, T\}\}$ where $\{S\} = \{S, S\}$ by axiom of pairing;
– binary unions $S \cup T$ since $2 \in \mathbb{N}$;
– binary products $S \times T = \{(x, y) \in \mathcal{P}(\mathcal{P}(S \cup T)) \mid x \in S, y \in T\}$;
– function sets $\operatorname{map}(I, S)$ of all $f \in \mathcal{P}(I \times S)$ such that, for each $i \in I$ there is a unique $f(i) \in S$ such that $(i, f(i)) \in f$;
– products $\prod_{i\in I} S_i = \{f \colon I \to \bigcup_{i\in I} S_i \mid f(i) \in S_i \text{ for each } i\}$;
– disjoint unions $\coprod_{i\in I} S_i = \bigcup_{i\in I}(\{i\} \times S_i)$.

***Motivation* XV.1.1.2.** — In category theory, we want to speak of large collections of mathematical objects, e. g., 'all groups', 'all topological spaces', or ultimately 'all sets'. However, if one is not careful, then these size issues can lead to paradoxes like Russel's paradox.(XV.1) Hence, the collection

(XV.1) *Russel's paradox* states that: There is no set S such that $x \in S$ if and only if x is a set; or equivalently, of all sets is not a set. There are generally two ways to handle this.

(i) The first one, popularised by Grothendieck in SGA 4 (1963/1964), employs *Grothendieck universes* which are postulated to be a set. Then the 'set of all sets' is possible by interpreting it as the 'set of all members in a Grothendieck universe'. These universes are defined in such a way that all ordinary operations in ZFC can be performed in it.

(ii) The second one, much more common in current mathematics, introduces a new collection type of *classes*, and there may be classes of sets which are not themselves sets, so-called *proper classes*. This resolves the 'set of all sets' by saying that this is a 'class of all sets', prohibiting recursion. The notion of classes is formalised in *von Neumann-Bernays-Gödel set theory* (NBG), which is a conservative extension of ZFC, i. e., any statement provable in ZFC is also provable in NBG, and any statement that is provable in NBG but not in ZFC must involve classes.

We will cover both.

Definition* XV.1.1.3.** — *A* ***Grothendieck universe *or just* ***universe*** *is a set $\mathcal{U}$ satisfying the following properties:*

(i) *transitivity: if $u \in \mathcal{U}$ and $v \in u$, then $v \in \mathcal{U}$ (or equivalently, $u \in \mathcal{U}$ implies $u \subseteq \mathcal{U}$);*
(ii) *power set: if $u \in \mathcal{U}$, then $\mathcal{P}(\mathcal{U}) \subseteq \mathcal{U}$;*
(iii) *unions: if $I \in \mathcal{U}$ and $u_i \in \mathcal{U}$ for all $i \in I$, then $\bigcup_{i\in I} u_i \in \mathcal{U}$;*
(iv) *natural numbers: $\mathbb{N} \in \mathcal{U}$.*

In (iv), the natural numbers $\mathbb{N}$ are interpreted as the set of von Neumann ordinals, i. e., $0 = \varnothing \in \mathbb{N}$ and $n = \{\varnothing, n-1\}$ for all $n \geq 1$, so that $\varnothing \in \mathcal{U}$ by (i).[(XV.2)]

***Fact* XV.1.1.4** (properties of universes). — *Any universe $\mathcal{U}$ enjoys the following properties:*

(i) *subsets: if $u \in \mathcal{U}$ and $v \subseteq u$, then $v \in \mathcal{U}$;*
(ii) *singletons: if $u \in \mathcal{U}$, then $\{u\} \in \mathcal{U}$;*
(iii) *axiom schema of replacement: if $I \in \mathcal{U}$ and $u_i \in \mathcal{U}$ for all $i \in I$, then $\{u_i \mid i \in I\} \in \mathcal{U}$;*
(iv) *axiom of unions: if $u \in \mathcal{U}$, then $\bigcup_{x\in u} x \in \mathcal{U}$;*
(v) *intersections: if $I \in \mathcal{U}$ and $u_i \in \mathcal{U}$ for all $i \in I$, then $\bigcap_{i\in I} u_i \in \mathcal{U}$;*
(vi) *disjoint unions: if $I \in \mathcal{U}$ and $u_i \in \mathcal{U}$ for all $i \in I$, then $\coprod_{i\in I} u_i \in \mathcal{U}$;*
(vii) *function sets: if $u, v \in \mathcal{U}$, then $\operatorname{map}(u, v) \in \mathcal{U}$;*
(viii) *products: if $I \in \mathcal{U}$ and $u_i \in \mathcal{U}$ for all $i \in I$, then $\prod_{i\in I} u_i \in \mathcal{U}$.*

Proof. — For (i), $u \in \mathcal{U}$ implies $\mathcal{P}(u) \in \mathcal{U}$. Since $v \subseteq u$, we have $v \in \mathcal{P}(u)$, so $v \in \mathcal{U}$ by transitivity. Similarly for (ii), since $\{u\} \in \mathcal{P}(u)$, if $u \in \mathcal{U}$ then $\{u\} \in \mathcal{U}$. Next, (ii) implies (iii): if $I \in \mathcal{U}$ and $u_i \in \mathcal{U}$ for all $i \in I$, then $\{u_i \mid i \in I\} = \bigcup_{i\in I}\{u_i\} \in \mathcal{U}$. (iv) follows from transitivity, and (v) follows from (i).

The remaining statements follow from (i) and the following fact: if $u, v \in \mathcal{U}$, then $u \cup v \in \mathcal{U}$. Indeed, since $2 \in \mathbb{N}$ and $\mathbb{N} \in \mathcal{U}$, we have $2 \in \mathcal{U}$, whence $u \cup v \in \mathcal{U}$. For more details, see the definition of the basic set operations in axiom XV.1.1.1. ∎

In the context of category theory, we amend ZFC by the following axiom.

---

there is no set of all sets. Indeed, by way of contradiction, suppose that such a set S exists. By the axiom schema of *unrestricted* comprehension, consider the subset $R = \{x \in S \mid x \notin x\}$ of S. If $R \in R$, then $R \notin R$, and if $R \notin R$, then $R \in R$, a contradiction.

(XV.2) Originally, SGA 4 required $\mathcal{U} \neq \varnothing$ and used the *pairing* property instead of (iv): if $u, v \in \mathcal{U}$, then $\{u, v\} \in \mathcal{U}$. Like in fact XV.1.1.4, one shows that $u \in \mathcal{U}$ and $\varnothing \in \mathcal{P}(u) \in \mathcal{U}$ implies $\varnothing \in \mathcal{U}$. Then by pairing and singletons, $\mathbb{N} \subseteq \mathcal{U}$, but not necessarily $\mathbb{N} \in \mathcal{U}$. That is why by axiom XV.1.1.5, one w. l. o. g. chooses a universe large enough such that $\mathbb{N} \in \mathcal{U}$. However, relaxing (iv) in this way also allows the set of all *hereditarily finite sets* to be a Grothendieck universe.

***Axiom* XV.1.1.5** (Grothendieck)**.** — Every set belongs to some universe.

***Definition* XV.1.1.6.** — *Let $\mathcal{U}$ be a universe. A set is a **$\mathcal{U}$-set** if it belongs to $\mathcal{U}$, and a set is a **$\mathcal{U}$-class** if it is a subset of $\mathcal{U}$.*(XV.3)

***Example* XV.1.1.7.** — The principle idea is to almost always consider $\mathcal{U}$-small objects only, given a universe $\mathcal{U}$. For example, the *category of all groups* contains all $\mathcal{U}$-small groups G whose underlying set is $\mathcal{U}$-small and whose group operation $G \times G \to G$ is also $\mathcal{U}$-small.

***Definition* XV.1.1.8.** — *A class is a **set** if it belongs to some other class. In particular, every set is a class and classes only consist of sets. A class that is not a set is called a **proper class**.*

***Axiom* XV.1.1.9** (von-Neumann-Bernays-Gödel set theory, NBG)**.** — NBG is a conservative extension of ZFC (axiom XV.1.1.1) and can be expressed in first-order logic. Besides sets, we introduce a new type of object: classes. The notions unions, power sets and functions (families) extend to classes. Moreover, we change the axioms of ZFC in the following way:

– The axiom of extensionality (equality of sets) is extended to classes.
– As a benefit of classes, the axiom schema of replacement can be replaced by a single axiom since we may now speak of the class of (all) functions.
– We add theorem XV.1.1.10 below as an axiom. This avoids Russel's paradox.
– If desired, we modify the axiom of choice to the *axiom of global choice* by not only choosing in a family of non-empty sets, but choosing in the class of *all* non-empty sets. This is strictly stronger then ordinary choice.

Every other axiom in ZFC is left untouched.

Von Neumann and Bernays realised that NBG is finitely axiomatisable by replacing theorem XV.1.1.10 with seven other class existence axioms, four of them handling basic operations on classes, and three other manipulating tuples. With these axioms, theorem XV.1.1.10 is provable within NBG.

***Theorem* XV.1.1.10** (class existence theorem)**.** — *For every formula $\phi(x_1, \dots, x_n)$ that quantifies only over sets, not classes, there exists a class consisting of all n-tuples of sets $(x_1, \dots, x_n)$ satisfying ϕ.*

***Lemma* XV.1.1.11.** — *Let $f \colon A \to B$ be an injective class function. If A is proper, then B is proper.*

Proof. — Suppose that B is a set. Then the image $f(A) \subseteq B$ is a set, and we have a bijection $f \colon A \to f(A)$. We obtain a function $f^{-1} \colon f(A) \to A$ with image $f^{-1}(f(A)) = A$. By the axiom of replacement, A is a set, a contradiction. ∎

***Example* XV.1.1.12.** — By Russel's paradox, the class C of all sets is a proper class. Similarly, the classes of all groups, A-modules, topological spaces, etc. are proper classes since we can always find an of C into these classes and use lemma XV.1.1.11. For example, given a set $S \in C$, we can associate the trivial group $\{S\}$ consisting of one element; this defines an injection from C to the class of all groups.

(XV.3) Originally, SGA 4 does not define $\mathcal{U}$-classes at all, as in the context of categories $\mathcal{C}$, it often does not matter what $\mathrm{Ob}(\mathcal{C})$ and $\mathrm{Mor}(\mathcal{C})$ are w. r. t. $\mathcal{U}$ (they are always very big, and one can make them $\mathcal{V}$-sets by choosing a larger category $\mathcal{V} \ni \{\mathcal{U}, \mathrm{Ob}(\mathcal{C}), \mathrm{Mor}(\mathcal{C})\}$). However, it generalises the idea of $\mathcal{U}$-sets to *$\mathcal{U}$-small sets* which are sets that are in bijection with a $\mathcal{U}$-set. This is sometimes necessary since some constructions like localisations in derived $\mathcal{U}$-categories (with $\mathcal{U}$-hom-sets) lead outside of $\mathcal{U}$, but are still $\mathcal{U}$-small. However, categories $\mathcal{C}$ with $\mathcal{U}$-small hom-sets ('$\mathcal{U}$-locally small') do not define hom-functors $\mathrm{Hom}_{\mathcal{C}}(-, X) \colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}_{\mathcal{U}}$ into the category of $\mathcal{U}$-sets. Instead, a second axiom in Bourbaki set theory (SGA does not use ZFC) says that one can find $Z \cong \mathrm{Hom}_{\mathcal{C}}(Y, X)$ with $Z \in \mathcal{U}$ and then just defines $\mathrm{Hom}_{\mathcal{C}}(Y, X) := Z$.

The following is one of the reasons for future size restrictions of categories.

***Lemma* XV.1.1.13.** — *Let $f\colon A \to B$ be a class function, which we identify with its graph. If A is a proper class, then f is a proper class. In particular, f cannot be an element of another class, and $\mathrm{map}(A, B)$ is ill-defined if A is proper.*

Proof. — Recall that the graph of f consists of a unique pair $(x, f(x)) \in A \times B$ for each $x \in A$. Consider the surjective function $f \to A$ onto the first entry in $(x, f(x))$, which exists by theorem XV.1.1.10. If f were a set, then A is a set by the axiom of replacement, a contradiction. ∎

***Remark* XV.1.1.14.** — Any Grothendieck universe $\mathcal{U}$ provides a model for NBG by defining *classes* to be $\mathcal{U}$-classes and *sets* to be $\mathcal{U}$-sets. However, the language of universes is strictly stronger than NBG. The reason is that universes allow stacking $\mathcal{U}_0 \in \mathcal{U}_1 \in \cdots$ to construct ever larger collections, but classes must not contain classes in NBG.

***Convention* XV.1.1.15.** — *We will mostly work within the language of classes and sets in NBG. If there is ever a need to pass to the language of Grothendieck universes, replace* sets *by* $\mathcal{U}$-small sets *and* classes *by* subsets of $\mathcal{U}$ *for any universe* $\mathcal{U}$*.*

XV.1.2. Categories.

Definition* XV.1.2.1.** — *A* ***category $\mathcal{C}$ *consists of the following:*

(i) *a class* $\mathrm{Ob}(\mathcal{C})$*;*
(ii) *a class* $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ *called* ***hom-set*** *for each* $X, Y \in \mathrm{Ob}(\mathcal{C})$*;*
(iii) *a map*

$$\circ\colon \mathrm{Hom}_{\mathcal{C}}(X, Y) \times \mathrm{Hom}_{\mathcal{C}}(Y, Z) \longrightarrow \mathrm{Hom}_{\mathcal{C}}(X, Z), \qquad (f, g) \longmapsto g \circ f$$

called ***composition*** *for each* $X, Y, Z \in \mathrm{Ob}(\mathcal{C})$*;*

satisfying the following properties:

(i) $\circ$ *is associative, i. e.,* $(h \circ g) \circ f = h \circ (g \circ f)$ *for all* $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$*,* $g \in \mathrm{Hom}_{\mathcal{C}}(Y, Z)$*,* $h \in \mathrm{Hom}_{\mathcal{C}}(Z, W)$*;*
(ii) *for each* $X \in \mathrm{Ob}(\mathcal{C})$*, there exists an* ***identity morphism*** *(or just* identity*)* $\mathrm{id}_X \in \mathrm{Hom}_{\mathcal{C}}(X, X)$ *such that* $f \circ \mathrm{id}_X = \mathrm{id}_Y \circ f = f$ *for all* $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$*.*

***Fact* XV.1.2.2.** — *Identity morphisms are unique.*

***Notation* XV.1.2.3.** — Let $\mathcal{C}$ be a category. The elements of $\mathrm{Ob}(\mathcal{C})$ are called **objects** and the elements of $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ are called **morphisms** or **arrows**, often written as $f\colon X \to Y$ and sometimes as $f\colon Y \leftarrow X$. For a morphism $f\colon X \to Y$, the object X is called the **domain** or **source** and the object Y is called the **codomain** or **target** of f. We denote the class of all morphisms by $\mathrm{Mor}(\mathcal{C}) \coloneqq \bigcup_{X,Y \in \mathrm{Ob}(\mathcal{C})} \mathrm{Hom}(X, Y)$.

If a morphism $f\colon X \to Z$ can be written as a composition of $g\colon X \to Y$ and $h\colon Y \to Z$, then we say that f **factorises** through any of g, h or Y.

We will often drop the index from the notation $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ if it is clear which category is meant. Other notations, which we will not use, include $\mathcal{C} \coloneqq \mathrm{Ob}(\mathcal{C})$ and $\mathcal{C}(X, Y) \coloneqq \mathrm{Hom}_{\mathcal{C}}(x, Y)$.

***Motivation* XV.1.2.4.** — The categorical point of view emphasises not only the objects of a mathematical theory, but also the relational morphisms between the objects. It is the radical thought that objects can be best described not through their construction and properties, but by their behaviour with other objects.

Some would even go as far as to only consider morphisms since there is an injection $\mathrm{Ob}(\mathcal{C}) \hookrightarrow \mathrm{Mor}(\mathcal{C})$, $X \mapsto \mathrm{id}_X$. This can be made precise by defining a category to just be a class of morphisms $\mathrm{Mor}(\mathcal{C})$ with a few special maps on $\mathrm{Mor}(\mathcal{C})$, indicating composition and identity morphisms.

Definition* XV.1.2.5.** — *Let $\mathcal{C}$ be a category. It is* ***locally small *if every hom-set* $\mathrm{Hom}(X, Y)$ *for* $X, Y \in \mathrm{Ob}(\mathcal{C})$ *is a set, otherwise, it is* ***large****. It is* ***small*** *if it is locally small and if* $\mathrm{Ob}(\mathcal{C})$ *is a set.*

***Warning* XV.1.2.6.** — Despite the name, hom-sets are not necessarily sets. But in most categories across mathematics, they are, i. e., most categories are locally small. For that reason, some literature define all categories to be locally small, e. g., SGA 4.

***Remark* XV.1.2.7.** — Given a Grothendieck universe $\mathcal{U}$, if a category $\mathcal{C}$ (or even a family of categories)

***Example* XV.1.2.8.** — All of the following categories are only locally small, not small. Moreover, the following categories describe sets with some structure.

(i) The category of sets Set with maps as morphisms.
(ii) The category of groups Grp with group homomorphisms as morphisms.
(iii) The category of left A-modules $A\text{-}\mathsf{LMod}$ with A-linear maps as morphisms. If $A = k$ is a field, then we denote it by $k\text{-}\mathsf{Vect}$, the category of all k-vector spaces. If $A = \mathbb{Z}$ are the integers, then we denote it by AbGrp, the category of abelian groups.
(iv) The category of topological spaces Top with continuous maps as morphisms.
(v) The category of pointed sets Set_* with base point-preserving maps. Similarly, the category of all pointed topological spaces Top_* with base point-preserving continuous maps.
(vi) The category of binary relations. The objects are relations (X, ρ) on sets X, and morphisms are relation-preserving maps $f\colon (X, \rho) \to (Y, \tau)$, i. e., maps $f\colon X \to Y$ such that if $(x, x') \in \rho$, then $(f(x), f(x')) \in \tau$.
(vii) Other categories include the category of models from logic, the category of graphs, the category of finite-state automata, the category of smooth manifolds, etc.

However, there are also more abstract categories whose objects do not have an underlying set.

(viii) The category of matrices. It has integers as objects and matrices as morphisms, i. e., $\mathrm{Hom}(n, m) = \mathrm{Mat}_{m\times n}(\mathbb{Z})$. Composition is defined as $A \circ B = BA$.
(ix) Let $(I, \leq)$ be a poset (or more generally a preorder, i. e., not necessarily anti-symmetric). This defines a category $\mathcal{C}$ with objects $\mathrm{Ob}(\mathcal{C}) = I$ and morphisms

$$\mathrm{Hom}(x, y) = \begin{cases} \{\mathrm{pt}\} & \text{if } x \leq y, \\ \varnothing & \text{otherwise.} \end{cases}$$

Composition exists since $\leq$ is transitive, and reflexivity implies the identity morphism. A particular example of this is to consider a topological space (X, τ) and set $I = \tau$ and $\leq = \subseteq$.

***Notation* XV.1.2.9.** — (i) The **empty category** $\mathsf{0} = \varnothing$ has no objects and hence no morphisms.
(ii) The **trivial category** $\mathsf{1} = \{\bullet\}$ has only one object and no morphisms except for the identity.
(iii) Let $\mathsf{2} = \{\bullet \to \bullet\}$ denote the category consisting of two objects a and b with only one non-identity morphism $a \to b$.

Definition* XV.1.2.10.** — *Let $\mathcal{C}$ be a category. The* ***opposite category $\mathcal{C}^{\mathrm{op}}$ *of* $\mathcal{C}$ *is defined by* $\mathrm{Ob}(\mathcal{C}^{\mathrm{op}}) = \mathrm{Ob}(\mathcal{C})$, $\mathrm{Hom}_{\mathcal{C}^{\mathrm{op}}}(X, Y) = \mathrm{Hom}_{\mathcal{C}}(Y, X)$ *and composition is given by* $g \circ^{\mathrm{op}} f := f \circ g$*. In other words, we only flip all morphisms. We will always write* $\circ = \circ^{\mathrm{op}}$*. Sometimes, we denote the corresponding object or morphism in* $\mathcal{C}^{\mathrm{op}}$ *of* $X \in \mathrm{Ob}(\mathcal{C})$ *or* $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ *by* X^{op} *and* f^{op}*, resp.*

***Remark* XV.1.2.11.** — Many constructions and properties in category theory can be performed in the opposite category. Logically translating this construction or property back to the original category, we call this situation the **dual** to the original one. In other words, we flip all arrows for the dual situation. Many notions in category use the prefix ‘co-’ for the dual notion. Everything that holds for one situation also holds for the dual situation, since the proof itself can be applied to the opposite category. Thus, the dual proof is always tacitly implied.

Definition* XV.1.2.12.** — *A morphism $f\colon X \to Y$ is an* ***isomorphism *if there exists a morphisms $f^{-1}\colon Y \to X$ such that $f \circ f^{-1} = \mathrm{id}_Y$ and $f^{-1} \circ f = \mathrm{id}_X$. Such f^{-1} are automatically unique, whence the notation. In this case, we also write $f\colon X \xrightarrow{\sim} Y$ and say that X and Y are isomorphic, written $X \cong Y$.*

***Fact* XV.1.2.13.** — *Isomorphisms form an equivalence relation.*

Definition* XV.1.2.14.** — *An* ***endomorphism *is a morphism of the form $X \to X$. If it is also an isomorphism, then it is an* ***automorphism****.*

Definition* XV.1.2.15.** — (i) *A morphism $f\colon X \to Y$ is a* ***monomorphism *if for any pair of parallel morphisms $g_1, g_2\colon Z \rightrightarrows X$ with $f \circ g_1 = f \circ g_2$, we have $g_1 = g_2$. In other words, f is right-cancellable. We often write $f\colon X \hookrightarrow Y$ in this case.*

(ii) *Dually, a morphism $f\colon X \to Y$ is an* ***epimorphism*** *if for any pair of parallel morphisms $g_1, g_2\colon Y \rightrightarrows Z$ with $g_1 \circ f = g_2 \circ f$, we have $g_1 = g_2$. In other words, f is left-cancellable. We often write $f\colon X \twoheadrightarrow Y$ in this case.*

***Example* XV.1.2.16.** — In Set, the monomorphisms are precisely the injective functions and the epimorphisms are precisely the surjective functions.

(i) Let $f\colon X \hookrightarrow Y$ be a monomorphism. If $f(x_1) = f(x_2)$ for $x_1, x_2 \in X$, consider $g_1, g_2\colon \{\mathrm{pt}\} \rightrightarrows X$ with $g_i(\mathrm{pt}) = x_i$ for $i = 1, 2$. This shows that $x_1 = x_2$, so f is injective. The converse is easy.

(ii) Let $f\colon X \twoheadrightarrow Y$ be a epimorphism. Consider the two maps $g_1, g_2\colon Y \rightrightarrows \{0, 1\}$ defined as $g_1 = 1$ and $g_2 = \chi_{f(X)}$ (the indicator function on $f(X)$). This shows that $g_1 = g_2$, so $Y \subseteq f(X)$. The converse is easy.

***Remark* XV.1.2.17.** — A morphism $f\colon X \to Y$ is a:

(i) monomorphism if and only if $f \circ -\colon \operatorname{Hom}(Z, X) \to \operatorname{Hom}(Z, Y)$ is injective for all objects Z;
(ii) epimorphism if and only if $- \circ f\colon \operatorname{Hom}(Y, Z) \to \operatorname{Hom}(X, Z)$ is injective for all objects Z.

In particular, the composition of two monomorphisms (resp. epimorphisms) is again a monomorphism (resp. epimorphism).

Definition* XV.1.2.18.** — *For morphisms $f\colon X \to Y$ and $g\colon Y \to X$, if $g \circ f = \mathrm{id}_X$, we say that f is a* ***left inverse*,* ***retraction*** *or* ***split monomorphism*** *of g and that g is a* ***right inverse****,* ***section*** *or* ***split epimorphism*** *of f. In this case, f is a monomorphism and g is an epimorphism.*

***Fact* XV.1.2.19.** — *Isomorphisms are monomorphisms and epimorphisms.*

***Warning* XV.1.2.20.** — A morphism that is both a monomorphisms and an epimorphisms is not necessarily an isomorphism. For example, in Ring, the inclusion $\mathbb{Z} \hookrightarrow \mathbb{Q}$ is both a monomorphism and epimorphism, but not an isomorphism: there is no ring map $\mathbb{Q} \to \mathbb{Z}$.

***2-out-of-6 property* XV.1.2.21.** — (i) *The class of isomorphisms in a category $\mathcal{C}$ satisfies the* 2-out-of-6 property*: Given morphisms*

$$X \xrightarrow{f} Y \xrightarrow{g} Z \xrightarrow{h} U,$$

if $g \circ f$ and $h \circ g$ are isomorphisms, then f, g, h and $h \circ g \circ f$ are isomorphisms.

(ii) *Conversely, suppose that a class of morphisms (so-called* weak equivalences*) $W \subseteq \operatorname{Mor}(\mathcal{C})$ contains all identities and satisfies the 2-out-of-6 property (i. e., if $g \circ f, h \circ g \in W$, then $f, g, h, h \circ g \circ f \in W$). Then W contains all isomorphisms*

Proof. — For (i), observe that $f \circ (g \circ f)^{-1}$ is a right inverse of g and that $(h \circ g)^{-1} \circ h$ is a left inverse of g. Moreover, we have $f \circ (g \circ f)^{-1} = (h \circ g)^{-1} \circ h$ by rearranging. Hence, g is an isomorphism. Thus, $g^{-1} \circ (g \circ f) = f$ and $(h \circ g) \circ g^{-1} = h$ are isomorphisms as well.

Conversely for (ii), let f be an isomorphism and consider $f^{-1} \circ f \circ f^{-1}$. Then $f^{-1} \circ f = \mathrm{id}$ and $f \circ f^{-1} = \mathrm{id}$ are in W, and so is f by the 2-out-of-6 property. ∎

***2-out-of-3 property* XV.1.2.22.** — *The class of isomorphisms in a category satisfies the* 2-out-of-3 property*: Given morphisms*

$$X \xrightarrow{f} Y \xrightarrow{g} Z,$$

if two of the morphisms f, g and $g \circ f$ are isomorphisms, then the third one is an isomorphism.

Proof. — This can be directly proven. Alternatively, it is a direct consequence of the 2-out-of-6 property XV.1.2.21. Namely, if f and g are isomorphisms, then applying the 2-out-of-6 property XV.1.2.21 to

$$X \xrightarrow{f} Y \xrightarrow{\mathrm{id}_Y} Y \xrightarrow{g} Z$$

shows that $g \circ \mathrm{id}_Y \circ f = g \circ f$ is an isomorphism. If f and $g \circ f$ are isomorphism, then applying the 2-out-of-6 property XV.1.2.21 to

$$X \xrightarrow{\mathrm{id}_X} X \xrightarrow{f} Y \xrightarrow{g} Z$$

shows that g is an isomorphism. Similarly for if g and $g \circ f$ are isomorphisms. ∎

***Definition* XV.1.2.23.** — *A category $\mathcal{C}$ is:*

(i) ***discrete*** *if all morphisms are identity morphisms;*
(ii) *a **groupoid** if all morphisms are isomorphisms;*
(iii) ***thin*** *if $\mathrm{Hom}_{\mathcal{C}}(X,Y)$ has at most one element for any $X, Y \in \mathrm{Ob}(\mathcal{C})$;*
(iv) ***non-empty*** *if $\mathrm{Ob}(\mathcal{C}) \neq \varnothing$;*
(v) ***finite*** *if $\mathrm{Mor}(\mathcal{C})$ and hence $\mathrm{Ob}(\mathcal{C})$ (via $\mathrm{Ob}(\mathcal{C}) \hookrightarrow \mathrm{Mor}(\mathcal{C})$, $X \mapsto \mathrm{id}_X$) is finite;*
(vi) ***connected*** *if the diagram of $\mathcal{C}$ is weakly connected in the sense of graph theory, i. e., for any two $X, Y \in \mathrm{Ob}(\mathcal{C})$, there is a sequence $X = X_0, \dots, X_n = Y$ of objects such that a morphism $X_i \to X_{i+1}$ or $X_{i+1} \to X_i$ exists for all $0 \leq i \leq n-1$.*

***Example* XV.1.2.24.** — Any monoid M (i. e., a set with a associative product with unit) defines a category with a single object and M as morphisms. Similarly, a group G defines a groupoid with a single object and G as morphisms.

***Definition* XV.1.2.25.** — *Let $\mathcal{C}$ be a category. The **arrow category** $\mathrm{Arr}(\mathcal{C})$ has $\mathrm{Mor}(\mathcal{C})$ as objects and commutative squares*

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "u"'] & X' \arrow[d, "v"] \\ Y \arrow[r, "g"] & Y' \end{tikzcd}$$

*as morphisms $(u, v)\colon f \to g$. The composition is pasting of squares and the identity morphisms are the obvious ones. Additionally, there are the **source map** and **target map***

$$\begin{aligned} \mathrm{s}\colon \mathrm{Ob}(\mathrm{Arr}(\mathcal{C})) &\longrightarrow \mathrm{Ob}(\mathcal{C}), \qquad (X \longrightarrow Y) \longmapsto X; \\ \mathrm{t}\colon \mathrm{Ob}(\mathrm{Arr}(\mathcal{C})) &\longrightarrow \mathrm{Ob}(\mathcal{C}), \qquad (X \longrightarrow Y) \longmapsto Y. \end{aligned}$$

***Definition* XV.1.2.26.** — (i) *An object I is **initial** if for any other object X, there is exactly one morphisms $I \to X$.*

(ii) *Dually, an object T is **terminal** if for any other object, there is exactly one morphisms $X \to T$.*

(iii) *An object is a **zero object** if it is both initial and terminal, and we denote the zero object by 0. A morphism $X \to Y$ is a **zero morphism** if it factorises through 0, i. e., $X \to 0 \to Y$. By abuse of notation, we will also denote this by $0\colon X \to Y$.*

***Fact* XV.1.2.27.** — *If they exist, initial and terminal objects and unique up to unique isomorphism.*

Proof. — Let I and I' be initial objects. Then there are unique morphisms $I \to I'$ and $I' \to I$. Then the compositions $I \to I' \to I$ and $I' \to I \to I'$ must be id_I and $\mathrm{id}_{I'}$, resp., since the identity morphisms are the only morphisms $I \to I$ and $I' \to I'$. Thus, $I \xrightarrow{\sim} I'$ and $I' \xrightarrow{\sim} I$ are the unique isomorphisms.

***Fact* XV.1.2.28.** — *Let I and T be initial and terminal objects of a category, resp. If any morphism $T \to I$ exists, then $I \cong T$ is the zero object.*

Proof. — This is the same proof as fact XV.1.2.27.

***Example* XV.1.2.29.** — (i) In Set, the initial object is $\varnothing$ and the terminal object is $\{\mathrm{pt}\}$. The same is true for Top.

(ii) In Top_*, the zero object is $(\{\mathrm{pt}\}, \mathrm{pt})$.

(iii) In A-LMod, the zero object is the zero module 0. Similarly, the zero object of Grp is the trivial group $\{1\}$.

(iv) The category of the poset $(\mathbb{Z}, \leq)$ has neither initial nor terminal object.

***Example* XV.1.2.30.** — Fix a right A-module R and a left A-module L. We define a category whose objects are $\mathbb{Z}$-linear maps $f\colon R \times L \to M$ for $M \in \mathrm{Ob}(\mathsf{AbGrp})$ with $f(ra, l) = f(r, al)$ for all $a \in A$. The morphisms are $\mathbb{Z}$-linear maps $M \to N$ such that the diagram

$$\begin{tikzcd} & R \times L \arrow[dl] \arrow[dr] & \\ M \arrow[rr] & & N \end{tikzcd}$$

commutes. Since any map $f\colon R \times L \to M$ factorises through the tensor product $R \otimes_A L \to M$, the object $R \times L \to R \otimes_A L$ is initial.

XV.1.3. Functors.

Definition* XV.1.3.1.** — *Let $\mathcal{C}$ and $\mathcal{D}$ be two categories. A* ***covariant functor *or just* ***functor*** *$F\colon \mathcal{C} \to \mathcal{D}$ consists of the following:*

(i) *a map $F\colon \mathrm{Ob}(\mathcal{C}) \to \mathrm{Ob}(\mathcal{D})$;*
(ii) *a map $F\colon \mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$ for all $X, Y \in \mathcal{C}$;*

such that:

(i) *$F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$ for all $X \in \mathrm{Ob}(\mathcal{C})$;*
(ii) *$F(g \circ f) = F(g) \circ F(f)$ for all morphisms $f\colon X \to Y$ and $g\colon Y \to Z$ in $\mathcal{C}$.*

These two properties are called ***functoriality*** *of F.*

Dually, a ***contravariant functor*** *$F\colon \mathcal{C} \to \mathcal{D}$ is a functor $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$. In particular, it satisfies $F(g \circ f) = F(f) \circ F(g)$.*

***Example* XV.1.3.2.** — (i) For two categories I and $\mathcal{C}$, let X be an object of $\mathcal{C}$. We define the **constant functor** $\Delta_X\colon I \to \mathcal{C}$ by $i \mapsto X$ on objects and $f \mapsto \mathrm{id}_X$ on morphisms. Note that we usually gather the domain I from context.

(ii) For a category $\mathcal{C}$, the **identity functor** $\mathrm{id}_{\mathcal{C}}\colon \mathcal{C} \to \mathcal{C}$ is defined in the obvious way. It induces a contravariant functor $\mathcal{C}^{\mathrm{op}} \to \mathcal{C}$.

(iii) For an object X of a category $\mathcal{C}$, the functor $1 \to \mathcal{C}$, $\mathrm{pt} \mapsto X$ is the **inclusion functor** of X. By abuse of notation, this is often denoted just by X.

(iv) For a category $\mathcal{C}$, any functor $\mathcal{C} \to \mathcal{C}$ is called an **endofunctor** of $\mathcal{C}$.

***Convention* XV.1.3.3.** — *We will avoid the distinction of covariant and contravariant functors as much as possible. Instead, we will use the notations $\mathcal{C} \to \mathcal{D}$ and $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$ and just say* functor *to both of them.*

***Remark* XV.1.3.4.** — (i) Two functors $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{E}$ can be composed in the obvious way: compose the map on objects and the map on hom-sets. We denote their composition by $G \circ F$, which is again a functor.

(ii) If $F\colon \mathcal{C} \to \mathcal{D}$ is a functor, then $F^{\mathrm{op}}\colon \mathcal{C}^{\mathrm{op}} \to \mathcal{D}^{\mathrm{op}}$ is a functor.

***Example* XV.1.3.5.** — (i) The fundamental group of a pointed topological space defines a functor $\pi_1\colon \mathsf{Top}_* \to \mathsf{Grp}$.

(ii) The dual space of a vector space defines a functor $(-)^*\colon k\text{-}\mathsf{Vect}^{\mathrm{op}} \to k\text{-}\mathsf{Vect}$.

***Fact* XV.1.3.6.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor. If f is an isomorphism in $\mathcal{C}$, then $F(f)$ is an isomorphism in $\mathcal{D}$.*

***Definition* XV.1.3.7.** — *A functor $F\colon \mathcal{C} \to \mathcal{D}$ is:*

(i) ***faithful*** *(resp.* ***full, fully faithful****) if $\mathrm{Hom}_{\mathcal{C}}(X,Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X),F(Y))$ is injective (resp. surjective, bijective) for all $X,Y \in \mathrm{Ob}(\mathcal{C})$;*

(ii) ***essentially surjective*** *if for each $Y \in \mathrm{Ob}(\mathcal{D})$ there exists $X \in \mathrm{Ob}(\mathcal{C})$ and an isomorphism $F(X) \xrightarrow{\sim} Y$;*

(iii) ***conservative*** *if a morphism f in $\mathcal{C}$ is an isomorphism whenever $F(f)$ is an isomorphism in $\mathcal{D}$.*

All of these properties are transitive, i. e., if $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{E}$ have one of these properties, then $G \circ F$ has this property as well.

***Lemma* XV.1.3.8.** — *Fully faithful functors $F\colon \mathcal{C} \to \mathcal{D}$ are conservative. Even stronger, if $F(X) \cong F(Y)$ for objects X and Y in $\mathcal{C}$, then there exists a unique isomorphism $X \xrightarrow{\sim} Y$ in $\mathcal{C}$.*

Proof. — Let f be a morphism in $\mathcal{C}$. If $F(f)$ is an isomorphism, then since F is full, there exists a morphism g in $\mathcal{C}$ such that $F(g)$ is the inverse of f. Since F is faithful, $F(g \circ f) = \mathrm{id}$ and $F(f \circ g) = \mathrm{id}$ implies that $g \circ f = \mathrm{id}$ and $f \circ g = \mathrm{id}$, whence g is the inverse of f.

If there is an isomorphism $F(X) \xrightarrow{\sim} F(Y)$, then by fully faithfulness, there is a corresponding morphism $X \to Y$ in $\mathcal{C}$, and this must be an isomorphism by the previous part. □

***Lemma* XV.1.3.9.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a faithful functor, and let f be a morphism in $\mathcal{C}$. If $F(f)$ is an epimorphism (resp. monomorphism), then so is f.*

Proof. — Let $F(f)\colon F(X) \to F(Y)$ be an epimorphism in $\mathcal{D}$ and let $g,h\colon Y \rightrightarrows Z$ be parallel morphisms such that $g \circ f = h \circ f$. Then $F(g) \circ F(f) = F(h) \circ F(f)$. Since $F(f)$ is an epimorphism, we have $F(g) = F(h)$, so $g = h$ because F is faithful. □

Definition* XV.1.3.10.** — *Let $\mathcal{C}$ be a category. Formally, a* **diagram** *of shape I in $\mathcal{C}$ is a functor $F\colon I \to \mathcal{C}$. The category I is called the* ***index category *of F. The diagram is* ***finite*** *(resp.* ***small****) if I is finite (resp. small).*

Informally, a **diagram** *in $\mathcal{C}$ is a directed graph I (including loops and parallel edges) with objects in $\mathcal{C}$ as vertices and morphisms in $\mathcal{C}$ as directed edges. A diagram in $\mathcal{C}$* **commutes** *if for any two objects, any composition of morphisms along any direct path between them agree. We usually do not include the composition morphisms and identity morphisms in the diagram.*

***Remark* XV.1.3.11.** — Diagrams are the main tool in category theory to visualise situations and the prove statements. They are what equations are to algebraists. Most of the time, the index categories are very small, often finite. We will rarely use the formal definition of a diagram, but often draw a diagram and show that it commutes, which is sometimes called *diagram chasing* (in the wider sense of category theory, not homological algebra). Checking commutativity is the same as checking functoriality in the formal sense.

***Example* XV.1.3.12.** — Here are two diagrams:

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "h"'] & Y \arrow[d, "g"] \\ V \arrow[r, "k"] & Z, \end{tikzcd} \qquad \begin{tikzcd} X \arrow[r, "f"] \arrow[rr, bend left, "l"] & Y \arrow[r, "g"] & Z. \end{tikzcd}$$

The first one commutes if $g \circ f = k \circ h$, the second one if $l = g \circ f$. Note that not all diagrams commute, e. g., $\bullet \rightrightarrows \bullet$ in general.

***Definition* XV.1.3.13.** — *Let $\mathcal{C}$ be a category. A **subcategory** $\mathcal{C}'$ of $\mathcal{C}$ is defined by:*

- $\mathrm{Ob}(\mathcal{C}') \subseteq \mathrm{Ob}(\mathcal{C})$ *and* $\mathrm{Hom}_{\mathcal{C}'}(X,Y) \subseteq \mathrm{Hom}_{\mathcal{C}}(X,Y)$ *for all* $X, Y \in \mathrm{Ob}(\mathcal{C}')$*;*
- *composition in $\mathcal{C}'$ is induced by $\mathcal{C}$;*
- *identity morphisms in $\mathcal{C}'$ are those from $\mathcal{C}$.*

*It comes with a natural **embedding functor** $\mathcal{C}' \hookrightarrow \mathcal{C}$, which is faithful.*

Moreover, a subcategory $\mathcal{C}'$ is:

(i) ***full*** *if* $\mathrm{Hom}_{\mathcal{C}'}(X,Y) = \mathrm{Hom}_{\mathcal{C}}(X,Y)$ *for all* $X, Y \in \mathrm{Ob}(\mathcal{C}')$*, i. e., $\mathcal{C}'$ has all morphisms of $\mathcal{C}$;*

(ii) ***replete*** *if for any* $X' \in \mathrm{Ob}(\mathcal{C}')$ *and any isomorphism* $f\colon X' \xrightarrow{\sim} X$ *in $\mathcal{C}$, both X' and f are in $\mathcal{C}'$, i. e., $\mathcal{C}'$ is closed under isomorphisms;*

(iii) ***saturated*** *if it is full and replete.*

Definition* XV.1.3.14.** — *A category $\mathcal{C}$ is **concrete** if it has a faithful **forgetful functor $\mathrm{For}\colon \mathcal{C} \to \mathsf{Set}$.

***Remark* XV.1.3.15.** — The term *forgetful functor* has no precise definition and is used for various functors in category theory. It simply indicates that something is 'forgotten'.

***Example* XV.1.3.16.** — Most categories in mathematics are concrete and their forgetful functors give the underlying set. For example, $\mathrm{For}\colon \mathsf{Top} \to \mathsf{Set}$ forgets the topology, or $\mathrm{For}\colon \mathsf{Grp} \to \mathsf{Set}$ forgets the group structure. Other forgetful functors include $\mathsf{AbGrp} \to \mathsf{Grp}$ or $\mathsf{Field} \to \mathsf{Ring}$. Note that most forgetful functors are not full.

***Lemma* XV.1.3.17.** — *Let $\mathcal{C}$ be a concrete category with forgetful functor* $\mathrm{For}\colon \mathcal{C} \to \mathsf{Set}$*. Then a morphism f of $\mathcal{C}$ is a monomorphism (resp. epimorphism) if* $\mathrm{For}(f)$ *is injective (resp. surjective).*

Proof. — Combine example XV.1.2.16 and lemma XV.1.3.9. □

***Definition* XV.1.3.18.** — *Let $(\mathcal{C}_i)_{i\in I}$ be a family of category indexed by a set I.*

(i) *The **product category** $\prod_{i\in I} \mathcal{C}_i$ has $\prod_{i\in I} \mathrm{Ob}(\mathcal{C}_i)$ as objects and $\prod_{i\in I} \mathrm{Hom}_{\mathcal{C}_i}(X_i, Y_i)$ as hom-sets.*

(ii) *The **disjoint union category** $\coprod_{i\in I} \mathcal{C}_i$ has*

$$\coprod_{i\in I} \mathrm{Ob}(\mathcal{C}_i) = \{(X,i) \mid i \in I, X \in \mathrm{Ob}(\mathcal{C}_i)\}$$

as objects and

$$\mathrm{Hom}_{\coprod_{i\in I}\mathcal{C}_i}(X,Y)=\begin{cases}\mathrm{Hom}_{\mathcal{C}_i}(X,Y) & \textit{if both } X,Y \textit{ in } \mathcal{C}_i,\\ \varnothing & \textit{otherwise,}\end{cases}$$

as hom-sets.

If I is finite, say, $(\mathcal{C}_1,\dots,\mathcal{C}_n)$, then we write $\mathcal{C}_1\times\cdots\times\mathcal{C}_n$ and $\mathcal{C}_1\amalg\cdots\amalg\mathcal{C}_n$ instead.

***Remark* XV.1.3.19.** — Given a family of functors $(F_i\colon \mathcal{C}_i\to\mathcal{D}_i)_{i\in I}$, we naturally obtain functors $\prod_{i\in I}F_i\colon \prod_{i\in I}\mathcal{C}_i\to\prod_{i\in I}\mathcal{D}_i$ and $\coprod_{i\in I}F_i\colon \coprod_{i\in I}\mathcal{C}_i\to\coprod_{i\in I}\mathcal{D}_i$.

***Definition* XV.1.3.20.** — *A **bifunctor** is a functor $F\colon \mathcal{C}_1\times\mathcal{C}_2\to\mathcal{D}$. This means that $F(X_1,-)\colon\mathcal{C}_2\to\mathcal{D}$ and $F(-,X_2)\colon\mathcal{C}_1\to\mathcal{D}$ are functors for all objects X_i in $\mathcal{C}_i$, and the diagram*

$$\begin{tikzcd} F(X_1,X_2) \arrow[r, "F(X_1,f_2)"] \arrow[d, "F(f_1,X_2)"'] \arrow[dr, dashed, "F(f_1,f_2)"] & F(X_1,Y_2) \arrow[d, "F(f_1,Y_2)"] \\ F(Y_1,X_2) \arrow[r, "F(Y_1,f_2)"] & F(Y_1,Y_2) \end{tikzcd}$$

commutes for all morphisms $f_i\colon X_i\to Y_i$ in $\mathcal{C}_i$.

***Example* XV.1.3.21.** — (i) Let $\mathcal{C}$ be a locally small category. Then the bifunctor

$$\mathrm{Hom}_{\mathcal{C}}(-,-)\colon \mathcal{C}^{\mathrm{op}}\times\mathcal{C}\longrightarrow \mathsf{Set}$$

is called the **hom-functor**. Note that $\mathcal{C}$ must be locally small for the functor to have a codomain. If Grothendieck universes are assumed, then one can choose a sufficiently large universe and drop the size restriction.

(ii) For an algebra A over a commutative ring k, we have bifunctors

$$-\otimes_A -\colon A\text{-}\mathsf{RMod}\times A\text{-}\mathsf{LMod}\longrightarrow k\text{-}\mathsf{Mod},\qquad \mathrm{Hom}_A(-,-)\colon A\text{-}\mathsf{LMod}^{\mathrm{op}}\times A\text{-}\mathsf{LMod}\longrightarrow k\text{-}\mathsf{Mod}.$$

The following construction generalises many categories arising from two functors.

***Definition* XV.1.3.22** (Lawvere 1963)**.** — *Let $F\colon\mathcal{C}\to\mathcal{E}$ and $G\colon\mathcal{D}\to\mathcal{E}$ be two functors. The **comma category**[(XV.4)] $F\downarrow G$ or F/G consists of the following:*

– the objects are triples (X,Y,f) with $X\in\mathrm{Ob}(\mathcal{C})$, $Y\in\mathrm{Ob}(\mathcal{D})$ and $f\colon F(X)\to G(Y)$ in $\mathcal{E}$;
– the morphisms $(X,Y,f)\to(X',Y',f')$ are pairs of morphisms (u,v) with $u\colon X\to X'$ in $\mathcal{C}$ and $v\colon Y\to Y'$ in $\mathcal{D}$ such that the diagram

$$\begin{tikzcd} F(X) \arrow[r, "f"] \arrow[d, "F(u)"'] & G(Y) \arrow[d, "G(v)"] \\ F(X') \arrow[r, "f'"] & G(Y') \end{tikzcd}$$

commutes.

Composition is the obvious one from $\mathcal{C}\times\mathcal{D}$. There are two faithful forgetful functors $\mathrm{For}_{\mathcal{C}}\colon(F\downarrow G)\to\mathcal{C}$, $(X,Y,f)\mapsto X$ and $\mathrm{For}_{\mathcal{D}}\colon(F\downarrow G)\to\mathcal{D}$, $(X,Y,f)\mapsto Y$.

***Example* XV.1.3.23.** — The arrow category $\mathrm{Arr}(\mathcal{C})$ is the comma category $\mathrm{id}_{\mathcal{C}}\downarrow\mathrm{id}_{\mathcal{C}}$.

(XV.4) Historically, a comma category was written as (F,G), hence the name.

***Fact* XV.1.3.24.** — *Let $F\colon \mathcal{C} \to \mathcal{E}$ and $G\colon \mathcal{D} \to \mathcal{E}$ be functors. If $\mathcal{C}$ and $\mathcal{D}$ are locally small, then the comma category $F \downarrow G$ is locally small. If $\mathcal{C}$ and $\mathcal{D}$ are small and if $\mathcal{E}$ is locally small, then $F \downarrow G$ is small.*

Proof. — Suppose that $\mathcal{C}$ and $\mathcal{D}$ are locally small. Let (X, Y, f) and (X', Y', f') be two objects in $F \downarrow G$. Then we have

$$\mathrm{Hom}_{F\downarrow G}((X, Y, f), (X', Y', f')) \subseteq \mathrm{Hom}_{\mathcal{C}}(X, X') \times \mathrm{Hom}_{\mathcal{D}}(Y, Y')$$

where the right-hand side is a product of sets and hence a set. Thus, the left-hand side is a set and $F \downarrow G$ is locally small.

Now, suppose that $\mathcal{C}$ and $\mathcal{D}$ are small and that $\mathcal{E}$ is locally small. We have

$$\mathrm{Ob}(F \downarrow G) \subseteq \mathrm{Ob}(\mathcal{C}) \times \mathrm{Ob}(\mathcal{D}) \times \coprod_{X\in\mathrm{Ob}(\mathcal{C})} \coprod_{Y\in\mathrm{Ob}(\mathcal{D})} \mathrm{Hom}_{\mathcal{E}}(F(X), G(Y)),$$

where the right-hand side is a set. Thus, $F \downarrow G$ is small. ◻

***Definition* XV.1.3.25.** — *Let $\mathcal{C}$ be a category and let A be an object of $\mathcal{C}$. Let $A\colon 1 \to \mathcal{C}$,* $\mathrm{pt} \mapsto A$ *be the inclusion of A.*

(i) *The* ***over category*** *or* ***slice category*** *$\mathcal{C}/A$ of $\mathcal{C}$ over A is the comma category $\mathrm{id}_{\mathcal{C}} \downarrow A$. Explicitly, it has morphisms $X \to A$ as objects and commutative triangles*

$$\begin{tikzcd} X \arrow[rr, "f"] \arrow[dr] & & Y \arrow[dl] \\ & A & \end{tikzcd}$$

as morphisms.

(ii) *Dually, the* ***under category*** *or* ***coslice category*** *$A/\mathcal{C}$ of $\mathcal{C}$ under A is the comma category $A \downarrow \mathrm{id}_{\mathcal{C}}$. Explicitly, it has morphisms $A \to X$ as objects and commutative triangles*

$$\begin{tikzcd} X \arrow[rr, "f"] & & Y \\ & A \arrow[ul] \arrow[ur] & \end{tikzcd}$$

as morphisms.

Both categories have forgetful functors $\mathrm{For} = \mathrm{s}\colon \mathcal{C}/A \to \mathcal{C}$ *and* $\mathrm{For} = \mathrm{t}\colon A/\mathcal{C} \to \mathcal{C}$*, resp.*

***Definition* XV.1.3.26.** — *For a category $\mathcal{C}$, let $\pi_0(\mathcal{C})$ be the set of connected components in $\mathcal{C}$. In other words, we define an equivalence relation on* $\mathrm{Ob}(\mathcal{C})$ *by $X \sim Y$ if X and Y are connected. Then $\pi_0(\mathcal{C})$ is the class of equivalence classes of* $\mathrm{Ob}(\mathcal{C})$ *w. r. t. $\sim$.*

***Remark* XV.1.3.27.** — If $A \in \mathrm{Ob}(\mathcal{C})$, then the objects of $\mathcal{C}/A$ and of $A/\mathcal{C}$ are both the connected component of A, i. e., the equivalence class $[A] \in \pi_0(\mathcal{C})$. In particular, $\mathcal{C}$ is connected if and only if $\pi_0(\mathcal{C})$ is consists of one element.

***Definition* XV.1.3.28.** — *Let $\mathcal{C}$ be a category and let X be an object of $\mathcal{C}$.*

(i) *Two monomorphisms $f\colon Y \hookrightarrow X$ and $g\colon Z \hookrightarrow X$ are isomorphic if there exists an isomorphism $h\colon Y \xrightarrow{\sim} Z$ such that the diagram*

$$\begin{tikzcd} Y \arrow[rr, "h", "\sim"'] \arrow[dr, hook, "f"'] & & Z \arrow[dl, hook', "g"] \\ & X & \end{tikzcd}$$

*commutes, i. e., $f \cong g$ in $\mathcal{C}/X$. Note that such an h is unique since g is a monomorphism. A **subobject** of X is an isomorphism class of monomorphisms with target X.*

(ii) *Dually, two epimorphisms $f\colon X \twoheadrightarrow Y$ and $g\colon X \twoheadrightarrow Z$ are isomorphic if $f \cong g$ in $X/\mathcal{C}$. A **quotient** of X is an isomorphism class of epimorphisms with source X.*

***Remark* XV.1.3.29.** — We can impose a partial order on the class of all subobjects of X by defining $(f\colon Y \hookrightarrow X) \leq (g\colon Z \hookrightarrow X)$ if there exists a morphism $f \to g$ in $\mathcal{C}/X$. Note that such a morphism $f \to g$ is unique since g is a monomorphism.

XV.1.4. Natural transformations.

***Definition* XV.1.4.1.** — *Let $F, G\colon \mathcal{C} \to \mathcal{D}$ be two functors. A **natural transformation** $\alpha\colon F \to G$ consists of a morphism $\alpha_X\colon F(X) \to G(X)$ for each $X \in \mathrm{Ob}(\mathcal{C})$ such that the diagram*

$$\begin{tikzcd} F(X) \arrow[r, "F(f)"] \arrow[d, "\alpha_X"'] & F(Y) \arrow[d, "\alpha_Y"] \\ G(X) \arrow[r, "G(f)"] & G(Y) \end{tikzcd}$$

*commutes for every morphism $f\colon X \to Y$ in $\mathcal{C}$. The commutativity of this square is called **naturality** of α. We often visualise natural transformations by the diagram*

$$\begin{tikzcd} \mathcal{C} \arrow[r, bend left, "F"] \arrow[r, bend right, "G"'] \arrow[r, phantom, "\Downarrow\alpha"] & \mathcal{D} \end{tikzcd}$$

and, if $\mathcal{C}$ is small, denote the class of all natural transformations by $\mathrm{Nat}(F, G)$.

*A natural transformation $\alpha\colon F \to G$ is a **natural isomorphism** if α_X is an isomorphism for each $X \in \mathrm{Ob}(\mathcal{C})$. In this case, we say that F and G are **isomorphic**.*

***Warning* XV.1.4.2.** — In order for $\mathrm{Nat}(F, G)$ to be defined, by lemma XV.1.1.13, $\mathrm{Ob}(\mathcal{C})$ must be a set since any natural transformation $\eta\colon F \to G$ is essentially a function $\mathrm{Ob}(\mathcal{C}) \to \mathrm{Mor}(\mathcal{D})$. However, there is no real expense if we further require $\mathcal{C}$ to be small, so we do that. Alternatively, by choosing sufficiently large Grothendieck universes, we can remove the smallness condition.

***Notation* XV.1.4.3.** — As with normal categories, we will use diagrams with categories as vertices and functors as directed edges to visualise situations involving functors. Apart from saying that such a diagram commutes, we also say that a diagram **quasi-commutes** if the composition of functors are equal up to natural isomorphism.

***Example* XV.1.4.4.** — Natural transformations formalise the idea that some map is 'natural', i. e., entails no choice. For example, the double dual space on k-vector spaces defines a natural isomorphism from $\mathrm{id}_{k\text{-}\mathsf{Vect}}$ to the double dual $(-)^{**}\colon k\text{-}\mathsf{Vect} \to k\text{-}\mathsf{Vect}$ since the diagram

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\wr"'] & W \arrow[d, "\wr"] \\ V^{**} \arrow[r, "f^{**}"] & W^{**} \end{tikzcd}$$

commutes for all k-linear maps $f\colon V \to W$. Note that the isomorphism $V \xrightarrow{\sim} V^{**}$, $v \mapsto (v^* \mapsto v^*(v))$ entails no choice.

On the other hand, there cannot be a natural isomorphism from the identity functor on k-$\mathsf{FinVect}$ to the single dual $(-)^*\colon k\text{-}\mathsf{FinVect}^{\mathrm{op}} \to k\text{-}\mathsf{FinVect}$ since the isomorphism $V \cong V^*$ is not natural. Although $V \to V^*$ is natural, the inverse map requires a choice of basis for V, and linear maps $f\colon V \to W$ almost never preserve bases. Another reason is that $\mathrm{id}_{k\text{-}\mathsf{FinVect}}$ is covariant and $(-)^*$ is contravariant, but this is besides the point.

***Example* XV.1.4.5.** — Let $\mathcal{C}$ be a locally small category. The hom-functor $\mathrm{Hom}_{\mathcal{C}}(-,-)\colon \mathcal{C}^{\mathrm{op}} \times \mathcal{C} \to \mathsf{Set}$ is natural in both arguments. This means that any morphism $f\colon X \to Y$ in $\mathcal{C}$ gives natural transformations

$$f_* \coloneqq \mathrm{Hom}_{\mathcal{C}}(-,f) = f \circ - \colon \mathrm{Hom}_{\mathcal{C}}(-,X) \longrightarrow \mathrm{Hom}_{\mathcal{C}}(-,Y),$$
$$f^* \coloneqq \mathrm{Hom}_{\mathcal{C}}(f,-) = - \circ f \colon \mathrm{Hom}_{\mathcal{C}}(Y,-) \longrightarrow \mathrm{Hom}_{\mathcal{C}}(X,-)$$

of functors $\mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ and $\mathcal{C} \to \mathsf{Set}$, resp. These are also called **hom-functors**.

***Definition* XV.1.4.6.** — *Let $\mathcal{C}$ be a small category and $\mathcal{D}$ be a category. The **functor category** $\mathrm{Fun}(\mathcal{C},\mathcal{D})$ from $\mathcal{C}$ to $\mathcal{D}$ has functors $\mathcal{C} \to \mathcal{D}$ as objects and natural transformations between functors $\mathcal{C} \to \mathcal{D}$ as morphisms. If $\alpha\colon F \to G$ and $\beta\colon G \to H$ are two natural transformations, then their composition $\beta \circ \alpha\colon F \to H$ is defined by $(\beta \circ \alpha)_X = \beta_X \circ \alpha_X$. Other notations include $[\mathcal{C},\mathcal{D}]$ and $\mathcal{D}^{\mathcal{C}}$, which we will not use often.*

Fix whether notation is used often.

***Warning* XV.1.4.7.** — Using any larger categories $\mathcal{C}$ for $\mathrm{Fun}(\mathcal{C},\mathcal{D})$ is not possible by lemma XV.1.1.13, because a functor $F\colon \mathcal{C} \to \mathcal{D}$ is effectively a function $F\colon \mathrm{Ob}(\mathcal{C}) \cup \mathrm{Mor}(\mathcal{C}) \to \mathrm{Ob}(\mathcal{D}) \cup \mathrm{Mor}(\mathcal{D})$. Alternatively, one can choose a sufficiently large Grothendieck universe so that everything is small, in which case the size restrictions disappear.

***Remark* XV.1.4.8.** — Note that the natural isomorphisms between functors $\mathcal{C} \to \mathcal{D}$ are precisely the isomorphisms in $\mathrm{Fun}(\mathcal{C},\mathcal{D})$.

***Notation* XV.1.4.9.** — For a small category $\mathcal{C}$, we write $\mathrm{End}(\mathcal{C}) \coloneqq \mathrm{Fun}(\mathcal{C},\mathcal{C})$ for the **endofunctor category** of $\mathcal{C}$.

***Fact* XV.1.4.10.** — *Let $\mathcal{C}$ be a small category and let $\mathcal{D}$ be a locally small category. Then $\mathrm{Fun}(\mathcal{C},\mathcal{D})$ is locally small. If $\mathcal{D}$ is even small, then $\mathrm{Fun}(\mathcal{C},\mathcal{D})$ is also small.*

Proof. — Let $\mathcal{D}$ be locally small. Since a natural transformation $\alpha\colon F \to G$ is a family $(\alpha_X)_{X \in \mathrm{Ob}(\mathcal{C})}$, we have

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C},\mathcal{D})}(F,G) = \mathrm{Nat}(F,G) \subseteq \prod_{X \in \mathrm{Ob}(\mathcal{C})} \mathrm{Hom}_{\mathcal{D}}(F(X),G(X)).$$

Since $\mathrm{Ob}(\mathcal{C})$ and $\mathrm{Hom}_{\mathcal{D}}(F(X),G(X))$ for all $X \in \mathrm{Ob}(\mathcal{C})$ are sets, the product on the right-hand side is a set, and so is the left-hand side. Thus, $\mathrm{Fun}(\mathcal{C},\mathcal{D})$ is locally small.

Let $\mathcal{D}$ be small. Since $\mathrm{Ob}(\mathcal{D})$ is a set, $\mathrm{Mor}(\mathcal{D}) = \bigcup_{X,Y \in \mathrm{Ob}(\mathcal{D})} \mathrm{Hom}_{\mathcal{D}}(X,Y)$ is also a set. A functor $F\colon \mathcal{C} \to \mathcal{D}$ consists of two maps $\mathrm{Ob}(\mathcal{C}) \to \mathrm{Ob}(\mathcal{D})$ and $\mathrm{Mor}(\mathcal{C}) \to \mathrm{Mor}(\mathcal{D})$, so

$$\mathrm{Ob}(\mathrm{Fun}(\mathcal{C},\mathcal{D})) \subseteq \mathcal{P}(\mathrm{Ob}(\mathcal{C}) \times \mathrm{Ob}(\mathcal{D})) \times \mathcal{P}(\mathrm{Mor}(\mathcal{C}) \times \mathrm{Mor}(\mathcal{D})).$$

The right-hand side is a set, and so is the left-hand side. Thus, $\mathrm{Fun}(\mathcal{C},\mathcal{D})$ is small.

***Remark* XV.1.4.11.** — Natural transformations can be composed in two ways.

(i) There is the **horizontal composition**. Let $F_1, G_1\colon \mathcal{C} \to \mathcal{D}$ and $F_2, G_2\colon \mathcal{D} \to \mathcal{E}$ be functors. If $\alpha_1\colon F_1 \to G_1$ and $\alpha_2\colon F_2 \to G_2$ are natural transformations, then we obtain the natural composition $\alpha_2\alpha_1\colon F_2 \circ F_1 \to G_2 \circ G_1$:

$$\begin{tikzcd} \mathcal{C} \arrow[r, bend left, "F_1"] \arrow[r, bend right, "G_1"'] \arrow[r, phantom, "\Downarrow\alpha_1"] & \mathcal{D} \arrow[r, bend left, "F_2"] \arrow[r, bend right, "G_2"'] \arrow[r, phantom, "\Downarrow\alpha_2"] & \mathcal{E} \end{tikzcd} \qquad \rightsquigarrow \qquad \begin{tikzcd} \mathcal{C} \arrow[rr, bend left, "F_2 \circ F_1"] \arrow[rr, bend right, "G_2 \circ G_1"'] \arrow[rr, phantom, "\Downarrow\alpha_2\alpha_1"] & & \mathcal{E}. \end{tikzcd}$$

This defines a bifunctor

$$\mathrm{Fun}(\mathcal{C},\mathcal{D}) \times \mathrm{Fun}(\mathcal{D},\mathcal{E}) \longrightarrow \mathrm{Fun}(\mathcal{C},\mathcal{E}).$$

(ii) There is the **vertical composition**, which is the composition in functor categories. Let $F, G, H \colon \mathcal{C} \to \mathcal{D}$ be functors. If $\alpha \colon F \to G$ and $\beta \colon G \to H$ are natural transformations, then we obtain the natural composition $\beta \circ \alpha \colon F \to H$:

$$\begin{tikzcd} \mathcal{C} \arrow[rr, bend left=50, "F"] \arrow[rr, "G" description] \arrow[rr, bend right=50, "H"'] \arrow[rr, phantom, shift left=3, "\Downarrow\alpha"] \arrow[rr, phantom, shift right=3, "\Downarrow\beta"] & & \mathcal{D} & \rightsquigarrow & \mathcal{C} \arrow[r, bend left=40, "F"] \arrow[r, bend right=40, "H"'] \arrow[r, phantom, "\Downarrow\beta\circ\alpha"] & \mathcal{D}. \end{tikzcd}$$

Both horizontal and vertical compositions are compatible. The **middle four interchange law** states that both compositions commute. It is easier to just see a picture:

$$\begin{tikzcd} \mathcal{C} \arrow[r, bend left=50, "F_1"] \arrow[r, "G_1" description] \arrow[r, bend right=50, "H_1"'] \arrow[r, phantom, shift left=3, "\Downarrow\alpha_1"] \arrow[r, phantom, shift right=3, "\Downarrow\beta_1"] & \mathcal{D} \arrow[r, bend left=50, "F_2"] \arrow[r, "G_2" description] \arrow[r, bend right=50, "H_2"'] \arrow[r, phantom, shift left=3, "\Downarrow\alpha_2"] \arrow[r, phantom, shift right=3, "\Downarrow\beta_2"] \arrow[d, rightsquigarrow] & \mathcal{E} & \rightsquigarrow & \mathcal{C} \arrow[r, bend left=50, "F_2\circ F_1"] \arrow[r, "G_2\circ G_1" description] \arrow[r, bend right=50, "H_2\circ H_1"'] \arrow[r, phantom, shift left=3, "\Downarrow\alpha_2\alpha_1"] \arrow[r, phantom, shift right=3, "\Downarrow\beta_2\beta_1"] \arrow[d, rightsquigarrow] & \mathcal{E} \\ \mathcal{C} \arrow[r, bend left=40, "F_1"] \arrow[r, bend right=40, "H_1"'] \arrow[r, phantom, "\Downarrow\beta_1\circ\alpha_1"] & \mathcal{D} \arrow[r, bend left=40, "F_2"] \arrow[r, bend right=40, "H_2"'] \arrow[r, phantom, "\Downarrow\beta_2\circ\alpha_2"] & \mathcal{E} & \rightsquigarrow & \mathcal{C} \arrow[r, bend left=40, "F_2\circ F_1"] \arrow[r, bend right=40, "G_2\circ G_1"'] \arrow[r, phantom, "\Downarrow\gamma"] & \mathcal{E} \end{tikzcd}$$

where

$$\gamma = (\beta_2\beta_1) \circ (\alpha_2\alpha_1) = (\beta_2 \circ \alpha_2)(\beta_1 \circ \alpha_1).$$

That both compositions indeed yield natural transformations and that the middle four interchange law holds follow by horizontal and vertical pasting of naturality squares.

***Remark* XV.1.4.12.** — Let Cat be the **category of small categories**, whose morphisms are functors between small categories, i. e., $\mathrm{Hom}_{\mathsf{Cat}}(\mathcal{C}, \mathcal{D}) = \mathrm{Fun}(\mathcal{C}, \mathcal{D})$. This is a 2-category, which by definition enjoys the following properties:

(i) There are 0-morphisms (small categories), 1-morphisms (functors) and 2-morphisms (natural transformations).

(ii) Each hom-set $\mathrm{Fun}(\mathcal{C}, \mathcal{D})$ is itself a category under vertical composition.

(iii) If we take $\mathrm{Ob}(\mathsf{Cat})$ as objects and 2-cells

$$\begin{tikzcd} \mathcal{C} \arrow[r, bend left=40, "F"] \arrow[r, bend right=40, "G"'] \arrow[r, phantom, "\Downarrow\alpha"] & \mathcal{D} \end{tikzcd}$$

as morphisms, then we again obtain a category with horizontal composition.

(iv) The identity morphisms in the categories described in (ii) and (iii) are compatible, i. e., the horizontal composition of vertical identity morphisms is the expected vertical identity morphism.

(v) The middle four exchange law holds.

This leads to a rich theory of higher categories, which is out of our scope.

Maybe add higher category theory.

Definition* XV.1.4.13.** — *Let $\alpha \colon F \to G$ be a natural transformation of functors $\mathcal{C} \to \mathcal{D}$. If $H \colon \mathcal{D} \to \mathcal{E}$ is a functor, then* ***whiskering *α and H is the horizontal composition $H\alpha \coloneqq \mathrm{id}_H \circ \alpha$:*

$$\begin{tikzcd} \mathcal{C} \arrow[r, bend left=40, "F"] \arrow[r, bend right=40, "G"'] \arrow[r, phantom, "\Downarrow\alpha"] & \mathcal{D} \arrow[r, "H"] & \mathcal{E} & \rightsquigarrow & \mathcal{C} \arrow[r, bend left=40, "H\circ F"] \arrow[r, bend right=40, "H\circ G"'] \arrow[r, phantom, "\Downarrow H\alpha"] & \mathcal{E}. \end{tikzcd}$$

Similarly, if $H\colon \mathcal{E} \to \mathcal{C}$ *is a functor, then* ***whiskering*** H *and* α *is the horizontal composition* $\alpha H := \alpha \circ \mathrm{id}_H$*:*

$$\begin{tikzcd} \mathcal{E} \arrow[r, "H"] & \mathcal{C} \arrow[r, bend left, "F"] \arrow[r, bend right, "G"'] \arrow[r, phantom, "\Downarrow\alpha"] & \mathcal{D} \arrow[r, rightsquigarrow] & \mathcal{C} \arrow[r, bend left, "F\circ H"] \arrow[r, bend right, "G\circ H"'] \arrow[r, phantom, "\Downarrow\alpha H"] & \mathcal{E}. \end{tikzcd}$$

***Observation* XV.1.4.14.** — Interestingly enough, natural transformations can be described 1-categorically. Let $F, G\colon \mathcal{C} \rightrightarrows \mathcal{D}$ be two functors. The natural transformations $\mathrm{Nat}(F,G)$ correspond to bifunctors $H\colon \mathcal{C} \times 2 \to \mathcal{D}$ with the property that the diagram

$$\begin{tikzcd} \mathcal{C} \arrow[r, hook, "\iota_0"] \arrow[dr, "F"'] & \mathcal{C} \times 2 \arrow[d, "H"] & \mathcal{C} \arrow[l, hook', "\iota_1"'] \arrow[dl, "G"] \\ & \mathcal{D} & \end{tikzcd}$$

commutes. Here, ι_0 and ι_1 denote the obvious embedding functors $\mathcal{C} \hookrightarrow \mathcal{C} \times 2$.

Indeed, write $2 = \{0 \to 1\}$. Since H is a bifunctor, the diagram

$$\begin{tikzcd} H(X,0) \arrow[r, "H(f,0)"] \arrow[d] & H(Y,0) \arrow[d] \\ H(X,1) \arrow[r, "H(f,1)"] & H(Y,1) \end{tikzcd}$$

in $\mathcal{D}$ commutes for all morphisms $f\colon X \to Y$ in $\mathcal{C}$. Since $H(-,0) = F$ and $H(-,1) = G$, this square is the naturality square. Thus, the collection of all 'morphisms' $\alpha_X\colon H(X,0) \to H(X,1)$ in $\mathcal{D}$ for $X \in \mathrm{Ob}(\mathcal{C})$ defines a natural transformation α.

***Lemma* XV.1.4.15.** — *The endomorphisms* $\mathrm{End}(\mathrm{id}_{\mathcal{C}}) = \mathrm{Hom}_{\mathrm{End}(\mathcal{C})}(\mathrm{id}_{\mathcal{C}}, \mathrm{id}_{\mathcal{C}})$ *as a hom-set of* $\mathrm{End}(\mathcal{C})$ *form a commutative monoid. Hence, the automorphisms* $\mathrm{Aut}(\mathrm{id}_{\mathcal{C}})$ *form an abelian group.*

Proof. — The class $\mathrm{End}(\mathrm{id}_{\mathcal{C}})$ is obviously a monoid. Let $\alpha, \beta\colon \mathrm{id}_{\mathcal{C}} \to \mathrm{id}_{\mathcal{C}}$ be two natural transformations and we need to check commutativity pointwise. For any $X \in \mathrm{Ob}(\mathcal{C})$, applying the naturality square of α to the morphism $\beta_X\colon X \to X$ gives the commutative square

$$\begin{tikzcd} X \arrow[r, "\beta_X"] \arrow[d, "\alpha_X"'] & X \arrow[d, "\alpha_X"] \\ X \arrow[r, "\beta_X"] & X. \end{tikzcd}$$

$\square$

***Lemma* XV.1.4.16.** — *Let* $\phi\colon \mathcal{C} \to \mathcal{D}$ *be a functor and let* I *be a small category. If* ϕ *is faithful (resp. fully faithful), then so is the functor*

$$\phi \circ -\colon\ \mathrm{Fun}(I, \mathcal{C}) \longrightarrow \mathrm{Fun}(I, \mathcal{D}), \qquad F \longmapsto \phi \circ F.$$

Proof. — Fix two functors $F, G\colon I \to \mathcal{C}$. For natural transformations $\alpha, \beta\colon \phi \circ F \to \phi \circ G$, suppose that $\phi \circ \alpha = \phi \circ \beta$. Then for each $X \in \mathrm{Ob}(I)$, we have $\phi(\alpha_X) = \phi(\beta_X)$, so $\alpha_X = \beta_X$ by faithfulness of ϕ. Thus, $\alpha = \beta$.

Now, suppose that ϕ is additionally full, and let $\alpha\colon \phi \circ F \to \phi \circ G$ be a natural transformation. Then for each $X \in \mathrm{Ob}(I)$, there is a unique morphism $\beta_X\colon F(X) \to G(X)$ such that $\phi(\beta_X) = \alpha_X$. These β_X define a natural transformation $\beta\colon F \to G$: for any morphism $f\colon X \to Y$ in I, the naturality of $\alpha_X = \phi(\beta_X)$ gives the commutative square

$$\begin{tikzcd} \phi(F(X)) \arrow[r, "\phi(F(f))"] \arrow[d, "\phi(\beta_X)"'] & \phi(F(Y)) \arrow[d, "\phi(\beta_Y)"] \\ \phi(G(X)) \arrow[r, "\phi(G(f))"] & \phi(G(Y)). \end{tikzcd}$$

Since ϕ is faithful, we obtain the commutative square

$$\begin{tikzcd} F(X) \arrow[r, "F(f)"] \arrow[d, "\beta_X"'] & F(Y) \arrow[d, "\beta_Y"] \\ G(X) \arrow[r, "G(f)"] & G(Y). \end{tikzcd}$$

XV.1.5. Equivalence of categories.

It is obvious how to define an isomorphism of categories.

***Motivation* XV.1.5.1.** — A functor $F\colon \mathcal{C} \to \mathcal{D}$ is an **isomorphism of categories** if there exists a functor $G\colon \mathcal{D} \to \mathcal{C}$ such that $G \circ F = \mathrm{id}_{\mathcal{C}}$ and $F \circ G = \mathrm{id}_{\mathcal{D}}$. In this case, we say that $\mathcal{C}$ and $\mathcal{D}$ are *isomorphic* and we write $\mathcal{C} \cong \mathcal{D}$.

In practice, however, this situation is too restrictive and virtually never appears: isomorphic categories have the same datum from a categorical view and cannot be distinguished.

***Example* XV.1.5.2.** — For a small category $\mathcal{C}$ and a category $\mathcal{D}$, we have a true isomorphism of categories

$$\mathrm{Fun}(\mathcal{C}, \mathcal{D})^{\mathrm{op}} \xrightarrow{\sim} \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathcal{D}^{\mathrm{op}}), \qquad F \longmapsto \mathrm{op} \circ F \circ \mathrm{op}$$

where the functor op swaps composition of morphisms in $\mathcal{C}$ and $\mathcal{D}^{\mathrm{op}}$, resp. (cf. remark XV.1.3.4).

***Example* XV.1.5.3.** — For small categories $\mathcal{C}$ and $\mathcal{D}$ and a category $\mathcal{E}$, we have a true isomorphism of categories

$$\mathrm{Fun}(\mathcal{C} \times \mathcal{D}, \mathcal{E}) \xrightarrow{\sim} \mathrm{Fun}(\mathcal{C}, \mathrm{Fun}(\mathcal{D}, \mathcal{E})), \qquad F \longmapsto (X \longmapsto F(X, -))$$

and similarly $\mathrm{Fun}(\mathcal{C} \times \mathcal{D}, \mathcal{E}) \xrightarrow{\sim} \mathrm{Fun}(\mathcal{D}, \mathrm{Fun}(\mathcal{C}, \mathcal{E}))$. This is sometimes called *currying*, and the inverse is called *uncurrying*. An important example is currying of the hom-functor.

***Convention* XV.1.5.4.** — *Most results in this subsection require the axiom of choice. However, we will explicitly say this again in each statement.*

We introduce a weaker, but much more useful, notion.

***Definition* XV.1.5.5.** — *A functor $F\colon \mathcal{C} \to \mathcal{D}$ is an **equivalence of categories** if there exists a functor $G\colon \mathcal{D} \to \mathcal{C}$ such that $\eta\colon \mathrm{id}_{\mathcal{C}} \xrightarrow{\sim} G \circ F$ and $\varepsilon\colon F \circ G \xrightarrow{\sim} \mathrm{id}_{\mathcal{D}}$ are natural isomorphisms. In this case, we say that G is the **quasi-inverse** of F, and vice versa. Moreover, we say that $\mathcal{C}$ and $\mathcal{D}$ are **equivalent** and write $\mathcal{C} \simeq \mathcal{D}$.*

***Fact* XV.1.5.6.** — *Equivalences of categories form an equivalence relation.*

***Lemma* XV.1.5.7** (universal property of subcategories)**.** — *Assume the axiom of choice. Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor and let $\iota\colon \mathcal{D}' \hookrightarrow \mathcal{D}$ be the embedding of a full subcategory $\mathcal{D}'$. Suppose that F is essentially surjective onto $\mathcal{D}'$, i. e., for each $X \in \mathrm{Ob}(\mathcal{C})$ there exists $Y \in \mathrm{Ob}(\mathcal{D}')$ and an isomorphism $F(X) \cong Y$. Then there exists a functor $F'\colon \mathcal{C} \to \mathcal{D}'$ such that there is a natural isomorphism $\alpha\colon F \xrightarrow{\sim} \iota \circ F'$, and F' is unique with this property up to unique isomorphism. More precisely, if $\beta\colon F \xrightarrow{\sim} \iota \circ G'$ is a natural isomorphism, then there exists a unique natural isomorphism $\gamma\colon G' \xrightarrow{\sim} F'$ such that $\alpha = (\iota\gamma) \circ \beta$.*

Proof. — By the axiom of choice, choose for each $X \in \mathrm{Ob}(\mathcal{C})$ some $Y \in \mathrm{Ob}(\mathcal{D}')$ and an isomorphism $\phi_X\colon Y \xrightarrow{\sim} F(X)$, and define $F'(X) := Y$. On morphisms $f\colon X \to X'$ in $\mathcal{C}$, define

$$F'(f)\colon F'(X) \underset{\sim}{\xrightarrow{\phi_X}} F(X) \xrightarrow{F(f)} F(X') \underset{\sim}{\xrightarrow{\phi_{X'}^{-1}}} F'(X').$$

This defines a functor $F'\colon \mathcal{C} \to \mathcal{D}'$ since F is functorial. The natural isomorphism $\alpha\colon F \xrightarrow{\sim} \iota \circ F'$ is given by $\alpha_X := \phi_X^{-1}\colon F(X) \xrightarrow{\sim} Y = \iota(F'(X))$. The uniqueness is easily seen since $\alpha = (\iota\gamma) \circ \beta$ means that $\alpha_X = \gamma_X \circ \beta_X$ for each $X \in \mathrm{Ob}(\mathcal{C})$, and if β_X is an isomorphism, then $\gamma_X = \alpha_X \circ \beta_X^{-1}$ necessarily. □

***Definition* XV.1.5.8.** — *A category $\mathcal{C}$ is **skeletal** if isomorphic objects in $\mathcal{C}$ are already equal. The **skeleton** of $\mathcal{C}$ is a subcategory $\mathrm{sk}(\mathcal{C})$ such that the embedding $\mathrm{sk}(\mathcal{C}) \hookrightarrow \mathcal{C}$ is an equivalence of categories.*

***Remark* XV.1.5.9.** — Observe that if η is a natural isomorphism of functors whose target is a skeletal category, then η is the identity. Hence, an equivalence of skeletal categories is an isomorphism of categories. This shows that the skeleton of a category, if it exists, is unique up to isomorphism.

***Lemma* XV.1.5.10.** — *Every category has a skeleton.*

Proof. — Let $\mathcal{C}$ be a category. By the axiom of choice, pick an object from each isomorphism class of $\mathrm{Ob}(\mathcal{C})$. This generates a full subcategory $\mathrm{sk}(\mathcal{C})$ of $\mathcal{C}$, and no two distinct objects in $\mathrm{sk}(\mathcal{C})$ are isomorphic. If $\iota\colon \mathrm{sk}(\mathcal{C}) \hookrightarrow \mathcal{C}$ is the embedding functor, applying lemma XV.1.5.7 to $\mathrm{id}_{\mathcal{C}}$ shows that there exists a functor $F'\colon \mathcal{C} \to \mathrm{sk}(\mathcal{C})$ such that $\iota \circ F' \cong \mathrm{id}_{\mathcal{C}}$. Conversely, observe that $\iota \circ (F' \circ \iota) = (\iota \circ F') \circ \iota \cong \mathrm{id}_{\mathcal{C}} \circ \iota = \iota = \iota \circ \mathrm{id}_{\mathrm{sk}(\mathcal{C})}$. Since ι is fully faithful and hence conservative (lemma XV.1.3.8), we thus have $F' \circ \iota \cong \mathrm{id}_{\mathrm{sk}(\mathcal{C})}$. □

***Proposition* XV.1.5.11** (characterisation of equivalences of categories)**.** — *Assuming the axiom of choice, a functor is an equivalence of categories if and only if it is fully faithful and essentially surjective.*

Proof. — Let $F\colon \mathcal{C} \to \mathcal{D}$ be an equivalence of categories with quasi-inverse $G\colon \mathcal{D} \to \mathcal{C}$. Let $\eta\colon \mathrm{id}_{\mathcal{C}} \xrightarrow{\sim} G \circ F$ and $\varepsilon\colon F \circ G \xrightarrow{\sim} \mathrm{id}_{\mathcal{D}}$ be natural isomorphisms. Consider

$$\mathrm{Hom}_{\mathcal{C}}(X, Y) \xrightarrow{F} \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y)) \xrightarrow{G} \mathrm{Hom}_{\mathcal{C}}(G(F(X)), G(F(X))),$$

which is $f \mapsto \alpha_Y^{-1} \circ f \circ \alpha_X$ by naturality of α. In particular, this map is bijective, and hence F is faithful. The same argument with β shows that G is faithful. For fullness of F, let $g \in \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$ and consider $f = \alpha_Y \circ G(g) \circ \alpha_X^{-1} \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$. Then $G(F(f)) = G(g)$, so $F(f) = g$ since G is faithful. Thus, F is full. For essential surjectivity, given any $Y \in \mathrm{Ob}(\mathcal{D})$, we have an isomorphism $\beta_Y\colon F(G(Y)) \xrightarrow{\sim} Y$.

For the converse, let $F\colon \mathcal{C} \to \mathcal{D}$ be a fully faithful and essentially surjective functor. By lemma XV.1.5.10, consider the skeleton $\mathrm{sk}(\mathcal{C})$ with the equivalence of categories $\iota_{\mathcal{C}}\colon \mathrm{sk}(\mathcal{C}) \hookrightarrow \mathcal{C}$, and let $\kappa_{\mathcal{C}}$ be the quasi-inverse of $\iota_{\mathcal{C}}$. Similarly, define $\mathrm{sk}(\mathcal{D})$, $\iota_{\mathcal{D}}$ and $\kappa_{\mathcal{D}}$. Consider the functor $H := \kappa_{\mathcal{D}} \circ F \circ \iota_{\mathcal{C}}\colon \mathrm{sk}(\mathcal{C}) \to \mathrm{sk}(\mathcal{D})$, which is again fully faithful and essentially surjective by the previous part. We want to show that this is an isomorphism of categories. For $Y \in \mathrm{Ob}(\mathrm{sk}(\mathcal{D}))$, there exists $X \in \mathrm{Ob}(\mathrm{sk}(\mathcal{C}))$ such that $H(X) \cong Y$, so $H(X) = Y$ in $\mathrm{sk}(\mathcal{D})$. If $H(X) = H(X')$ for $X, X' \in \mathrm{Ob}(\mathrm{sk}(\mathcal{C}))$, then $X \cong X'$ by lemma XV.1.3.8 and hence $X = X'$ in $\mathrm{sk}(\mathcal{C})$. Thus, H is bijective on objects, and fully faithfulness implies that H is bijective on morphisms, as desired. Now, set $G = \iota_{\mathcal{C}} \circ H^{-1} \circ \kappa_{\mathcal{D}}$, which is clearly a quasi-inverse of F. □

***Definition* XV.1.5.12.** — *The **essential image** of a functor $F\colon \mathcal{C} \to \mathcal{D}$ is the smallest saturated subcategory of $\mathcal{D}$ containing $F(\mathrm{Ob}(\mathcal{C}))$.*

***Corollary* XV.1.5.13.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a fully faithful functor, and let $\mathcal{D}'$ be the essential image of F. Then F induces an equivalence of categories $F'\colon \mathcal{C} \xrightarrow{\sim} \mathcal{D}'$, and if $\iota\colon \mathcal{D}' \hookrightarrow \mathcal{D}$ is the embedding, then $F \cong \iota \circ F'$.*

Proof. — The induced F' is fully faithful like F, but now even essentially surjective. Thus, F' is an equivalence of categories, and we clearly have $F \cong \iota \circ F'$ (in fact, $F_X = (\iota \circ F')_X$ for all $X \in \mathrm{Ob}(\mathcal{C})$). ◻

***Counterexample* XV.1.5.14.** — Corollary XV.1.5.13 may fail when F is not full, i. e., $F\colon \mathcal{C} \to \mathcal{D}$ may not factor through a full subcategory of $\mathcal{D}$ that is equivalent to $\mathcal{C}$.

Let $\mathcal{C}$ be the category consisting of a single object c and a single idempotent morphism $c \to c$ other than id_c. Consider the obvious functor $F\colon \mathsf{2} \to \mathcal{C}$ mapping the non-trivial morphism of $\mathsf{2}$ to $c \to c$. Since every hom-set of $\mathsf{2}$ is a singleton, F is faithful. The only full subcategories of $\mathcal{C}$ are $\mathsf{0}$ and $\mathcal{D}$ itself, neither of them being equivalent to $\mathsf{2}$.

***Example* XV.1.5.15.** — For k a field, let $\mathcal{C}$ be the category with $\mathrm{Ob}(\mathcal{C}) = \mathbb{N}$ and $\mathrm{Hom}_{\mathcal{C}}(n,m) = \mathrm{Mat}_{m\times n}(k)$. This category is a skeleton of the category of finite-dimensional k-vector spaces. The embedding ι is given by $\iota(n) = k^n$ for $n \in \mathbb{N}$ and $\iota(A) = A$ for $A \in \mathrm{Mat}_{m\times n}(k)$, and ι is an equivalence of categories.

***Example* XV.1.5.16.** — Let $(\mathcal{C}_i)_{i\in I}$ be a family of categories indexed by a set I. If $I = \varnothing$, then we have equivalences $\prod_{i\in\varnothing} \mathcal{C}_i \simeq \mathsf{1}$ and $\coprod_{i\in\varnothing} \mathcal{C}_i \simeq \mathsf{0}$.

Definition* XV.1.5.17.** — *A category is* ***essentially small *if it is equivalent to a small category.*

***Remark* XV.1.5.18.** — A category $\mathcal{C}$ is essentially small if and only if $\mathcal{C}$ is locally small and there is a subset $S \subseteq \mathrm{Ob}(\mathcal{C})$ such that any object of $\mathcal{C}$ is isomorphic to an object in S. In other words, $\mathcal{C}$ is essentially small if and only if $\mathrm{sk}(\mathcal{C})$ is small.

Skip definition of half-full functors, Definition 1.3.17, Proposition 1.3.18.

XV.1.6. Yoneda lemma and representable functors.

Definition* XV.1.6.1.** — *Let $\mathcal{C}$ be a small category. The* ***Yoneda embedding *of $\mathcal{C}$ is the functor*

$$\mathrm{h}_{\mathcal{C}}\colon \mathcal{C} \longrightarrow \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}), \qquad X \longmapsto \mathrm{Hom}_{\mathcal{C}}(-,X).$$

There is also the contravariant Yoneda embedding

$$\mathrm{h}^{\mathcal{C}}\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathrm{Fun}(\mathcal{C}, \mathsf{Set}), \qquad X \longmapsto \mathrm{Hom}_{\mathcal{C}}(X,-).$$

Both are obtained from the hom-functor $\mathrm{Hom}_{\mathcal{C}}\colon \mathcal{C}^{\mathrm{op}} \times \mathcal{C} \to \mathsf{Set}$.

The name 'embedding' will be justified in corollary XV.1.6.7.

***Remark* XV.1.6.2.** — Note that $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ and $\mathrm{Fun}(\mathcal{C}, \mathsf{Set})$ are not small, but locally small.

***Remark* XV.1.6.3.** — Since $\mathrm{Hom}_{\mathcal{C}}(X,-) = \mathrm{Hom}_{\mathcal{C}^{\mathrm{op}}}(-,X^{\mathrm{op}})$ where $X = X^{\mathrm{op}}$ in $\mathcal{C}^{\mathrm{op}}$, we see that $\mathrm{h}^{\mathcal{C}}$ is just the usual Yoneda embedding

$$\mathrm{h}^{\mathcal{C}} = \mathrm{h}_{\mathcal{C}^{\mathrm{op}}}\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathrm{Fun}((\mathcal{C}^{\mathrm{op}})^{\mathrm{op}}, \mathsf{Set}), \qquad X^{\mathrm{op}} \longmapsto \mathrm{Hom}_{\mathcal{C}^{\mathrm{op}}}(-,X^{\mathrm{op}}).$$

Thus, any result about the contravariant Yoneda embedding is completely dual.

***Notation* XV.1.6.4.** — We will also write $\mathrm{h}_{\mathcal{C}}(X) = \mathrm{Hom}_{\mathcal{C}}(-,X)$ and $\mathrm{h}^{\mathcal{C}}(X) = \mathrm{Hom}_{\mathcal{C}}(X,-)$ if $\mathcal{C}$ is locally small.

***Yoneda lemma* XV.1.6.5.** — *Let $\mathcal{C}$ be a small category.*

(i) *For any functor $F\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ and any $X \in \mathrm{Ob}(\mathcal{C})$, we have an isomorphism of sets*

$$\mathrm{Nat}(\mathrm{h}_{\mathcal{C}}(X), F) = \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(\mathrm{h}_{\mathcal{C}}(X), F) \xrightarrow{\sim} F(X), \qquad \eta \longmapsto \eta_X(\mathrm{id}_X).$$

This isomorphism is natural in F and X, i. e., these isomorphisms define a natural isomorphism of functors $\mathcal{C}^{\mathrm{op}} \times \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \to \mathsf{Set}$.

(ii) *Dually, for any functor* $F\colon \mathcal{C} \to \mathsf{Set}$ *and any* $X \in \mathrm{Ob}(\mathcal{C})$*, we have an isomorphism of sets*

$$\mathrm{Nat}(\mathrm{h}^{\mathcal{C}}(X), F) = \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C},\mathsf{Set})}(\mathrm{h}^{\mathcal{C}}(X), F) \xrightarrow{\sim} F(X), \qquad \eta \longmapsto \eta_X(\mathrm{id}_X)$$

which is natural in F *and* X*, i. e., it defines a natural isomorphism of functors* $\mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}} \times \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$.

Proof. — It is not surprising that this is an isomorphism: the value $\eta_X(\mathrm{id}_X)$ completely characterises η since for any morphism $f\colon Y \to X$ in $\mathcal{C}$, naturality of η shows that

$$\begin{tikzcd} \mathrm{id}_X \arrow[r, mapsto, "-\circ f"] \arrow[d, mapsto, "\eta_X"'] & f \arrow[d, mapsto, "\eta_Y"] \\ \eta_X(\mathrm{id}_X) \arrow[r, mapsto, "F(f)"] & \eta_Y(f), \end{tikzcd} \tag{XV.1.6.5.1}$$

so $\eta_Y = F(-)(\eta_X(\mathrm{id}_X))$.

We need to exhibit an inverse $\phi\colon F(X) \to \mathrm{Nat}(\mathrm{h}_{\mathcal{C}}(X), F)$. For $s \in F(X)$, a natural transformation $\phi(s)$ consists of maps $\phi(s)_Y\colon \mathrm{Hom}_{\mathcal{C}}(Y, X) \to F(Y)$ with $Y \in \mathrm{Ob}(\mathcal{C})$, which we define by

$$\phi(s)_Y\colon \mathrm{Hom}_{\mathcal{C}}(Y, X) \longrightarrow \mathrm{Hom}_{\mathsf{Set}}(F(X), F(Y)) \longrightarrow F(Y), \qquad f \longmapsto F(f) \longmapsto F(f)(s).$$

Then $\phi(s)$ is indeed a natural transformation since for any morphism $g\colon Z \to Y$ in $\mathcal{C}^{\mathrm{op}}$, the diagram

$$\begin{tikzcd} \mathrm{Hom}_{\mathcal{C}}(Y, X) \arrow[r, "-\circ g"] \arrow[d, "\phi(s)_Y"'] & \mathrm{Hom}_{\mathcal{C}}(Z, X) \arrow[d, "\phi(s)_Z"] \\ F(Y) \arrow[r, "F(g)"] & F(Z) \end{tikzcd}$$

commutes. One easily checks that ϕ is indeed the inverse.

For naturality, it suffices to check naturality in F and in X separately. If $\alpha\colon F \to G$ is a natural transformation, then the diagram

$$\begin{tikzcd} \mathrm{Nat}(\mathrm{h}_{\mathcal{C}}(X), F) \arrow[r, "\alpha\circ-"] \arrow[d, "\wr"'] & \mathrm{Nat}(\mathrm{h}_{\mathcal{C}}(X), G) \arrow[d, "\wr"] \\ F(X) \arrow[r, "\alpha_X"] & G(X), \end{tikzcd} \qquad \begin{tikzcd} \eta \arrow[r] \arrow[d] & \alpha \circ \eta \arrow[d] \\ \eta_X(\mathrm{id}_X) \arrow[r] & (\alpha_X \circ \eta_X)(\mathrm{id}_X) \end{tikzcd}$$

commutes. If $f\colon X \to Y$ is a morphism in $\mathcal{C}$, then the diagram

$$\begin{tikzcd} \mathrm{Nat}(\mathrm{h}_{\mathcal{C}}(X), F) \arrow[r, "-\circ \mathrm{h}_{\mathcal{C}}(f)"] \arrow[d, "\wr"'] & \mathrm{Nat}(\mathrm{h}_{\mathcal{C}}(Y), F) \arrow[d, "\wr"] \\ F(X) \arrow[r, "F(f)"] & F(Y), \end{tikzcd} \qquad \begin{tikzcd} \eta \arrow[r] \arrow[d] & \eta \circ \mathrm{h}_{\mathcal{C}}(f) \arrow[d] \\ \eta_X(\mathrm{id}_X) \arrow[r] & F(f)(\eta_X(\mathrm{id}_X)) \end{tikzcd}$$

commutes since $(\mathrm{h}_c(f))_Y = f \circ -$ and by diagram (XV.1.6.5.1) (naturality of η), we have

$$F(f)(\eta_X(\mathrm{id}_X)) = \eta_Y(f) = \eta_Y(f \circ \mathrm{id}_Y) = (\eta \circ \mathrm{h}_{\mathcal{C}}(f))_Y(\mathrm{id}_Y).$$

***Remark* XV.1.6.6.** — If we do not insist on a bijection of classes but instead spell out the correspondence explicitly ('Any natural transformation $\eta\colon \mathrm{Hom}_{\mathcal{C}}(-, X) \to F$ is uniquely determined by $\eta_X(\mathrm{id}_X) \in F(X)$, and any $s \in F(X)$ defines a unique natural transformation η such that $\eta_X(\mathrm{id}_X) = s$.'), then the Yoneda lemma XV.1.6.5 also holds for locally small categories $\mathcal{C}$ as the proof does not require $\mathcal{C}$ to be small at all. This is also true for naturality if we state naturality elementwise. Alternatively, the Yoneda lemma XV.1.6.5 always holds if we choose a sufficiently large Grothendieck universe.

***Corollary* XV.1.6.7.** — *For $\mathcal{C}$ a small category, the Yoneda embeddings $\mathrm{h}_{\mathcal{C}}$ and $\mathrm{h}^{\mathcal{C}}$ are fully faithful (justifying the name 'embedding').*

Proof. — Taking $F = \mathrm{h}_{\mathcal{C}}(Y)$, we obtain

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(\mathrm{h}_{\mathcal{C}}(X), \mathrm{h}_{\mathcal{C}}(Y)) \cong \mathrm{h}_{\mathcal{C}}(Y)(X) = \mathrm{Hom}_{\mathcal{C}}(X, Y).$$

Skip Corollary 1.4.6 on smallness of special case of comma categories.

***Corollary* XV.1.6.8.** — *Let $f\colon X \to Y$ be a morphism in a small category $\mathcal{C}$. If the hom-functor $f \circ -\colon \mathrm{Hom}_{\mathcal{C}}(-, X) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(-, Y)$ (resp. $- \circ f\colon \mathrm{Hom}_{\mathcal{C}}(Y, -) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(X, -)$) is a natural isomorphism, then f is an isomorphism.*

Proof. — By hypothesis, $\mathrm{h}_{\mathcal{C}}(f)\colon \mathrm{h}_{\mathcal{C}}(X) \to \mathrm{h}_{\mathcal{C}}(Y)$ is a (natural) isomorphism. Since $\mathrm{h}_{\mathcal{C}}$ is fully faithful (corollary XV.1.6.7), we see that f is an isomorphism.

It is a good idea to generalise corollary XV.1.6.8 to arbitrarily sized categories.

***Lemma* XV.1.6.9.** — *Let X and Y be objects in $\mathcal{C}$ with a bijection $\phi_W\colon \mathrm{Hom}_{\mathcal{C}}(W, X) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(W, Y)$ for each $W \in \mathrm{Ob}(\mathcal{C})$ that is natural in W, i. e., the corresponding naturality square commutes. Then there exists a unique isomorphism $f\colon X \xrightarrow{\sim} Y$ such that $\phi_W = f \circ -$, namely $f = \phi_X(\mathrm{id}_X)$.*

The dual statement holds for bijections $\psi_W\colon \mathrm{Hom}_{\mathcal{C}}(Y, W) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(X, W)$, giving the isomorphism $\psi_Y(\mathrm{id}_Y)\colon X \xrightarrow{\sim} Y$.

Proof. — Let $f = \phi_X(\mathrm{id}_X)\colon X \to Y$ and $g = \phi_Y^{-1}(\mathrm{id}_Y)\colon Y \to X$. By naturality, the square

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{C}}(X, X) \arrow[r, "\phi_X", "\sim"'] \arrow[d, "-\circ g"'] & \mathrm{Hom}_{\mathcal{C}}(X, Y) \arrow[d, "-\circ g"] & \mathrm{id}_X \arrow[r, mapsto] \arrow[d, mapsto] & f \arrow[d, mapsto] \\
\mathrm{Hom}_{\mathcal{C}}(Y, X) \arrow[r, "\phi_Y", "\sim"'] & \mathrm{Hom}_{\mathcal{C}}(Y, Y), & g \arrow[r, mapsto] & f \circ g = \mathrm{id}_Y
\end{tikzcd}$$

commutes, showing that $f \circ g = \mathrm{id}_Y$ by diagram chasing. A similar argument with ϕ_W^{-1} shows that $g \circ f = \mathrm{id}_X$, whence f is an isomorphism. If $\phi_W = f' \circ -$ for some other morphism $f'\colon X \to Y$, then $f' = \phi_X(\mathrm{id}_X) = f$.

The Yoneda lemma can be seen as an ultra abstract generalisation of Cayley's theorem from group theory.

***Cayley's theorem* XV.1.6.10.** — *Any group is isomorphic to a subgroup of a permutation group.*

Proof. — Let G be a group, and let $BG = \{\mathrm{pt}\}$ be the groupoid with a single object and G as morphisms.(XV.5) Observe that functors $F\colon BG^{\mathrm{op}} \to \mathsf{Set}$ correspond to right G-sets $F(\mathrm{pt})$, and natural transformations $\alpha\colon F \to G$ of functors $BG^{\mathrm{op}} \to \mathsf{Set}$ correspond to right G-equivariant maps $\alpha_{\mathrm{pt}}\colon F(\mathrm{pt}) \to G(\mathrm{pt})$. Hence, $\mathrm{h}_{BG}(\mathrm{pt}) = \mathrm{Hom}_{BG}(-, \mathrm{pt})$ is precisely G itself as a right G-set, fully faithfulness of the Yoneda embedding (corollary XV.1.6.7) gives the bijection

$$\begin{aligned} G = \mathrm{Hom}_{BG}(\mathrm{pt}, \mathrm{pt}) &\xrightarrow{\sim} \mathrm{Hom}_{\mathrm{Fun}(BG^{\mathrm{op}},\mathsf{Set})}(\mathrm{h}_{BG}(\mathrm{pt}), \mathrm{h}_{BG}(\mathrm{pt})) = \mathrm{End}_G(G), \\ g &\longmapsto (x \longmapsto gx). \end{aligned}$$

In fact, this bijection is a group homomorphism, and $\mathrm{End}_G(G) = \mathrm{Aut}_G(G)$. Thus, we have an injective group homomorphism $G \xrightarrow{\sim} \mathrm{Aut}_G(G) \hookrightarrow \mathrm{Sym}(G)$ where $\mathrm{Sym}(G)$ is the permutation group of the underlying set G.

(XV.5) The notation BG is from homotopy theory and denotes the classifying space of a topological group. In our case, G is discrete.

Definition* XV.1.6.11.** — *Let $\mathcal{C}$ be a locally small category. A functor $F\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ (resp. $F\colon \mathcal{C} \to \mathsf{Set}$) is* ***representable *if there is a natural isomorphism $\eta\colon \mathrm{Hom}_{\mathcal{C}}(-,X) \xrightarrow{\sim} F$ (resp. $\eta\colon \mathrm{Hom}_{\mathcal{C}}(X,-) \xrightarrow{\sim} F$) for an object X in $\mathcal{C}$. In this case, η is called a* ***representation*** *of F and is uniquely determined by an element $s \in F(X)$ by the Yoneda lemma XV.1.6.5. We call (X,s) the* ***universal element*** *for F and call X a* ***representative*** *of F.*

***Remark* XV.1.6.12.** — Suppose that $\mathcal{C}$ is small. If X and X' are two representatives of F with natural isomorphisms $\mathrm{h}_{\mathcal{C}}(X) \cong F \cong \mathrm{h}_{\mathcal{C}}(X')$, then there exists a unique isomorphism $X \cong X'$ by corollary XV.1.6.7 and lemma XV.1.3.8. Thus, the pair $(X, \mathrm{h}_{\mathcal{C}}(X) \cong F)$ of representative and representation is unique up to unique isomorphism. We can generalise to locally small $\mathcal{C}$ by virtue of lemma XV.1.6.9.

***Lemma* XV.1.6.13** (characterisation of representability)**.** — *Let $\mathcal{C}$ be a locally small category. A functor $F\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ is representable if and only if there exist $X \in \mathrm{Ob}(\mathcal{C})$ and $s \in F(X)$ such that $\mathrm{Hom}_{\mathcal{C}}(Y,X) \xrightarrow{\sim} F(Y)$, $f \mapsto F(f)(s)$ is bijective for all $Y \in \mathrm{Ob}(\mathcal{C})$.*

Proof. — Any pair (X,s) with $X \in \mathrm{Ob}(\mathcal{C})$ and $s \in F(X)$ defines a unique natural transformation $\eta\colon \mathrm{h}_{\mathcal{C}}(X) \to F$ by the Yoneda lemma XV.1.6.5 (remark XV.1.6.6). By the proof of the Yoneda lemma XV.1.6.5, for each $Y \in \mathrm{Ob}(\mathcal{C})$ the map $\eta_Y\colon \mathrm{Hom}_{\mathcal{C}}(Y,X) \to F(Y)$ is of the form $f \mapsto F(f)(s)$. Hence, η is a natural isomorphism if and only if all η_Y are bijective.

***Definition* XV.1.6.14.** — *Let $\{\mathrm{pt}\}\colon \mathsf{1} \to \mathsf{Set}$ is the inclusion of the singleton set.*

(i) *Let $F\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ be a functor. The* ***category of elements*** $\mathrm{el}(F)$ *of F is the comma category $\{\mathrm{pt}\} \downarrow F$. Explicitly, it consists of the following:*

- *the objects are pairs (X,s) with $X \in \mathrm{Ob}(\mathcal{C})$ and $s \in F(X)$;*
- *the morphisms $(X,s) \to (Y,t)$ are morphisms $f\colon X \to Y$ in $\mathcal{C}$ such that $F(f)(t) = s$.*

(ii) *Dually, let $F\colon \mathcal{C} \to \mathsf{Set}$ be a functor. The* ***category of elements*** $\mathrm{el}(F)$ *of F is the comma category $F \downarrow \{\mathrm{pt}\}$. Explicitly, it consists of the following:*

- *the objects are pairs (X,s) with $X \in \mathrm{Ob}(\mathcal{C})$ and $s \in F(X)$;*
- *the morphisms $(X,s) \to (Y,t)$ are morphisms $f\colon X \to Y$ in $\mathcal{C}$ such that $F(f)(s) = t$ (we are covariant now).*

Both categories come with the faithful forgetful functor $\mathrm{For}\colon \mathrm{el}(F) \to \mathcal{C}$.

***Example* XV.1.6.15.** — Over and under categories are special cases of categories of elements. For an object A of $\mathcal{C}$, we have an isomorphism of categories

$$\mathrm{el}(\mathrm{h}_{\mathcal{C}}(A)) \xrightarrow{\sim} \mathcal{C}/A, \qquad (X, s\colon X \longrightarrow A) \longmapsto s.$$

Indeed, a morphism $f\colon (X,s) \to (Y,t)$ in $\mathrm{el}(\mathrm{h}_{\mathcal{C}}(A))$ satisfies $\mathrm{Hom}_{\mathcal{C}}(f,A)(t) = t \circ f = s$ and is mapped to the commutative triangle

$$\begin{tikzcd} X \arrow[rr, "f"] \arrow[dr, "s"'] & & Y \arrow[dl, "t"] \\ & A. & \end{tikzcd}$$

Similarly, we have an isomorphism of categories $\mathrm{el}(\mathrm{h}^{\mathcal{C}}(A)) \xrightarrow{\sim} A/\mathcal{C}$.

***Proposition* XV.1.6.16.** — *Let $\mathcal{C}$ be a locally small category. A functor $F\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ (resp. $F\colon \mathcal{C} \to \mathsf{Set}$) is representable if and only if $\mathrm{el}(F)$ has a terminal (resp. initial) object. In particular, the universal element for F is the terminal (resp. initial) object in $\mathrm{el}(F)$.*

Proof. — An object (X,s) in $\mathrm{el}(F)$ is terminal if and only if there exists exactly one morphism $(Y,t) \to (X,s)$ for each object (Y,t) in $\mathrm{el}(F)$ (with $Y \in \mathrm{Ob}(\mathcal{C})$ and $t \in F(Y)$). Such a morphism $(Y,t) \to (X,s)$ is uniquely described by a morphism $f\colon Y \to X$ in $\mathcal{C}$ such that $F(f)(s) = t$. So, this condition is equivalent to saying that $\mathrm{Hom}_{\mathcal{C}}(Y,X) \to F(X)$, $f \mapsto F(f)(s)$ is bijective for any $Y \in \mathrm{Ob}(\mathcal{C})$. This is equivalent to F being representable by lemma XV.1.6.13. ◻

***Example* XV.1.6.17.** — Representable functors generalise the universal property of initial and terminal objects. Let $\mathcal{C}$ be a locally small category. A terminal object X of $\mathcal{C}$ is characterised as a representative of the constant functor $\Delta_{\{\mathrm{pt}\}}\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$. Indeed, if $\mathrm{Hom}_{\mathcal{C}}(-,X) \cong \Delta_{\{\mathrm{pt}\}}$, then $\mathrm{Hom}_{\mathcal{C}}(Y,X) \cong \{\mathrm{pt}\}$ for all $Y \in \mathrm{Ob}(\mathcal{C})$, so X is terminal. Dually, an initial object of $\mathcal{C}$ is characterised as a representative of $\Delta_{\{\mathrm{pt}\}}\colon \mathcal{C} \to \mathsf{Set}$.

The language of representable functors is the right language to talk about universal properties (like Kan extensions, adjoint functors and limits).

***Example* XV.1.6.18.** — Fix a right A-module R and a left A-module L. Recall from example XV.1.2.30 that the category of all bilinear maps $R \times L \to M \in \mathrm{Ob}(\mathsf{AbGrp})$ has the (canonical) initial object $R \times L \to R \otimes_A L$. Hence, the functor

$$F\colon \mathsf{AbGrp} \longrightarrow \mathsf{Set}, \qquad M \longmapsto \{\text{bilinear } R \times L \longrightarrow M\}$$

has a representation $F \cong \mathrm{Hom}_{\mathbb{Z}}(R \otimes_A L, -)$. The representative $R \otimes_A L$ with the universal element $s\colon R \times L \to R \otimes_A L$ appear in the universal property of tensor products: for each bilinear map $f\colon R \times L \to M$, that is, $f \in F(M)$, there is a unique $\mathbb{Z}$-linear map $R \otimes_A L \to M$ such that the following diagram commutes:

$$\begin{tikzcd} R \times L \arrow[r, "s"] \arrow[dr, "f"'] & R \otimes_A L \arrow[d, dashed, "\exists!"] \\ & M. \end{tikzcd}$$

***Example* XV.1.6.19.** — For a group G and a commutative ring k, consider the functor

$$F = \mathrm{Hom}_{\mathsf{Grp}}(G, (-)^{\times})\colon k\text{-}\mathsf{Alg} \longrightarrow \mathsf{Set}.$$

This characterises the group algebra $k[G]$ via the representation $F \cong \mathrm{Hom}_{k\text{-}\mathsf{Alg}}(k[G], -)$, and the universal element is the canonical map $s\colon G \hookrightarrow k[G]^{\times}$. It encodes the following universal property: for each group homomorphism $f\colon G \to A^{\times}$ with A a k-algebra, there exists a unique k-algebra map $k[G] \to A$ such that the following diagram commutes:

$$\begin{tikzcd} G \arrow[r, hook, "s"] \arrow[d, "f"'] & k[G] \arrow[d, dashed, "\exists!"] \\ A^{\times} \arrow[r, hook] & A \end{tikzcd}$$

***Example* XV.1.6.20.** — Representable functors are ubiquitous. Here are a couple of representatives of functors.

(i) The forgetful functor $\mathrm{For}\colon \mathsf{Top} \to \mathsf{Set}$ is represented by $\{\mathrm{pt}\}$ with universal element pt.
(ii) The forgetful functor $\mathrm{For}\colon \mathsf{Grp} \to \mathsf{Set}$ is represented by $\mathbb{Z}$ with universal element 1.
(iii) The forgetful functor $\mathrm{For}\colon \mathsf{Ring} \to \mathsf{Set}$ is represented by $\mathbb{Z}[x]$ with universal element x.
(iv) The forgetful functor $\mathrm{For}\colon A\text{-}\mathsf{Mod} \to \mathsf{Set}$ is represented by A with universal element 1.
(v) Taking units $(-)^{\times}\colon \mathsf{Ring} \to \mathsf{Set}$ is represented by $\mathbb{Z}[x, x^{-1}]$ with universal element x.

***Warning* XV.1.6.21.** — In the following subsections, we will cover other ways to express universal properties. There are two approaches in their definitions:

(i) A definition which formulates everything in the language of representable functors. The proofs rely on the Yoneda lemma XV.1.6.5 (remark XV.1.6.6) and are very elegant/short. However, this only works if we assume various smallness conditions in order for the functors' codomain to be Set (e. g., $\mathcal{C}$ is locally small, or I is small and $\mathcal{C}$ is locally small for $\mathrm{Fun}(I, \mathcal{C})$ to be locally small).

(ii) A most general definition which does not rely on representable functors and has no size restrictions. The proofs amount to diagram chasing and are longer.

Fix this formatting (Warning)!

For instructive reasons, we often provide both proofs. Adopting Grothendieck universe, one can circumvent the size restrictions of representable functors by always passing to a sufficiently large universe. However, we avoid this abstract nonsense.

XV.1.7. Adjoint functors.

Like representable functors, adjoint functors are ubiquitous.

> The slogan is 'Adjoint functors arise everywhere'.
>
> —Saunders MacLane, *Categories for the Working Mathematician*

Definition* XV.1.7.1** (Kan 1958).*** — *Let $L\colon \mathcal{C} \to \mathcal{D}$ and $R\colon \mathcal{D} \to \mathcal{C}$ be two functors. The pair of functors (L, R) is **adjoint** or is an **adjunction** if there is a bijection of classes*

$$\mathrm{Hom}_{\mathcal{D}}(L(X), Y) \cong \mathrm{Hom}_{\mathcal{C}}(X, R(Y))$$

for all $X \in \mathrm{Ob}(\mathcal{C})$ and $Y \in \mathrm{Ob}(\mathcal{D})$ such that this bijection is natural in X and Y, i. e., the corresponding naturality square

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{D}}(L(X), Y) \arrow[r, "\sim"] \arrow[d, "g\circ - \circ L(f)"'] & \mathrm{Hom}_{\mathcal{C}}(X, R(Y)) \arrow[d, "R(g)\circ - \circ f"] \\
\mathrm{Hom}_{\mathcal{D}}(L(X'), Y') \arrow[r, "\sim"] & \mathrm{Hom}_{\mathcal{C}}(X', R(Y'))
\end{tikzcd}$$

*commutes for all $f\colon X' \to X$ in $\mathcal{C}$ and all $g\colon Y \to Y'$ in $\mathcal{D}$. In this case, L is the **left adjoint** of R and R is the **right adjoint** of L, and we write $L \dashv R$. We will express this situation in diagrams as*

$$\begin{tikzcd}
\mathcal{C} \arrow[r, bend left, "L"] \arrow[r, phantom, "\perp"] & \mathcal{D}. \arrow[l, bend left, "R"]
\end{tikzcd}$$

*Moreover, we call the bijection the **adjunction isomorphism**, and if f is an element of either side, its image under the adjunction isomorphism is the **adjunct** of f.*

***Example* XV.1.7.2.** — For sets X, Y and Z, we have an isomorphism

$$\mathrm{Hom}_{\mathsf{Set}}(X \times Y, Z) \cong \mathrm{Hom}_{\mathsf{Set}}(X, \mathrm{Hom}_{\mathsf{Set}}(Y, Z))$$

which is natural in X, Y and Z. Hence, the functors $- \times Y$ and $\mathrm{Hom}_{\mathsf{Set}}(Y, -)$ are adjoint. The same works for, e. g., Top.

***Example* XV.1.7.3.** — Let R be a (not necessarily commutative) algebra over a commutative ring k. For a k-module K and for left R-modules M and N, recall that the *tensor-hom adjunction*

Maybe insert reference.

$$\mathrm{Hom}_R(N \otimes_k K, M) \cong \mathrm{Hom}_k(K, \mathrm{Hom}_R(N, M))$$

shows that $N \otimes_k -$ and $\mathrm{Hom}_k(N, -)$ are adjoint functors $R\text{-}\mathsf{LMod} \to R\text{-}\mathsf{LMod}$. Let $\mathrm{For}\colon R\text{-}\mathsf{LMod} \to k\text{-}\mathsf{Mod}$ yield the underlying k-module structure. With $N = R$, we obtain the adjunction

$$\mathrm{Hom}_R(R \otimes_k K, M) \cong \mathrm{Hom}_k(K, \mathrm{Hom}_R(R, M)) \cong \mathrm{Hom}_k(K, \mathrm{For}(M)).$$

Hence, the extension of scalars $R \otimes_k -$ is a left adjoint of the restriction of scalars For.

Similarly, swapping the roles of R and k, for a right R-module K, a left R-module N and a k-module M, we have the tensor-hom adjunction

$$\mathrm{Hom}_k(K \otimes_R N, M) \cong \mathrm{Hom}_R(N, \mathrm{Hom}_k(K, M)).$$

With $K = R$, we obtain

$$\mathrm{Hom}_k(\mathrm{For}(N), M) \cong \mathrm{Hom}_k(R \otimes_k N, M) \cong \mathrm{Hom}_R(R, \mathrm{Hom}_k(R, M)).$$

Hence, the coextension of scalars $\mathrm{Hom}_k(R, M)$ is a right adjoint of the restriction of scalars For.

***Example* XV.1.7.4.** — The forgetful functor $\mathrm{For}\colon \mathsf{Top} \to \mathsf{Set}$ admits both left and right adjoints. Let $D\colon \mathsf{Set} \to \mathsf{Top}$ be the functor that endows a set with the discrete topology, and let $I\colon \mathsf{Set} \to \mathsf{Top}$ be the functor that endows a set with the indiscrete topology. Then we have natural bijections

$$\mathrm{Hom}_{\mathsf{Top}}(D(X), Y) \cong \mathrm{Hom}_{\mathsf{Set}}(X, \mathrm{For}(Y)), \qquad \mathrm{Hom}_{\mathsf{Set}}(\mathrm{For}(X), Y) \cong \mathrm{Hom}_{\mathsf{Top}}(X, I(Y)).$$

Hence, we obtain adjunctions $D \dashv \mathrm{For} \dashv I$.

***Example* XV.1.7.5.** — In various contexts in mathematics, a forgetful functor almost always has a left adjoint which is some kind of *free* construction. Interestingly, right adjoints of forgetful functors are rarer. Here some examples:

(i) $\mathrm{For}\colon \mathsf{Ring} \to \mathsf{AbGrp}$. The free ring of an abelian group A is $\bigoplus_{n \geq 0} A^{\otimes n}$.
(ii) $\mathrm{For}\colon A\text{-}\mathsf{Mod} \to \mathsf{Set}$. The free A-module on a set X is $A^{\oplus X}$.
(iii) $\mathrm{For}\colon \mathsf{CommRing} \to \mathsf{Set}$. The free commutative ring on a set X is the polynomial ring $\mathbb{Z}[X]$.
(iv) $\mathrm{For}\colon k\text{-}\mathsf{Alg} \to \mathsf{Set}$. The free k-algebra on a set X is the free k-algebra $k\langle X \rangle$.
(v) $(-)^\times\colon \mathsf{Ring} \to \mathsf{Grp}$. The free ring on a group G is the group ring $\mathbb{Z}[G]$.

There is an alternative definition of adjunctions.

Insert reference to representation theory defining (co)induction (Frobenius reciprocity). Or see [Rie16, Ex. 4.1.11].

Proposition* XV.1.7.6.** — *A pair (L, R) of functors $\mathcal{C} \rightleftarrows \mathcal{D}$ is adjoint if and only if there exist natural transformations $\eta\colon \mathrm{id}_{\mathcal{C}} \to R \circ L$ and $\varepsilon\colon L \circ R \to \mathrm{id}_{\mathcal{D}}$ satisfying the* ***triangle identities

$$(R\varepsilon) \circ (\eta R) = \mathrm{id}_R, \qquad (\varepsilon L) \circ (L\eta) = \mathrm{id}_L,$$

or in diagrams:

$$\begin{tikzcd} R \arrow[r, Rightarrow, "\eta R"] \arrow[dr, Rightarrow, "\mathrm{id}_R"'] & R \circ L \circ R \arrow[d, Rightarrow, "R\varepsilon"] \\ & R, \end{tikzcd} \qquad \begin{tikzcd} L \arrow[r, Rightarrow, "L\eta"] \arrow[dr, Rightarrow, "\mathrm{id}_L"'] & L \circ R \circ L \arrow[d, Rightarrow, "\varepsilon L"] \\ & L. \end{tikzcd}$$

Here, the juxtaposition of a natural transformation and a functor is whiskering. In this case, η is the ***unit*** *and ε is the* ***counit*** *of the adjunction $L \dashv R$. Moreover, adjuncts are given by the commutative diagrams*

$$\begin{tikzcd} \mathrm{Hom}_{\mathcal{D}}(L(X), Y) \arrow[r, "R"] \arrow[dr, "\sim"'] & \mathrm{Hom}_{\mathcal{C}}(R(L(X)), R(Y)) \arrow[d, "- \circ \eta_X"] \\ & \mathrm{Hom}_{\mathcal{C}}(X, R(Y)) \end{tikzcd}$$

and

$$\begin{tikzcd} \mathrm{Hom}_{\mathcal{C}}(X, R(Y)) \arrow[r, "L"] \arrow[dr, "\sim"'] & \mathrm{Hom}_{\mathcal{D}}(L(X), L(R(Y))) \arrow[d, "\varepsilon_Y \circ -"] \\ & \mathrm{Hom}_{\mathcal{D}}(L(X), Y) \end{tikzcd}$$

for all $X \in \mathrm{Ob}(\mathcal{C})$ and $Y \in \mathrm{Ob}(\mathcal{D})$.

Proof. — Let $L \dashv R$ be an adjunction. Under the adjunction isomorphism

$$\mathrm{Hom}_{\mathcal{D}}(L(X), L(X)) \cong \mathrm{Hom}_{\mathcal{C}}(X, R(L(X))),$$

the adjunct of $\mathrm{id}_{L(X)}$ gives a morphism $\eta_X \colon X \to R(L(X))$ for each $X \in \mathrm{Ob}(\mathcal{C})$. These assemble into a natural transformation $\eta \colon \mathrm{id}_{\mathcal{C}} \to R \circ L$ by naturality of the adjunction isomorphism.

To see the adjunct formula, for $f \in \mathrm{Hom}_{\mathcal{D}}(L(X), Y)$, the naturality of the adjunction isomorphism gives the commutative square

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{D}}(L(X), L(X)) \arrow[r, "f\circ -"] \arrow[d, "\wr"'] & \mathrm{Hom}_{\mathcal{D}}(L(X), Y) \arrow[d, "\wr"] \\
\mathrm{Hom}_{\mathcal{C}}(X, R(L(X))) \arrow[r, "R(f)\circ -"] & \mathrm{Hom}_{\mathcal{C}}(X, R(Y)),
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathrm{id}_{L(X)} \arrow[r, mapsto] \arrow[d, mapsto] & f \arrow[d, mapsto] \\
\eta_X \arrow[r, mapsto] & R(f) \circ \eta_X.
\end{tikzcd}$$

Thus, the adjunct of f is $R(f) \circ \eta_X$.

Applying this adjunction formula to $\varepsilon_Y \in \mathrm{Hom}_{\mathcal{C}}(L(R(Y)), Y)$ (with $X = R(Y)$), we obtain $R(\varepsilon_Y) \circ \eta_{R(Y)} = \mathrm{id}_{R(Y)}$ for any $Y \in \mathrm{Ob}(\mathcal{D})$, which is the triangle inequality $(R\varepsilon) \circ (\eta R) = \mathrm{id}_R$.

For the converse, assume the triangle identities. We define the adjunction maps

$$\mathrm{Hom}_{\mathcal{D}}(L(X), Y) \xrightarrow{R} \mathrm{Hom}_{\mathcal{C}}(R(L(X)), R(Y)) \xrightarrow{-\circ\eta_X} \mathrm{Hom}_{\mathcal{C}}(X, R(Y)),$$
$$\mathrm{Hom}_{\mathcal{C}}(X, R(Y)) \xrightarrow{L} \mathrm{Hom}_{\mathcal{D}}(L(X), L(R(Y))) \xrightarrow{\varepsilon_Y\circ -} \mathrm{Hom}_{\mathcal{D}}(L(X), Y).$$

Direct computation with the naturality of the (co)units and triangle identities shows that these are mutually inverse, e. g., for $f \colon L(X) \to Y$ we have

$$\varepsilon_Y \circ L(R(f) \circ \eta_X) = \varepsilon_Y \circ L(R(f)) \circ L(\eta_X) = f \circ \varepsilon_{L(X)} \circ L(\eta_X) = f$$

by fist using naturality of ε and then the triangle identity $(\varepsilon L) \circ (L\eta) = \mathrm{id}_L$. Naturality of the adjunction isomorphism follows easily from functoriality of L and R.

***Observation* XV.1.7.7.** — Let $L \dashv R$ be an adjunction $\mathcal{C} \rightleftarrows \mathcal{D}$ with (co)units (η, ε). For any $Y, Y' \in \mathrm{Ob}(\mathcal{D})$, consider the diagram

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{C}}(R(Y), R(Y')) \arrow[r, "L"] \arrow[dr, "\sim"] & \mathrm{Hom}_{\mathcal{D}}(L(R(Y)), L(R(Y'))) \arrow[d, "\varepsilon_{Y'}\circ -"] \\
\mathrm{Hom}_{\mathcal{D}}(Y, Y') \arrow[u, "R"] \arrow[r, "-\circ\varepsilon_Y"] & \mathrm{Hom}_{\mathcal{D}}(L(R(Y)), Y').
\end{tikzcd} \tag{XV.1.7.7.1}$$

The upper right triangle commutes because this is the adjunct diagram from proposition XV.1.7.6. The outer square commutes since for any morphism $f \colon Y \to Y'$, we have $\varepsilon_{Y'} \circ L(R(f)) = f \circ \varepsilon_Y$ by naturality of ε. Hence, the lower left triangle, i. e., the diagram

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{D}}(Y, Y') \arrow[r, "R"] \arrow[dr, "-\circ\varepsilon_Y"'] & \mathrm{Hom}_{\mathcal{C}}(R(Y), R(Y')) \arrow[d, "\wr"] \\
& \mathrm{Hom}_{\mathcal{D}}(L(R(Y)), Y'),
\end{tikzcd}$$

commutes. Dually, the diagram

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{C}}(X, X') \arrow[r, "L"] \arrow[dr, "\eta_{X'}\circ -"'] & \mathrm{Hom}_{\mathcal{D}}(L(X), L(X')) \arrow[d, "\wr"] \\
& \mathrm{Hom}_{\mathcal{C}}(X, R(L(X')))
\end{tikzcd}$$

commutes for all $X, X' \in \mathrm{Ob}(\mathcal{C})$.

***Remark* XV.1.7.8.** — If $\mathcal{C}$ and $\mathcal{D}$ are locally small, then hom-functors are set-valued and we can instead define that $L \dashv R$ is an adjunction of functors $\mathcal{C} \rightleftarrows \mathcal{D}$ if there is a natural isomorphism

$$\mathrm{Hom}_{\mathcal{D}}(L(-), -) \cong \mathrm{Hom}_{\mathcal{C}}(-, R(-))$$

of bifunctors $\mathcal{C}^{\mathrm{op}} \times \mathcal{D} \to \mathsf{Set}$.

XV.1.8. Properties of adjoint functors.

***Proposition* XV.1.8.1** (adjoints from object-wise representation). — *Let $L\colon \mathcal{C} \to \mathcal{D}$ be a functor. If each object $Y \in \mathrm{Ob}(\mathcal{D})$ has an object $R(Y) \in \mathrm{Ob}(\mathcal{C})$ such that there is a bijection of classes*

$$\mathrm{Hom}_{\mathcal{D}}(L(X), Y) \cong \mathrm{Hom}_{\mathcal{C}}(X, R(Y))$$

for all $X \in \mathrm{Ob}(\mathcal{C})$ and natural in X, then $R\colon \mathrm{Ob}(\mathcal{D}) \to \mathrm{Ob}(\mathcal{C})$ can be uniquely extended to a functor $R\colon \mathcal{D} \to \mathcal{C}$ which is right adjoint to L. The dual statement for the existence of left adjoints also holds.

Proof with the Yoneda lemma. — The functor $\mathrm{Hom}_{\mathcal{D}}(L(-), Y)\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ is represented by $R(Y)$ for each Y. Moreover, naturality says that for any morphism $g\colon Y \to Y'$ in $\mathcal{D}$ the square

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{D}}(L(-), Y) \arrow[r, "g\circ-"] \arrow[d, "\wr"'] & \mathrm{Hom}_{\mathcal{D}}(L(-), Y') \arrow[d, "\wr"] \\
\mathrm{Hom}_{\mathcal{C}}(-, R(Y)) \arrow[r, "\mathrm{Hom}_{\mathcal{C}}(-,R(g))"] & \mathrm{Hom}_{\mathcal{C}}(-, R(Y'))
\end{tikzcd}$$

commutes and hence defines the natural transformation $\mathrm{Hom}_{\mathcal{C}}(-, R(g))$. The Yoneda lemma XV.1.6.5 yields a unique $R(g)\colon R(Y) \to R(Y')$, extending R to a functor. ◻

Proof with (co)units. — Define the counit $\varepsilon\colon L \circ R \to \mathrm{id}_{\mathcal{D}}$ in the same way as in the proof of proposition XV.1.7.6, which is natural since the adjunction isomorphism is natural in X. Then for morphisms $g\colon Y \to Y'$ in $\mathcal{D}$, define $R(g)$ as the adjunct of $g \circ \varepsilon_Y \in \mathrm{Hom}_{\mathcal{D}}(L(R(Y)), Y')$. Note that this definition is forced by observation XV.1.7.7 if R shall be a right adjoint of L.

To show functoriality $R(h \circ g) = R(h) \circ R(g)$ for another morphism $h\colon Y' \to Y''$ in $\mathcal{D}$, we show equality under the adjunction isomorphism. By naturality of the given isomorphism in X, we the commutative square

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{D}}(L(R(Y')), Y'') \arrow[r, "-\circ L(R(g))"] \arrow[d, "\wr"'] & \mathrm{Hom}_{\mathcal{D}}(L(R(Y)), Y'') \arrow[d, "\wr"] & h \circ \varepsilon_{Y'} \arrow[r, mapsto] \arrow[d, mapsto] & h \circ \varepsilon_Y \circ L(R(g)) \arrow[d, mapsto] \\
\mathrm{Hom}_{\mathcal{C}}(R(Y'), R(Y'')) \arrow[r, "-\circ R(g)"] & \mathrm{Hom}_{\mathcal{C}}(R(Y), R(Y'')), & R(h) \arrow[r, mapsto] & R(h) \circ R(g),
\end{tikzcd}$$

so the adjunct of $R(h) \circ R(g)$ is $h \circ \varepsilon_Y \circ L(R(g))$. But by naturality of ε, we have $\varepsilon_Y \circ L(R(g)) = g \circ \varepsilon_{L(R(Y))}$, and hence the adjunct of $h \circ \varepsilon_Y \circ L(R(g))$ is also the adjunct of $R(h \circ g)$, as desired. Finally, we see that the isomorphism is also natural in Y, given by post-composition, i. e., the following diagram commutes:

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{D}}(L(X), Y) \arrow[r, "g\circ-"] \arrow[d, "\wr"'] & \mathrm{Hom}_{\mathcal{D}}(L(X), Y') \arrow[d, "\wr"] \\
\mathrm{Hom}_{\mathcal{C}}(X, R(Y)) \arrow[r, "R(g)\circ-"] & \mathrm{Hom}_{\mathcal{C}}(X, R(Y')).
\end{tikzcd}$$

◻

***Proposition* XV.1.8.2** (uniqueness of adjoints)**.** — *If a functor $L\colon \mathcal{C} \to \mathcal{D}$ admits a right adjoint, then it is unique up to natural isomorphism. In more detail, if $L \dashv R$ and $L \dashv R'$ are two adjunctions, then there exists a unique natural isomorphism $\phi\colon R \xrightarrow{\sim} R'$ such that:*

(i) *either, if ε and ε' are the counits of $L \dashv R$ and $L \dashv R'$, resp., we have $\varepsilon' \circ (L\phi) = \varepsilon$;*
(ii) *or equivalently, the diagram*

$$\begin{tikzcd} \mathrm{Hom}_{\mathcal{C}}(X, R(Y)) \arrow[rr, "\phi_Y \circ -", "\sim"'] & & \mathrm{Hom}_{\mathcal{C}}(X, R'(Y)) \\ & \mathrm{Hom}_{\mathcal{D}}(L(X), Y) \arrow[ul, "\sim"] \arrow[ur, "\sim"'] & \end{tikzcd}$$

commutes for all $X \in \mathrm{Ob}(\mathcal{C})$ and $Y \in \mathrm{Ob}(\mathcal{D})$ and is natural in X and Y.

The dual statement holds for left adjoints.

Proof with the Yoneda lemma. — The adjunction isomorphisms of $L \dashv R$ and $L \dashv R'$ give rise to a natural isomorphism

$$\alpha\colon\ \mathrm{Hom}_{\mathcal{D}}(-, R(Y)) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{C}}(L(-), Y) \xrightarrow{\sim} \mathrm{Hom}_{\mathcal{D}}(-, R'(Y))$$

of functors $\mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ for all $Y \in \mathrm{Ob}(\mathcal{D})$. By the uniqueness of representable functors (remark XV.1.6.12), we obtain a unique isomorphism $\phi_Y\colon R(Y) \xrightarrow{\sim} R'(Y)$. These assemble into a natural isomorphism $\phi\colon R \to R'$ by the naturality of the Yoneda lemma XV.1.6.5.

For uniqueness, note that in the proof of the Yoneda lemma XV.1.6.5, we have seen that for each $X \in \mathrm{Ob}(\mathcal{C})$, the component α_X is of the form $f \mapsto s \circ f$ where $s \in \mathrm{Hom}_{\mathcal{C}}(R(Y), R'(Y))$ is the unique element corresponding to α. Thus, $s = \phi_Y$. ◻

Proof with (co)units. — Let (η, ε) and (η', ε') be the (co)unit pairs for $L \dashv R$ and $L \dashv R'$, resp. Consider the natural transformations

$$\varPhi\colon R \xrightarrow{\eta' R} R' \circ L \circ R \xrightarrow{R'\varepsilon} R', \qquad \varPsi\colon R' \xrightarrow{\eta R'} R \circ L \circ R' \xrightarrow{R\varepsilon'} R.$$

To show that these are mutually inverse, we show equality, e. g., $\varPsi_Y \circ \varPhi_Y = \mathrm{id}_{R(Y)}$ for each $Y \in \mathrm{Ob}(\mathcal{D})$, by showing equality of the adjuncts of both sides. Recall from proposition XV.1.7.6 that the adjunct formula is $\varepsilon_Y \circ L(-)$ on morphisms $R(Y) \to R(Y)$. We have the following identities:

- $\varepsilon_Y \circ L(R(\varepsilon'_Y)) = \varepsilon'_Y \circ \varepsilon_{L(R'(Y))}$ by naturality of ε applied to $\varepsilon'_Y\colon L(R'(Y)) \to Y$;
- $\varepsilon_{L(R'(Y))} \circ L(\eta_{R'(Y)}) = \mathrm{id}_{L(R'(Y))}$ by the triangle identity $(\varepsilon L) \circ (L\eta) = \mathrm{id}_L$;
- $\varepsilon'_Y \circ L(R'(\varepsilon_Y)) = \varepsilon_Y \circ \varepsilon'_{L(R(Y))}$ by naturality of ε' applied to $\varepsilon_Y\colon L(R(Y)) \to Y$;
- $\varepsilon'_{L(R(Y))} \circ L(\eta'_{R(Y)}) = \mathrm{id}_{L(R(Y))}$ by the triangle identity $(\varepsilon' L) \circ (L\eta') = \mathrm{id}_L$.

Thus,

$$\begin{aligned} &\varepsilon_Y \circ L(R(\varepsilon'_Y) \circ \eta_{R'(Y)} \circ R'(\varepsilon_Y) \circ \eta'_{R(Y)}) \\ &\qquad = \varepsilon_Y \circ L(R(\varepsilon'_Y)) \circ L(\eta_{R'(Y)}) \circ L(R'(\varepsilon_Y)) \circ L(\eta'_{R(Y)}) \\ &\qquad = \varepsilon'_Y \circ \varepsilon_{L(R'(Y))} \circ L(\eta_{R'(Y)}) \circ L(R'(\varepsilon_Y)) \circ L(\eta'_{R(Y)}) \\ &\qquad = \varepsilon'_Y \circ L(R'(\varepsilon_Y)) \circ L(\eta'_{R(Y)}) \\ &\qquad = \varepsilon_Y \circ \varepsilon'_{L(R(Y))} \circ L(\eta'_{R(Y)}) \\ &\qquad = \varepsilon_Y, \end{aligned}$$

as desired.

For uniqueness, just observe that the condition $\varepsilon'_Y \circ L(\varPhi_Y) = \varepsilon_Y$ for all $Y \in \mathrm{Ob}(\mathcal{D})$ says that the adjunct of $\varPhi_Y$ is ε_Y by proposition XV.1.7.6. Since the adjunction map is bijective, this shows uniqueness of $\varPhi_Y$. ◻

***Proposition* XV.1.8.3** (horizontal composition of adjoints). — *Given an adjunction $L_1 \dashv R_1$ of functors $\mathcal{C} \rightleftarrows \mathcal{D}$ and an adjunction $L_2 \dashv R_2$ of functors $\mathcal{D} \rightleftarrows \mathcal{E}$, then we have an adjunction $L_2 \circ L_1 \dashv R_1 \circ R_2$ of functors $\mathcal{C} \rightleftarrows \mathcal{D}$. Or as a diagram:*

$$\begin{tikzcd} \mathcal{C} \arrow[r, bend left, "L_1"] \arrow[r, phantom, "\perp"] & \mathcal{D} \arrow[l, bend left, "R_1"] \arrow[r, bend left, "L_2"] \arrow[r, phantom, "\perp"] & \mathcal{E} \arrow[l, bend left, "R_2"] \end{tikzcd} \quad \rightsquigarrow \quad \begin{tikzcd} \mathcal{C} \arrow[r, bend left, "L_2\circ L_1"] \arrow[r, phantom, "\perp"] & \mathcal{E}. \arrow[l, bend left, "R_2\circ R_1"] \end{tikzcd}$$

Proof. — We have bijections

$$\mathrm{Hom}_{\mathcal{E}}(L_2(L_1(X)), Y) \cong \mathrm{Hom}_{\mathcal{D}}(L_1(X), R_2(Y)) \cong \mathrm{Hom}_{\mathcal{C}}(X, R_1(R_2(Y)))$$

for all $X \in \mathrm{Ob}(\mathcal{C})$ and $Y \in \mathrm{Ob}(\mathcal{E})$ natural in X and Y. ∎

***Warning* XV.1.8.4.** — There is no vertical composition of adjoints! If $F \dashv G \dashv H$, then we do *not* have $F \dashv H$ since they are both functors in the same direction.

***Proposition* XV.1.8.5.** — *Let $L \dashv R$ be an adjunction of functors $\mathcal{C} \rightleftarrows \mathcal{D}$ with (co)units (η, ε).*

(i) *The right adjoint R is fully faithful if and only if the counit $\varepsilon\colon L \circ R \Rightarrow \mathrm{id}_{\mathcal{D}}$ is a natural isomorphism.*

(ii) *Dually, the left adjoint L is fully faithful if and only if the unit $\eta\colon \mathrm{id}_{\mathcal{C}} \Rightarrow R \circ L$ is a natural isomorphism.*

Proof. — Recall from observation XV.1.7.7 that the diagram

$$\begin{tikzcd} \mathrm{Hom}_{\mathcal{D}}(Y, Y') \arrow[r, "R"] \arrow[dr, "-\circ\varepsilon_Y"'] & \mathrm{Hom}_{\mathcal{C}}(R(Y), R(Y')) \arrow[d, "\wr"] \\ & \mathrm{Hom}_{\mathcal{D}}(L(R(Y)), Y') \end{tikzcd}$$

commutes for all $Y, Y' \in \mathrm{Ob}(\mathcal{D})$. In particular, R is fully faithful if and only if ε_Y is an isomorphism for all $Y \in \mathrm{Ob}(\mathcal{D})$, i. e., if and only if ε is a natural isomorphism. ∎

***Corollary* XV.1.8.6.** — *Let $L \dashv R$ be an adjunction. Then the following are equivalent:*

(i) *L is an equivalence of categories.*
(ii) *R is an equivalence of categories.*
(iii) *L and R are fully faithful.*

In this case, L and R are mutually quasi-inverse, and their (co)units are natural isomorphisms. The adjoint pair $L \dashv R$ of equivalences is sometimes called an ***adjoint equivalence****.*

Proof. — Suppose that R is an equivalence of categories, i. e., it is fully faithful and essentially surjective by proposition XV.1.5.11 (this holds even without the axiom of choice). Then the counit ε is a natural isomorphism. To show that $L\colon \mathrm{Hom}_{\mathcal{C}}(X, X') \to \mathrm{Hom}_{\mathcal{D}}(L(X), L(X'))$ is bijective for any $X, X' \in \mathrm{Ob}(\mathcal{C})$, there are isomorphisms $R(Y) \cong X$ and $X' \cong R(Y')$ for some $Y, Y' \in \mathrm{Ob}(\mathcal{D})$. Therefore, it suffices to show that $L\colon \mathrm{Hom}_{\mathcal{C}}(R(Y), R(Y')) \to \mathrm{Hom}_{\mathcal{D}}(L(R(Y)), L(R(Y')))$ is bijective. But this is the case since this map appears in diagram (XV.1.7.7.1) from observation XV.1.7.7 and ε is an isomorphism:

$$\begin{tikzcd} \mathrm{Hom}_{\mathcal{C}}(R(Y), R(Y')) \arrow[r, "L"] \arrow[dr, "\sim"] & \mathrm{Hom}_{\mathcal{D}}(L(R(Y)), L(R(Y'))) \arrow[d, "\varepsilon_{Y'}\circ-"] \\ \mathrm{Hom}_{\mathcal{D}}(Y, Y') \arrow[u, "R"] \arrow[r, "-\circ\varepsilon_Y"] & \mathrm{Hom}_{\mathcal{D}}(L(R(Y)), Y'). \end{tikzcd}$$

Thus, L is fully faithful and we showed that (ii) implies (iii). All other statements are either dual or immediate from proposition XV.1.8.5.

The converse also holds.

***Proposition* XV.1.8.7.** — *Any pair $L\colon \mathcal{C} \to \mathcal{D}$ and $R\colon \mathcal{D} \to \mathcal{C}$ of mutually quasi-inverse equivalences can be promoted to an adjoint equivalence $L \dashv R$. This means that there are natural isomorphisms $\eta\colon \mathrm{id}_{\mathcal{C}} \xrightarrow{\sim} R \circ L$ and $\varepsilon\colon L \circ R \xrightarrow{\sim} \mathrm{id}_{\mathcal{D}}$ satisfying the triangle identities.*

Proof. — Let $\eta\colon \mathrm{id}_{\mathcal{C}} \xrightarrow{\sim} R \circ L$ be a natural isomorphism given by the equivalence. Then the map

$$\mathrm{Hom}_{\mathcal{D}}(L(X), Y) \xrightarrow{R} \mathrm{Hom}_{\mathcal{C}}(R(L(X)), R(Y)) \xrightarrow{-\circ\eta_X} \mathrm{Hom}_{\mathcal{C}}(X, R(Y))$$

is a bijection since R is fully faithful and since η_X is an isomorphism. This map is also natural in X since R is functorial and since η is natural. By proposition XV.1.8.1, this defines an adjunction $L \dashv R$.

Alternative proof sketch. — Given natural isomorphism $\eta\colon \mathrm{id}_{\mathcal{C}} \xrightarrow{\sim} R \circ L$ and $\xi\colon L \circ R \to \mathrm{id}_{\mathcal{D}}$ from the equivalence of categories, we define the natural isomorphism

$$\varepsilon\colon L \circ R \xrightarrow[\sim]{(L\circ R)\xi^{-1}} L \circ R \circ L \circ R \xrightarrow[\sim]{L\eta^{-1}R} L \circ R \xrightarrow[\sim]{\xi} \mathrm{id}_{\mathcal{D}}.$$

It remains to check the triangle identities for (η, ε), which we will not do here. For an explicit and tedious calculation, see [Rie16, Proposition 4.3.5] for a conventional calculation or [nLa26, adjoint equivalence] for a more visual calculation with string diagrams.

***Remark* XV.1.8.8.** — Hence, if $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{C}$ are mutually quasi-inverse equivalences of categories, then G is both left and right adjoint of F by symmetry.

§ XV.2. LIMITS

XV.2.1. Definition of limits.

As with representable functors and adjunctions, limits arise everywhere. Limits are the most common incarnation of universal properties across mathematics and are the primary (if not only) way to construct new objects. The most common and visual definition is the following.

XV.2.1.1. — Recall that a *diagram* is a functor $F\colon I \to \mathcal{C}$ where I is called the *index category*. Moreover, F is *finite* or *small* if I is. Also, recall that the *constant functor* $\Delta_X\colon I \to \mathcal{C}$ for $X \in \mathrm{Ob}(\mathcal{C})$ maps every object of I to X (and every morphism of I to id_X).

***Definition* XV.2.1.2.** — *A functor $\Delta\colon \mathcal{C} \to \mathrm{Fun}(I, \mathcal{C})$, $X \mapsto \Delta_X$ is called a **diagonal functor**. If I is discrete and small, then $\mathrm{Fun}(I, \mathcal{C}) \cong \prod_{i\in\mathrm{Ob}(I)} \mathcal{C}$ and $\Delta\colon \mathcal{C} \to \prod_{i\in I} \mathcal{C}$ is given by $X \mapsto (X)_{i\in\mathrm{Ob}(I)}$.*

***Definition* XV.2.1.3.** — (i) *Let $F\colon I^{\mathrm{op}} \to \mathcal{C}$ be a diagram (we use this opposite convention because of the Yoneda lemma XV.1.6.5). For an object X of $\mathcal{C}$, a **cone** over F is a family of morphisms $(\alpha_i\colon X \to F(i))_{i\in\mathrm{Ob}(I)}$ which are compatible with all morphism $F(f)\colon F(j) \to F(i)$ for all morphisms $f\colon i \to j$ in I, i. e., the diagram*

$$\begin{tikzcd} & X \arrow[dl, "\alpha_i"'] \arrow[dr, "\alpha_j"] & \\ F(j) \arrow[rr, "F(f)"] & & F(i) \end{tikzcd}$$

commutes for all morphisms $f\colon i \to j$ in I. Equivalently, a cone *over F is a natural transformation $\alpha\colon \Delta_X \to F$.*

(ii) *Dually, let $F\colon I \to \mathcal{C}$ be a diagram. For an object X of $\mathcal{C}$, a* ***cocone*** *under F is a family of morphisms $(\beta_i\colon X \to F(i))_{i\in \mathrm{Ob}(I)}$ such that the diagram*

$$\begin{tikzcd} F(i) \arrow[rr, "F(f)"] \arrow[dr] & & F(j) \arrow[dl] \\ & X & \end{tikzcd}$$

commutes for all morphisms $f\colon i \to j$ in I. Equivalently, a cocone *under F is a natural transformation $\beta\colon F \to \Delta_X$.*

***Remark* XV.2.1.4.** — Let $F\colon I^{\mathrm{op}} \to \mathcal{C}$ be a diagram, and let $\Delta\colon \mathcal{C} \to \mathrm{Fun}(I^{\mathrm{op}}, \mathcal{C})$ be the diagonal functor. Then the cones over F are precisely the objects in the comma category $\Delta \downarrow F$ where $F\colon 1 \to \mathrm{Fun}(I^{\mathrm{op}}, \mathcal{C})$ is the inclusion of F. Thus, we obtain a **category of cones** over F. Dually, the **category of cocones** under F is $F \downarrow \Delta$.

Let us explicitly describe the morphisms in the category of cones. A morphism between two cones $(\alpha_i\colon X \to F(i))_{i\in \mathrm{Ob}(I)}$ and $(\beta_i\colon X \to F(i))_{i\in \mathrm{Ob}(I)}$ over F are morphisms $f\colon X \to Y$ in $\mathcal{C}$ such that

$$\begin{tikzcd} F(X) \arrow[rr, "F(f)"] \arrow[dr, "\alpha_i"'] & & F(Y) \arrow[dl, "\beta_i"] \\ & F(i) & \end{tikzcd}$$

commutes for all $i \in \mathrm{Ob}(I)$. Equivalently, a morphism of cones $\alpha\colon \Delta_X \to F$ and $\beta\colon \Delta_Y \to F$ is a natural transformation $\phi\colon \Delta_X \to \Delta_Y$ such that $\beta \circ \phi = \alpha$, or as a diagram,

$$\begin{tikzcd} \Delta_X \arrow[rr, Rightarrow, "\phi"] \arrow[dr, Rightarrow, "\alpha"'] & & \Delta_Y \arrow[dl, Rightarrow, "\beta"] \\ & F. & \end{tikzcd}$$

Note that by naturality of ϕ and by definition of constant functors, we have $\phi_i = \phi_j\colon X \to Y$ for all $i, j \in \mathrm{Ob}(I)$. Hence, we have $\phi = \Delta_f$ for the morphism $f = \phi_i\colon X \to Y$ in $\mathcal{C}$.

Definition* XV.2.1.5.** — (i) *Let $F\colon I^{\mathrm{op}} \to \mathcal{C}$ be a diagram. A* ***limit cone *or* ***universal cone*** *over F is a terminal object in the category of cones over F. The vertex of the limit cone is called* ***limit*** *or* ***projective limit*** *in $\mathcal{C}$ and we write $\lim_I F = \lim_{I^{\mathrm{op}}} F = \lim_{i\in I} F(i)$. Explicitly, a* limit cone *over F is a cone*

$$\begin{tikzcd} & \lim_I F \arrow[dl] \arrow[dr] & \\ F(j) \arrow[rr, "F(f)"] & & F(i) \end{tikzcd}$$

such that for any other cone over F with vertex X, there exists a unique morphism $X \to \lim_I F$ such that the diagram

$$\begin{tikzcd} & X \arrow[d, dashed, "\exists!"] \arrow[ddl] \arrow[ddr] & \\ & \lim_I F \arrow[dl] \arrow[dr] & \\ F(j) \arrow[rr, "F(f)"] & & F(i) \end{tikzcd}$$

commutes for all $f\colon i \to j$ in I.

(ii) *Let $F\colon I \to \mathcal{C}$ be a diagram. Dually, a* ***colimit cone***[XV.6] *or* ***universal cocone*** *under F is an initial object in the category of cocones under F. The vertex of the colimit cone is called* ***colimit*** *or* ***inductive limit*** *in $\mathcal{C}$ and we write* $\operatorname{colim}_I F = \operatorname{colim}_{i\in I} F$*. Explicitly,* a colimit cone *over F is a cone*

$$\begin{tikzcd} F(i) \arrow[rr, "F(f)"] \arrow[dr] & & F(j) \arrow[dl] \\ & \operatorname{colim}_I F & \end{tikzcd}$$

such that for any other cocone under F with vertex X, there exists a unique morphism $\operatorname{colim}_I F \to X$ *such that the diagram*

$$\begin{tikzcd} F(i) \arrow[rr, "F(f)"] \arrow[dr] \arrow[ddr] & & F(j) \arrow[dl] \arrow[ddl] \\ & \operatorname{colim}_I F \arrow[d, dashed, "\exists!"] & \\ & X & \end{tikzcd}$$

commutes for all $f\colon i \to j$ in I.

***Remark* XV.2.1.6.** — Let $\phi\colon J \to I$ be a functor and let $\phi^{\mathrm{op}}\colon J^{\mathrm{op}} \to I^{\mathrm{op}}$ be induced by ϕ. If $F\colon I^{\mathrm{op}} \to \mathcal{C}$ is a diagram, then there exists a canonical morphism $\lim_I F \to \lim_J(F\circ\phi^{\mathrm{op}})$ if the limits exist. This morphism is natural in ϕ, i. e., let $\phi_1, \phi_2\colon J \rightrightarrows I$ be two functors with a natural transformation $\alpha\colon \phi_1 \to \phi_2$. Then the triangle

$$\begin{tikzcd} & \lim_I F \arrow[dl] \arrow[dr] & \\ \lim_J(F\circ\phi_1^{\mathrm{op}}) \arrow[rr, "\lim_J(F\alpha^{\mathrm{op}})"] & & \lim_J(F\circ\phi_2^{\mathrm{op}}) \end{tikzcd}$$

commutes. Dually, if $F\colon I \to \mathcal{C}$ is a diagram, then there exists a canonical morphism $\operatorname{colim}_J(F\circ\phi) \to \operatorname{colim}_I F$ if the colimits exist, and this morphism is natural in ϕ.

***Remark* XV.2.1.7.** — Limits and colimits are formally dual. For a diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$, let $F^{\mathrm{op}}\colon I \to \mathcal{C}^{\mathrm{op}}$ be the dual diagram. Then $(\lim_I F)^{\mathrm{op}} \cong \operatorname{colim}_I F^{\mathrm{op}}$.

***Definition* XV.2.1.8.** — *For a diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$ (resp. $F\colon I \to \mathcal{C}$), we say that the limit* $\lim_I F$ *(resp. colimit* $\operatorname{colim}_I F$*), if it exists, is* finite *or* small *if F is.*

***Construction* XV.2.1.9.** — There is again an (almost) equivalent definition via the Yoneda lemma, but it only works for locally small categories $\mathcal{C}$. We first consider a Set-valued small diagram $F\colon I^{\mathrm{op}} \to \mathsf{Set}$. Then the limit of F is *formally defined* as

$$\lim_I F \coloneqq \operatorname{Hom}_{\operatorname{Fun}(I^{\mathrm{op}},\mathsf{Set})}(\Delta_{\{\mathrm{pt}\}}, F) \tag{XV.2.1.9.1}$$

where $\Delta_{\{\mathrm{pt}\}}\colon I^{\mathrm{op}} \to \mathsf{Set}$ maps everything to the singleton set $\{\mathrm{pt}\}$. Conceptually, every $\alpha \in \lim_I F$ gives a family of elements $\alpha_i(\mathrm{pt}) \in F(i)$ for $i \in \operatorname{Ob}(I)$ that are compatible under the image of F. In other words, chasing an element $x \in X$ through any cone $\beta\colon \Delta_X \to F$, the limit $\lim_I F$ gives all possible choices for $(\beta_i(x))_{i\in\operatorname{Ob}(I)}$.

[XV.6] We say *colimit cone* instead of *colimit cocone* since the latter is a mouthful.

We now extend this definition to general diagrams by means of the Yoneda embedding. Let $F\colon I^{\mathrm{op}} \to \mathcal{C}$ be a small diagram in a locally small category $\mathcal{C}$. Then there is a functor

$$\hat{F}\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathsf{Set}, \qquad X \longmapsto \lim_I(\mathrm{h}^{\mathcal{C}}(X) \circ F) = \mathrm{Hom}_{\mathrm{Fun}(I^{\mathrm{op}},\mathsf{Set})}(\Delta_{\{\mathrm{pt}\}}, \mathrm{Hom}_{\mathcal{C}}(X, F)).$$

Note that any $\alpha \in \lim_I(\mathrm{h}^{\mathcal{C}}(X) \circ F)$ carries the same datum $(\alpha_i(\mathrm{pt})\colon X \to F(i))_{i\in\mathrm{Ob}(I)}$ as a cone $\Delta_X \to F$ over F. Hence, the category of elements $\mathrm{el}(\hat{F})$ is the category of cones over F. By proposition XV.1.6.16, a *limit* of F is a representative of $\hat{F}$, i. e.,

$$\hat{F} = \mathrm{Hom}_{\mathrm{Fun}(I^{\mathrm{op}},\mathsf{Set})}(\Delta_{\{\mathrm{pt}\}}, \mathrm{Hom}_{\mathcal{C}}(-, F)) \cong \mathrm{Hom}_{\mathcal{C}}(-, \lim_I F)$$

where $\mathrm{Hom}_{\mathcal{C}}(X, F) = \mathrm{h}^{\mathcal{C}}(X) \circ F$ is a functor $I^{\mathrm{op}} \to \mathsf{Set}$. By the Yoneda lemma XV.1.6.5, this natural isomorphism gives rise to a natural transformation $\Delta_{\{\mathrm{pt}\}} \to \mathrm{Hom}_{\mathcal{C}}(\lim_I F, F)$ (corresponding to $\mathrm{id}_{\lim_I F}$), which is a *limit cone* $\Delta_{\lim_I F} \to F$ over F.

This general definition recovers the definition in equation (XV.2.1.9.1): for a small diagram $F\colon I^{\mathrm{op}} \to \mathsf{Set}$, plugging in $X = \{\mathrm{pt}\}$ gives a natural bijection

$$\begin{aligned} \mathrm{Hom}_{\mathrm{Fun}(I^{\mathrm{op}},\mathsf{Set})}(\Delta_{\{\mathrm{pt}\}}, F) &\cong \mathrm{Hom}_{\mathrm{Fun}(I^{\mathrm{op}},\mathsf{Set})}(\Delta_{\{\mathrm{pt}\}}, \mathrm{h}^{\mathsf{Set}}(\{\mathrm{pt}\}) \circ F) \\ &\cong \mathrm{Hom}_{\mathsf{Set}}(\{\mathrm{pt}\}, \lim_I F) \\ &\cong \lim_I F \end{aligned}$$

because of the representation $\mathrm{id}_{\mathsf{Set}} \cong \mathrm{h}^{\mathsf{Set}}(\{\mathrm{pt}\})$.

Dually, let $F\colon I \to \mathcal{C}$ be a small diagram in a locally small category $\mathcal{C}$. Then, since $\mathrm{h}^{\mathcal{C}^{\mathrm{op}}} = \mathrm{h}_{\mathcal{C}}$ by remark XV.1.6.3, a representation

$$\check{F} = \lim_I(\mathrm{h}_{\mathcal{C}}(-) \circ F^{\mathrm{op}}) = \mathrm{Hom}_{\mathrm{Fun}(I^{\mathrm{op}},\mathsf{Set})}(\Delta_{\{\mathrm{pt}\}}, \mathrm{Hom}_{\mathcal{C}}(F^{\mathrm{op}}, -)) \cong \mathrm{Hom}_{\mathcal{C}}(\operatorname*{colim}_I F, -) \quad \text{(XV.2.1.9.2)}$$

defines the *colimit* $\mathrm{colim}_I F$ of F and the *colimt cone* $F \to \Delta_{\mathrm{colim}_I F}$ under F.

***Remark* XV.2.1.10.** — Equivalently, we could define the colimit of a small diagram $F\colon I \to \mathcal{C}$ in construction XV.2.1.9 as a representative of the functor

$$\check{F}(X) = (\operatorname*{colim}_I(\mathrm{h}_{\mathcal{C}}(X)^{\mathrm{op}} \circ F))^{\mathrm{op}} \cong \lim_I(\mathrm{h}_{\mathcal{C}}(X) \circ F^{\mathrm{op}})$$

using remark XV.2.1.7. Dually, the same works for limits of a small diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$.

***Example* XV.2.1.11.** — (i) Let $\mathcal{C}$ be a category. Then $\lim_{\mathcal{C}} \mathrm{id}_{\mathcal{C}}$ is the initial object and $\mathrm{colim}_{\mathcal{C}} \mathrm{id}_{\mathcal{C}}$ is the terminal object in $\mathcal{C}$, provided they exist.

(ii) Let t be the terminal object of I. Then for any diagram $F\colon I \to \mathcal{C}$, we have $\mathrm{colim}_I F \cong F(t)$ since $F \to \Delta_{F(t)}$ is already a cocone, and the universality of this cocone is automatic. Dually, t is the initial object of I^{op}, so for any diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$, we have $\lim_I F \cong F(t)$.

***Observation* XV.2.1.12** ((co)limits are adjoint to the constant functor)**.** — Suppose that a category $\mathcal{C}$ admits all limits of shape I with I small. For any natural transformation $\alpha\colon F \to G$ of diagrams $I^{\mathrm{op}} \to \mathcal{C}$, the limit cone over F defines a cone over G by naturality of α. Hence, we obtain a unique morphism $\lim_I F \to \lim_I G$ by the universal property of $\lim_I G$. Thus, there is a functor $\lim_I\colon \mathrm{Fun}(I^{\mathrm{op}}, \mathcal{C}) \to \mathcal{C}$. Dually, there is $\mathrm{colim}_I\colon \mathrm{Fun}(I, \mathcal{C}) \to \mathcal{C}$.

Let $\Delta\colon \mathcal{C} \to \mathrm{Fun}(I^{\mathrm{op}}, \mathcal{C})$ be the diagonal functor. Then there is a bijection

$$\mathrm{Hom}_{\mathrm{Fun}(I^{\mathrm{op}},\mathcal{C})}(\Delta_X, Y) \cong \mathrm{Hom}_{\mathcal{C}}(X, \lim_I Y)$$

natural in X and Y, because each cone $\Delta_X \to Y$ corresponds to a unique morphism $X \to \lim_I Y$ by the universal property. Thus, the functors $\Delta \dashv \lim_I$ are adjoint. Dually, the functors $\mathrm{colim}_I \dashv \Delta$ are adjoint.

XV.2.2. Properties of limits.

***Fact* XV.2.2.1** (uniqueness of limits). — *Given a diagram F, if limits over F (resp. colimits under F) exist, then they are unique up to unique isomorphism.*

Proof. — This follows from the uniqueness of terminal objects in categories, e. g., the categories of cones over a diagram. Alternatively, perform the same proof as in fact XV.1.2.27.

Proof with the Yoneda lemma. — If F is a small diagram, then this follows by uniqueness of representatives (remark XV.1.6.12).

***Definition* XV.2.2.2.** — (i) *A category $\mathcal{C}$ **admits limits** (resp. **admits colimits**) of shape I if every diagram $I^{\mathrm{op}} \to \mathcal{C}$ (resp. $I \to \mathcal{C}$) has a limit (resp. colimit) in $\mathcal{C}$.*

(ii) *A category $\mathcal{C}$ is **complete** (resp. **cocomplete**) if it admits all small limits (resp. small colimits) in $\mathcal{C}$.*

***Fact* XV.2.2.3.** — *Let $(\mathcal{C}_s)_{s\in S}$ be a family of categories admitting limits (resp. colimits) of shape I, indexed by a set S. Then the product $\prod_{s\in S} \mathcal{C}_s$ also admits limits (resp. colimits) of shape I.*

Proof. — Let $F\colon I^{\mathrm{op}} \to \prod_{s\in S} \mathcal{C}_s$ be a diagram of shape I. Then $\pi_s \circ F\colon I^{\mathrm{op}} \to \mathcal{C}_s$ is a diagram in $\mathcal{C}_s$ admitting a limit. One can now check that $\lim_I F$ is $\prod_{s\in S} \lim_I(\pi_s \circ F)$.

***Proposition* XV.2.2.4** (Set is (co)complete). — (i) *For any small diagram $F\colon I^{\mathrm{op}} \to \mathsf{Set}$, its limit is given by*

$$\lim_I F \cong \left\{ (x_i)_{i\in \mathrm{Ob}(I)} \in \prod_{i\in\mathrm{Ob}(I)} F(i) \;\middle|\; F(f)(x_j) = x_i \text{ for all } f \in \mathrm{Hom}_I(i,j) \right\}.$$

(ii) *For any small diagram $F\colon I \to \mathsf{Set}$, its colimit is given by*

$$\operatorname*{colim}_I F \cong \coprod_{i\in\mathrm{Ob}(I)} F(i) \Big/ \sim$$

where the equivalence relation $\sim$ is generated by $x_i \sim x_j \in F(i) \times F(j)$ if $F(f)(x_i) = x_j$ for some $f \in \mathrm{Hom}_I(i,j)$.

Proof. — Since Set is locally small, we use equation (XV.2.1.9.1) as the definition for limits. For each $i \in \mathrm{Ob}(I)$, we have a map

$$\lim_I F = \mathrm{Hom}_{\mathrm{Fun}(I^{\mathrm{op}},\mathsf{Set})}(\Delta_{\{\mathrm{pt}\}}, F) \longrightarrow \mathrm{Hom}_{\mathsf{Set}}(\Delta_{\{\mathrm{pt}\}}(i), F(i)) \cong F(i)$$

for each $i \in \mathrm{Ob}(I)$. This induces an injective map $\lim_I F \to \prod_{i\in\mathrm{Ob}(I)} F(i)$, whose image is precisely the set given in (i).

By equation (XV.2.1.9.2) and (i), we have

$$\begin{aligned}
\mathrm{Hom}_{\mathcal{C}}(\operatorname*{colim}_I F, S) &\cong \lim_I \mathrm{Hom}_{\mathrm{Fun}(I^{\mathrm{op}},\mathsf{Set})}(F, S)\\
&\cong \left\{ (p(i))_i \in \prod_{i\in\mathrm{Ob}(I)} \mathrm{Hom}_{\mathsf{Set}}(F(i), S) \;\middle|\; p(i) = p(j)\circ F(f) \text{ for all } f\colon i \longrightarrow j \right\}\\
&\cong \left\{ p \in \mathrm{Hom}_{\mathsf{Set}}\left(\coprod_{i\in\mathrm{Ob}(I)} F(i), S\right) \;\middle|\; p(x) = p(y) \text{ whenever } x \sim y \right\}\\
&\cong \mathrm{Hom}_{\mathsf{Set}}\left(\coprod_{i\in\mathrm{Ob}(I)} F(i) \Big/ \sim, S\right)
\end{aligned}$$

for any set S, and the isomorphisms are natural in S. Now, (ii) follows from the Yoneda lemma XV.1.6.5. ◻

***Warning* XV.2.2.5.** — The category Set admits all *small* limits, but not all *large* limits. Let $I = \mathrm{Ob}(\mathsf{Set})$ be the discrete category of all sets. Then the limit of the large diagram $F\colon I^{\mathrm{op}} \to \mathsf{Set}$ is the initial object $\lim_I F = \prod_S S = \varnothing$. Indeed, any cone over F with vertex S must, in particular, give a map $S \to \varnothing$, which is only possible if $S = \varnothing$ (there is only one cone over F).

However, if $I = \mathrm{Ob}(\mathsf{Set}) \setminus \{\varnothing\}$ is the discrete category of all non-empty sets, then the limit of the large diagram $F\colon I^{\mathrm{op}} \to \mathsf{Set}$ cannot exist. Otherwise, the limit cone of F gives a non-surjective[XV.7] map $\pi\colon \lim_I F \to \mathcal{P}(\lim_I F)$, say, f misses $S \in \mathcal{P}(\lim_I F)$. By the axiom of global choice, choose an element of each non-empty set except $\mathcal{P}(\lim_I F)$. These choices and $S \in \mathcal{P}(\lim_I F)$ give rise to a cone $\Delta_{\{\mathrm{pt}\}} \to F$ which cannot factor through $\lim_I F$ because π misses S.

This phenomenon is not specific to Set. In general, a category cannot admit all large limits.

***Proposition* XV.2.2.6** (Freyd). — *If a category $\mathcal{C}$ admits products indexed by sets of cardinality $\#\,\mathrm{Mor}(\mathcal{C})$, then it is already **thin**, i. e., there is at most one morphism $X \to Y$ for any $X, Y \in \mathrm{Ob}(\mathcal{C})$.*

Proof. — By way of contradiction, suppose that there are two parallel morphisms $X \rightrightarrows Y$. By the universal property of $\prod_{\mathrm{Mor}(\mathcal{C})} Y$, the hom-set $\mathrm{Hom}_{\mathcal{C}}(X, \prod_{\mathrm{Mor}(\mathcal{C})} Y)$ is in bijection with the class of all families of morphisms $(f_i\colon X \to Y)_{i \in \mathrm{Mor}(\mathcal{C})}$. Since there are two choices for each f_i, we see that the cardinality of $\mathrm{Hom}_{\mathcal{C}}(X, \prod_{\mathrm{Mor}(\mathcal{C})} Y)$ is at least $2^{\#\,\mathrm{Mor}(\mathcal{C})} = \#\mathcal{P}(\mathrm{Mor}(\mathcal{C}))$, but as a subclass of $\mathrm{Mor}(\mathcal{C})$ it can only have cardinality at most $\#\,\mathrm{Mor}(\mathcal{C})$, a contradiction. ◻

Proof using the Yoneda lemma. — Suppose that $\mathcal{C}$ is small. By way of contradiction, suppose that $\#\,\mathrm{Hom}_{\mathcal{C}}(X, Y) \geq 2$. By proposition XV.2.2.8, we have

$$\#\,\mathrm{Mor}(\mathcal{C}) \geq \#\,\mathrm{Hom}_{\mathcal{C}}\left(X, \prod_{\mathrm{Mor}(\mathcal{C})} Y\right) = \prod_{\mathrm{Mor}(\mathcal{C})} \#\,\mathrm{Hom}_{\mathcal{C}}(X, Y) \geq 2^{\#\,\mathrm{Mor}(\mathcal{C})} = \#\mathcal{P}(\mathrm{Mor}(\mathcal{C})),$$

which is absurd. ◻

***Definition* XV.2.2.7.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor.*

(i) *Let $J\colon I^{\mathrm{op}} \to \mathcal{C}$ be a diagram, and suppose that $\lim_I J$ is a limit of J with limit cone $\alpha\colon \Delta_{\lim_I J} \to J$. Then F **preserves** the limit $\lim_I J$ if $F(\lim_I J)$ is a limit of $F \circ J$ with limit cone $F\alpha$ (this is a whiskering $\Delta_{F(\lim_I J)} \to F \circ J$). In particular, the canonical morphism $F(\lim_I J) \xrightarrow{\sim} \lim_I (F \circ J)$ is an isomorphism.*

(ii) *Dually, let $J\colon I \to \mathcal{C}$ be a diagram, and suppose that $\mathrm{colim}_I J$ is a limit of J with colimit cone $\alpha\colon J \to \Delta_{\mathrm{colim}_I J}$. Then F **preserves** the colimit $\mathrm{colim}_I J$ if $F(\mathrm{colim}_I J)$ is a colimit of $F \circ J$ with colimit cone $F\alpha$ (this is a whiskering $F \circ J \to \Delta_{F(\lim_I J)}$). In particular, the canonical morphism $\mathrm{colim}_I (F \circ J) \xrightarrow{\sim} F(\mathrm{colim}_I J)$ is an isomorphism.*

(iii) *The functor F is **continuous** (resp. **cocontinuous**) if it preserves all small limits (resp. small colimits) in $\mathcal{C}$.*

***Proposition* XV.2.2.8** (hom-functors preserve limits). — *Let $\mathcal{C}$ be a locally small category. If a diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$ (resp. $F\colon I \to \mathcal{C}$) has a limit (resp. colimit) in $\mathcal{C}$, then there is an isomorphism*

$$\mathrm{Hom}_{\mathcal{C}}(X, \lim_I F) \cong \lim_{i \in I} \mathrm{Hom}_{\mathcal{C}}(X, F(i)),$$
$$\mathrm{Hom}_{\mathcal{C}}(\operatorname*{colim}_I F, X) \cong \lim_{i \in I} \mathrm{Hom}_{\mathcal{C}}(F(i), X),$$

(XV.7) This is well-known from the theory of cardinalities (Cantor's diagonal argument). For any map $f\colon S \to \mathcal{P}(S)$, consider $T = \{s \in S \mid s \notin f(s)\}$. Then $T \neq f(s)$ for all $s \in S$.

resp., of sets for all $X \in \mathrm{Ob}(\mathcal{C})$ which is natural in X. Note that the right-hand (co)limits are taken in Set.

Proof. — Let $\alpha\colon \Delta_{\lim_I F} \to F$ be the limit cone of $\lim_I F$. We need to show that the cone $\mathrm{Hom}_{\mathcal{C}}(X, \alpha)$ over $\mathrm{Hom}_{\mathcal{C}}(X, F)$ with $\mathrm{Hom}_{\mathcal{C}}(X, \lim_I F)$ at the tip is universal.

Let $\beta\colon \Delta_S \to \mathrm{Hom}_{\mathcal{C}}(X, F)$ be any other cone. Each $s \in S$ gives a family of morphisms $\beta_i(s)\colon X \to F(i)$ compatible under the image of $\mathrm{Hom}_{\mathcal{C}}(X, F)$, i. e., post-composition with morphisms $F(i) \to F(j)$, so these morphisms assemble into a cone $\Delta_X \to F$. By the universal property of $\lim_I F$, all $\beta_i(s)$ factor through a single morphism $X \to \lim_I F$, so we define the map of cones $f\colon S \to \mathrm{Hom}_{\mathcal{C}}(X, \lim_I F)$ by defining $f(s)$ to be that morphism. One easily checks that f is unique such that it induces a map of cones from β to $\mathrm{Hom}_{\mathcal{C}}(X, \alpha)$. Thus, $\mathrm{Hom}_{\mathcal{C}}(X, \lim_I F)$ is a limit.

By a diagram chase in the naturality square, one also verifies naturality (use that the canonical isomorphism is explicitly given by $f \mapsto (\alpha_i \circ f)_{i \in \mathrm{Ob}(I)}$, cf. proposition XV.2.2.4). □

Proof using the Yoneda lemma. — Suppose that the diagram F is small. Then this follows immediately from construction XV.2.1.9 (i. e., by definition). □

***Example* XV.2.2.9.** — Let V be a k-vector space. Then the hom-functor $\mathrm{Hom}_k(V, -)\colon k\text{-}\mathsf{Vect} \to k\text{-}\mathsf{Vect}$ is cocontinuous if and only if V is finite-dimensional. Obviously, it is continuous.

***Proposition* XV.2.2.10** (limits in functor categories are computed pointwise)**.** — *For the functor category* $\mathrm{Fun}(\mathcal{D}, \mathcal{C})$ *(with $\mathcal{D}$ small), let* $\mathrm{ev}_X\colon \mathrm{Fun}(\mathcal{D}, \mathcal{C}) \to \mathcal{C}$, $G \mapsto G(X)$ *be the evaluation functor at* $X \in \mathrm{Ob}(\mathcal{D})$.

Let $F\colon I^{\mathrm{op}} \to \mathrm{Fun}(\mathcal{D}, \mathcal{C})$ *be a diagram such that* $\mathrm{ev}_X \circ F\colon I^{\mathrm{op}} \to \mathcal{C}$ *has a limit for all* $X \in \mathrm{Ob}(\mathcal{D})$. *Then F has a limit in* $\mathrm{Fun}(\mathcal{D}, \mathcal{C})$, *which is described pointwise by the isomorphism*

$$(\lim_I F)(X) \cong \lim_I (\mathrm{ev}_X \circ F) = \lim_{i \in I} F(i)(X)$$

for each $X \in \mathrm{Ob}(\mathcal{D})$ *which is natural in X. In particular,* ev_X *preserves* $\lim_I F$.

The dual statement with colimits also holds.

Proof. — Define $G\colon \mathcal{D} \to \mathcal{C}$, $X \mapsto \lim_I(\mathrm{ev}_X \circ F)$, which is a functor. Indeed, note that we can uncurry F to a bifunctor $F\colon I^{\mathrm{op}} \times \mathcal{D} \to \mathcal{C}$, that is, F is functorial in both entries. If $f\colon X \to Y$ is a morphism in $\mathcal{D}$, then the diagram

$$\begin{tikzcd}
 & G(X) \arrow[dl] \arrow[dr] & \\
F(i)(X) \arrow[rr] \arrow[d, "F(i)(f)"'] & & F(j)(X) \arrow[d, "F(j)(f)"] \\
F(i)(Y) \arrow[rr] & & F(j)(Y)
\end{tikzcd}$$

commutes for all morphisms $i \to j$ in I. Hence, $\Delta_{G(X)} \to \mathrm{ev}_Y \circ F$ is a cone and we obtain a unique morphism $G(X) \to G(Y)$ by the universal property of $G(Y)$. Functoriality follows from a similar argument.

We need to show that G is a limit of F with limit cone $\alpha\colon \Delta_G \to F$. Let $\beta\colon \Delta_H \to F$ be a cone. Then pointwise, β gives rise to cones $\mathrm{ev}_X \beta$ in $\mathcal{C}$ with tip $G(X)$. By the universal property of $G(X)$, we obtain a unique morphism $H(X) \to G(X)$. Moreover, for every morphism $f\colon X \to Y$ in $\mathcal{D}$ the naturality square

$$\begin{tikzcd}
H(X) \arrow[r, "H(f)"] \arrow[d] & H(Y) \arrow[d] \\
G(X) \arrow[r, "G(f)"] & G(Y)
\end{tikzcd}$$

commutes since both compositions give a morphism from the cone $\Delta_{H(X)} \to \mathrm{ev}_Y \circ F$ to the limit cone $\Delta_{G(Y)} \to \mathrm{ev}_Y \circ F$, so they must be equal. Thus, these $H(X) \to G(X)$ assemble into the natural transformation $H \to G$. One checks that this transformation $H \to G$ is unique such that β factorises through it and α. □

Proof with the Yoneda lemma. — Suppose that $\mathcal{C}$ is locally small so that $\mathrm{Fun}(\mathcal{D}, \mathcal{C})$ is locally small. We check that $G\colon \mathcal{D} \to \mathcal{C}$, $X \mapsto \lim_{i \in I} F(i)(X)$ is a functor. Any morphism $f\colon X \to Y$ in $\mathcal{D}$ induces a morphism $F(i)(X) \to F(i)(Y)$ for all $i \in \mathrm{Ob}(I)$. Hence, we obtain a natural transformation (morphism of representations)

$$\mathrm{Hom}_{\mathcal{C}}(-, G(X)) \cong \lim_{i \in I} \mathrm{Hom}_{\mathcal{C}}(-, F(i)(X)) \longrightarrow \lim_{i \in I} \mathrm{Hom}_{\mathcal{C}}(-, F(i)(Y)) \cong \mathrm{Hom}_{\mathcal{C}}(-, G(Y))$$

of functors $\mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$. The isomorphisms above follow from proposition XV.2.2.8 (or by definition assuming I is small). By the Yoneda lemma XV.1.6.5, this induces a unique morphism $G(f)\colon G(X) \to G(Y)$.

For each functor $H\colon \mathcal{D} \to \mathcal{C}$, we have an isomorphism

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{D},\mathcal{C})}(H, G) = \mathrm{Hom}_{\mathrm{Fun}(\mathcal{D},\mathcal{C})}(H, \lim_{i \in I} F(i)(-)) \cong \lim_{i \in I} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{D},\mathcal{C})}(H, F(i)(-))$$

by proposition XV.2.2.8 (or by definition assuming I is small) which is natural in H. Hence, G represents the functor on the right-hand side and thus, $G \cong \lim_I F$ by construction XV.2.1.9. □

***Corollary* XV.2.2.11.** — *For a small category $\mathcal{C}$, the functor categories $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ and $\mathrm{Fun}(\mathcal{C}, \mathsf{Set})$ admit all limits and colimits.*

Proof. — Combine propositions XV.2.2.4 and XV.2.2.10. □

The following statement is very powerful, one of the main reasons why we consider adjoint functors at all.

***Theorem* XV.2.2.12** (right adjoints preserve limits, RAPL). — *Let $L \dashv R$ be adjoints $\mathcal{C} \rightleftarrows \mathcal{D}$. If a diagram $F\colon I^{\mathrm{op}} \to \mathcal{D}$ has a limit in $\mathcal{D}$, then R preserves $\lim_I F$, i. e., we have a canonical isomorphism*

$$R(\lim_I F) \xrightarrow{\sim} \lim_I (R \circ F).$$

The dual statement with L and colimits also holds.

Proof. — Let $\alpha\colon \Delta_{\lim_I F} \to F$ be the limit cone, and we need to show that $R\alpha\colon \Delta_{R(\lim_I F)} \to R \circ F$ is a limit cone over $R \circ F$. Let $\beta\colon \Delta_X \to R \circ F$ be a cone in $\mathcal{C}$. Then there is an adjunct cone $L\beta\colon \Delta_{L(X)} \to F$ over F in $\mathcal{D}$ which uniquely factorises through α. Let $L(X) \to \lim_I F$ be the resulting unique morphism in $\mathcal{D}$. Then its adjunct $X \to R(\lim_I F)$ gives a morphism in $\mathcal{C}$ through which β factorises uniquely. □

Proof with the Yoneda lemma. — Suppose that $\mathcal{C}$ and $\mathcal{D}$ are locally small. In the following we will use proposition XV.2.2.8 or require that I is small. Then we have a chain of natural isomorphisms

$$\begin{aligned}\mathrm{Hom}_{\mathcal{C}}(-, R(\lim_I F)) &\cong \mathrm{Hom}_{\mathcal{D}}(L(-), \lim_I F) \cong \lim_I \mathrm{Hom}_{\mathcal{D}}(L(-), F) \\ &\cong \lim_I \mathrm{Hom}_{\mathcal{C}}(-, R \circ F) \cong \mathrm{Hom}_{\mathcal{C}}(-, \lim_I (R \circ F)).\end{aligned}$$

By the Yoneda lemma XV.1.6.5, we obtain an isomorphism $f\colon R(\lim_I F) \xrightarrow{\sim} \lim_I (R \circ F)$, which is given by evaluating the above chain at $R(\lim_I F)$ and tracing the identity. For each $i \in \mathrm{Ob}(I)$, we obtain by naturality a commutative square

$$\begin{tikzcd}
\mathrm{Hom}_{\mathcal{C}}(R(\lim_I F), R(\lim_I F)) \arrow[r, "\sim"] \arrow[d, "R(\alpha_i)\circ -"'] & \mathrm{Hom}_{\mathcal{C}}(R(\lim_I F), \lim_I(R\circ F)) \arrow[d] \\
\mathrm{Hom}_{\mathcal{C}}(R(\lim_I F), R(F(i))) \arrow[r, equal] & \mathrm{Hom}_{\mathcal{C}}(R(\lim_I F), R(F(i)))
\end{tikzcd}$$

where $\alpha\colon \Delta_{\lim_I F} \to F$ is the limit cone. Chasing the identity of $R(\lim_I F)$, we see that f must be the canonical morphism. ◻

***Corollary* XV.2.2.13** (equivalences of categories preserve and create limits). — *Let $E\colon \mathcal{C} \rightsquigarrow \mathcal{D}$ be an equivalence of categories, and let $F\colon I^{\mathrm{op}} \to \mathcal{C}$ be a diagram. Then $\lim_I F$ exists in $\mathcal{C}$ if and only if $\lim_I(E\circ F)$ exists in $\mathcal{D}$, in which case we have an isomorphism*

$$E(\lim_I F) \xrightarrow{\sim} \lim_I(E\circ F).$$

The dual statement for colimits also holds.

Proof. — By proposition XV.1.8.7, we can promote E to an adjoint equivalence, in which case E is a right adjoint of a quasi-inverse $\mathcal{D} \rightsquigarrow \mathcal{C}$. If $\lim_I F$ exists, then the preservation statement follows from theorem XV.2.2.12. Conversely, suppose that $\lim_I(E\circ F)$ exists. Since E is essentially surjective, there exists $X \in \mathrm{Ob}(\mathcal{C})$ such that $\lim_I(E\circ F) \cong E(X)$, which is also a limit of $E\circ F$. Since E is fully faithful, we can lift the limit cone over $E\circ F$ with tip $E(X)$ uniquely to a cone over F with tip X. Again by lifting along E, we argue that $X \cong \lim_I F$ is the limit of F. ◻

***Proposition* XV.2.2.14** (limits commute). — *Let $F\colon I^{\mathrm{op}} \times J^{\mathrm{op}} \to \mathcal{C}$ be a bifunctor with small index categories I and J, which induces functors $F_I\colon J^{\mathrm{op}} \to \mathrm{Fun}(I^{\mathrm{op}},\mathcal{C})$ and $F_J\colon I^{\mathrm{op}} \to \mathrm{Fun}(J^{\mathrm{op}},\mathcal{C})$. Then there are isomorphisms*

Remove smallness conditions.

$$\lim_{i\in I}\lim_{j\in J} F(i,j) = \lim_I \lim_J F_I \cong \lim_{I\times J} F \cong \lim_J \lim_I F_J = \lim_{j\in J}\lim_{i\in I} F(i,j)$$

if the limits on the left-hand and right-hand side exist. In particular, F has a limit.

The dual statement for colimits also holds.

Proof. — We check that $\lim_I \lim_J F_I$ has the universal property of the limit of F; the other isomorphism follows by symmetry. Let $\alpha\colon \Delta_X \to F$ be a cone. Then we obtain a cone $\alpha_I\colon (\Delta_X)_I \to F_I$ indexed by J^{op} where $(\Delta_X)_I\colon J^{\mathrm{op}} \to \mathrm{Fun}(I^{\mathrm{op}},\mathcal{C})$, $j \mapsto \Delta_X^{I^{\mathrm{op}}}$ (see example XV.1.5.3). By the universal property, α_I uniquely factorises through a natural transformation $\Delta_X^{I^{\mathrm{op}}} \to \lim_J F_I$, which again is a cone indexed by I^{op}. Again by the universal property, this uniquely factorises through a morphism $X \to \lim_I \lim_J F_I$. Uniqueness can be easily checked. ◻

Maybe add proof with Yoneda.

***Remark* XV.2.2.15.** — There is a much more elegant proof if $\mathcal{C}$ admits all limits of shape I and J with I and J small. Then there is a functor $\lim_I\colon \mathrm{Fun}(I^{\mathrm{op}},\mathcal{C}) \to \mathcal{C}$. Let $F\colon I^{\mathrm{op}} \times J^{\mathrm{op}} \to \mathcal{C}$ be a bifunctor, which induces $F_I\colon J^{\mathrm{op}} \to \mathrm{Fun}(I^{\mathrm{op}},\mathcal{C})$. Then F_I has a limit since $\mathrm{Fun}(I^{\mathrm{op}},\mathcal{C})$ admits all limits of shape J by proposition XV.2.2.10. Since $\lim_I$ is is a right adjoint to the diagonal functor Δ, it preserves limits by theorem XV.2.2.12. Thus, we have a canonical isomorphism

$$\lim_I(\lim_J F_I) \cong \lim_J(\lim_I \circ F_I) \cong \lim_J(\lim_I F_J)$$

since $(\lim_I \circ F_I)(j) = \lim_I(\mathrm{ev}_j \circ F_J) \cong (\lim_I F_J)(j)$ for all $j \in \mathrm{Ob}(J)$ as limits in functor categories are computed pointwise (proposition XV.2.2.10).

***Fact* XV.2.2.16.** — *Let $\mathcal{C}$ be a category that admits all colimits (resp. small colimits) of shape I, and let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor preserving such colimits. Then for any $A \in \mathrm{Ob}(\mathcal{D})$, the comma category $F \downarrow A$ admits all colimits (resp. small colimits) of shape I. In particular,* $\mathrm{For}\colon F \downarrow A \to \mathcal{C}$ *preserves such colimits.*

The dual statement on limits and $A \downarrow F$ also holds true.

Proof. — Let $G\colon I \to F \downarrow A$ be a diagram. We obtain a diagram $\mathrm{For} \circ G\colon I \to \mathcal{C}$ which has a colimit $\mathrm{colim}_I(\mathrm{For} \circ G)$. Since F preserves colimits, we obtain a morphism $F(\mathrm{colim}_I(\mathrm{For} \circ G)) \to A$. It is now easy to see that it satisfies the universal property of $\mathrm{colim}_I G$. ∎

XV.2.3. Special (co)limits.

***Definition* XV.2.3.1.** — *Let $I = \mathsf{0}$ be the empty category. The limit over the diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$ is the **terminal object** of $\mathcal{C}$. Dually, the colimit over the diagram $F\colon I \to \mathcal{C}$ is the **initial object** of $\mathcal{C}$.*

***Definition* XV.2.3.2.** — *Let I be a discrete category and let $(X_i)_{i \in \mathrm{Ob}(I)}$ be a family of objects in $\mathcal{C}$. The limit over a diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$, $i \mapsto X_i$ is the **product** of $(X_i)_{i\in\mathrm{Ob}(I)}$. We write $\prod_{i\in I} X_i \coloneqq \lim_I F$ and it comes with projection morphisms $\pi_i\colon \prod_{i\in I} X_i \to X_i$ for each $i \in I$. The universal property is visualised by the diagram*

$$\begin{tikzcd} & T \arrow[dl] \arrow[d, dashed, "\exists!"] \arrow[dr] & \\ X_i & \prod_{i\in I} X_i \arrow[l, "\pi_i"'] \arrow[r, "\pi_j"'] & X_j. \end{tikzcd}$$

*Dually, the colimit under a diagram $F\colon I \to \mathcal{C}$, $i \mapsto X_i$ is the **coproduct** of the objects $(F(i))_{i\in\mathrm{Ob}(I)}$. We write $\coprod_{i\in I} X_i \coloneqq \mathrm{colim}_I F$ and it comes with inclusion morphisms $\iota_i\colon X_i \to \coprod_{i\in I} X_i$ for each $i \in I$. The universal property is visualised by the diagram*

$$\begin{tikzcd} X_i \arrow[r, "\iota_i"] \arrow[dr] & \coprod_{i\in I} X_i \arrow[d, dashed, "\exists!"] & X_j \arrow[l, "\iota_j"'] \arrow[dl] \\ & T. & \end{tikzcd}$$

***Notation* XV.2.3.3.** — If $I = \{1, \dots, n\}$ is finite, then we usually write $X_1 \times \dots \times X_n \coloneqq \prod_{i\in I} X_i$ and $X_1 \amalg \dots \amalg X_n \coloneqq \coprod_{i\in I} X_i$.

***Warning* XV.2.3.4.** — In general, projections are not generally epimorphisms, not even in Set. If $(X_i)_{i\in I}$ is a family such that one but not all X_i is empty, then $\prod_{i\in I} X_i = \varnothing$ by the same argument as in warning XV.2.2.5. However, if all X_i are non-empty, then $\prod_{i\in I} X_i$ is the Cartesian product and hence all projections are surjective, i. e., epimorphisms (example XV.1.2.16).

***Remark* XV.2.3.5.** — By proposition XV.2.2.8, for any locally small category $\mathcal{C}$ and a family of objects $(X_i)_{i\in I}$ of $\mathcal{C}$, we have bijections

$$\mathrm{Hom}_{\mathcal{C}}\left(Y, \prod_{i\in I} X_i\right) \cong \prod_{i\in I} \mathrm{Hom}_{\mathcal{C}}(Y, X_i), \qquad \mathrm{Hom}_{\mathcal{C}}\left(\coprod_{i\in I} X_i, Y\right) \cong \prod_{i\in I} \mathrm{Hom}_{\mathcal{C}}(X_i, Y)$$

if the product and coproduct of $(X_i)_{i\in I}$ in $\mathcal{C}$ exist.

Definition* XV.2.3.6.** — *Let $I = \{\bullet \leftleftarrows \bullet\}$ be a parallel pair of arrows. The limit over the diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$ is the* ***equaliser *of a pair of morphisms (f, g) in $\mathcal{C}$, for which we write $\mathrm{Eq}(f, g) := \lim_I F$. The universal property is visualised by the diagram*

$$\begin{tikzcd} T \arrow[d, dashed, "\exists!"'] \arrow[dr] & & \\ \mathrm{Eq}(f,g) \arrow[r] & X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y. \end{tikzcd}$$

Dually, the colimit over the diagram $F\colon I \to \mathcal{C}$ is the ***coequaliser*** *of a pair of morphisms (f, g) in $\mathcal{C}$, for which we write $\mathrm{Coeq}(f, g) := \mathrm{colim}_I F$. The universal property is visualised by the diagram*

$$\begin{tikzcd} X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \arrow[r] \arrow[dr] & \mathrm{Coeq}(f,g) \arrow[d, dashed, "\exists!"] \\ & & T. \end{tikzcd}$$

***Example* XV.2.3.7.** — For $\mathcal{C} = \mathsf{Set}$ and two maps $f, g\colon X \rightrightarrows Y$, we have $\mathrm{Eq}(f, g) = \{x \in X \mid f(x) = g(x)\}$ and $\mathrm{Coeq}(f, g) = Y/\sim$ where $y_1 \sim y_2$ if there exists some $x \in X$ with $f(x) = y_1$ and $g(x) = y_2$.

***Remark* XV.2.3.8.** — By proposition XV.2.2.8, for any locally small category $\mathcal{C}$ and parallel morphisms $f, g\colon X \rightrightarrows Y$, we have bijections

$$\begin{aligned} \mathrm{Hom}_{\mathcal{C}}(Z, \mathrm{Eq}(f, g)) &\cong \{u \in \mathrm{Hom}_{\mathcal{C}}(Z, X) \mid f \circ u = g \circ u\}, \\ \mathrm{Hom}_{\mathcal{C}}(\mathrm{Coeq}(f, g), Z) &\cong \{u \in \mathrm{Hom}_{\mathcal{C}}(Y, Z) \mid u \circ f = u \circ g\}. \end{aligned}$$

if the equaliser and coequaliser of (f, g) in $\mathcal{C}$ exist.

***Fact* XV.2.3.9.** — *Let $f, g\colon X \rightrightarrows Y$ be two parallel morphisms in a category.*

(i) *If the equaliser $\mathrm{Eq}(f, g)$ exists, then the canonical morphism $\mathrm{Eq}(f, g) \to X$ is an monomorphism.*

(ii) *Dually, if the coequaliser $\mathrm{Coeq}(f, g)$ exists, then the canonical morphism $Y \to \mathrm{Coeq}(f, g)$ is an epimorphism.*

Proof. — Straightforward by definition and the universal property of (co)equalisers. ∎

***Lemma* XV.2.3.10.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor.*

(i) *Suppose that $\mathcal{C}$ admits and F preserves equalisers (resp. coequalisers). If F is conservative, then F is faithful.*

(ii) *Conversely, suppose that morphisms in $\mathcal{C}$ that are monomorphisms and epimorphisms are isomorphisms. If F is faithful, then F is conservative.*

Proof. — Under the assumptions of (i) with kernels, let $f, g\colon X \rightrightarrows Y$ be morphisms with $F(f) = F(g)$. Let $u\colon \mathrm{Eq}(f, g) \to X$ be the canonical morphism. Then $F(\mathrm{Eq}(f, g)) \cong \mathrm{Eq}(F(f), F(g))$, whence $F(u)$ is an isomorphism. Since F is conservative, u is an isomorphism. By definition, $f \circ u = g \circ u$, so $f = g$, that is, F is faithful. The statement with cokernels follows dually.

Under the assumptions of (ii), let f be a morphism such that $F(f)$ is an isomorphism. By lemma XV.1.3.9 (F reflects monomorphisms and epimorphisms), we see that f is a monomorphism and epimorphism. Hence, f is an isomorphism, i. e., F is conservative. ∎

***Definition* XV.2.3.11.** — *Let $I = \{\bullet \leftarrow \bullet \rightarrow \bullet\}$ be a so-called* span. *The limit of a diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$ is the* **pullback** *or* **fibre product**(XV.8) *of the two morphsims $f\colon X \to Z$ and $g\colon Y \to Z$ in $\mathcal{C}$. We write $X \times_Z Y := \lim_I F$ (although f and g cannot be neglected). The universal property is visualised by the diagram*

$$\begin{tikzcd} T \arrow[dr, dashed, "\exists!"] \arrow[drr, bend left] \arrow[ddr, bend right] & & \\ & X \times_Z Y \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner", very near start] & X \arrow[d] \\ & Y \arrow[r] & Z. \end{tikzcd}$$

Dually, the colimit of a diagram $F\colon I \to \mathcal{C}$ is the **pushout** *of two morphisms $X \leftarrow Z \to Y$, for which we write $X \amalg_Z Y := \operatorname{colim}_I F$. The universal property is visualised by the diagram*

$$\begin{tikzcd} Z \arrow[r] \arrow[d] & X \arrow[d] \arrow[ddr, bend left] & \\ Y \arrow[r] \arrow[drr, bend right] & X \amalg_Z Y \arrow[ul, phantom, "\ulcorner", very near start] \arrow[dr, dashed, "\exists!"] & \\ & & T. \end{tikzcd}$$

***Definition* XV.2.3.12.** — *A commutative square*

$$\begin{tikzcd} Y \arrow[r] \arrow[d] & X_0 \arrow[d] \\ X_1 \arrow[r] & Z \end{tikzcd}$$

in $\mathcal{C}$ is a:

(i) ***Cartesian square*** *if $Y \cong X_0 \times_Z X_1$ is a pullback;*
(ii) ***co-Cartesian square*** *if $Z \cong X_0 \amalg_Z X_1$ is a pushout.*

***Remark* XV.2.3.13.** — Let $X \to Z \leftarrow Y$ be morphisms. Then the product $X \times Y$ induces morphisms $X \times Y \rightrightarrows Z$. By definition of pullbacks, $X \times_Z Y \to X \times Y$ is the equaliser of $X \times Y \rightrightarrows Z$.

Dually, for morphisms $X \leftarrow Z \to Y$, the pushout $X \amalg_Z Y$ is the coequaliser of $Z \rightrightarrows X \amalg Y$.

***Remark* XV.2.3.14.** — We can generalise pullbacks and pushouts in a category $\mathcal{C}$ via (co)products in (co)slice categories.

(i) Let Z be an object of $\mathcal{C}$, and recall that $\mathcal{C}/Z$ is the slice category whose objects are morphisms $X \to Z$. Then the *pullback* or *fibre product* of a family of morphisms $(X_i \to Z)_{i\in I}$ indexed by a set I is $\mathrm{For}(\prod_{i\in I} X_i \to Z)$. Here, $\mathrm{For}\colon \mathcal{C}/Z \to \mathcal{C}$ is the forgetful functor. We recover the usual pullback if $I = \{0, 1\}$.

(ii) Dually, the *pushout* of a family of morphisms $(Z \to X_i)_{i\in I}$ in $\mathcal{C}$ indexed by a set I is $\mathrm{For}(\coprod_{i\in I} Z \to X_i)$ where $\mathrm{For}\colon Z/\mathcal{C} \to \mathcal{C}$ is the forgetful functor.

***Lemma* XV.2.3.15** (pullbacks preserve monomorphisms)**.** — (i) *Let*

$$\begin{tikzcd} X \times_Z Y \arrow[r] \arrow[d, dashed, "f^*(g)"'] \arrow[dr, phantom, "\lrcorner", very near start] & X \arrow[d, "g"] \\ Y \arrow[r, "f"] & Z \end{tikzcd}$$

(XV.8) Namesake is the special pullback in Top of a continuous map $f\colon X \to Y$ along the inclusion of a point $\{y\} \hookrightarrow Y$. Then $X \times_Y \{y\}$ is the fibre $f^{-1}(y)$.

be a pullback square. If g is a monomorphism (resp. isomorphism), then $f^*(g)$ is.

(ii) *Dually, let*

$$\begin{tikzcd} Z \arrow[r, "f"] \arrow[d, "g"'] & X \arrow[d, dashed, "f_*(g)"] \\ Y \arrow[r] \arrow[ur, phantom, "\ulcorner"] & X \amalg_Z Y \end{tikzcd}$$

be a pushout square. If g is an epimorphism (resp. isomorphism), then $f_*(g)$ is.

Proof. — We denote $X \times_Z Y \to X$ by $g^*(f)$. Firstly, suppose that g is a monomorphism. Let $h_1, h_2 \colon T \rightrightarrows X \times_Z Y$ be morphisms such that $f^*(g) \circ h_1 = f^*(g) \circ h_2$. Since g is a monomorphism, we also have $g^*(f) \circ h_1 = g^*(f) \circ h_2$. The universal property of $X \times_Z Y$ implies $h_1 = h_2$.

Now, suppose that g is an isomorphism. Then

$$\begin{tikzcd} Y \arrow[r, "g^{-1} \circ f"] \arrow[d, "\mathrm{id}_Y"'] \arrow[dr, phantom, "\lrcorner", very near start] & X \arrow[d, "g"] \\ Y \arrow[r, "f"] & Z \end{tikzcd}$$

is also a pullback square. Hence, the canonical morphism $f^*(g) \colon X \times_Z Y \to Y$ induced by the universal property of Y is an isomorphism. ∎

***Lemma* XV.2.3.16** (pasting of pullbacks). — *Consider a commutative diagram*

$$\begin{tikzcd} X \arrow[r] \arrow[d] & Y \arrow[r] \arrow[d] & Z \arrow[d] \\ X' \arrow[r] & Y' \arrow[r] & Z'. \end{tikzcd}$$

(i) *Suppose that the right-hand square is Cartesian. Then the left-hand square is Cartesian if and only if the outer square is.*

(ii) *Suppose that the left-hand square is co-Cartesian. Then the right-hand square is co-Cartesian if and only if the outer square is.*

Proof. — Consider the diagram

$$\begin{tikzcd} T \arrow[dr, dashed, "f"'] \arrow[drr, dashed, "g"] \arrow[drrr, dashed, bend left, "h"] \arrow[ddr, bend right] & & & \\ & X \arrow[r] \arrow[d] & Y \arrow[r] \arrow[d] & Z \arrow[d] \\ & X' \arrow[r] & Y' \arrow[r] & Z'. \end{tikzcd}$$

Suppose that the left-hand square is a pullback and suppose that h exists such that everything commutes. Then the universal property of Y gives a unique g, and the universal property the left-hand square gives a unique f. Conversely, suppose that the outer square is a pullback and suppose that g (and hence h) exists such that everything commutes. Then the universal property of the outer square gives a unique f, and the universal property of Y shows that the triangle with morphisms f and g commutes. ∎

***Definition* XV.2.3.17.** — *Let $f \colon X \to Y$ be a morphism.*

(i) *Suppose that the pullback $X \times_Y X$ induced by f exists, and let $\pi_1, \pi_2 \colon X \times_Y X \rightrightarrows X$ be the projections. The canonical morphism $\Delta_f \colon X \to X \times_Y X$ such that $\pi_1 \circ \Delta_f = \pi_2 \circ \Delta_f = \mathrm{id}_X$ is the* ***diagonal morphism*** *of f. Visually, we have the commutative diagram*

$$\begin{tikzcd}
X \arrow[dr, dashed, "\Delta_f"] \arrow[drr, bend left, "\mathrm{id}_X"] \arrow[ddr, bend right, "\mathrm{id}_X"'] & & \\
& X \times_Y X \arrow[r, "\pi_1"] \arrow[d, "\pi_2"'] \arrow[dr, phantom, "\lrcorner", very near start] & X \arrow[d, "f"] \\
& X \arrow[r, "f"] & Y.
\end{tikzcd}$$

(ii) *Dually, suppose that the pushout $Y \amalg_X Y$ induced by f exists, and let $\iota_1, \iota_w \colon Y \rightrightarrows Y \amalg_X Y$ be the inclusions. The canonical morphism $\Delta^f \colon Y \amalg_X Y \to Y$ such that $\Delta^f \circ \iota_1 = \Delta^f \circ \iota_2 = \mathrm{id}_Y$ is the* ***codiagonal morphism*** *of f. Visually, we have the commutative diagram*

$$\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "f"'] & Y \arrow[d, "\iota_1"] \arrow[ddr, bend left, "\mathrm{id}_Y"] & \\
Y \arrow[r, "\iota_2"] \arrow[drr, bend right, "\mathrm{id}_Y"'] & Y \amalg_X Y \arrow[dr, dashed, "\Delta^f"] \arrow[ul, phantom, "\ulcorner", very near start] & \\
& & Y.
\end{tikzcd}$$

Skip definition after [KS06, Not. 2.2.8].

XV.2.4. Existence of limits.

In fact, remark XV.2.3.13 holds in full generality, i. e., we can construct any (co)limit from (co)products and (co)equalisers.

XV.2.4.1. — Recall that the arrow category $\mathrm{Arr}(\mathcal{C})$ of a category $\mathcal{C}$ comes with the domain and codomain maps $\mathrm{s}, \mathrm{t} \colon \mathrm{Mor}(\mathcal{C}) \to \mathrm{Ob}(\mathcal{C})$ given by $(X \to Y) \mapsto X$ and $(X \to Y) \mapsto Y$, resp.

***Theorem* XV.2.4.2** (existence of limits). — (i) *Let $F \colon I^{\mathrm{op}} \to \mathcal{C}$ be a diagram. Suppose that the products $\prod_{i \in \mathrm{Ob}(I)} F(i)$ and $\prod_{f \in \mathrm{Mor}(I)} F(\mathrm{t}(f))$ exist in $\mathcal{C}$. Then F has a limit if and only if the diagram*

$$\prod_{i \in \mathrm{Ob}(I)} F(i) \underset{\prod_{f \in \mathrm{Mor}(I)} \pi_{\mathrm{s}(f)}}{\overset{\prod_{f \in \mathrm{Mor}(I)} (F(f) \circ \pi_{\mathrm{t}(f)})}{\rightrightarrows}} \prod_{f \in \mathrm{Mor}(I)} F(\mathrm{t}(f)) \tag{XV.2.4.2.1}$$

has an equaliser. In this case, $\lim_I F$ is the equaliser.

(ii) *Dually, let $F \colon I \to \mathcal{C}$ be a diagram. Suppose that the coproducts $\coprod_{i \in \mathrm{Ob}(I)} F(i)$ and $\coprod_{f \in \mathrm{Mor}(I)} F(\mathrm{s}(f))$ exist in $\mathcal{C}$. Then F has a colimit if and only if the diagram*

$$\coprod_{f \in \mathrm{Mor}(I)} F(\mathrm{s}(f)) \underset{\coprod_{f \in \mathrm{Mor}(I)} \iota_{\mathrm{t}(f)}}{\overset{\coprod_{f \in \mathrm{Mor}(I)} (F(f) \circ \iota_{\mathrm{s}(f)})}{\rightrightarrows}} \coprod_{i \in \mathrm{Ob}(I)} F(i)$$

has a coequaliser. In this case, $\mathrm{colim}_I F$ is the coequaliser.

Proof. — We first construct diagram (XV.2.4.2.1). For each $f \in \mathrm{Mor}(I)$, the following morphisms exist:

$$\prod_{i\in\mathrm{Ob}(I)} F(i) \xrightarrow{\pi_{\mathrm{t}(f)}} F(\mathrm{t}(f)) \xrightarrow{F(f)} F(\mathrm{s}(f)), \qquad \prod_{i\in\mathrm{Ob}(I)} F(i) \xrightarrow{\pi_{\mathrm{s}(f)}} F(\mathrm{s}(f)).$$

By the universal property of $\prod_{f\in\mathrm{Mor}(I)} F(\mathrm{t}(f))$, we obtain the morphisms into $\prod_{f\in\mathrm{Mor}(I)} F(\mathrm{t}(f))$.

Let $X \to \prod_{i\in\mathrm{Ob}(I)} F(i)$ be a morphism giving a cone over diagram (XV.2.4.2.1). By construction, we obtain morphisms $\alpha_j : X \to \prod_{i\in\mathrm{Ob}(I)} F(i) \twoheadrightarrow F(j)$ for each $j \in \mathrm{Ob}(I)$ such that $F(f) \circ \alpha_j = \alpha_i$ for each morphism $f : i \to j$ of I. Thus, the α_j constitute a cone $\alpha : \Delta_X \to F$ over F. Conversely, if $\alpha : \Delta_X \to F$ is a cone over F, then the components $\alpha_i : X \to F(i)$ for each $i \in \mathrm{Ob}(I)$ induce a morphism $X \to \prod_{i\in\mathrm{Ob}(I)} F$, and naturality of α shows that this morphism gives a cone over diagram (XV.2.4.2.1). The statement follows. ◻

Proof with the Yoneda lemma. — Assume that I is small. If $\mathcal{C} = \mathsf{Set}$, then this follows from the explicit description of limits in proposition XV.2.2.4 and example XV.2.3.7. Hence, for general locally compact categories $\mathcal{C}$ the statement is true for $\mathrm{Hom}_{\mathcal{C}}(X, F) : I^{\mathrm{op}} \to \mathsf{Set}$ and

$$\prod_{i\in\mathrm{Ob}(I)} \mathrm{Hom}_{\mathcal{C}}(X, F(i)) \underset{\prod_{f\in\mathrm{Mor}(I)}(\pi_{\mathrm{s}(f)} \circ -)}{\overset{\prod_{f\in\mathrm{Mor}(I)}(F(f) \circ \pi_{\mathrm{t}(f)} \circ -)}{\rightrightarrows}} \prod_{f\in\mathrm{Mor}(I)} \mathrm{Hom}_{\mathcal{C}}(X, F(\mathrm{t}(f)))$$

for all $X \in \mathrm{Ob}(\mathcal{C})$ and natural in X. In particular, the equaliser of the above diagram is canonically isomorphic to $\mathrm{Hom}_{\mathcal{C}}(X, \lim_I F)$. By the uniqueness of representatives (remark XV.1.6.12), the statement is true for F and diagram (XV.2.4.2.1). ◻

***Corollary* XV.2.4.3.** — (i) *A category admits all (co)limits (resp. is (co)complete) if and only if it admits all (co)products (resp. small (co)products) and all (co)equalisers.*

(ii) *A functor preserves all (co)limits (resp. is (co)continuous) if and only if it preserves all (co)products (resp. small (co)products) and all (co)equalisers.*

***Corollary* XV.2.4.4.** — *For a category, the following statements are equivalent:*

(i) *It admits all finite (co)limits.*
(ii) *It admits a terminal/initial object, all binary (co)products and all (co)equalisers.*
(iii) *It admits a terminal/initial object and all pullbacks/pushouts.*

Proof. — The equivalence of (i) and (ii) is clear since terminal objects are empty products.

For the equivalence of (ii) and (iii), suppose that the category $\mathcal{C}$ admits a terminal object T. Then binary products $Z_1 \times Z_2$ of $Z_1, Z_2 \in \mathrm{Ob}(\mathcal{C})$ and equalisers $\mathrm{Eq}(f, g)$ of morphisms $f, g : X \rightrightarrows Y$ in $\mathcal{C}$ are pullbacks

$$\begin{tikzcd} Z_1 \times Z_2 \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner"] & Z_1 \arrow[d] \\ Z_2 \arrow[r] & T, \end{tikzcd} \qquad \begin{tikzcd} \mathrm{Eq}(f,g) \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner"] & X \arrow[d, "f\times g"] \\ Y \arrow[r, "\Delta_Y"] & Y \times Y, \end{tikzcd}$$

whose universal properties can be verified easily. Conversely, a pullback can be written as an equaliser of a product of morphisms, see remark XV.2.3.13. ◻

XV.2.5. Base change.

***Definition* XV.2.5.1.** — *Let $\mathcal{C}$ be a category admitting fibre products.*

(i) *For any morphism $f\colon Y \to Z$ in $\mathcal{C}$, the* ***base change functor*** $f^*\colon \mathcal{C}/Z \to \mathcal{C}/Y$ *induced by f is defined by taking the fibre product of $X \to Z$ along f to obtain $X \times_Z Y \to Y$, i. e., we have the pullback square*

$$\begin{tikzcd} X \times_Z Y \arrow[r] \arrow[d, dashed, "f^*(p)"'] \arrow[dr, phantom, "\lrcorner", very near start] & X \arrow[d, "p"] \\ Y \arrow[r, "f"] & Z. \end{tikzcd}$$

(ii) *A colimit of a diagram $F\colon I \to \mathcal{C}$ is* ***stable under base change*** *or* ***universal*** *if all base change functors preserve that colimit. In other words, we have the canonical isomorphism*

$$\operatorname*{colim}_{i \in I}(F(i) \times_Z Y) \xrightarrow{\sim} \Big(\operatorname*{colim}_{i \in I} F(i)\Big) \times_Z Y$$

for any morphisms $Y \to Z$ and $\operatorname{colim}_{i \in I} F(i) \to Z$ in $\mathcal{C}$.

***Lemma* XV.2.5.2.** — *Small colimits in* Set *are stable under base change.*

Proof. — Let Z be a set and consider a map $Y \to Z$ of sets. By proposition XV.2.2.4, Set/Z admits all binary products (pullbacks in Set), so the functor $- \times_Z Y\colon \mathsf{Set}/Z \to \mathsf{Set}/Z$ is well-defined. Define the functor

$$\mathcal{H}om_Z(Y, -)\colon \mathsf{Set}/Z \longrightarrow \mathsf{Set}/Z, \qquad \mathcal{H}om_Z(Y, W) := \coprod_{z \in Z} \operatorname{Hom}_{\mathsf{Set}}(Y_z, W_z)$$

where Y_z is the fibre of $z \in Z$ along $Y \to Z$. Note that $\mathcal{H}om_Z(Y, W)$ is a set over Z by sending the component $\operatorname{Hom}_{\mathsf{Set}}(Y_z, W_z)$ to z, and $\mathcal{H}om_Z(Y, -)$ is functorial by mapping between components of the disjoint union.

These two functors are adjoint, i. e., there is an adjunction isomorphism

$$\operatorname{Hom}_{\mathsf{Set}/Z}(X \times_Z Y, W) \cong \operatorname{Hom}_{\mathsf{Set}/Z}(X, \mathcal{H}om_Z(Y, W)). \tag{XV.2.5.2.1}$$

Indeed, any map $A \to B$ of sets over Z restricts on fibres to a map $A_z \to B_z$ for any $z \in Z$, and conversely, any family $(A_z \to B_z)_{z \in Z}$ of maps defines a map $A \to B$ of sets over Z. Hence, we only need to show the adjunction isomorphism

$$\operatorname{Hom}_{\mathsf{Set}}(X_z \times Y_z, W_z) \cong \operatorname{Hom}_{\mathsf{Set}}(X_z, \operatorname{Hom}_{\mathsf{Set}}(Y_z, W_z))$$

for each $z \in Z$, which is example XV.1.7.2. Taking the product over all $z \in Z$ gives equation (XV.2.5.2.1). Thus, $- \times_Z Y$ is left adjoint and preserves colimits by theorem XV.2.2.12. ◻

***Corollary* XV.2.5.3.** — *For small categories $\mathcal{C}$, small colimits in* $\operatorname{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ *are stable under base change.*

Proof. — This follows since (co)limits in $\operatorname{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ are computed pointwise (proposition XV.2.2.10). For example, we have $(X \times_Z Y)(U) = X(U) \times_{Z(U)} Y(U)$ for all $U \in \operatorname{Ob}(\mathcal{C})$. ◻

***Counterexample* XV.2.5.4.** — In $\mathbb{Z}$-Mod, not even finite colimits are stable under base change. Recall that the trivial module 0 is the empty colimit (the initial object). If finite colimits were stable under base change, we would have $0 \times M \cong 0$ for all $\mathbb{Z}$-modules M (the colimit on the right-hand side is taken over the empty index category). In fact, for any morphism $f\colon M \to 0$ the diagram

$$\begin{tikzcd} & M \arrow[dl] \arrow[d, dashed] \arrow[dr, "\mathrm{id}_M"] & \\ 0 & 0 \times M \cong 0 \arrow[l, two heads] \arrow[r, two heads] & M \end{tikzcd}$$

commutes by the universal property of products. Since 0 is initial, the dashed morphism is f and admits a right inverse $0 \to M$. Thus, $f\colon M \xrightarrow{\sim} 0$ is an isomorphism and $\mathbb{Z}$-$\mathsf{Mod} \cong \mathbf{0}$.

XV.2.6. Kan extensions.

As before, Kan extensions are ubiquitous.

> *All Concepts Are Kan Extensions* [section title].
>
> The notion of Kan extensions subsumes all the other fundamental concepts of category theory.
>
> —Saunders MacLane, *Categories for the Working Mathematician*

***Definition* XV.2.6.1.** — *Let $\phi\colon I \to J$ be a functor. For each functor $G\colon J \to \mathcal{C}$, it induces a functor $\phi^*(G) := G \circ \phi\colon I \to \mathcal{C}$.*

(i) *A **left Kan extension** of a functor $F\colon I \to \mathcal{C}$ along ϕ is, if it exists, a functor $\operatorname{Lan}_\phi F\colon J \to \mathcal{C}$ together with a natural transformation $\eta_F\colon F \to \phi^*(\operatorname{Lan}_\phi F)$ such that each natural transformation $F \to \phi^*(G)$ factorises uniquely through η_F:*

$$\begin{tikzcd} I \arrow[rr, "F"] \arrow[dr, "\phi"'] & \arrow[d, phantom, "\Downarrow"] & \mathcal{C} \\ & J \arrow[ur, "G"'] & \end{tikzcd} \quad = \quad \begin{tikzcd} I \arrow[rr, "F"] \arrow[dr, "\phi"'] & \arrow[d, phantom, "\Downarrow\eta_F"] & \mathcal{C} \\ & J. \arrow[ur, "\operatorname{Lan}_\phi F"] \arrow[ur, bend right=60, "G"'] \arrow[ur, phantom, bend right=20, "\Downarrow\exists!"'] & \end{tikzcd}$$

(ii) *Dually, a **right Kan extension** of a functor $F\colon I \to \mathcal{C}$ along ϕ is, if it exists, a functor $\operatorname{Ran}_\phi F\colon J \to \mathcal{C}$ together with a natural transformation $\varepsilon_F\colon \phi^*(\operatorname{Ran}_\phi F) \to F$ such that each natural transformation $\phi^*(G) \to F$ factorises uniquely through ε_F:*

$$\begin{tikzcd} I \arrow[rr, "F"] \arrow[dr, "\phi"'] & \arrow[d, phantom, "\Uparrow"] & \mathcal{C} \\ & J \arrow[ur, "G"'] & \end{tikzcd} \quad = \quad \begin{tikzcd} I \arrow[rr, "F"] \arrow[dr, "\phi"'] & \arrow[d, phantom, "\Uparrow\varepsilon_F"] & \mathcal{C} \\ & J. \arrow[ur, "\operatorname{Ran}_\phi F"] \arrow[ur, bend right=60, "G"'] \arrow[ur, phantom, bend right=20, "\Uparrow\exists!"'] & \end{tikzcd}$$

The left and right Kan extensions solve the extension problems

$$\begin{tikzcd} I \arrow[r, "F"] \arrow[d, "\phi"'] \arrow[dr, phantom, "\Swarrow\eta_F"] & \mathcal{C} \\ J, \arrow[ur, dashed, "\operatorname{Lan}_\phi F"'] & \end{tikzcd} \qquad \begin{tikzcd} I \arrow[r, "F"] \arrow[d, "\phi"'] \arrow[dr, phantom, "\Nearrow\varepsilon_F"] & \mathcal{C} \\ J, \arrow[ur, dashed, "\operatorname{Ran}_\phi F"'] & \end{tikzcd}$$

by giving the indicated natural transformations induced by $\mathrm{id}_{\operatorname{Lan}_\phi F}$ and $\mathrm{id}_{\operatorname{Ran}_\phi F}$, resp.

***Observation* XV.2.6.2.** — If I and J are small, then $\phi^*\colon \operatorname{Fun}(J, \mathcal{C}) \to \operatorname{Fun}(I, \mathcal{C})$ is a well-defined functor. In this case, we can give an adjunction style definition for a functor $F\colon I \to \mathcal{C}$.

(i) A *left Kan extension* of F along ϕ is a functor $\operatorname{Lan}_\phi F\colon J \to \mathcal{C}$ such that the there is a bijection

$$\operatorname{Hom}_{\operatorname{Fun}(I,\mathcal{C})}(F, \phi^*(G)) \cong \operatorname{Hom}_{\operatorname{Fun}(J,\mathcal{C})}(\operatorname{Lan}_\phi F, G) \tag{XV.2.6.2.1}$$

which is natural in $G\colon J \to \mathcal{C}$.

(ii) Dually, a *right Kan extension* of a functor F along ϕ is a functor $\operatorname{Ran}_\phi F\colon J \to \mathcal{C}$ such that there is a bijection

$$\operatorname{Hom}_{\operatorname{Fun}(I,\mathcal{C})}(\phi^*(G), F) \cong \operatorname{Hom}_{\operatorname{Fun}(J,\mathcal{C})}(G, \operatorname{Ran}_\phi F)$$

which is natural in $G\colon J \to \mathcal{C}$.

If the Kan extensions exist for all F, then we obtain functors $\operatorname{Lan}_\phi\colon \operatorname{Fun}(I,\mathcal{C}) \to \operatorname{Fun}(J,\mathcal{C})$ and $\operatorname{Ran}_\phi\colon \operatorname{Fun}(I,\mathcal{C}) \to \operatorname{Fun}(J,\mathcal{C})$ by proposition XV.1.8.1 which form adjunctions

$$\begin{tikzcd} \operatorname{Fun}(I,\mathcal{C}) \arrow[r, bend left=50, "\operatorname{Lan}_\phi"] \arrow[r, bend right=50, "\operatorname{Ran}_\phi"'] & \operatorname{Fun}(J,\mathcal{C}) \arrow[l, "\phi^*" description] \end{tikzcd}.$$

In this case, η (resp. ε) is precisely the unit (resp. counit) of the adjunction $\operatorname{Lan}_\phi \dashv \phi^*$ (resp. $\phi^* \dashv \operatorname{Ran}_\phi$).

***Fact* XV.2.6.3.** — *Left (resp. right) Kan extensions are unique up to unique natural isomorphism.*

Proof. — Either use the universal property like for limits or, in the case of adjunctions, use lemma XV.1.6.9. ∎

***Remark* XV.2.6.4.** — Left and right Kan extensions are mutually dual. Since there is an isomorphism of categories $\operatorname{Fun}(I,\mathcal{C})^{\mathrm{op}} \xrightarrow{\sim} \operatorname{Fun}(I^{\mathrm{op}},\mathcal{C}^{\mathrm{op}})$, $F \mapsto F^{\mathrm{op}}$, we have a bijection

$$\begin{aligned} &\operatorname{Hom}_{\operatorname{Fun}(J^{\mathrm{op}},\mathcal{C}^{\mathrm{op}})}(G^{\mathrm{op}}, \operatorname{Ran}_{\phi^{\mathrm{op}}} F^{\mathrm{op}}) \cong \operatorname{Hom}_{\operatorname{Fun}(I^{\mathrm{op}},\mathcal{C}^{\mathrm{op}})}((\phi^{\mathrm{op}})^*(G^{\mathrm{op}}), F^{\mathrm{op}}) \\ &\quad\cong \operatorname{Hom}_{\operatorname{Fun}(I,\mathcal{C})^{\mathrm{op}}}((\phi^*)^{\mathrm{op}}(G), F) \cong \operatorname{Hom}_{\operatorname{Fun}(I,\mathcal{C})}(F, \phi^*(G)) \cong \operatorname{Hom}_{\operatorname{Fun}(J,\mathcal{C})}(\operatorname{Lan}_\phi F, G) \\ &\quad\cong \operatorname{Hom}_{\operatorname{Fun}(J,\mathcal{C})^{\mathrm{op}}}(G, \operatorname{Lan}_\phi^{\mathrm{op}} F) \cong \operatorname{Hom}_{\operatorname{Fun}(J^{\mathrm{op}},\mathcal{C}^{\mathrm{op}})}(G^{\mathrm{op}}, (\operatorname{Lan}_\phi^{\mathrm{op}} F)^{\mathrm{op}}) \end{aligned}$$

that is natural in $G \in \operatorname{Fun}(J,\mathcal{C})^{\mathrm{op}}$. If left and right Kan extensions extend to functors on $\operatorname{Fun}(I,\mathcal{C})$, then uniqueness of adjoints up to natural isomorphism (proposition XV.1.8.2) implies that the diagram

$$\begin{tikzcd} \operatorname{Fun}(I,\mathcal{C})^{\mathrm{op}} \arrow[r, "\operatorname{Lan}_\phi^{\mathrm{op}}"] \arrow[d, "\wr"'] & \operatorname{Fun}(J,\mathcal{C})^{\mathrm{op}} \arrow[d, "\wr"] \\ \operatorname{Fun}(I^{\mathrm{op}},\mathcal{C}^{\mathrm{op}}) \arrow[r, "\operatorname{Ran}_{\phi^{\mathrm{op}}}"] & \operatorname{Fun}(J^{\mathrm{op}},\mathcal{C}^{\mathrm{op}}) \end{tikzcd}$$

quasi-commutes.

***Proposition* XV.2.6.5.** — *Let $\phi\colon I \to J$ be a functor.*

(i) *Let $F\colon I \to \mathcal{C}$ be a diagram such that $\operatorname{Lan}_\phi F$ exists, and let $\eta_F\colon F \to \phi^* \operatorname{Lan}_\phi F$ be the adjunct of $\mathrm{id}_{\operatorname{Lan}_\phi F}$ (unit). Then the canonical morphism*

$$\operatorname*{colim}_I F \xrightarrow{\operatorname{colim}_I \eta_F} \operatorname*{colim}_I \phi^* \operatorname{Lan}_\phi F \longrightarrow \operatorname*{colim}_J \operatorname{Lan}_\phi F \tag{XV.2.6.5.1}$$

is an isomorphism in $\mathcal{C}$ where the second morphism is from remark XV.2.1.6.

(ii) *Dually, let $F\colon I^{\mathrm{op}} \to \mathcal{C}$ is a diagram such that $\operatorname{Ran}_{\phi^{\mathrm{op}}} F$ exists, and let $\varepsilon_F\colon (\phi^{\mathrm{op}})^* \operatorname{Ran}_{\phi^{\mathrm{op}}} F \to F$ be the adjunct of $\mathrm{id}_{\operatorname{Ran}_{\phi^{\mathrm{op}}} F}$ (counit). Then the canonical morphism*

$$\lim_J \operatorname{Ran}_{\phi^{\mathrm{op}}} F \longrightarrow \lim_I (\phi^{\mathrm{op}})^* \operatorname{Ran}_{\phi^{\mathrm{op}}} F \xrightarrow{\lim_I \varepsilon_F} \lim_I F$$

is an isomorphism in $\mathcal{C}$ where the first morphism is from remark XV.2.1.6.

In other words, we have isomorphisms

$$\lim_{j\in J} \lim_{\phi(i)\to j} F(i) \xrightarrow{\sim} \lim_{i\in I} F(i), \qquad \operatorname*{colim}_{i\in I} F(i) \xrightarrow{\sim} \operatorname*{colim}_{j\in J} \operatorname*{colim}_{\phi(i)\to j} F(i).$$

Proof by universal properties. — Observe that there is a natural isomorphism $\phi^*(\Delta^J_X) \cong \Delta^I_X$ of constant functors for all $X \in \mathrm{Ob}(\mathcal{C})$. Let $\alpha\colon F \to \Delta^I_X \cong \phi^*(\Delta^J_X)$ be a cone. The universal property of $\mathrm{Lan}_\phi F$ shows that α factorises through $\eta_F\colon F \to \phi^*(\mathrm{Lan}_\phi F)$ and the cone $\phi^*(\beta)$ for a unique cone $\beta\colon \mathrm{Lan}_\phi F \to \Delta^J_X$. Then the universal property of $\mathrm{colim}_J \mathrm{Lan}_\phi F$ gives a unique morphism $\mathrm{colim}_J \mathrm{Lan}_\phi F \to X$. This shows that $\mathrm{colim}_J \mathrm{Lan}_\phi F$ satisfies the same universal property as $\mathrm{colim}_I F$ and hence they are canonically isomorphic. □

Proof by adjunction. — Suppose that I and J are small so that $\mathrm{Fun}(I, \mathcal{C})$ and $\mathrm{Fun}(J, \mathcal{C})$ are defined. Using the adjunction $\mathrm{colim} \dashv \Delta$ (observation XV.2.1.12), we have the bijections

$$\begin{aligned}\mathrm{Hom}_{\mathcal{C}}(\operatorname*{colim}_J \mathrm{Lan}_\phi F, X) &\cong \mathrm{Hom}_{\mathrm{Fun}(J,\mathcal{C})}(\mathrm{Lan}_\phi F, \Delta^J_X) \cong \mathrm{Hom}_{\mathrm{Fun}(I,\mathcal{C})}(F, \phi^*(\Delta^J_X)) \\ &\cong \mathrm{Hom}_{\mathrm{Fun}(I,\mathcal{C})}(F, \Delta^I_X) \cong \mathrm{Hom}_{\mathcal{C}}(\operatorname*{colim}_I F, X)\end{aligned}$$

which are natural in $X \in \mathrm{Ob}(\mathcal{C})$. By lemma XV.1.6.9, we obtain an isomorphism $\mathrm{colim}_I F \xrightarrow{\sim} \mathrm{colim}_J \mathrm{Lan}_\phi F$ by chasing the identity of $\mathrm{colim}_J \mathrm{Lan}_\phi F$ through the bijections. Recall from the adjunct formula in proposition XV.1.7.6 that the second isomorphism is $\phi^*(-) \circ \eta_F$ (this only requires naturality in one entry, not both). Hence, the isomorphism is indeed the morphism in equation (XV.2.6.5.1). □

In general, Kan extensions are rather poorly behaved. The following describes a nicely behaved Kan extension.

***Theorem* XV.2.6.6.** — *Let $\phi\colon I \to J$ and $F\colon I \to \mathcal{C}$ be functors. By abuse of notation, let $j\colon 1 \to J$ denotes the inclusion functor of $j \in \mathrm{Ob}(J)$.*

(i) *Suppose that* $\lim_{j\downarrow\phi} F = \lim_{j\to\phi(i)} F(i)$ *exists for all* $j \in \mathrm{Ob}(J)$. *Then* $\mathrm{Ran}_\phi F\colon J \to \mathcal{C}$ *exists and is computed pointwise by*

$$\mathrm{Ran}_\phi F(j) \cong \lim_{j\to\phi(i)} F(i)$$

for all $j \in \mathrm{Ob}(J)$.

(ii) *Dually, suppose that* $\mathrm{colim}_{\phi\downarrow j} F = \mathrm{colim}_{\phi(i)\to j} F(i)$ *exists for all* $j \in \mathrm{Ob}(J)$. *Then* $\mathrm{Lan}_\phi F\colon J \to \mathcal{C}$ *exists and is computed pointwise by*

$$\mathrm{Lan}_\phi F(j) \cong \operatorname*{colim}_{\phi(i)\to j} F(i)$$

for all $j \in \mathrm{Ob}(J)$.

In this case, $\mathrm{Ran}_\phi F$ *and* $\mathrm{Lan}_\phi F$ *are called* ***pointwise right Kan extension*** *and* ***pointwise left Kan extension****, resp.*

Proof by universal properties. — We first construct the functor $\mathrm{Ran}_\phi F$. Define $\mathrm{Ran}_\phi F(j) = \lim_{j\to\phi(i)} F(i)$ for each $j \in \mathrm{Ob}(J)$. Each morphism $f\colon j \to j'$ in J induces a functor $(j' \downarrow \phi) \to (j \downarrow \phi)$ sending $j' \to \phi(i)$ to $j \to j' \to \phi(i)$, so a diagram indexed by $j\downarrow\phi$ is also indexed by $j'\downarrow\phi$. Hence, the morphisms $\mathrm{Ran}_\phi F(j) = \lim_{j\to\phi(i)} F(i) \to F(i)$ for each $(j \to \phi(i)) \in \mathrm{Ob}(j\downarrow\phi)$ form a cone indexed by $j'\downarrow\phi$, and the universal property gives a morphism $\mathrm{Ran}_\phi F(f)\colon \mathrm{Ran}_\phi F(j) \to \mathrm{Ran}_\phi F(j')$.

To prove that $\mathrm{Ran}_\phi F$ is a Kan extension, define $\varepsilon_F\colon \phi^*(\mathrm{Ran}_\phi F) \to F$ by the universal cone, i.e., $(\varepsilon_F)_i\colon \mathrm{Ran}_\phi F(\phi(i)) = \lim_{j\to\phi(i)} F(i) \to F(i)$. Let $\alpha\colon \phi^*(G) \to F$ be a natural transformation where $G\colon J \to \mathcal{C}$. To show the universal property, we need to construct a natural transformation $\theta\colon G \to \mathrm{Ran}_\phi F$ such that $\alpha = \varepsilon_F \circ \phi^*(\theta)$. For each $j \in \mathrm{Ob}(J)$, we define the component $\theta_j\colon G(j) \to$

$\operatorname{Ran}_\phi F(j) = \lim_{j\to\phi(i)} F(i)$ by constructing the cone

$$\begin{tikzcd} & G(j) \arrow[dl] \arrow[dr] & \\ G(\phi(i)) \arrow[rr] \arrow[d, "\alpha_i"'] & & G(\phi(i')) \arrow[d, "\alpha_{i'}"] \\ F(i) \arrow[rr] & & F(i') \end{tikzcd}$$

over F indexed by $j \downarrow \phi$ and using the universal property of $\operatorname{Ran}_\phi F(j)$ to obtain θ_j. Naturality of θ follows by functoriality of $\lim_{j\to\phi(i)}$. The condition $\alpha = \varepsilon_F \circ \phi^*(\theta)$ is obviously satisfied (consider $\mathrm{id}_{G(\mathrm{id}_{\phi(i)})}$). For all $j \to \phi(i)$, we have the commutative diagram

$$\begin{tikzcd} G(j) \arrow[rr] \arrow[d, dashed, "\theta_j"'] & & G(\phi(i)) \arrow[d, "\theta_{\phi(i)}"] \arrow[dl, bend right, "\alpha_i"'] \\ \operatorname{Ran}_\phi F(j) \arrow[rr] \arrow[dr] & & \operatorname{Ran}_\phi F(\phi(i)) \arrow[dl, "(\eta_F)_i"] \\ & F(i), & \end{tikzcd}$$

which shows that the choice for θ_j is unique. ◻

Proof with adjunctions. — Let I and J be small. To prove that $\operatorname{Ran}_\phi F$ is a Kan extension, we show equation (XV.2.6.2.1). We first define

$$\Phi\colon \operatorname{Hom}_{\mathrm{Fun}(I,\mathcal{C})}(\phi^*(G), F) \longrightarrow \operatorname{Hom}_{\mathrm{Fun}(J,\mathcal{C})}(G, \operatorname{Ran}_\phi F).$$

Let $u \in \operatorname{Hom}_{\mathrm{Fun}(I,\mathcal{C})}(\phi^*(G), F)$. Then for each $j \in \mathrm{Ob}(J)$ and each $(j \to \phi(i)) \in \mathrm{Ob}(j \downarrow \phi)$, the diagram

$$\begin{tikzcd} G(j) \arrow[r] \arrow[dr, dashed, "\Phi(u)_j"'] & G(\phi(i)) \arrow[r, "u_i"] & F(i) \\ & \lim_{j\to\phi(i)} F(i) = \operatorname{Ran}_\phi F(j) \arrow[ur] & \end{tikzcd}$$

commutes. The morphisms $\Phi(u)_j$ are functorial in j and hence define a natural transformation $\Phi(u) \in G \to \operatorname{Ran}_\phi F$. Now, we define

$$\Psi\colon \operatorname{Hom}_{\mathrm{Fun}(J,\mathcal{C})}(G, \operatorname{Ran}_\phi F) \longrightarrow \operatorname{Hom}_{\mathrm{Fun}(I,\mathcal{C})}(\phi^*(G), F).$$

Let $v \in \operatorname{Hom}_{\mathrm{Fun}(J,\mathcal{C})}(G, \operatorname{Ran}_\phi F)$. Then for each $i \in \mathrm{Ob}(I)$, we define a morphism

$$\Psi(v)_i\colon G(\phi(i)) \xrightarrow{v_{\phi(i)}} \operatorname{Ran}_\phi F(\phi(i)) = \lim_{\phi(i)\to\phi(i')} F(i') \longrightarrow F(i).$$

These morphisms are functorial in i and hence define a natural transformation $\Psi(v)\colon \phi^*(G) \to F$. It is now easy to check that Φ and Ψ are mutually inverse. ◻

***Corollary* XV.2.6.7.** — *Let $\mathcal{C}$ be a complete (resp. cocomplete) category, let I be a small and J a locally small category. Then the (pointed) Kan extension $\operatorname{Ran}_\phi F$ (resp. $\operatorname{Lan}_\phi F$) of $F\colon I \to \mathcal{C}$ along $\phi\colon I \to J$ exists.*

Proof. — By the size hypothesis on ϕ, for each $j \in \mathrm{Ob}(J)$ the category $j \downarrow \phi$ is small. Since $\mathcal{C}$ is complete, it admits all limits of shape $j \downarrow \phi$. Theorem XV.2.6.6 implies the statement. ∎

***Warning* XV.2.6.8.** — The size condition on I in corollary XV.2.6.7 cannot be improved. If corollary XV.2.6.7 were true for all large (or even locally small) categories I, then proposition XV.2.6.5 below shows that any complete category $\mathcal{C}$ admits all limits of that size. By proposition XV.2.2.6, every complete category would be thin, which is clearly not the case for $\mathcal{C} = \mathsf{Set}$. Even assuming Grothendieck universes, by proposition XV.2.2.6, there will always be limits which do not exist in $\mathcal{C}$ if $\mathcal{C}$ is not thin, however large the universe was chosen.

***Corollary* XV.2.6.9.** — *Let $\phi\colon I \to J$ be a fully faithful functor of small categories, and suppose that the pointwise right (resp. left) Kan extension functor Ran_ϕ (resp. Lan_ϕ) exists. Then the Kan extension is fully faithful and there is a natural isomorphism $\varepsilon\colon \phi^* \circ \mathrm{Ran}_\phi \xRightarrow{\sim} \mathrm{id}_{\mathrm{Fun}(I,\mathcal{C})}$ (resp. $\eta\colon \mathrm{id}_{\mathrm{Fun}(I,\mathcal{C})} \xRightarrow{\sim} \phi^* \circ \mathrm{Lan}_\phi$).*

Proof. — There is an isomorphism of categories

$$i/I = (i \downarrow \mathrm{id}_I) \cong (\phi \circ i \downarrow \phi \circ \mathrm{id}_I) = (\phi(i) \downarrow \phi).$$

Hence, there is an isomorphism

$$\phi^* \mathrm{Ran}_\phi F(i) = \mathrm{Ran}_\phi F(\phi(i)) \cong \lim_{\phi(i) \to \phi(i')} F(i') \cong \lim_{i \to i'} F(i') = F(i)$$

which is natural in i (the last equality follows since id_i is the initial object in i/I). In other words, we have a natural isomorphism $\phi^* \mathrm{Ran}_\phi F \xRightarrow{\sim} F$ for all $F \in \mathrm{Fun}(I, \mathcal{C})$, giving a natural isomorphism (counit) $\phi^* \circ \mathrm{Ran}_\phi \xRightarrow{\sim} \mathrm{id}_{\mathrm{Fun}(I,\mathcal{C})}$. Proposition XV.1.8.5 implies that Ran_ϕ is fully faithful. ∎

As an application, we recast some concepts as Kan extensions.

***Proposition* XV.2.6.10** (limits are Kan extensions)**.** — *Let $\phi\colon I \to \mathbb{1}$ be the functor to the terminal category. The colimits are left Kan extensions along ϕ and limits are right Kan extensions along ϕ.*

Proof with universal properties. — Let $F\colon I \to \mathcal{C}$ be a functor. For any $X \in \mathrm{Ob}(\mathcal{C})$, the composition $\phi^*(X) = X \circ \phi\colon I \to \mathbb{1} \to \mathcal{C}$ is the constant functor Δ_X, where $X\colon \mathbb{1} \to \mathcal{C}$, $\mathrm{pt} \mapsto X$ by abuse of notation. Hence, the universal property of left Kan extensions is the universal property of colimits: each natural transformation $F \to \Delta_X$ factors uniquely through the natural transformation $F \to \phi^*(\mathrm{Lan}_\phi F)$. Thus, $\mathrm{Lan}_\phi F$ is the colimit $\mathrm{colim}_I F$ considered as a functor $\mathbb{1} \to \mathcal{C}$. ∎

Proof with adjunctions. — Suppose that I is small. Note that $\phi^* = \Delta\colon \mathcal{C} \to \mathrm{Fun}(I, \mathcal{C})$, $X \mapsto \Delta_X$ is the constant functor. The statement now follows from the adjunction triples $\mathrm{colim}_I \dashv \Delta \dashv \lim_I$ (observation XV.2.1.12) and $\mathrm{Lan}_\phi \dashv \phi^* \dashv \mathrm{Ran}_\phi$ (observation XV.2.6.2) and the uniqueness of adjoints (proposition XV.1.8.2).

If we only consider a specific diagram $F\colon I \to \mathcal{C}$, then all referenced statements above can be specialised to F and the proof works the same (use lemma XV.1.6.9 for uniqueness). ∎

***Proposition* XV.2.6.11** (adjunctions are Kan extensions)**.** — *Let $L\colon \mathcal{C} \to \mathcal{D}$ and $R\colon \mathcal{D} \to \mathcal{C}$ be functors. If $L \dashv R$, then $R \cong \mathrm{Lan}_L \mathrm{id}_\mathcal{C}$ and $L \cong \mathrm{Ran}_R \mathrm{id}_\mathcal{D}$, and the Kan extensions are pointwise.*

Proof with universal properties. — Let $\eta\colon \mathrm{id}_\mathcal{C} \to R \circ L$ and $\varepsilon\colon L \circ R \to \mathrm{id}_\mathcal{D}$ be the unit and counit, resp. We check that R satisfies the universal property of $\mathrm{Lan}_L \mathrm{id}_\mathcal{C}$ with η. Let $\alpha\colon \mathrm{id}_\mathcal{C} \to F \circ L$ be a natural transformation for a functor $F\colon \mathcal{D} \to \mathcal{C}$. We first show that $\theta = (F\varepsilon) \circ (\alpha R)$ satisfies $\alpha = (\theta L) \circ \eta$ by doing a similar calculation as in the (co)units proof of proposition XV.1.8.2. We have the following identities:

– $\alpha_{R(L(X))} \circ \eta_X = F(L(\eta_X)) \circ \alpha_X$ by naturality of α applied to $\eta_X\colon X \to R(L(X))$;
– $\varepsilon_{L(X)} \circ L(\eta_X) = \mathrm{id}_{L(X)}$ by the triangle inequality $(\varepsilon L) \circ (L\eta) = \mathrm{id}_L$.

Hence, for each $X \in \mathrm{Ob}(\mathcal{C})$ we have

$$F(\varepsilon_{L(X)}) \circ \alpha_{R(L(X))} \circ \eta_X = F(\varepsilon_{L(X)}) \circ F(L(\eta_X)) \circ \alpha_X = F(\mathrm{id}_{L(X)}) \circ \alpha_X = \mathrm{id}_{F(L(X))} \circ \alpha_X = \alpha_X,$$

as desired. For uniqueness, we have the following identities:

– $\theta_Y \circ R(\varepsilon_Y) = F(\varepsilon_Y) \circ \theta_{L(R(Y))}$ by naturality of θ applied to $\varepsilon_Y\colon L(R(Y)) \to Y$;
– $R(\varepsilon_Y) \circ \eta_{R(Y)} = \mathrm{id}_{R(Y)}$ by the triangle inequality $(R\varepsilon) \circ (\eta R) = \mathrm{id}_R$.

Hence, for each $Y \in \mathrm{Ob}(\mathcal{D})$ we have

$$F(\varepsilon_Y) \circ \alpha_{R(Y)} = F(\varepsilon_Y) \circ \theta_{L(R(Y))} \circ \eta_{R(Y)} = \theta_Y \circ R(\varepsilon_Y) \circ \eta_{R(Y)} = \theta_Y,$$

as expected. ◻

Proof with adjunctions. — Suppose that $\mathcal{C}$ and $\mathcal{D}$ are small, and let (η, ε) be the (co)units. We need to show that the maps

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{D},\mathcal{C})}(R, F) \cong \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C},\mathcal{C})}(\mathrm{id}_{\mathcal{C}}, F \circ L), \qquad \alpha \longmapsto (\alpha L) \circ \eta, \qquad (F\varepsilon) \circ (\alpha R) \longleftarrow \alpha,$$

which are natural in F, are mutually inverse. This proves that R is $\mathrm{Lan}_L \mathrm{id}_{\mathcal{C}}$.

For each $\alpha\colon \mathrm{id}_{\mathcal{C}} \to F \circ L$, we show $((F\varepsilon) \circ (\alpha R))L \circ \eta = \alpha$ like in the proof above. The other identity $(F\varepsilon) \circ ((\alpha L) \circ \eta)R = \alpha$ is proven with a similar argument. ◻

XV.2.7. Final functors.

***Definition* XV.2.7.1.** — (i) *A functor $\phi\colon J \to I$ is **final**[(XV.9)] if the comma category $i \downarrow \phi$ is connected (hence non-empty) for each $i \in \mathrm{Ob}(I)$.*

(ii) *Dually, a functor $\phi\colon J \to I$ is **initial** if $\phi^{\mathrm{op}}\colon J^{\mathrm{op}} \to I^{\mathrm{op}}$ is final, i. e., $\phi \downarrow i$ is connected for each $i \in \mathrm{Ob}(I)$.*

***Fact* XV.2.7.2.** — *Let $L \dashv R$ be adjoint functors $I \rightleftarrows J$. Then R is final and L is initial.*

Proof. — Let $\eta\colon \mathrm{id}_I \to R \circ L$ be the unit. For each $i \in \mathrm{Ob}(I)$, the category $i \downarrow R$ is non-empty since it contains $\eta_i\colon i \to R(L(i))$. Given any object $f\colon i \to R(j)$ of $i \downarrow R$, let $\tilde{f}\colon L(i) \to j$ be its adjunct. By the adjunct formula from proposition XV.1.7.6, we have $R(\tilde{f}) \circ \eta_i = f$, that is, f and η_i are connected by $R(\tilde{f})$. Thus, $i \downarrow R$ is connected and R is final. ◻

***Fact* XV.2.7.3.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor, and let T be a terminal object of $\mathcal{C}$. Then $F(T)$ is terminal if and only if F is final. The dual statement holds for initial objects in $\mathcal{C}$.*

Proof. — Let $Y \in \mathrm{Ob}(\mathcal{D})$ be arbitrary. If F is final, then $Y \downarrow F$ is connected and, in particular, non-empty. Let $f\colon Y \to F(X)$ be an object of $Y \downarrow F$. Then there exists a morphism $u\colon X \to T$ in $\mathcal{C}$, so $F(u) \circ f$ is a morphism $Y \to F(T)$. If there are two morphisms $v, v'\colon Y \rightrightarrows F(T)$, then connectivity gives a morphism $u\colon T \to T$ such that $F(u) \circ v = v'$. But T is terminal, so $u = \mathrm{id}_T$.

Conversely, suppose that $F(T)$ is terminal. Then $Y \downarrow F$ is non-empty as it contains the unique morphism $v\colon Y \to F(T)$. For any object $f\colon Y \to F(X)$ in $Y \downarrow F$, the unique morphism $u\colon X \to T$ gives a morphism $F(u) \circ f\colon Y \to F(T)$, which must equal v by uniqueness. Thus, f and v are connected. ◻

(XV.9) Historically, final functors were called *cofinal* and initial functors were called *co-cofinal*, a notion imported from order theory (a subset B of a preordered set A is *cofinal* if for each $a \in A$, there is some $b \in B$ such that $a \leq b$). Here, 'co-' is used in the Latin sense of *cum* ('with') and not in the sense of *contra* ('against'). Modern usage avoids these terms since there is nothing dual about the prefix 'co-' in 'cofinal'.

Before we come to a characterisation of final functors, namely that changing diagrams along ϕ does not change their colimit, we need a couple of lemmas.

***Fact* XV.2.7.4.** — *Let $\Delta_X \colon I \to \mathcal{C}$ be a constant functor with connected index category I. Then* $\lim_I \Delta_X \cong \operatorname{colim}_I \Delta_X \cong X$*, if the (co)limits exist.*

Proof. — For any cone $\alpha \colon \Delta_Y \to \Delta_X$ over Δ_X, we have $\mathrm{id}_X \circ \alpha_i = \alpha_j$ whenever there exists a morphism $i \to j$ in I. Since I is connected, induction shows that each α_i is some given $f \colon X \to Y$. The same is true for colimits.

***Lemma* XV.2.7.5.** — *Consider the small constant diagram $\Delta_S \colon I \to \mathsf{Set}$. Then $\operatorname{colim}_I \Delta_S \cong \coprod_{\pi_0(I)} S$. In particular, if $S = \{\mathrm{pt}\}$, then $\operatorname{colim}_I \Delta_S \cong \pi_0(I)$.*

Proof. — Consider $\pi_0(I)$ as a discrete category with the functor $\pi \colon I \to \pi_0(I)$ that collapses connected components. Then Δ_S factorises through π; let $\tilde{\Delta}_S \colon \pi_0(I) \to \mathsf{Set}$ denote the induced constant functor. Since $(\pi \downarrow [i]) \cong I/i$ is connected for any $[i] \in \pi_0(I)$, the colimit $\operatorname{colim}_{\pi\downarrow[i]} \Delta_S \cong S$ exists, and so the Kan extension $\operatorname{Lan}_\pi \Delta_S \cong \tilde{\Delta}_S$ exists by theorem XV.2.6.6. By propositions XV.2.6.5 and XV.2.2.4, we obtain $\operatorname{colim}_I \Delta_S \cong \operatorname{colim}_{\pi_0(I)} \tilde{\Delta}_S \cong \coprod_{\pi_0(I)} S$.

***Lemma* XV.2.7.6.** — *Let $F \colon I \to \mathsf{Set}$ be a small diagram. Then $\operatorname{colim}_I F \cong \pi_0(\mathrm{el}(F))$.*

Proof. — Recall that $\mathrm{el}(F)$ has pairs (i, x) with $i \in \mathrm{Ob}(I)$ and $x \in F(i)$ as objects, and a morphism $(i, x) \to (j, y)$ is a morphism $f \colon i \to j$ in I such that $F(f)(x) = y$. By proposition XV.2.2.4, we have

$$\operatorname*{colim}_I F \cong \coprod_{i \in \mathrm{Ob}(I)} F(i)/{\sim} \cong \coprod_{(i,x) \in \mathrm{Ob}(\mathrm{el}(F))} \{(i,x)\}/{\sim} = \mathrm{Ob}(\mathrm{el}(F))/{\sim} = \pi_0(\mathrm{el}(F))$$

where the equivalence relation $\sim$ is generated by $(x \sim y) = ((i, x) \sim (j, y))$ with $x \in F(i)$ and $y \in F(j)$ such that there exists $f \colon i \to j$ with $F(f)(x) = y$.

***Corollary* XV.2.7.7.** — *Let $F \colon \mathcal{C} \to \mathcal{D}$ and $G \colon \mathcal{C} \to \mathcal{E}$ be two functors from a small category $\mathcal{C}$ to locally small categories $\mathcal{D}$ and $\mathcal{E}$. Let $A \in \mathrm{Ob}(\mathcal{D})$ and $B \in \mathrm{Ob}(\mathcal{E})$. Then there is an isomorphism*

$$\operatorname*{colim}_{(G(X)\to B)\in(G\downarrow B)} \mathrm{Hom}_{\mathcal{D}}(A, F(X)) \cong \operatorname*{colim}_{(A\to F(X))\in(A\downarrow F)} \mathrm{Hom}_{\mathcal{E}}(G(X), B).$$

Proof. — We define the category I as follows.

– The objects of I are triples (X, s, t) with $X \in \mathrm{Ob}(\mathcal{C})$ and morphisms $s \colon A \to F(X)$ in $\mathcal{D}$ and $t \colon G(X) \to B$ in $\mathcal{E}$.

– The morphisms $(X, s, t) \to (X', s', t')$ of I are morphisms $f \colon X \to X'$ such that the diagrams

$$\begin{tikzcd} G(X) \arrow[rr, "G(f)"] \arrow[dr, "t"'] & & G(X') \arrow[dl, "t'"] \\ & B, & \end{tikzcd} \qquad \begin{tikzcd} F(X) \arrow[rr, "F(f)"] & & F(X') \\ & A \arrow[ul, "s"] \arrow[ur, "s'"'] & \end{tikzcd}$$

commute.

Consider the functors $\mathrm{Hom}_{\mathcal{D}}(A, F) \colon (G \downarrow B) \to \mathsf{Set}$ and $\mathrm{Hom}_{\mathcal{E}}(G, B) \colon (A \downarrow F)^{\mathrm{op}} \to \mathsf{Set}$. By the size hypotheses, $G \downarrow B$ and $A \downarrow F$ is small, so the colimits of these functors exist. Then we have isomorphisms of categories $I \cong \mathrm{el}_{G\downarrow B}(\mathrm{Hom}_{\mathcal{D}}(A, F))$ and $I^{\mathrm{op}} \cong \mathrm{el}_{(A\downarrow F)^{\mathrm{op}}}(\mathrm{Hom}_{\mathcal{E}}(G, B))$. By lemma XV.2.7.6, we obtain

$$\operatorname*{colim}_{(G(X)\to B)\in(G\downarrow B)} \mathrm{Hom}_{\mathcal{D}}(A, F(X)) \cong \pi_0(I) \cong \pi_0(I^{\mathrm{op}}) \cong \operatorname*{colim}_{(A\to F(X))\in(A\downarrow F)} \mathrm{Hom}_{\mathcal{E}}(G(X), B).$$

***Lemma* XV.2.7.8.** — *Let I be a small category, and consider the hom-functor $\mathrm{Hom}_I(i,-)\colon I \to \mathsf{Set}$. Then $\mathrm{colim}_{j\in I}\mathrm{Hom}_I(i,j) \cong \{\mathrm{pt}\}$.*

Proof. — We show that

$$\{\mathrm{pt}\} \xrightarrow{\mathrm{pt}\mapsto \mathrm{id}_i} \mathrm{Hom}_I(i,i) \longrightarrow \underset{j\in I}{\mathrm{colim}}\,\mathrm{Hom}_I(i,j).$$

is surjective. For any $[f] \in \mathrm{colim}_{j\in I}\mathrm{Hom}_I(i,j)$, let $f\colon i \to j$ be the morphism in I whose image in the colimit is $[f]$, see proposition XV.2.2.8. Then there is a commutative diagram

$$\begin{tikzcd}
\mathrm{Hom}_I(i,i) \arrow[rr] \arrow[dr, "f\circ -"'] & & \mathrm{colim}_{j\in I}\mathrm{Hom}_I(i,j) \\
& \mathrm{Hom}_I(i,j), \arrow[ur] &
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathrm{id}_i \arrow[rr, mapsto] \arrow[dr, mapsto] & & {[f]} \\
& f. \arrow[ur, mapsto] &
\end{tikzcd}$$

***Proposition* XV.2.7.9** (characterisation of finality)**.** — *Let $\phi\colon J \to I$ be a functor between small categories. Then the following are equivalent.*

(i) *ϕ is final, or equivalently, ϕ^{op} is initial.*

(ii) *For any diagram $F\colon I^{\mathrm{op}} \to \mathsf{Set}$, the canonical morphism $\lim_I F \to \lim_J(F\circ\phi^{\mathrm{op}})$ is an isomorphism.*

(iii) *Dually, for any diagram $F\colon I \to \mathsf{Set}$, the canonical morphism $\mathrm{colim}_J(F\circ\phi) \to \mathrm{colim}_I F$ is an isomorphism.*

(iv) *For any diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$ in a complete category $\mathcal{C}$, the canonical morphism $\lim_I F \to \lim_J(F\circ\phi^{\mathrm{op}})$ is an isomorphism.*

(v) *Dually, for any diagram $F\colon I \to \mathcal{C}$ in a cocomplete category $\mathcal{C}$, the canonical morphism $\mathrm{colim}_J(F\circ\phi) \to \mathrm{colim}_I F$ is an isomorphism.*

(vi) *$\mathrm{colim}_{j\in J}\mathrm{Hom}_I(i,\phi(j)) \cong \{\mathrm{pt}\}$ for all $i \in \mathrm{Ob}(I)$.*

Proof. — Obviously, (iv) implies (ii) and (v) implies (iii) (using proposition XV.2.2.4). Further, (iii) implies (vi) by lemma XV.2.7.8. Also, (iv) and (v) are equivalent since they are formally dual.

We show that (ii) implies (iv). Let $F\colon I^{\mathrm{op}} \to \mathcal{C}$ with $\mathcal{C}$ complete. Then

$$\mathrm{Hom}_{\mathcal{C}}(X, \lim_I F) \cong \lim_I \mathrm{Hom}_{\mathcal{C}}(X,F) \xrightarrow{\sim} \lim_J \mathrm{Hom}_{\mathcal{C}}(X, F\circ\phi^{\mathrm{op}}) \cong \mathrm{Hom}_{\mathcal{C}}(X, \lim_J(F\circ\phi^{\mathrm{op}}))$$

is an isomorphism that is natural in $X \in \mathrm{Ob}(\mathcal{C})$ by (ii) and proposition XV.2.2.8. Then corollary XV.1.6.8 implies (iv). Note that this does not work for showing that (iii) implies (v).

Next, we show that (vi) implies (i) by utilising corollary XV.2.7.7. Consider the functors $\phi\colon J \to I$ and $\Delta_{\{\mathrm{pt}\}}\colon J \to \mathsf{Set}$. Hence with (vi), for each $i \in \mathrm{Ob}(I)$, there is an isomorphism

$$\{\mathrm{pt}\} \cong \underset{j\in J\cong(\Delta_{\{\mathrm{pt}\}}\downarrow\{\mathrm{pt}\})}{\mathrm{colim}} \mathrm{Hom}_I(i,\phi(j)) \cong \underset{(i\to\phi(j))\in(i\downarrow\phi)}{\mathrm{colim}} \mathrm{Hom}_{\mathsf{Set}}(\Delta_{\{\mathrm{pt}\}}(j), \{\mathrm{pt}\}) \cong \underset{(i\to\phi(j))\in(i\downarrow\phi)}{\mathrm{colim}} \{\mathrm{pt}\}.$$

On the other hand, by lemma XV.2.7.5, we have

$$\underset{(i\to\phi(j))\in(i\downarrow\phi)}{\mathrm{colim}} \{\mathrm{pt}\} \cong \underset{(i\to\phi(j))\in(i\downarrow\phi)}{\mathrm{colim}} \Delta_{\{\mathrm{pt}\}}(j) \cong \pi_0(i\downarrow\phi).$$

Thus, $i\downarrow\phi$ is connected. This implies (i).

Finally, we show that (i) implies (v). We need to construct an inverse to $\alpha\colon \mathrm{colim}_J(F\circ\phi) \to \mathrm{colim}_I F$. Let $i \in \mathrm{Ob}(I)$. Each object $i \to \phi(j)$ in $i\downarrow\phi$ induces a morphism $F(i) \to F(\phi(j))$, which assemble into a natural transformation $\Delta_{F(i)} \to F\circ\phi$ of functors $(i\downarrow\phi) \to \mathcal{C}$. Since $i\downarrow\phi$ is small, so that the colimits exist, we obtain a morphism

$$\underset{(i\to\phi(j))\in(i\downarrow\phi)}{\mathrm{colim}} \Delta_{F(i)} \longrightarrow \underset{(i\to\phi(j))\in(i\downarrow\phi)}{\mathrm{colim}} F(\phi(j)) \cong \underset{i\downarrow\phi}{\mathrm{colim}}(F\circ\phi\circ\mathrm{For}_i) \longrightarrow \underset{J}{\mathrm{colim}}(F\circ\phi)$$

where $\mathrm{For}_i\colon (i\downarrow\phi) \to J$ is the forgetful functor. Since $i\downarrow\phi$ is connected by (i), we have $\operatorname{colim}_{i\to\phi(j)} \Delta_{F(i)} \cong F(i)$. The universal property of colimits gives $\beta\colon \operatorname{colim}_I F \to \operatorname{colim}_J(F\circ\phi)$.

We check that α and β are mutually inverse. Let $\lambda\colon F \to \Delta_{\operatorname{colim}_I F}$ and $\mu\colon F\circ\phi \to \Delta_{\operatorname{colim}_J(F\circ\phi)}$ be the colimit cones. Note that by construction, we have

$$\left(F(i) \xrightarrow{\lambda_i} \operatorname*{colim}_I F \xrightarrow{\beta} \operatorname*{colim}_J(F\circ\phi)\right) = \left(F(i) \xrightarrow{F(f)} F(\phi(j)) \xrightarrow{\mu_j} \operatorname*{colim}_J(F\circ\phi)\right)$$

for some (hence any since $i\downarrow\phi$ is connected) morphism $f\colon i\to\phi(j)$ in I. On the other hand, we have

$$\left(F(\phi(j)) \xrightarrow{\mu_j} \operatorname*{colim}_J(F\circ\phi) \xrightarrow{\alpha} \operatorname*{colim}_I F\right) = \left(F(\phi(j)) \xrightarrow{\lambda_{\phi(j)}} \operatorname*{colim}_I F\right).$$

Hence,

$$\alpha\circ\beta\circ\lambda_i = \alpha\circ\mu_j\circ F(f) = \lambda_{\phi(j)}\circ F(f) = \lambda_i,$$
$$\beta\circ\alpha\circ\mu_j = \beta\circ\lambda_{\phi(j)} = \mu_j\circ F(\mathrm{id}_{\phi(j)}) = \mu_j$$

for all $i\in\mathrm{Ob}(I)$ and $j\in\mathrm{Ob}(J)$. By the universal properties of colimits, this implies $\alpha\circ\beta = \mathrm{id}_{\operatorname{colim}_I F}$ and $\beta\circ\alpha = \mathrm{id}_{\operatorname{colim}_J(F\circ\phi)}$. ∎

***Corollary* XV.2.7.10.** — *Let $\phi\colon J\to I$ be a final functor of small categories. Then $\pi_0(J)\cong\pi_0(I)$. In particular, I is connected if and only if J is connected.*

Proof. — Let $\Delta^I_{\{\mathrm{pt}\}}$ and $\Delta^J_{\{\mathrm{pt}\}}$ be the constant functors into Set. Then $\Delta^J_{\{\mathrm{pt}\}} \cong \Delta^I_{\{\mathrm{pt}\}}\circ\phi$. By proposition XV.2.7.9, we have $\operatorname{colim}_J \Delta^J_{\{\mathrm{pt}\}} \cong \operatorname{colim}_J(\Delta^I_{\{\mathrm{pt}\}}\circ\phi) \cong \operatorname{colim}_I \Delta^I_{\{\mathrm{pt}\}}$. So, $\pi_0(J)\cong\pi_0(I)$ due to lemma XV.2.7.5. ∎

***Fact* XV.2.7.11.** — *Let $\psi\colon K\to J$ and $\phi\colon J\to I$ be functors of small categories.*

(i) *If ϕ and ψ are final, then $\phi\circ\psi$ is final.*
(ii) *If $\phi\circ\psi$ and ψ are final, then ϕ is final.*
(iii) *If $\phi\circ\psi$ is final and if ϕ is fully faithful, then ϕ and ψ are final.*

The dual statement holds for initial functors.

Proof. — For any Set-valued diagram $F\colon I\to\mathsf{Set}$, we have morphisms

$$\operatorname*{colim}_K(F\circ\phi\circ\psi) \xrightarrow{f} \operatorname*{colim}_J(F\circ\phi) \xrightarrow{g} \operatorname*{colim}_I F$$

of sets. By proposition XV.2.7.9, g is an isomorphism if and only if ϕ is final, and $g\circ f$ an isomorphism if and only if $\phi\circ\psi$ is final. Moreover, if ψ is final, then f is an isomorphism. This implies (i) and (ii).

For (iii), since ϕ is fully faithful, we have an isomorphism of categories $j\downarrow\psi \cong \phi(j)\downarrow\phi\circ\psi$ for all $j\in\mathrm{Ob}(J)$. Since $\phi\circ\psi$ is final, the category $j\downarrow\psi$ is connected, whence ψ is final. Then ϕ is final by (ii). ∎

***Definition* XV.2.7.12.** — (i) *A category I is **finally small** if there exists a final functor $J\to I$ with J small.*

(ii) *A category I is **initially small** if I^{op} is finally small, i. e., there exists an initial functor $J\to I$ with J small.*

***Remark* XV.2.7.13.** — If $\mathcal{C}$ is a cocomplete (resp. complete) category and if I is a finally (resp. initially) small category, then $\mathcal{C}$ admits all colimits (resp. limits) of shape I by proposition XV.2.7.9.

***Corollary* XV.2.7.14.** — *Let I be finally (resp. initially) small. Then there exists a small full subcategory J of I such that the embedding functor $J \to I$ is final (resp. initial).*

Proof. — Let $f\colon K \to I$ be final with K small. Let J be the full subcategory of I with $\mathrm{Ob}(J) = f(\mathrm{Ob}(K))$. Then J is small. Let $\phi\colon J \to I$ be the fully faithful embedding functor and let $\psi\colon K \to J$ be induced by f. Since $\phi \circ \psi$ is final, fact XV.2.7.11 implies that ϕ is final. ∎

XV.2.8. Free cocompletions and Yoneda extensions.

***Motivation* XV.2.8.1.** — A category $\mathcal{C}$ might not have all (co)limits. One way to mitigate is that if $\mathcal{C}$ is small, then we can pass to the (co)complete category $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ (corollary XV.2.2.11) via the Yoneda embedding $\mathrm{h}_{\mathcal{C}}\colon \mathcal{C} \to \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$, $X \mapsto \mathrm{Hom}_{\mathcal{C}}(-, X)$. Then $\mathcal{C}$ is a full subcategory of $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ and Moreover, if $\mathcal{C}$ is complete, then $\mathrm{h}_{\mathcal{C}}$ is continuous (proposition XV.2.2.8), but even if $\mathcal{C}$ is cocomplete, then $\mathrm{h}_{\mathcal{C}}$ might still be not cocontinuous.

***Definition* XV.2.8.2.** — *Let $\mathcal{C}$ be a small category.*

(i) *The category $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ is the **free cocompletion** of $\mathcal{C}$. We consider $\mathcal{C}$ as a full subcategory via the Yoneda embedding $\mathrm{h}_{\mathcal{C}}\colon \mathcal{C} \to \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$. Hence, we will write $X := \mathrm{h}_{\mathcal{C}}(X) = \mathrm{Hom}_{\mathcal{C}}(-, X)$ for $X \in \mathrm{Ob}(\mathcal{C})$ and $X(A) := \mathrm{Hom}_{\mathcal{C}}(A, X)$ for $A \in \mathrm{Ob}(\mathcal{C})$. In this sense, given a small diagram $F\colon I \to \mathcal{C}$ in $\mathcal{C}$, the **formal colimit** of F in $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ is $\mathrm{colim^f}_I F := \mathrm{colim}_I(\mathrm{h}_{\mathcal{C}} \circ F)$.*

(ii) *Dually, the category $\mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}$ is the **free completion** of $\mathcal{C}$. We also consider $\mathcal{C}$ as the full subcategory of $\mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}$ via the Yoneda embedding $(\mathrm{h}^{\mathcal{C}})^{\mathrm{op}}\colon \mathcal{C} \to \mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}$. Given a small diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$ in $\mathcal{C}$, the **formal limit** of F in $\mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}$ is $\mathrm{lim^f}_I F := \mathrm{lim}_I((\mathrm{h}^{\mathcal{C}})^{\mathrm{op}} \circ F)$.*

***Remark* XV.2.8.3.** — For consistency reasons, we could have defined the limit of a small diagram $F\colon I^{\mathrm{op}} \to \mathcal{C}$ by a representation $\lim_I((\mathrm{h}^{\mathcal{C}})^{\mathrm{op}}(-) \circ F) \cong \mathrm{Hom}_{\mathcal{C}}(-, \lim_I F)$ agreeing with construction XV.2.1.9.

Moreover, $(\mathrm{h}^{\mathcal{C}})^{\mathrm{op}}\colon \mathcal{C} \to \mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}$ is again fully faithful, and the Yoneda lemma XV.1.6.5 holds:

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}}(F, (\mathrm{h}^{\mathcal{C}})^{\mathrm{op}}(X)) \cong F(X)$$

for every $X \in \mathrm{Ob}(\mathcal{C})$ and every functor $F\colon \mathcal{C} \to \mathsf{Set}$ and this isomorphism is natural in F and X, i. e., a natural isomorphism of functors $\mathrm{Fun}(\mathcal{C}, \mathsf{Set}) \times \mathcal{C} \to \mathsf{Set}$.

***Remark* XV.2.8.4.** — Let $F\colon I \to \mathcal{C}$ be a small diagram with $\mathcal{C}$ small. Then $\mathrm{colim}_I F$, as a representative of a covariant functor $\mathcal{C} \to \mathsf{Set}$, can be considered as an object of $\mathrm{Fun}(\mathcal{C}, \mathsf{Set})$, but $\mathrm{colim^f}_I F$ is an object of $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$. Dually for small diagrams $F\colon I^{\mathrm{op}} \to \mathcal{C}$, the limit $\lim_I F$ is an object of $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$, but $\mathrm{lim^f}_I F$ is an object of $\mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}$.

***Remark* XV.2.8.5.** — Both notions of formal (co)limits are dual. Let $F\colon I^{\mathrm{op}} \to \mathcal{C}$ be a small diagram, and let $F^{\mathrm{op}}\colon I \to \mathcal{C}^{\mathrm{op}}$ be induced by F. Then

$$(\operatorname*{lim^f}_I F)^{\mathrm{op}} = (\lim_I((\mathrm{h}^{\mathcal{C}})^{\mathrm{op}} \circ F))^{\mathrm{op}} \cong \operatorname*{colim}_I(\mathrm{h}^{\mathcal{C}} \circ F^{\mathrm{op}}) = \operatorname*{colim}_I(\mathrm{h}_{\mathcal{C}^{\mathrm{op}}} \circ F^{\mathrm{op}}) = \operatorname*{colim^f}_I F^{\mathrm{op}}.$$

***Observation* XV.2.8.6.** — Let $\mathcal{C}$ be a small category. Let $F\colon I \to \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ be a small diagram, and let $X \in \mathrm{Ob}(\mathcal{C})$. Then by the Yoneda lemma XV.1.6.5 and proposition XV.2.2.10, we have

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})}(X, \operatorname*{colim}_I F) \cong (\operatorname*{colim}_I F)(X) \cong \operatorname*{colim}_{i \in I} F(i)(X) \cong \operatorname*{colim}_{i \in I} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})}(X, F(i)).$$

Dually for small diagrams $F\colon I^{\mathrm{op}} \to \mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}$, we have

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}}(\lim_I F, X) \cong (\lim_I F)(X) \cong \lim_{i \in I} F(i)(X) \cong \lim_{i \in I} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}}(F(i), X).$$

Note that the isomorphism is generally false if $X \in \mathrm{Ob}(\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}))$ (resp. $X \in \mathrm{Ob}(\mathrm{Fun}(\mathcal{C}, \mathsf{Set})^{\mathrm{op}}))$.

On the other hand, proposition XV.2.2.8 implies that for small diagrams $F\colon I \to \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ and $G \in \mathrm{Ob}(\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}))$, we have

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(\operatorname*{colim}_I F, G) \cong \lim_I \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(F, G),$$

and dually for small diagrams $F\colon I \to \mathrm{Fun}(\mathcal{C}, \mathsf{Set})$ and $G \in \mathrm{Ob}(\mathrm{Fun}(\mathcal{C}, \mathsf{Set}))$, we have

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C},\mathsf{Set})}(G, \lim_I F) \cong \lim_I \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C},\mathsf{Set})}(G, F).$$

Hence, for small diagrams $F\colon I \to \mathcal{C}$ and $G\colon J \to \mathcal{C}$ we have

$$\begin{aligned}\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(\operatorname*{colim^f}_{i\in I} F(i), \operatorname*{colim^f}_{j\in J} G(j)) &\cong \lim_{i\in I} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(F(i), \operatorname*{colim^f}_{j\in J} G(j)) \\ &\cong \lim_{i\in I} \operatorname*{colim}_{j\in J} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(F(i), G(j)) \cong \lim_{i\in I} \operatorname*{colim}_{j\in J} \mathrm{Hom}_{\mathcal{C}}(F(i), G(j))\end{aligned}$$

where the last isomorphism holds since the Yoneda embedding is fully faithful (corollary XV.1.6.7). Dually, for small diagrams $F\colon I^{\mathrm{op}} \to \mathcal{C}$ and $G\colon J^{\mathrm{op}} \to \mathcal{C}$, we have

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C},\mathsf{Set})}(\operatorname*{lim^f}_{i\in I} F(i), \operatorname*{lim^f}_{j\in J} G(j)) \cong \lim_{j\in J} \lim_{i\in I} \mathrm{Hom}_{\mathcal{C}}(G(j), F(i)).$$

***Remark* XV.2.8.7.** — Let $\mathcal{C}$ be a category. Suppose that $\mathcal{C}$ is cocomplete, and let $F\colon I \to \mathcal{C}$ be a small diagram. Then there is a canonical morphism

$$\operatorname*{colim^f}_I F = \operatorname*{colim}_I \mathrm{Hom}_{\mathcal{C}}(X, F) \longrightarrow \mathrm{Hom}_{\mathcal{C}}(X, \operatorname*{colim}_I F) = \mathrm{h}_{\mathcal{C}}(\operatorname*{colim}_I F).$$

Dually, suppose that $\mathcal{C}$ is complete, and let $F\colon I^{\mathrm{op}} \to \mathcal{C}$ be a small diagram. Then there is a canonical morphism

$$(\mathrm{h}^{\mathcal{C}})^{\mathrm{op}}(\lim_I F) = \mathrm{Hom}_{\mathcal{C}}(\lim_I F, X) \longrightarrow \lim_I \mathrm{Hom}_{\mathcal{C}}(F, X) = \operatorname*{lim^f}_I F.$$

The following statements relates formal colimits and limits by setting $G = \mathrm{id}_{\mathcal{C}}$.

***Proposition* XV.2.8.8.** — *Let $\mathcal{C}$ be a small category, and let $F\colon I \to \mathcal{C}$ be a small diagram. Assume that $\mathrm{colim^f}_I F$ is represented by some $X \in \mathrm{Ob}(\mathcal{C})$, that is, $\mathrm{colim^f}_I F \cong X$. Then for any functor $G\colon \mathcal{C} \to \mathcal{D}$, we have $\mathrm{colim}_I(G \circ F) \cong G(X)$.*

Proof. — Let $Y \in \mathrm{Ob}(\mathcal{D})$ be arbitrary, and define the functor

$$\mathcal{H}om(G, Y)\colon \mathcal{C}^{\mathrm{op}} \longrightarrow \mathsf{Set}, \qquad Z \longmapsto \mathrm{Hom}_{\mathcal{D}}(G(Z), Y).$$

Then we have isomorphisms

$$\begin{aligned}\mathrm{Hom}_{\mathcal{D}}(\operatorname*{colim}_I(G \circ F), Y) &\cong \lim_I \mathrm{Hom}_{\mathcal{D}}(G \circ F, Y) = \lim_I(\mathcal{H}om(G, Y) \circ F^{\mathrm{op}}) \\ &\cong \lim_I \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(F, \mathcal{H}om(G, Y)) \cong \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(\operatorname*{colim^f}_I F, \mathcal{H}om(G, Y)) \\ &\cong \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(X, \mathcal{H}om(G, Y)) \cong \mathcal{H}om(G, Y)(X) = \mathrm{Hom}_{\mathcal{D}}(F(X), Y)\end{aligned}$$

which are natural in Y by using observation XV.2.8.6 and the Yoneda lemma XV.1.6.5. Finally, corollary XV.1.6.8 implies the statement. ∎

Our next goal is to show that extending functors to the free cocompletion makes them cocontinuous. But first, a very technical lemma.

***Lemma* XV.2.8.9.** — *Let $\mathcal{C}$ be a small category. Let $A\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ be a functor, and we denote the inclusion $1 \to \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ by A as well. Then there is an equivalence of categories $\phi\colon (\mathrm{id}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})} \downarrow A) = \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})/A \xrightarrow{\sim} \mathrm{Fun}((\mathrm{h}_{\mathcal{C}} \downarrow A)^{\mathrm{op}}, \mathsf{Set})$ such that the diagram*

$$\begin{tikzcd} & \mathrm{h}_{\mathcal{C}} \downarrow A \arrow[dl, "\mathrm{For}"'] \arrow[dr, "\mathrm{h}_{\mathrm{h}_{\mathcal{C}}\downarrow A}"] & \\ \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})/A \arrow[rr, "\phi", "\sim"'] & & \mathrm{Fun}((\mathrm{h}_{\mathcal{C}} \downarrow A)^{\mathrm{op}}, \mathsf{Set}) \end{tikzcd} \tag{XV.2.8.9.1}$$

quasi-commutes where $\mathrm{For}\colon (\mathrm{h}_{\mathcal{C}} \downarrow A) \to \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})/A$, $(X, \mathrm{h}_{\mathcal{C}}(X) \to A) \mapsto (\mathrm{h}_{\mathcal{C}}(X) \to A)$.

Proof. — We construct ϕ as follows. Let $t\colon G \to A$ be an object of $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})/A$. For each object $s\colon X \to A$ in $\mathrm{h}_{\mathcal{C}} \downarrow A$, define $\phi(t)\colon (\mathrm{h}_{\mathcal{C}} \downarrow A)^{\mathrm{op}} \to \mathsf{Set}$ by

$$\phi(t)(s) := \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})/A}(\mathrm{For}(s), t) = \{u \in G(X) \mid t_X(u) = s \in A(X)\}$$

(the equality uses the Yoneda lemma XV.1.6.5). It is obvious that this hom-functor is functorial. If $G = Y \in \mathrm{Ob}(\mathcal{C})$, then

$$\phi(t)(s) = \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})/A}(\mathrm{For}(s), \mathrm{For}(t)) \cong \mathrm{Hom}_{(\mathrm{h}_{\mathcal{C}}\downarrow A)^{\mathrm{op}}}((X, s), (Y, t)),$$

so diagram (XV.2.8.9.1) quasi-commutes.

For the quasi-inverse $\psi\colon \mathrm{Fun}((\mathrm{h}_{\mathcal{C}} \downarrow A)^{\mathrm{op}}, \mathsf{Set}) \to \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})/A$, let $F\colon (\mathrm{h}_{\mathcal{C}} \downarrow A)^{\mathrm{op}} \to \mathsf{Set}$ be a functor. We need to construct an object in $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})/A$, which we denote as $\psi(F) \to A$ by abuse of notation. For each $X \in \mathrm{Ob}(\mathcal{C})$, we define $\psi(F) \to A$ by

$$\psi(F)(X) := \{(x, s) \mid s \in A(X), x \in F(X \xrightarrow{s} A)\} \longrightarrow A(X), \qquad (x, s) \longmapsto s$$

(where we used the Yoneda lemma XV.1.6.5 to associate $s\colon X \to A$ to $s \in A(X)$). Functoriality of ψ in F can be readily checked.

We omit the proof that ϕ and ψ are indeed quasi-inverse. ◻

The following can be considered as a dual of the Yoneda lemma XV.1.6.5.(XV.10) Essentially, while the Yoneda lemma XV.1.6.5 states that the free cocompletion contains the category fully faithfully, the co-Yoneda lemma XV.2.8.10 states that the image of the Yoneda embedding is dense.

***Co-Yoneda lemma* XV.2.8.10** (density formula, presheaves are colimits of representables)**.** — *Let $\mathcal{C}$ be a small category. Let $A\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ be a functor, and let $\mathrm{For}_A\colon (\mathrm{h}_{\mathcal{C}} \downarrow A) \to \mathcal{C}$ be the forgetful functor where A is understood as an object of $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$. Then*

$$\operatorname*{colim^f}_{\mathrm{h}_{\mathcal{C}}\downarrow A} \mathrm{For}_A = \operatorname*{colim^f}_{\mathcal{C}\ni V\to A} V \cong A.$$

Proof. — Note that by the Yoneda lemma XV.1.6.5, we have an isomorphism of categories $(\mathrm{h}_{\mathcal{C}}\downarrow A) \cong (\{\mathrm{pt}\} \downarrow A) = \mathrm{el}(A)$. For any $B\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$, we have a map

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(A, B) \longrightarrow \lim_{V\to A} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(V, B) \xrightarrow{\sim} \lim_{V\to A} B(V)$$

by the universal property of limits and the Yoneda lemma XV.1.6.5. By proposition XV.2.2.8, an element of $\lim_{V\to A} B(V)$ is a family $(x_{(V,s)})_{(V,s)} \in \prod_{(V,s)\in\mathrm{Ob}(\mathrm{el}(A))} B(V)$ such that for all $f\colon V \to W$ in $\mathcal{C}$ we have $B(f)(x_{(W,t)}) = x_{(V,s)}$ where $A(f)(s) = t$. But such a family defines a unique natural transformation $\eta\colon A \to B$ by $\eta_V(s) = x_{(V,s)}$, the compatibility property of the family is exactly naturality of η. Hence, we obtain a bijection

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(A, B) \xrightarrow{\sim} \lim_{V\to A} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(V, B) \cong \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(\operatorname*{colim^f}_{V\to A} V, B)$$

which is natural in B, and corollary XV.1.6.8 finishes. ◻

Move to some place when needed.

***Proposition* XV.2.8.11.** — *Let $\mathcal{C}$ be a small category. Let $F\colon I \to \mathcal{C}$ be a small diagram. Setting $A := \operatorname{colim}^{\mathrm{f}}_I F$, the induced functor*

$$\tilde{F}\colon I \longrightarrow (\mathrm{h}_{\mathcal{C}} \downarrow A), \qquad \tilde{F}(i)\colon \mathrm{h}_{\mathcal{C}}(F(i)) \longrightarrow \operatorname*{colim}_I(\mathrm{h}_{\mathcal{C}} \circ F) = A$$

is final. Note that $\mathrm{h}_{\mathcal{C}} \downarrow A$ is generally not essentially small.

Proof. — The situation is the quasi-commutative diagram

$$\begin{tikzcd} I \arrow[r, "\tilde{F}"] & (\mathrm{h}_{\mathcal{C}} \downarrow A) \arrow[r, "\mathrm{For}"] \arrow[dr, "\mathrm{h}_{\mathrm{h}_{\mathcal{C}}\downarrow A}"'] & \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})/A \arrow[r, "\mathrm{s}"] \arrow[d, "\phi", "\wr"'] & \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \\ & & \mathrm{Fun}((\mathrm{h}_{\mathcal{C}} \downarrow A)^{\mathrm{op}}, \mathsf{Set}) & \end{tikzcd}$$

by lemma XV.2.8.9 where ϕ is an equivalence of categories (recall that s denotes the forgetful functor that yields the domain). Hence, we obtain

$$\phi^{-1}(\operatorname*{colim}_I(\mathrm{h}_{\mathrm{h}_{\mathcal{C}}\downarrow A} \circ \tilde{F})) \cong \operatorname*{colim}_I(\phi^{-1} \circ \mathrm{h}_{\mathrm{h}_{\mathcal{C}}\downarrow A} \circ \tilde{F}) \cong \operatorname*{colim}_I(\mathrm{For} \circ \tilde{F}) \cong (\operatorname*{colim}_I(s \circ \mathrm{For} \circ \tilde{F}) \longrightarrow A) \cong \mathrm{id}_A$$

by corollary XV.2.2.13 and fact XV.2.2.16. Since id_A is the terminal object in $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})/A$, we see that $\operatorname{colim}_I(\mathrm{h}_{\mathrm{h}_{\mathcal{C}}\downarrow A} \circ \tilde{F})$ is the terminal object in $\mathrm{Fun}((\mathrm{h}_{\mathcal{C}} \downarrow A)^{\mathrm{op}}, \mathsf{Set})$ as well. Using that $\mathrm{h}_{\mathrm{h}_{\mathcal{C}}\downarrow A}$ is fully faithful (corollary XV.1.6.7) and observation XV.2.8.6, for all objects $(X, s\colon X \to A)$ in $\mathrm{h}_{\mathcal{C}} \downarrow A$ we have

$$\begin{aligned} \operatorname*{colim}_I \mathrm{Hom}_{\mathrm{h}_{\mathcal{C}}\downarrow A}((X,s), \tilde{F}) &\cong \operatorname*{colim}_I \mathrm{Hom}_{\mathrm{Fun}((\mathrm{h}_{\mathcal{C}}\downarrow A)^{\mathrm{op}}, \mathsf{Set})}((X,s), \mathrm{h}_{\mathrm{h}_{\mathcal{C}}\downarrow A} \circ \tilde{F}) \\ &\cong \mathrm{Hom}_{\mathrm{Fun}((\mathrm{h}_{\mathcal{C}}\downarrow A)^{\mathrm{op}}, \mathsf{Set})}((X,s), \operatorname*{colim}_I(\mathrm{h}_{\mathrm{h}_{\mathcal{C}}\downarrow A} \circ \tilde{F})) \cong \{\mathrm{pt}\}. \end{aligned}$$

Proposition XV.2.7.9 implies that $\tilde{F}$ is final. ◻

***Theorem* XV.2.8.12.** — *Let $\mathcal{C}$ be a small category and let $\mathcal{D}$ be a cocomplete category. Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor. Then the pointwise left Kan extension $\mathrm{Lan}_{\mathrm{h}_{\mathcal{C}}} F\colon \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \to \mathcal{D}$ exists, is cocontinuous and satisfies $(\mathrm{Lan}_{\mathrm{h}_{\mathcal{C}}} F) \circ \mathrm{h}_{\mathcal{C}} \cong F$.*

*Conversely, if a functor $\tilde{F}\colon \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \to \mathcal{D}$ satisfies $\tilde{F} \circ \mathrm{h}_{\mathcal{C}} \cong F$ and preserves small colimits in $\mathcal{C}$ (i. e., for any small diagram $G\colon I \to \mathcal{C}$ we have $\tilde{F}(\operatorname{colim}^{\mathrm{f}}_I G) \cong \operatorname{colim}_I(\tilde{F} \circ \mathrm{h}_{\mathcal{C}} \circ G)$), then there is a unique natural isomorphism $\tilde{F} \cong \mathrm{Lan}_{\mathrm{h}_{\mathcal{C}}} F$. We call $\mathrm{Lan}_{\mathrm{h}_{\mathcal{C}}} F$ the **Yoneda extension** of F.*

Proof. — For brevity, write $\tilde{F} := \mathrm{Lan}_{\mathrm{h}_{\mathcal{C}}} F$. By corollary XV.2.6.7, $\tilde{F}$ exists and is pointwise, i. e., $\tilde{F}(A) \cong \operatorname{colim}_{U \to A} F(U)$ for all $A \in \mathrm{Ob}(\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}))$ where the limit is indexed by $\mathrm{h}_{\mathcal{C}} \downarrow A$. Note that for all $V \in \mathrm{Ob}(\mathcal{C})$, we have $\tilde{F}(\mathrm{h}_{\mathcal{C}}(V)) \cong \operatorname{colim}_{U \to V} F(U) \cong F(V)$ which is natural in V (we used that $\mathrm{h}_{\mathcal{C}}$ is fully faithful, corollary XV.1.6.7).

It remains to show that $\tilde{F}$ is cocontinuous. Recall from the proof of proposition XV.2.8.8 that for $Y \in \mathrm{Ob}(\mathcal{D})$ we defined $\mathcal{H}om(F, Y)\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ by

$$\mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})}(U, \mathcal{H}om(F, Y)) \cong \mathcal{H}om(F, Y)(U) := \mathrm{Hom}_{\mathcal{D}}(F(U), Y).$$

for all $U \in \mathrm{Ob}(\mathcal{C})$. For any $A\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$, we hence obtain

$$\begin{aligned} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})}(A, \mathcal{H}om(F, Y)) &\cong \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})}(\operatorname*{colim^{\mathrm{f}}}_{U \to A} U, \mathcal{H}om(F, Y)) \\ &\cong \lim_{U \to A} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})}(U, \mathcal{H}om(F, Y)) \cong \lim_{U \to A} \mathrm{Hom}_{\mathcal{D}}(F(U), Y) \\ &\cong \mathrm{Hom}_{\mathcal{D}}(\operatorname*{colim}_{U \to A} F(U), Y) \cong \mathrm{Hom}_{\mathcal{D}}(\tilde{F}(A), Y), \end{aligned}$$

(XV.10) The duality arises if we consider the Yoneda lemma XV.1.6.5 as an end formula. Then the co-Yoneda lemma XV.2.8.10 can be written as a coend formula.

using observation XV.2.8.6 at various places. Plugging in $A = \operatorname{colim}_I G$ for any small diagram $G\colon I \to \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$, we obtain

$$\begin{aligned}\mathrm{Hom}_{\mathcal{D}}(\tilde{F}(\operatorname*{colim}_I G), Y) &\cong \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(\operatorname*{colim}_I G, \mathcal{H}om(F, Y))\\ &\cong \lim_I \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(G, \mathcal{H}om(F, Y)) \cong \lim_I \mathrm{Hom}_{\mathcal{D}}(\tilde{F}(G), Y) \cong \mathrm{Hom}_{\mathcal{D}}(\operatorname*{colim}_I \tilde{F}(G), Y).\end{aligned}$$

Since everything was natural in Y, corollary XV.1.6.8 implies that $\operatorname{colim}_I \tilde{F}(G) \cong \tilde{F}(\operatorname{colim}_I G)$ is the canonical isomorphism.

For uniqueness, suppose that a functor $\tilde{F}\colon \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \to \mathcal{D}$ satisfies $\tilde{F} \circ \mathrm{h}_{\mathcal{C}} \cong F$ and preserves small colimits in $\mathcal{C}$. Then for any functor $A\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$, we have $A \cong \operatorname{colim}^{\mathrm{f}}{}_{U\to A} U$ by the co-Yoneda lemma XV.2.8.10. Hence,

$$\tilde{F}(A) \cong \tilde{F}(\operatorname*{colim^f}_{U\to A} U) \cong \operatorname*{colim}_{U\to A} \tilde{F}(\mathrm{h}_{\mathcal{C}}(U)) \cong \operatorname*{colim}_{U\to A} F(U).$$

By theorem XV.2.6.6, $\tilde{F}$ is a pointwise left Kan extension along $\mathrm{h}_{\mathcal{C}}$ and uniqueness of Kan extensions finishes. ∎

***Notation* XV.2.8.13.** — Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor of small categories. We write $\hat{F} := \mathrm{Lan}_{\mathrm{h}_{\mathcal{C}}}(\mathrm{h}_{\mathcal{D}} \circ F)\colon \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \to \mathrm{Fun}(\mathcal{D}^{\mathrm{op}}, \mathsf{Set})$ for the pointwise left Kan extension associated to $\mathrm{h}_{\mathcal{D}} \circ F\colon \mathcal{C} \to \mathrm{Fun}(\mathcal{D}^{\mathrm{op}}, \mathsf{Set})$. In particular, for all functors $A\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ and all $V \in \mathrm{Ob}(\mathcal{D})$ we have

$$\hat{F}(A)(V) \cong (\operatorname*{colim}_{U\to A} \mathrm{h}_{\mathcal{D}}(F(U)))(V) \cong \operatorname*{colim}_{U\to A} \mathrm{Hom}_{\mathcal{D}}(V, F(U)).$$

By theorem XV.2.8.12, $\hat{F}$ is cocontinuous.

Skip [KS06, Not. 2.7.3] due to Universe usage ($\hat{\mathcal{C}}$ is not small for $\mathrm{Fun}(\hat{\mathcal{C}}, \mathcal{A})$).

***Proposition* XV.2.8.14.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor of small categories. Then there is a natural isomorphism $\hat{F} \cong \mathrm{Lan}_{F^{\mathrm{op}}}$ of functors $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \to \mathrm{Fun}(\mathcal{D}^{\mathrm{op}}, \mathsf{Set})$.*

Proof. — First of all, $\mathrm{Lan}_{F^{\mathrm{op}}}\colon \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \to \mathrm{Fun}(\mathcal{D}^{\mathrm{op}}, \mathsf{Set})$ exists and is pointwise by corollary XV.2.6.7. By corollary XV.2.7.7 and the Yoneda lemma XV.1.6.5, for all functors $A\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ and all $V \in \mathrm{Ob}(\mathcal{D})$ we have isomorphisms

$$\begin{aligned}\hat{F}(A)(V) &\cong \operatorname*{colim}_{\mathrm{h}_{\mathcal{C}}(U)\to A} \mathrm{Hom}_{\mathcal{D}}(V, F(U)) \cong \operatorname*{colim}_{V\to F(U)} \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}},\mathsf{Set})}(\mathrm{h}_{\mathcal{C}}(U), A)\\ &\cong \operatorname*{colim}_{F^{\mathrm{op}}(U)\to V} A(U) \cong \mathrm{Lan}_{F^{\mathrm{op}}} A(V)\end{aligned}$$

which are natural in A (and in V). This implies $\hat{F} \cong \mathrm{Lan}_{F^{\mathrm{op}}}$. ∎

Maybe add subsection on idempotent completion, also called Cauchy completion. See [KS06, Exer. 2.9].

XV.2.9. Filtered colimits.

XV.2.9.1. — Recall that a *directed set* is a non-empty poset $(I, \leq)$ (i. e., $\leq$ is reflexive, anti-symmetric and transitive) such that for any $i, j \in I$, there exists some $k \in I$ with $i \leq k$ and $j \leq k$. In other words, any finite subset of I has an upper bound. Dually, a *codirected set* is a poset $(I, \leq)$ such that the opposite $(I^{\mathrm{op}}, \leq^{\mathrm{op}})$ is directed.

***Motivation* XV.2.9.2.** — A non-trivial question is under which circumstances limits can commute with colimits (cf. proposition XV.2.2.14). Historically, one of the first observations was that finite limits commute with *sequential colimits* indexed by $\mathbb{N}$ with its total order. This result was subsequently generalised to *directed colimits* (more common *direct limit*) indexed by directed sets I, and ultimately to *filtered colimits*. Of course, the dual result on cofiltered limits also holds.

***Counterexample* XV.2.9.3.** — Let $\mathrm{For}\colon \mathsf{AbGrp} \to \mathsf{Set}$ be the forgetful functor. Since For is right adjoint to the free functor $\mathsf{Set} \to \mathsf{AbGrp}$, $X \mapsto \mathbb{Z}^{\oplus X}$, it preserves limits (theorem XV.2.2.12). However, it does not preserve all colimits. For instance, the coproduct of two A-modules M and N is the direct sum $M \oplus N$, but the coproduct of their underlying sets $\mathrm{For}(M)$ and $\mathrm{For}(N)$ is their disjoint union. The reason is that for small index categories I, the functor $\mathrm{colim}_I\colon \mathrm{Fun}(I, \mathsf{Set}) \to \mathsf{Set}$ does not preserve finite limits in general. Otherwise, for any small diagram $F\colon I \to \mathsf{AbGrp}$, the addition map $+$ would induce

$$\operatorname*{colim}_{i\in I} \mathrm{For}(F(i)) \times \operatorname*{colim}_{i\in I} \mathrm{For}(F(i)) \cong \operatorname*{colim}_{i\in I} \mathrm{For}(F(i) \times F(i)) \longrightarrow \operatorname*{colim}_{i\in I} \mathrm{For}(F(i)),$$

and $\mathrm{colim}_{i\in I} \mathrm{For}(F(i))$ would have the structure of an abelian group.

***Definition* XV.2.9.4.** — *A category I is **filtered** if the following properties are satisfied:*

(i) *I is non-empty.*
(ii) *For any $i, j \in \mathrm{Ob}(I)$, there exists $k \in \mathrm{Ob}(I)$ and morphisms $i \to k$ and $j \to k$.*
(iii) *For any parallel morphisms $f, g\colon i \rightrightarrows j$, there exists a morphism $h\colon j \to k$ such that $h \circ f = h \circ g$.*

These are visualised by the diagrams

$$\begin{tikzcd} i \arrow[dr, dashed] & & j \arrow[dl, dashed] \\ & k, & \end{tikzcd} \qquad \begin{tikzcd} i \arrow[r, shift left] \arrow[r, shift right] \arrow[dr, dashed] & j \arrow[d, dashed] \\ & k. \end{tikzcd}$$

*A diagram $F\colon I \to \mathcal{C}$ is **filtered** if I is filtered. A colimit $\mathrm{colim}_I F$ is **filtered** if F is filtered.*

*Dually, a category I is **cofiltered** if I^{op} is filtered, i. e.,*

(i) *I is non-empty.*
(ii) *For any $i, j \in \mathrm{Ob}(I)$, there exists $k \in \mathrm{Ob}(I)$ and morphisms $k \to i$ and $k \to j$.*
(iii) *For any parallel morphisms $f, g\colon i \rightrightarrows j$, there exists a morphism $h\colon k \to i$ such that $f \circ h = g \circ h$.*

*We similarly define **cofiltered** diagrams and **cofiltered** limits.*

***Example* XV.2.9.5.** — Recall that a preordered set $(I, \leq)$ can be understood as a category I by setting $\mathrm{Ob}(I) = I$ and $\mathrm{Hom}_I(i, j) = \{(i, j) \mid i \leq j\}$. This category is thin. Conversely, any thin category is a preordered set. Obviously, preordered sets are filtered categories.

***Fact* XV.2.9.6.** — (i) *A category is filtered if it has a terminal object.*
(ii) *A category is filtered if it admits all finite colimits.*
(iii) *A product of filtered categories is filtered.*
(iv) *Filtered categories are connected.*

The dual statement on cofiltered categories also holds true.

Proof. — By definition.

***Lemma* XV.2.9.7.** — (i) *Let $F\colon I \to \mathcal{C}$ be a small diagram. Let $\mathcal{J}$ be the category of finite finite subcategories of I ordered by inclusion. Then $\mathcal{J}$ is small and filtered, and we have a canonical isomorphism*

$$\operatorname*{colim}_{I} F \cong \operatorname*{colim}_{J\in\mathcal{J}} \operatorname*{colim}_{J} F|_J$$

where $F|_J\colon J \to \mathcal{C}$ denotes the restriction of F to $J \in \mathcal{J}$, if the colimits exist.

(ii) *Dually, let* $F\colon I^{\mathrm{op}} \to \mathcal{C}$ *be a small diagram. Let* $\mathcal{J}$ *be the poset of finite subsets of* $\mathrm{Ob}(I)$ *ordered by inclusion. Then* $\mathcal{J}^{\mathrm{op}}$ *is small and cofiltered, and we have a canonical isomorphism*

$$\lim_I F \cong \lim_{J\in\mathcal{J}} \lim_J F|_{J^{\mathrm{op}}}$$

where $F|_{J^{\mathrm{op}}}\colon J^{\mathrm{op}} \to \mathcal{C}$ *denotes the restriction of* F *to* $J^{\mathrm{op}} \in \mathcal{J}^{\mathrm{op}}$*, if the limits exist.*

Proof. — Since $\mathrm{Mor}(\mathcal{J}) \subseteq \mathcal{P}(\mathrm{Mor}(I))$ is a set, $\mathcal{J}$ is small. Also, since $\mathcal{J}$ is thin, we can easily check that it is filtered.

Let us prove the isomorphism. Let K be the full subcategory of $\mathcal{J} \times I$ consisting of objects (J, i) with $i \in \mathrm{Ob}(J)$. Then the morphisms are $(J, i) \to (J', i')$ with $J \hookrightarrow J'$ and $i \to i'$ in I. Then the forgetful functor $\phi\colon K \to I$ induced by $\mathcal{J} \times I \to I$ is final. Indeed, for any $i \in \mathrm{Ob}(I)$, the category $i \downarrow \phi$ consists of objects $f\colon i \to \phi(J, i') = i'$. In particular, the diagram

$$\begin{tikzcd}
i \arrow[r, "f"] \arrow[d] & \phi(J, i') \arrow[d] \\
\phi(\{i\}, i) \arrow[r, "\phi(f)"] & \phi(J \amalg \{i\}, i')
\end{tikzcd}$$

commutes for any f, so $i \downarrow \phi$ is connected. Thus, propositions XV.2.6.5 and XV.2.7.9 imply that

$$\operatorname*{colim}_I F \cong \operatorname*{colim}_J (F \circ \phi) \cong \operatorname*{colim}_{J\in\mathcal{J}} \operatorname*{colim}_{\mathrm{For}_{\mathcal{J}}(k)\to J} F(\phi(k))$$

where the inner colimit is indexed by $\mathrm{For}_{\mathcal{J}} \downarrow J$ and $\mathrm{For}_{\mathcal{J}}\colon K \to \mathcal{J}$ is the (other) forgetful functor.

Furthermore, for each $J \in \mathcal{J}$ the functor

$$\psi_J\colon J \longrightarrow (\mathrm{For}_{\mathcal{J}} \downarrow J), \qquad j \longmapsto (\mathrm{For}_{\mathcal{J}}(J, j) = J \xrightarrow{\mathrm{id}_J} J)$$

is final. Indeed, first note that $(\mathrm{For}_{\mathcal{J}} \downarrow J) \cong \{(J', j') \mid J' \hookrightarrow J\}$. Under this identification, for any object $k = (J', j')$ in $\mathrm{For}_{\mathcal{J}} \downarrow J$, the category $k \downarrow \psi_J$ is connected: its objects are of the form $k = (J', j') \to (J, j) = \psi_J(j)$ and are connected to $k \to (J, j') = \psi_J(j')$ via the morphism $(J, j') \to (J, j)$ (we have $j' \in \mathrm{Ob}(J') \subseteq \mathrm{Ob}(J)$). Thus, proposition XV.2.7.9 implies that

$$\operatorname*{colim}_{\mathrm{For}_{\mathcal{J}}(k)\to J} F(\phi(k)) \cong \operatorname*{colim}_J (F \circ \phi \circ \psi_J) = \operatorname*{colim}_J F|_J,$$

***Corollary* XV.2.9.8.** — (i) *If a category admits all small filtered colimits and all finite colimits (resp. finite coproducts), then it is cocomplete (resp. admits all small coproducts).*

(ii) *If a functor preserves small filtered colimits and all finite colimits (resp. finite coproducts), then it is cocontinuous (resp. preserves small coproducts).*

The dual statement on limits also holds true.

Cf. corollary XV.2.4.3 which states that if a category admits all small coproducts and all finite colimits (coequalisers suffice), then it is cocomplete.

***Notation* XV.2.9.9.** — Especially in algebra, the image of a (co)filtered diagram is often called a **(co)filtered system**.

***Lemma* XV.2.9.10.** — *A category* I *is filtered if and only if for any functor* $\phi\colon J \to I$ *with* J *finite, there exists* $i \in \mathrm{Ob}(I)$ *such that* $\lim_{j\in J} \mathrm{Hom}_I(\phi(j), i) \neq \varnothing$.

Proof. — Assume that I is filtered, and let $\phi\colon J \to I$ be a functor with J finite. Since I is filtered and since J is finite, we find some $i_1 \in \mathrm{Ob}(I)$ such that there are morphisms $f_j\colon \phi(j) \to i_1$ for all $j \in \mathrm{Ob}(J)$. Similarly, for each $j \in \mathrm{Ob}(J)$ there is a morphism $g_j\colon i_1 \to k_j$ such that the composition

$$\phi(j) \xrightarrow{\phi(h)} \phi(j') \xrightarrow{f_{j'}} i_1 \xrightarrow{g_j} k_j$$

is the same for all morphisms $h\colon j \to j'$ in J. Again, there are morphisms $h_j\colon k_j \to i_2$ for each $j \in \mathrm{Ob}(J)$, and there is a morphism $i_2 \to i_3$ such that the composition

$$i_1 \xrightarrow{g_j} k_j \xrightarrow{h_j} i_2 \longrightarrow i_3$$

is the same for all h_j. Then the family of compositions

$$\phi(j) \xrightarrow{f_j} i_1 \xrightarrow{g_j} k_j \xrightarrow{h_j} i_2 \longrightarrow i_3$$

over all $j \in \mathrm{Ob}(J)$ defines an element of $\lim_{j\in J} \mathrm{Hom}_I(\phi(j), i_3)$ by proposition XII.2.1.5.

Conversely, assume the other property. We check the definition of filteredness for I. Applying the property to any functor $\phi\colon J \to I$ with J finite, we see that I is non-empty. For any $i, j \in \mathrm{Ob}(I)$, define $\phi\colon 1 \amalg 1 \to I$. Then there is some $k \in \mathrm{Ob}(I)$ such that $\mathrm{Hom}_I(i,k) \times \mathrm{Hom}_I(j,k) \neq \varnothing$, so there are morphisms $i \to k$ and $j \to k$. Finally, for any parallel morphisms $f, g\colon i \rightrightarrows j$, define $\phi\colon \{\bullet \rightrightarrows \bullet\} \to I$. Then there is some $k \in \mathrm{Ob}(I)$ such that $\mathrm{Eq}(- \circ f, - \circ g) \neq \varnothing$ where both precompositions are maps $\mathrm{Hom}_I(j,k) \to \mathrm{Hom}_I(i,k)$. Any element $h\colon j \to k$ in the equaliser satisfies $h \circ f = h \circ g$. $\square$

This leads to a very explicit description of filtered colimits in Set (cf. proposition XV.2.2.4, where the equivalence relation must be transitively generated). Note that I being filtered is crucial.

***Proposition* XV.2.9.11.** — *Let $F\colon I \to \mathsf{Set}$ be a small filtered diagram. On $\coprod_{i\in\mathrm{Ob}(I)} F(i)$, for $x \in F(i)$ and $y \in F(j)$ define $x \sim y$ if there are $s\colon i \to k$ and $t\colon j \to k$ such that $F(s)(x) = F(t)(y)$. Then $\sim$ is an equivalence relation and $\mathrm{colim}_I F \cong \coprod_{i\in\mathrm{Ob}(I)} F(i)/\sim$.*

Proof. — Reflexivity and symmetry of $\sim$ is obvious. Suppose that $x_1 \sim x_2$ and $x_2 \sim x_3$ with $x_\lambda \in F(i_\lambda)$ for $\lambda = 1, 2, 3$. Then we have the diagram of solid morphisms

$$\begin{tikzcd}
F(i_1) \arrow[dr] & & F(i_2) \arrow[dl] \arrow[dr] & & F(i_3) \arrow[dl] \\
& F(j_1) \arrow[dr, dashed] & & F(j_2) \arrow[dl, dashed] & \\
& & F(k) & &
\end{tikzcd}$$

where x_1 and x_2 are sent to the same element, and x_2 and x_3 are sent to the same element. By applying lemma XV.2.9.10 to the span $j_1 \leftarrow i_2 \to j_2$, we obtain the dashed morphisms such that everything commutes. A diagram chase shows that $x_1 \sim x_3$, so $\sim$ is transitive. The representation of $\mathrm{colim}_I F$ follows from proposition XV.2.2.8. $\square$

***Remark* XV.2.9.12.** — There is no real dual statement to proposition XV.2.9.11 because limits in Set are already much better behaved. Compare: for a small diagram $F\colon I^{\mathrm{op}} \to \mathsf{Set}$, we have $(x_i)_{i\in\mathrm{Ob}(I)} \in \lim_I F$ satisfying $F(f)(x_j) = x_i$ for all $f\colon i \to j$ in I.

***Corollary* XV.2.9.13.** — *Let $F\colon I \to \mathsf{Set}$ be a small filtered diagram.*

(i) *Let* $S \subseteq \operatorname{colim}_I F$ *be a finite subset. Then there exists some* $i \in \mathrm{Ob}(I)$ *such that* S *is contained in the image of the canonical map* $F(i) \to \operatorname{colim}_I F$.

(ii) *Suppose that* $x, y \in F(i)$ *have the same image under* $F(i) \to \operatorname{colim}_I F$. *Then there exists a morphism* $f\colon i \to j$ *such that* $x, y \in F(i)$ *have the same image under* $F(f)\colon F(i) \to F(j)$.

Proof. — We identify $\operatorname{colim}_I F \cong \coprod_{i\in\mathrm{Ob}(I)} F(i)/{\sim}$ as in proposition XV.2.9.11. If $x \in F(i)$ and $y \in F(j)$, then there exist morphisms $i \to k$ and $j \to k$ such that the images of x and y lie in $F(k)$. Hence, lifting every element of S and applying this observation proves (i).

For (ii), proposition XV.2.9.11 implies that there are morphisms $s, t\colon i \rightrightarrows k$ such that $F(s)(x) = F(t)(y)$. Then there is a morphism $u\colon k \to j$ such that $f \coloneqq u \circ s = u \circ t$, so $F(f)(x) = F(f)(y)$. ∎

Add [KS06, Cor. 3.1.5] later when covering monads.

XV.2.10. Filtered colimits commute with finite limits.

This is the most important property of filtered colimits.

***Theorem* XV.2.10.1** (filtered colimits commute with finite limits)**.** — *Let* I *be a small category. Then* I *is filtered if and only if for any finite category* J *and any diagram* $F\colon I \times J^{\mathrm{op}} \to \mathsf{Set}$, *we have a canonical isomorphism*

$$\operatorname*{colim}_{i\in I} \lim_{j\in J} F(i,j) \xrightarrow{\sim} \lim_{j\in J} \operatorname*{colim}_{i\in I} F(i,j).$$

The dual statement about cofiltered limits commuting with finite colimits also holds true.

Proof. — Suppose that I is filtered. In order to show that colim_I preserves finite limits, by virtue of corollary XV.2.4.4, we show that it preserves all equalisers and binary products (the terminal object is obviously preserved).

We show that colim_I preserves equalisers. Let $\alpha, \beta\colon F \rightrightarrows G$ be two natural transformations of functors $I \to \mathsf{Set}$. Note that $\mathrm{Eq}(\alpha,\beta)$ being the equaliser means that $\mathrm{Eq}(\alpha,\beta)(i) \to F(i) \rightrightarrows G(i)$ is an equaliser diagram for all $i \in \mathrm{Ob}(I)$. We need to show that ϕ in the diagram

$$\begin{tikzcd}
\operatorname{colim}_I \mathrm{Eq}(\alpha,\beta) \arrow[d, dashed, "\phi"'] \arrow[dr] & & \\
\mathrm{Eq}(\operatorname{colim}_I \alpha, \operatorname{colim}_I \beta) \arrow[r, hook, "\iota"] & \operatorname{colim}_I F \arrow[r, shift left, "\operatorname{colim}_I \alpha"] \arrow[r, shift right, "\operatorname{colim}_I \beta"'] & \operatorname{colim}_I G
\end{tikzcd}$$

is bijective.

The map ϕ is surjective. Indeed, represent $x \in \mathrm{Eq}(\operatorname{colim}_I \alpha, \operatorname{colim}_I \beta)$ by some $x_i \in F(i)$ such that $\alpha_i(x_i), \beta_i(x_i) \in G(i)$ have the same image in $\operatorname{colim}_I G$. By corollary XV.2.9.13, there is $f\colon i \to j$ such that $G(f)(\alpha_i(x_i)) = G(f)(\beta_i(x_i)) \in G(j)$. Setting $x_j \coloneqq F(f)(x_i)$, we have $\alpha_j(x_j) = \beta_j(x_j) \in G(j)$ by naturality of α and β, whence $x_j \in \mathrm{Eq}(\alpha_j, \beta_j) = \mathrm{Eq}(\alpha,\beta)(j)$. Since $x_j = F(f)(x_i)$, the images of x_i and x_j in $\operatorname{colim}_I F$ agree by proposition XV.2.9.11, and since ι is injective, we must have $\phi(x_j) = x$.

The map ϕ is also injective. Indeed, let $x, y \in \operatorname{colim}_I \mathrm{Eq}(\alpha,\beta)$ with $\phi(x) = \phi(y)$. By corollary XV.2.9.13, we can represent x and y by $x_i, y_i \in \mathrm{Eq}(\alpha,\beta)(i)$, resp., for the same $i \in \mathrm{Ob}(I)$. Since x_i and y_i have the same image in $\operatorname{colim}_I F$, there exists $f\colon i \to j$ such that $F(f)(x_i) = F(f)(y_i) \in F(j)$ by corollary XV.2.9.13. Hence, $F(f)(x_i) = F(f)(y_i) \in \mathrm{Eq}(\alpha,\beta)(j)$, and $x = y$ by proposition XV.2.9.11.

In the case of binary products, let $F, G\colon I \to \mathsf{Set}$ be two functors, and we need to show that $\phi\colon \operatorname{colim}_I(F \times G) \to (\operatorname{colim}_I F) \times (\operatorname{colim}_I G)$ is bijective. The proof is very similar.

For surjectivity, let $x \in \operatorname{colim}_I F$ and $y \in \operatorname{colim}_I G$. Since I is filtered, we may represent x and y by $x_i \in F(i)$ and $y_i \in G(i)$, resp., for the same $i \in \mathrm{Ob}(I)$. Hence, $(x_i, y_i) \in (F \times G)(i)$ and $\phi(x_i, y_i) = (x, y)$.

For injectivity, let $x, y \in \operatorname{colim}_I(F \times G)$ with $\phi(x) = \phi(y)$. Since I is filtered, we may represent x and y by (x_1, x_2) and (y_1, y_2) in $F(i) \times G(i) = (F \times G)(i)$ for the same $i \in \mathrm{Ob}(I)$. By assumption, x_1 and y_1 have the same image in $\operatorname{colim}_I F$. So, corollary XV.2.9.13 gives some $f\colon i \to j$ such that $F(f)(x_1) = F(f)(y_1) \in F(j)$. By a similar argument and since I is filtered, we may choose f such that also $G(f)(x_2) = G(f)(y_2) \in G(j)$. Thus, $x = y$ by proposition XV.2.9.11.

Conversely, suppose that colim_I preserves finite limits. We appeal to lemma XV.2.9.10 to show that I is filtered. Let $\phi\colon J \to I$ be a functor with J finite. By lemma XV.2.7.8, we have

$$\operatorname*{colim}_{i\in I} \lim_{j\in J} \mathrm{Hom}_I(\phi(j), i) \cong \lim_{j\in J} \operatorname*{colim}_{i\in I} \mathrm{Hom}_I(\phi(j), i) \cong \lim_{j\in J}\{\mathrm{pt}\} \cong \{\mathrm{pt}\}$$

Hence, on the left-hand side, there exists some $i \in \mathrm{Ob}(I)$ such that $\lim_{j\in J} \mathrm{Hom}_I(\phi(j), i) \neq \varnothing$.

***Corollary* XV.2.10.2.** — *In* $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$*, small filtered colimits commute with finite limits.*

Proof. — This follows from theorem XV.2.10.1 since (co)limits in functor categories like $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ are computed pointwise (proposition XV.2.2.10).

***Warning* XV.2.10.3.** — It is not true that filtered colimits commute with finite limits in *any* category. Consider the one-point-compactification $\mathbb{N}^* := \mathbb{N} \amalg \{\infty\}$ where $\mathbb{N}$ is discrete, and let $\mathcal{C}$ be the partially ordered set of all closed subsets of $\mathbb{N}^*$ ordered by inclusion. This means that the objects of $\mathcal{C}$ are all finite subsets of $\mathbb{N}$ and all sets $U \amalg \{\infty\}$ where $U \subseteq \mathbb{N}$.

Let $(C_i)_{i\in I}$ be the family of all finite subsets of $\mathbb{N}$. Then $\operatorname{colim}_{i\in I} C_i = \mathbb{N}^*$ since the only closed subset of $\mathbb{N}^*$ containing $\mathbb{N}$ is $\mathbb{N}^*$. Also, observe that $\{\infty\} \cap U = \{\infty\} \cap U$ for any $U \subseteq \mathbb{N}^*$ since any cone with tip S has the property $S \subseteq \{\infty\} \cap U$ and the product is maximal with this property. Hence,

$$\{\infty\} \times \operatorname*{colim}_{i\in I} C_i = \{\infty\} \times \mathbb{N}^* = \{\infty\} \neq \varnothing = \operatorname*{colim}_{i\in I} \varnothing = \operatorname*{colim}_{i\in I}(\{\infty\} \times C_i).$$

Skip [KS06, Prop. 3.1.7], generalise with monads. Also Skip [KS06, Prop. 3.1.8(i)] on finality.

***Lemma* XV.2.10.4.** — *Let* $\psi\colon K \to I$ *and* $\phi\colon J \to I$ *be functors.*

(i) *Suppose that* ψ *is final. If* K *is filtered and if* $\phi \downarrow \psi(k)$ *is filtered for all* $k \in \mathrm{Ob}(K)$*, then* J *is filtered.*

(ii) *Dually, suppose that* ψ *is initial. If* K *is cofiltered and if* $\psi(k) \downarrow \phi$ *is cofiltered for all* $k \in \mathrm{Ob}(K)$*, then* J *is cofiltered.*

Proof. — Let $F\colon A \to J$ be a finite diagram. Since ψ is final, $\phi(F(a)) \downarrow \psi$ is connected for all $a \in \mathrm{Ob}(A)$, so there exists some $\phi(F(a)) \to \psi(k_a)$ for some $k_a \in \mathrm{Ob}(K)$. Since K is filtered, there exists some $k \in \mathrm{Ob}(K)$ such that we obtain morphisms $\phi(F(a)) \to \psi(k)$. This is functorial in a, so we obtain a diagram $A \to (\phi \circ F \downarrow \psi(k))$. Since $\psi \downarrow \psi(k)$ is filtered, the diagram admits a cocone, which gives a cocone of F under the forgetful functor $\mathrm{For}\colon (\phi \downarrow \psi(k)) \to J$.

Alternative proof. — Assume that I, J and K are small. Let $F\colon J \to \mathsf{Set}$ be any diagram. Using corollary XV.2.6.7 and proposition XV.2.6.5 for the first isomorphism and proposition XV.2.7.9 (ψ is final) for the second, we have

$$\operatorname*{colim}_J F \cong \operatorname*{colim}_{i\in I} \operatorname*{colim}_{\phi(j)\to i} F(j) \cong \operatorname*{colim}_{k\in K} \operatorname*{colim}_{\phi(j)\to\psi(k)} F(j).$$

By theorem XV.2.10.1, we see that colim_J commutes with finite limits, so J is filtered.

XV.2.11. Filtered colimits commute with small products.

Small filtered colimits also commute with a small products in a much larger class of categories.

***Construction* XV.2.11.1.** — Throughout, we assume $\mathcal{C}$ to admit all small products and all small filtered colimits. Let $(I_s)_{s\in S}$ be a family of small filtered categories indexed by a set S. Then the product $K = \prod_{s\in S} I_s$ is also small filtered, and let $\pi_s : K \to I_s$ be the projection. Let $(F_s : I_s \to \mathcal{C})_{s\in S}$ be a family of functors, which induces the functor $\prod_{s\in S} F_s : K \to \mathcal{C}$. Consider $\prod_{s\in S} \lim_{I_s} F_s$ in $\mathcal{C}$. The morphisms $F_s(i_s) \to \operatorname{colim}_{I_s} F_s$ for each $i_s \in \mathrm{Ob}(I_s)$ induce $\prod_{s\in S} F_s(i_s) \to \prod_{s\in S} \operatorname{colim}_{I_s} F_s$ for all $(i_s)_{s\in S} \in \mathrm{Ob}(K)$. We obtain a morphism

$$\operatorname*{colim}_K \prod_{s\in S} (F_s \circ \pi_s) = \operatorname*{colim}_{(i_s)_s \in K} \prod_{s\in S} F_s(i_s) \longrightarrow \prod_{s\in S} \operatorname*{colim}_{i_s\in I_s} F_s(i_s).$$

We say that a category $\mathcal{C}$, which admits small products and small filtered colimits, satisfies the **IPC-property** (inductive limits product commutation property) if this morphism is an isomorphism.

For example, if $I_s = I$ for all $s \in S$, then $K = I^S$ and $\prod_{s\in S} F_s$ is a functor $I \times S \to \mathcal{C}$, and we obtain

$$\operatorname*{colim}_{(i_s)_{s\in S} \in I^S} \prod_{s\in S} F(i_s, s) \longrightarrow \prod_{s\in S} \operatorname*{colim}_{i\in I} F(i, s).$$

***Fact* XV.2.11.2.** — *If $\mathcal{C}_j$ satisfy the IPC-property for all $j \in J$, then $\prod_{j\in J} \mathcal{C}_j$ satisfies the IPC-category.*

Proof. — For any functor $F : K \to \prod_{j\in J} \mathcal{C}_j$ can be decomposed into a family of functors $F_j : K \to \mathcal{C}_j$ by post-composition with the projections $\prod_{j\in J} \mathcal{C}_j \twoheadrightarrow \mathcal{C}_j$, and vice versa. ∎

***Proposition* XV.2.11.3** (characterisation of the IPC-property). — (i) *The category* Set *satisfies the IPC-property.*

(ii) *Let $\mathcal{C}$ be a category admitting small products and small filtered colimits. If a conservative functor $\mathcal{C} \to \prod_{j\in J} \mathsf{Set}$ (with J a set) preserves small products and small filtered colimits, then $\mathcal{C}$ satisfies the IPC-property.*

Proof. — For (i), let $(F_s : I_s \to \mathsf{Set})_{s\in S}$ be a family of small filtered diagrams, and write $K = \prod_{s\in S} I_s$ (which is filtered). We need to show that the canonical morphism $\phi\colon \operatorname{colim}_{(i_s)_s\in K} \prod_{s\in S} F_s(i_s) \to \prod_{s\in S} \operatorname{colim}_{I_s} F_s$ is bijective.

For surjectivity, let $(x_s)_{s\in S} \in \prod_{s\in S} \operatorname{colim}_{I_s} F_s$. Each x_s is represented by some $x_{i_s} \in F_s(i_s)$ where $i_s \in \mathrm{Ob}(I_s)$. Let $k = (i_s)_{s\in S} \in \mathrm{Ob}(K)$, let $y_k = (x_{i_s})_{s\in S} \in \prod_{s\in S} F_s(i_s)$, and let y be the image of y_k in $\operatorname{colim}_K \prod_{s\in S} F_s$. Then $\phi(y) = (x_s)_{s\in S}$ by the definition of the colimit $\operatorname{colim}_K \prod_{s\in S} F_s$.

For injectivity, let $y, y' \in \operatorname{colim}_K \prod_{s\in S} F_s$ with $\phi(y) = \phi(y')$. By corollary XV.2.9.13, there exists a common $k = (i_s)_{s\in S} \in \mathrm{Ob}(K)$ such that y and y' are represented by elements $y_k = (x_{i_s})_{s\in S}$ and $y'_k = (x'_{i_s})_{s\in S}$, resp., of $\prod_{s\in S} F_s(i_s)$. Then for each $s \in S$, both x_{i_s} and x'_{i_s} have the same image in $\operatorname{colim}_{I_s} F_s$. By corollary XV.2.9.13, there exists some $f_s : i_s \to i'_s$ such that both x_{i_s} and x'_{i_s} have the same image under $F_s(f_s)$. Hence, y_k and y'_k have the same image under $\prod_{s\in S} F_s(f_s)$, so $y_k = y'_k$ by proposition XV.2.9.11.

For (ii), let $\phi\colon \mathcal{C} \to \prod_{j\in J} \mathsf{Set}$ be the functor from the hypothesis. Note that $\prod_{j\in J} \mathsf{Set}$ satisfies the IPC-property by (i). Let $(F_s : I_s \to \mathcal{C})_{s\in S}$ be a family of small filtered diagrams, and write $K = \prod_{s\in S} I_s$. Then for the family $(\phi \circ F_s)_{s\in S}$, the fact that ϕ preserves small products and small filtered colimits gives rise to the isomorphism

$$\phi\left(\operatorname*{colim}_{(i_s)_s\in K} \prod_{s\in S} F_s(i_s)\right) \cong \operatorname*{colim}_{(i_s)_s\in K} \prod_{s\in S} \phi(F_s(i_s)) \cong \prod_{s\in S} \operatorname*{colim}_{i_s\in I_s} \phi(F_s(i_s)) \cong \phi\left(\prod_{s\in S} \operatorname*{colim}_{i_s\in I_s} F_s(i_s)\right),$$

which is $\phi(\psi)$ where $\psi\colon \operatorname{colim}_{(i_s)_s\in K} \prod_{s\in S} F_s(i_s) \to \prod_{s\in S} \operatorname{colim}_{i_s\in I_s} F_s(i_s)$. Since ϕ is conservative, ψ is an isomorphism, showing that $\mathcal{C}$ satisfies the IPC-property. ∎

***Corollary* XV.2.11.4.** — *If $\mathcal{C}$ is a small category, then the functor category* $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ *satisfies the IPC-property.*

Proof. — Consider the evaluation functor

$$\prod_{X \in \mathrm{Ob}(\mathcal{C})} \mathrm{ev}_X \colon \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \longrightarrow \prod_{X \in \mathrm{Ob}(\mathcal{C})} \mathsf{Set}, \qquad F \longmapsto \prod_{X \in \mathrm{Ob}(\mathcal{C})} F(X),$$

which is conservative. Using that (co)limits are computed pointwise in functor categories (proposition XV.2.2.10) and that products commute (proposition XV.2.2.14), for any family of functors $(F_s)_{s\in S}$ indexed by a set S, we have

$$\prod_{X \in \mathrm{Ob}(\mathcal{C})} \mathrm{ev}_X\Big(\prod_{s\in S} F_s\Big) \cong \prod_{X \in \mathrm{Ob}(\mathcal{C})} \prod_{s \in S} (\mathrm{ev}_X \circ F_s) \cong \prod_{s\in S}\Big(\prod_{X \in \mathrm{Ob}(\mathcal{C})} \mathrm{ev}_X\Big)(F_s),$$

i. e., $\prod_{X \in \mathrm{Ob}(\mathcal{C})} \mathrm{ev}_X$ preserves small products. Similarly, using that Set satisfies the IPC-property (proposition XV.2.11.3), for any small filtered diagram $G \colon I \to \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$, we have

$$\prod_{X \in \mathrm{Ob}(\mathcal{C})} \mathrm{ev}_X(\operatorname*{colim}_I G) \cong \prod_{X \in \mathrm{Ob}(\mathcal{C})} \operatorname*{colim}_I(\mathrm{ev}_X \circ G) \cong \operatorname*{colim}_I \Big(\prod_{X \in \mathrm{Ob}(\mathcal{C})} \mathrm{ev}_X\Big)(G),$$

i. e., $\prod_{X \in \mathrm{Ob}(\mathcal{C})} \mathrm{ev}_X$ preserves small filtered colimits. Then $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ satisfies the IPC-property by proposition XV.2.11.3. ∎

Skip [KS06, Cor. 3.1.13] that R-Mod is IPC.

XV.2.12. Exact functors.

Generally in category theory, *exactness* means that something commutes with finite (co)limits.

***Definition* XV.2.12.1.** — *Let $\mathcal{C}$ be a category admitting all finite limits. A functor $F \colon \mathcal{C} \to \mathcal{D}$ is:*

(i) ***left exact*** *if it preserves finite limits in $\mathcal{D}$;*
(ii) *dually,* ***right exact*** *if it preserves finite colimits in $\mathcal{D}$;*
(iii) ***exact*** *if it is both right and left exact.*

There is a slightly more general definition which also works if the domain category does not admit all finite limits.

XV.2.12.2. — Let $F \colon \mathcal{C} \to \mathcal{D}$ be a functor and let $U \in \mathrm{Ob}(\mathcal{D})$. Recall that the comma category $F \downarrow U$ consists of pairs (X, u) where $X \in \mathrm{Ob}(\mathcal{C})$ and $u \colon F(X) \to U$ in $\mathcal{D}$. Similarly, the comma category $U \downarrow F$ consists of pairs (X, u) where $X \in \mathrm{Ob}(\mathcal{C})$ and $u \colon U \to F(X)$ in $\mathcal{D}$. The forgetful functors $\mathrm{For} \colon (F \downarrow U) \to \mathcal{C}$ and $\mathrm{For} \colon (U \downarrow F) \to \mathcal{C}$ are faithful (see definition XV.1.3.22).

***Definition* XV.2.12.3.** — *A functor $F \colon \mathcal{C} \to \mathcal{D}$ is:*

(i) ***flat*** *if the comma category $U \downarrow F$ is cofiltered for all $U \in \mathrm{Ob}(\mathcal{C})$;*
(ii) *dually,* ***coflat*** *if $F^{\mathrm{op}} \colon \mathcal{C}^{\mathrm{op}} \to \mathcal{D}^{\mathrm{op}}$ is flat, i. e., if the comma category $F \downarrow U$ is filtered for all $U \in \mathrm{Ob}(\mathcal{C})$.*(XV.11)

***Proposition* XV.2.12.4** (flat functors preserve limits)**.** — *Let $F \colon \mathcal{C} \to \mathcal{D}$ be a flat functor. If a finite diagram $G \colon I^{\mathrm{op}} \to \mathcal{C}$ has a limit in $\mathcal{C}$, then the limit $\lim_I(F \circ G) \cong F(\lim_I G)$ exists in $\mathcal{D}$.*

The dual statement about coflat functors and finite colimits also holds true.

(XV.11) Some literature calls flat (resp. coflat) functors left (resp. right) exact. We do not inflate the notion of exactness.

Proof using universal properties. — We need to show that $F(\lim_I G)$ satisfies the universal property of $\lim_I(F\circ G)$. Let $\alpha\colon \Delta_U \to F\circ G$ be a cone over $F\circ G$. This defines a finite diagram $I^{\mathrm{op}} \to (U\downarrow F)$, $i \mapsto (G(i), \alpha_i)$, which admits a cone with, say, tip $(X, f\colon U \to F(X))$ because $U\downarrow F$ is cofiltered. Passing to $\mathcal{C}$ by means of For, the cone in $U\downarrow F$ gives a cone $\Delta_X \to G$ in $\mathcal{C}$, which factorises uniquely through $g\colon X \to \lim_I G$. Passing back to $U\downarrow F$, we see that the cone in $U\downarrow F$ factorises through $(\lim_I G, F(g)\circ f)$. This factorisation is unique since any factorisation of the cone in $U\downarrow F$ gives a factorisation of the cone in $\mathcal{C}$ under For. ◻

Proof using the Yoneda lemma. — Suppose that $\mathcal{C}$ is small and $\mathcal{D}$ is locally small so that $U\downarrow F$ is small. Applying corollary XV.2.7.7 to F and $\mathrm{id}_{\mathcal{C}}$, we have

$$\mathrm{Hom}_{\mathcal{D}}(U, F(X)) \cong \operatorname*{colim}_{(Z\to X)\in(\mathrm{id}_{\mathcal{C}}\downarrow X)} \mathrm{Hom}_{\mathcal{D}}(U, F(Z)) \cong \operatorname*{colim}_{(U\to F(Z))\in(F\downarrow U)} \mathrm{Hom}_{\mathcal{C}}(Z, X)$$

for all $X \in \mathrm{Ob}(\mathcal{C})$ and all $U \in \mathrm{Ob}(\mathcal{D})$. Then for all $U \in \mathrm{Ob}(\mathcal{D})$ and all finite diagrams $G\colon I^{\mathrm{op}} \to \mathcal{C}$ we have isomorphisms

$$\begin{aligned}\lim_{i\in I} \mathrm{Hom}_{\mathcal{D}}(U, F(G(i))) &\cong \lim_{i\in I} \operatorname*{colim}_{U\to F(Z)} \mathrm{Hom}_{\mathcal{C}}(Z, G(i)) \cong \operatorname*{colim}_{U\to F(Z)} \lim_{i\in I} \mathrm{Hom}_{\mathcal{C}}(Z, G(i)) \\ &\cong \operatorname*{colim}_{U\to F(Z)} \mathrm{Hom}_{\mathcal{C}}(Z, \lim_{i\in I} G(i)) \cong \mathrm{Hom}_{\mathcal{D}}(U, F(\lim_{i\in I} G(i)))\end{aligned}$$

which are natural in U using proposition XV.2.2.8 and theorem XV.2.10.1. Hence, $F(\lim_I G)$ is a representative of the left-hand side functor, and thus $F(\lim_{i\in I} G(i)) \cong \lim_{i\in I} F(G(i))$. ◻

***Corollary* XV.2.12.5.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor, and suppose that $\mathcal{C}$ admits all finite limits (resp. colimits). Then F is left (resp. right) exact if and only if it is flat (resp. coflat).*

Proof. — Suppose that F is left exact, and let $U \in \mathrm{Ob}(\mathcal{D})$. Then $U\downarrow F$ admits all finite limits by fact XV.2.2.16, and is hence cofiltered by fact XV.2.9.6. Hence, F is flat. The converse follows from proposition XV.2.12.4. ◻

***Corollary* XV.2.12.6.** — *Let $\mathcal{C}$ be a category admitting all finite limits, and let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor. Then the following are equivalent:*

(i) *F is left exact.*
(ii) *F preserves the terminal object, binary products and equalisers.*
(iii) *F preserves the terminal object and fibre products.*

The dual statement on right exact functors also holds true.

Proof. — This follows from corollary XV.2.4.4. ◻

***Example* XV.2.12.7.** — For a ring R, let For: R-$\mathsf{Mod} \to \mathsf{Set}$ be the forgetful functor. One can check that For preserves the terminal object (the zero module is a singleton set), binary products and equalisers. Hence, For is left exact. However, it is not right exact because it does not preserve the initial object.

Maybe move this elsewhere where completeness of R-Mod is covered, e. g., monads.

***Lemma* XV.2.12.8.** — *Right (resp. left) adjoints are flat (resp. coflat). In particular, if the domain category admits all finite limits, then right (resp. left) adjoints are left (resp. right) exact.*

Proof. — Let $L \dashv R$ be an adjoint pair $\mathcal{C} \rightleftarrows \mathcal{D}$. Since for any $V \in \mathrm{Ob}(\mathcal{D})$ and any $U \in \mathrm{Ob}(\mathcal{C})$ we have the adjunction isomorphism $\mathrm{Hom}_{\mathcal{C}}(U, R(V)) \cong \mathrm{Hom}_{\mathcal{C}}(L(U), V)$, we have an isomorphism $(U\downarrow R) \cong L(U)/\mathcal{D}$ for any $U \in \mathrm{Ob}(\mathcal{C})$. This category $L(U)/\mathcal{D}$ is cofiltered since it admits an initial element, namely $\mathrm{id}_{L(U)}$. The second statement follows from proposition XV.2.12.4. ◻

***Fact* XV.2.12.9.** — *Flat (resp. coflat) functors are final (resp. initial). In particular, right (resp. left) adjoints are final (resp. initial).*

Proof. — The first statement follows since (co)filtered categories are connected. The second statement follows from lemma XV.2.12.8. ∎

***Fact* XV.2.12.10.** — (i) *Let $\mathcal{C}$ be a locally small category admitting finite limits and finite colimits. Then the hom-functor* $\mathrm{Hom}_{\mathcal{C}}\colon \mathcal{C}^{\mathrm{op}} \times \mathcal{C} \to \mathsf{Set}$ *is left-exact in both arguments.*

(ii) *A small product of (co)flat functors is (co)flat. In particular, if all domain categories admit all finite limits, then a small product of left (resp. right) exact functors is left (resp. right) exact.*

(iii) *Let $\mathcal{C}$ be a category admitting limits of shape I for I small. Then* $\lim_I\colon \mathrm{Fun}(I^{\mathrm{op}}, \mathcal{C}) \to \mathcal{C}$ *is coflat. The dual statement for* $\mathrm{colim}_I\colon \mathrm{Fun}(I, \mathcal{C}) \to \mathcal{C}$ *being flat also holds true.*

(iv) *Let I be a small filtered category. Then* $\mathrm{colim}_I\colon \mathrm{Fun}(I, \mathsf{Set}) \to \mathsf{Set}$ *is exact. Dually,* $\lim_I\colon \mathrm{Fun}(I^{\mathrm{op}}, \mathsf{Set}) \to \mathsf{Set}$ *is exact for I small cofiltered.*

Proof. — (i) follows from proposition XV.2.2.8 and corollary XV.2.12.5.

For (ii), let $F_i\colon \mathcal{C}_i \to \mathcal{D}_i$ be flat functors for each $i \in I$. Then $\prod_{i\in I} F_i$ is flat because $(\prod_{i\in I} F_i \downarrow (U_i)_i) \cong \prod_{i\in I}(F_i \downarrow U_i)$ for all $(U_i)_i \in \mathrm{Ob}(\prod_{i\in I} \mathcal{D}_i)$ and the product of filtered categories is again filtered, .

(iii) follows from lemma XV.2.12.8 and the fact that $\lim_I$ admits a right adjoint, namely the diagonal functor $\Delta\colon \mathcal{C} \to \mathrm{Fun}(I^{\mathrm{op}}, \mathcal{C})$ (observation XV.2.1.12).

(iv) follows from theorem XV.2.10.1 and proposition XV.2.2.14. ∎

***Lemma* XV.2.12.11.** — *Flat (resp. coflat) functors preserve monomorphisms (resp. epimorphisms). In particular, left (resp. right) exact functors preserve monomorphisms (resp. epimorphisms).*

Proof. — Let $F\colon \mathcal{C} \to \mathcal{D}$ be a flat functor. Observe that a morphism $f\colon X \to Y$ in $\mathcal{C}$ is a monomorphism if and only if the square

$$\begin{tikzcd} X \arrow[r, "\mathrm{id}_X"] \arrow[d, "\mathrm{id}_X"'] & X \arrow[d, "f"] \\ X \arrow[r, "f"] & Y \end{tikzcd}$$

is a pullback. The claim now follows since flat functors preserve limits (proposition XV.2.12.4). ∎

***Corollary* XV.2.12.12.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor of small categories. If $u\colon A \to B$ is an epimorphism in* $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$*, then* $\hat{F}(u)\colon \hat{F}(A) \to \hat{F}(B)$ *(see notation XV.2.8.13) is an epimorphism in* $\mathrm{Fun}(\mathcal{D}^{\mathrm{op}}, \mathsf{Set})$.

Proof. — Recall from theorem XV.2.8.12 that $\hat{F}$ is cocontinuous. By corollary XV.2.12.5, $\hat{F}$ is coflat and thus preserves epimorphisms. ∎

***Notation* XV.2.12.13.** — Let $\mathcal{C}$ be category admitting all limits of small shape I and all finite limits. We say that limits of shape I are **left exact** (resp. **right exact**, resp. **exact**) if $\lim_I\colon \mathrm{Fun}(I^{\mathrm{op}}, \mathcal{C}) \to \mathcal{C}$ is left exact (resp. right exact, resp. exact). We dually define the concept for $\mathrm{colim}_I\colon \mathrm{Fun}(I, \mathcal{C}) \to \mathcal{C}$.

***Lemma* XV.2.12.14.** — *Let I be a small connected category. Let $\mathcal{C}$ be a category admitting finite limits and colimits of shape I. If colimits of shape I are exact, then they are stable under base change.*

Proof. — Let $F\colon I \to \mathcal{C}$ be a colimit and let $Y \to Z$ and $\mathrm{colim}_{i\in I} F(i) \to Z$ be morphisms in $\mathcal{C}$. Then

$$\operatorname*{colim}_{i\in I}(F(i) \times_Z Y) \cong (\operatorname*{colim}_I F) \times_{\mathrm{colim}_I Z} (\operatorname*{colim}_I Y) \cong (\operatorname*{colim}_I F) \times_Z Y$$

where the second isomorphism follows since I is connected (fact XV.2.7.4). ∎

***Lemma* XV.2.12.15.** — *Let $\phi\colon J \to I$ be a coflat (resp. flat) functor. If I is filtered (resp. cofiltered), then J is filtered (resp. cofiltered).*

Proof. — Apply lemma XV.2.10.4 to $\psi = \mathrm{id}_I\colon I \to I$. □

***Fact* XV.2.12.16** (flatness is stable under composition)**.** — *If $F\colon \mathcal{C} \to \mathcal{D}$ and $G\colon \mathcal{D} \to \mathcal{E}$ are (co)flat functors, then $G \circ F$ is (co)flat.*

Proof. — We prove the coflat case. The category $G \downarrow Z$ is filtered for all $Z \in \mathrm{Ob}(\mathcal{E})$. Moreover, the embedding functor $F'\colon (G \circ F \downarrow Z) \to (G \downarrow Z)$ is again coflat. Indeed, for any $Y \in \mathrm{Ob}(\mathcal{D})$ and any $g\colon G(Y) \to Z$ in $\mathcal{E}$, we have an isomorphism $(F' \downarrow Y) \xrightarrow{\sim} (F \downarrow Y)$ where we identified Y with (Y, g): the objects of $F' \downarrow Y$ are of the form $(X, f\colon G(F(X)) \to Z, h\colon F(X) \to Y)$ where h is a morphism in $G \downarrow Z$, that is, $g \circ G(h) = f$. By forgetting f, we obtain the isomorphism. Now, $F \downarrow Y$ is filtered by hypothesis. By lemma XV.2.12.15, $G \circ F \downarrow Z$ is filtered. □

***Proposition* XV.2.12.17.** — *Let $\mathcal{C}$ be a small category admitting finite colimits. Then a functor $A\colon \mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ is left exact if and only if $\mathrm{h}_{\mathcal{C}} \downarrow A$ is filtered where A is understood as an object of* $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$.

Proof. — The proof is similar to proposition XV.2.12.4. Suppose that $\mathrm{h}_{\mathcal{C}} \downarrow A$ is filtered. Recall from the co-Yoneda lemma XV.2.8.10 that $A \cong \mathrm{colim}_{(Y \to A) \in (\mathrm{h}_{\mathcal{C}} \downarrow A)} \mathrm{Hom}_{\mathcal{C}}(-, Y)$. Since $\mathrm{Hom}_{\mathcal{C}}(-, Y)$ preserves finite limits and finite limits commute with the small filtered colimit indexed by $\mathrm{h}_{\mathcal{C}} \downarrow A$ (theorem XV.2.10.1), A preserves finite limits.

Conversely, suppose that A is left exact. Then for any finite diagram $F\colon I \to (\mathrm{h}_{\mathcal{C}} \downarrow A)$, write $F(i) = (X_i, f_i)$ where $X_i \in \mathrm{Ob}(\mathcal{C})$ and $f_i \in A(X) \cong \mathrm{Hom}_{\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})}(\mathrm{Hom}_{\mathcal{C}}(-, X), A)$ by the Yoneda lemma XV.1.6.5. If u is the image of $(f_i)_i$ under $\lim_{i \in I} A(X_i) \cong A(\mathrm{colim}_{i \in I} X_i)$, one can then directly check that $(\lim_{i \in I} X_i, u)$ is the colimit of F. Hence, $\mathrm{h}_{\mathcal{C}} \downarrow A$ is filtered. □

***Proposition* XV.2.12.18.** — *Let $F\colon \mathcal{C} \to \mathcal{D}$ be a functor of small categories. Then F is flat if and only if $\hat{F}\colon \mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set}) \to \mathrm{Fun}(\mathcal{D}^{\mathrm{op}}, \mathsf{Set})$ (see notation XV.2.8.13) is left exact. Note that by theorem XV.2.8.12, $\hat{F}$ is always right exact.*

Proof. — Recall from proposition XV.2.8.14 that $\hat{F} \cong \mathrm{Lan}_{F^{\mathrm{op}}}$. Since $\mathrm{Lan}_{F^{\mathrm{op}}}$ is pointwise by theorem XV.2.8.12, we can write

$$\mathrm{Lan}_{F^{\mathrm{op}}} A(U) \cong \operatorname*{colim}_{F^{\mathrm{op}}(V^{\mathrm{op}}) \to U^{\mathrm{op}}} A(V) = \operatorname*{colim}_{U \to F(V)} A(V).$$

It follows from theorem XV.2.10.1 that $\mathrm{Lan}_{F^{\mathrm{op}}}$ preserves finite limits in $\mathrm{Fun}(\mathcal{C}^{\mathrm{op}}, \mathsf{Set})$ if and only if the indexing category $U \downarrow F$ of the colimit is filtered for all $U \in \mathrm{Ob}(\mathcal{C})$. Here, we implicitly used that limits of functors $\mathcal{C}^{\mathrm{op}} \to \mathsf{Set}$ are computed pointwise (proposition XV.2.2.10). □

Skip [KS06] 3.3.14, 3.3.15, 3.3.16, 3.3.17, 3.3.18, 3.3.19 on small functors and Kan extensions.
Move this elsewhere, maybe where comma categories are defined.

XV.2.13. Properties of comma categories.

***Lemma* XV.2.13.1.** — *Consider a quasi-commutative diagram of functors*

$$\begin{tikzcd}
\mathcal{C}_1 \arrow[r, "F_1"] \arrow[d, "\phi"'] & \mathcal{E}_1 \arrow[d, "\chi"] & \mathcal{D}_1 \arrow[l, "G_1"'] \arrow[d, "\psi"] \\
\mathcal{C}_2 \arrow[r, "F_2"] & \mathcal{E}_2 & \mathcal{D}_2. \arrow[l, "G_2"']
\end{tikzcd}$$

This defines a functor $\theta\colon (F_1 \downarrow G_1) \to (F_2 \downarrow G_2)$.

(i) *If ϕ and ψ are faithful, then θ is faithful.*

(ii) *If ϕ and ψ are fully faithful and if χ is faithful, then θ is fully faithful.*

(iii) *If ϕ and ψ are equivalences of categories and if χ is fully faithful, then θ is an equivalence of categories.*

Proof. — The functor θ maps a morphism $(X, Y, f) \to (X', Y', f')$ in $F_1 \downarrow G_1$ to a morphism $(\phi(X), \psi(Y), \chi(f)) \to (\phi(X'), \psi(Y'), \chi(f'))$ in $F_2 \downarrow G_2$ (by abuse of notation, we write $\chi(f)$ and $\chi(f')$ instead of the composition with the natural isomorphisms) by sending $(u, v) \mapsto (\phi(u), \psi(v))$ such that the diagrams

$$\begin{tikzcd}
F_1(X) \arrow[r, "f"] \arrow[d, "F_1(u)"'] & G_1(Y) \arrow[d, "G_1(v)"] \\
F_1(X') \arrow[r, "f'"] & G_1(Y')
\end{tikzcd} \tag{XV.2.13.1.1}$$

and

$$\begin{tikzcd}
& \chi(F_1(X)) \arrow[r, "\chi(f)"] \arrow[d, "\wr"'] & \chi(G_1(Y)) \arrow[d, "\wr"] \arrow[dr, "\chi(G_1(v))"] & \\
& F_2(\phi(X)) \arrow[dl, "\sim"'] \arrow[d, "F_2(\phi(u))"] & G_2(\psi(Y)) \arrow[d, "G_2(\psi(v))"'] & \chi(G_1(Y')) \arrow[dl, "\sim"] \\
\chi(F_1(X)) \arrow[dr, "\chi(F_1(u))"'] & F_2(\phi(X')) \arrow[d, "\wr"'] & G_2(\psi(Y')) \arrow[d, "\wr"] & \\
& \chi(F_1(X')) \arrow[r, "\chi(f')"] & \chi(G_1(Y')) &
\end{tikzcd} \tag{XV.2.13.1.2}$$

commute (the triangles and the outer octagon commute by naturality of the natural isomorphisms).

(i) is easy. For (ii), fully faithfulness of ϕ and ψ give for any morphism in $F_2 \downarrow G_2$ a unique pair of morphisms $(u\colon X \to X', v\colon Y \to Y')$ such that $(\phi(u), \phi(v))$ is that morphism in $F_2 \downarrow G_2$. In particular, diagram (XV.2.13.1.2) commutes. Since χ is faithful, diagram (XV.2.13.1.1) commutes. For (iii), we have a quasi-commutative diagram

$$\begin{tikzcd}
\mathscr{C}_1 \arrow[r, "F_1"] \arrow[d, "\phi", "\wr"'] \arrow[dd, bend right=60, equal, "\mathrm{id}_{\mathscr{C}_1}"'] & \mathscr{E}_1 \arrow[d, "\chi"] & \mathscr{D}_1 \arrow[l, "G_1"'] \arrow[d, "\psi"', "\wr"] \arrow[dd, bend left=60, equal, "\mathrm{id}_{\mathscr{D}_1}"] \\
\mathscr{C}_2 \arrow[r, "F_2"] \arrow[d, "\phi^{-1}", "\wr"'] & \mathscr{E}_2 & \mathscr{D}_2 . \arrow[l, "G_2"'] \arrow[d, "\psi^{-1}"', "\wr"] \\
\mathscr{C}_1 \arrow[r, "F_1"] & \mathscr{E}_1 \arrow[u, "\chi"'] & \mathscr{D}_1 . \arrow[l, "G_1"']
\end{tikzcd} \tag{XV.2.13.1.3}$$

Note that the bottom two squares define a morphism $\theta^{-1}\colon (F_2 \downarrow G_2) \to (F_1 \downarrow G_1)$ by sending

$$\begin{tikzcd}
\chi(F_1(\phi^{-1}(X))) \arrow[r, "\sim"] \arrow[d, "\chi(F_1(\phi^{-1}(u)))"'] & F_2(X) \arrow[r, "f"] \arrow[d, "F_2(u)"'] & G_2(Y) \arrow[r, "\sim"] \arrow[d, "G_1(v)"] & \chi(G_1(\psi^{-1}(Y))) \arrow[d, "\chi(G_1(\psi^{-1}(v)))"] \\
\chi(F_1(\phi^{-1}(X'))) \arrow[r, "\sim"] & F_2(X') \arrow[r, "f'"] & G_2(Y') \arrow[r, "\sim"] & \chi(G_1(\psi^{-1}(Y')))
\end{tikzcd}$$

to

$$\begin{tikzcd}
F_1(\phi^{-1}(X)) \arrow[r, "\chi^{-1}(f)"] \arrow[d, "F_1(\phi^{-1}(u))"'] & G_1(\psi^{-1}(Y)) \arrow[d, "G_1(\psi^{-1}(v))"] \\
F_1(\phi^{-1}(X')) \arrow[r, "\chi^{-1}(f')"] & G_1(\psi^{-1}(Y'))
\end{tikzcd}$$

where $\chi^{-1}(f)$ and $\chi^{-1}(f')$ are abuse of notation. Diagram (XV.2.13.1.3) shows that $\theta^{-1}\circ\theta \cong \mathrm{id}_{F_1\downarrow G_1}$. A similar diagram shows that $\theta\circ\theta^{-1} \cong \mathrm{id}_{F_2\downarrow G_2}$.

Add [KS06, Prop. 3.4.3, 3.4.4, 3.4.5, Cor. 3.4.6] later if necessary.

§ XV.3. CATEGORICAL ALGEBRA

Categorical algebra tries to generalise concepts from algebra from a categorical point of view. We will only scratch the surface of this deep field, which is a part of *higher category theory*.

XV.3.1. Monoidal categories, functors and natural transformations.

The following mimics the definition of monoids from algebra.

Definition* XV.3.1.1** (MacLane 1963). — *A* ***monoidal category(XV.12) *is a category $\mathcal{C}$ together with the following datum:*

(i) *a bifunctor $-\otimes- \colon \mathcal{C}\times\mathcal{C}\to\mathcal{C}$ called the* ***tensor product****;*
(ii) *a natural isomorphism a of functors $\mathcal{C}\times\mathcal{C}\times\mathcal{C}\to\mathcal{C}$ of the form $a_{X,Y,Z}\colon (X\otimes Y)\otimes Z \xrightarrow{\sim} X\otimes(Y\otimes Z)$ called the* ***associator****;*
(iii) *an object $1\in\mathrm{Ob}(\mathcal{C})$ called the* ***unit object****;*
(iv) *a natural isomorphism λ of functors $\mathcal{C}\to\mathcal{C}$ of the form $\lambda_X\colon 1\otimes X\xrightarrow{\sim} X$ called the* ***left unitor****;*
(v) *a natural isomorphism ρ of functors $\mathcal{C}\to\mathcal{C}$ of the form $\rho_X\colon X\otimes 1\xrightarrow{\sim} X$ called the* ***right unitor****;*

such that the following coherence conditions are satisfied, i. e., the following diagrams commute:

(i) *the* ***triangle identity*** *(which are different from* triangle identities *for adjunctions)*

$$\begin{tikzcd}
(X\otimes 1)\otimes Y \arrow[rr, "a_{X,1,Y}"] \arrow[dr, "\rho_X\otimes Y"'] & & X\otimes(1\otimes Y) \arrow[dl, "X\otimes\lambda_Y"] \\
& X\otimes Y &
\end{tikzcd}$$

for any $X,Y,Z\in\mathrm{Ob}(\mathcal{C})$;
(ii) *the* ***pentagon identity***

$$\begin{tikzcd}
& ((X\otimes Y)\otimes Z)\otimes W \arrow[dl, "a_{X,Y,Z}\otimes W"'] \arrow[dr, "a_{X\otimes Y,Z,W}"] & \\
(X\otimes(Y\otimes Z))\otimes W \arrow[d, "a_{X,Y\otimes Z,W}"'] & & (X\otimes Y)\otimes(Z\otimes W) \arrow[d, "a_{X,Y,Z\otimes W}"] \\
X\otimes((Y\otimes Z)\otimes W) \arrow[rr, "X\otimes a_{Y,Z,W}"] & & X\otimes(Y\otimes(Z\otimes W))
\end{tikzcd} \tag{XV.3.1.1.1}$$

for any $X,Y,Z,W\in\mathrm{Ob}(\mathcal{C})$.

We say that a monoidal category $\mathcal{C}$ is ***unital*** *to emphasise the existence of $(1,\lambda,\rho)$ together with the commutativity of the triangle identity.*

A monoidal category $\mathcal{C}$ is ***strict*** *if associator, left and right unitor are identities and not just isomorphisms. For example, this means that $(X\otimes Y)\otimes Z = X\otimes(Y\otimes Z)$ and $(f\otimes g)\otimes h = f\otimes(g\otimes h)$ for all $X,Y,Z\in\mathrm{Ob}(\mathcal{C})$ and all morphisms f,g,h in $\mathcal{C}$.*

(XV.12) Older literature calls this *tensor category*. Modern usage requires the tensor product to satisfy a universal property in the traditional sense.

***Example* XV.3.1.2.** — The following are monoidal categories.

(i) A monoid M understood as a discrete category with $a \otimes b := ab$ for all $a, b \in M$.

(ii) For a commutative ring k, the category of k-modules $k\text{-}\mathsf{Mod}$ with tensor product $\otimes_k$ and the unit object k.

(iii) For a (non-commutative) k-algebra A, the category of (A, A)-bimodules $(A, A)\text{-}\mathsf{BMod} \cong (A \otimes_k A^{\mathrm{op}})\text{-}\mathsf{Mod}$ with tensor product $\otimes_A$ and unit object A.

(iv) For a small category $\mathcal{C}$, the category of endofunctors $\mathrm{End}(\mathcal{C})$ with tensor product $\circ$ and unit object $\mathrm{id}_{\mathcal{C}}$. In fact, this category is strict monoidal.

(v) A category $\mathcal{C}$ admitting finite products with tensor product $\times$ and the terminal object as the unit object. Such categories are called **cartesian categories**.

(vi) A category $\mathcal{C}$ admitting finite coproducts with tensor product $\amalg$ and the initial object as the unit object. Such categories are called **cocartesian categories**.

(vii) For a group G and a field k, the category of G-modules $G\text{-}\mathsf{Rep}/k \cong kG\text{-}\mathsf{Mod}$ over k. Recall that the tensor product satisfies $g(v \otimes w) = gv \otimes gw$.

***Definition* XV.3.1.3.** — *Let $\mathcal{C}$ be a monoidal category. Define the following functors for $\mathcal{C}$:*

(i) *the tensor product $- \otimes^r - : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ by $X \otimes^r Y := Y \otimes X$;*

(ii) *the associator a^r by*

$$a^r_{X,Y,Z} \colon (X \otimes^r Y) \otimes^r Z = Z \otimes (Y \otimes X) \xrightarrow[\sim]{a^{-1}_{Z,Y,X}} (Z \otimes Y) \otimes X = X \otimes^r (Y \otimes^r Z);$$

(iii) *the left unitor λ^r by $\lambda^r_X := \rho_X \colon 1 \otimes^r X = X \otimes 1 \xrightarrow{\sim} X$;*

(iv) *the right unitor ρ^r by $\rho^r_X := \lambda_X \colon X \otimes^r 1 = 1 \otimes X \xrightarrow{\sim} X$.*

Then the monoidal category $(\mathcal{C}, \otimes^r, a^r, 1, \lambda^r, \rho^r)$ is the ***reverse*** *monoidal category of $\mathcal{C}$.*

***Remark* XV.3.1.4.** — If $\mathcal{C}$ is a monoidal category, then the opposite category $\mathcal{C}^{\mathrm{op}}$ is obviously monoidal. Note that it is different from the reverse monoidal category.

Definition* XV.3.1.5.** — *A* ***monoidal functor *$\mathcal{C} \to \mathcal{D}$ of monoidal categories is a functor $F \colon \mathcal{C} \to \mathcal{D}$ with:*

(i) *a natural isomorphism μ of bifunctors of the form $\mu_{X,Y} \colon F(X \otimes Y) \xrightarrow{\sim} F(X) \otimes F(Y)$;*

(ii) *an isomorphism $\eta \colon F(1) \xrightarrow{\sim} 1$ in $\mathcal{D}$;*

such that the following diagrams commute:

(i) *associativity:*

$$\begin{tikzcd} F((X \otimes Y) \otimes Z) \arrow[rr, "F(a_{X,Y,Z})"] \arrow[d, "\mu_{X\otimes Y,Z}"'] & & F(X \otimes (Y \otimes Z)) \arrow[d, "\mu_{X,Y\otimes Z}"] \\ F(X \otimes Y) \otimes F(Z) \arrow[d, "\mu_{X,Y}\otimes F(Z)"'] & & F(X) \otimes F(Y \otimes Z) \arrow[d, "F(X)\otimes\mu_{Y,Z}"] \\ (F(X) \otimes F(Y)) \otimes F(Z) \arrow[rr, "a_{F(X),F(Y),F(Z)}"] & & F(X) \otimes (F(Y) \otimes F(Z)) \end{tikzcd} \tag{XV.3.1.5.1}$$

for any $X, Y, Z \in \mathrm{Ob}(\mathcal{C})$;

(ii) *unitality:*

$$\begin{tikzcd} F(1 \otimes X) \arrow[r, "F(\lambda_X)"] \arrow[d, "\mu_{1,X}"'] & F(X) \\ F(1) \otimes F(X) \arrow[r, "\eta\otimes F(X)"] & 1 \otimes F(X), \arrow[u, "\lambda_{F(X)}"'] \end{tikzcd} \qquad \begin{tikzcd} F(X \otimes 1) \arrow[r, "F(\rho_X)"] \arrow[d, "\mu_{X,1}"'] & F(X) \\ F(X) \otimes F(1) \arrow[r, "F(X)\otimes\eta"] & F(X) \otimes 1 \arrow[u, "\rho_{F(X)}"'] \end{tikzcd} \tag{XV.3.1.5.2}$$

for any $X \in \mathrm{Ob}(\mathcal{C})$.

We say that a monoidal functor $F : \mathcal{C} \to \mathcal{D}$ is ***strict*** *if μ and η are identities and not just isomorphisms.*

Definition* XV.3.1.6.** — *A* ***monoidal natural transformation *between monoidal functors $F, G : \mathcal{C} \to \mathcal{D}$ is a natural transformation $\theta : F \to G$ such that the following diagrams commute:*

(i) *associativity:*

$$\begin{tikzcd} F(X \otimes Y) \arrow[r, "\mu^F_{X,Y}", "\sim"'] \arrow[d, "\theta_{X\otimes Y}"'] & F(X) \otimes F(Y) \arrow[d, "\theta_X \otimes \theta_Y"] \\ G(X \otimes Y) \arrow[r, "\mu^G_{X,Y}", "\sim"'] & G(X) \otimes G(Y) \end{tikzcd}$$

for any $X, Y \in \mathrm{Ob}(\mathcal{C})$;

(ii) *unitality:*

$$\begin{tikzcd} F(1) \arrow[rr, "\theta_1"] \arrow[dr, "\sim", "\eta^F"'] & & G(1) \arrow[dl, "\sim"', "\eta^G"] \\ & 1. & \end{tikzcd}$$

If θ is even a natural isomorphism, then we say that it is a ***monoidal natural isomorphism***.

Definition* XV.3.1.7.** — *A* ***monoidal equivalence *of monoidal categories $\mathcal{C}$ and $\mathcal{D}$ is a monoidal functor $F : \mathcal{C} \to \mathcal{D}$ such that there exists a monoidal functor $G : \mathcal{D} \to \mathcal{C}$, the monoidal quasi-inverse, such that $G \circ F \cong \mathrm{id}_{\mathcal{C}}$ and $F \circ G \cong \mathrm{id}_{\mathcal{D}}$ as monoidal natural isomorphisms.*

***Lemma* XV.3.1.8.** — *Let $F : \mathcal{C} \to \mathcal{D}$ be a monoidal functor. If F is an equivalence of ordinary categories, then there exists a monoidal quasi-inverse $G : \mathcal{D} \to \mathcal{C}$ such that there are monoidal natural isomorphisms $G \circ F \cong \mathrm{id}_{\mathcal{C}}$ and $F \circ G \cong \mathrm{id}_{\mathcal{D}}$.*

Proof. — By proposition XV.1.8.7, we can pick an ordinary quasi-inverse $G : \mathcal{D} \to \mathcal{C}$ such that $G \dashv F$ are adjoint. Then define the natural isomorphisms $\mu^G_{X,Y} : G(X \otimes Y) \xrightarrow{\sim} G(X) \otimes G(Y)$ as the adjunct of

$$X \otimes Y \overset{\eta_X \otimes \eta_Y}{\longrightarrow} F(G(X)) \otimes F(G(Y)) \overset{\mu^{F,-1}_{G(X),G(Y)}}{\longrightarrow} F(G(X) \otimes G(Y))$$

and the isomorphism $\eta^G : G(1) \xrightarrow{\sim} 1$ as the adjunct of $\eta^{F,-1} : 1 \xrightarrow{\sim} F(1)$. One can use adjunctions to check that G is monoidal and that the natural isomorphisms are monoidal as well. ◻

Historically, the original definition of monoidal categories required even more coherence axioms for unitors, which were later reduced by Kelly.

***Lemma* XV.3.1.9** (Kelly 1964)**.** — *Let $\mathcal{C}$ be a monoidal category.*

(i) *The diagrams*

$$\begin{tikzcd} (1 \otimes X) \otimes Y \arrow[rr, "a_{1,X,Y}"] \arrow[dr, "\lambda_X \otimes Y"'] & & 1 \otimes (X \otimes Y) \arrow[dl, "\lambda_{X \otimes Y}"] \\ & X \otimes Y, & \end{tikzcd} \qquad \begin{tikzcd} (X \otimes Y) \otimes 1 \arrow[rr, "a_{X,Y,1}"] \arrow[dr, "\rho_{X \otimes Y}"'] & & X \otimes (Y \otimes 1) \arrow[dl, "X \otimes \rho_Y"] \\ & X \otimes Y & \end{tikzcd}$$

commute for any $X, Y \in \mathrm{Ob}(\mathcal{C})$.

(ii) *We have* $\lambda_1 = \rho_1 \colon 1 \otimes 1 \xrightarrow{\sim} 1$. *In fact,* $\mathcal{C}$ *being unital is equivalent to giving an object* $1 \in \mathrm{Ob}(\mathcal{C})$ *such that the functors* $X \mapsto X \otimes 1$ *and* $X \mapsto 1 \otimes X$ *are fully faithful, and an isomorphism* $\sigma \colon 1 \otimes 1 \xrightarrow{\sim} 1$.

Proof. — For (i), consider the diagram

$$\begin{tikzcd}
& (1\otimes(1\otimes X))\otimes Y \arrow[r, "a_{1,1\otimes X,Y}"] \arrow[d, "(1\otimes\lambda_X)\otimes Y"] & 1\otimes((1\otimes X)\otimes Y) \arrow[d, "1\otimes(\lambda_X\otimes Y)"'] \arrow[dr, "1\otimes a_{1,X,Y}"] & \\
((1\otimes 1)\otimes X)\otimes Y \arrow[ur, "a_{1,1,X}\otimes Y"] \arrow[r, "(\rho_1\otimes X)\otimes Y"] \arrow[drr, "a_{1\otimes 1,X,Y}"'] & (1\otimes X)\otimes Y \arrow[r, "a_{1,X,Y}"] & 1\otimes(X\otimes Y) & 1\otimes(1\otimes(X\otimes Y)) \arrow[l, "1\otimes\lambda_{X\otimes Y}"'] \\
& & (1\otimes 1)\otimes(X\otimes Y). \arrow[u, "\rho_1\otimes(X\otimes Y)"] \arrow[ur, "a_{1,1,X\otimes Y}"'] &
\end{tikzcd}$$

The upper left triangle commutes by applying $-\otimes Y$ to the triangle identity. The upper square and lower left quadrilateral commute by naturality of a. The lower right triangle commutes by the triangle identity once again. The outer pentagon commutes by the pentagon identity. Since every arrow is an isomorphism, we conclude that the upper right triangle commutes, which is the desired triangle after applying $1 \otimes -$. But $1 \otimes -$ is an equivalence of categories (it is isomorphic to $\mathrm{id}_{\mathcal{C}}$ by λ), so the statement follows.

Let us prove (ii). By naturality of λ, the square

$$\begin{tikzcd}
1\otimes(1\otimes 1) \arrow[r, "\lambda_{1\otimes 1}"] \arrow[d, "1\otimes\lambda_1"'] & 1\otimes 1 \arrow[d, "\lambda_1"] \\
1\otimes 1 \arrow[r, "\lambda_1"] & 1
\end{tikzcd}$$

commutes. Since λ_1 is an isomorphism, we obtain $1 \otimes \lambda_1 = \lambda_{1\otimes 1}$. Using (i) and the triangle identity, we obtain $\lambda_1 \otimes 1 = \lambda_{1\otimes 1} \circ a_{1,1,1} = (1 \otimes \lambda_1) \circ a_{1,1,1} = \rho_1 \otimes 1$. Since $-\otimes 1$ is an equivalence of categories and, in particular, faithful, we obtain $\lambda_1 = \rho_1$. Of course, $-\otimes 1$ and $1 \otimes -$ are equivalences and, in particular, fully faithful.

Conversely, suppose that $-\otimes 1$ and $1 \otimes -$ are fully faithful, and let $\sigma \colon 1 \otimes 1 \xrightarrow{\sim} 1$ be an isomorphism. Then under $-\otimes 1$, there exists a unique isomorphism $\rho_X \colon X \otimes 1 \xrightarrow{\sim} X$, natural in X, that is the preimage of $(X \otimes \sigma) \circ a_{X,1,1} \colon (X \otimes 1) \otimes 1 \xrightarrow{\sim} X \otimes (1 \otimes 1) \xrightarrow{\sim} X \otimes 1$. We similarly obtain $\lambda_X \colon 1 \otimes X \xrightarrow{\sim} X$ such that $1 \otimes \lambda_X = (\sigma \otimes X) \circ a^{-1}_{1,1,X}$. Next, consider the diagram

$$\begin{tikzcd}
& (X\otimes(1\otimes 1))\otimes Y \arrow[r, "a_{X,1\otimes 1,Y}"] \arrow[d, "(X\otimes\sigma)\otimes Y"] & X\otimes((1\otimes 1)\otimes Y) \arrow[d, "X\otimes(\sigma\otimes Y)"'] \arrow[dr, "X\otimes a_{1,1,Y}"] & \\
((X\otimes 1)\otimes 1)\otimes Y \arrow[ur, "a_{X,1,1}\otimes Y"] \arrow[r, "(\rho_X\otimes 1)\otimes Y"] \arrow[drr, "a_{X\otimes 1,1,Y}"'] & (X\otimes 1)\otimes Y \arrow[r, "a_{X,1,Y}"] & X\otimes(1\otimes Y) & X\otimes(1\otimes(1\otimes Y)) \arrow[l, "X\otimes\lambda_{1\otimes Y}"'] \\
& & (X\otimes 1)\otimes(1\otimes Y). \arrow[u, "\rho_X\otimes(1\otimes Y)"] \arrow[ur, "a_{X,1,1\otimes Y}"'] &
\end{tikzcd}$$

The upper left (resp. right) triangle commutes by definition or ρ (resp. λ). The middle square and the lower left quadrilateral commute by naturality of a. Since the outer pentagon commutes by the pentagon identity, we conclude that the lower right triangle commutes, which gives the desired diagram because λ_Y is a natural isomorphism. Alternatively, note that the outer square

$$\begin{tikzcd}
(X\otimes 1)\otimes(1\otimes Y) \arrow[r, "\rho_X\otimes(1\otimes Y)"] \arrow[d, "(X\otimes 1)\otimes\lambda_Y"'] & X\otimes(1\otimes Y) \arrow[d, "X\otimes\lambda_Y"] \\
(X\otimes 1)\otimes Y \arrow[r, "\rho_X\otimes Y"] \arrow[ur, "a_{X,1,Y}"] & X\otimes Y
\end{tikzcd}$$

commutes since $\otimes$ is a bifunctor. We have seen that the upper triangle commutes, whence the lower triangle commutes. ◻

***Fact* XV.3.1.10** (unit object is essentially unique). — *Let $(1, \lambda, \rho)$ and $(1', \lambda', \rho')$ be two unit objects with their unitors. Then there exists a unique isomorphism $\sigma\colon 1' \xrightarrow{\sim} 1$ such that the diagrams*

$$\begin{tikzcd} 1'\otimes X \arrow[rr, "\sigma\otimes X"] \arrow[dr, "\lambda'_X"'] & & 1\otimes X \arrow[dl, "\lambda_X"] \\ & X, & \end{tikzcd} \qquad \begin{tikzcd} X\otimes 1' \arrow[rr, "X\otimes\sigma"] \arrow[dr, "\rho'_X"'] & & X\otimes 1 \arrow[dl, "\rho_X"] \\ & X, & \end{tikzcd}$$

commute for any $X \in \mathrm{Ob}(\mathcal{C})$.

Proof. — Suppose that such a σ exists. Taking $X = 1$ in the left triangle, we obtain $\lambda'_1 = \lambda_1 \circ (\sigma \otimes 1) = \rho_1 \circ (\sigma \otimes 1)$ by lemma XV.3.1.9. Naturality of ρ further gives $\rho_1 \circ (\sigma \otimes 1) = \sigma \circ \rho_{1'}$. Thus, we must define $\sigma := \lambda'_1 \circ \rho_{1'}^{-1}$. Suppose that σ is defined that way. In the diagram

$$\begin{tikzcd} (1'\otimes 1)\otimes X \arrow[dr, "a_{1',1,X}"] \arrow[drr, bend left, "\lambda'_1\otimes X"] \arrow[ddr, bend right, "\rho_{1'}\otimes X"'] & & \\ & 1'\otimes(1\otimes X) \arrow[r, "\lambda'_{1\otimes X}"] \arrow[d, "1'\otimes\lambda_X"'] & 1\otimes X \arrow[d, "\lambda_X"] \\ & 1'\otimes X \arrow[r, "\lambda'_X"] & X, \end{tikzcd}$$

the square commutes by naturality of λ', the left triangle commutes by triangle identity, and the upper triangle commutes by lemma XV.3.1.9. Hence, $\lambda_X \circ (\lambda'_1 \otimes X) = \lambda'_X \circ (\rho_{1'} \otimes X)$. Thus,

$$\lambda_X \circ (\sigma \otimes X) = \lambda_X \circ (\lambda'_1 \otimes X) \circ (\rho_{1'}^{-1} \otimes X) = \lambda'_X.$$

A similar argument shows the commutativity of the other triangle. ◻

The following explains the name 'monoidal categories'

***Proposition* XV.3.1.11.** — *The monoid of endomorphisms $\mathrm{Hom}_{\mathcal{C}}(1, 1)$ on 1 w. r. t. composition is commutative.*

Proof. — The endomorphisms $\mathrm{Hom}_{\mathcal{C}}(1, 1)$ form a monoid w. r. t. $\circ$ with unit id_1. By naturality of ρ and λ, we have $f \otimes \mathrm{id}_1 = \rho_1^{-1} \circ f \circ \rho_1$ and $\mathrm{id}_1 \otimes g = \lambda_1^{-1} \circ g \circ \lambda_1$. By lemma XV.3.1.9, write $\sigma = \lambda_1 = \rho_1$. Bifunctoriality of $\otimes$ then implies

$$\sigma^{-1} \circ (f \circ g) \circ \sigma = (f \otimes \mathrm{id}_1) \circ (\mathrm{id}_1 \otimes g) = f \otimes g = (\mathrm{id}_1 \otimes g) \circ (f \otimes \mathrm{id}_1) = \sigma^{-1} \circ (g \circ f) \circ \sigma.$$

Thus, $f \circ g = g \circ f$. ◻

XV.3.2. Strictification and coherence theorem.

Morally, it should not matter how a tensor product $X_1 \otimes \cdots \otimes X_n$ is bracketed, and in which way a unit 1 in such a tensor product is absorbed. This is the content of *coherence theorems*.

Definition* XV.3.2.1.** — *Let I be any category. The* ***free strict monoidal category *of I is $\mathcal{S}(I) := \coprod_{n\geq 0} I^n$. Explicitly, the objects of $\mathcal{S}(I)$ are finite sequences $(x_1, \dots, x_n)$, and morphisms $f\colon (x_1, \dots, x_n) \to (y_1, \dots, y_m)$ only exist if both sequences have the same length $n = m$, in which case $f = (f_1, \dots, f_n)$ for morphisms $f_i\colon x_i \to y_i$ in I. This turns into a monoidal category by defining $(x_1, \dots, x_n) \otimes (y_1, \dots, y_p) := (x_1, \dots, x_n, y_1, \dots, y_p)$ as concatenation and the empty sequence $\varnothing$ as the unit object.*

***Lemma* XV.3.2.2** (universal property of free strict monoidal categories). — *Let $F\colon I \to \mathcal{C}$ be a diagram in a monoidal category $\mathcal{C}$. Then there exists a monoidal functor $\phi\colon \mathcal{S}(I) \to \mathcal{C}$ such that $\phi \circ \iota \cong F$ where $\iota\colon I \hookrightarrow \mathcal{S}(I)$ is the canonical embedding. Moreover, ϕ is unique up to unique monoidal natural isomorphism.*

Proof. — By induction on n, define $\phi(\varnothing) = 1$ and $\phi(i_1, \dots, i_n) = \phi(i_1, \dots, i_{n-1}) \otimes F(i_n)$. Then $\eta = \mathrm{id}_1$. By induction on m, define $\mu_{(i_1,\dots,i_n),(j_1,\dots,j_m)}$ by

$$\begin{aligned}\phi((i_1, \dots, i_n) \otimes (j_1, \dots, j_m)) &= \phi(i_1, \dots, i_n, j_1, \dots, j_m) = \phi(i_1, \dots, i_n, j_1, \dots, j_{m-1}) \otimes F(j_m)\\ &= \phi((i_1, \dots, i_n) \otimes (j_1, \dots, j_{m-1})) \otimes F(j_m) \cong (\phi(i_1, \dots, i_n) \otimes \phi(j_1, \dots, j_{m-1})) \otimes F(j_m)\\ &\cong \phi(i_1, \dots, i_n) \otimes (\phi(j_1, \dots, j_{m-1}) \otimes F(j_m)) = \phi(i_1, \dots, i_n) \otimes \phi(j_1, \dots, j_m).\end{aligned}$$

One can now check that ϕ is monoidal and that it satisfies $\phi \circ \iota = F$.

For uniqueness, let $\phi'\colon \mathcal{S}(I) \to \mathcal{C}$ be another such functor. Then a natural isomorphism $\theta\colon \phi \to \phi'$ is defined by $\theta_i\colon \phi(i) \cong F(i) \cong \phi'(i)$ for all $i \in \mathrm{Ob}(I)$ and then inductively defining $\theta_{(i_1,\dots,i_n)}$ by

$$\phi(i_1, \dots, i_n) \cong \phi(i_1, \dots, i_{n-1}) \otimes \phi(i_n) \cong \phi'(i_1, \dots, i_{n-1}) \otimes \phi'(i_n) \cong \phi'(i_1, \dots, i_{n-1}).$$

Note that θ is unique such that it is compatible with the isomorphisms $\phi \circ \iota \cong F \cong \phi' \circ \iota$. One can now check that θ is monoidal. ◻

***Observation* XV.3.2.3.** — By lemma XV.3.2.2, we can define the tensor product

$$X_1 \otimes \cdots \otimes X_n := (\cdots((X_1 \otimes X_2) \otimes X_3) \otimes \cdots) \otimes X_n$$

for $X_1, \dots, X_n \in \mathrm{Ob}(\mathcal{C})$, which is independent of the bracketing up to isomorphism. Indeed, e. g., for $X_1 \otimes X_2 \otimes X_3 \otimes X_4$, consider the discrete diagram $F\colon I = \{i_1, i_2, i_3, i_4\} \to \mathcal{C}$, $F(i_j) = X_j$. Then

$$\begin{aligned}X_1 \otimes (X_2 \otimes (X_3 \otimes X_4)) &\cong \phi(i_1) \otimes (\phi(i_2) \otimes (\phi(i_3) \otimes \phi(i_4))) \cong \phi(i_1) \otimes (\phi(i_2) \otimes \phi(i_3, i_4))\\ &\cong \phi(i_1) \otimes \phi(i_2, i_3, i_4) \cong \phi(i_1, i_2, i_3, i_4),\end{aligned}$$

and this holds for any bracketing of $X_1 \otimes X_2 \otimes X_3 \otimes X_4$. The same argument works for tensor products of morphisms. A similar argument works for cancelling the unit object in tensor products, i. e.,

$$X_1 \otimes \cdots \otimes X_{n-1} \otimes 1 \otimes X_n \otimes \cdots \otimes X_{n+m} \cong X_1 \otimes \cdots \otimes X_{n-1} \otimes X_n \otimes \cdots \otimes X_{n+m}.$$

This only gives some isomorphism between objects induced by ϕ, but it does not say that any isomorphism between those objects constructed from the structure maps of the monoidal category are the same.

***Construction* XV.3.2.4.** — Let $\mathcal{C}$ be a monoidal category. The **strictifiction** of $\mathcal{C}$ is a category $\mathrm{str}(\mathcal{C})$ of finite sequences $(X_1, \dots, X_n)$ of objects in $\mathcal{C}$ with morphisms

$$\mathrm{Hom}_{\mathrm{str}(\mathcal{C})}((X_1, \dots, X_n), (Y_1, \dots, Y_m)) = \mathrm{Hom}_{\mathcal{C}}(F(X_1, \dots, X_n), F(Y_1, \dots, Y_m))$$

where

$$F(X_1, \dots, X_n) := (\cdots(X_1 \otimes X_2) \otimes \cdots) \otimes X_n.$$

Note that $F(\varnothing) = 1$ for the empty sequence.

We want to make $\mathrm{str}(\mathcal{C})$ strict monoidal. To this end, define

$$(X_1, \dots, X_n) \otimes (Y_1, \dots, Y_m) := (X_1, \dots, X_n, Y_1, \dots, Y_m).$$

Then $\varnothing$ is the unit object.

Then $F(X_1, \dots, X_n) = F(X_1, \dots, X_{n-1}) \otimes X_n$. Further define isomorphisms $\mu_{S,T} \colon F(S \otimes T) \xrightarrow{\sim} F(S) \otimes F(T)$ for $S, T \in \mathrm{Ob}(\mathrm{str}(\mathcal{C}))$ by $\mu_{\varnothing,S} = \lambda^{-1}_{F(S)}$, $\mu_{S,\varnothing} = \rho^{-1}_{F(S)}$ for empty sequences, and inductively by

$$\mu_{S,(X)} = \mathrm{id}_{F(S \otimes (X))}, \qquad \mu_{S,T\otimes(X)} = a_{F(S),F(T),X} \circ (\mu_{S,T} \otimes X)$$

for $X \in \mathrm{Ob}(\mathcal{C})$. Then for morphisms $f \colon S \to T$ and $f' \colon S' \to T'$ in $\mathrm{str}(\mathcal{C})$, which correspond to morphisms $\tilde{f} \colon F(S) \to F(T)$ and $\tilde{f}' \colon F(S') \to F(T')$ in $\mathcal{C}$, define $f \otimes g$ in $\mathrm{str}(\mathcal{C})$ as the dashed arrow such that the square

$$\begin{tikzcd}
F(S \otimes S') \arrow[r, "\mu_{S,S'}"] \arrow[d, dashed, "f\otimes f'"'] & F(S) \otimes F(S') \arrow[d, "\tilde{f}\otimes\tilde{f}'"] \\
F(T \otimes T') \arrow[r, "\mu_{T,T'}"] & F(T) \otimes F(T')
\end{tikzcd}$$

commutes. This defines the tensor product on morphisms in $\mathrm{str}(\mathcal{C})$ as well.

***Lemma* XV.3.2.5.** — *Let $\mathcal{C}$ be a monoidal category. Then $F \colon \mathrm{str}(\mathcal{C}) \to \mathcal{C}$ with the natural isomorphism μ from construction XV.3.2.4 and the identity $\eta = \mathrm{id}_{F(\varnothing)}$ is a monoidal functor.*

Proof. — The map F extends to morphisms by being the identity on morphisms of $\mathrm{str}(\mathcal{C})$. By definition, η is a natural isomorphism. Since $\mu_{\varnothing,S} = \lambda^{-1}_{F(S)}$ and $\mu_{S,\varnothing} = \rho^{-1}_{F(S)}$, unitality is clear.

We prove associativity by induction. Let $S, T, U \in \mathrm{Ob}(\mathrm{str}(\mathcal{C}))$. For $U = \varnothing$, consider the diagram

$$\begin{tikzcd}
F(S \otimes T) \arrow[r, equal, "F(a_{S,T,\varnothing})"] \arrow[d, "\rho^{-1}_{(S\otimes T)}"'] & F(S \otimes T) \arrow[d, "\mu_{S,T}"] \\
F(S \otimes T) \otimes 1 \arrow[ur, "\rho^{-1}_{F(S\otimes T)}"'] \arrow[d, "\mu_{S,T}\otimes 1"'] & F(S) \otimes F(T) \arrow[d, "F(S)\otimes\rho^{-1}_{F(T)}"] \\
(F(S) \otimes F(T)) \otimes 1 \arrow[ur, "\rho^{-1}_{F(S)\otimes F(T)}"] \arrow[r, "a_{F(S),F(T),1}"'] & F(S) \otimes (F(T) \otimes 1).
\end{tikzcd}$$

The lower triangle commutes by lemma XV.3.1.9 and the parallelogram commutes commute by naturality of ρ. Hence, the outer associativity diagram commutes.

Now, suppose that the associativity diagram commutes for S, T, U, and we want to show it for $S, T, U \otimes (X)$. For simplicity, we will drop all subscripts of associators. Consider the diagram

$$\begin{tikzcd}
 & F(S \otimes T \otimes U) \otimes X \arrow[r, equal, "F(a)\otimes X"] \arrow[dl, "\mu_{S\otimes T,U\otimes(X)}"'] \arrow[d, "\mu_{S\otimes T,U}\otimes X"] & F(S \otimes T \otimes U) \otimes X \arrow[ddd, "\mu_{S,T\otimes U}\otimes X"] \\
F(S \otimes T) \otimes (F(U) \otimes X) \arrow[d, "\mu_{S,T}\otimes(F(U)\otimes X)"'] & (F(S \otimes T) \otimes F(U)) \otimes X \arrow[l, "a"'] \arrow[d, "(\mu_{S,T}\otimes F(U))\otimes X"] & \\
(F(S) \otimes F(T)) \otimes (F(U) \otimes X) \arrow[dddr, bend right, "a"'] & ((F(S) \otimes F(T)) \otimes F(U)) \otimes X \arrow[l, "a"'] \arrow[d, "a"] & \\
 & (F(S) \otimes (F(T) \otimes F(U))) \otimes X \arrow[d, "a"] & (F(S) \otimes F(T \otimes U)) \otimes X \arrow[l, "(F(S)\otimes\mu_{T,U})\otimes X"'] \arrow[d, "a"] \\
 & F(S) \otimes ((F(T) \otimes F(U)) \otimes X) \arrow[d, "a"] & F(S) \otimes (F(T \otimes U) \otimes X) \arrow[l, "F(S)\otimes(\mu_{T,U}\otimes X)"'] \arrow[dl, "F(S)\otimes\mu_{T,U\otimes(X)}"] \\
 & F(S) \otimes (F(T) \otimes (F(U) \otimes X)). &
\end{tikzcd}$$

The upper left and lower right triangles commute by definition of μ. The upper left and lower right squares commute by naturality of a. The bottom left pentagon is the pentagon identity, and the upper right hexagon is the induction hypothesis. Finally, using that $a \circ (\mu_{S,T\otimes U} \otimes X) = \mu_{S,T\otimes U\otimes(X)}$ once again, we see that the outer diagram commutes, as desired. ◻

***Strictification theorem* XV.3.2.6** (MacLane 1963)**.** — *Any monoidal category is monoidally equivalent to a strict one.*

Proof. — We show that the functor $F\colon \mathrm{str}(\mathcal{C}) \to \mathcal{C}$ from construction XV.3.2.4 is a monoidal equivalence of categories. Note that F is monoidal by lemma XV.3.2.5. Since any morphism $f\colon S \to T$ in $\mathrm{str}(\mathcal{C})$ is $F(f) = f\colon F(S) \to F(T)$ in $\mathcal{C}$, we see that F is fully faithful. For any $X \in \mathrm{Ob}(\mathcal{C})$, we have $F(X) = X$, so F is essentially surjective. From proposition XV.1.5.11, it follows that F is an equivalence of categories. By lemma XV.3.1.8, F is a monoidal equivalence.

We explicitly construct the monoidal quasi-inverse and the monoidal natural isomorphism. The functor $G\colon \mathcal{C} \to \mathrm{str}(\mathcal{C})$, $X \mapsto (X)$ is the quasi-inverse since $F \circ G = \mathrm{id}_{\mathcal{C}}$ and there is a natural isomorphism $\theta\colon G \circ F \rightsquigarrow \mathrm{id}_{\mathrm{str}(\mathcal{C})}$ where $\theta_S\colon G(F(S)) \to S$ corresponds to $\mathrm{id}_{F(S)}\colon F(G(F(S))) = F(S) \to F(S)$. Using θ, we obtain isomorphisms $\theta_\varnothing\colon G(1) \rightsquigarrow \varnothing$ and $\theta_{(X,Y)}\colon G(X \otimes Y) \rightsquigarrow (X,Y)$. By applying F to the associativity and unitality diagrams of G, they commute in $\mathcal{C}$ (all multiplication morphisms μ are identities), so the original diagrams commute in $\mathrm{str}(\mathcal{C})$ by definition. Hence, G is monoidal. Moreover, θ is obviously monoidal because $\mu^{G\circ F}_{X,Y} = \mathrm{id}_{X\otimes Y}$. Thus, F is a monoidal equivalence. ◻

Maybe add alternative proof [Eti+15, Thm. 2.8.5].

We obtain the following convenient result as a corollary.

***Coherence theorem* XV.3.2.7** (MacLane 1963)**.** — *Let $X_1, \dots, X_n$ be objects of a monoidal category $\mathcal{C}$. Let P_1 and P_2 be two expressions of tensor products of $X_1, \dots, X_n$ (in this order) which are parenthesised arbitrarily with arbitrary insertions of tensor products with 1. Let $f, g\colon P_1 \rightrightarrows P_2$ be two isomorphisms, obtained from composing associators, unitors and their inverses possibly tensored with identity morphisms, that transform P_1 into P_2. Then $f = g$.*

In other words, any 'formal' diagram built from a, λ, ρ, a^{-1}, λ^{-1}, ρ^{-1} and identity morphisms by composing and tensoring commutes.

Proof. — By the Strictification theorem XV.3.2.6, let $F\colon \mathcal{C} \rightsquigarrow \mathrm{str}(\mathcal{C})$ be a monoidal equivalence of $\mathcal{C}$ to its strictification $\mathrm{str}(\mathcal{C})$. We want to show that the diagram $f, g\colon P_1 \rightrightarrows P_2$ commutes. Applying F to this diagram and construct over each associator morphism and unitor morphism a rectangle as in diagrams (XV.3.1.5.1) and (XV.3.1.5.2). This yields a huge prism in $\mathrm{str}(\mathcal{C})$ whose top face commutes as the morphisms are all identity morphisms and whose rectangular sides commute. Hence, the bottom face $F(f), F(g)\colon F(P_1) \rightrightarrows F(P_2)$ commutes, and thus $f, g\colon P_1 \rightrightarrows P_2$ commutes as well. ◻

XV.3.3. Braidings.

Braidings capture the notion of commutativity.

Definition* XV.3.3.1.** — *Let $\mathcal{C}$ be a monoidal category. A* ***braiding *on $\mathcal{C}$ is a natural isomorphism B of the form $B_{X,Y}\colon X \otimes Y \rightsquigarrow Y \otimes X$ for all $X, Y \in \mathrm{Ob}(\mathcal{C})$ such that the* ***hexagon identities***

$$\begin{tikzcd}
(X \otimes Y) \otimes Z \arrow[r, "a_{X,Y,Z}"] \arrow[d, "B_{X,Y\otimes Z}"'] & X \otimes (Y \otimes Z) \arrow[r, "B_{X,Y\otimes Z}"] & (Y \otimes Z) \otimes X \arrow[d, "a_{Y,Z,X}"] \\
(Y \otimes X) \otimes Z \arrow[r, "a_{Y,X,Z}"] & Y \otimes (X \otimes Z) \arrow[r, "Y\otimes B_{Z,X}"] & Y \otimes (Z \otimes X),
\end{tikzcd}$$

$$\begin{tikzcd}
X \otimes (Y \otimes Z) \arrow[r, "a^{-1}_{X,Y,Z}"] \arrow[d, "X\otimes B_{Y,Z}"'] & (X \otimes Y) \otimes Z \arrow[r, "B_{X\otimes Y,Z}"] & Z \otimes (X \otimes Y) \arrow[d, "a^{-1}_{Z,X,Y}"] \\
X \otimes (Z \otimes Y) \arrow[r, "a^{-1}_{X,Z,Y}"] & (X \otimes Z) \otimes Y \arrow[r, "B_{X,Z}\otimes Y"] & (Z \otimes X) \otimes Y
\end{tikzcd}$$

commute for all $X, Y, Z \in \mathrm{Ob}(\mathcal{C})$. A monoidal category $\mathcal{C}$ is ***braided*** *if it admits a braiding.*

A braided monoidal category $\mathcal{C}$ whose braiding satisfies $B_{Y,X} \circ B_{X,Y} = \mathrm{id}_{X\otimes Y}$ for all $X, Y \in \mathrm{Ob}(\mathcal{C})$ is ***symmetric****.*

***Remark* XV.3.3.2.** — Note that if B is a braiding on $\mathcal{C}$, then B^{-1} defined by $B^{-1}_{X,Y} : Y \otimes X \xrightarrow{\sim} X \otimes Y$ is also a braiding on $\mathcal{C}$.

***Example* XV.3.3.3.** — The categories Set, k-Vect, A-Mod and G-Rep/k are symmetric monoidal categories.

The coherence conditions of braidings and unitors is automatic.

***Lemma* XV.3.3.4.** — *Let $\mathcal{C}$ be e braided monoidal category. Then the diagrams*

$$\begin{tikzcd}
1 \otimes X \arrow[rr, "B_{1,X}"] \arrow[dr, "\lambda_X"'] & & X \otimes 1 \arrow[dl, "\rho_X"] \\
& X, &
\end{tikzcd}
\qquad
\begin{tikzcd}
X \otimes 1 \arrow[rr, "B_{X,1}"] \arrow[dr, "\rho_X"'] & & 1 \otimes X \arrow[dl, "\lambda_X"] \\
& X &
\end{tikzcd}$$

commute for all $X \in \mathrm{Ob}(\mathcal{C})$. In particular, $B_{1,X} = B^{-1}_{X,1}$.

Proof. — Consider the diagram

$$\begin{tikzcd}
(X \otimes 1) \otimes 1 \arrow[r, "a_{X,1,1}"] \arrow[dd, "B_{X,1}\otimes 1"'] \arrow[dr, "\rho_X\otimes 1"] & X \otimes (1 \otimes 1) \arrow[r, "B_{X,1\otimes 1}"] \arrow[d, "X\otimes\lambda_1"] & (1 \otimes 1) \otimes X \arrow[d, "\lambda_1\otimes X"'] \arrow[dd, bend left, "a_{1,1,X}"] \\
& X \otimes 1 \arrow[r, "B_{X,1}"] & 1 \otimes X \\
(1 \otimes X) \otimes 1 \arrow[r, "a_{1,X,1}"] \arrow[ur, "\lambda_X\otimes 1"] & 1 \otimes (X \otimes 1) \arrow[r, "1\otimes B_{X,1}"] \arrow[u, "\lambda_{X\otimes 1}"'] & 1 \otimes (1 \otimes X). \arrow[u, "\lambda_{1\otimes X}"]
\end{tikzcd}$$

The outer hexagon is the hexagon identity. The right and lower middle triangles commute by lemma XV.3.1.9, the upper middle triangle is the triangle identity. The upper square commutes by naturality of B, and the lower square by naturality of λ. Hence, the left square commutes. Since $- \otimes 1$ is an equivalence, we obtain the desired commutative triangle.

The second triangle is proved similarly, and $B_{1,X} = B^{-1}_{X,1}$ follows from the two triangles. ∎

***Yang-Baxter equation* XV.3.3.5.** — *Let be a braided monoidal category $\mathcal{C}$. Then the diagram*

$$\begin{tikzcd}
(X \otimes Y) \otimes Z \arrow[r, "a_{X,Y,Z}"] \arrow[d, "B_{X,Y}\otimes Z"'] \arrow[dddr, dashed, "B_{X\otimes Y,Z}"] & X \otimes (Y \otimes Z) \arrow[r, "X\otimes B_{Y,Z}"] & X \otimes (Z \otimes Y) \arrow[r, "a^{-1}_{X,Z,Y}"] & (X \otimes Z) \otimes Y \arrow[d, "B_{X,Z}\otimes Y"] \\
(Y \otimes X) \otimes Z \arrow[d, "a_{Y,X,Z}"'] \arrow[dddr, dashed, "B_{Y\otimes X,Z}"] & & & (Z \otimes X) \otimes Y \arrow[d, "a_{Z,X,Y}"] \\
Y \otimes (X \otimes Z) \arrow[d, "Y\otimes B_{X,Z}"'] & & & Z \otimes (X \otimes Y) \arrow[d, "Z\otimes B_{X,Y}"] \\
Y \otimes (Z \otimes X) \arrow[r, "a^{-1}_{Y,Z,X}"] & (Y \otimes Z) \otimes X \arrow[r, "B_{Y,Z}\otimes X"] & (Z \otimes Y) \otimes X \arrow[r, "a_{Z,Y,X}"] & Z \otimes (Y \otimes X)
\end{tikzcd}$$

commutes for all $X, Y, Z \in \mathrm{Ob}(\mathcal{C})$. In particular, if $\mathcal{C}$ is strict, then the Yang-Baxter equation

$$(B_{Y,Z} \otimes X) \circ (Y \otimes B_{X,Z}) \circ (B_{X,Y} \otimes Z) = (Z \otimes B_{X,Y}) \circ (B_{X,Z} \otimes Y) \circ (X \otimes B_{Y,Z})$$

is always satisfied.

Proof. — Consider the stated diagram with the dashed morphisms. Then the upper and lower hexagons are the hexagon identities. The middle parallelogram follows from naturality of B. Hence, the outer diagram commutes. ◻

XV.3.4. Duality.

Duality is reminiscent of adjoint functors.

***Definition* XV.3.4.1.** — *Let $\mathcal{C}$ be a monoidal category. A pair of objects (X, Y) are a **dual pair** if there exist morphisms $\eta\colon 1 \to Y \otimes X$ and $\varepsilon\colon X \otimes Y \to 1$ such that the diagrams*

$$\begin{tikzcd}
1 \otimes X \arrow[r, "\lambda_X"] & X & X \otimes 1 \arrow[l, "\rho_X"'] \arrow[d, "X\otimes\eta"] \\
(X \otimes Y) \otimes X \arrow[u, "\varepsilon\otimes X"] \arrow[rr, "a_{X,Y,X}"] & & X \otimes (Y \otimes X),
\end{tikzcd}
\qquad
\begin{tikzcd}
1 \otimes Y \arrow[r, "\lambda_Y"] \arrow[d, "\eta\otimes Y"'] & Y & Y \otimes 1 \arrow[l, "\rho_Y"'] \\
(Y \otimes X) \otimes Y \arrow[rr, "a_{Y,X,Y}"] & & Y \otimes (X \otimes Y) \arrow[u, "Y\otimes\varepsilon"']
\end{tikzcd}$$

*commute. We write $X^* := Y$ and say that X^* is a **right dual** of X and that X is a **left dual** of X^*.*

***Lemma* XV.3.4.2.** — *Suppose that $X \in \mathrm{Ob}(\mathcal{C})$ has a right dual $X^* \in \mathrm{Ob}(\mathcal{C})$. Then we have adjunctions $(- \otimes X^*) \dashv (- \otimes X)$ and $(X \otimes -) \dashv (X^* \otimes -)$.*

Proof. — We only prove the first adjunction; the second is similar. For any $Y, Z \in \mathrm{Ob}(\mathcal{C})$, define

$$\begin{aligned}
\mathrm{Hom}_{\mathcal{C}}(Y \otimes X^*, Z) &\longrightarrow \mathrm{Hom}_{\mathcal{C}}(Y, Z \otimes X), & u &\longmapsto (Y \cong Y \otimes 1 \xrightarrow{Y\otimes\eta} Y \otimes X^* \otimes X \xrightarrow{v\otimes X} Z \otimes X), \\
\mathrm{Hom}_{\mathcal{C}}(Y, Z \otimes X) &\longrightarrow \mathrm{Hom}_{\mathcal{C}}(Y \otimes X^*, Z), & v &\longmapsto (Y \otimes X^* \xrightarrow{v\otimes X^*} Z \otimes X \otimes X^* \xrightarrow{Z\otimes\varepsilon} Z \otimes 1 \cong Z).
\end{aligned}$$

One can check that these two maps are mutually inverse. ◻

***Fact* XV.3.4.3** (duals are essentially unique)**.** — *Let (Y, η, ε) and $(Y', \eta', \varepsilon')$ be two right duals of X. Then there exists a unique isomorphism $f\colon Y \xrightarrow{\sim} Y'$ such that the diagrams*

$$\begin{tikzcd}
& 1 \arrow[dl, "\eta"'] \arrow[dr, "\eta'"] & \\
Y \otimes X \arrow[rr, "f\otimes X"] & & Y' \otimes X,
\end{tikzcd}
\qquad
\begin{tikzcd}
X \otimes Y \arrow[rr, "X\otimes f"] \arrow[dr, "\varepsilon"'] & & X \otimes Y' \arrow[dl, "\varepsilon'"] \\
& 1 &
\end{tikzcd}$$

commute.

The dual statement on right duals also holds true.

Proof. — In the following, we suppress all associators and unitors. Suppose that such an isomorphism $f\colon Y \xrightarrow{\sim} Y'$ exists. Since Y is a right dual of X, we have $f = f \circ (Y \otimes \varepsilon) \circ (\eta \otimes Y)$. Note that by bifunctoriality of $\otimes$, we have

$$\begin{tikzcd}
Y \otimes X \otimes Y \arrow[r, "Y\otimes\varepsilon"] \arrow[d, "f\otimes X\otimes Y"'] & Y \arrow[d, "f"] \\
Y' \otimes X \otimes Y \arrow[r, "Y'\otimes\varepsilon"] & Y'.
\end{tikzcd}$$

Hence, $f = (Y' \otimes \varepsilon) \circ (f \otimes X \otimes Y) \circ (\eta \otimes Y)$. Applying $- \otimes Y$ to the triangle $\eta' = (f \otimes X) \circ \eta$, we see that $f = (Y' \otimes \varepsilon) \circ (\eta' \otimes Y)$.

Thus, define

$$f\colon Y \xrightarrow{\eta' \otimes Y} Y' \otimes X \otimes Y \xrightarrow{Y' \otimes \varepsilon} Y', \qquad g\colon Y' \xrightarrow{\eta \otimes Y'} Y \otimes X \otimes Y' \xrightarrow{Y \otimes \varepsilon'} Y.$$

Our aim is to show that f and g are mutually inverse. Consider the diagram

$$\begin{tikzcd}
Y \arrow[r, "\eta \otimes Y"] \arrow[d, "\eta' \otimes Y"'] & Y \otimes X \otimes Y \arrow[d, "Y \otimes X \otimes \eta' \otimes Y"'] \arrow[dr, equal] & \\
Y' \otimes X \otimes Y \arrow[r, "\eta \otimes Y' \otimes X \otimes Y"'] \arrow[d, "Y' \otimes \varepsilon"'] & Y \otimes X \otimes Y' \otimes X \otimes Y \arrow[r, "Y \otimes \varepsilon' \otimes X \otimes Y"'] \arrow[d, "Y \otimes X \otimes Y' \otimes \varepsilon"'] & Y \otimes X \otimes Y \arrow[d, "Y \otimes \varepsilon"] \\
Y' \arrow[r, "\eta \otimes Y'"] & Y \otimes X \otimes Y' \arrow[r, "Y \otimes \varepsilon'"] & Y.
\end{tikzcd}$$

The top right triangle commutes since Y' is a right dual of X. The three squares all commute by the bifunctoriality of $\otimes$, e. g., the top left square is

$$\begin{tikzcd}
1 \otimes 1 \otimes Y \arrow[r, "\eta \otimes 1 \otimes Y"] \arrow[d, "1 \otimes \eta' \otimes Y"'] & Y \otimes X \otimes 1 \otimes Y \arrow[d, "Y \otimes X \otimes \eta' \otimes Y"] \\
1 \otimes Y' \otimes X \otimes Y \arrow[r, "\eta \otimes Y' \otimes X \otimes Y"] & Y \otimes X \otimes Y' \otimes X \otimes Y.
\end{tikzcd}$$

Hence, the outer diagram commutes. The top composition is $(Y \otimes \varepsilon) \circ (\eta \otimes Y) = \mathrm{id}_Y$ since Y is a right dual of X. The bottom composition is $g \circ f = \mathrm{id}_Y$. One similarly shows that $f \circ g = \mathrm{id}_{Y'}$. ∎

Alternative proof. — For any right dual X^* of X, lemma XV.3.4.2 shows that there is a bijection $\mathrm{Hom}_{\mathcal{C}}(X^*, Z) \cong \mathrm{Hom}_{\mathcal{C}}(1, X \otimes Z)$ that is natural in Z. By lemma XV.1.6.9, there is a unique (canonical) isomorphism $Y \cong X^* \cong Y'$. ∎

***Construction* XV.3.4.4.** — Let $\mathcal{C}$ be a monoidal category. Let $f\colon X \to Y$ be a morphism, and suppose that X and Y have right duals. Then the **right dual** of f is defined as

$$f^*\colon Y^* \xrightarrow{\eta_X \otimes Y^*} X^* \otimes X \otimes Y^* \xrightarrow{X^* \otimes f \otimes Y^*} X^* \otimes Y \otimes Y^* \xrightarrow{X^* \otimes \varepsilon_Y} X^*.$$

Similarly, for a morphism $f^*\colon X^* \to Y^*$, the **left dual** of f is defined as

$$f\colon Y \xrightarrow{Y \otimes \eta_X^*} Y \otimes X^* \otimes X \xrightarrow{Y \otimes f^* \otimes X} Y \otimes Y^* \otimes X \xrightarrow{\varepsilon_Y^* \otimes X} X.$$

If $\mathcal{C}$ admits all right duals, then right duality defines a functor $(-)^*\colon \mathcal{C}^{\mathrm{op}} \to \mathcal{C}^r$ where $\mathcal{C}^r$ is the reverse monoidal category. Similarly, if $\mathcal{C}$ admits all left duals, then left duality defines a functor $\mathcal{C}^{\mathrm{op}} \to \mathcal{C}^r$ as well.

***Definition* XV.3.4.5.** — *An object in a monoidal category is **rigid** if it admits left and right duals. A monoidal category is **rigid** if all objects are rigid.*

***Example* XV.3.4.6.** — The category of finite-dimensional k-vector spaces is rigid. In fact, the left and right duals coincide. For a finite-dimensional vector space V, its dual is $V^* = \mathrm{Hom}_k(V, k)$. Then $\varepsilon\colon V \otimes V^* \to k$, $v \otimes f \mapsto f(v)$ is the evaluation map and $\eta\colon k \to V^* \otimes V$, $\lambda \mapsto \lambda \sum_{i=1}^n v_i^* \otimes v$ is the coevaluation map where $\{v_1, \dots, v_n\}$ is a basis of V and $\{v_1^*, \dots, v_n^*\}$ is the dual basis of V^*.

The category k-Vect of all k-vector spaces does not admit all duals, however. Indeed, if a map $\eta\colon k \to V^* \otimes V$ exists, then $V \to V \otimes V^* \otimes V \to V$ is the identity on V, but the image of the first map $V \otimes \eta$ is one-dimensional.

XV.3.5. Categorification of algebraic constructions.

***Definition* XV.3.5.1.** — *Let $\mathcal{T}$ be a monoidal category. An **action** of $\mathcal{T}$ onto some small category $\mathcal{C}$ is a monoidal functor $F\colon \mathcal{T} \to \mathrm{End}(\mathcal{C})$.*

More generally (for not necessarily small categories $\mathcal{C}$), an action *of $\mathcal{T}$ on $\mathcal{C}$ is given by:*

(i) *a bifunctor $-\otimes-\colon \mathcal{T}\times\mathcal{C}\to\mathcal{C}$;*

(ii) *and a natural isomorphism a of the form $a_{X,Y,W}\colon (X\otimes Y)\otimes W \xrightarrow{\sim} X\otimes(Y\otimes W)$ for $X, Y \in \mathrm{Ob}(\mathcal{T})$ and $W \in \mathrm{Ob}(\mathcal{C})$ such that the pentagon identity (diagram (XV.3.1.1.1)) is satisfied;*

(iii) *a natural isomorphism ξ of the form $\xi_W\colon 1\otimes W \xrightarrow{\sim} W$ for all $W\in\mathrm{Ob}(\mathcal{C})$ such that the triangle identities*

$$\begin{tikzcd}
(1\otimes X)\otimes W \arrow[rr, "a_{1,X,W}"] \arrow[dr, "\lambda_X\otimes W"'] & & 1\otimes(X\otimes W) \arrow[dl, "\xi_{X\otimes W}"] \\
& X\otimes W, &
\end{tikzcd}
\qquad
\begin{tikzcd}
(X\otimes 1)\otimes W \arrow[rr, "a_{X,1,W}"] \arrow[dr, "\rho_X\otimes W"'] & & X\otimes(1\otimes W) \arrow[dl, "X\otimes\xi_{1\otimes W}"] \\
& X\otimes W &
\end{tikzcd}$$

commute for any $X\in\mathrm{Ob}(\mathcal{T})$ and any $W\in\mathrm{Ob}(\mathcal{C})$.

The equivalence is given by $X\otimes W := F(X)(W)$, $a_{X,Y,W} := \mu_{X,Y}(W)$ and $\xi_W := \eta_W\colon F(1)(W) \xrightarrow{\sim} \mathrm{id}_{\mathcal{C}}(W)$.

***Example* XV.3.5.2.** — (i) A monoidal category $\mathcal{C}$ acts trivially on itself.

(ii) For a small category $\mathcal{C}$, the monoidal category $\mathrm{End}(\mathcal{C})$ acts on $\mathcal{C}$ via $F\otimes X := F(X)$ for $F\colon\mathcal{C}\to\mathcal{C}$ and $X\in\mathrm{Ob}(\mathcal{C})$.

***Definition* XV.3.5.3.** — *Let $\mathcal{C}$ be a monoidal category. A **monoid object** (or just **monoid**) in $\mathcal{C}$ is an object A together with a multiplication morphism $\mu\colon A\otimes A\to A$ and a unit morphism $\eta\colon 1\to A$ such that the following diagrams commute:*

(i) *associativity:*

$$\begin{tikzcd}
(A\otimes A)\otimes A \arrow[rr, "a_{A,A,A}"] \arrow[d, "\mu\otimes A"'] & & A\otimes(A\otimes A) \arrow[d, "A\otimes\mu"] \\
A\otimes A \arrow[r, "\mu"] & A & A\otimes A, \arrow[l, "\mu"']
\end{tikzcd}$$

(ii) *unitality:*

$$\begin{tikzcd}
1\otimes A \arrow[r, "\eta\otimes A"] \arrow[dr, "\lambda_A"'] & A\otimes A \arrow[d, "\mu"] & A\otimes 1 \arrow[l, "A\otimes\eta"'] \arrow[dl, "\rho_A"] \\
& A. &
\end{tikzcd}$$

*A **morphism of monoids** $f\colon A\to B$ is a morphism such that the diagrams*

$$\begin{tikzcd}
A\otimes A \arrow[r, "f\otimes f"] \arrow[d, "\mu^A"'] & B\otimes B \arrow[d, "\mu^B"] \\
A \arrow[r, "f"] & B,
\end{tikzcd}
\qquad
\begin{tikzcd}
& 1 \arrow[dl, "\eta^A"'] \arrow[dr, "\eta^B"] & \\
A \arrow[rr, "f"] & & B
\end{tikzcd}$$

commute.

***Example* XV.3.5.4.** — (i) For a commutative ring k, a monoid object in k-Mod is an (associative, non-commutative) k-algebra.

(ii) As a special case, a monoid object in the category AbGrp of abelian groups is a ring.

(iii) A monoid object in the cartesian category of monoids is a commutative monoid. Indeed, a monoid object M has a monoid structure $\cdot$ and a monoid structure $\circ$ given by the monoidal category. Since $\circ\colon M \times M \to M$ is a homomorphism of monoids, it must satisfy $(a \cdot b) \circ (c \cdot d) = (a \circ c) \cdot (b \circ d)$. Now, commutativity follows from the Eckmann-Hilton argument.(XV.13)

***Definition* XV.3.5.5.** — *Let $\mathcal{T}$ be a monoidal category acting on a category $\mathcal{C}$. Let (A, μ^A, η^A) be a monoid in $\mathcal{T}$. A **left A-module** in $\mathcal{C}$ is an object M in $\mathcal{C}$ with a scalar multiplication morphism $\mu^M\colon A \otimes M \to M$ such that the following diagrams commute:*

$$\begin{tikzcd} (A\otimes A)\otimes M \arrow[rr, "a_{A,A,M}"] \arrow[d, "\mu^A\otimes M"'] & & A\otimes(A\otimes M) \arrow[d, "A\otimes\mu^M"] \\ A\otimes M \arrow[r, "\mu^M"] & M & A\otimes M, \arrow[l, "\mu^M"'] \end{tikzcd} \qquad \begin{tikzcd} 1\otimes M \arrow[rr, "\eta^A\otimes M"] \arrow[dr, "\xi_M"'] & & A\otimes M \arrow[dl, "\mu^M"] \\ & M. & \end{tikzcd}$$

*A **morphism of left A-modules** over A is a morphism $f\colon M \to N$ of left A-modules such that the diagram*

$$\begin{tikzcd} A\otimes M \arrow[r, "A\otimes f"] \arrow[d, "\mu^M"'] & A\otimes N \arrow[d, "\mu^N"] \\ M \arrow[r, "f"] & N \end{tikzcd}$$

commutes. This defines a category $A\text{-}\mathsf{LMod}(\mathcal{C})$ of left A-modules in $\mathcal{C}$.

*We similarly define a **right A-module** and **morphisms of right A-modules**, giving the category $A\text{-}\mathsf{RMod}_{\mathcal{C}}$ of right A-modules in $\mathcal{C}$.*

XV.3.6. Monads.

***Definition* XV.3.6.1.** — *Let $\mathcal{C}$ be a category. A **monad** in $\mathcal{C}$ is monoid object in the (strict) monoidal category* $\mathrm{End}(\mathcal{C})$ *of endofunctors, provided that $\mathcal{C}$ is small.*

*More generally, a **monad** in $\mathcal{C}$ is an endofunctor $T\colon \mathcal{C} \to \mathcal{C}$ together with a multiplication natural transformation $\mu\colon T \to T$ and a unit natural transformation $\eta\colon \mathrm{id}_{\mathcal{C}} \to T$ such that the following diagrams commute:*

$$\begin{tikzcd} T\circ T\circ T \arrow[r, Rightarrow, "\mu T"] \arrow[d, Rightarrow, "T\mu"'] & T\circ T \arrow[d, Rightarrow, "\mu"] \\ T\circ T \arrow[r, Rightarrow, "\mu"] & T, \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, Rightarrow, "\eta T"] \arrow[dr, Rightarrow, "\mathrm{id}_T"'] & T\circ T \arrow[d, Rightarrow, "\mu"] & T \arrow[l, Rightarrow, "T\eta"'] \arrow[dl, Rightarrow, "\mathrm{id}_T"] \\ & T. & \end{tikzcd}$$

***Definition* XV.3.6.2.** — *Let $\mathcal{C}$ be a category, and let $T\colon \mathcal{C} \to \mathcal{C}$ be a monad on $\mathcal{C}$. By convention, we call a left T-module a **T-algebra**, and the category $T\text{-}\mathsf{Mod}(\mathcal{C})$ of T-algebras is called the **Eilenberg-Moore category**, written* $\mathrm{EM}(T)$. *Explicitly, this means that a T-algebra A satisfies the commutative diagrams*

$$\begin{tikzcd} T(T(A)) \arrow[r, "T(\mu^A)"] \arrow[d, "\mu^T_A"'] & T(A) \arrow[d, "\mu^A"] \\ T(A) \arrow[r, "\mu^A"] & A, \end{tikzcd} \qquad \begin{tikzcd} A \arrow[rr, "\eta^T_A"] \arrow[dr, "\mathrm{id}_A"'] & & T(A) \arrow[dl, "\mu^A"] \\ & A. & \end{tikzcd}$$

(XV.13) The *Eckmann-Hilton argument* goes as follows. Let M be a monoid with two monoid structures $*$ and $\circ$ satisfying $(a*b)\circ(c*d) = (a\circ c)*(b\circ d)$. Then $*\colon M\times M \to M$ is a homomorphism w. r. t. $\circ$ and vice versa. Then $1_\circ = 1_\circ \circ 1_\circ = (1_\circ * 1_*)\circ(1_\circ * 1_*) = (1_\circ \circ 1_*) * (1_\circ \circ 1_*) = 1_* * 1_* = 1_*$ and $a\circ b = (a*1)\circ(b*1) = (a\circ 1)*(b\circ 1) = a*b$. Hence, the monoid structures agree. Finally, $ab = (1a)(b1) = (1b)(a1) = ba$ and $(ab)c = (ab)(1c) = (a1)(bc) = a(bc)$, so M is commutative and associative.

Monads arise often from adjunctions.

***Lemma* XV.3.6.3.** — *Let $L \dashv R$ be an adjoint pair of functors $\mathcal{C} \rightleftarrows \mathcal{D}$, and let $\eta \colon \mathrm{id}_{\mathcal{C}} \colon R \circ L$ and $\varepsilon \colon L \circ R \to \mathrm{id}_{\mathcal{D}}$ be the unit and counit.*

(i) *The endofunctor $T := R \circ L$ together with $\mu^T := R\varepsilon L \colon T \circ T \to T$ and $\eta^T := \eta$ is a monad in $\mathcal{C}$.*

(ii) *For $Y \in \mathrm{Ob}(\mathcal{D})$, the object $R(Y) \in \mathrm{Ob}(\mathcal{C})$ together with $\mu^Y := R(\varepsilon_Y) \colon T(R(Y)) \to R(Y)$ is a T-algebra. Moreover, we obtain the* ***comparison functor*** *$K^T \colon \mathcal{D} \to \mathrm{EM}(T)$, $Y \mapsto (R(Y), \mu^Y)$.*

Proof. — For (i), note that unitality

$$\begin{tikzcd}
R \circ L \arrow[r, Rightarrow, "\eta(R\circ L)"] \arrow[dr, Rightarrow, "\mathrm{id}_R L"'] & (R \circ L) \circ (R \circ L) \arrow[d, Rightarrow, "R\varepsilon L"] & R \circ L \arrow[l, Rightarrow, "(R\circ L)\eta"'] \arrow[dl, Rightarrow, "R\mathrm{id}_L"] \\
& R \circ L &
\end{tikzcd}$$

of T follows from the triangle identities of $L \dashv R$. Regarding associativity, we need to verify that for any $X \in \mathrm{Ob}(\mathcal{C})$ the diagram

$$\begin{tikzcd}
R(S(S(Y))) \arrow[r, "R(\varepsilon_{S(Y)})"] \arrow[d, "R(S(\varepsilon_Y))"'] & R(S(Y)) \arrow[d, "R(\varepsilon_Y)"] \\
R(S(Y)) \arrow[r, "R(\varepsilon_Y)"] & R(Y)
\end{tikzcd} \tag{XV.3.6.3.1}$$

commutes where $\varepsilon \colon S := L \circ R \to \mathrm{id}_{\mathcal{D}}$ and $Y := L(X)$. But this diagram commutes since it is the image of a naturality square of ε under R.

For (ii), recall that $T \otimes R(Y) := T(R(Y)) = R(L(R(Y)))$. Again, write $\varepsilon \colon S := L \circ R \to id_{\mathcal{D}}$. Then the diagram

$$\begin{tikzcd}
R(Y) \arrow[rr, "\eta_{R(Y)}"] \arrow[dr, "\mathrm{id}_{R(Y)}"'] & & R(S(Y)) \arrow[dl, "R(\varepsilon_Y)"] \\
& R(Y) &
\end{tikzcd}$$

is literally the triangle identity of $L \dashv R$, and the other diagram for $R(Y)$ to be a T-algebra is literally diagram (XV.3.6.3.1). Finally, any morphism $f \colon Y \to Z$ in $\mathcal{D}$ induces a morphism $R(f) \colon R(Y) \to R(Z)$ such that the diagram

$$\begin{tikzcd}
R(S(Y)) \arrow[r, "R(S(f))"] \arrow[d, "R(\varepsilon_Y)"'] & R(S(Z)) \arrow[d, "R(\varepsilon_Z)"] \\
R(Y) \arrow[r, "R(f)"] & R(Z),
\end{tikzcd}$$

which is naturality of ε. □

Definition* XV.3.6.4.** — *An adjoint pair $L \dashv R$ of functors $\mathcal{C} \rightleftarrows \mathcal{D}$ is* ***monadic *if the comparison functor $K^{R\circ L} \colon \mathcal{D} \to \mathrm{EM}(R \circ L)$ is an equivalence of categories. A functor $R \colon \mathcal{D} \to \mathcal{C}$ is* ***monadic*** *if it admits a left adjoint $L \colon \mathcal{C} \to \mathcal{D}$ such that $L \dashv R$ is monadic.*

The following gives a converse.

***Lemma* XV.3.6.5.** — *Let (T, μ^T, η^T) be a monad in a category $\mathcal{C}$.*

(i) *For any $X \in \mathrm{Ob}(\mathcal{C})$, the object $T(X)$ with $\mu^T_X \colon T(T(X)) \to T(X)$ is a T-algebra called the* ***free*** *T-algebra of X.*

(ii) *The free functor* $F\colon \mathcal{C} \to \mathrm{EM}(T)$, $X \mapsto (T(X), \mu^T_X)$ *is left adjoint to the forgetful functor* $\mathrm{For}\colon \mathrm{EM}(T) \to \mathcal{C}$.

Proof. — (i) is obvious: evaluate the coherence diagrams of T at X. For (ii), we need to exhibit an adjunction isomorphism

$$\mathrm{Hom}_{\mathrm{EM}(T)}((T(X), \mu^T_X), (Y, \mu^Y)) \cong \mathrm{Hom}_{\mathcal{C}}(X, Y).$$

We check that

$$u \longmapsto (X \xrightarrow{\eta^T_X} T(X) \xrightarrow{u} Y), \qquad (T(X) \xrightarrow{T(v)} T(Y) \xrightarrow{\mu^Y} Y) \longleftarrow v$$

are mutually inverse. Note that the image of v is indeed a morphism of T-algebras because in the diagram

$$\begin{tikzcd}
T(T(X)) \arrow[r, "T(T(v))"] \arrow[d, "\mu^T_X"'] & T(T(Y)) \arrow[r, "T(\mu^Y)"] \arrow[d, "\mu^T_Y"] & T(Y) \arrow[d, "\mu^Y"] \\
T(X) \arrow[r, "T(v)"] & T(Y) \arrow[r, "\mu^Y"] & Y,
\end{tikzcd}$$

the left square commutes by naturality of μ^T and the right square commutes since Y is a T-algebra. Given $v\colon X \to Y$, we have

$$\begin{tikzcd}
X \arrow[r, "v"] \arrow[d, "\eta^T_X"'] & Y \arrow[r, "\mathrm{id}_Y"] \arrow[d, "\eta^T_Y"'] & Y \\
T(X) \arrow[r, "T(v)"] & T(Y) \arrow[ur, "\mu^Y"']
\end{tikzcd}$$

where the square is the naturality square of η^T and where the triangle commutes since Y is a T-algebra. On the other hand, given $u\colon T(X) \to Y$ in $\mathrm{EM}(T)$, we have

$$\begin{tikzcd}
T(X) \arrow[r, "T(\eta^T_X)"] \arrow[dr, "\mathrm{id}_{T(X)}"'] & T(T(X)) \arrow[r, "T(u)"] \arrow[d, "\mu^T_X"] & T(Y) \arrow[d, "\mu^Y"] \\
 & T(X) \arrow[r, "u"] & Y
\end{tikzcd}$$

where the square commutes since u is a T-algebra morphisms and where the triangle commutes since T is a monad. Note that the unit is η^T and the counit $\varepsilon\colon F \circ \mathrm{For} \to \mathrm{id}_{\mathrm{EM}(T)}$ is given by $\varepsilon_{(Y,\mu^Y)} := \mu^Y : (T(Y), \mu^T_Y) \to (Y, \mu^Y)$. ◻

***Remark* XV.3.6.6** (Kleisli category)**.** — There is an alternative construction to lemma XV.3.6.5. Let (T, μ, η) be a monad in $\mathcal{C}$. We define the **Kleisli category** $\mathrm{Kl}(T)$ of T by $\mathrm{Ob}(\mathrm{Kl}(T)) = \mathrm{Ob}(\mathcal{C})$ and $\mathrm{Hom}_{\mathrm{Kl}(T)}(X, Y) = \mathrm{Hom}_{\mathcal{C}}(X, T(Y))$. For $X \in \mathrm{Ob}(\mathrm{Kl}(T))$, the identity is $\mathrm{id}_X = \eta_X$, and composition of **Kleisli morphisms** $f\colon X \to T(Y)$ and $g\colon Y \to T(Z)$ is given by the **Kleisli composition** $\mu_Z \circ T(g) \circ f\colon X \to T(Z)$.

Like lemma XV.3.6.5, there is an adjunction $L \dashv R$ of functors $\mathcal{C} \rightleftarrows \mathrm{Kl}(T)$ as follows, [see Rie16, Lemma 5.2.12]. The functor $L\colon \mathcal{C} \to \mathrm{Kl}(T)$ is the identity on objects and maps $f\colon X \to Y$ to $\eta_Y \circ f\colon X \to T(Y)$. The functor $R\colon \mathrm{Kl}(T) \to \mathcal{C}$ maps $X \mapsto T(X)$ and maps Kleisli morphisms $f\colon X \to T(Y)$ to the Kleisli extension $\mu_Y \circ T(f)\colon T(X) \to T(Y)$.

In fact, both the Eilenberg-Moore and Kleisli categories are universal and in the following sense dual. Let $A \dashv B$ be an adjunction $\mathcal{D} \rightleftarrows \mathcal{C}$ that again give T by lemma XV.2.12.8. Then there exists a unique functor $\mathrm{Kl}(T) \to \mathcal{D}$ such that the diagram

$$\begin{tikzcd}
\mathrm{Kl}(T) \arrow[rr, dashed, "\exists!"] \arrow[dr, shift left, "R"] & & \mathcal{D} \arrow[dl, shift left, "A"] \\
 & \mathcal{C} \arrow[ul, shift left, "L"] \arrow[ur, shift left, "B"] &
\end{tikzcd}$$

commutes (not only quasi-commutes), i. e., the pair $L \dashv R$ is initial in the category of adjunctions $A \dashv B$ giving back T. Similarly, the pair $F \dashv$ For from lemma XV.3.6.5 is terminal. [See Rie16, Proposition 5.2.13]. Thus, we obtain a unique functor $K\colon \mathrm{Kl}(T) \to \mathrm{EM}(T)$. This functor K is fully faithful and its image is the (full) subcategory of all free T-algebras, [see Rie16, Lemma 5.2.14].

XV.3.7. Monadicity theorem.

We want to characterise monadic functors.

***Definition* XV.3.7.1.** — *In a category $\mathcal{C}$, a diagram*

$$\begin{tikzcd} X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \arrow[r, "h"] & Z \end{tikzcd}$$

with $h \circ f = h \circ g$ is a ***split coequaliser*** *if there are sections $s\colon Z \to Y$ of h and $t\colon Y \to X$ of f, i. e., $h \circ s = \mathrm{id}_Z$ and $f \circ t = \mathrm{id}_Y$, with $s \circ h = g \circ t$.*

***Fact* XV.3.7.2.** — *A split coequaliser is a coequaliser diagram. In particular, any functor preserves this colimit.*

Proof. — Consider the diagram

$$\begin{tikzcd} X \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & Y \arrow[l, bend right=60, "t"'] \arrow[r, "h"] \arrow[dr, "k"'] & Z \arrow[l, bend right=60, "s"'] \arrow[d, dashed, "\exists!"] \\ & & W \end{tikzcd}$$

where $X \rightrightarrows Y \to Z$ forms a split coequaliser diagram and where $k \circ f = k \circ g$. Then $k \circ s\colon Z \to W$ for the dashed arrow makes the triangle commute: $k \circ s \circ h = k \circ g \circ t = k \circ f \circ t = k$. The morphism $k \circ s$ is unique since h is a (split) epimorphism. Finally, any functor preserves the relations defining a split coequaliser, so it preserves split coequalisers.

***Beck's monadicity theorem* XV.3.7.3** (1968[XV.14])**.** — *A functor $R\colon \mathcal{D} \to \mathcal{C}$ is monadic if and only if:*

(i) *R admits a left adjoint.*
(ii) *R is conservative.*
(iii) *If for a parallel pair of morphisms (f, g) in $\mathcal{D}$, the pair $(R(f), R(g))$ has a split coequaliser $\mathrm{Coeq}(R(f), R(g))$ in $\mathcal{C}$, then $\mathrm{Coeq}(f, g)$ in $\mathcal{D}$ exists which is preserved by R.*

Proof. — Assume that $R\colon \mathcal{D} \to \mathcal{C}$ is monadic. By definition, it admits a left adjoint $L\colon \mathcal{C} \to \mathcal{D}$ such that $K^{R\circ L}\colon \mathcal{D} \xrightarrow{\sim} \mathrm{EM}(R \circ L)$ is an equivalence of categories. It suffices to show that for any monad T in $\mathcal{C}$, the forgetful functor For: $\mathrm{EM}(T) \to \mathcal{C}$ satisfies all the statements. Indeed, since $K^{R\circ L}$ is conservative and preserves and creates colimits (corollary XV.2.2.13), the properties of For for $T = R \circ L$ transfer to $R = \mathrm{For}^{R\circ L} \circ K^{R\circ L}$.

Finish the proof, see [Rie16]

By lemma XV.3.6.5, For has a left adjoint, and one can easily check that it is conservative.

(XV.14) Often wrongly attributed to Barr and Beck.

SYMBOLS

TERMS

BIBLIOGRAPHY

[Ahl21] Lars Ahlfors. *Complex analysis. An introduction to the theory of analytic functions of one complex variable.* English. 3rd edition, reprint of the 1979 edition. Vol. 385. AMS Chelsea Publ. Providence, RI: American Mathematical Society (AMS), 2021. ISBN: 978-1-4704-6767-8.

[AM16] Michael F. Atiyah and I. G. Macdonald. *Introduction to commutative algebra.* English. Student economy edition. Addison-Wesley Ser. Math. Boulder: Westview Press, 2016. ISBN: 978-0-8133-5018-9; 978-0-8133-4544-4.

[DS05] Fred Diamond and Jerry Shurman. *A first course in modular forms.* English. Vol. 228. Grad. Texts Math. Berlin: Springer, 2005. ISBN: 0-387-23229-X.

[Eti+15] Pavel Etingof, Shlomo Gelaki, Dmitri Nikshych and Victor Ostrik. *Tensor categories.* English. Vol. 205. Math. Surv. Monogr. Providence, RI: American Mathematical Society (AMS), 2015. ISBN: 978-1-4704-2024-6. DOI: 10.1090/surv/205.

[FB09] Eberhard Freitag and Rolf Busam. *Complex analysis. Transl. from the German by Dan Fulea.* English. 2nd ed. Universitext. Berlin: Springer, 2009. ISBN: 978-3-540-93982-5; 978-3-540-93983-2. DOI: 10.1007/978-3-540-93983-2.

[GW20] Ulrich Görtz and Torsten Wedhorn. *Algebraic geometry I. Schemes. With examples and exercises.* English. 2nd edition. Springer Stud. Math. – Master. Wiesbaden: Springer Spektrum, 2020. ISBN: 978-3-658-30732-5; 978-3-658-30733-2. DOI: 10.1007/978-3-658-30733-2.

[Kas95] Christian Kassel. *Quantum groups.* English. Vol. 155. Grad. Texts Math. New York, NY: Springer-Verlag, 1995. ISBN: 0-387-94370-6.

[Kna10] A. W. Knapp. 'Prerequisites for the Langlands program'. English. In: *Automorphic forms and the Langlands program. Selected papers of the conference on Langlands and geometric Langlands program, Guangzhou, China, June 18–21, 2007.* Somerville, MA: International Press; Beijing: Higher Education Press, 2010, pp. 1–9. ISBN: 978-1-57146-141-4.

[KS06] Masaki Kashiwara and Pierre Schapira. *Categories and sheaves.* English. Vol. 332. Grundlehren Math. Wiss. Berlin: Springer, 2006. ISBN: 3-540-27949-0. DOI: 10.1007/3-540-27950-4.

[Mat80] Hideyuki Matsumura. *Commutative algebra. 2nd ed.* English. Vol. 56. Math. Lect. Note Ser. The Benjamin/Cummings Publishing Company, Reading, MA, 1980.

[Mat86] Hideyuki Matsumura. *Commutative ring theory. Transl. from the Japanese by M. Reid.* English. Vol. 8. Camb. Stud. Adv. Math. Cambridge University Press, Cambridge, 1986.

[Neu99] Jürgen Neukirch. *Algebraic number theory. Transl. from the German by Norbert Schappacher.* English. Vol. 322. Grundlehren Math. Wiss. Berlin: Springer, 1999. ISBN: 3-540-65399-6.

[nLa26] nLab authors. *HomePage.* https://ncatlab.org/nlab/show/HomePage. Revision 317. May 2026.

[Rie16] Emily Riehl. *Category theory in context*. English. Aurora Dover Mod. Math Orig. Mineola, NY: Dover Publications, 2016. ISBN: 978-0-486-80903-8.

[SS03] Elias M. Stein and Rami Shakarchi. *Complex analysis*. English. Vol. 2. Princeton Lect. Anal. Princeton, NJ: Princeton University Press, 2003. ISBN: 0-691-11385-8.

[Wei94] Charles A. Weibel. *An introduction to homological algebra*. English. Vol. 38. Camb. Stud. Adv. Math. Cambridge: Cambridge University Press, 1994. ISBN: 0-521-43500-5.

CONTENTS